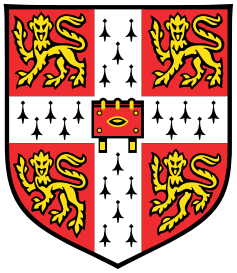
Deep forward and reverse phenotyping for genetic discovery in pulmonary arterial hypertension.



Emilia Maria Świetlik

Department of Medicine
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

Murray Edwards College June 2021

Ad maiorem Dei gloriam. . .

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text. It is not substantially the same as any that I have submitted, or am concurrently submitting, for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or is being concurrently submitted, for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. This dissertation does not exceed the prescribed limit of 60 000 words.

Emilia Maria Świetlik
June 2021

Abstract

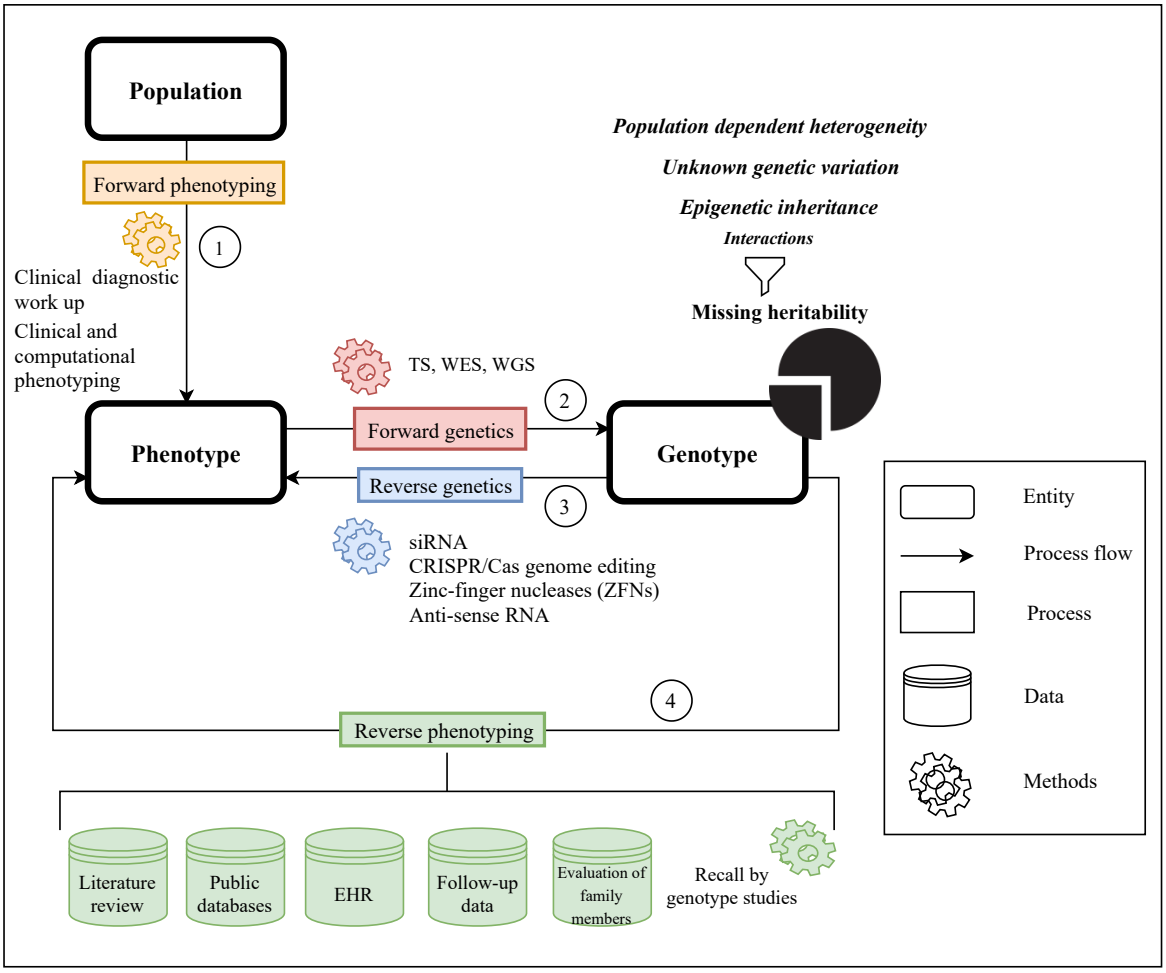
Deep forward and reverse phenotyping for genetic discovery in pulmonary arterial hypertension.

Emilia Maria Świetlik

Pulmonary arterial hypertension (PAH) is a rare disease characterised by constriction and obliteration of small pulmonary arteries, which leads to increased pulmonary vascular resistance and in consequence, right ventricular failure and death. The accurate clinical diagnosis of pulmonary hypertension became possible in the 1950s due to the invention of right heart catheterisation, but it was not until 2000 when the landmark discovery of the causative role of bone morphogenetic protein receptor type 2 (*BMPR2*) mutations shed new light on the pathogenesis of PAH. Since then, several genes have been discovered, which now account for around 25% of cases with the clinical diagnosis of idiopathic PAH (IPAH). Despite the ongoing efforts, for the majority of patients, the cause of the disease remains elusive, a phenomenon often referred to as "missing heritability". The objective of my PhD project was to expand our understanding of the genetic architecture of PAH by studying a large international cohort of deeply phenotyped patients with idiopathic and heritable PAH who had been whole-genome sequenced. The approach I have used in this thesis differs from previously published studies. Rather than relying solely on case labels in keeping with the diagnostic classification of PAH, I clustered patients into homogeneous groups based on deep phenotype information in the hope that such groups would be enriched for rare deleterious variants. I took a two-fold approach. Firstly, using domain knowledge, I stratified the patients based on age, acute nitric oxide challenge response, transfer factor for carbon monoxide (KCO), the presence of small cardiac defects and poor- or super- survivor status. Secondly, to account for complex phenotypes, I annotated patients with Human Phenotype Ontology terms and devised computational clusters based on ontological similarity. Once patients were grouped into new phenotypic clusters, I deployed a Bayesian model comparison method, BeviMed, for case-control analysis. The BeviMed analysis identified 59 significant

gene-tag associations with posterior probability (PP) above 0.75 (when prior set to 0.001), including three associations with a new risk gene, Kinase insert domain (*KDR*), encoding vascular endothelial growth factor receptor 2 (VGFR2), and a strong association with a new candidate gene, *COL6A5*. While *BMPR2*, *TBX4*, *EIF2AK4*, *ACVRL1* and *AQP1* showed the highest association (PP $\geq$ 0.99), I also confirmed significant associations in the majority of other previously identified genes. High impact variants in the *KDR* were associated with a significantly reduced KCO (KCO lower tertile, log(BF) = 11.362, PP = 0.989), older age at diagnosis (tag: old age, log(BF) = 9.249, PP = 0.912) and PAM 2 (Lin) cluster (tag: PAM2, log(BF) = 8.048, PP = 0.758) under the autosomal dominant mode of inheritance. Moderate and high impact variants in *COL6A5* were strongly associated with divisive hierarchical clustering cluster 2 (tag: Divisive HC 2, log(BF) = 10.627, PP = 0.976) under the autosomal dominant mode of inheritance. Computationally derived phenotypes led to the discovery of three additional new candidate genes (*OR6T1*, *EYS*, *HPSE2*), albeit at lower significance levels. The refined phenotype approach corroborated many previously reported genotype-phenotype associations. Individuals with rare variants in *BMPR2*, *TBX4* (high impact), *EIF2AK4* (biallelic) and *SOX17* had a significantly younger age of disease onset (tag: young age). Patients with familial pulmonary arterial hypertension were harbouring deleterious variants in *AQP1* (log(BF) = 10.023, PP = 0.958). Additionally, high impact variants in *BMPR2* were associated with preserved KCO (tag: KCO higher tertile, log(BF) = 99.923, PP = 1), while biallelic *EIF2AK4* mutations showed association with reduced KCO (tag: KCO < 50% predicted, log(BF) = 29.741, PP = 1).

Finally, I characterised patients harbouring mutations in two new pertinent genes, namely *KDR* and *GDF2*. I found that patients harbouring protein-truncating variants in *KDR* showed mild fibrotic changes on high resolution computed tomography, which were further characterised as patchy bronchocentric fibrosis on wedge biopsy from one patient. *GDF2* mutation carriers were similar to PAH patients without mutations and showed no features of hereditary haemorrhagic telangiectasia (HHT) or vascular anomaly syndromes. *GDF2* mutation carriers were significantly older and had less severe haemodynamics than *BMPR2* mutation carriers, and did not constitute a distinct cluster. Additionally, I found a significant correlation between the prodomain-bound form of the bone morphogenic protein 10 (pBMP10) levels and the risk of developing systemic hypertension among patients with IPAH. In summary, deep clinical and computational phenotyping allowed me to devise homogenous groups of patients which were enriched for rare deleterious variants in known and new PAH risk genes.



Schematic representation of the concepts of forward phenotyping (1) and genetics (2) and reverse genetics (3) and phenotyping (4). Public databases include, but are not limited to, Online Mendelian Inheritance in Man (OMIM), Human Phenotype Ontology (HPO), DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources (DECIPHER), ClinVar databases. Forward genetics - "genotype to phenotype" approach; reverse genetics - analysis of the impact of induced variation within a specific gene on gene function; reverse phenotyping - clinical assessment directed by genetic results. Abbreviations: TS: targeted sequencing; WES: whole-exome sequencing; WGS: whole-genome sequencing; siRNA: small interfering RNA; EHR: Electronic Healthcare Records.

Fig. 1 Graphical abstract.

Acknowledgements

I wish to express my sincere appreciation to my supervisor, Professor Nicolas Morrell, for his support and inspiration. Being his student has been an honour and a joy, and without him, none of this would have been possible. I would also like to thank my co-supervisor, Dr Stefan Gräf, for his aid and direction. I am grateful to all the people whose assistance was a milestone in the completion of this project. I am indebted to Dr Daniel Greene, who has been generous with his time and statistical expertise. I wish to pay special regards to my colleagues from the Bioinformatics Team, Dr Marta Bleda, Dr Matthias Haimel, Dr Kathryn Auckland and Mr Tobias Tilly, for their time and support and to Dr Paul Upton and Dr Joshua Hodgson for fruitful collaboration on parts of this project. Furthermore, I would like to thank my mentors Professor Anna Doboszyńska and Dr Joanna Pepke-Żaba, for being wonderful teachers and friends and guiding me on the right path. Lastly, I would like to thank my parents, Krystyna and Mirosław, for opening all the doors which led me here and my siblings Piotrek, Justyna and Tomi for all their encouragement and support during my PhD.

Table of contents

List of figures **xvii**

List of tables **xxiii**

Nomenclature **xxix**

1 Introduction **1**

1.1 Background . . . 2
1.1.1 Definitions . . . 2
1.1.2 Epidemiology . . . 3
1.1.3 Genetic basis of rare diseases . . . 6
1.2 Forward phenotyping . . . 12
1.2.1 Clinical phenotype derivation . . . 14
1.2.2 The rationale for using computationally derived phenotypes . . . 25
1.3 Forward genetics . . . 27
1.3.1 Concepts . . . 27
1.3.2 Methodology . . . 28
1.4 Reverse genetics . . . 34
1.4.1 Concepts . . . 34
1.4.2 Methodology . . . 35
1.5 Forward and reverse genetics studies in PAH . . . 37
1.5.1 Rare genetic variation . . . 37
1.5.2 Common Genetic Variation . . . 49
1.6 Reverse phenotyping . . . 52
1.6.1 Concepts . . . 52
1.6.2 Theoretical Basis . . . 52
1.6.3 Methodology . . . 52

1.6.4 Studies . . . 54
1.7 Missing heritability in the postgenomic era . . . 57
1.7.1 Epigenetic Inheritance . . . 58
1.8 Clinical utility and therapeutic implications of genetic testing in PAH . . . 61
1.8.1 Clinical utility of genetic testing . . . 61
1.8.2 Pharmacogenetics . . . 61
1.8.3 Knowledge, attitudes and barriers towards genetic testing and counselling . . . 64
1.9 Summary of chapters and aims of study . . . 64

2 Methods **67**
2.1 Study design . . . 67
2.2 Ethics . . . 68
2.3 Setting . . . 68
2.4 Participants . . . 68
2.5 Data sources . . . 69
2.5.1 Clinical data capture, processing, quality control and release . . . 69
2.5.2 Clinical data verification and analysis . . . 70
2.5.3 Ethnicity . . . 73
2.5.4 Census . . . 73
2.5.5 Risk stratification . . . 73
2.5.6 Assessment of comorbidities . . . 74
2.5.7 Standardised vocabulary for description of phenotypic abnormalities 74
2.5.8 External data sets . . . 75
2.5.9 Bias . . . 77
2.6 Statistical methods . . . 78
2.6.1 Methods used to handle missing data . . . 78
2.6.2 Descriptive statistics . . . 83
2.6.3 Survival analysis . . . 83
2.7 Clinical phenotyping . . . 83
2.7.1 The analysis of computerised tomography (CT) scans . . . 83
2.7.2 Stratification . . . 84
2.8 Computational phenotyping . . . 84
2.8.1 Clustering algorithms . . . 85
2.9 Genetic data . . . 87

2.9.1 Whole-genome sequencing (WGS), short read alignment, and variant calling . . . 87
2.9.2 The PAH domain samples for the genetic analysis . . . 87
2.9.3 "Labelling" of explained cases . . . 88
2.9.4 Selection of gene biotypes and rare variants for association testing . 89
2.9.5 Genetic association between rare variants and selected diagnostic and phenotypic tags . . . 90
2.9.6 Conclusions . . . 93

3 Results **95**
3.1 Clinical characterisation of study participants . . . 95
3.1.1 The NBR and I/HPAH Cohort studies . . . 95
3.1.2 The PAH USBB study . . . 105
3.1.3 Discussion . . . 110
3.1.4 Conclusions . . . 113
3.2 Forward phenotyping . . . 114
3.2.1 Clinical phenotypes . . . 114
3.2.2 Computational phenotypes . . . 142
3.2.3 Discussion . . . 153
3.2.4 Conclusions . . . 154
3.3 Forward genetics . . . 155
3.3.1 Characterisation of study cohorts and tag definitions . . . 155
3.3.2 Rare variants in previously established genes . . . 156
3.3.3 Rare variant association testing . . . 159
3.3.4 Rare high impact variants in the new PAH candidate gene *KDR* . . 161
3.3.5 Clinical characterisation of *KDR* mutation carriers (reverse phenotyping) . . . 166
3.3.6 Additional *KDR* cases in the US PAH cohorts . . . 167
3.3.7 Other putative genes . . . 169
3.3.8 Discussion . . . 173
3.3.9 Conclusions . . . 178
3.4 Reverse phenotyping . . . 179
3.4.1 Clinical characterisation of *GDF2* carriers . . . 179
3.4.2 Plasma BMP9 and pBMP10 levels in PAH patients . . . 181
3.4.3 Discussion . . . 189

3.4.4 Conclusions 192

4 Final remarks **193**
4.1 Conclusions 193
4.2 Limitations, challenges and future directions 194

References **197**

Appendix A Appendix Figures **259**

Appendix B Appendix Tables **273**

List of figures

1 Graphical abstract. ix

1.1 Potential factors contributing to missing heritability in PAH. Abbreviations: CNV: copy number variants; PAH: pulmonary arterial hypertension 8
1.2 Genetic landscape of pulmonary arterial hypertension. 10
1.3 Phenotypic and genetic heterogeneity in PAH. 11
1.4 Forward phenotyping strategies. 12
1.5 Visual representation of the rationale behind the extreme phenotype approach. Modified from Amanat et al. 2020 [13]. 13
1.6 Schematic representation of the relationship between artificial intelligence (AI), machine learning (ML) and deep learning. 26
1.7 Druggable molecular pathways involved in the pathogenesis of PAH. 63

2.1 Data collection, harmonisation, verification pipeline (The NBR and Cohort studies). 71
2.2 Data collection, harmonisation and verification pipeline (The US PAH Biobank study). 76
2.3 Heatmap showing percentage of missing data per variable (rows) and per centre (columns) among core variables. 79
2.4 Schematic representation of imputation procedure. Abbreviations: df: data frame; res: results 80
2.5 Outflux-influx plot. 81
2.6 The convergence of the algorithm over 20 interactions. 82
2.7 Design of the genetic association study. 91

3.2 Breakdown of diagnoses in the study cohort. 96
3.1 Consort diagram of the NBR and Cohort studies. 97

3.3 Kaplan-Meier curves by age tertiles (A), Charlson comorbidity score (B), incident or prevalent label (C) overall survival, (D) transplant free survival . 104
3.4 Kaplan-Meier curves by Reveal risk group (A) and ERS risk group (B). . . 105
3.5 The US PAH Biobank study participants who underwent whole-exome sequencing (WES). 106
3.6 Use of drugs associated with PAH in the entire NBR/Cohort study (A) and the entire US PAHBB cohort (C), and among patients with drugs and toxins induced PAH from the NBR/Cohort study (B) and the US PAHBB (D). . . . 108
3.7 Population structure over time (I/HPAH and PVOD/PCH) in (A) the NBR/-Cohort studies, (B) the US PAH Biobank. 109
3.8 Charlson comorbidity score in patients diagnosed with IPAH/HPAH/PVOD/PCH (A) The NBR/Cohort studies, (B) The US PAHBB. 111
3.9 Analysis flow chart for abnormalities in pulmonary function tests in I/HPAH patients. 116
3.10 Missingness rate and pattern in pulmonary function tests. Missing data shown in red, existing data shown in blue. 117
3.11 Distribution of % predicted values of pulmonary function tests and FEV1/FVC ratio. (A) FEV1 % pred., (B) FVC % pred., (C) FEV1/FVC ratio, (D) TLC % pred., (E) KCO % pred. In red are the patients with values below 80% predicted for (A) FEV1 (39%), (B) FVC (22%), (D) TLC (17%), (E) KCO (67%) and (C) below 0.7 for FEV1/FVC ratio (24%). 118
3.12 Kaplan-Meier curves showing survival according to TLC (A), FEV1/FVC ratio (C), and KCO (E) threshold, and ROC curves showing the trade-off between sensitivity and specificity for TLC (B), FEV1/FVC ratio (D) and KCO (F). 122
3.13 Overlap between patients fulfilling various diagnostic criteria used to define acute vasoreactivity 130
3.14 Long-term survival according to the vasoresponse status. Two patients in non-responder group were excluded from the survival analysis due to missing date of diagnosis. 131
3.15 Long-term survival according to the presence or absence of small cardiac atrial defects. 140
3.16 Bar plot indicating the number of human phenotype ontology (HPO) terms annotated to the cases in (A) the NBR/Cohort studies, (B) the US PAHBB study. 143

3.17 Heat map showing cumulative number of term occurrences per body system. 144
3.18 Machine learning methods used to discern ontological phenotypes in the NBR/Cohort studies. 146
3.19 Affinity propagation clustering and cluster selection. 147
3.20 Results of the partitioning around medoids clustering on the similarity matrix derived using Lin method. (A) Two-dimensional projection of four clusters, of which three were statistically significant (cluster 1, cluster 2 and cluster 4), (B) Kaplan-Meier survival curves by cluster. 148
3.21 Kaplan-Meier survival curves by clusters devised by divisive hierarchical clustering. 150
3.22 The results of partitioning around medoids algorithm performed on a combined data set (The NBR/Cohort studies and the US PAHBB). 152
3.23 Consort diagram of the genotype-phenotype study. (A) Case-control study flow chart, (B) Number of PAH domain samples in the analysis broken down by maximal unrelated set and related set, and cases (included and excluded) and unaffected relatives. 155
3.24 Rare variant association study results revealing established and new genotype-phenotype links. 162
3.25 Established and new genotype-phenotype links 163
3.26 Summary of rare single nucleotide variants (SNVs) and small insertions and deletions (indels) identified in the novel PAH candidate gene *KDR*. 164
3.27 Chest computerised tomography (CT) scans of patients carrying high impact *KDR* mutations. 168
3.28 Mutations in *GDF2*. 182
3.29 Plasma samples collected from control subjects and patients with PAH were assayed for (A) BMP9 and (B) pBMP10. 184
3.30 Correlation between plasma pBMP10 and plasma BMP9 levels. 185
3.31 Correlations between plasma BMP9 and pBMP10 levels and BMI in cases and controls. 186
3.32 Kapplan-Meier survival curves for plasma BMP9 (A) and pBMP10 levels (B). 187
3.33 Plasma gradients of BMP9 and pBMP10. 188

A.1 Comparison of ROC curves of ERS risk score: (A) ERS risk score + KCO, (B) ERS risk score + TLC, (C) ERS risk score + FEV1/FVC ratio 260
A.2 Heat map showing missingness in reported autoimmune screen. 261

A.3 Cluster of patients with Hereditary Hemorrhagic Telangiectasia (HHT). Ontological similarity p-value< 0.001. 262
A.4 An exemplary cluster produced by affinity propagation clustering driven by substance addition phenotype. 263
A.5 Visualisation of the results of direct and statistical methods for choosing the optimal number of clusters. Most methods pointed towards two clusters. Lin semantic similarity used. 264
A.6 Comparison between dendrograms created by different linkage methods on Lin semantic similarity matrix. The plot shows no correlation between dendrograms. 265
A.7 Characterisation and survival analysis of the cohort based on transfer coefficient for carbon monoxide (KCO) and age at diagnosis. Distribution of and Kaplan-Meier survival curves for KCO below and above the 50% predicted threshold (A, B), KCO tertiles (C D), and age tertiles (E, F). 266
A.8 Summary of large deletions identified in previously established disease genes. Deletions are indicated by light blue boxes. The protein-coding genes, annotated in the displayed region by Ensembl (GRCh37, version 75), are depicted in the bottom panels. The affected genomic regions, with the disease gene locus highlighted in red and the magnified view focusing on the gene loci, are shown for *BMPR2* (A, B), *GDF2* (C, D) and *TBX4* (E, F). 267
A.9 Chest computerised tomography (CT) scans of daughter of W000229 carrying the inherited nonsense variant in *KDR*. (A) The scan shows bibasal reticular ground glass changes with mild traction bronchiectasis. (B) Upper left lobe shows sub pleural reticular ground glass changes in keeping with interstitial fibrosis. Further ground glass changes are visible in the right lung. 268
A.10 (A) IDH3G mutations superimposed on the structure of IDH3 holoenzyme, (B) Enzymatic activity assay of mutant and wild type holoenzymes, (C) Enzymatic activity assay results showing that the mutant holoenzymes exhibit comparable activity and kinetic properties as the wild-type IDH3 holoenzyme. 269
A.11 BMP9 and pBMP10 levels do not correlate with age for males or females in cases and controls. The levels of BMP9 and pBMP10 were assessed regarding any correlation to age for males or females using a Spearman test for combined cases and controls, cases only or controls only. 270

A.12 ROC curves for pBMP10 measurements as a predictor of systemic hypertension (HTN) and diabetes mellitus type 2 in I/HPAH patients. ROC analysis for pBMP10 levels as a predictor of (A) systemic hypertension (HTN) and (B) diabetes mellitus type 2. The diagonal line indicates the line of no discrimination. 271

List of tables

1.1 Haemodynamic definitions of pulmonary hypertension (PH) 2
1.2 Clinical classification of pulmonary hypertension. 4
1.3 Definitions and changes to the classification of Group 1 PAH. 5
1.4 Different types of auto-antibodies found in PAH. 23
1.5 Examples of successful studies that used reverse phenotyping. 55

2.1 NIHR BioResource - Rare Diseases domain definitions and number of maximal unrelated individuals per project. 88

3.1 Study cohort characteristics (The NBR and PAH Cohort studies) 99
3.1 Study cohort characteristics (The NBR and PAH Cohort studies) 100
3.2 Clinical characteristics by age strata (The NBR and Cohort studies). 101
3.3 Clinical characteristics of poor- and super-survivors (The NBR and Cohort studies). 103
3.4 Risk stratification by ESC/ERS and REVEAL risk scores. 110
3.5 Age at diagnosis (median and IQR) by time period under investigation for the NBR/Cohort study and the US PAH Biobank. 112
3.6 Comparison of baseline clinical characteristics and missingness rates between the NBR/Cohort study and The US PAH Biobank. 112
3.7 Characteristics of the study population - abnormalities in lung function tests. 115
3.8 Clinical characteristics of study population by total lung capacity threshold. 119
3.9 Clinical characteristics of study population according to FEV1/FVC ratio threshold at 0.7. 120
3.10 Clinical characteristics of study population according to severity of reduction of KCO 121
3.11 Result of Cox regression analysis relating overall survival to individually selected variables at baseline. Complete cases analysis. 123

3.12 Result of Cox regression analysis relating overall survival to individually selected variables at baseline. Imputed data analysis. 124
3.13 ROC analysis on complete and imputed data. 125
3.14 Correlation between baseline TLC and FEV1/FVC ratio and the change in 6MWD and NTproBNP. 125
3.15 Comparison of clinical characteristics between vasoresponders and non-responders. 132
3.16 Differences between patients with reported positive and negative autoimmune screen. The NBR and Cohort Studies) 136
3.17 Comparison between patients with positive and negative in house autoimmune screen. The NBR and Cohort studies. 137
3.18 Differences between the patients with and without small cardiac defects. The NBR and Cohort studies 139
3.19 Strength of ontological similarity in clusters defined by mutation in PAH risk genes or presence of a clinical syndrome with known genetic background (HHT). The NBR and Cohort study. 145
3.20 Clinical characteristics of patients in four clusters devised by partitioning around medoids algorithm (PAM). The NBR and Cohort studies. 149
3.21 Clinical characteristics of clusters produced by divisive hierarchical clustering on Lin similarity matrix. The NBR and Cohort studies. 151
3.22 Clinical characterization of the NBR study population. 157
3.23 BeviMed analysis result comparison. 160
3.23 BeviMed analysis result comparison. 161
3.24 Gene changes for IPAH patients harboring protein-truncating variants (PTV) in the *KDR* gene. 165
3.25 Differences between patients harbouring missense and PTV *KDR* variants . 167
3.26 Clinical characteristics of patients harbouring deleterious variants in *COL6A5*.170
3.27 Gene changes for IPAH patients harboring PTV and missense variants in the *IDH3G* gene. 172
3.28 Comparison of high impact likely loss-of-function variants in *KDR* in the Human large-scale sequencing reference populations gnomAD and TOPMed with the NBR non-PAH controls and PAH cases. 173
3.29 Gene changes, demographics and pulmonary haemodynamic data for PAH patients harbouring *BMP9* mutations. 180

3.30 Clinical characteristics of *BMPR2* and *GDF2* mutation carriers and patients without these mutations. . . . 181
3.31 Clinical phenotypes of BMP10 mutation carriers. . . . 183
3.32 Differences in BMP9 and pBMP10 levels and activity at various anatomical locations. . . . 189

B.1 Exclusion criteria related to pulmonary function tests in pivotal randomised controlled trials leading to drug approval in PAH. . . . 274
B.2 Landmark forward genetics studies in Group 1 PAH. . . . 275
B.2 Landmark forward genetics studies in Group 1 PAH. . . . 276
B.3 Supporting evidence for the role of risk genes in PAH pathogenesis. . . . 277
B.3 Supporting evidence for the role of risk genes in PAH pathogenesis. . . . 278
B.3 Supporting evidence for the role of risk genes in PAH pathogenesis. . . . 279
B.3 Supporting evidence for the role of risk genes in PAH pathogenesis. . . . 280
B.4 Eligibility criteria for A. The NIHR Bioresource Rare Disease (NBR) study PAH domain B. The Cohort PAH study. . . . 281
B.5 Summary of electronic case report forms (eCRFs) constructed to capture phenotypic information . . . 282
B.6 ICD-10 codes used to calculate Charlson and Elixhauser scores. . . . 283
B.7 Inclusion criteria to The US PAH Biobank. . . . 284
B.8 Reporting proforma for CT scan revision. . . . 285
B.9 Definitions or clinical characteristics (in case of computational phenotypes) of labels and the number of unrelated cases and controls included in the genetic association analysis with BeviMed. . . . 286
B.9 Definitions or clinical characteristics (in case of computational phenotypes) of labels and the number of unrelated cases and controls included in the genetic association analysis with BeviMed. . . . 287
B.10 Breakdown of cases from the NBR and the PAH Cohort studies per centre. . 288
B.11 International registries . . . 289
B.11 International registries . . . 290
B.11 International registries . . . 291
B.11 International registries . . . 292
B.12 Differences in clinical characteristics between patients with complete and missing pulmonary function tests. . . . 293

B.13 Clinical characteristics of study population by FEV1 threshold – complete cases. . . . 294
B.14 Clinical characteristics of study population by FVC severity. . . . 295
B.15 Summary of imaging analysis by FEV1/FVC ratio. . . . 296
B.15 Summary of imaging analysis by FEV1/FVC ratio. . . . 297
B.16 Summary of imaging analysis by TLC threshold. . . . 298
B.17 Summary of imaging analysis by KCO threshold. . . . 300
B.17 Summary of imaging analysis by KCO threshold. . . . 301
B.18 Correlations between pulmonary function tests predictive of survival in univariate analysis and other relevant clinical parameters. . . . 302
B.19 Comparison of patients with and without post-baseline values. . . . 303
B.20 The number of patients who had the acute vasoreactivity test performed by diagnosis and response status. . . . 304
B.21 Reported autoantibody screen results. . . . 305
B.22 Autoimmune screening by in house immunofluorescence. . . . 306
B.23 Clinical characteristics of patients in clusters selected for BeviMed analysis. 307
B.23 Clinical characteristics of patients in clusters selected for BeviMed analysis. 308
B.23 Clinical characteristics of patients in clusters selected for BeviMed analysis. 309
B.24 Cophenetic correlation coefficient and agglomerative coefficient for various linkage methods - agglomerative hierarchical clustering. . . . 310
B.25 Clinical differences between patients with present and missing transfer coefficient results. . . . 311
B.26 Clinical characteristics of unrelated individuals used in gene-tag association analysis by KCO threshold. . . . 312
B.26 Clinical characteristics of unrelated individuals used in gene-tag association analysis by KCO threshold. . . . 313
B.27 Clinical characteristics of study individuals unrelated individuals used in gene-tag association analysis by KCO tertiles. . . . 314
B.27 Clinical characteristics of study individuals unrelated individuals used in gene-tag association analysis by KCO tertiles. . . . 315
B.28 Results of Cox regression analysis relating overall survival to selected variables at baseline. . . . 316
B.29 Clinical characteristics of all unrelated individuals used in gene-tag association by age tertiles. . . . 318

B.30 Comparison of clinical characteristics between patients harbouring *BMPR2*, *EIF2AK4*, *KDR* and without mutation. 320
B.31 Summary of CT imaging analysis. 321
B.31 Summary of CT imaging analysis. 322
B.32 Variants driving associations with computational phenotypes. 323
B.33 Clinical characteristics of patients and controls in whom BMP9 and pBMP10 measurements were undertaken 325
B.34 Clinical characteristics at the time of sampling of I/HPAH cases by BMP9 concentration tertiles. 326
B.34 Clinical characteristics at the time of sampling of I/HPAH cases by BMP9 concentration tertiles. 327
B.35 Clinical characteristics at the time of sampling of I/HPAH cases by pBMP10 concentration tertiles. 328

Nomenclature

Acronyms / Abbreviations

ABCC8 ATP Binding Cassette Subfamily C Member 8

ACE angiotensin-converting enzyme

ACMG American College of Medical Genetics and Genomics

ACVRL1 Activin A Receptor Like Type 1

ADGRG6 Adhesion G Protein-Coupled Receptor G6

AGTR1 Angiotensin II type 1 receptor

ALK ALK Receptor Tyrosine Kinase

APAH Associated Pulmonary Arterial Hypertension

APOE Apolipoprotein E

AQP1 Aquaporin 1

ASD Atrial Septal Defect

ATP13A3 ATPase 13A3

ATP1A2 ATPase Na+/K+transporting subunit alpha 2

BAC Bacterial Artificial Chromosome

BAS Balloon Atrial Septostomy

BMP Bone Morphogenic Protein

BMPR2 Bone Morphogenetic Protein Receptor Type 2

BOECs blood outgrowth endothelial cells

BPD Bronchopulmonary dysplasia

BRAF B-Raf Proto-Oncogene, Serine/Threonine Kinase

CADD Combined Annotation Dependent Depletion

CAV1 Caveolin 1

CBLN2 Cerebellin 2 Precursor

CCB Calcium Channel Blocker

ceRNAs competing endogenous RNAs

CHD Congenital heart disease

CI Cardiac Index

circRNAs circular RNAs

CNV Copy Number Variant

CO cardiac output

COL18A1 Collagen Type XVIII Alpha 1 Chain

COL6A5 Collagen Type 6 Alpha 5 Chain

COL6A5 Collagen Type VI Alpha 5 Chain

COMPERA Comparative Prospective Registry of Newly Initiated Therapies for Pulmonary Hypertension

COPD Chronic Obstructive Pulmonary Disease

COS Fibroblast-like cell lines derived from monkey kidney tissue

CpcPH Combined pre- and post-capillary PH

CT Computerised Tomography

CTEPH Chronic Thromboembolic Pulmonary Hypertension

CXR Chest radiogram

DECIPHER DatabasE of genomiC varIation and Phenotype in Humans using Ensembl Resources

DLCO Diffusing capacity of the lungs for carbon monoxide

DLNO Diffusing capacity of the lungs for nitric oxide

EC Endothelial Cells

ECHO Echocardiogram

EDN1 Endothelin 1

EHR Electronic Health Records

EIF2AK4 Eukaryotic Translation Initiation Factor 2 Alpha Kinase 4

ELISA Enzyme Linked Immunosorbent Assay

EMR Electronic Medical Records

EMSA Electrophoretic Mobility Shift Assay

EndoMT Endothelial-to-Mesenchymal Transition

ENG Endoglin

eNOS endothelial Nitric Oxide Synthase

ER Endoplasmic Reticulum

ER Eyes Shut Homolg

ESCs Embryonic stem cells

ERS European Respiratory Society

ESC European Society of Cardiology

ESCs Endoplasmic Reticulum

ExAC Exome Aggregation Consortium

FBLN2 Fibulin 2

FEV1 Forced expiratory volume in 1 second

FOAD Fetal Origins of Adult Diseases

FVC Forced Vital Capacity

GATA2 GATA-binding factor 2

GDF2 Growth Differentiation Factor 2

GDF3 Growth Differentiation Factor 3

GDF6 Growth Differentiation Factor 6

GERP Genomic Evolutionary Rate Profiling

GGCX Gamma-Glutamyl Carboxylase

GGO Graound-Glass Opacity

gnomAD The Genome Aggregation Database

GRC Genome Reference Consortium

GWAS Genome-wide association study

HBB Haemoglobin Subunit Beta

HHT Hereditary Hemorrhagic Telangiectasia

HPO Human Phenotype Ontology

HPSE2 Heparanase 2 (Inactive)

I/HPAH Idiopathic/Hereditary Pulmonary Arterial Hypertension

ICD-10 International Classification of Diseases

IDH3G Isocitrate Dehydrogenase (NAD(+)) 3 Non-Catalytic Subunit Gamma

ILD Interstitial Lung Disease

IpcPH Isolated post-capillary PH

iPSC Klippel-Feil Syndrome

iPSC induced Pluripotent Stem Cells

KCNA5 Potassium Voltage-Gated Channel Subfamily A Member 5

KCNK3 Potassium Two Pore Domain Channel Subfamily K Member 3

KCO Transfer Coefficient for Carbon Monoxide

KDR Kinase Insert Domain Receptor

KFS induced Pluripotent Stem Cells

KIR Killer-cell Immunoglobulin-like Receptor

KLF4 Kruppel Like Factor 4

KLK1 Kallikrein 1

KRAS KRAS Proto-Oncogene, GTPase

lncRNAs long non-coding RNAs

LoF Loss of Function

LOINC Logical Observation Identifiers

MAF Minor Allele Frequency

MAPT Microtubule Associated Protein Tau

MCT monocrotaline

MEOX1 Mesenchyme Homeobox 1

MHC Major Histocompatibility Complex

miRNA microRNA

ML machine learning

MLPA Multiplex Ligation-dependent Probe Amplification

mPAP mean Pulmonary Artery Pressure

MR Mendelian Randomisation

mRNA messenger RNA

mtDNA mitochondrial DNA

MTHFR Methylenetetrahydrofolate Reductase

MYO18B Myosin XVIIIB

NBR National BioResource Rare diseases study

ncRNA non coding RNA

NGS Next Generation Sequencing

NIH National Institutes of Health

NIHR National Institute for Health Research

NO nitric oxide

NRAS NRAS Proto-Oncogene, GTPase

NBR Online Mendelian Inheritance in Man

OR Odds Ratio

ORPHANET The portal for rare diseases and orphan drugs

PAAT Pulmonary Artery Acceleration Time

PAB Pulmonary Artery Banding

PAEC Pulmonary Artery Endothelial Cells

PAH Pulmonary Arterial Hypertension

PASMC Pulmonary Artery Smooth Muscle Cells

PAWP Pulmonary Artery Wedge Pressure

PDGFD Platelet Derived Growth Factor D

PFO Patent Foramen Ovale

PFT Pulmonary Function Test

PH Pulmonary Hypertension

piRNAs piwi-interacting RNAs

PolyPhen-2 Polymorphism Phenotyping v2

PPARG Peroxisome Proliferator-Activated Receptor Gamma

PPHN Persistent Pulmonary Hypertension of the Newborn

PPH Primary Pulmonary Hypertension

PTPN11 Protein Tyrosine Phosphatase Non-Receptor Type 11

PTV Protein-Truncating Variants

PVOD Pulmonary veno-occlusive disease

PVR Pulmonary Vascular Resistance

RAA Renin-Angiotensin-Aldosterone

RAF1 Raf-1 Proto-Oncogene, Serine/Threonine Kinase

RbG Recall by genotype

REVEL Rare Exome Variant Ensemble Learner

RHC Right Heart Catheterisation

RIT1 Ras Like Without CAAX 1

RVAS Rare Variant Association Study

RV Right Ventricle

RVSP Right Ventricle Systolic Pressure

SHOC2 SHOC2 Leucine Rich Repeat Scaffold Protein

SIFT Sorting Intolerant From Tolerant

SIRT3 Sirtuin 3

SKAT-O (Sequence) Kernel Association Test - optimal

SKAT (Sequence) Kernel Association Test

SLE Systemic Lupus Erythematosus

SMAD1 SMAD Family Member 1

SMAD4 SMAD Family Member 4

SMAD9 SMAD Family Member 9

SMC Smooth Muscle Cell

SNOMED-CT SNOMED Clinical Terms; a systematically organized computer processable collection of medical terms

SNP Single Nucleotide Polymorphisms

SNVs Single Nucleotide Variants

SOD2 superoxide dismutase 2

SOS1 SOS Ras/Rac Guanine Nucleotide Exchange Factor 1

SOS2 SOS Ras/Rac Guanine Nucleotide Exchange Factor 2

SOX17 SRY-Box Transcription Factor 17

SPARK Simons Foundation Powering Autism Research for Knowledge

SPS Small patella syndrome

STAT5 Signal transducer and activator of transcription 5

SuHx Sugen 5416-hypoxia

SV Structural Variant

TALENs Transcription Activator-Like Effector Nucleases

TAPSE Tricuspid Annular Plane Systolic Excursion

TBX4 T-Box Transcription Factor 4

TCA Tricarboxylic acid

TERT telomerase reverse transcriptase

TET2 Tet Methylcytosine Dioxygenase 2

TGF-B Transforming Growth Factor Beta

TLC Total Lung Capacity

TopBP1 DNA Topoisomerase II Binding Protein 1

TopMed Trans-Omics for Precision Medicine

TSP1 Thrombospondin 1

UCP2 uncoupling protein 2

UGT2B UDP-glucuronosyltransferases 2B subfamily

UniProt Universal Protein Resource

VA alveolar volume

VEGFR2 Vascular Endothelial Growth Factor Receptor 2

VEGF Vascular Endothelial Growth Factor

VOC Volatile organic compounds

VUS Variant of Unknown Significance

WES Whole-Exome Sequencing

WGS Whole-Genome Sequencing

WSPH World Symposium on Pulmonary Hypertension

WT Wild Type

ZFN Zinc Finger Nucleases

Chapter 1

Introduction

Parts of this Introduction chapter have been previously published. (1) Swietlik EM, Prapa M, Martin JM, Pandya D, Auckland K, Morrell NW, Gräf S. 'There and Back Again'-Forward Genetics and Reverse Phenotyping in Pulmonary Arterial Hypertension, Genes (Basel). 2020 Nov 26;11(12):1408. PMID: 33256119, (2) Swietlik EM, Morrell NW, Chapter 1. The Genetic Basis of Pulmonary Arterial Hypertension, Pulmonary Vascular Disorders, Nova Publishers, 2021 (in press)

This chapter aims to present the definition and the epidemiology of pulmonary arterial hypertension and the overview of the genetic basis of rare diseases together with the approaches to disentangle genetic architecture of pulmonary arterial hypertension. To this end I present four concepts and corresponding analytical approaches, namely:

- forward phenotyping – a process of describing phenotypes for case-control genetic analysis; this process can be divided into three stages: clinical diagnosis, clinical phenotyping and computational phenotyping
- forward genetics – molecular genetic approach used to elucidate the genetic underpinnings of the phenotype of interest
- reverse genetics – an experimental approach that involves targeted modifications in specific genes to analyse phenotypic impact and validate *in silico* predictions
- reverse phenotyping – a process whereby the genetic marker data are used to define the disease subgroup definitions and characteristics.

1.1 Background

1.1.1 Definitions

Pulmonary hypertension (PH) is defined as an elevation of mean pulmonary artery pressure (mPAP) equal to or greater than 25 mmHg measured by right heart catheterisation (RHC) in the supine position at rest [153]. This threshold was originally proposed at the 1^{st} World Symposium on Pulmonary Hypertension (WSPH) in Geneva in 1973 [204] and has now been updated in line with new data showing that a normal mPAP is 14 ± 3.3 mmHg, which suggests an upper limit of normal at 20 mmHg (14 mmHg + 2SD). This new definition has been endorsed by the 6^{th} WSPH, along with the inclusion of pulmonary vascular resistance (PVR) and pulmonary artery wedge pressure (PAWP) cut-offs into a new haemodynamic definition, which categorises PH based on haemodynamic criteria into three groups: pre-capillary PH, isolated post-capillary PH (IpcPH) and combined pre- and post-capillary PH (CpcPH) [471] (Table 1.1).

Table 1.1 Haemodynamic definitions of pulmonary hypertension (PH)

Haemodynamic definition	Characteristics	Clinical groups#
Pre-capillary PH	PAP $>$ 20 mmHg, PAWP $\leq$ 15 mmHg, PVR $\geq$ 3 WU	1, 3, 4 and 5
Isolated post-capillary PH (IpcPH)	PAP $>$ 20 mmHg, PAWP $>$ 15 mmHg, PVR $<$ 3 WU	2 and 5
Combined pre-and post-capillary PH (CpcPH)	PAP $>$ 20 mmHg, PAWP $>$ 15 mmHg, PVR $\geq$3 WU	2 and 5

Abbreviations: mPAP: mean pulmonary arterial pressure; PAWP: pulmonary arterial wedge pressure; PVR: pulmonary vascular resistance; WU: Wood Units. #: group 1: PAH; group 2: PH due to left heart disease; group 3: PH due to lung diseases and/or hypoxia; group 4: PH due to pulmonary artery obstructions; group 5: PH with unclear and/or multifactorial mechanisms.

These updates are also expected to feature in the new ERS/ESC guidelines, due to be published in 2021. Beyond the haemodynamic definition, the disease classification

is the second component of the clinical diagnosis and has also evolved over time. The clinical classification (Table 1.2) groups patients according to the driving pathophysiological mechanism, haemodynamic features (Table 1.1) and response to vasodilators [471]. Despite comprehensive and multidimensional criteria for group inclusion and exclusion, there is a persistent heterogeneity within the groups and subgroups and a growing notion that some patients may fit into more than one category. These observations result in ongoing changes to the disease classification. For example, the 2^{nd} WSPH, in 1998, recognised primary and secondary forms of PH and coined the term familial PAH (FPAH) to denote cases in which at least two family members were affected. The landmark discovery of the role of bone morphogenetic protein receptor type II (*BMPR2*) in the pathogenesis of both familial and sporadic cases led to the introduction of the term heritable PAH (HPAH) by the 4^{th} WSPH in 2008, to describe both patients with familial disease and sporadic cases in which causative mutations were identified. Likewise, based on recent genetic advances, the 6^{th} WSPH proposed "PAH with overt features of venous/capillaries (PVOD/PCH) involvement" (Group 1.6) as a subgroup of Group 1. Similarly, "PAH long-term responders to calcium channel blockers", characterised by distinct disease pathobiology [208] and prognosis [473], have been recognised as a distinct subgroup (Group 1.5) [471] (Tables 1.2 and 1.3). As the diagnostic algorithm evolved over time, one of the changes was recognition of the role and importance of genetic testing and counselling, particularly emphasised in the ERS/ESC guidelines from 2015 [153].

1.1.2 Epidemiology

PH poses a significant healthcare burden, particularly in countries with increasing ageing populations. The condition is estimated to affect 1% of people globally and up to 10% of those over the age of 65. In Western countries, elderly patients with left heart and/or lung diseases are most commonly affected; conversely, in the developing world, PH is frequently associated with congenital heart diseases (CHD) and infectious diseases, like schistosomiasis, HIV and rheumatic fever, and affects the younger population [217].

Idiopathic and heritable PAH, on the other hand, are rare diseases [363], with an estimated prevalence ranging from 10 to 52 cases per million [394, 229, 148]. The interest in this condition was triggered by the epidemic of PAH in 1965, attributed to aminorex, which in turn led to the first registry on primary pulmonary hypertension (PPH) in 1980 sponsored by the National Institutes of Health (NIH). This registry established basic demographics (female predominance 1.7 times; mean age of onset 36 years), clinical characteristics (dyspnea 60%,

Table 1.2 Clinical classification of pulmonary hypertension.

1.PAH
1.1 Idiopathic PAH
1.2 Heritable PAH
1.3 Drug- and toxin-induced PAH
1.4 PAH associated with:
1.4.1 Connective tissue disease
1.4.2 HIV infection
1.4.3 Portal hypertension
1.4.4 Congenital heart disease
1.4.5 Schistosomiasis
1.5 PAH long-term responders to calcium channel blockers
1.6 PAH with overt features of venous/capillaries (PVOD/PCH) involvement
1.7 Persistent PH of the newborn syndrome
2. PH due to left heart disease
2.1 PH due to heart failure with preserved LVEF
2.2 PH due to heart failure with reduced LVEF
2.3 Valvular heart disease
2.4 Congenital/acquired cardiovascular conditions leading to post-capillary PH
3. PH due to lung diseases and/or hypoxia
3.1 Obstructive lung disease
3.2 Restrictive lung disease
3.3 Other lung disease with mixed restrictive/obstructive pattern
3.4 Hypoxia without lung disease
3.5 Developmental lung disorders
4. PH due to pulmonary artery obstructions
4.1 Chronic thromboembolic PH
4.2 Other pulmonary artery obstructions
5. PH with unclear and/or multifactorial mechanisms
5.1 Haematological disorders
5.2 Systemic and metabolic disorders
5.3 Others
5.4 Complex congenital heart disease

Abbreviations: PAH: pulmonary arterial hypertension; HIV: human immunodeficiency virus; PVOD/PCH: pulmonary veno-occlusive disease/pulmonary capillary haemangiomatosis; PH: pulmonary hypertension; LVEF: left ventricle ejection fraction

Table 1.3 Definitions and changes to the classification of Group 1 PAH.

WSPH Proceedings and ERS/ESC Guidelines	**Definition of Group 1**	**Comments**	**Changes to the classification**
1st WSPH, Geneva, 1973	No haemodynamic definition mentioned		
2nd WSPH, Evian, 1998	No haemodynamic definition mentioned, but RHC recommended for diagnosis		Introduction of the terms primary (PPH) and secondary (related to other conditions) pulmonary hypertension, recognition of familial forms of PH
3rd WSPH, Venice, 2003	mPAP(R) > 25 mmHg; mPAP(E) > 30 mmHg; PAWP < 15 mmHg; PVR > 3 WU		Abandonment of the term PPH, the introduction of terms idiopathic and familial PAH as well as associated PAH, *BMPR2* and *ACVRL1* implicated in the pathogenesis of PAH
4th WSPH, Dana Point, 2008	mPAP(R) $\geq$ 25 mmHg; PAWP $\leq$ 15 mmHg	Exercise-induced PH removed from the definition as although (R) mPAP has been shown to be stable across age groups, (E) mPAP increases with age hence based on the available data it was not possible to define a cutoff	Introduction of the terms idiopathic (no family history, no precipitating risk factor) and hereditary (encompassing familial cases with or without identified germline mutations and PAH). Inclusion of PH associated with Schistosomiasis and PH associated with chronic hemolytic anaemia to Group 1
ERS/ESC Guidelines, 2009	mPAP(R) $\geq$ 25 mmHg; PAWP $\leq$ 15 mmHg; CO normal or reduced	No definition of PH on exercise	
5th WSPH, Nice, 2013	mPAP(R) $\geq$ 25 mmHg; PAWP $\leq$15 mmHg; PVR > 3 WU	Introduction of PVR to the definition, a recommendation to report PVR in WU; fluid challenge may be helpful to unmask occult LV diastolic dysfunction	*SMAD9*, *CAV1* and *KCNK3* included as risk genes for HPAH
ERS/ESC Guidelines, 2015	mPAP(R) $\geq$ 25 mmHg; PAWP $\leq$ 15 mmHg	The clinical significance of a mPAP between 21 and 24 mmHg is unclear	Group 1$'$ PVOD/PCH has been expanded and includes idiopathic, heritable, drug-, toxin- and radiation-induced and associated forms; PPHN includes a heterogeneous group of conditions that may differ from classical PAH. As a consequence, PPHN has been sub categorised as Group 1$''$
6th WSPH, Nice, 2018	mPAP(R) $\geq$ 20 mmHg; PAWP $\leq$ 15 mmHg; PVR $\geq$ 3 WU	PVR $\geq$ 3 WU should be used as a diagnostic criterion for all forms of pre-capillary PH	PAH long-term responses to calcium channel blockers established as a subtype of Group 1; PAH with overt features of venous/capillaries (PVOD/PCH) involvement established as a subtype of Group 1

Abbreviations: WSPH: World Symposium on Pulmonary Hypertension; ERS/ESC: European Respiratory Society and European Society of Cardiology; mPAP: mean pulmonary artery pressure; (R): resting; (E): exercise; PAWP: pulmonary artery wedge pressure; PVR: pulmonary vascular resistance; LV: left ventricle; *BMPR2*: Bone morphogenetic protein receptor; type II; *ACVRL1*: Activin A receptor like type 1; *SMAD9*: SMAD Family Member 9; *CAV1*: Caveolin 1; *KCNK3*: Potassium two-pore domain channel subfamily K member 3; PVOD/PCH: pulmonary veno-occlusive disease/pulmonary capillary haemangiomatosis; PAH: pulmonary arterial hypertension; PPH: primary pulmonary hypertension; PH: pulmonary hypertension; PPHN: persistent pulmonary hypertension of the newborn; WU: Wood Units; RHC: right heart catheterisation

fatigue 19%, syncope 13%, Raynaud phenomenon 10%, positive antinuclear antibodies in 29%, decreased diffusing capacity for carbon monoxide (DLCO)) and survival (2.8 years) of patients with what was then termed PPH [97].

Subsequent registries accrued an increasing wealth of clinical and epidemiological information in various ethnicities and established clinical trajectory and patient characteristics over time. Invariably across populations and time, the registries showed that female predominance remained a stable feature but more recent data showed an increase in age at diagnosis as well as the higher fraction of patients with comorbidities, both features leading to worse outcomes [218]. Likewise decreased DLCO has been associated with worse survival but not with response to PAH therapies [532, 550]. Low DLCO is also a feature of PVOD/PCH with or without biallelic *EIF2AK4* mutations [194, 293, 357].

The major limitations of registry data are a lack of a control group, variability in diagnostic and therapeutic protocols between participating centres and countries, inclusion bias, missingness and lack of data verification. A complementary approach is to use electronic health service records to learn about disease characteristics from real-life data. Such a study has been recently performed in Ontario, Canada [573], where information about the incidence, prevalence, comorbidities, mortality and prescribing patterns for groups 1 to 4 PH between 1993 and 2012 has been reported. The study showed that both the incidence and prevalence of PH among adults are on the increase and highlighted groups 2 and 3 as the most common and lethal forms of PH, increasing the risk of death by sevenfold. Closer inspection of the data revealed that 35% of patients could be classified to more than one group, with the most common overlap between groups 2 and 3. Importantly, the authors also called attention to the irregularities in prescribing medication approved only for the treatment of group 1 PAH [573]. The evidence from this study suggests an emerging epidemic of PH with substantial adverse health and economic implications.

1.1.3 Genetic basis of rare diseases

Rare diseases, such as PAH, are enriched with underlying genetic causes and are defined as life-threatening or chronically debilitating disorders with a prevalence of less than 1 in 2000 [363]. Although individually characterised by low prevalence, in total, rare diseases pose a significant burden to health care systems and a diagnostic challenge. Over 6000 rare diseases have been reported to date (ORPHANET [379]) and new genotype-phenotype associations are discovered every month[1]. Despite the several-fold increase in genetic diagnoses in the

[1] https://www.orpha.net/consor/cgi-bin/Education_Home.php?lng=EN

area of rare diseases [576], the cause of the disease remains elusive in a significant proportion of cases. Since the 2008 World Symposium on Pulmonary Hypertension (WSPH), the term "heritable PAH" (HPAH) has been used to describe both familial PAH and sporadic PAH with an identified underlying pathogenic variant [472] (Table 1.2 and 1.3). Family history is a vital disease component directly linked to the proportion of variation attributable to genetic factors, known as heritability [139, 577]. Familial studies have been used historically as a tool for gene mapping, with the classical example of twin studies commonly used to disentangle the relative contribution of genes and environment to complex human traits [379]. In statistical terms, heritability is defined as a proportion of the phenotypic variance that can be attributed to the variance of genotypic values:

$$H^2(broad\ sense) = \frac{Var(G)}{Var(P)}$$

H^2 estimates are specific to the population, disease and circumstances on which they are estimated [514]. In theory, total genotypic variance (Var(G)) can be divided into multiple components: $Var(A)$ - total additive variance (breeding values); $Var(D)$ - dominance variance (interactions between alleles at the same locus); $Var(I)$ - epistatic variance (interactions between alleles at different loci); $Var(G \times E)$ - variation arising from interactions between genes and the environment.

$$Var(G) = Var(A) + Var(D) + Var(I) + Var(G \times E)$$

In practice, total genotypic variation is difficult to measure, although estimates can be made, and the components of genotypic variation are nearly unattainable. Hence genetic studies usually refer to heritability in its narrow sense, which is the proportion of the phenotypic variance that can be attributed to favourable or unfavourable alleles.

$$h^2(narrow\ sense) = \frac{Var(A)}{Var(P)}$$

Recently, genome-wide association studies (GWAS) have enabled the estimation of additive heritability attributed to common genetic variation (single nucleotide polymorphisms - SNPs), albeit with a typically small effect size [342]. This has led to the issue of "missing heritability" [337], whereby SNP-based estimates were not sufficient to explain prior heritability predictions arising from twin or familial recurrence studies. To that effect, several hypotheses have been proposed (Fig. 1.1) including the complex interplay between genes and environment and the often overlooked potential contribution of structural variation [164]. To date, about 75%

of patients with a clinical diagnosis of idiopathic pulmonary arterial hypertension (IPAH) have no defined genetic cause of the disease. In the following subsections of the Introduction, I will present the role of forward and reverse genomic and phenomic approaches as well as other "omic" technologies in the search for missing heritability of PAH (Fig. 1.1).

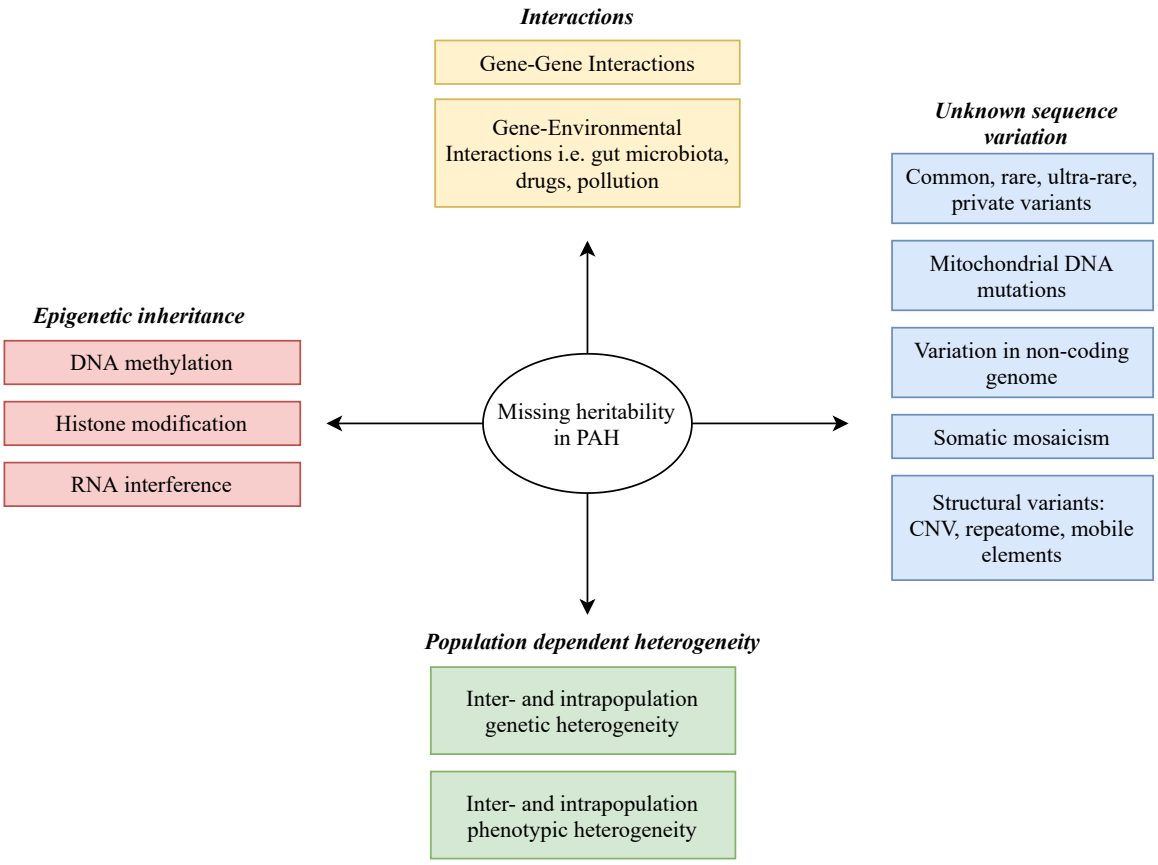
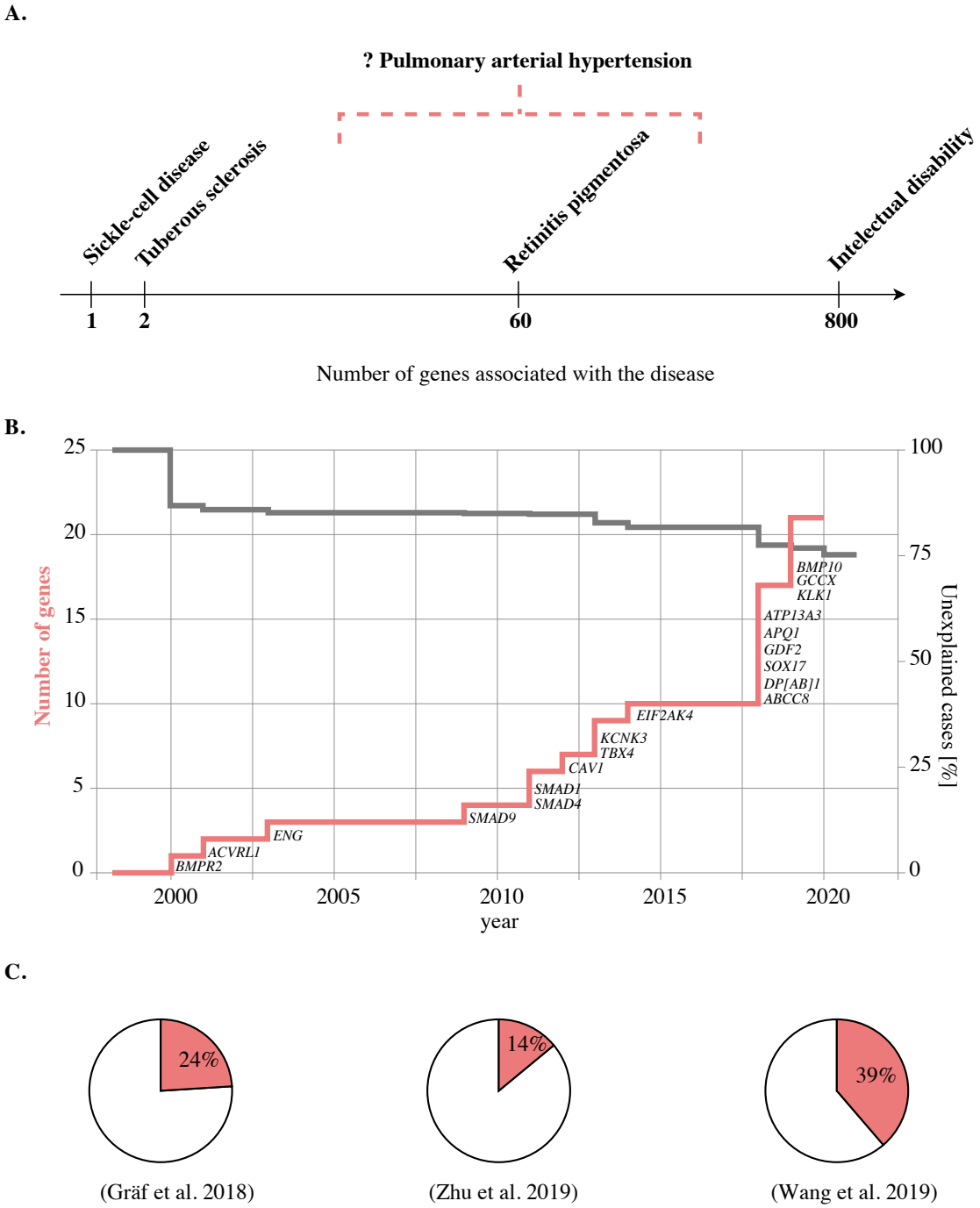


Fig. 1.1 Potential factors contributing to missing heritability in PAH. Abbreviations: CNV: copy number variants; PAH: pulmonary arterial hypertension

Genetic and phenotypic heterogeneity

Discoveries in the field of rare diseases have been hampered by genetic and phenotypic heterogeneity of these entities. Genetic heterogeneity is a situation in which sequence variation in two or more genes results in the same or very similar phenotype. The degree of genetic heterogeneity varies between different diseases (Fig. 1.2 A). For instance, sickle cell anaemia has only been associated with mutations in one gene [259], Haemoglobin subunit β (HBB) and tuberous sclerosis [398] in two genes, retinitis pigmentosa [57] with over 60 and intellectual disability with over 800 [77]. PAH also shows high genetic heterogeneity, with

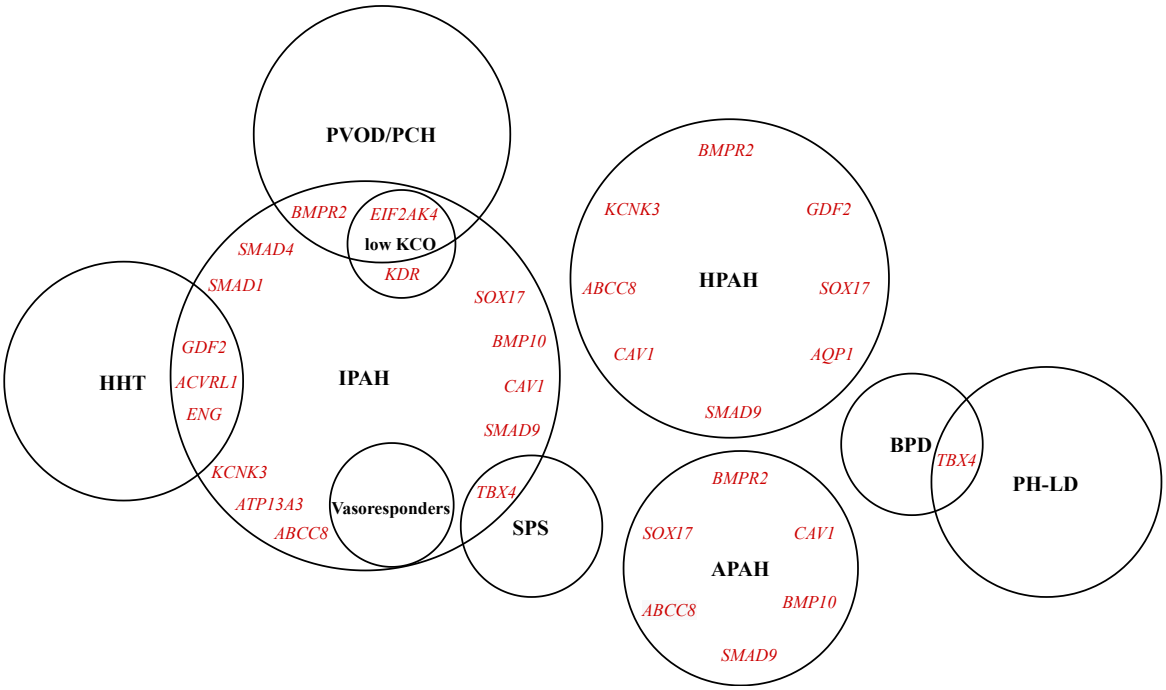
so far approximately 20 risk genes reported [484] and more expected to be found (Fig. 1.2 B and C).



(A) Genetic heterogeneity in various genetic disorders, (B) Genetic discoveries in PAH, (C) Proportion of explained cases by cohort (Gräf et al., 2018 - I/HPAH [179]; Zhu et al., 2019 - Group 1 PAH [598]; Wang et al., 2019 - IPAH [564]).'?' denotes uncertainty around number of genes involved in the pathogenesis of PAH

Fig. 1.2 Genetic landscape of pulmonary arterial hypertension.

As alluded to above, genotypic heterogeneity is further complicated by phenotypic heterogeneity (see Chapter 1.1.1). The clinical classification aims to categorise patients into groups according to pathophysiological mechanisms, haemodynamics and therapeutic management. Despite this, there is persistent heterogeneity within groups and subgroups, and not all patients fit easily into a single category, which can be a reflection of genetic pleiotropy (Fig. 1.3).



Abbreviations: I/H/APAH: idiopathic/hereditary/associated pulmonary arterial hypertension; PVOD/PCH: pulmonary veno-occlusive disease/pulmonary capillary haemangiomatosis; SPS: small patella syndrome; BPD: bronchopulmonary dysplasia; HHT: hereditary hemorrhagic telangiectasia, PH-LD: pulmonary hypertension due to lung diseases and/or hypoxia.

Fig. 1.3 Phenotypic and genetic heterogeneity in PAH.

Other phenotypes previously reported in the literature, but not explicitly included in the guidelines, have been summarised in An Official American Thoracic Society Statement: Pulmonary Hypertension Phenotypes [118]. They consist of “severe” PH in respiratory disease, maladaptive right ventricular (RV) hypertrophy, PH in elderly individuals, PAH in children and PAH with metabolic syndrome [500] and are awaiting clinical validation and confirmation of utility both in clinical practice and research settings.

1.2 Forward phenotyping

A precise definition of the phenotype of interest is a cornerstone of any genetic study (Fig. 1, Process 1, Fig. 1.4). I define a process of describing phenotypes for case-control genetic analysis as forward phenotyping and divide this process into 3 stages: clinical diagnosis, clinical phenotyping and computational phenotyping (Fig. 1.4). As described above, clinical diagnosis relies on clustering patients based on observable and measurable traits, signs and symptoms, which are the product of genetic, epigenetic and environmental factors. As a consequence, clinical phenotypes can be dynamic and reactive, which are useful and desirable in the clinical setting but unsuitable for genetic studies. A distinction must be made between clinical and research diagnosis, particularly diagnosis for genetic analysis. The former is usually spread over time and obtained in three interrelated stages: history taking, physical examination, differential diagnosis and confirmation. In contrast, a research diagnosis is typically obtained during a single encounter, meaning that a decision is made immediately as to whether or not the patient should be recruited. To make this feasible and reliable, standardised checklists and operating procedures need to be in place, diagnostic criteria should follow simple inclusion/exclusion rules and phenotypes need to be described using controlled vocabulary to avoid ambiguity. Additionally, the data must be in a format amenable to computational analysis. Finally, the validity of phenotypes is confirmed in test cohorts, through functional studies and ultimately via reverse phenotyping (see below). The accuracy and precision of phenotype measurements are of paramount importance. In genetic studies, the diagnosis misclassification or admixture of phenocopies can significantly affect power to detect an association [64]. Equally, categorising biologically continuous phenotypes (i.e., age, mPAP, diffusing capacity of the lungs for carbon monoxide (DLCO)) is prone to errors due to flaws in quantification methods and arbitrary thresholds.

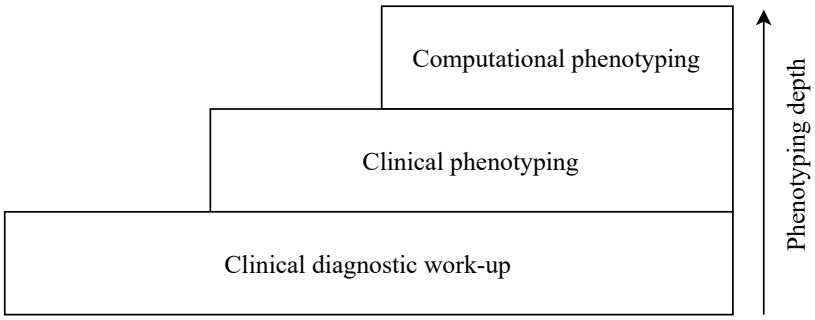


Fig. 1.4 Forward phenotyping strategies.

Phenotype optimisation for genetic studies aims at finding homogenous groups of patients that likely share the same genetic architecture. This can be approached through various strategies. For example, an extreme phenotype strategy aims at identifying rare variants with large effect sizes through recruitment of patients with traits at either end of the phenotypic spectrum (Fig. 1.5). These phenotypes can be based on family history, age of onset, outcome, severity scores, biomarker levels, disease trajectory or response to treatment [420, 235]. Such stratification was proven to increase the power to detect novel disease risk genes [80] and to be cost-effective [308]. Other strategies include covariate-based methods, which jointly estimate the effect of multiple variables and data reduction techniques. Alternatively, intermediate phenotypes can be used. Intermediate phenotypes are features closer to underlying biology that are at least as heritable as the phenotype itself, stable over time, and are associated with the disease of interest [458].

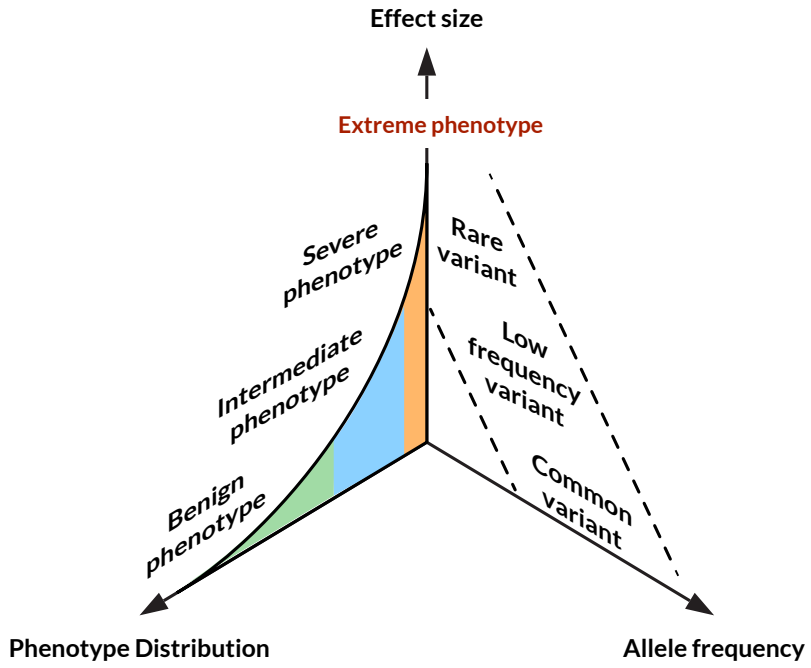


Fig. 1.5 Visual representation of the rationale behind the extreme phenotype approach. Modified from Amanat et al. 2020 [13].

Although clinical phenotyping remains the most widely used method of patient stratification both in clinical practice and research, it requires substantial domain knowledge and is time-consuming. Computational phenotyping based on clinical and/or "omics" datasets using machine learning might be an alternative due to unparalleled diagnostic precision, accuracy, speed and potential to define composite phenotypes, whereby multiple variables are modelled

simultaneously. Two recent publications exemplify the power of computational tools in identifying disease phenotypes. Based on blood cytokine profiles, a prospective observational study of Group 1 PAH discovered and validated four immune phenotypes; importantly, these phenotypes differed in clinical outcomes despite the fact that demographics, PAH aetiologies, comorbidities, and treatments were similar across clusters [502]. Likewise, clinical data mining using the Comparative Prospective Registry of Newly Initiated Therapies for Pulmonary Hypertension (COMPERA) again revealed four clusters with differing survival and response to therapy [220]. A viable alternative approach for clinical data mining is utilising phenotype ontologies, such as Human Phenotype Ontology (HPO), which allow standardised, highly granular and precise phenotyping across different disease domains [268]. Use of ontologies to define phenotypes has already proven useful in identifying novel candidate genes for rare disorders [270]. Ontology-based analysis of phenotypes has been further facilitated by the implementation of methods for manipulation, visualisation and computation of semantic similarity between ontological terms and sets of terms [184].

1.2.1 Clinical phenotype derivation

The rationale for patient stratification based on the current PH classification

As delineated above, the rationale for clinical classification is grouping patients into categories that share common pathophysiology, haemodynamics, clinical presentation, and response to treatment. Importantly these features must be discernible in a clinical setting. The classification considers both genetic and environmental causes based on family history, genetic data and large epidemiological studies. From 2008, familial cases and patients with *de novo* mutations in PAH risk genes have been clustered together under the term HPAH although in clinical practice, the term FPAH is still used to denote familial cases with or without mutations in risk genes. In clinical settings, history of exposure to drugs and toxins is usually collected through questionnaires, which should cover prescription and illicit drugs as well as industrial exposures (i.e. solvents). The latest classification update, which still awaits guidelines endorsement, recognised two PAH subphenotypes: PAH long-term responders to calcium channel blockers (CCB) and PAH with overt features with venous/capillary (PVOD/PCH) involvement. The former category was first recognised in a seminal study by Rich et al. [439], which showed that patients who acutely responded to high doses of CCB (17 out of 64) had a dramatic improvement in haemodynamics and survival. In 2005, Sitbon et al. [473] demonstrated in an extensive series of 557 PH patients that acute vasoresponsiveness may be observed in 12.5% of IPAH, but overall only

6.8% will have a long-term clinical and haemodynamic improvement on CCB. The study defined acute vasoreactivity criteria most predictive of long-term response as a decrease of mPAP of ≥ 10 mmHg to reach an absolute value of mPAP ≤ 40 mmHg with an increased or unchanged cardiac output (CO). Long term response was defined as an improvement in WHO functional class to I or II and sustained haemodynamic improvement after at least one year on CCB only. Since then, several studies have examined acute vasoreactivity and its impact on long-term survival in other groups of PH; these studies concluded that results of acute vasoreactivity testing are mostly misleading and that long term responses are rare [362]. Therefore, vasoreactivity testing is recommended only for patients with I/HPAH and drug-induced PAH [153]. Although the pathobiology of the phenomenon is largely unknown, patients with HPAH secondary to *BMPR2* mutations and sporadic *BMPR2* mutation carriers usually do not show acute response to nitric oxide [124]. Some new insights into the molecular basis of acute vasoreactivity recently came from two studies showing a distinct transcriptomic pattern (microarray of cultured lymphocytes of vasoresponsive patients n = 8) and differences in genetic variants (whole exome sequencing (WES)) between vasoresponders and non-responders [208]. These findings were not replicated in the to date largest GWAS in PAH [436]. Concurrently retrospective studies suggested a sex discrepancy in response to endothelin receptor antagonists (ERA) [150] and phosphodiesterase inhibitors, and more recently a few studies have found genomic associations with response to ERAs [42]. Further research is needed on a large number of patients to explore the molecular bases of response to particular classes of medication. The second phenotype, PVOD/PCH has recently been reclassified as an ultra-rare form of Group 1 PAH [471]. PVOD and PCH often show significant phenotypic overlap. Indeed, 73% of patients diagnosed with PVOD are found to have capillary proliferation and 80% of patients with PCH demonstrate typical venous and arterial changes [290] and are therefore referred to as PVOD/PCH. Clinically, PVOD/PCH is characterised by early-onset, significantly reduced DLCO and patchy centrilobular ground-glass opacities, septal lines and lymph node enlargement seen on high-resolution computed tomography. The disease outcome is dismal, with rapid progression and frequent pulmonary oedema in response to PAH medication. Similarly to PAH, PVOD/PCH can present as either a sporadic or familial disease [104, 555]. It was the latter that triggered a family-based study into the genetic basis of this condition. Familial linkage mapping, WES, and Sanger sequencing were employed and identified biallelic *EIF2AK4* mutations in affected siblings. Subsequently, biallelic *EIF2AK4* mutations were also identified in 25% of sporadic PVOD cases [127], 11.1% of HPAH cases (one of nine cases) [46] and 1.04% of I/HPAH cases (nine of 864 cases) [194]. Harbouring *EIF2AK4* mutations confer a poor prognosis irrespective

of clinical diagnosis, and importantly, radiological assessment was unable to distinguish reliably between PVOD/PCH patients and patients with IPAH [194].

The rationale for patient stratification based on pulmonary function testing

The respiratory system consists of two functional units. The conducting unit, responsible for conducting the air in and out of the lungs, including the trachea, bronchi, bronchioles and terminal bronchioles, and the respiratory unit, which participates in gas exchange and consists of the respiratory bronchioles, alveolar ducts and alveoli. Various pulmonary function tests (PFTs), spirometry, body plethysmography and the diffusion capacity, are used to measure ventilation and gas exchange. These measurements can be quantitative (normal/abnormal/change post-bronchodilator) or qualitative (i.e. intrathoracic or extrathoracic airway obstruction, intrapulmonary or extrapulmonary restriction) and together with clinical information, as well as biochemical, histopathological and radiological data, help to establish the diagnosis. The normal values depend on age, sex and height of the patient and are based on large populational studies of healthy never-smoking adults. The normal range of values is arbitrarily defined as a 90% confidence interval (the 90% of data points closest to mean), meaning that 5% of a normal population will fall above and 5% will fall below and this spread is described by the mean $\pm$ 1.645 RSD (relative standard deviation). RSD from the predicted mean is referred to as the standardised residual (SR). Traditionally, from 0 to – 1.645 SR is normal, – 1.645 SR is the lower limit of normal (LLN), from – 1.645 to – 2.5 SR represents mild deviation from normality, and from – 2.5 to – 3.5 SR moderate abnormality; beyond – 3.5 SR represents severe abnormality. Values of 120% and 80% of the predicted normal value are approximately equal to $\pm$ 1.645 SRs. The interpretative strategies of lung volumes are a subject of multiple guidelines and should incorporate clinical information [489]. Diffusion capacity assessed by single breath maneuver can be described by the following equations, which understanding is vital for correct interpretation of the results.

$$\frac{(KCO \times VA)}{Pb*} = DLCO$$

$$\frac{DLCO}{VA} = \frac{kco}{Pb*} = KCO$$

KCO (min^{-1}) is the rate of uptake of carbon monoxide (CO) from the alveoli during a 10s breath-hold at full inflation (to total lung capacity - TLC); KCO is a rate constant per unit pressure (efficiency index); *kco* is rate constant for CO uptake by alveolar capillaries; VA is the alveolar volume seen by CO during the breath-hold, VA is 91$\pm$8% (1.645 SRs) of

TLC [442], but decreases significantly in the presence of airflow obstruction, because of gas mixing delays during the short breath hold time; Pb* (barometric pressure minus water vapour pressure at 37°C) normalises the uptake (mmol or ml per min) per unit pressure (kPa or mmHg), so that the product, DLCO, has the units of conductance, mmol per min per kPa in SI units or ml per min per mmHg in traditional units. The values of DLCO need to be interpreted in conjunction with KCO and VA as primary measurements. A decrease in KCO is caused by diseased microvasculature with or without accompanying diffuse alveolar damage. Anaemia and a high carboxyhemoglobin secondary to smoking reduce the KCO, however, corrections can be made, and patients should be asked not to smoke for 12 - 24 h before testing [226].

The association between PAH and abnormal lung mechanics has been described for decades [133, 162]. Even in never-smokers, obstructive [133, 63, 485], restrictive [133, 63, 485] and mixed [249] ventilatory impairments have been reported. Although more prominent amongst patients with Eisenmenger physiology [63, 249, 326], such abnormalities were repeatedly described in patients with I/HPAH in the absence of concomitant parenchymal lung disease. Furthermore, conditions associated with PAH may predispose to underlying lung disease not considered sufficient to cause PH. For example, up to 90% of scleroderma patients exhibit evidence of interstitial changes on computed tomography (CT) scan, and 40 - 75% of these patients exhibit PFT abnormalities [490]. Thus, mild obstructive or restrictive lung abnormalities have been considered as part of the PAH phenotype. Therefore, pivotal randomised clinical trials included patients with forced expiratory volume in 1 second (FEV1) and total TLC as low as 50 - 55% and 60 - 70% of predicted, respectively (Table B.1). Conversely, patients with severe PH related to chronic lung diseases exhibit limited improvements with PAH-targeted therapies and poorer prognosis compared to IPAH, with a 3-year survival approaching 30 - 40% [58, 233, 216]. Additionally, the relationship between PFT abnormalities and pulmonary hemodynamic impairments is not straightforward amongst patients with chronic obstructive pulmonary disease and idiopathic pulmonary fibrosis, since severe PH has been described in patients with relatively mild obstructive/restrictive disease [462]. In contrast, decreased diffusion capacity for carbon monoxide (DLCO) is a characteristic feature of pulmonary hypertension with mean readings as low as $52(\pm 21)$% predicted [438, 232, 519, 533] and lower in males than females [438]. Reduced DLCO proved to be an independent predictor of survival, although it does not appear to influence response to treatment with PAH specific therapies [532, 550]. Low DLCO has also been found in patients with PVOD/PCH with or without biallelic *EIF2AK4* mutations [194, 293, 357]. Although some reduction in DLCO is one of the typical features of pulmonary hypertension,

PVOD patients show the lowest DLCO values when compared to IPAH or CTEPH. The DLNO/DLCO (diffusion capacity for nitric oxide to diffusion capacity for carbon monoxide) ratio in these patients is highest in PVOD, indicative of proportionally greater involvement of the pulmonary capillary compartment in the disease process [173]. In contrast to these findings, DLCO is relatively preserved in *BMPR2* mutation carriers [532] and shows intermediate values in *BMPR2* mutation carriers with coexisting *EIF2AK4* heterozygous mutations [119]. Several genome-wide association studies (GWAS) investigated the influence of common genetic variation of pulmonary ventilatory function and gas exchange. A recent study in individuals selected from extremes of the lung function distribution in the UK Biobank, increased the yield of independent signals for lung function from 54 to 97. The 97 signals showed enrichment in lung development, elastic fibers and epigenetic regulation pathways [558]. Another study [516] looked specifically into DLCO and DLCO/VA (KCO) and estimated that heritability of DLCO and KCO ranged between 23% and 28% in unrelated, and 45% and 49% in related individuals. The study identified a significant association between KCO and rs17280293 (*ADGRG6*). ADGRG6 is a G protein-coupled receptor, expressed in endothelial cells (ECs), which plays an important role in angiogenesis by controlling ECs proliferation, migration and tube formation. It regulates the expression of VEGFR2 via the transcription factors STAT5 and GATA2 [95]. The deletion of *ADGRG6* is embryonically lethal [559] or may result in postnatal pulmonary hypertension and early death [430]. Previous studies have also shown that variation at *ADGRG6* locus is associated with spirometric measures of lung function [200, 480]. Sakornsakolpat et al. [451] found genome-wide associations between DLCO and rs3009947 (*TGFB2*) and rs112878080 (*CHRNA3*). Thus, whether pulmonary function impairment in I/HPAH is coincidental, resulting from previous smoking or concomitant lung diseases, or rather represents a distinct phenotype associated with underlying genotype and specific outcome amongst patients with I/HPAH, remains debated and has been investigated in this thesis.

The rationale for patient stratification based on presence of small/coincidental heart defects

Patent foramen ovale (PFO) is a congenital interatrial connection that in the majority of cases closes at birth but in a substantial proportion of individuals may remain open into adulthood. The incidence of PFO in an otherwise healthy population ranges from 20 to 34%, with equal prevalence among sexes [195, 350]. Although most patients are asymptomatic, in some cases, PFO might be associated with cryptogenic stroke [276, 569], migraine [474], air embolism, decompression sickness and platypnea-orthodeoxia syndrome [189]. In clinical practice the

detection of PFO is possible by the means of transesophageal ECHO (TEE) in 72% and transthoracic ECHO in 57% of cases confirmed by RHC or autopsy [272]. In the context of I/HPAH small cardiac defects such as PFO or atrial septal defect (ASD) < 2 cm and ventricular septal defects < 1 cm are not thought to account for the development of elevated PVR and closing of these defects is contraindicated. Likewise palliative creation of interatrial right-to-left shunt via balloon atrial septostomy (BAS) has been reported beneficial in WHO functional class IV with right heart failure refractory to medical therapy or with severe syncopal episodes. BAS also proved helpful as bridging therapy in patients awaiting lung transplantation with unsatisfactory clinical response to medical treatment. BAS was shown to improve cardiac index (CI) and decrease right atrial pressure (RAP) with concomitant improvement in 6-minute walk distance (6MWD). The impact of BAS on long-term survival has not been established in randomised control studies (RCTs) [453, 279, 452]. Studies in PH reported prevalence of PFO ranging from 17 to 27%, which seems to be in keeping with the prevalence in the general population and does not impact on survival in patients with PH [385, 157, 464]. In these studies patients with PFO were younger at diagnosis and have higher RAP and mPAP and more dysfunctional RV [157, 464], which may suggest that in some cases the PFO was not congenital but rather reopened due to increased RA pressures. Although CHDs account for approximately 28% of all congenital anomalies [109] their genetic underpinnings remain poorly understood. CHD are frequently associated with major chromosomal syndromes [568], *de novo* mutations [586], aneuploidy and copy number variants [172] as well as uncommon syndromic forms such as *TBX5* for Holt-Oram syndrome [36, 55]. These genetic abnormalities, however, account for no more than 10% of CHDs. Several GWAS studies have been conducted to further elucidate the genetic component of CHDs. For ASDs 4p16 was identified as a risk locus [88, 593], for tetralogy of Fallot, regions of interest have been reported on chromosomes 1, 12 and 13 [88, 477], and for transposition of the great arteries, four SNPs on chromosome 20p12.1, pointing to the *MACROD2* locus, were detected [283]. Although no GWAS or rare variant association studies (RVAS) have been performed to specifically address the genetic architecture of PFO, several studies have suggested common genetic causes for PFO and accompanying disease which may be relevant to PAH. For example mutations in *ATP1A2* encoding Na^+/K^+-ATPase $\alpha 2$, a plasma membrane protein belonging to the P-type ATPase family responsible for maintaining the sodium and potassium gradients across cellular membrane, has been associated with hemiplegic migraine with or without PFO [156, 401, 261]. Previous studies have reported the presence and activity of the $\alpha 2$-subunit of the Na^+/K^+-ATPase in pulmonary vascular smooth muscle cells (PASMCs) [137, 338], and the decrease in expression and/or activity of different

types of K^+ channels in PASMCs of IPAH patients [62, 287]. Likewise, PAH has been linked to mutations in potassium channels such as *KCNK3* [322] and genes encoding ATP-ases such as *ATP13A3* [179] and *ATP1A2*, although the evidence for the latter association is much weaker [359]. Similarly, common variants c.677C>T and c.1298A>C in gene *MTHFR* encoding methylenetetrahydrofolate reductase frequently occur in people of Hispanic descent (25%) and North American Cauasians (15%) and have been linked to birth abnormalities, glaucoma, cancer and some mental health disorders. More specifically, a recent study showed an association between endothelial dysfunction (low L-Arginine/ADMA ratio used as surrogate) and impaired activity of *MTHFR* with the spectrum of interatrial septal defects [463]. To date, several clinical and translational research studies have investigated the hypothesis that *MTHFR* aberrations lead to endothelial dysfunction which in turn influences interatrial septal phenotype. For example higher plasma levels of homocysteine were found in PFO carriers [388], conversely lack of endothelial nitric oxide synthase was shown to affect normal postnatal heart development [300]. Although epidemiological data do not support higher incidence of PFO among patients with PAH, the common genetic predisposition to these two conditions is likely and has been investigated in this thesis.

The rationale for patient stratification based on the presence of autoimmune phenomena

Autoimmune disorders may affect connective tissue in a variety of organs such as joints, tendons, ligaments and blood vessels. WHO PH Group 1.4.1 consists of patients with PAH associated with connective tissue diseases (PAH-CTD) but autoimmune phenomena are increasingly recognised in patients with IPAH. This is suggested by a high prevalence of autoimmune thyroid disease among IPAH patients [412], infiltration of macrophages and lymphocytes in remodelled pulmonary arteries [546], presence of tertiary lymphoid tissues in the lungs [400], circulating autoantibodies [191, 509, 508] and inflammatory mediators [502]. Although several types of autoantibodies have been reported in IPAH and PAH-CTD, their overall prevalence and impact on disease pathogenesis and progression remain mostly unexplored beyond isolated reports and small studies. Additionally, impairment of BMPR2 signalling pathway has been seen in patients with HPAH and IPAH irrespective of mutation status. Presence of circulating autoantibodies to BMPR2 has been previously excluded in patients with mixed connective tissue disease with PAH [455]. In the PH spectrum, the role of circulating autoantibodies has been mostly researched in the context of connective tissue diseases. In most instances, the role of single antibodies was explored rather than panels. Autoantibodies against nuclear antigens, ECs and fibroblasts were most

commonly investigated. Anti-U1 ribonucleoprotein (RNP) antibodies were found to be associated with a distinct phenotype and better prognosis in patients with systemic sclerosis-associated PAH (SSc-PAH) [475]. A clinical study on 84 anti-centromeric proteins (CENP)-A positive patients with SSc found that different antibody subspecificities showed associations with clinical features and risk of developing PH [399]. Anti-SSA/Ro autoantibodies in high titers were also seen in patients with IPAH without a diagnosis of connective tissue disease suggestive of possible role of these antibodies in the pathogenesis of PAH [191]. Inflammatory injury to the vascular endothelium is believed to be the mainstay of PAH pathobiology. Autoantibodies against ECs have been reported in SSc-PAH and correlated with digital ulcerations highlighting their deleterious impact on the microvasculature. Antibodies against endothelin receptor type A (ETRA) were suggested as a mechanistic biomarker of PAH in systemic lupus erythematosus (SLE) [192] and SSc [39]. Fibroblasts lead to pulmonary arteries remodelling through the collagen synthesis. IgG and IgM anti-fibroblast antibodies were found in PAH patients with distinct reactivity patterns in IPAH and SSc-PAH [218]. Table 1.4 summarises different types of auto-antibodies found in PAH, their role and association with survival.

Genetic predisposition to producing autoantibodies and/or to autoimmune diseases is an area of ongoing research, initiated primarily by a frequent phenomenon of familial autoimmunity [14]. The Human Leukocyte Antigen (*HLA*) locus, located on the short arm of chromosome 6 at 6p21.3 [185] plays a major genetic role in autoimmunity. Evidence suggests that specific *HLA* alleles might enhance autoimmune response and predispose to autoimmunity throughout the expressed repertoire of T cells [10]. To that end a plethora of *HLA* alleles have been found to be associated with autoimmune conditions [539]. Further research also established that several other genes at the major histocompatibility complex (MCH) and non-MCH may contribute to the development of autoimmunity. The evidence so far indicates that: (1) the genetic predisposition to autoimmune phenomena is polygenic, (2) the functional variations are in both coding and non-coding regions, (3) most of the associations are modest and (4) explain predisposition as well as protection [15]. While GWAS has established disease loci that can be grouped into meaningful mechanistic pathways such as regulation of interferon pathway (*IRF5*, *IRF7*, *IRF8*, *ITGAM*), B-cell signaling (*BANK1*, *BLK*, *LYN*, *PLCG2*), antigen presentation (*HLA*, *DEC205*) and B and T-cell activation (*PTPN22*, *TYK*) [190], even when combined these variants are responsible only for a small fraction of known genetic liability, this suggests a potential role of rare pathogenic variants. Indeed several highly penetrant forms of autoimmune diseases have been reported [11], although some studies dispute the overall rare variant contribution to missing heritability of autoim-

mune diseases [230]. No study has examined the association of rare variants with a specific phenotype of PAH with autoantibody positivity.

Table 1.4 Different types of auto-antibodies found in PAH.

Autoantibody	PAH type or sub-type	Prevalence of PAH in patients with antibody positivity; % (total number)	Prevalence of ILD in patients with antibody positivity % (total number)	Role/Correlation with clinical parameters	Association with survival	Evidence
Anti-ETRA	SLE-PAH	-	-	Proliferation of SMC, disruption of endothelial barrier, upregulation of PAH-related markers, aggravation of right ventricle hypertrophy	NA	(Guo et al. 2015; Becker et al. 2014) [192, 39]
Anti-ETRA	SSc-PAH	-	-	Increased vascular endothelial reactivity and induction of pulmonary vasculopathy	Yes	(Becker et al. 2014) [39]
Anti-centromere **ANA** **Scl-70** **Anti-U1 RNP** **RNA pol III**	SSc-PAH	37% (162) 24% 7% 5% 6%	NA NA NA NA NA	-	No	(Hinchcliff et al. 2015) [211]
Anti-EC	SSc-PAH	-	-	Activation of EC, expression of adhesion molecules, trigger apoptosis, correlated with PAH and digital ulcerations	NA	(Negi et al. 1998) [375]
Anti-fibroblasts	SSc-PAH, IPAH	-	-	Activation of fibroblasts, induction of collagen synthesis	NA	(Negi et al. 1998) [375]
Anti-CENP-A	SSc-PAH	4.7% (84)	20.3% (84)	Subsepcificities of anti-CENP-A antibodies associated with clinical features in opposite ways	NA	(Perosa et al. 2016) [399]
Anti-SSA/Ro	IPAH	-	-	Case report of 53-year-old female of African American descent with high titers of Anti-SSA/Ro and no signs or symptoms of autoimmune disease but PAH	NA	(Guerreso and Conner 2016) [191]
Anti-U1 RNP	CTD-PAH, SSc-PAH	11% (342)	NA	Associated with younger age and better functional status, not associated with haemodynamics	Unclear	(Sobanski et al. 2016) [475]
Anti-U3 RNP	SSc-PAH	31% (86%)	36% (97)	Anti-U3 RNP antibodies more common in males of African American origin and associated with younger age, PAH and skeletal muscle involvement	Yes	(Aggarwal et al. 2009) [5]
Anti-centromere	lcSSc	28% (87) 19% (306)	48% (87) 13% (306)	Patients with Anti-Th/To more often had PAH and parenchymal changes than patients with anti-centromere antibodies. Patients with Anti-Th/To antibodies had reduced survival in comparison with anticentromere positive patients	Yes	(Mitri et al. 2003) [355]

Abbreviations: Anti-ETRA: endothelin receptor A antibody; ANA: antinuclear antibodies; scl-70: antibody to topoisomerase I; Anti-U1 RNP: small nuclear ribonucleoprotein that associates with U1 spliceosomal RNA antibodies, RNA pol III: RNA polymerase III antibodies; Anti-EC: endothelial cells antibodies; Anti-CENP-A: anticentromere A antibody; Anti-SSA/Ro: Anti-Sjögren's-syndrome-related antigen A; SLE: systemic lupus erythematosus; PAH: pulmonary arterial hypertension; SSc: systemic sclerosis; lcSSc: limited cutaneous systemic sclerosis; CTD: connective tissue disease

The rationale for stratifying patients by age

It is estimated that between 2000 and 2050, the proportion of the world's population over the age of 60 years will double from 11% to 22%, with the absolute number increasing from 605 million to 2 billion [332]. The change in the population demographics is likely to impact on epidemiology of both common and rare conditions, in fact recent registries have shown a considerable shift in PAH demographics [517, 312]. Particularly in the ageing Western populations, PAH is now diagnosed in elderly patients, with a significant burden of comorbidities, a weaker response to treatment and poorer survival [218]. While rare diseases that manifest early in life are more likely to be mendelian disorders, the late-onset diseases are predominantly polygenic. The genetic causes for the latter group are difficult to discern due to the complexity of genetic mechanisms involved as well as the interaction between these mechanisms and the environment (see Chapter 1.1.3).

GWAS have highlighted several genes to be implicated in ageing and age associated cardiovascular diseases and further functional studies have followed. Among others, apolipoprotein E (*APOE*) [374], methylenetetrahydrofolate reductase (*MTHFR*) [107], nitric oxide synthase 3 (*NOS3*), transforming growth factor β1 (*TGFB1*) [376], insulin (*INS*) [336], vascular endothelial growth factor A (*VEGFA*) [584], and calmodulin 1 (*CALM1*) [307] have been reported. Additionally, ageing is associated with the accumulation of molecular damage resulting in changes in gene expression, epigenetics, and structural DNA changes such as telomere shortening [66]. The molecular damage is likely to have both intrinsic and environmental causes. Interestingly, expression and activity of telomerase reverse transcriptase (*TERT*) which not only regulates telomere length but also cell growth was found to be increased in lungs from patients with idiopathic PAH and from mice with PH induced by hypoxia or serotonin-transporter overexpression (SM22-5HTT(+) mice), mainly within PA-SMCs. The *TERT* inhibitor imetelstat, reduced hypoxia-induced PH in wild-type mice or partially reversed established PH in SM22-5HTT(+) mice while simultaneously decreasing *TERT* expression [369]. Similarly, somatic mutations in Tet-methylcytosine-dioxygenase-2 (*TET2*), a key enzyme in DNA demethylation, occur in cardiovascular disease and are associated with clonal hematopoiesis, inflammation, and adverse vascular remodelling. In the context of PAH, germline and somatic mutations were found in individuals with clinically idiopathic disease without mutations in known PAH risk genes. Patients with *TET2* mutations were significantly older, did not show response in acute NO challenge test, had lower PVR and increased inflammatory markers (IL-1β) when compared with those without the mutation [415]. While some complex traits (such as height) may vary little in their genetic architecture over time [61] an alternative scenario in which a different set of genes is contributing to the

variance in phenotype over time is more likely in case of rare phenotypes and warrants age stratification for causative gene discovery.

The rationale for stratification based on survival

Related to the age at disease onset and response to therapy is survival. As mentioned before many studies have investigated the impact of various disease features on survival but less is known about genetic underpinnings of survival in PAH patients. A recent GWAS in I/HPAH has shown that the *HLA-DPA1/DPB1* rs2856830 genotype was not only associated with I/HPAH but also with a beneficial effect on survival (CC genotype 13.5 vs TT genotype 6.97 years) irrespective of baseline disease severity, age and sex [436]. No genetic study in PAH focused on sampling (investigating) longevity extremes has been conducted in this patient population.

1.2.2 The rationale for using computationally derived phenotypes

The role of computer algorithms and early artificial intelligence (AI) information systems in medicine, especially in clinical decision making, have been under exploration since the 1960s. Since then, advances in AI, machine learning and deep learning have revolutionised diagnostic and prognostic approaches, particularly in radiology and cardiovascular medicine, thanks to the unparalleled speed and precision in recognising patterns in large data sets. John McCarthy coined the term AI in 1956 to name "the science and engineering of making intelligent machines". Nowadays, AI is usually defined as a computer system able to perform tasks that normally require human intelligence, such as visual perception, speech recognition, decision-making, and translation between languages and can be understood as an umbrella term encompassing machine learning and deep learning in a fashion resembling a set of Russian dolls nested within each other (Fig. 1.6).

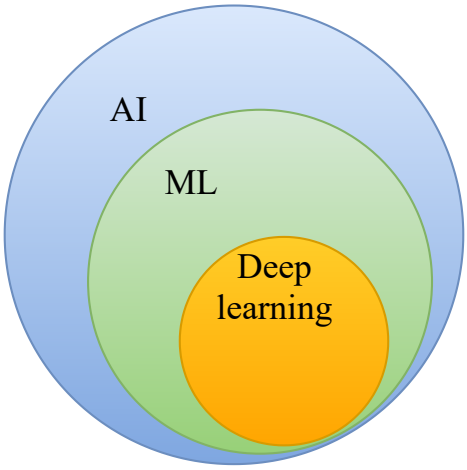


Fig. 1.6 Schematic representation of the relationship between artificial intelligence (AI), machine learning (ML) and deep learning.

Machine learning is a subset of AI defined as a "field of study that gives computers the ability to learn without being explicitly programmed" [27]. Finally, deep learning is a subset of machine learning referring to deep artificial neural networks, where "deep" is a technical term referring to the number of layers in a neural network. Deep artificial neural networks are a set of algorithms that have set new records in accuracy for many important medical problems. For example, neural network classifiers were shown to facilitate the detection of congestive heart failure patterns on chest radiographs (CXR) and highlight relevant abnormal features in CXR [461]. In echocardiography, machine learning approaches have been employed in automatic calculations of aortic valve area in aortic stenosis or to help differentiate prognostic phenotypes [409]. In the era of wearable devices, the automated analysis of ECG gained traction. A study by Turner et al. used a deep learning algorithm to detect arrhythmia trained on over 4000 ECG Holter recordings, including rare conditions and showed a correct recognition rate of 98.5% and accuracy of 92% [537].

Along with diagnostic applications, risk assessment and predictions are a mainstay of personalised medicine. Przewlocka-Kosmala et al. used an ML classifier to identify prognostic phenotypes among patients with heart failure and preserved ejection fraction based on pre- and post-exercise ECHO imaging [419]. Another study employed ML to devise a cardiovascular risk prediction score that outperformed Framingham score and discovered new cardiovascular risk predictors and previously unknown interactions between different individual features [8]. Unsupervised ML has also allowed clustering patients with pulmonary hypertension into distinct, more homogenous phenotypes beyond diagnostic classification, and has highlighted additional features impacting survival [132]. Importantly, several AI

algorithms have already been approved for medical use by the American Food and Drug Administration with imaging and cardiovascular algorithms being the most represented[2].

One of the main aims of genomics is to describe gene function by identifying the genotype-phenotype associations. This may have diagnostic and prognostic implications and an impact on the choice of treatment modality (pharmacogenomics). While large sequencing efforts provide ever-improving quality data, multidimensional clinical and biological data sets for genetic studies still lag behind. The most common limitations are missingness, variable phenotype depth and variable coding approaches which render the data not always amendable for computational analysis. Multiple approaches have been deployed to close this gap.

Along with biobank phenotypic proformas, electronic medical records (EMR) are frequently linked to biobank data to enhance clinical information depth and breadth. To this end ML algorithms are often employed to recover disease phenotypes from structured (SNOMED-CT, ICD10 codes, LOINC terms) and unstructured (free text) EMR. When deep phenotyping (precise and comprehensive description of phenotypic abnormalities) is required, the Human Phenotype Ontology (HPO) is being increasingly adopted. The HPO is a structured, comprehensive, controlled vocabulary database designed to fully annotate human disease phenotypes. The HPO has been used for developing algorithms and computational tools for differential diagnostics [270, 37, 476], for the prioritisation of candidate disease-associated genes in exome sequencing studies [469], and has been employed for translational research, including pharmacogenomics [178]. In summary, using computational methods to capture composite phenotypes is a valid strategy for genetic complex trait research and has proved superior to various single trait approaches [310].

1.3 Forward genetics

1.3.1 Concepts

Forward genetics is a classic molecular genetics approach used to elucidate the genetic underpinnings of a mutant phenotype of interest [403]. Forward genetics is typically considered a "phenotype to genotype" approach as mutant phenotypes are first observed before their corresponding genes are identified (Fig. 1, Process 2). In humans, forward genetics approaches most commonly include family-based linkage studies, GWAS and RVAS.

[2]https://medicalfuturist.com/fda-approvals-for-algorithms-in-medicine

1.3.2 Methodology

Study Design

The two main approaches for studying the underlying genetics of PAH are family-based studies and case-control studies. The former is based on studying inheritance patterns of genetic polymorphisms, the second involves comparing genotype frequencies between cases and controls. Family-based studies are effective when parental samples along with phenotype information are available and the disease in question has a high penetrance; they are particularly useful for studying dichotomous traits and are robust to population stratification [244]. Case-control studies are a viable alternative if the above criteria are not met, although they have their own challenges which need to be addressed. To name the most important:

1. Selection of cases (recognition of selection bias, incident vs. prevalent cases recruitment)
2. Case definition (precise definition of the phenotype that can be ascertained in a research setting)
3. Selection of controls (healthy vs. disease controls, matched in respect to age, sex and ethnicity, having a comparable evaluation of presence or absence of the phenotype in question)

Power calculations in genetic studies are an absolute necessity as ignoring this basic step can lead to both underpowered (risking false rejection of null hypothesis and characterised by wide sampling distributions for sample estimates) and overpowered (wasteful and often unethical) experiments. The factors which are limiting statistical power to detect new genotype-phenotype associations are phenotypic variance, phenocopies, the effect size of risk alleles and minor allele frequency (MAF), with the last two factors driving the difference of sample sizes between GWAS and RVAS.

Statistical Methods

Prior to the widespread use of GWAS, the most important tool in genetics were linkage studies in families. These were particularly useful in single-gene disorders in which implicated genes have large effect sizes. GWAS on the other hand compares the frequency of common SNPs between unrelated cases and controls. The associated SNPs are then considered markers of relevant regions that influence the risk of the phenotype. In fact, power calculations

provided evidence that GWAS are better than linkage studies at detecting variation with small effect sizes [441]. Multiple statistical methods can be applied in GWAS, for example, Pearson X^2 test, normal approximation to Fisher's Exact Test, logistic regression, categorical model tests, Cochran–Armitage Trend test, and allele tests. The best method depends on the mode of inheritance and trait frequency. Importantly, the assumptions used in various tests may differ; these assumptions directly impact the results, as tests that assume the same mode of inheritance should yield the same results (i.e., Cochran–Armitage Trend Additive test, Logistic Regression Additive test and Allele test). Due to a large number of comparisons, adjustment for multiple testing is necessary; therefore, the p-value threshold is Bonferroni corrected (which encourages high type II errors). Additionally, genotype and phenotype misclassification errors impact on power in GWAS. Epistatic scenarios and modelling gene–environment interactions require yet another set of methods and are computationally challenging although feasible [344]. GWAS is unsuitable for single-variant testing due to the potentially low prevalence of mutation carriers and small effect size, which would both require unfeasibly large sample sizes. Instead, gene and region-based aggregation approaches have been developed which compare mutation frequencies between cases and controls within the boundaries of the gene. These techniques are appropriate when different variants exhibit an equal risk of disease and thus have the same phenotypic impact. For instance, several variants may result in loss-of-function (LoF, e.g., nonsense, frameshift, essential splice site), and thus analysis would determine the association by counting the presence of LoF variants between cases and controls. Prior to association testing, quality control and filtering methods are utilised, namely sequencing quality scores, MAF filters [404] (usually MAF of 1:10,000 for autosomal dominant disorders and MAF of 1:1,000 for autosomal recessive disorders) and *in silico* predictions. Predictions include deleteriousness scores for missense variants such as PolyPhen-2 [4], Sorting Intolerant From Tolerant (SIFT) [4, 470], and rare exome variant ensemble learner (REVEL) [242], conservation scores such as Genomic Evolutionary Rate Profiling (GERP) [85], PhyloP [410] or PhastCons [468], or the Combined Annotation Dependent Depletion (CADD) score [267], which combines several metrics in one score. Analysis of the protein-coding region, consisting of $\sim$20,000 genes, requires adjustment for multiple-testing. This can be done using the Bonferroni correction, where $\alpha = (0.05/20{,}000) \approx 2.5 \times 10^6$. Where several models are applied, this adjustment must be made more stringent by dividing by the number of models tested. Region-based collapsing approaches hinge on the notion that different regions within genes may vary in their tolerance to missense variation. An alternative approach, particularly useful in smaller studies, is collapsing variants that belong to the same gene set (i.e., genes that belong

to the same pathway). Candidate gene testing is a powerful approach to avoid overcorrection and, therefore, false-negative results. This proved useful when investigating members of the transforming growth factor-β (TGF-β) pathway, such as *SMAD9* [371], which did not reach statistical significance in the exome-wide analysis [179]. More recently, the same approach revealed an association between *TET2* and PAH, which was further supported by experimental evidence [415] but did not reach exome-wide significance [598]. Complex genetic models such as recessive inheritance pose additional challenges. In the recessive mode of inheritance, the MAF threshold must be more lenient as heterozygotes are unaffected (higher MAF in reference populations); also, variants in *cis* configuration (affecting the same allele) might be wrongly counted (as oppose to trans variants, which are those present on opposing alleles). Similarly, testing for digenic inheritance is particularly problematic due to the large number of possible combinations requiring testing and adjusting for [417].

A number of statistical methods have been developed to test for rare variant associations. Burden tests [20, 367, 306] aggregate the information found within a predefined genetic region into a summary dose variable. In weighted burden tests [334], variants are weighted according to their frequency or functional significance. Adaptive burden tests [243] aim to account for bidirectional effects by selecting appropriate weights. Variance component (kernel) tests such as (Sequence) Kernel Association Test (SKAT) [578] allow to test risk and protective variants simultaneously but are underpowered when most variants are causal, and effects are unidirectional. Omnibus tests such as SKAT-O [297], which combines burden tests with the variance-component test, might be particularly useful when there is little knowledge of the underlying disease architecture. In addition to frequentist approaches, a Bayesian statistical framework offers a robust alternative. Bayesian model comparison methods such as BeviMed [183] allow for the testing of associations between rare Mendelian disease and a genomic locus by comparing support for a model where disease risks depend on genotypes at rare variant sites in the locus and a genotype-independent "null" model. The prior probability in such models can vary across variants (reflective of external biological information, i.e., depending on MAF, conservation scores, gene ontologies, expression in the tissue of interest) or be constant for all genes/variants reflecting the prior belief of the overall proportion of variants that are associated with a given phenotype.

Last but not least, an essential step in rare variant discovery is to ascertain the pathogenicity of a given variant and its causative role in the disease. Not all damaging variants are pathogenic and *in silico* approaches alone are not enough to predict if the variant is disease-causing [120]. Viability and phenotyping inferred from knockout mice screens, as well as essentiality screens on human cell lines, may further help predict variant impact [65]. To aid

both research and clinical decision making, the American College of Medical Genetics and the Association for Molecular Pathology (ACMG) issued recommendations that combine and weigh the computational, functional, population and clinical evidence to determine pathogenicity [440]. Other initiatives such as ClinGen and ClinVar aim to define the clinical relevance of genes and variants reported in the literature for use in precision medicine and research [431].

Molecular Genetic Techniques

Molecular genetics techniques used for genetic diagnosis, including the detection of specific gene mutations and copy number variants, have been recently summarised [177]. Traditional methods used to identify candidate genes involved in the pathogenesis of PAH include linkage analysis, but more recently next-generation sequencing (NGS) has taken centre stage. The advent of NGS technologies has opened a plethora of opportunities both for clinical diagnostics and research. Such technologies that fall under the umbrella of NGS include targeted panel sequencing, whole-exome sequencing (WES) and whole-genome sequencing (WGS). Nevertheless, the need for more conventional methods such as Sanger sequencing and multiplex ligand-dependent probe amplification (MLPA) remains. Disease-specific panels include a set of genes or regions of genes that are known to be causative of a specific phenotype. This is particularly beneficial in the clinical context when assessing highly heterogeneous traits, such as intellectual disability. Although these panels are not always consistent across laboratories, efforts are being made to produce guidance around their design and development [38]. Targeted panel testing has been introduced in PAH including known and candidate disease genes [481]. WES includes $< 2\%$ of the genome i.e., the protein-coding regions only. This method is clinically useful given that 85% of all described disease-causing sequence variants are in this region [425]. For diseases that are more genetically heterogeneous, WES has proven to be a fruitful method, especially when incorporating segregation analysis, which increases the diagnostic yield from 23.6% in probands to 31% in child–parent trios [433]. WES has been used for both the identification and discovery of candidate genes in PAH and has been applied to family-based [25] and case-control studies [599]. Limitations of WES include poor coverage of some exons such as GC rich regions and low confidence to identify structural variation [587].

WGS is a high-throughput sequencing technology predominantly used in the research setting [538]. WGS is massively parallel, short reads from extracted DNA fragments are aligned against a reference to form a contiguous sequence. The cost of this technology is

halving approximately every two years[3]; however, the $1,000 genome often referred to is still some way off unless sequencing is performed at scale [460]; as an example, WGS is currently being introduced to the UK National Health Service in collaboration with Genomics England Ltd.

European Respiratory Society and European Society of Cardiology (ERS/ESC) guidelines published in 2015 recommended sequential testing starting with bone morphogenetic protein receptor type 2 (*BMPR2*) sequencing and MLPA in patients with sporadic/familial PAH and *EIF2AK4* sequencing in sporadic/familial PVOD/PCH [153], and this approach has been successfully used in many clinical and research settings across the world [35]. With the decreasing cost of WES/WGS, a common practice is now that of virtual panel testing whereby a selected number of genes are chosen for bioinformatics analysis based on the individual's phenotype. This also allows for data reanalysis when a novel disease-gene is identified and/or another condition is suspected (emerging phenotype over time). In the UK, this is coordinated at a national level via the Genomics England PanelApp tool; virtual disease gene panels applied to WGS data are continuously curated and include a PAH panel[4]. As more patients with PAH are being tested via NGS methods, a diagnostic benefit is starting to emerge.

Despite the benefits of NGS technologies, there are some challenges that require attention and systematic solutions, among these, storage and handling of big data remains a significant consideration, as well as the management of incidental findings and the reporting of variants of unknown significance (VUS).

Along with the growing number of research projects using NGS to uncover the genetic basis of various diseases, there has been an ongoing effort to aggregate and harmonise WES and WGS data from large-scale disease and population projects and to make them publicly available as a reference variome. This started in 2012 with a founder project called Exome Aggregation Consortium (ExAC) which harvested WES data from over 60,000 individuals. This was followed by The Genome Aggregation Database (gnomAD), of which three versions have been released so far, covering 71,702 genomes from unrelated individuals aligned against human genome assembly 38 (Genome Reference Consortium Human Build 38 - GRCh38 (v3)) [254]. Also, in 2012, came the announcement of the 100,000 genomes project by the UK Government. The project sequenced the genomes of 100,000 NHS patients with particular focus on those with rare disease(s) and cancer. Another useful resource is the Trans-Omics for Precision Medicine (TopMed) program, which aims to

[3]https://www.genome.gov

[4]https://panelapp.genomicsengland.co.uk/panels/193/

sequence over 120,000 well-phenotyped individuals as well as collect other omics datasets [507]. The assertion of ethnic diversity is an important consideration, and several initiatives such as KoVariome [265], Genes and Health [136] and BioBank Japan[5] are addressing this issue. Furthermore, new disease cohort genomic databases are being established. Examples include the NIHR BioResource Rare Disease Study (NBR) [538], the Inflammatory Bowel Disease BioResource[6], Genetic Links to Anxiety and Depression (GLAD)[7] and the Eating Disorders Genetics Initiative (EDGI)[8]. The advantages of these data sets are numerous; first, they provide allele frequencies for diverse populations, second, they help to address the overestimation of disease penetrance arising from the historical focus on multiplex pedigrees [142], and third, through acknowledging variable penetrance, they help to identify genetic and environmental disease modifiers [69].

Reference Genome

The Human Genome Project was completed in 2003 [84, 240] and since then, successive iterations of the human reference genome have been published, updated, and refined by the GRC. Recent versions include GRCh37 (hg19) and GRCh38 (hg38) released in 2009 and 2013, respectively. These are both composite genomes, i.e., derived from the sequence of several anonymous donors; the make-up of these two assemblies is largely similar, with approximately 93% of the primary assembly composed of sequences from 11 genomic clone libraries.

To date, most large-scale PAH studies have aligned their data to the GRCh37 reference genome (Table B.2), with only a couple of recent studies using GRCh38 as a reference [598, 415]. Even though GRCh38 was released seven years ago, the transition from GRCh37 to GRCh38 has been a long process. The comparison of the two reference panels demonstrated that GRCh38 provides a more accurate analysis of human sequencing data due to the improved annotation of the exome and the additional reads aligned to GRCh38, which indicate better structural and sequence representation [193]. In addition, Pan et al. [389] noted that GRCh38 had better genome coverage, with a 5% increase in the number of SNVs identified. In comparison to GRCh37, GRCh38 altered 8000 nucleotides, corrected several misassembled regions, filled in gaps, and increased the number of genes and protein-coding transcripts [193]. Additionally, GRCh38 is the first human reference genome to

[5] http://biobankjp.org
[6] https://www.ibdbioresource.nihr.ac.uk
[7] https://gladstudy.org.uk
[8] https://edgiuk.org

contain sequence-based representations for the centromeres [457]. Whilst GRCh37 is a single representation of multiple genomes, with only three regions containing alternative sequences UDP-glucuronosyltransferases 2B subfamily (UGT2B) on chromosome 4, the major histocompatibility complex (MHC) region on chromosome 6, and the *MAPT* gene on chromosome 17) [81], GRCh38 includes 261 alternate loci across 178 genomic regions, providing a more robust representation of human population variation [457]. The increased level of alternative sequence representation requires new analysis methods to support their inclusion yet at present, most tools and pipelines do not make use of these [81]. Despite the advantages, GRCh38 still contains gaps and errors at repetitive and structurally diverse regions [291]. Additionally, as it is a mosaic haploid representation of the human genome [457], in which poor alignment can affect the detection of alleles in regions of high variation, such as the MHC locus and KIR, it is unlikely to truly represent human diversity [291]. It does, however, provide the starting point for a more inclusive population-based reference genome, or pan-genome [309] and as such, will play an evolving role in the generation of individual diploid genome assemblies and graph-based representations of genome-wide population variation [391, 426, 31], thereby providing unique opportunities for data analysis.

1.4 Reverse genetics

1.4.1 Concepts

Reverse genetic techniques are the experimental approaches that involve targeted modifications to specific genes in order to analyse phenotypic impact (Fig. 1, Process 3) and are critical in validating *in silico* predictions. Most deleterious mutations found in PAH patients are LoF, suggesting a knockout or knockdown approach to analysing the phenotypic effect of a mutation. The main benefit of direct gene mutagenesis is the permanence and potentially ubiquitous effect of that alteration. In comparison, gene knockdown is a transient change that can be created after the developmental stage in the animal life cycle. This is particularly useful for understanding LoF in genes that are embryonically lethal [44]. Conditional and inducible systems are also excellent means of circumventing this issue, permitting temporally and spatially controlled gene disruption [207, 223, 92, 415]. Inducible systems display higher efficiency and limited side effects compared to stably-expressed mutations and have the added benefit of reversibility [252].

1.4.2 Methodology

Techniques for genetic modification have permitted targeted genetic analysis using *in vitro* and *in vivo* models of PAH. Modification of a specific gene, through mutagenesis or knockout, can be carried out using a wide range of tools, including Zinc-finger nucleases (ZFNs) [545], transcription activator-like effector nucleases (TALENs) [251], and CRISPR/Cas9 [427]. Similarly, transient genetic changes can be achieved through gene expression knockdown, such as RNAi [264], or using conditional/inducible expression systems, such as Cre/LoxP and Flp/FRT sites or the Tet inducible system [304]. Generating models with specific mutations is also critical for the identification of therapeutic targets and drug testing. Specific techniques and their utility in animal models have been extensively reviewed [551, 304, 547].

In vitro

As PAH is a disease of the pulmonary circulation, most studies seek to model genetic changes in physiologically relevant cell lines, including, but not limited to, pulmonary artery endothelial cells (PAECs), PASMCs, blood outgrowth endothelial cells (BOECs) [530] and human umbilical vein endothelial cells (HUVECs). PAH is characterised by molecular and cellular deviations in pulmonary vascular function [365, 228], including hyperproliferation [67, 585, 583, 428], impaired migration [159, 449, 67] and aberrant apoptosis [316, 583]. Traditionally, research has focused on using 2D cellular monolayers for analysis of these features; however, with rapidly evolving technologies, future analyses may utilise multi-layer, 3D models such as organoids [524] and hydrogels [102] to better replicate the pathobiology of PAH.

In vivo

Understanding the pathobiology of PAH is dependent upon robust animal models for functional analysis and exploration of potential therapeutics. To this effect, a wide range of animal models have been developed across species, including pig, sheep, dog and most commonly rodents [384]. The most widely accepted rat models for PAH are pharmacologically induced, namely monocrotaline (MCT) and Sugen 5416-hypoxia (SuHx) models [325]. MCT is a toxin that induces damage to PAECs, causes RV and pulmonary arterial medial hypertrophy, and alters pulmonary artery pressures [175, 7]. In the SuHx rat model, SU5416, a Vascular Endothelial Growth Factor Receptor (VEGFR) antagonist that promotes endothelial cell apoptosis and smooth muscle cell (SMC) proliferation is administered, followed by a period of hypoxia [325], which results in vascular changes that model human IPAH [513].

The downside of both above-mentioned models is the potential introduction of non-PAH related symptoms such as alveolar oedema, pulmonary vein occlusion, acute lung injury, liver toxicity and emphysema [271, 340].

The chronically hypoxic mouse model is also widely used in PAH research, with mice housed at 10% oxygen for a variable length of time [506, 321, 325]. While mice display changes in pulmonary artery pressure, they display limited vascular remodelling, which consists of muscularization of vessels and medial hypertrophy of muscular resistance vessels [491]. Alternatively, pulmonary artery banding (PAB) has been used in various species to increase pulmonary artery pressures and induce RV hypertrophy. The main benefit of this technique is the lack of pharmacologic or environmental modification of the animal [6]. In practicality, the popularity of each model varies across PAH research groups; a recent poll showed that two-thirds of groups used two or more models [295], suggesting that no one model truly replicates the disease.

In comparison, genetic models tend to be more favourable as they are free from the systemic effects of pharmacological models. Several heterozygous *BMPR2* mutant models have been developed. While some models have shown elevated mPAP and PVR compared to WT littermates [44], others did not replicate these findings [482, 317]. Alternate studies have shown the development of pulmonary vascular lesions [570] and often, an environmental hit may be required for development of PAH [525, 482, 317]. More pronounced phenotypes are seen in homozygous deletions; as *BMPR2* is critical for embryogenesis, conditional deletion in pulmonary endothelial cells has been used and showed incomplete penetrance of the mutation, with a subset of mice displaying elevated right ventricle systolic pressure (RVSP), RV hypertrophy and inflammation [223]. The following genetic models have also been developed: *AQP1* (*Aqp1* null mice [324] and *Aqp1* with a COOH-terminal tail deletion [585]), *TET2* (conditional, haematopoietic heterozygous and homozygous deletion in mice, generated using the Vav1-Cre system) [415], *BMPR1a* (several conditional knockout mice models) [296, 298, 121], *SOX17* (conditional deletion in mice, using Demo1-Cre in descendants of the splanchnic mesenchyme) [289], *UCP2* (knockout mouse model created using a plasmid vector carrying modified Ucp2) [588], *CAV1* (modified mouse cav-1 targeting construct) [594], *EIF2AK4* (modified mouse *Gcn2* targeting construct) [590], *SIRT3* (heterozygous and homozygous mice, generated using targeted embryonic stem cell clones harbouring a mutated *SIRT3*) [392], *ATP13A3* (generation of a protein truncating mutation in mouse *Atp13a3* using CRISPR/Cas9) [299], *SMAD9* (Disrupted Smad8 allele generated in mice using LoxP/cre and Frt sites) [225], *KCNK3* (heterozygous and homozygous rats harbouring a CRISPR/Cas9-generated exonic, inframe deletion) [286] and *SMAD1* (mice with *Smad1*

deletion in endothelial cells or SMCs, generated using L1Cre or Tagln-Cre, respectively) [199], *KDR* (cell-specific conditional deletion of *Kdr* in mice) [574].

In humans, the severity of disease corresponds to the extent of pulmonary vascular changes and impairment of RV function. However, rodent models may display milder and isolated symptoms of PAH and, thus, require thorough characterisation to validate the model. Echocardiography is useful for the assessment of RV function but estimates of pulmonary pressures are inferior to direct measurements obtained through RHC [325]. Alternatively, cardiac MRI is another noninvasive tool for assessment of RV function; it supersedes echocardiography in its ability to produce high quality, three-dimensional images from which the RV can be directly measured. However, it also has limited availability and bears higher running costs [395]. Close-chested RHC is preferred over open-chested RHC, as it is less invasive and preserves the negative intrathoracic pressure associated with breathing. Open-chested RHC is usually a terminal procedure, thus preventing longitudinal assessment of rodents; however, it also permits a more comprehensive assessment of haemodynamic characteristics and analysis using pressure–volume loops. Tissue harvesting permits further characterisation of the heart and lungs, including evaluation of RV hypertrophy [325].

It is important that measures are taken to recognise and limit bias in experimentation and phenotyping. This includes ensuring randomisation of groups when carrying out tests that require a control and treatment group, such as selecting mice to be pharmacological models or therapeutic treatment. Additionally, mice should be matched for all possible qualities, including strain, age and sex. During phenotype assessment, researchers should aim to avoid implicit bias by using blinded observers during all procedures. Furthermore, the data collected must be comprehensive and relevant for PAH, including RHC data and histological analysis [418].

1.5 Forward and reverse genetics studies in PAH

1.5.1 Rare genetic variation

Soon after describing sporadic PPH in the early 1950s, the first case of familial PPH was reported [113], and other confirmatory publications followed [320]. Subsequently, linkage studies identified a region on chromosome 2q31-32 [380, 368] associated with familial cases. Over the last two decades, forward genetics approaches have associated PAH with numerous genes which were further experimentally studied (Table B.2); the level of evidence supporting the causal role of these genes, however, is variable and depends on multiple

factors (Table B.3). PAH is considered to be a monogenic condition transmitted in autosomal dominant fashion with incomplete penetrance. Heterozygous germline mutations in *BMPR2*, a member of the TGF-β superfamily, are the most common genetic cause of PAH [108, 241], accounting for over 80% of familial PAH, and approximately 25% of idiopathic PAH [179]. Additional mutations within the TGF-β/BMP signalling pathway, such as activin A like type 1 (*ACVRL1*), endoglin (*ENG*), *SMAD* family members (*SMAD1*, *SMAD4*, *SMAD9*), caveolin 1 (*CAV1*), growth differentiation factor 2 (*GDF2*), loss of function variants in channelopathy genes, potassium two pore domain channel subfamily K member 3 (*KCNK3*), ATP binding cassette subfamily C member 8 (*ABCC8*), ATPase type 13A3 (*ATP13A3*), variants within developmental transcription factors, SRY-box transcription factor 17 (*SOX17*) and T-box transcription factor 4 (*TBX4*), and newly reported risk genes, γ-glutamyl carboxylase (*GGCX*), kallikrein 1 (*KLK1*) [598] have all been identified as individually rare causes of PAH. Infrequent cases of autosomal recessive transmission in *KCNK3* [373] and *GDF2* [561] have been associated with early disease onset and severe phenotype. A subtype of PAH, PVOD/PCH is linked to biallelic mutations in *EIF2AK4* [127].

Autosomal dominant mode of inheritance

TGF-β Pathway

In 2000, the International Primary Pulmonary Hypertension (PPH) Consortium demonstrated that familial PAH (FPAH) is caused by mutations in *BMPR2*, located on chromosome 2, encoding a TGF-β type II receptor [241]. They established a panel of eight kindreds, in which at least two members had the typical manifestations of PAH. Sequence variants were detected in seven probands; these variants, including two frameshift, two nonsense and three missense mutations, were distributed across the gene and each of the amino acid substitutions occurred at a highly conserved and functionally important site of the BMPR2 protein. They observed segregation of the mutations with the disease phenotype in seven of the eight families studied. As control subjects, they screened 150 normal chromosomes from the same population and 64 normal chromosomes from ethnically diverse subjects and observed no *BMPR2* mutations [241]. The predicted functional impact of these mutations, their segregation with the phenotype, and the absence of these variants in healthy controls provided strong support for the role of *BMPR2* and the TGF-β signalling pathway in the pathobiology of PAH. The role of *BMPR2* mutations has been subsequently reported in IPAH. Thomson et al. [521] investigated *BMPR2* gene mutations in 50 unrelated IPAH patients with no family history of the disease. In 13 patients (26%), 11 novel heterozygous mutations in *BMPR2* were identified, these included three missense, three nonsense and

five frameshift. They also sequenced both parents for five of the 13 probands; paternal transmission was observed for three families, whereas the remaining two mutations arose spontaneously. *BMPR2* mutations were not observed in 150 normal chromosomes [521]. Screening of other disease subtypes revealed *BMPR2* mutations among patients with PAH associated with congenital heart disease (PAH-CHD) [405] and PVOD [447].

The *BMPR2* gene encodes for the bone morphogenic type 2 receptor, which consists of four highly conserved domains: N-terminal ligand binding, transmembrane, serine/threonine kinase, and a cytoplasmic tail. Presence of a long cytoplasmic tail, encoded by exon 12, distinguishes BMPR2 from other TGF-β receptors. BMPR2 associates with a type 1 receptor (ALK1) in response to the main circulating BMP ligands (BMP9 and BMP10), with endoglin (ENG) serving as a co-receptor. Type 1 receptors then phosphorylate downstream SMAD proteins required for transcriptional processes [544]. BMPR2 and its co-receptors are abundantly expressed in pulmonary vascular endothelium which leads to a predominantly lung-specific phenotype of patients harbouring *BMPR2* mutations. Functional genomics studies and transgenic animal models have shed light on the effects of disrupting the *BMPR2* signalling pathway. In pulmonary artery endothelial cells (PAECs), *BMPR2* knockout leads to increased permeability, heightened proliferation and enhanced apoptosis. The loss of *BMPR2* signalling also impacts the secretion of vasodilatory and inflammatory molecules by endothelial cells, such as IL-6, IL-8, E-selectin [544] and eNOS [159], facilitates a transition from mitochondrial oxidative phosphorylation to glycolysis [67] and may promote endothelial-to-mesenchymal transition [428]. Additionally, loss of *BMPR2* in other cell types, like PASMC, fibroblasts, immune cells, cardiomyocytes and hematopoietic stem cells, may contribute to disease development. PASMCs with *BMPR2* haploinsufficiency show a hyperproliferative phenotype due to the loss of antiproliferative Smad1/5 signalling [583]. A flagship illustration of the effect of *BMPR2* mutations on cardiomyocyte metabolism was a study investigating a disruption of RV function in a mouse harbouring a heterozygous *Bmpr2* C-terminal truncating mutation (R899X), which led to impaired hypertrophy and lipid accumulation in the RV [207]. There is also a growing body of evidence that PAH is associated with myeloproliferative disorders [3]. Furthermore, abnormalities in the bone marrow and hematopoietic progenitor cells are a frequent finding in PAH patients without hematologic disorder. It was previously demonstrated that low dose lipopolysaccharide could instigate the development of PH in genetically susceptible mice [483]. A follow-up study showed that *Bmpr2*$^{+/-}$ mice could be rescued by hematopoietic stem cell transplantation, suggesting hematopoietic stem cells involvement in susceptibility to PH [92].

Large cohort studies have proved useful in defining the relative contribution of *BMPR2* mutations in various PAH subtypes. Gräf et al. [179] reported rare heterozygous *BMPR2* mutations in 160 of 1,048 PAH cases (15.3%); the frequency of *BMPR2* mutations in FPAH, IPAH and anorexigen-exposed PAH were 75.9%, 12.2% and 8.3%, respectively. Fourteen percent of *BMPR2* mutations resulted in the deletion of larger protein-coding regions, ranging from 5 kb to 3.8 Mb in size. Additionally, 52% of the observed *BMPR2* mutations were newly identified in their study [179], suggesting that nearly two decades after the first *BMPR2* mutation was identified, the use of WGS has allowed for closer study of *BMPR2*, including large deletions around the *BMPR2* locus, and the TGF-β pathway. Another large-scale study in a more heterogeneous group of patients (Group 1 PAH) [598] reported *BMPR2* mutations in 180 of 2572 cases (7%); the frequency of *BMPR2* mutations in FPAH and IPAH patients were 62.4% and 9.3%, respectively. Taken together, over 600 distinct mutations in *BMPR2* have been identified in PAH patients [330, 328, 327, 179, 598] of which around 70 – 80% are identified in FPAH and 10 – 20% in IPAH [124]. Importantly, impaired BMPR2 signalling was shown to be a universal feature of PAH and pointed towards other key members of the canonical BMPR2 signalling pathway as potential culprits for the disease [21, 294, 74]. Mutations in *ACVRL1* and *ENG* have been reported in PAH patients and in patients with PH in association with hereditary hemorrhagic telangiectasia (HHT). HHT is a rare autosomal dominant genetic disorder characterised by arteriovenous malformations and multiple telangiectasias [110]; it is frequently linked to defects in *ACVRL1* and *ENG* and as HHT and PAH may co-present in families, suggests a common molecular aetiology [71, 202, 531]. Of note, PH secondary to high cardiac output from arteriovenous fistulas is much more common in HHT, and such phenocopies, if unrecognised, may introduce significant bias to the studies [203]. Conversely, I/HPAH associated with *ACVRL1* and *ENG* can occur without clinical features of HHT [203, 149, 202], as the latter shows age-related penetrance. In a large case-control study, which employed deep phenotyping prior to association analysis, *ACVRL1* was associated with HPAH [503] but fell just below the cut-off for significance when studied in unselected patients with Group 1 PAH [598].

Two studies using targeted sequencing of *BMPR2* signalling intermediates provided further evidence supporting the role of this pathway in the pathogenesis of PAH. Shintani et al. [467] screened 23 patients with IPAH for mutations in *SMAD1*, *SMAD2*, *SMAD3*, *SMAD4*, *SMAD5*, *SMAD6* and *SMAD9* (*SMAD8*) and identified a nonsense mutation in *SMAD9* in a child who was diagnosed at eight years of age and his unaffected father. The results of immunoblotting and co-immunoprecipitation assays indicated that the *SMAD9* mutant disturbs the downstream signalling of TGF-β/BMP. In a latter study, *SMAD1*, *SMAD4*,

SMAD5 and *SMAD9* were screened by direct sequencing in a cohort of 324 PAH cases (188 IPAH and 136 anorexigen-induced PAH) [371]. Four gene defects in three genes were observed. A novel missense variant in *SMAD1* was observed in an IPAH patient, a predicted splice-site mutation and a missense variant in *SMAD4* were observed in two IPAH patients and a novel missense variant in *SMAD9* was observed in a patient of Japanese origin. These four variants were absent in the 960 European and 284 French control samples and the *SMAD9* variant was excluded from the panel of 340 Japanese controls. A case-control study using WGS detected two cases harbouring protein-truncating variants in *SMAD1*, of which one co-existed with a protein-truncating variant in *BMPR2*, and eight *SMAD9* variants, two of which co-occurred with protein-truncating variants in *BMPR2* and *GDF2*; statistical analysis did not reveal significant association with studied phenotypes [503]. In another large cohort study (n = 2572 cases, 72% European), deleterious variants were observed in *SMAD1* (two cases), *SMAD4* (two cases) and *SMAD9* (13 cases) but were not statistically significant [598]. Taken together, these findings demonstrate that variations within the SMAD family have a small effect size, suggesting that a second genetic or environmental hit is needed, or that they perturb other non-investigated pathways. Besides *BMPR2* mutations, *CAV1* mutations are a rare cause of PAH. Variants in *CAV1* were initially implicated in PAH pathogenesis by exome sequencing of a three-generation family with autosomal dominant HPAH who were negative for established variants in the TGF-β family [25]. The study identified a frameshift mutation in *CAV1*; all PAH patients and several unaffected family members carried the *CAV1* mutation, suggesting incomplete penetrance. Subsequent evaluation of an additional 62 unrelated HPAH and 198 IPAH patients identified an independent *de novo* *CAV1* mutation in a child with IPAH. The experimental analysis of *CAV1* mutations have revealed impaired protein trafficking in mouse embryonic fibroblasts carrying a disrupted copy of *CAV1*. Furthermore, patient fibroblast cells displayed reductions in both CAV1 and associated accessory proteins [86]. Two separate studies have since identified *CAV1* mutations in PAH patients; the first identified a novel heterozygous frameshift mutation in an adult PAH patient with a paediatric-onset daughter who died at nine years old [597], and the second identified deleterious variants in 10 patients with I/F/APAH, with three related cases carrying the same likely gene damaging mutation [598]. WGS in an I/HPAH cohort did not detect deleterious variants in *CAV1* [179]. Whilst these findings highlight the importance of caveolae in the homeostasis of the pulmonary vasculature, the link between *CAV1* mutations and PAH requires further study.

Bone morphogenetic protein (BMP)-9 (encoded by *GDF2*) and BMP10 are ligands involved in the TGF-β signalling pathway. Wang et al. [561] identified a novel homozygous

nonsense mutation in the *GDF2* gene in a five-year-old Hispanic child with severe PAH. Genetic testing revealed that both parents were heterozygous for the same mutation, indicating that the child inherited the *GDF2* mutant allele from each parent. This study was the first to report a novel homozygous nonsense mutation in *GDF2* in an IPAH patient, suggestive of the causative role of *GDF2* mutations in PAH. Further evidence came from the NBR study which identified associations between rare heterozygous missense (n = 7) and frameshift variants (n = 1) in adult-onset IPAH (88% European) [179]. *In vitro* analysis of seven *in silico* predicted deleterious missense mutations in *GDF2*, from the NBR study revealed that all variants displayed dysfunctional processing, secretion or protein stability. In addition, functional studies found a predicted benign mutation to be deleterious [215].

The identification of *GDF2* mutations has since been independently replicated in a Chinese cohort [564]. Wang et al. [564] performed an exome-wide gene-based burden analysis on two independent case-control studies. The discovery analysis, containing 251 IPAH patients, identified rare heterozygous mutations in *BMPR2* (49 cases), *ACVRL1* (15 cases), *TBX4* (10 cases), *SMAD1* (two cases), *BMPR1B* (one case), *KCNK3* (one case) and *SMAD9* (one case). In a gene-based burden analysis (cases: n = 251; controls: n = 1884), only three genes (*BMPR2*, *GDF2* and *ACVRL1*) had an exome-wide significant enrichment of mutations in IPAH cases when compared to healthy controls. *GDF2* mutations were identified in 17 cases (6.8%) and ranked second to BMPR2 (56 cases, 22.3%). To validate the risk effects of *GDF2*, they performed WES in an independent replication cohort of 80 IPAH cases and in a second gene-based burden analysis (cases: n = 80; controls: n = 8624), *BMPR2*, *GDF2* and *ACVRL1* were again identified as the top three disease-associated genes. Within this analysis, five additional *GDF2* heterozygous mutations were identified. Among the 331 IPAH patients, they identified 22 cases carrying 21 distinct rare heterozygous mutations in *GDF2*, only two of which had been reported previously [179], accounting for 6.7% of IPAH cases. Six mutations were selected for functional studies, using mutant *GDF2* plasmids transfected into HEK293E cells/PAECs. Mutants displayed reduced BMP9 levels, reduced BMP activity and increased apoptosis [564]. An independent cohort confirmed the exome-wide association of *GDF2* among 1832 PAH and 812 IPAH cases of European ancestry [598]; twenty-four *GDF2* variants were observed in 28 cases, only two of which had been reported previously, and 75% of these occurred in IPAH cases.

Additionally, gene panel sequencing of 263 PAH patients (180 IPAH, 11 FPAH, 13 drug and toxin-induced PAH and 59 sporadic PVOD) revealed two (1.2%) *GDF2* mutations in adult PAH cases [128]; due to the close similarity of BMP9 and BMP10 (a close paralogue of BMP9 that encodes an activating ligand for ACVRL1), the BMP10 gene was also included

in the capture design. Two mutations were identified in BMP10, a truncating mutation and a predicted loss of function variant were identified in two severely affected IPAH patients. Two rare missense variants in *BMP10* were identified in patients with IPAH in an independent cohort [215]. These results emphasise the role of *GDF2* in the pathobiology of PAH and suggest that *BMP10* might act as a predisposing risk factor.

Channelopathies

Mutations in ion channels have been implicated in numerous diseases, collectively called channelopathies. These diseases include, but are not limited to, cardiac, respiratory, neurological and endocrine disorders. So far at least four genes encoding ion channels have been implicated in the pathogenesis of PAH.

The first channelopathy described in PAH was caused by a genetic defect in *KCNK3* in patients with familial and sporadic PAH [322]. *KCNK3* belongs to a family of mammalian potassium channels and encodes for a two-pore potassium channel which is expressed in PASMCs; this channel plays a role in the regulation of resting membrane potential and pulmonary vascular tone and vascular remodelling [373]. In studying a family in which multiple members had PAH, Ma et al. [322] identified a novel heterozygous missense variant in *KCNK3* as a disease-causing candidate gene within the family. WES was used to study an additional 10 probands with FPAH and two novel heterozygous *KCNK3* variants were identified and segregated with the disease. In addition, three novel *KCNK3* variants were identified in 230 patients with IPAH. These five variants were predicted to be damaging. In summary, *KCNK3* mutations were identified in three of 93 unrelated patients (3.2%) with FPAH and in three of 230 patients (1.3%) with IPAH [322]. Functional analysis was conducted on six mutations reported in *KCNK3*, all of which resulted in LoF. Patch clamp variants showed a reduction in current, with rescue of channel function, in a subset of variants by administration of a phospholipase inhibitor, ONO-RS-082 [322]. In addition, Bohnen et al. [52] introduced a heterozygous, single amino acid variant into *KCNK3* using site-directed mutagenesis and transfected hPASMCs with this mutant via a pcDNA3.1 plasmid. The physiological impact of this mutation was assessed using whole-patch clamp tests to observe changes to membrane potential and current, with experiments demonstrating loss of channel function in mutant cells compared to WT *KCNK3*.

In another study using targeted sequencing, two *KCNK3* mutations were observed in three patients from two families. One of these mutations, a homozygous missense variant in *KCNK3*, was identified in a patient belonging to a consanguineous Romani family; his affected mother and asymptomatic father were carriers of the same *KCNK3* mutation. This

is the first report of a young patient with severe PAH carrying a homozygous mutation in *KCNK3* [373]. Of note, in the Ma et al. [322] study, as the pedigree suggested an autosomal dominant mode of inheritance, homozygous variants were excluded from the analysis. The two biggest case-control studies, reported by Gräf et al. [179] and Zhu et al. [598], identified heterozygous *KCNK3* mutations in only four (0.4%) and three (0.1%) cases, respectively, and did not show statistically significant associations.

Conversely, Gräf et al. [179] detected statistically significant deleterious variants in two new channel genes: *ATP13A3* and *AQP1*. Utilising a rigorous case-control comparison using a tiered search for variants, they searched for high-impact protein-truncating variants (PTVs) overrepresented in cases and identified a higher frequency of PTVs in *ATP13A3* (six cases). ATP13A3 is a poorly characterised member of the P-type ATPase family of proteins that transport a variety of cations across membranes [280]; *ATP13A3* is thought to play a role in polyamine transport [333]. Within PAH cases, Gräf et al. [179] identified three heterozygous frameshift variants, two stop gain, two splice region variants and four heterozygous likely pathogenic missense variants in *ATP13A3*. These variants were predicted to lead to loss of ATPase catalytic activity, and to destabilise the conformation of the catalytic domain; six variants were predicted to cause protein truncation, suggesting that loss of function of *ATP13A3* contributes to PAH pathogenesis. Further functional studies showed that siRNA-mediated knockdown of expression resulted in a reduction in cell proliferation and increased apoptosis, as well as the loss of endothelial integrity in hPAECs [313]. Likewise, mutant *Atp13a3* mice, carrying a protein-truncating variant introduced using CRISPR/Cas9, demonstrated haemodynamic measurements indicative of PAH. This included reduced pulmonary artery acceleration time (PAAT) and increased RVSP, with rats also exhibiting reduced polyamine levels in their lungs [299].

In the NBR study, *AQP1* ranked second in a combined rare PTV and missense variant case-control analysis. Along with statistical evidence, familial segregation of *AQP1* variants with the phenotype was shown in three families [179]. *AQP1* belongs to the aquaporins family, a family of water-specific membrane channel proteins that facilitate water transport in response to osmotic gradients [497]. *AQP1* has also been associated with cell migration and vascularity. *AQP1* is highly expressed in microvascular endothelial cells, vascular smooth muscle cells and non-vascular endothelium [527], and has recently been proposed as a novel promoter of tumour angiogenesis [83]. A recent study demonstrated that depletion of *AQP1* reduced proliferation, migratory potential, and increased apoptosis of PASMCs [459]. Saadoun et al. [449] produced an *AQP1* knockout mouse model using targeted gene disruption; when implanted with melanoma cells, the developing tumours displayed impaired

angiogenesis. In-depth analysis of *APQ1* function in endothelial cells, produced from mouse aortic cells, using cell migration, wound healing, proliferation and cord development assays, revealed critical roles of *AQP1* in cell migration. *AQP1* null cells showed reduced ability in all four assays. This was corroborated by the plasmid expression of *AQP1* and another water channel protein, *AQP4*, in non-endothelial cells, resulting in increased migration. Functional analysis has confirmed the role of AQP1 in cell migration under hypoxic conditions. Yun et al. [585] used rat pulmonary-artery-derived PASMCs, infected with adenovirus vectors carrying either WT *AQP1* or *AQP1* with a 37 amino acid terminal deletion removing the COOH tail. Overexpression of *AQP1* increased levels of β-catenin and its downstream targets, including c-Myc and cyclin D1. However, loss of the COOH tail produced no such results, suggesting COOH-mediated upregulation of β-catenin. Both *AQP1* and β-catenin displayed elevated levels under hypoxic conditions. This was further corroborated by siRNA-mediated silencing of β-catenin, in cells subject to hypoxia, and consequent infection with adenoviral constructs. Cells were rescued from the migration and proliferation associated with hypoxic conditions. Another large case-control study [598] identified patients harbouring *ATP13A3* and *AQP1* but neither of these genes reached genome-wide significance.

Mutations in *ABCC8* have recently been identified as a potential second potassium channelopathy in PAH [51]. *ABCC8* encodes SUR1 (sulfonylurea receptor 1), a regulatory subunit of the ATP-sensitive potassium channel; *ABCC8* is highly expressed in the human brain and endocrine pancreas and moderately expressed in human lungs [28]. Mutations in *ABCC8* have previously been related to type II diabetes mellitus and congenital hyperinsulinism [141]. In exome-sequencing a cohort of 99 paediatric- and 134 adult-onset Group 1 PAH patients (182 IPAH and 52 HPAH), Bohnen et al. [51] identified a *de novo* heterozygous predicted deleterious missense variant in *ABCC8* in a child with IPAH. All individuals within this cohort and the second cohort of 680 adult-onset PAH patients (NBR study) were screened for rare or novel variants in *ABCC8*. Eleven heterozygous predicted damaging *ABCC8* variants were identified, seven of these were observed in the original cohort, including one familial case. Eight mutations were analysed in COS cells using patch-clamp tests and rubidium flux assays, as a measure of ion efflux. Loss of channel function was seen in all variants, with the rescue of function by the administration of the SUR1 activator drug, diazoxide [51]. In a study of 2572 PAH cases, Zhu et al. [598] identified rare deleterious variants in newly reported risk genes and nearly two-thirds of these variants were in *ABCC8* (26 variants in 29 IPAH/APAH patients). In a more recent study, utilising a custom NGS targeted sequencing panel of 21 genes, Lago-Docampo et al. [282] identified 11 rare variants in *ABCC8* within a cohort of 624 paediatric and adult patients from the Spanish PAH registry.

To date, *ABCC8* variants have been identified in patients with IPAH, FPAH, PAH-CHD and APAH and account for 0.5 - 1.7% of cases.

Transcription Factors

An emerging category of HPAH is the one that can be labelled as a disorder of transcriptional regulation. So far, two non-Smad transcription factors have been reported as causative genes in PH, namely *TBX4* and *SOX17*. Interestingly, variants within developmental transcription factors are enriched in paediatric patients.

TBX4, is an evolutionarily conserved transcription factor that plays a critical role during embryogenesis [466]. *TBX4*, expressed in the atrium of the heart, limbs, and the mesenchyme of the lung and trachea, encodes a transcription factor in the T-box gene family [18]. Deletions and LoF mutations in *TBX4* cause a variety of developmental lung disorders [255] and have been identified as a prominent risk factor in small patella syndrome (SPS) [53], childhood-onset PAH [262] and more recently, persistent pulmonary hypertension in neonates [151]. In a 2013 study incorporating array-comparative hybridisation and direct sequencing, three *TBX4* mutations and three novel *TBX4* microdeletions were detected in six out of 20 children with I/HPAH and interestingly, features of SPS were detected in all living *TBX4* mutation carriers [262].

Zhu et al. [597] performed exome sequencing on a cohort of 155 paediatric- and 257 adult-onset PAH patients. Within 13 probands (12 paediatric- and one adult-onset), they identified 13 likely pathogenic/predicted highly deleterious *TBX4* variants; eight of these variants were inherited from an unaffected parent, whereas one was *de novo*. This pattern is consistent with the incomplete penetrance observed for *BMPR2* mutation carriers [292]. Similar frequencies of rare, deleterious *BMPR2* mutations were observed in paediatric- and adult-onset I/FPAH patients; however, there was significant enrichment of rare, predicted deleterious *TBX4* mutations in paediatric- (10 of 130 patients) compared with adult-onset (0 of 178 patients) IPAH patients. In comparison to *BMPR2* mutation carriers, *TBX4* carriers had a 20-year younger age of onset, with a mean age of onset of 28.2 ± 15.4 years and 7.9 ± 9.0, respectively. After *BMPR2* mutations (10%), variants in *TBX4* (7.7%) conferred the highest degree of genetic risk of paediatric-onset IPAH [597]. Similar estimates of *BMPR2* (12.5% of I/FPAH patients) and *TBX4* mutation carriers (7.5% of I/FPAH patients) were observed in a study of 66 paediatric patients [303]. Additionally, in a 2019 study, examining 263 PAH and PVOD/PCH patients (paediatric and adult cases), *TBX4* mutations were the second most frequent mutations after *BMPR2* in both paediatric and adult cases [128]. Indicative of bimodal age distribution, pathogenic *TBX4* variants have also been reported in adult-onset

PAH. Gräf et al. [179] identified deleterious heterozygous rare variants in 14 cases, Navas Tejedor et al. [373], in a Spanish cohort of 136 adult-onset PAH patients, identified three pathogenic mutations, and Kerstjens-Frederikse et al. [262], in a much smaller adult cohort (n = 49), detected a rare *TBX4* mutation. In a more recent study, 448 index patients were screened for PAH predisposing genes; 20 patients (nine childhood-onset) from 17 unrelated families carried heterozygous mutations in the *TBX4* gene, bringing the frequency of *TBX4* mutations in France to 6% and 3% in childhood- and adult-onset PAH, respectively [522]. Within this cohort, SPS was present in 80% of cases [522] and interestingly, all patients showed decreased DLCO and 87% had parenchymal abnormalities. These findings suggest that *TBX4* mutations may occur with or without skeletal abnormalities and whilst such mutations are mainly associated with childhood-onset PAH, the prevalence of PAH in adult *TBX4* mutation carriers could be up to 3%, depending on the population studied.

A recent study of 1038 PAH patients observed that rare heterozygous variants in *SOX17* were significantly overrepresented in the I/HPAH cohort [179]. *SOX17* is a member of the conserved SOX family of transcription factors. These transcription factors play a pivotal role in cardiovascular development and feature prominently in the aetiology of human vascular disease; they are involved in the regulation of embryonic development, the determination of cell fate and participate in vasculogenesis and remodelling [144]. In the NBR study [179], deleterious variants in *SOX17* were detected in less than 1% of the studied population and were characterised by younger age at diagnosis; familial segregation was shown in one patient. Interestingly, a much smaller WES study of 256 PAH-CHD patients (15 familial and 241 sporadic cases; 56% pediatric-onset) showed a significant association between *SOX17* and PAH-CHD. In this cohort *SOX17* variants explained the disease in 3% of European PAH-CHD patients and were more commonly found among patients with pediatric-onset disease [599].

A Japanese study also demonstrated familial segregation of *SOX17* variants [212]. This study whole-exome sequenced 12 patients with PAH, 12 asymptomatic family members and 128 index cases and identified *SOX17* mutations in four PAH patients (three of these had congenital heart defects, i.e., atrial septal defect or patent ductus arteriosus) and one asymptomatic family member. Interestingly, the same heterozygous missense mutation in *SOX17* (c.397C) was observed in a Japanese patient [212] and in a patient with PAH from the NBR study [179], suggesting that this base position may be a pan-ethnic mutational hot spot. Taken together, these data strongly implicate *SOX17* as a new risk gene for PAH-CHD and suggest that this gene has a pleiotropic effect. Replication analyses in other PAH cohorts, with specific PAH subclasses, are needed to confirm the precise role of *SOX17*.

New genes

In recent years, new risk genes for PAH have emerged from large WES and WGS studies. Zhu et al. [598] performed targeted gene sequencing alongside WES in a large cohort from the National Biological Sample and Data Repository for PAH (US PAH Biobank (US PAHBB): n = 2,572). Despite screening for 11 established PAH risk genes and seven recently reported risk genes, they failed to identify rare deleterious variants in known risk genes for 86% of the PAH Biobank cases (Group 1 PAH). They performed gene-based case-control association analysis and to prevent confounding by genetic ancestry, only participants of European ancestry (cases: n = 1,832; controls: n = 7,509 gnomAD WGS subjects and n = 5,262 unaffected parents from the Pediatric Cardiac Genomics Consortium) were included. Using a variable threshold method, they identified two genes that exceeded the Bonferroni-corrected cut-off for significance: *BMPR2* and *KLK1*. *KLK1*, which encodes kallikrein 1, also known as tissue kallikrein, has not previously been associated with pulmonary hypertension.

The analysis was then repeated using 812 European IPAH cases and significant associations were observed for *BMPR2*, *KLK1* and *GGCX*. *GGCX* encodes γ-glutamyl carboxylase and has previously been implicated in coagulation factor deficiencies and vascular calcification [106], but again, never in PAH. In a final analysis, the entire PAH Biobank cohort was screened for rare deleterious variants in *KLK1* and *GGCX*; twelve cases carried *KLK1* variants (all European), whereas 28 cases carried *GGCX* variants (19 European, six African, three Hispanic), accounting for $\sim$0.4% and $\sim$0.9% of PAH Biobank cases, respectively. Carriers of *KLK1* and *GGCX* had a later mean age of onset and relatively moderate disease phenotype compared to *BMPR2* carriers. Both *KLK1* and *GGCX*, expressed in the lung and vascular tissues, play an important role in vascular haemodynamics and inflammation. Whilst Zhu et al. [598] identified *KLK1* and *GGCX* as new candidate risk genes for IPAH, suggesting new pathogenic mechanisms outside of the TGF-β/BMP signalling pathway, further research needs to be conducted to better understand these findings, especially in larger cohorts of similar phenotypic characteristics.

Limiting the case-control analysis to putative risk gene candidates (rather than all protein coding genes) is a viable approach, providing that adequate information about the pathways involved in disease pathogenesis is available. Based on the data from preclinical studies suggesting epigenetic dysregulation in PAH, Potus et al. [415] screened patients from the US PAHBB (n = 2,572) with Group 1 PAH for rare deleterious variants in *TET1*, *TET2* and *TET3* genes, and found enrichment of rare, predicted deleterious germline variants in *TET2*. Additionally, heterozygous and homozygous *Tet2* knockout mice developed PAH, pulmonary artery remodeling, which were ascertained through echocardiography and RHC. Patients

carrying *TET2* mutations were found to have a later age of onset compared to other PAH patients; this finding was also replicated in mice.

Autosomal Recessive Mode of Inheritance

As alluded above, PVOD/PCH is a rare form of Group 1 PAH, in which 25% of cases can be explained by biallelic mutations in *EIF2AK4*. *EIF2AK4* encodes a member of a family of kinases that phosphorylate the alpha subunit of eukaryotic translation initiation factor-2 (EIF2), resulting in the downregulation of protein synthesis [445]. The encoded protein responds to amino acid deprivation by binding uncharged transfer RNAs. It may also be activated by glucose deprivation and viral infection [99].

Animal models of homozygous *Eif2ak4* knockout mice displayed increased levels of protein carbonylation, a symptom of oxidative damage that was induced by dietary leucine deprivation. Additionally, these mice exhibited reduced viability compared to controls, with some dying shortly after birth. These findings suggest that the gene may elicit protective effects against oxidative damage [16].

Additionally, single cases of homozygous mutations in *GDF2* [561], *KCNK3* [373] and *ATP13A3* [35] have been reported. These cases have been uniformly characterised by early disease onset and high disease severity.

1.5.2 Common Genetic Variation

Complex pathobiology, low penetrance, heterogeneous phenotype and variable disease trajectory allow for common sequence variation that contributes to PAH risk and natural history. In HPAH, several common genetic variants were shown to impact the disease. Firstly, *BMPR2* mutations are a significant source of sex-related bias in disease penetrance (42% in females vs 14% in males) [292]; one explanation for this could be gene polymorphisms involved in estrogen metabolism [23]. Females harbouring deleterious variants in *BMPR2* show a significant reduction in *CYP1B1* gene expression and as a consequence, a lower 2-hydroxyestrone to 16α-hydroxyestrone ratio leading to activation of mitogenic pathways [23]. Moreover, direct estrogen receptor α binding to *BMPR2* promoter results in reduced *BMPR2* gene expression in females and may contribute to the increased prevalence of PAH [24]. Secondly, variation in *BMPR2* expression in HPAH, caused by nonsense-mediated decay positive (NMD^+) *BMPR2* mutations, is conditional upon individual polymorphisms of the wild type (WT) allele [198]. Thirdly, a SNP in TGF-β1 modulates age at disease onset in patients harbouring *BMPR2* mutations [407].

In IPAH, a SNP (rs11246020) in the Sirtuin3 (*SIRT3*), mitochondrial deacetylase, in either homozygote or heterozygote fashion, was associated with increased acetylation of mitochondrial proteins compared to the IPAH patients or disease comparator group with the WT genotype [392]. *SIRT3*, by deacetylating and thus activating multiple enzymes and electron transport chain complexes, plays a significant role in cell bioenergetics, and its polymorphism has been associated with susceptibility to metabolic syndrome [213] and IPAH, but not APAH [392]. Similarly, uncoupling protein 2 (UCP2), shown to conduct calcium from the endoplasmic reticulum (ER) to mitochondria and suppressing mitochondrial function, has been implicated in the pathogenesis of PAH [114]. These findings have important therapeutic implications as common variants in *SIRT3* and *UCP2* predicted response to dichloroacetate, pyruvate dehydrogenase kinase inhibitor, in a PAH phase 2 clinical trial [353]. These findings are further supported by *Sirt3* knockout mice (129/Sv strain), which displayed elevated mPAP and PVR along with RV hypertrophy and reduced exercise tolerance in a dose dependent manner when compared with controls [392]. However these results were not replicated in C57BL/6 mice [567], which may indicate that the differences in background metabolism result in variable response to gene dosing.

The renin–angiotensin–aldosterone (RAA) system has been implicated in the pathogenesis of both systemic and pulmonary hypertension. For example, polymorphisms in the gene encoding angiotensin-converting enzyme (ACE) have been associated with IPAH [2] and with diaphragmatic hernia with persistent pulmonary hypertension [479]. Likewise, common variation in angiotensin II type 1 receptor (*AGTR1*) was associated with age at diagnosis in PAH. These findings indicate that RAA might be a therapeutic target [80].

Considering the important role of vasoactive, angiogenic and angiostatic substances in the pathogenesis and therapy of PAH, studies investigating the impact of common variation on gene expression of these substances are warranted. Recently, Villar et al. [32] found a recurrent SNP (rs397751713) in the promoter region of the endothelin-1 (*END1*) gene that had important regulatory consequences in both IPAH and APAH patients; rs397751713, which consists of an adenine deletion, allows transcription factors (Peroxisome proliferator-activated receptor γ (PPARG) and Kruppel Like Factor 4 (KLF4)) to bind to the promoter. Both of these transcription factors are linked to PAH pathogenesis and have been suggested as potential therapeutic targets [501, 180, 339, 465]. In a similar fashion, a polymorphism in a potent angiostatic factor, endostatin, Collagen Type XVIII α 1 Chain (*COL18A1*), was studied. A SNP (rs12483377) in *COL18A1* was observed at an increased frequency in PAH patients (MAF 21.6) relative to published controls (MAF 7.5), or patients with scleroderma without PAH (MAF 12.5). Rs12483377 encodes uncharged amino acid asparagine (N) at

residue 104 in place of negatively charged aspartic acid (D); carriers of single minor allele A (genotype AG), had lower serum endostatin levels than those who carried WT (genotype GG) and showed better survival even after adjusting for confounders [101].

PAH is characterised, among other features, by accumulation of DNA lesions during the disease progression; however, the mechanisms underlying this phenomenon remain unclear. Common variation in the DNA Topoisomerase II Binding Protein 1 (*TopBP1)*, an enzyme required for DNA replication and rescue of stalled replication forks and checkpoint control, has also been found in IPAH patients. Immunohistochemical analysis of vascular lesions demonstrated a reduction in TopBP1 protein. Analysis of cultured pulmonary microvascular endothelial cells from IPAH lungs, using q-PCR and Western blotting, showed a reduction in *TopBP1* mRNA and protein. Further analysis of cells for increased susceptibility to DNA damage, by administering hydroxyurea and staining nuclei for phosphorylated histone foci, revealed an excess of DNA strand breaks and apoptosis. However, patient cells could be rescued by transfection with a plasmid containing wild-type human *TopBP1*, displaying a reduction in DNA damage-induced apoptosis [105].

Along with studies looking at polymorphisms in a particular gene or its promoter region, two GWAS have been carried out in PAH. The first consisted of a discovery and validation cohort totalling 625 patients (*sans BMPR2* mutations) and 1,525 healthy individuals. A genome-wide significant association was found at Cerebellin 2 Precursor (*CBLN2*) locus; rs2217560-G allele was more frequent in cases than in controls in both the discovery and validation cohorts with a combined odds ratio (OR) for I/HPAH of 1.97 [95% CI 1.59;2.45] (p-value $= 7.47 \times 10^{-7}$). Although the rs2217560 genotype did not correlate with *CBLN2* mRNA expression in monocytes, higher *CBLN2* mRNA levels were seen in lung tissue explanted from PAH patients than in control lung samples. Further analysis showed that increasing concentrations of CBLN2 led to inhibition of proliferation of PASMCs [165].

To date, the largest PAH GWAS of multiple international I/HPAH cohorts, totalling 2085 cases and 9659 controls, identified two novel loci associated with PAH: an enhancer near *SOX17* (rs10103692, OR 1.80 (95% CI 1.55;2.08), p-value $= 5.13 \times 10^{-15}$), and a locus within *HLA-DPA1/DPB1* (rs2856830, OR 1.56 (95% CI 1.42;1.71), p-value $= 7.65 \times 10^{-20}$) [436]. CRISPR-mediated inhibition of the enhancer reduced SOX17 expression; this finding corroborates and extends the previous discovery of the association of rare variants in *SOX17* with PAH [179]. The *HLA-DPA1/DPB1* rs2856830 genotype was not only associated with I/HPAH but also with a beneficial effect on survival (CC genotype 13.5 vs TT genotype 6.97 years) irrespective of baseline disease severity, age and sex [436]. The latter study did not replicate previous results at genome-wide significance.

1.6 Reverse phenotyping

1.6.1 Concepts

The concept of reverse phenotyping (Fig. 1, Process 4) refers to the use of genetic marker data to refine the disease subgroup definitions [458]. The method was first used in phenotyping patients with sarcoidosis [188] and has proven successful in many other diseases characterised by high heterogeneity [288].

1.6.2 Theoretical Basis

Heterogeneity provides a unique challenge in the diagnosis and treatment of rare diseases, and is further complicated by the accuracy of reporting. Knowledge about the genetic makeup of the individual allows targeted treatments, thereby enhancing efficacy and decreasing the risk of potential side effects [377]. Conversely, knowledge about the phenotypic characteristics of mutation carriers is indispensable to match study design and methodological approaches to disease attributes [571].

1.6.3 Methodology

Studying the complexity of genotype–phenotype associations is of particular interest to medical genetics. Evidence for such studies can be retrieved from published peer-review literature and publicly available databases, and facilitated by computational tools such as the R package VarformPDB (Disease-Gene-Variant Relations Mining from the Public Databases and Literature), which captures and compiles the genes and variants related to a disease, a phenotype or a clinical feature from public databases including HPO (Human Phenotype Ontology), Orphanet, OMIM (Online Mendelian Inheritance in Man), ClinVar, and UniProt (Universal Protein Resource) and PubMed abstracts. Alternatively, the development of large biobanks that link rich electronic health record (EHR) data and dense genetic information have made it possible to enhance clinical characterisation of mutation carriers in a cost-effective fashion. Besides deep clinical, often longitudinal phenotyping, multi-omic approaches can be employed thanks to the concurrent collection of patients' biological samples. A number of efforts have been made to facilitate EHR-based genetic and genomic research, such as the establishment of the UK Biobank[9], the Electronic Medical Record

[9]https://www.ukbiobank.ac.uk

and Genomics Network[10] or the Precision Medicine Initiative Cohort Program[11], which allow both forward genetics and reverse phenotyping applications. Mining EHRs for genetic research purposes possesses a number of advantages over classical population or family-based genetic association studies; large samples can be collected over a short period of time in a cost-effective fashion. Moreover, the robustness of EHRs allows for the investigation of a spectrum of phenotypes in a hypothesis-free manner. Familial relationships reported in EHRs have also been used for the estimation of disease heritability shown to be consistent with the literature [411].

Multiple tools are being used by the Clinical Genetics community to address the issue of reverse phenotyping when interpreting NGS data. This includes DECIPHER [138], a publicly-available database that houses a collection of case-level evidence for 36,801 individuals with genetic diseases, for comparison of phenotypic and genotypic data. Projects affiliated to DECIPHER can deposit and share patients, variants and phenotypes to invite collaboration and increase diagnostic yield. Another example (also accessible via DECIPHER) is that of Matchmaker Exchange [406] or Genematcher [56], for mining other databases based on matching genotype and phenotype data. VarSome is a data aggregator that encompasses platforms and bioinformatic tools with the aim of consolidating information required for variant analysis. VarSome draws information from 30 databases [274] and is also embedded into another web-based tool, VariantValidator, which enables validation, mapping and formatting of sequence variants using HGVS nomenclature [145].

Knowledge of age-dependent and/or reduced penetrance is vital in reverse-phenotyping as new clinical features may emerge in an individual over time. This is applicable to relatives of PAH patients found to carry presumed pathogenic variants in known disease-associated genes. A detailed evaluation of multiple family members alongside longitudinal follow-up of these individuals is vital to study the disease trajectory which has often higher definition than snapshot phenotyping at diagnosis.

Apart from mining existing databases and health records, recall by genotype (RbG) studies will be a relevant and valuable source of information regarding the biological mechanics of genotype–phenotype associations. By making use of the random allocation of alleles at conception (mendelian randomization - MR) these studies enhance the ability to draw causal inferences in population-based studies and minimise the problems related to confounders; additionally, the focus on phenotypic assessment of a specific carrier subgroup can enhance the understanding of pathobiology in a cost-effective manner. RbG studies can focus on

[10] https://www.genome.gov

[11] https://www.nih.gov/precision-medicine-initiative-cohort-program

a single variant or multiple variants. The former considers rare and large effect loci to understand the biological pathways of interest, and as a result, the groups are relatively small so the phenotype can be studied in detail, in the latter, genetic variation is used as a proxy of relevant exposure (proviso there is a credible association between genetic markers and exposure). To increase power for studies in which a single variant has a small effect size deployment of aggregate genetic risk scores can be beneficial [87].

1.6.4 Studies

The reversal of the usual hypothesis-driven paradigm for the refined diagnosis was first used by Grunewald and Eklund [188] who by deep clinical phenotyping of patients with Lofgren's syndrome, showed that erythema nodosum predominantly occurs in women and bilateral periarticular arthritis in men, and that the acute sarcoidosis subgroup can be further characterised according to the presence of *HLA-DRB1*0301/DQB1*0201*. The reverse phenotyping method has been subsequently adopted in other disease domains (Table 1.5) and has led to new disease discovery, such as the Koolen–DeVries syndrome, due to a 17q21.31 microdeletion involving the *KANSL1* gene [273], and the Potocki–Lupski Syndrome, due to a 17p11.2 microduplication [414, 413]. Recently, a retrospective study of 111 patients with nephrotic syndrome, who underwent WES, showed that reverse phenotyping increased the diagnostic accuracy in patients referred with the diagnosis of steroid-resistant nephrotic syndrome [288].

Table 1.5 Examples of successful studies that used reverse phenotyping.

Gene	Condition	Study designs	Data sources	Results	References
*HLA-DRB1*0301/DQB1*0201*	Sarcoidosis	Mix of retrospective and prospective cases with acute onset sarcoidosis	Health records	Patients with acute onset sarcoidosis carrying *HLA-DRB1*0301/DQB1*0201* genotype have good prognosis, manifestations of Lofgren's syndrome differ between men and women.	(Iannuzzi and Baughman 2007; Grunewald and Eklund 2007)[237, 188]
BMPR2	PAH	Retrospective analysis of 169 PAH patients	WES, Health records	Patients with missense mutations that escape nonsense-mediated decay have more severe disease than those with truncating mutations.	(Austin et al. 2009)[26]
BMPR2	PAH	Retrospective analysis of 171 patients		Missense variants in the cytoplasmic tail appear to confer less severe phenotype than other *BMPR2* variants with a later age of onset, milder haemodynamics, and more vasoreactivity.	(Girerd et al. 2015)[169]
BMPR2	PAH	A large individual participant data meta-analysis	Literature review	Patients harbouring deleterious *BMPR2* mutations have earlier disease onset, worse haemodynamics, are less likely to respond to NO challenge and have lower survival when compared to those without *BMPR2* mutations.	(Evans et al. 2016)[124]
BMPR2	PAH	Retrospective analysis of 44 PAH patients who underwent lung transplantation between 2005 and 2014	Histology, immunohistochemistry, morphometry of explanted lungs; French registry database	*BMPR2* mutation carriers are more prone to haemoptysis; haemoptysis is closely correlated to bronchial arterial remodelling and angiogenesis; pronounced changes in the systemic vasculature correlate with increased pulmonary venous remodelling, creating a distinctive profile in PAH patients harbouring a *BMPR2* mutation.	(Ghigna et al. 2016)[167]
multiple genes	Nephrotic syndrome	A retrospective analysis of all patients diagnosed with nephrotic syndrome between 2000 and 2018	WES, personalised diagnostic workflow	Reverse phenotyping after WES increased the diagnostic accuracy in patients referred with the diagnosis of steroid-resistant nephrotic syndrome.	(Landini et al. 2020)[288]

Abbreviations: *BMPR2*: Bone morphogenic protein receptor type 2; PAH: pulmonary arterial hypertension; WES: whole-exome sequencing

The reverse phenotyping of *BMPR2* mutation carriers has been studied in PAH patients. A large individual participant data meta-analysis found that patients with *BMPR2* mutations have earlier disease onset, worse haemodynamics, are less likely to respond to nitric oxide challenge and have lower survival when compared to those without *BMPR2* mutations [124]. The histological analysis of lungs explanted from those patients also revealed a higher degree of bronchial artery hypertrophy/dilatation, which correlated with the frequency of haemoptysis at presentation [167]. Patients with missense mutations in *BMPR2* that escape nonsense-mediated decay have more severe disease than those with truncating mutations, suggestive of a dominant-negative impact of mutated protein on downstream signalling [26]. However, missense variants in the cytoplasmic tail appear to confer less severe phenotype than other *BMPR2* variants with a later age of onset, milder haemodynamics, and more vasoreactivity [169].

Other good candidates for reverse phenotyping are patients with a mutation in transcription factors, i.e., *TBX4*, which by virtue of gene function would be associated with a more complex and heterogeneous phenotype. Patients with mutations in *TBX4* present with severe PAH associated with bronchial and parenchymal changes, low DLCO, with or without skeletal abnormalities [522], and bimodal age of onset [598].

Interestingly, the penetrance of *TBX4* mutations for skeletal abnormalities is much higher than for PAH [262]. Descriptions of paediatric subjects with mutations in *TBX4*, and likely loci nearby including *TBX2*, characterised by skeletal dysplasias, developmental delay and hearing loss have been reported in CHD and PH patients [30, 383]. Interestingly, the US PAHBB study confirmed a causative role of *TBX4* not only across various age groups (12/266; 4.6% - paediatric-onset; 11/2345; 0.47%—adult-onset) but also across different PH phenotypes: 2 (one adult; one paediatric)/23; 8.7% PAH-CTD, 3 (two paediatric and one adult-onset)/23; 13% PAH-CHD [598]. Not all patients harbouring deleterious variants in *TBX4* fit neatly into Group 1 PH, some individuals with heterozygous mutations in *TBX4* or large deletions encompassing *TBX4* were characterised by death in infancy secondary to acinar dysplasia [166], congenital alveolar dysplasia and pulmonary hypoplasia [496]. The studies dissecting phenotype by mutation type or looking for subtle features (i.e. SPS) among patients harbouring deleterious variants in *TBX4* who were initially diagnosed with PAH are lacking.

1.7 Missing heritability in the postgenomic era

As already alluded to, missing heritability is the proportion of phenotypic variance not explained by the additive effects of known variants. There are multiple sources of missing heritability beyond the rare and common sequence variation described above (Fig. 1.1). In the following paragraphs, studies which contribute to a better understanding of the genetic landscape in PAH, beyond germline, rare, and common sequence variation will be discussed.

Unknown Genetic Variation

Recently, large case-control studies have allowed us to discover new genes involved in the pathogenesis of PAH; all these studies, however, assumed that PAH is a monogenic condition and examined only protein-coding space. Although this permitted the discovery of more than 16 genes associated with PAH, and increased the diagnostic yield, most PAH cases remain unexplained, suggesting additional variation in non-coding regions. It is now known that the remaining 98% of non-coding space plays a crucial role in the regulation of gene expression and requires interpretation using its 3D structure that allows regulatory elements to function across long distances [187]. Regulatory elements such as enhancers, promoters and actively transcribed genes are located in open-chromatin regions so they can be accessed by transcriptional machinery. A number of methods for the identification of non-coding regulatory elements (NCRE) have been developed over the last 20 years [397]. A recent study investigated variation in non-coding regions by researching differences in chromatin marks and gene expression in PAECs from patients with I/HPAH and controls. Using ChiP-Seq profiling of the active histone marks (H3K27ac, H3K4me1, H3K4me3) and RNA-Seq, it was shown that while PAECs from PAH patients and healthy controls are similar in terms of gene expression, they undergo large-scale remodelling of the active chromatin regions, which leads to differential gene expression in response to external stimuli. Overall, this study not only validated and expanded our knowledge regarding the genes involved in the pathogenesis of PAH, but also highlighted that steady-state expression analyses are of limited value in systems where the mechanism involves an aberrant response to stimuli [435].

So far, mostly germline mutations have been implicated in the pathogenesis of PAH but the ever-increasing age of onset may suggest a role of acquired somatic mutations in disease development. While rare somatic mutations of the *BMPR2* gene within pulmonary vasculature cells could cause or aggravate *BMPR2* haploinsufficiency, previous work has failed to find this to be true among those with a germline *BMPR2* mutation [329]. Conversely, a study by Aldred et al. [9] showed a high frequency of genetically abnormal subclones

of PAEC from patients with PAH including one patient harbouring a deleterious *BMPR2* variant. Similarly, Drake et al. [111] demonstrated that a somatic deletion on chromosome 13, encompassing *SMAD9*, in PAECs from a patient with PAH associated with CHD could have contributed to the pathogenesis of PAH, and potentially also to CHD [111]. Likewise, postzygotic mutations have been implicated in the pathogenesis of HHT [47] and HHT with concomitant PAH [125, 348]. Somatic mutations in *TET2* have recently been reported in PAH cases, and the phenotype was replicated in the mouse model [415].

Impaired bioenergetics is an established feature of PAH [96, 437] whether its origins can be traced back to mitochondrial DNA (mtDNA) remains an open question, but a recent study revealed that mitochondrial haplogroups influence the risk of PAH and that susceptibility to PAH emerged as a result of selective enrichment of specific haplogroups upon the migration of populations out of Africa [131]. Studies into the role of mtDNA in PAH will require novel bioinformatics tools and variant prioritisation approaches to account for differences between nuclear and mitochondrial genomes [59].

1.7.1 Epigenetic Inheritance

Epigenetics are heritable changes in gene function that occur without a change in the sequence of DNA and include DNA methylation, histone modification and RNA interference. All three of these modifications have been shown to be involved in the development of PAH. Epigenetic changes can be heritable [40, 75], *de novo* and may be modified by environmental factors such as diet, smoking, and drugs.

DNA Methylation

DNA methylation is a well-described form of epigenetic modification which involves the transfer of a methyl group to cytosine residues in dinucleotide CpG sequences of DNA resulting in gene silencing. In PAH, the best described is hypermethylation of the gene encoding superoxide dismutase 2 (*SOD2*), an enzyme involved in H_2O_2 regulation which also acts as a tumour suppressor gene. Hypermethylation of *SOD2* leads to uncontrolled cell proliferation in multiple cancers; similarly reduced expression of *SOD2* has been shown in the rat model of PH [17] and in IPAH [54]. Further evidence comes from a recent study which found rare germline and somatic mutations in the key regulator of DNA demethylation, *TET2*, in 0.39% of PAH cases [415]. The phenotype was replicated in the $Tet^{-/-}$ mouse and reversed with IL-1β blockade.

Histone Acetylation

Histone acetylation is the most common form of histone modification [236], with a proven impact on PAH development. In rat and human PH histone deacetylases, HDAC1 and HDAC5 concentrations are elevated and histone deacetylase inhibitor can reverse hyperproliferation of PASMCs and decrease mPAP and RV hypertrophy [592].

RNA Interference

A large number of non-coding RNAs (ncRNAs), including long non-coding RNAs (lncRNAs), miRNAs, circular RNAs (circRNAs), and piwi-interacting RNAs (piRNAs) are involved in the regulation of a variety of cellular processes. For example, so far, almost 27,000 lncRNAs have been identified in the human genome [221] and reported regulating gene expression via diverse mechanisms [487]. In PAH, both miRNA and lncRNAs have been studied and reported to impact cells relevant to PAH pathogenesis, including PASMCs, PAECs and fibroblasts. Regulatory effects of lncRNA in PAH range from enhanced proliferation of PASMCs (H19 [495], PAXIP-AS1 [247], MALAT1 [560], Inc-Ang362 [302], Tug1 [582], HOXA-AS3 [589], MEG3[580], TCONS_00034812 [316], and UCA1 [600]), over suppressed proliferation and/or migration of PASMCs (MEG3 [499, 595], CASC2 [176], and LnPRT [73]) to induction of EndoMT in PAECs (MALT1 [579] and GATA6-AS [378]). Significant causative links exist between multiple miRNAs and *BMPR2* expression and signalling [253, 421, 60], as well as hypoxia [68]. Importantly, although attractive as a therapeutic target, miRNAs have not been shown to have the potential to reverse the disease, which is most likely due to the high degree of redundancy and "fine-tuning" rather than "switch on/off" mode of action of these molecules.

Although still limited, our knowledge of ncRNA in the pathobiology of cardiovascular diseases [540, 523] including PAH, is rapidly growing. The major limitation of studies so far is that the focus has been mostly on identifying the role of preselected ncRNA with very limited effort to investigate ncRNA in a global manner accounting for multiple interactions.

Interactions

Although difficult to estimate, gene and environment covariation and gene-environment interactions may play a significant role in heritability, expressivity and penetration of the disease. The most commonly studied factors are drugs, diet, toxins, radiation and stress, and some of these factors have been implicated in the development of PAH. Definitive associations between PAH and drugs and toxins based on outbreaks, epidemiological case-control studies

or large multicentre series have been confirmed for aminorexigenes, methamphetamines, dasatinib and toxic rapeseed oil, yet some others require further validation [471]. Interestingly, several studies reported that *BMPR2* expression and degradation can be affected by viral proteins and cocaine [76, 117, 100]. Similarly, a number of volatile organic compounds (VOCs) identified in exhaled breath condensate in association with PAH, are exogenous [343] and may be considered pollutants due to exposure to cigarette smoke, air pollution and radiation [196]. Contribution of volatile compounds was also reported in the pathogenesis of PVOD [361].

A specific example of the environmental effects on phenotype is the foetal origins of adult diseases (FOAD) hypothesis. This hypothesis was based on the studies which reported that intrauterine exposure increased the risk of specific cardiovascular and metabolic diseases [33, 34, 197] as well as impacted on lung function [492, 45, 341] in adult life. In the area of PH, prenatal exposure to antidepressants and persistent pulmonary hypertension of the newborn (PPHN) has been a subject of multiple studies. A meta-analysis revealed that the risk for PPHN in infants exposed to SSRIs during late pregnancy is small, although significantly increased [186]. In systemic hypertension, the transmission of gut microbiota from parent to offspring was shown to influence the disease risk in adulthood [345] and similarly, bacterial translocation may contribute to PAH development [429]. Air pollution was also shown to correlate with PAH severity and outcomes [478].

With the advent of NGS, the notion of digenic or oligogenic inheritance has gained traction. HPAH has been historically considered a monogenic condition but incomplete penetrance may indicate that other germline or somatic variants are required for the disease to develop. Therefore, at least in a proportion of cases, the inheritance may be digenic or even oligogenic. In the true digenic model, both genes are required to develop the disease. Conversely, in the composite class model, a variant in one gene is sufficient to produce the phenotype, but an additional variant in a second gene impacts the disease phenotype or alters the age of onset [163]. The latter model seems to be plausible in PAH, where co-occurrence of the variants in different PAH risk genes has been reported to impact on disease onset and penetrance [119]. Patients harbouring deleterious variants in more than one PAH risk gene have been reported in case reports [562], small [416] and large cohorts of HPAH patients [179, 51, 564]. Finally, the phenomenon of synergistic heterozygosity, whereby the effect of susceptibility genes is enhanced by modifier genes, in which common variants may influence the disease onset and severity, has been described in PAH [407].

Population Dependent Heterogeneity

It is well recognised that although rare diseases can occur in any population, some ethnic groups are characterised by their higher incidence, i.e., sickle cell disease is more common in African, African-Americans and Mediterranean populations and similarly, Tay-Sachs disease occurs mostly in people of Ashkenazi Jewish or French Canadian ancestry. National and international PAH registries have shown differences in demographic characteristics between European and East Asian patients. East Asian IPAH cohorts resemble those of European cohorts from 40 years ago; additionally, they show more pronounced female predominance, a younger age of onset and lower comorbidity burden, and importantly, they demonstrate that significantly more cases can be explained by variation in known PAH risk genes (24% [179, 598] vs. 39% [564]).

1.8 Clinical utility and therapeutic implications of genetic testing in PAH

1.8.1 Clinical utility of genetic testing

The utility of genetic diagnosis cannot be overestimated since it explains aetiology, informs prognosis and treatment decisions, and allows risk stratification of family members. Genetic testing has the potential to mitigate the disease course [301, 366, 349, 79] and has been recommended by current practice guidelines [301, 366, 349, 79]. The influence of the mutation on outcomes has been well described in *BMPR2* and *EIF2AK4* mutation carriers. While *BMPR2* and *ACVRL1* mutation carriers present at a younger age, with more severe haemodynamics and worse survival than patients without pathogenic PAH mutations [124, 170, 504], *EIF2AK4* mutation carriers present at a younger age than non-mutation carriers but the mutation status does not seem to affect prognosis [360]. Identification of *EIF2AK4* mutations allows diagnosis confirmation not only in PVOD/PCH cases but also in patients who clinically presented as I/HPAH, eliminating the need to perform a lung biopsy [152]. Due to the dismal prognosis related to PVOD/PCH diagnosis, early referral for transplantation is warranted [194].

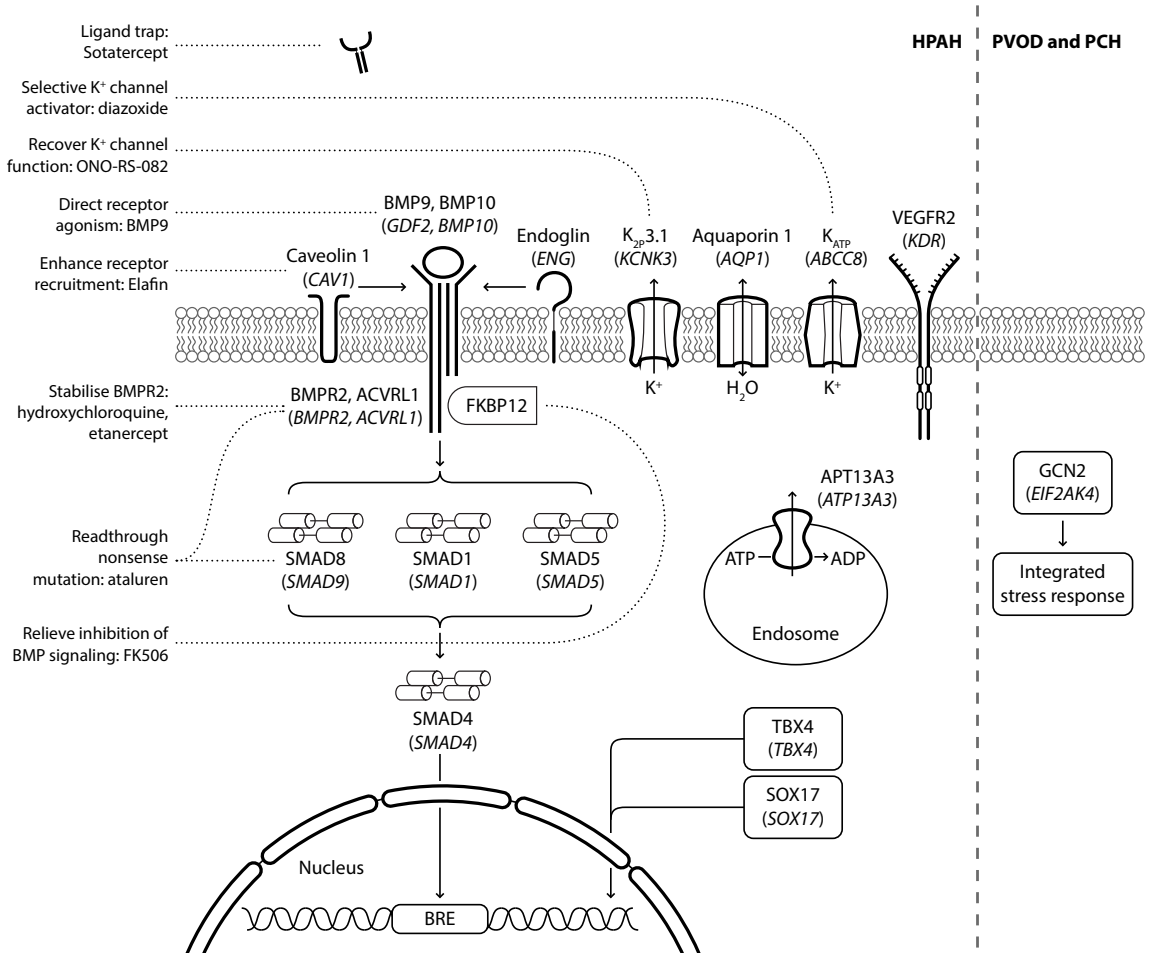
1.8.2 Pharmacogenetics

Knowledge about the genetic makeup of the individual allows targeted treatments enhancing efficacy and decreasing the risk of potential side effects [377]. Despite that only a small

number of cardiovascular drugs are provided with pharmacogenetic information in drug labels in an attempt to guide prescription practices [396]. Given the central role of BMP signalling in the pathogenesis of PAH, it is not surprising that therapies to enhance or rescue the BMPR2 pathway have gained most traction (Fig. 1.7). In preclinical studies, BMP9 administration in heterozygous *BMPR2* knockout mice had a positive effect on haemodynamics, suggestive of a possibility to overcome BMPR2 haploinsufficiency by up-titration of the ligand [353, 544]. Likewise, ataluren reads through nonsense mutations in *BMPR2* and *SMAD9* and restores the full-length proteins [112]. Chloroquine prevents progression of experimental PH through inhibition of lysosomal degradation of BMPR2, leading to increased receptor density at the surface of the endothelial cells [116]. Similarly, a TNF-β inhibitor, etanercept, reduces inflammation, receptor shedding and proteasomal degradation of BMPR2 [234].

As reduced BMPR2 signalling is a known phenomenon even in patients without the mutation, targeting *BMPR2* modifier genes might be an effective rescue mechanism. Enzastaurin was shown to rescue the *BMPR2* modifier gene, fragile histidine triad (*FHIT*) and reverse animal PH [103]. A further promising and direct approach under investigation is the enhancement of BMPR2/ALK1 signalling by supplementation of BMP9-based proteins [318]. Ligand traps inhibiting negative BMPR2 regulation are another therapeutic option. Recently, Acceleron announced positive results from the sotatercept Phase 2 PULSAR trial in patients with PAH [29]. Sotatercept is a novel recombinant fusion protein that inhibits TGF-β superfamily members including GDF11 and activin A and B and is claimed to restore the imbalance in the BMPR2 signalling pathway. Both EMA and FDA granted Sotatercept a breakthrough therapy status allowing expedited development process. In the future, patients with deleterious variants in *CAV1*, may benefit from elafin, a peptidase inhibitor 3, encoded by PI3 gene [29]. In sugen-hypoxia rats, elafin reduced elastase activity and reversed PH, as well as improved endothelial function by increasing apelin levels [381]. Tacrolimus (FK508), a potent immunosuppressant, was shown to increase BMP signalling via blocking FK-binding protein 12 in an animal PH model. The preclinical study confirmed the safety and tolerability of this agent in PAH [486]. Gremlin 1 secreted by vascular endothelium may inhibit BMPR2 signalling. Neutralising antibodies interfering with gremlin 1 proved effective in ameliorating chronic hypoxia/sugen-induced PAH in mice [82].

Besides the BMP pathway, modulating ion channel functions might be an effective therapeutic method. There is evidence that reduced potassium channel conductance in some *KCNK3* mutants can be recovered by the phospholipase A2 inhibitor ONO-RS-082 [323]. Similarly, *ABCC8* variants can be rescued by the SUR1 activator, diazoxide [51]. Fig. 1.7 presents druggable molecular pathways involved in the pathogenesis of PAH.



Abbreviations: HPAH: heritable pulmonary arterial hypertension, BMP: bone morphogenetic protein, *BMPR2*: bone morphogenetic protein receptor type 2, *ACVRL1* (also known as *ALK1*): activin receptor like 1, *ENG*: endoglin, *CAV1*: caveolin 1, *KCNK3*: Potassium Two Pore Domain Channel Subfamily K Member 3, *ABCC8*: ATP binding cassette subfamily C member 8, *SOX17*: SRY-Box Transcription Factor 17, *SMAD*: SMAD Family Member; *TBX4*: T-Box Transcription Factor 4, PVOD: pulmonary veno-occlusive disease, PCH: pulmonary capillary haemangiomatosis, *ATP13A3*: cation-transporting ATPase13A3, BRE: BMP-responsive element, GCN2: general control nonderepressible 2, *EIF2AK4*: Eukaryotic Translation Initiation Factor 2 Alpha Kinase 4, FKBP12: 12kDa FK506-binding protein

Fig. 1.7 Druggable molecular pathways involved in the pathogenesis of PAH.

1.8.3 Knowledge, attitudes and barriers towards genetic testing and counselling

Despite recent developments in understanding the genetic background of PAH and the potential therapeutic implications, almost 80% of physicians caring for PAH patients never or rarely refer their patients to genetic counsellors or order genetic testing [246]. At least in the US, the most frequent reasons for that were reported to be lack of insurance coverage and limited access to genetic counsellors. Interestingly, the most important driver for genetic testing was patient inquiry [246]. This situation might be different in Europe where healthcare is publicly funded [171], but it varies from country to country. In any case, a forward-thinking and innovative regulatory environment is necessary for integrating research and technological advances into clinical practice. In this respect, NHS England's Genomic Medicine Service is likely to revolutionise routine patients' care in the UK. Similar small and large scale initiatives are booming around the world. To give an example, the use of rapid WGS pioneered by Dr Stephen Kingsmore and his team at Rady Children's Institute for Genomic Medicine enables precise diagnoses for critically-ill newborns within 26 hrs [354].

1.9 Summary of chapters and aims of study

This thesis follows the traditional layout consisting of Introduction, Methods, Results and Discussion. In the Introduction, I presented the research approaches to uncover the genetic architecture of PAH starting with forward phenotyping, forward and reverse genetics, and finally reverse phenotyping (Fig. 1) and additionally, I summarised the state of the art knowledge of genetic architecture of PAH as well as potential sources of "missing heritability". In the Methods section, I firstly describe the study setup and the approaches employed in clinical data collection, verification and harmonisation as well as expansion of HPO terminology to enable standardised annotation of the phenotypes of PAH cases. I then focus on statistical and computational methods used in forward phenotyping that resulted in the identification of stable homogeneous groups that were subsequently used in a case-control rare variant association study. Following that, I describe a Bayesian model comparison method used for rare Mendelian variant discovery. In the Results section I firstly present the cohort characteristics, secondly report the results of clinical and computational phenotyping, and thirdly, delineate the rationale behind using selected phenotypes in gene discovery. In the most pertinent section of the results chapter, I detail the results of the case-control study and the discovery of the new PAH risk gene (*KDR*, encoding for VEGFR2), before detailing the

outcomes of the reverse phenotyping for the two most pertinent genetic findings (*GDF2*, *KDR*).

I discuss each result section separately. Firstly, I talk about how the study cohort compares to the US PAHBB, published registries and biobank studies, and secondly, comment on bias, study limitations and suggest paths forward. Next, I describe the results of forward phenotyping (both clinical and computational), provide evidence for the utility of such phenotypes in genetic studies and future perspectives revolving mostly around intermediate phenotypes derived through multi-omics methods. Finally, I discuss the main finding of this research project, which is the identification of *KDR* as a new PAH risk gene, and other candidate genes, before detailing reverse phenotyping.

I conclude by restating the aims of the study, summarising my main research findings and explaining their significance and contribution to the field while acknowledging also the limitations. Finally, I will outline my plans for future research.

The aims of this Thesis were as follows:

1. To collect, harmonise and verify deep phenotype data on patients with PAH recruited to the NBR and the PAH Cohort study and the US PAHBB study [Chapters 2.1, 2.2, 2.3, 2.4, 2.5, 3.1].

2. To enrich the Human Phenotype Ontology database for terms relevant to PAH and associated diseases which would enable discovery of computational phenotypes [Chapter 2.5.7]

3. To establish distinct clinical and computational phenotypes of patients with PAH that carry diagnostic, therapeutic and prognostic information as well as allow genotype-phenotype association analysis in enriched disease cohorts [Chapters 3.2.1, 3.2.2].

4. To define specific genotype-phenotype associations and their utility in clinical practice [Chapter 3.4].

Chapter 2

Methods

This chapter aims to present methods employed to collect, harmonise and verify clinical and genetic data for my PhD project. Moreover, I present approaches I used to manage missingness and standardise clinical information so it can be used for ontological phenotyping. Additionally, I describe machine learning algorithms used for computational phenotyping and the Bayesian framework applied in the genetic case-control analysis.

2.1 Study design

The National Institute for Health Research BioResource Rare Diseases study (NBR) was the pilot study for the 100,000 Genomes project and was established to identify genetic causes of rare diseases through whole genome sequencing and deep phenotyping. The study recruited to 18 domains defined either as a single disease or a group of disorders (Table 2.1); the PAH domain recruited patients with idiopathic or heritable PAH, pulmonary veno-occlusive disease (PVOD) and pulmonary capillary haemangiomatosis (PCH). Adult and paediatric onset cases were allowed as well as incident and prevalent cases; unaffected relatives of childhood onset cases were also recruited. Clinical information was recorded only at the time of initial diagnosis.

The Cohort study of idiopathic and heritable PAH is an observational, prospective and longitudinal study of patients with idiopathic and heritable PAH. Follow-up information is collected as a part of routine clinical care approximately every six months (since March 2019 at 1 and 3 years since diagnosis). The study allows recruitment of both incident and prevalent cases (since March 2019 only adult incident cases allowed, whereas both prevalent and incident paediatric cases are permitted). Patients consented to the study agree to have blood taken for next-generation sequencing and other omics studies. Alongside cases, healthy

relatives are invited to participate. Relatives are seen on a yearly basis for blood and urine sampling, epidemiology questionnaire, and clinical assessment. Adult healthy controls are recruited for comparison studies (Table B.4).

2.2 Ethics

The NBR and The Cohort I/HPAH studies were reviewed by local Research Ethics Committees and received favourable opinions (REC REF: 13/EE/0325 and 13/EE/0203). All participants provided written informed consent. Institutional review boards at Cincinnati Children's Hospital Medical Center (CHMC) and at the US PAH Biobank approved the validation cohort studies and written informed consent was obtained at enrolment (see Chapter 2.5.8). Patients diagnosed at Royal Papworth Hospital with I/HPAH donated blood samples for measurement of transpulmonary gradients of BMPR2 ligands with informed consent and approval of the local research ethics committee (Reference number 15/EE/0201) (see Chapter 2.5.8).

2.3 Setting

Patients diagnosed with I/HPAH, PVOD or PCH, relatives of index cases and unrelated healthy controls were recruited at nine UK centres and by international collaborators at University of Paris, University of Giessen, and hospitals in Graz, Pavia and Amsterdam between 3^{rd} December 2013 and 28^{th} February 2018 for the NBR study and 14^{th} January 2014 until present day for Cohort I/HPAH study. Patients recruited from UK centres who were participants in the Cohort study were followed up for a median 6.9 years.

2.4 Participants

In total, 1394 patients, 90 relatives, and 118 healthy controls were recruited to the NBR and the I/HPAH Cohort studies (as per 15^{th} July 2020). All patients provided written informed consent. Both prevalent and incident cases were allowed. Prevalent cases were defined as diagnosed earlier than six months before the study initiation. Patients consented only to the NBR study had a snapshot of their clinical information recorded from the time of diagnosis, whereas patients in the Cohort study were followed longitudinally as part of their clinical PAH care. All cases were diagnosed between January 2002 to July 2020, and diagnostic

classification was made according to international guidelines using a multidisciplinary assessment that included echocardiography, comprehensive blood testing, pulmonary function testing, overnight oximetry, isotope perfusion scanning, high-resolution computed tomography and right heart catheterisation. Patients with PAH associated with anorexigen exposure were considered as IPAH, whereas HPAH was defined by the presence of a positive family history of PAH. Clinical, functional and hemodynamic characteristics at the time of PAH diagnosis were prospectively entered into the database. The date of diagnosis corresponded to that of confirmatory right heart catheterisation. Following diagnosis, subsequent treatments and follow-ups were at the discretion of the treating physician, according to contemporary guidelines. In most centres, patients were seen every 3-6 months with an assessment of functional status and exercise capacity. Right heart catheterisation was repeated when considered necessary by the responsible clinician. Study visits were performed every 6 months. Relatives were seen on a yearly basis for clinical assessment (as per study protocol), blood and urine sampling and epidemiology questionnaire, whereas healthy controls were sampled only once and had clinical information recorded at the time of sampling.

2.5 Data sources

2.5.1 Clinical data capture, processing, quality control and release

Pseudonymised results of routinely performed clinical tests reported in either clinical case notes or electronic medical records (EMR) were manually entered by data officers at each center into a web-based *OpenClinica* (OC) data capture system (Community edition). Data available in batches (data dumps from local research registries and data from international collaborators) were received in Excel spreadsheets. I subjected this data to verification procedures, harmonised them with *OpenClinica* standards and automatically uploaded them. Twenty one electronic Clinical Case Report Forms (eCRFs) (Table B.5) distributed across seven events (Diagnostic, Continuous data, Follow-up, Epidemiology questionnaire, Suspension, Relatives, Unrelated healthy control) were constructed to accommodate routinely available clinical information; the OC DataSet export feature was used for automated file exports. All data items were selected for export and data were exported separately per event. The automated pipeline transferred the files to a release folder on an internal server and executed the load process. The load process was written in Python and loaded each file separately into the MySQL database. After loading the data, the version mapping was performed with a Python script using the mapping instructions. R language was used to

extract data from the MySQL database and to convert units based on mapping instructions, and finally the results were stored as an R object. I applied a set of rules to the data in order to verify the outliers. I then sent data queries to the centres and the responses were stored in an Excel file subsequently used to confirm true outliers. Logical inconsistencies have been corrected directly on *OpenClinica*. Identifiers were validated for consistency and the cleaned data was stored as an R object, ready for further analysis. Data dumps were scheduled to happen on a bi-weekly basis.

The Clinical School Computing Service (CSCS) provided support and computational resources to host a virtual server. Access to the system was controlled through user accounts and authorisations, importantly access to all data was restricted to study coordinators and data analysts, whereas sites had access only to their own data. Curated and verified data were made available on request to accompany biological samples for external projects. Standard operating procedures were made available on a public facing system called TROI. TROI also hosted a virtual machine with an OC installation to test eCRF changes and automatic uploads (Fig. 2.1).

2.5.2 Clinical data verification and analysis

To aid data analysis and improve data quality, I introduced a number of quality assurance procedures, listed below.

Data checks were written in the R language and consisted of:

1. Logical verification checks i.e. systolic/mean/diastolic pulmonary artery pressure sPAP $>$ mPAP $>$ dPAP; FVC [L] $>$ FEV1 [L], CO $>$ CI unless BSA $<$1

2. Checks for discrepancies between:

   (a) Comorbidities and ICD10 codes using the System of Ontology-based Re-coding and Technical annotation (SORTA) [390].

   (b) Comorbidities and medication i.e. comorbidity reported but lack of medication or vice versa.

   (c) Reported and calculated acute vasoresponse test results.

   (d) Reported lung emphysema, thromboembolic disease and fibrosis and free text search in CT reports for these abnormalities.

   (e) Reported lung function abnormalities and provided numerical values.

3. Missingness assessment and handling:

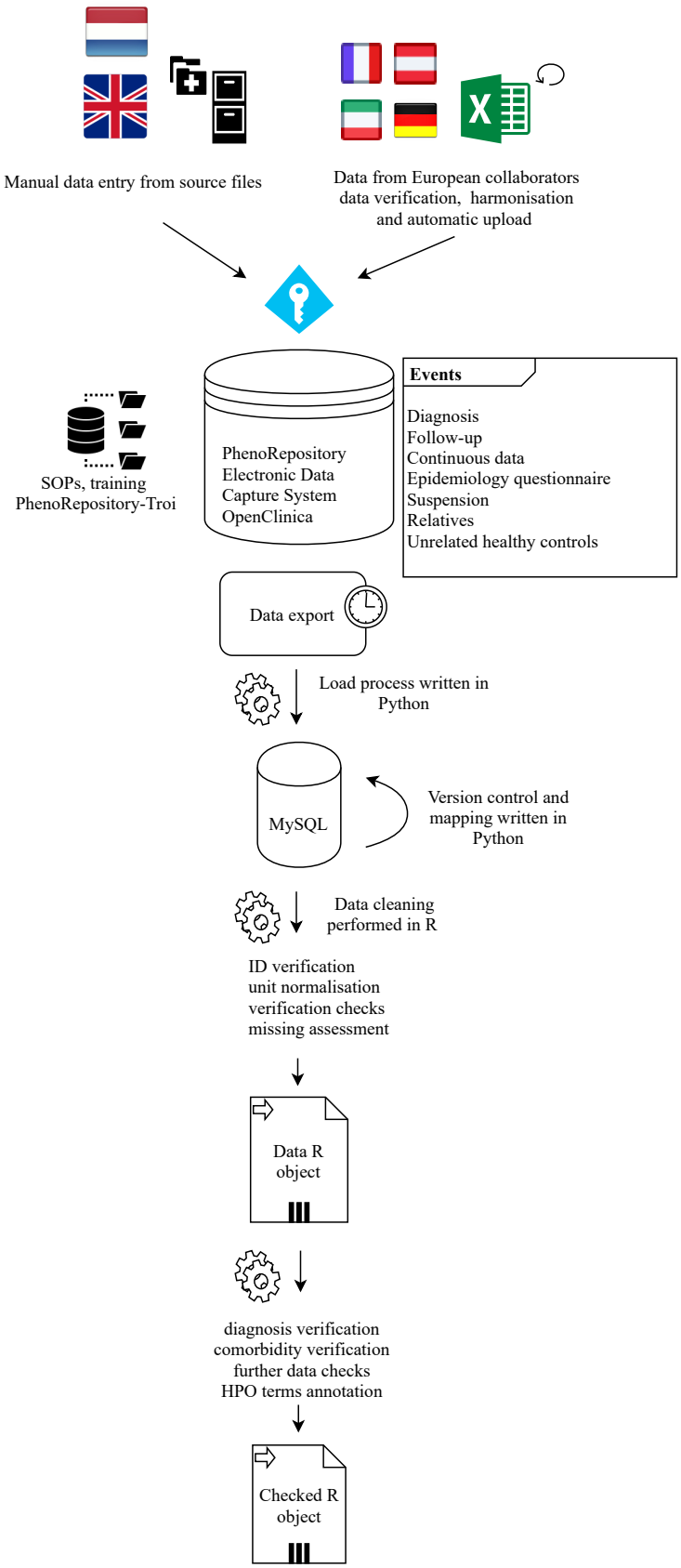


Fig. 2.1 Data collection, harmonisation, verification pipeline (The NBR and Cohort studies).

(a) Thirty-six core variables (of diagnostic and prognostic significance) were identified and a missing rate below 10% was established. Missingness tables were reported back to the centres for clarification.

(b) Subsequently, missingness in the remaining variables was assessed and checked for logical inconsistencies, i.e. PVR reported but PAWP and CO missing, CI missing while CO and BSA available.

(c) The proportion of missing data, missing data mechanisms and missing data patterns were assessed with the "*VIM*" [277] and "*Amelia"* [222] R packages.

4. Diagnosis verification:

(a) Based on the clinical information provided, a diagnosis verification algorithm was implemented in R to red-flag patients with a reported diagnosis of I/HPAH but clinical features indicative of a different diagnosis. Red-flagged features were the following:

i. Not meeting hemodynamic diagnostic criteria.
ii. History of pulmonary embolism or CT scan report suggestive of CTEPH.
iii. History of structural heart abnormality (PFO/ASD, VSD, PDA) or CT scan or heart MRI scan indicative of such abnormality.
iv. Positive autoimmune screen and/or history of connective tissue disease (CTD).
v. History of left heart disease.
vi. History of chronic obstructive pulmonary disease.
vii. History of anorexigen use.
viii. History of lung fibrosis or CT scan with significant fibrotic changes.
ix. History of sarcoidosis.

Excel spreadsheets were sent to the centres to verify the diagnosis. Patients identified as non-I/HPAH/PVOD/PCH were suspended, and the correct diagnosis was recorded. Patients with PFO or small ASD, which were deemed haemodynamically irrelevant were allowed and labelled as IPAH. Patients with larger intracardiac defects with significant shunt had diagnoses changed to PAH-CHD. In all but three individuals diagnostic labels were assigned. Patients with missing diagnostic labels were excluded from the analysis. All diagnostic tests including PFTs were performed in treatment naive patients, maximally 4 weeks apart. Spirometry, gas transfer and lung volumes assessed by body plethysmography were measured in each

centre according to European Respiratory Society/American Thoracic Society guidelines [90, 1, 331, 565]. The DLCO per alveolar volume was calculated by dividing the DLCO by VA. Quality control followed accepted guidelines, with DLCO gas analyser calibrations prior to each session, volume calibrations prior to each patient test and biological calibrations daily. The European Community of Coal and Steel (ECCS) prediction equations were used to define normal PFT values [423] in all but one centre (Glasgow); for this centre, % predicted values were recalculated to ECCS.

2.5.3 Ethnicity

NBR study collected and coded the self-reported ethnicity information as per The Office of National Statistics: White: A – British, B – Irish, C – Any other White background, Mixed: D – White and Black Caribbean, E – White and Black African, F – White and Asian, G – Any other mixed background, Asian or Asian British: H – Indian, J – Pakistani, K - Bangladeshi, L – Any other Asian background, Black or Black British: M – Caribbean, N – African, P – Any other Black background, Other Ethnic Groups: R – Chinese, S – Any other ethnic group, Z – Not stated. Genetic ethnicity was also inferred using Principal Component Analysis (PCA) [538].

2.5.4 Census

I collected information about participants' status every six months for the UK (this was ascertained via the National Health System Digital Spine portal or an equivalent local system) and once a year for international cases. I provided each centre with Excel spreadsheets with all included patients and their latest reported status (active or suspended) and asked each centre to update both the spreadsheet and *OpenClinica* suspension event in case of death, transplantation, and study withdrawal. I subsequently cross-checked *OpenClinica* data dump with the spreadsheets and reported any discrepancies back to the centres.

2.5.5 Risk stratification

Two risk assessment strategies were applied to the data: the Reveal risk score [43] and abbreviated ESC/ERS risk scores [219] were calculated in all patients who had the necessary minimum phenotypic information available. The REVEAL strata were defined as follows: score 0-7 – low risk, 8 – average risk, 9 – moderate risk, 10-11 – high risk, $\geq$12 – very high risk. For ESC/ERS risk score assignment I followed a validation strategy proposed by

Kylhammar et al. [281] and graded low, intermediate and high-risk criteria with cut offs of 1, 2 and 3 respectively. For each patient, the mean score was derived and rounded to define the risk group as per Hoeper et al. [219].

2.5.6 Assessment of comorbidities

I used two previously reported methods to measure patients' comorbidities based on ICD-10 codes. Coding for the Charlson and Elixhauser scores were based on work by Quan et al. [422]. Weights for the Charlson score were based on the original formulation by Charlson et al. [72], while the weights for the Elixhauser score were based on work by Moore et al. [364] and van Walraven et al. [552]. Finally, the categorisation of scores and weighted scores were based on work by Menendez et al. [351]. ICD-10 codes used to calculate Charlson and Elixhauser scores are summarised in (Table B.6).

2.5.7 Standardised vocabulary for description of phenotypic abnormalities

Abnormalities defined either by assessment of categorical or continuous variables (sex- and age-adjusted where necessary) were semiautomatically annotated with appropriate HPO codes. The System of Ontology-based Re-coding and Technical annotation (SORTA) [390] was used to suggest HPO terms for comorbidities based on lexical similarity; similarity score 100% was defined as a perfect match, terms not meeting perfect match criteria were manually annotated (90% of the terms). In cases where HPO terms were not available or lacked specificity, HPO terms were developed along with is-a relationships [269]. In total, over 250 terms were suggested as additions to HPO ontology[1]; all of these terms were discussed in HPO Lung meetings and were accepted with minor changes. HPO modifications were predominantly within the abnormality of pulmonary circulation, abnormality of pulmonary vasculature, abnormality of pulmonary interstitial morphology, abnormality of cardiovascular system electrophysiology, whereas some terms required creation of a new class or changes to the relations with other terms. Each HPO term was developed according to the predefined template consisting of: preferred term label, synonyms, definition, literature references, parent term, list of diseases characterised by the term and the authors nano-attribution (ORCID).

[1] https://groups.io/g/human-phenotype-ontology/message/24

2.5.8 External data sets

The US PAH Biobank

Phenotypic information of 2,572 individuals from the US PAHBB enrolling at 37 centres from the United States was obtained in an Excel spreadsheet and was accompanied by Whole-Exome Sequencing (WES) data. The US PAH Biobank founded by the National Institutes of Health/National Heart, Lung and Blood Institute aims at collecting genetic data, biological samples and clinical information from at least 3,000 WHO Group 1 PAH patients meeting inclusion criteria summarised in (Table B.7). The PAH Biobank phenotypic information consisted of the following diagnostic data: diagnostic group and subgroup, demographics, self-reported race and ethnicity, haemodynamics, WHO functional class, 6-minute walk test, symptoms, drug exposure, medication, smoking status, comorbidities and census. Self-reported race information was categorised into five groups: American Indian or Alaska Native, Asian, Black or African American, Native Hawaiian or Other Pacific Islander, White, the multiracial category was not recorded but the selection of one or more races was allowed. Additionally, self-reported ethnicity was categorised as Hispanic or Latino or Spanish Origin, or Not Hispanic or Latino or Spanish Origin. Data were harmonised and verified in the fashion depicted in Fig. 2.2.

Measurement of levels of BMPR2 ligands (BMP9 and pBMP10) in peripheral blood

BMP9 and pBMP10 levels were measured by ELISA in controls (n = 120) and patients with idiopathic PAH (n = 260) (including some mutation carriers) recruited to the PAH Cohort study (ligand measurements were performed by Dr Paul Upton, Department of Medicine, University of Cambridge).

Sampling for transpulmonary and transcardiac gradients BMPs measurements

I sampled 18 patients with IPAH/HPAH diagnosed at Royal Papworth Hospital during elective right heart catheterisation between 2015 and 2017, allowing for simultaneous sample collection from the superior vena cava, proximal portion of pulmonary artery and radial artery, and haemodynamic measurements. Exclusion criteria included left ventricular systolic and or diastolic dysfunction, significant valvular heart disease, chronic kidney disease stage 4 or 5, chronic liver disease, liver failure or alcohol abuse, current illicit substance use, active infection and peripheral arterial vascular disease. Patients were sampled between 9:30 am

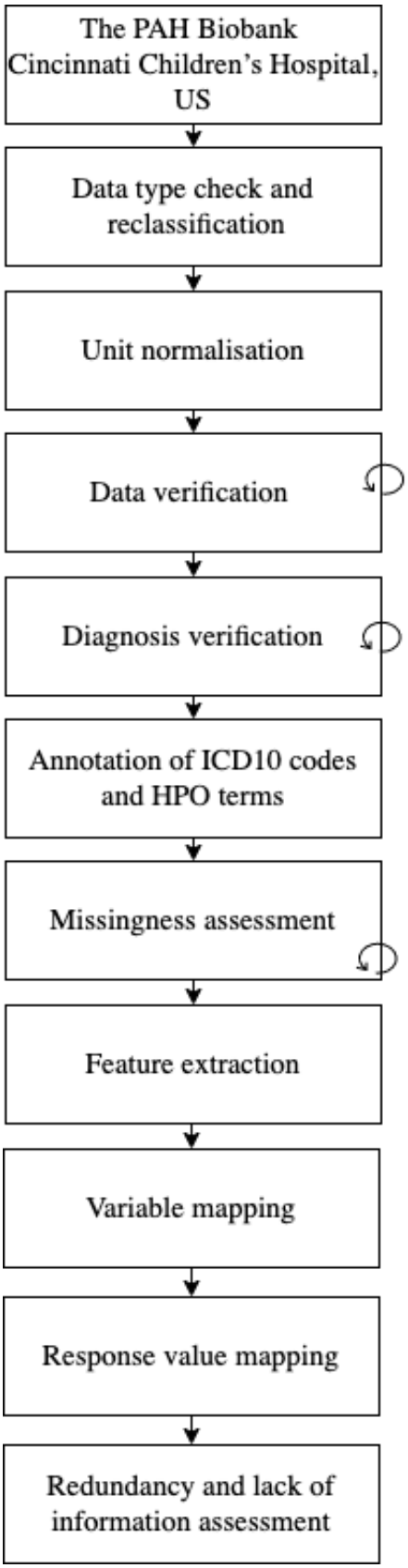


Fig. 2.2 Data collection, harmonisation and verification pipeline (The US PAH Biobank study).

and 12:30 pm. (BMP9 and pBMP10 levels were measured by ELISA by Dr Joshua Hodgson, Department of Medicine, University of Cambridge).

Measurement of autoantibodies in peripheral blood

A comprehensive autoimmune screen, based on 19 autoantibodies previously strongly associated with autoimmune conditions, was performed on 473 patients from the PAH Cohort study. The measurements by immunofluorescence were done at Dr Eoin McKinney's lab at University of Cambridge as a part of a large autoimmunity screen in healthy populations. Positivity was defined as two standard deviations above the cohort mean.

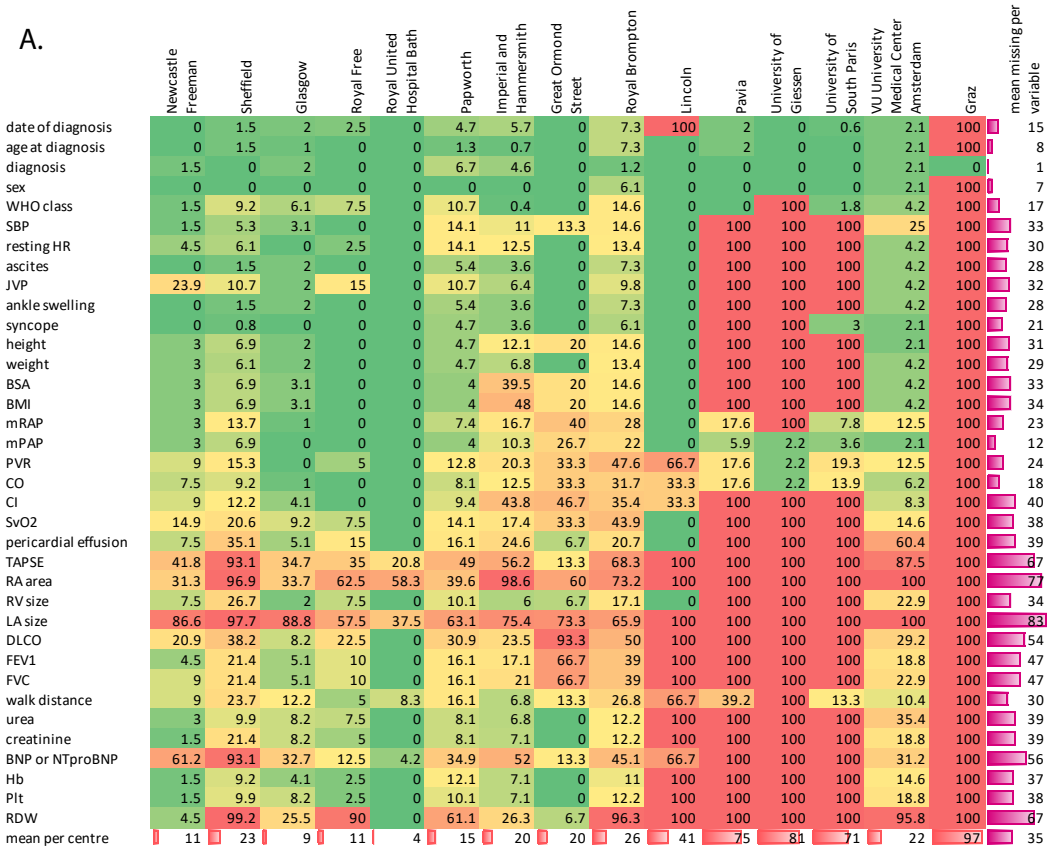
2.5.9 Bias

Observational studies like those presented in this thesis are prone to many problems that can bias the results and lead to a spurious inference. Firstly, selection bias can occur when there is a systematic difference between those who participate in the study and those who do not. This type of bias may affect the generalisability of the results. The NBR study aimed to recruit all patients with I/HPAH and PVOD/PCH in the UK but also allowed inclusion of legacy samples from deceased patients who were included in the study under other ethics (this also applied to all cases recruited by international collaborators). Recruitment of these cases can bias the results in two ways. Firstly, reasons for recruitment for the other studies are unknown. Secondly, data missingness rate in these cases was significantly higher than in NBR recruits, leading to information bias. Effect of confounders cannot be underestimated in this study; the most critical confounders were sex, age at diagnosis and whether the case was prevalent or incident. Regression-based approaches were used to correct for confounders. Last but not least, measurement error needs to be considered when data are pooled from many centres which follow different protocols. Also, the precision of some measurements could have improved over time, making comparisons between measurements taken at different time points difficult (see Chapter 3.2.1). When possible and appropriate I sought this information for particular analyses and made necessary allowances.

2.6 Statistical methods

2.6.1 Methods used to handle missing data

I analysed the clinical data set for missing rate and pattern as a whole, per centre and per patient. In cases where the reasons for missingness were not available, I sent inquiries to the centers. I set up a system of weekly monitoring of missingness rates and feedback loops to the data entry officers at each centre. I devised a list of core variables necessary for ascertainment of diagnosis for which the allowed missingness rate was set at 10%. As a result of these steps, overall missingness dropped from 60% in January 2018 to 48% in July 2020 and the missingness for core variables decreased from 35% to 24% for the same period for diagnostic visit (Fig. 2.3).



(A) Data release 27.01.2018, (B) Data release 15.07.2020.

Fig. 2.3 Heatmap showing percentage of missing data per variable (rows) and per centre (columns) among core variables.

Multiple imputation by chained equations

Multivariate imputation was performed using the multiple imputation by chained equations (MICE) framework implemented in "*mice*" and "*miceadds*" R-packages [548] in the order presented in Fig. 2.4.

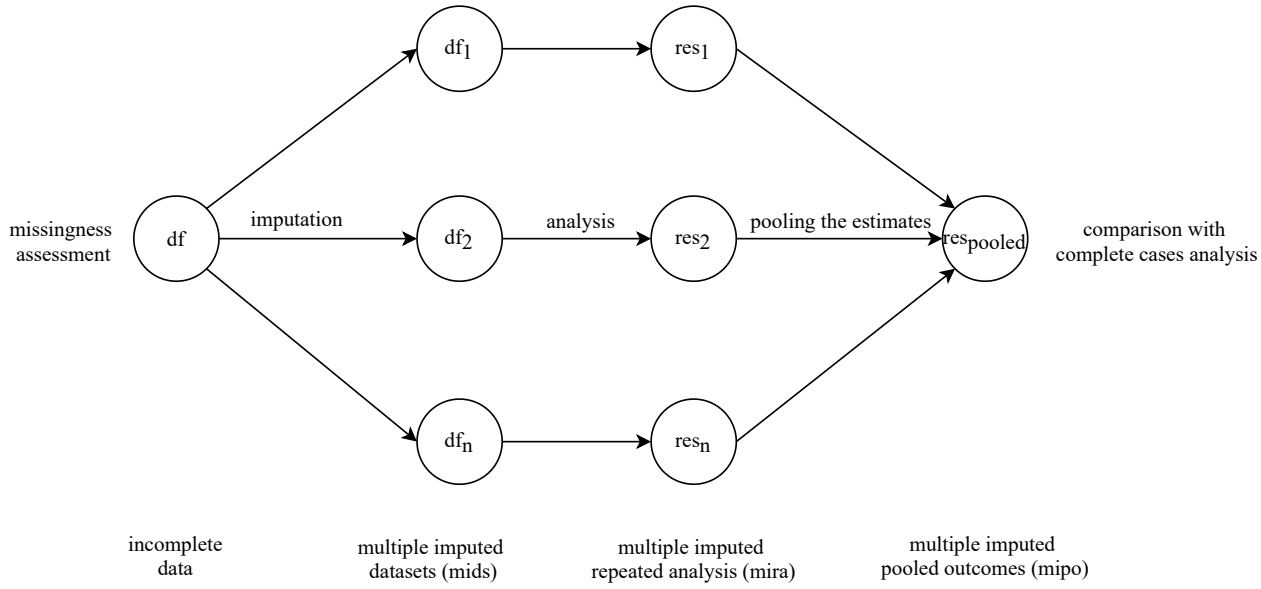


Fig. 2.4 Schematic representation of imputation procedure. Abbreviations: df: data frame; res: results

This method is based on the principle of a repeated chain of regression equations through the incomplete variables, wherein each imputation model, the respective incomplete variable is modelled as a function of the remaining variables [548]. Imputation model included all variables that were necessary in the analysis model, including cumulative baseline hazard function and variables that predicted both the incomplete variable and if the incomplete variable was missing like the centre and whether the case was incident or prevalent. Quality of predictors was assessed using outflux-influx plot (Fig. 2.5).

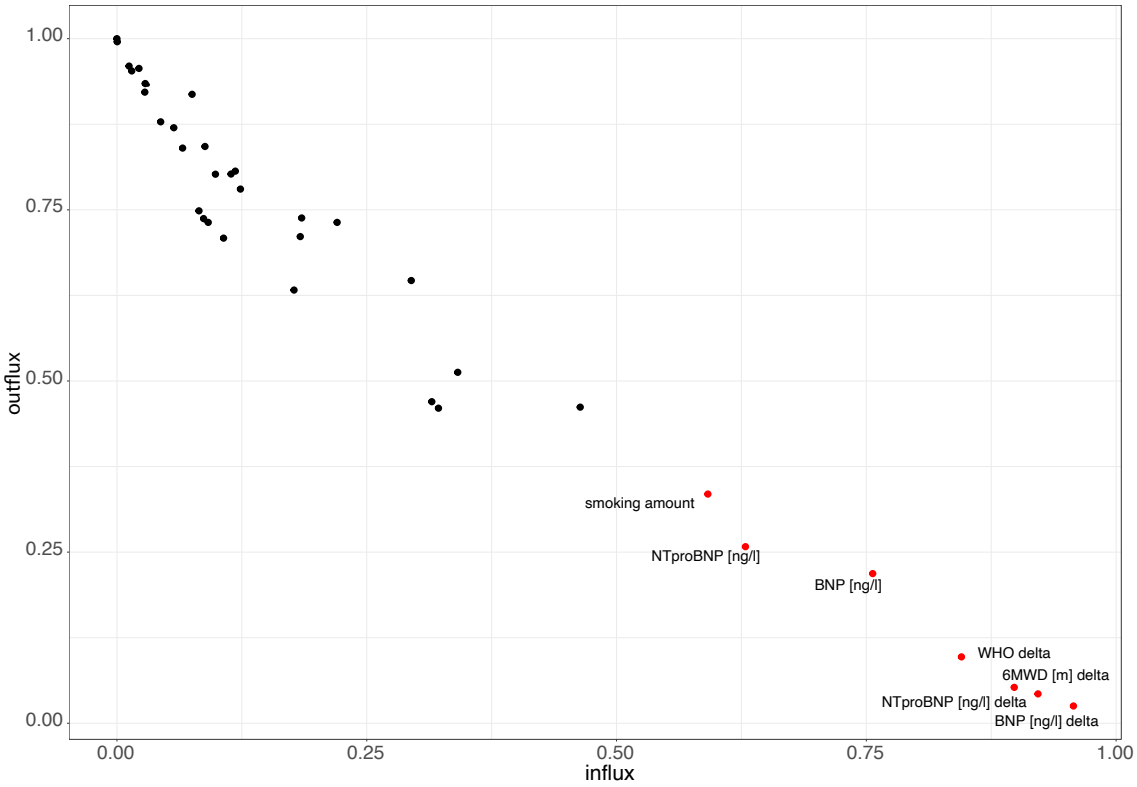
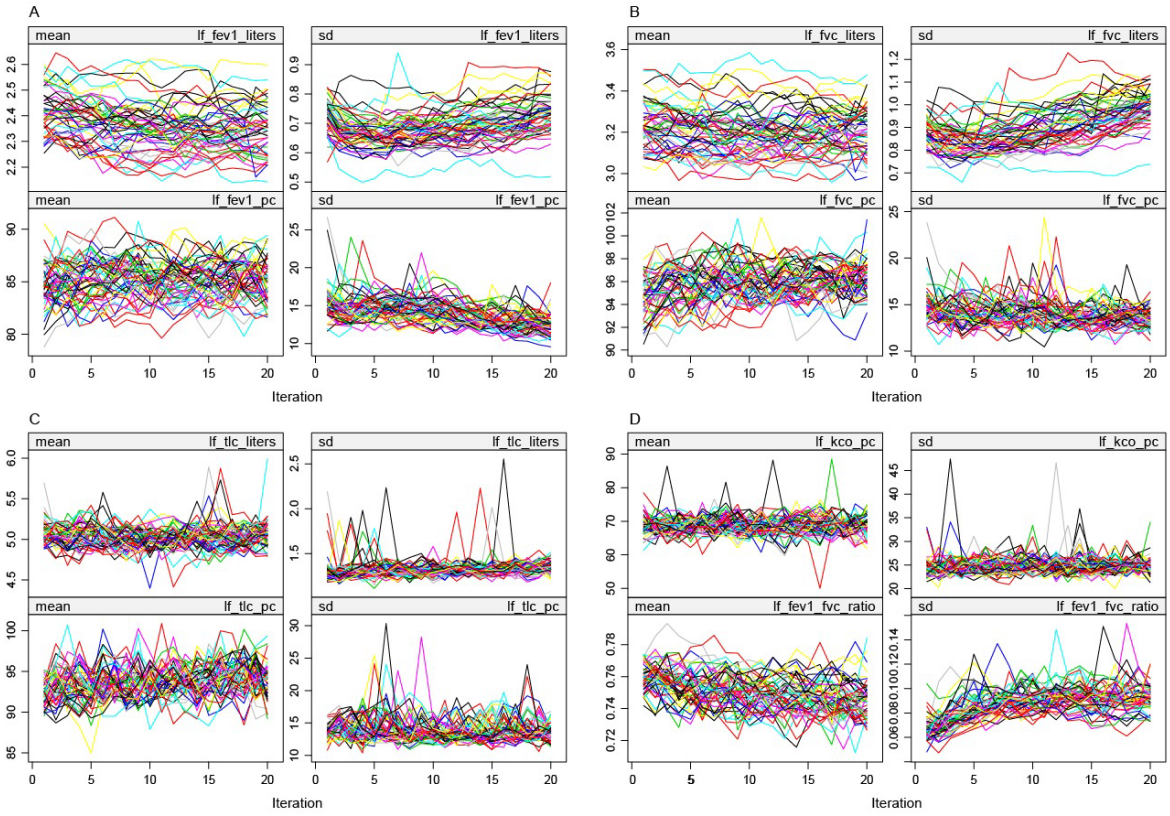


Fig. 2.5 Outflux-influx plot.

Numerical data were imputed with predictive mean matching (pmm), factors with two levels were imputed using logistic regression, factors with more than two levels with multinomial logit model and ordered factors with more than two levels with the ordered logit model. Transformed variables (tertiles, BMI, ratios, score sums) were imputed as just another variable as well as passively with good concordance. The visiting sequence was set to "monotone" to speed up convergence. The number of iterations was set to 20. Following the rule of thumb proposed by White et al. [572], the number of imputations should be at least equal to the percentage of incomplete cases; the procedure was performed at $m = 50$. The convergence of the algorithm was checked and the means and standard deviations of imputed values were plotted over 20 iterations (Fig. 2.6). The streams of numerical and factor variables intermingled and showed no trends at later iterations. Distribution of imputed data was also visually assessed against the distribution of observed data and showed plausible values.



(A) FEV1, (B) FVC, (C) TLC, (D) KCO.

Fig. 2.6 The convergence of the algorithm over 20 interactions.

Statistical analysis performed on complete cases was also repeated on each imputed data set. Estimates and standard errors from those analyses were combined using Rubin's rules [446], which take into account the variability in the estimates between the imputed data sets, reflecting the uncertainty associated with the missing values.

Loss to follow-up

Patients who died or were transplanted were suspended on the day of the event, patients who withdrew from the study were censored on the date of the last visit, and the reason for withdrawal was recorded. Missing follow-up data for patients who died before the follow-up could happen were not imputed, and these patients were excluded from relevant analyses.

2.6.2 Descriptive statistics

Statistical analysis and data visualisation were performed in R [424]. Summary statistics are shown as mean ($\pm$SD) or median [IQR] according to data distribution. The number of available data points is reported in tables. Comparisons between the categorical variables were performed using Fisher's exact and Chi-square test, comparisons between continuous non-normally distributed variables were performed with the Mann-Whitney test or the Kruskal-Wallis test. Adjustment for multiple comparisons was performed when appropriate. *P*-values of < 0.05 were considered significant.

2.6.3 Survival analysis

Analysis of survival times and of the factors that influence them was done using "*survival*" [520] and "*survminer*" [257] R packages.

The date of diagnosis was the starting point to determine the follow-up duration. The cut-off date was assessed via the National Health System Digital Spine portal. Patients lost to follow-up were censored as of the time of the last visit. As the study cohort consisted of both prevalent and incident cases, the date of entry to the study did not always coincide with the date of diagnosis. Therefore, data was left-truncated at the date of diagnosis. Patients were excluded if left truncation times (time interval between the date of diagnosis and date of study entry) were longer than 10 years. Kaplan-Meier plots were used to visualise survival curves and a log-rank test was used to compare survival curves for two or more groups. Cox proportional hazards regression was performed to describe the effect of variables on survival. Results are given as hazard ratio estimates with 95% confidence interval Cox proportional hazards models assumptions were tested and met.

2.7 Clinical phenotyping

2.7.1 The analysis of computerised tomography (CT) scans

Diagnostic chest CT scans were performed in 613 study participants and the reports were transcribed to electronic case report forms. Of these, 269 scans were available for repeated analysis; these were anonymised and transferred to the University of Sheffield, Sheffield, UK. Those CT scans were obtained between 2002 and 2018 (CT pulmonary angiogram (CTPA) n = 241, high resolution computed tomography no CTPA, n = 28). Slice thickness was less than 5 mm for all studies, typically $\leq$ 1mm. Images were analysed on the open-

source software Horos (Annapolis, MD USA). Cardiac and vascular measurements were taken by one observer (Marcella Cgliano) and reviewed by the Consultant Radiologist (Dr Andrew Swift). Thoracic radiological features were scored semi-quantitatively by two independent Cardiothoracic Radiologist observers (Dr Andrew Swift, Dr Smitha Rajaram), each with nine years experience in pulmonary hypertension imaging, who were blinded to the underlying diagnosis, mutation and smoking status. For consistency and reproducibility, all measurements were reported on a customised proforma (Table B.8).

Inter-Rater Reliability

To measure the magnitude of agreement between CT scan readers, 22 randomly selected scans were assessed by both radiologists with good interobserver agreement. For categorical variables, the weighted (ordinal data) and unweighted (for non-ordinal data) Cohen's Kappa for two readers were calculated and for continuous variables, the intraclass correlation coefficient (ICC) was computed with the "*irr*" R package [158].

2.7.2 Stratification

Although many collected and assessed variables relevant to the analysis were continuous and treated as such, ultimately the genotype-phenotype analysis required grouping patients into homogenous case clusters. I used three approaches for stratification; patients were stratified by tertiles, cut-offs found through ROC analysis, and those previously reported in the literature.

2.8 Computational phenotyping

To calculate the similarity between pairs of individuals, I first calculated relative information content (IC) of HPO terms based on their rarity in the study sample collection. The IC of term t is defined by the negative natural logarithm of the probability of the term:

$$\mathrm{IC(t)} = -\log p_t$$

where p_t is the number of terms annotated to the total number of terms in the ontology. If the probability decreases then the information content increases and consequently the specificity of the informativeness increases too. Thus, the IC tends to increase as we move away from the root to more specific terms. Semantic similarity is a distance metric that measures the

likeness of the terms meaning or semantic content and relies on their “is a” relationship. Computationally, semantic similarity can be estimated by defining topological similarity, by using ontologies to define the distance between terms/concepts. I used two methods to calculate semantic similarity, namely Resnik [432] and Lin [311]. Both methods are based on the notion of IC, but while Resnik only considers the IC of the most informative common ancestor of two terms (lowest common subsumer (lcs)), Lin considers the IC of lcs as well as the IC of the two compared concepts. I used the "*ontologyX*" R library for this analysis [184].

$$S_{\text{Resnik}}(t_1, t_2) = \max_{t \in \text{anc}(t_1) \cap \text{anc}(t_2)} \text{IC}(t)$$

$$S_{\text{Lin}}(t_1, t_2) = \frac{\max_{t \in \text{anc}(t_1) \cap \text{anc}(t_2)} \text{IC}(t)}{\text{IC}(t_1) + \text{IC}(t_2)}$$

The choice of the similarity measure depends on its application [402]. Suppose sensitivity to sharing rare ancestral terms when evaluating the similarity between the sets of terms is desired, then Resnik’s method should be preferred as it is equal to the IC of the most informative ancestral term and so sharing a rare term significantly contributes to the similarity. Conversely, if rare terms are expected to contribute to the noise rather than carry important information, Lin’s method is more robust as it averages over the similarities of multiple terms.

I used the `get_sim_p()` function from the "*ontologySimilarity*" R package [182] to calculate the statistical significance of within-group ontological similarity. I took a two-fold approach to validate the concept of clustering based on ontology similarity. Firstly, to compare and validate the similarity estimated by both methods (Resnik and Lin) I assessed the phenotypic closeness of groups of index cases with clinical syndromic diagnosis such as HHT. Secondly, to gauge the similarity between patients harbouring pathogenic variants in known PAH risk gene mutations, I repeated the above analysis grouping patients by mutation status.

2.8.1 Clustering algorithms

I then proceeded to cluster analysis using three different machine learning approaches, K-medoids algorithm, more specifically partitioning around medoids (PAM), hierarchical clustering, and affinity propagation clustering. The K-medoids algorithm is a clustering approach for partitioning data into k groups represented by medoids, which are objects within the clusters for which average dissimilarity between them and all other members of the given

group is minimal. Compared to k-means, k-medoids clustering is less sensitive to noise and outliers, as it uses medoids as cluster centres rather than means (used in k-means). The PAM algorithm proceeds in two stages: the "build phase", in which k objects, which become medoids, are selected, a dissimilarity matrix is built and every object is assigned to the closest medoid and the "swap phase", in which the algorithm searches if any of the objects of the cluster decrease the average dissimilarity coefficient. If such an object is found, the algorithm reverses into 'build phase' and every object is again assigned to its closest medoid. For this analysis I used the R packages "*cluster*" [335] , "*fpc*" [209], and "*factoextra*" [258] for visualisation.

In contrast to PAM, hierarchical clustering (HCA) is a method of grouping objects which does not require pre-specifying the number of clusters to be produced. HCA can be subdivided in to two types:

1. Agglomerative clustering (also called agglomerative nesting (AGNES)), in which each observation is initially considered a cluster of its own (leaf). The most similar clusters are successively linked together (by agglomeration or leakage method) until there is just one big cluster (root).

2. Divisive clustering begins with the root, in which all objects are included in one cluster. Then the most heterogeneous clusters are successively divided until all observations are in their own cluster.

While divisive clustering is particularly suited to identifying large clusters, agglomerative clustering is good at identifying small clusters. The HCA result is represented as a multilevel hierarchy called a dendrogram; the height of the fusion, provided on the vertical axis of the dendrogram, indicates clusters' similarity. The higher the height of the fusion (cophenetic distance), the less similar the objects are.

For this analysis I have investigated both divisive and agglomerative clustering and single, complete, average and Ward's linkage methods. I computed cophenetic distances for hierarchical clustering to verify the cluster tree and checked their correlation with original distance. To compare dendrograms obtained by different linkage methods I used the `cor.dendlist()` function from the "*dendextend*" R package [154] and finally to compare the results of agglomerative and divisive clustering I assessed divisive and agglomerative coefficients. I determined the optimal number of clusters based on elbow and silhouette methods and gap statistic [260, 526].

Affinity propagation clustering devised by Frey and Dueck [146] is a method similar to PAM in that it selects exemplars to represent clusters but does not require *a priori*

specification of the number of clusters. For affinity propagation clustering and exemplar-based agglomerative clustering, I used the "*apcluster*" R package [50]. After devising clusters by all three clustering algorithms and assessing statistical significance of their within-group ontological similarity, I plotted them using the "*ontologyPlot*" R library [181] and manually assessed them.

2.9 Genetic data

2.9.1 Whole-genome sequencing (WGS), short read alignment, and variant calling

Samples were received as either DNA extracted from whole blood or as whole blood EDTA samples that were extracted at the central DNA extraction and QC laboratory in Cambridge (UK). Following testing for adequate DNA concentration, DNA degradation, and purity, next-generation paired-end whole-genome sequencing was performed on cases and controls using Illumina HiSeq2500 and HiSeq X (Illumina Inc, San Diego, USA) generating three batches with different read lengths: 100bp (377 samples), 125bp (3,154 samples) and 150bp (9,656 samples). Reads were aligned against the Genome Reference Consortium human genome build 37[2] using the Illumina Isaac Aligner version SAAC00776.15.01.27 [538] and variants were called using the Illumina Starling software version 2.1.4.2[3]. The variants were then left-aligned, normalised with bcftools and loaded into our Hbase database to produce multi-sample variant calls to undertake the genetic association studies [538].

2.9.2 The PAH domain samples for the genetic analysis

Table 2.1 summarises the number of unrelated samples (estimated using the R/Bioconductor package "*Genesis*" [174] and referred to as “maximal unrelated set”) by study domain [538]. The PAH domain comprised 1,148 (1,123 unrelated) individuals. Of these, 1,122 cases (following exclusion of three cases with unknown diagnosis and 23 unaffected relatives) were available for the analysis. Based on the estimated family networks, only one affected member of each family with the given tag (selected from either the maximal unrelated or the related set) is considered for the genotype-phenotype association testing with BeviMed,

[2]GRCh37, https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.13/)

[3]https://support.illumina.com/help/BS_App_TS_Amplicon_OLH_15055858/Content/Source/Informatics/Apps/IsaacVariantCaller_appENR.htm.

which results in a variable number of cases per tag depending on tag availability and family structure.

Table 2.1 NIHR BioResource - Rare Diseases domain definitions and number of maximal unrelated individuals per project.

Project acronym	Project	Number of individuals
GEL	Genomics England Ltd	3058
BPD	Bleeding, Thrombotic and Platelet Disorders	986
PID	Primary Immune Disorders	1027
CNTRL	Processed Controls	50
IRD	Inherited Retinal Disorders	717
NDD	Neurological and Developmental Disorders	518
EDS	Ehlers Danlos Syndrome	15
HCM	Hypertrophic Cardiomyopathy	239
PMG	Primary Membranoproliferative Glomerulonephritis	181
SRNS	Steroid Resistant Nephrotic Syndrome	234
CSVD	Cerebral Small Vessel Disease	134
NPD	Neuropathic Pain Disorder	185
ICP	Intrahepatic Cholestasis of Pregnancy	267
LHON	Leber Hereditary Optic Neuropathy	54
MPMT	Multiple Primary Tumours	554
SMD	Stem Cell & Myeloid Disorders	153
PAH	Pulmonary arterial hypertension	1123
UKBio	UK BioBank	764
	Total	10259

2.9.3 "Labelling" of explained cases

For this analysis I made the assumption (widely supported by the literature) that PAH is inherited in Mendelian fashion, meaning that the disease is driven by a single gene. As a consequence of this assumption patients with rare deleterious variants in previously established PAH disease genes (*BMPR2*, *ACVRL1*, *ENG*, *CAV1*, *SMAD1*, *SMAD9*, *KCNK3*, *EIF2AK4*, *TBX4*, *AQP1*, *ATP13A3*, *GDF2*, *SOX17*) that were deemed disease-causing by a genetic multidisciplinary team according to the ACMG Standards and Guidelines [440], were excluded from the association testing for other genes. For example, a patient with a rare deleterious variant in *BMPR2*, which was assessed as pathogenic explaining the phenotype, was not included in the association testing for other genes. Of note, I also reported the results of association testing between rare variants in known PAH risk genes and phenotypes.

2.9.4 Selection of gene biotypes and rare variants for association testing

The following biotypes of genes defined by Ensembl[4] were used for the analysis:

1. lincRNA (long intergenic ncRNA): transcripts that are long intergenic non-coding RNA locus with a length >200bp. Requires lack of coding potential and may not be conserved between species
2. miRNA: a small RNA (22bp) that silences the expression of target mRNA
3. miscRNA: miscellaneous RNA, a non-coding RNA that cannot be classified
4. Protein coding: gene/transcript that contains an open reading frame (ORF)
5. rRNA: the RNA component of a ribosome
6. snoRNA: small RNA molecules that are found in the cell nucleolus and are involved in the post-transcriptional modification of other RNAs
7. snRNA: small RNA molecules that are found in the cell nucleus and are involved in the processing of pre-messenger RNAs
8. TR gene: T cell receptor gene that undergoes somatic recombination[5]
   (a) TR C gene: Constant chain T cell receptor gene that undergoes somatic recombination before transcription
   (b) TR D gene: Diversity chain T cell receptor gene that undergoes somatic recombination before transcription
   (c) TR J gene: Joining chain T cell receptor gene that undergoes somatic recombination before transcription
   (d) TR V gene: Variable chain T cell receptor gene that undergoes somatic recombination before transcription
9. IG gene: Immunoglobulin gene that undergoes somatic recombination[6]
   (a) IG C gene: Constant chain immunoglobulin gene that undergoes somatic recombination before transcription

[4] https://www.ensembl.org/info/genome/genebuild/biotypes.html
[5] annotated in collaboration with IMGT, http://www.imgt.org/
[6] annotated in collaboration with IMGT, http://www.imgt.org/

(b) IG D gene: Diversity chain immunoglobulin gene that undergoes somatic recombination before transcription

(c) IG J gene: Joining chain immunoglobulin gene that undergoes somatic recombination before transcription

(d) IG V gene: Variable chain immunoglobulin gene that undergoes somatic recombination before transcription

10. Mt_rRNA: mitochondrial ribosomal RNAs

11. Mt_tRNA: mitochondrial t-RNAs

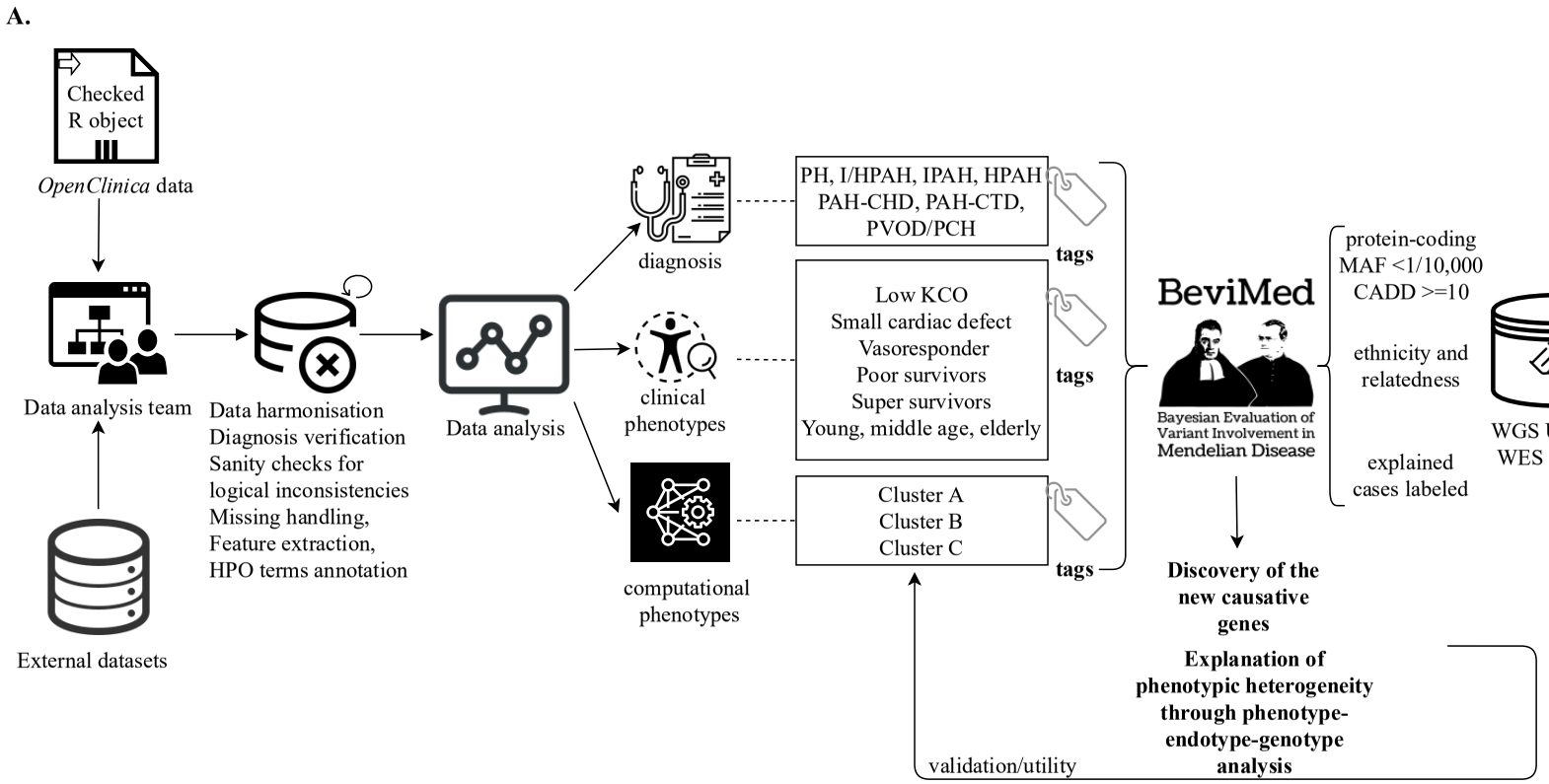
Rare variants were extracted from each gene as previously described [538] including a PMAFx (for a given variant, the probability that the minor allele count is at least the observed minor allele count, given that MAF = 1/X) < 0.05 with x = 1,000 for the recessive and x = 10,000 for the dominant association model, and a CADD Phred score $\geq$10.

To increase power in scenarios where only variants of particular consequence types were associated with the disease risk, association models were fitted to different subsets of variants according to the consequences provided by Ensembl[7]: the High category, comprised only variants of "high" impact, including likely loss-of-function variants and large deletions; the Moderate category contained variants of impact "moderate", including missense variants or consequence "non_coding_transcript_exon_variant"; the combined category of Moderate and High, combined the respective consequence types.

2.9.5 Genetic association between rare variants and selected diagnostic and phenotypic tags

A schematic of the analysis pipeline is depicted in Fig. 2.7 A. I hypothesised that phenotypically homogenous groups of patients would also share a similar genetic aetiology. I defined cases as individuals carrying a particular tag whereas the individuals from the non-PAH domains served as controls (Fig. 2.7 B, Table 2.1). Rare variants were extracted as described above. To study genotype-phenotype associations I used the Bayesian model comparison method BeviMed [518, 538]. This method is more powerful and informative than existing rare variant association tests under both dominant and recessive modes of inheritance. Furthermore, it has previously been shown to be highly effective at identifying known and novel

[7] https://www.ensembl.org/info/genome/variation/prediction/predicted_data.html



(A) Overview of the analytical approach. Using deep phenotyping, data tags were assigned to patients who shared phenotypic features. Rare sequence variants, called from whole-genome sequencing data, were filtered and explained cases were labelled. BeviMed was applied to a set of unrelated individuals to estimate the posterior probability of gene-tag associations. (B) Schematic representation of the definition of cases, exemplified by the KCO lower tertile tag. Cases were defined as individuals carrying a particular tag, whereas patients with missing information or those without a tag were removed from the gene-tag association testing. Individuals from non-PAH domains served as controls. KCO: transfer coefficient for carbon monoxide, MAF: minor allele frequency.

Fig. 2.7 Design of the genetic association study.

loci involved in rare diseases, including the corresponding modes of inheritance and consequence types responsible [518, 538]. BeviMed is a method for testing association between rare Mendelian disease and a genomic locus by comparing support for a model where disease risk is associated with genotypes at rare variant sites in the locus and a genotype-independent "null" model. The prior probability of the association model π can be set to incorporate external biological information, conservation scores, gene ontological annotation, expression in the tissue of interest, etc., or be set to the same value for all genes. In the current analysis, I have chosen a prior of 0.001 to reflect our assumption that there are around 30 of 32,606 genes that are associated with PAH. I have also calculated the posterior probability for a prior of 0.01 (estimating that 300 genes might be involved in the pathogenesis of PH) and for a prior of 0.00033 (assuming that only ten genes are implicated in the disease) to test the impact of the prior on posterior probability. Posterior probability (PP) is a metric that integrates the evidence from observed data (quantified by the Bayes factor (BF)) with the subjectively chosen prior probability that a gene is truly associated with the phenotype. The BF is the ratio of the likelihood of H1 to the likelihood of H0 and it reflects the strength of the evidence in favour of one hypothesis over the other. Because π in the genomic analysis is typically small (ditto), a large BF is required to provide convincing evidence of genotype-phenotype association (PP close to 1). The need for a large BF, therefore, reflects the small number of expected true genotype-phenotype associations and is comparable to FDR adjustment in frequentist approaches with the difference that frequentists adjust for the number of comparisons. Rather than relying on significance thresholds I ranked the findings according to the logarithm of the BF, thus independently of the prior probability of association. I truncated the list around the level of the lowest previously reported PAH risk gene and provided the posterior probabilities for when the prior probability of association was set at 0.001, which I believe is best justified in this case.

The present analysis was restricted to canonical transcripts as annotated by Ensembl. Importantly, since I selected rare variants that do not correlate strongly with coarse population structure, I did not correct for ethnicity. The association testing was only performed on unrelated individuals (Table B.9, Table 2.1). To increase power in scenarios where only variants of particular consequence types were associated with the disease risk, association models were fitted to different subsets of variants according to the impact categories defined by Ensembl (high, moderate and the combination of both; see above).

2.9.6 Conclusions

In Chapter 2, I described the steps I took to collect, harmonise, manage missingness and verify phenotypic information gathered on patients with PAH recruited to the NBR and PAH Cohort studies and the PAH USBB study. In the NBR and PAH Cohort studies the results of routinely performed tests were collected from EHR or case notes, while additional deep phenotype data such as BMPR2 ligand measurements and CT scan assessments were obtained through sub-studies. In the US PAHBB study, all clinical information was transcribed from registry proformas [Chapters 2.1, 2.2, 2.3, 2.4]. To standardise phenotype description, I annotated all clinical information with HPO terms and where the terms were missing I enriched the HPO ontology for relevant labels and their definitions [Chapter 2.5.7]. I subsequently described the methods I used to cluster the patients into homogenous groups [Chapter 2.8], which were then used for genetic case-control analysis [Chapter 2.9.5].

In Chapter 2 I addressed aims 1 and 2 of my Thesis.

Chapter 3

Results

In this chapter I describe the main findings of my PhD project. I start by presenting the descriptive characteristics of the NBR and PAH Cohort studies' population and compare it to the PAH US Biobank. I then provide the results of clinical and computational phenotyping and the rational for using selected phenotypes in genotype-phenotype analysis. After that I describe the most pertinent genetic findings (*KDR*, *COL6A5*) and the outcomes of reverse phenotyping of *KDR* and *GDF2* mutation carriers.

3.1 Clinical characterisation of study participants

3.1.1 The NBR and I/HPAH Cohort studies

In total 1,394 cases, 90 relatives and 118 healthy controls entered both studies (status from 15^{th} July 2020), Fig. 3.1. Breakdown of recruitment of cases per centre is shown in Table B.10. The data set was analysed for missing rate and pattern. Data was found to have a multivariate (many variables had missingness) missing pattern. Although the *a priori* threshold of the allowed missing rate for core variables was set at 10%, only 10 out of 36 variables stayed below this threshold. The missing rate was higher for prevalent cases when compared with the incident (due to changes in diagnostic protocols over time and difficulty accessing old paper patients records), in centres that supplied the highest proportion of legacy samples (Imperial and Hammersmith Hospital and Royal Papworth Hospital), shared care centres (Royal Brompton Hospital), and in cases recruited outside of the UK, which was related to different data collection protocols. Further differences in missingness between centres were related to the number of patients recruited (small centres had better data completeness) and local protocols (i.e. PH centre in Sheffield routinely performs incremental shuttle walk

test rather than 6-minute walk test (6MWT) and cardiac MRI rather than ECHO). Through diagnosis verification exercise, 77 (5.5%) patients were found to have a diagnosis other than I/HPAH or PVOD/PCH and were subsequently suspended (Fig. 3.1 and Fig. 3.2).

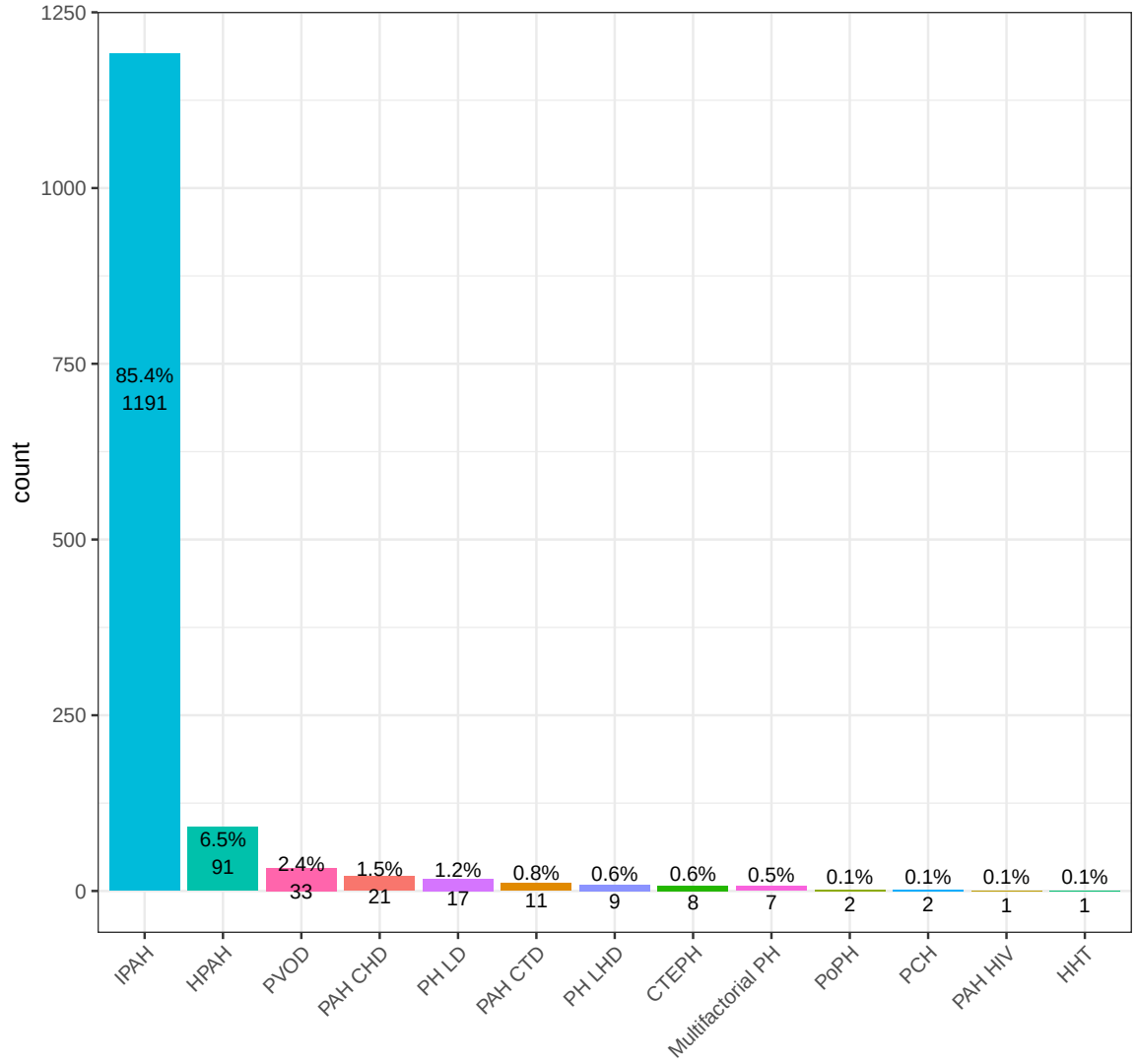


Fig. 3.2 Breakdown of diagnoses in the study cohort.

Baseline characteristics of patients included in the study are summarised in Table 3.1. In brief, the study population was similar to previously reported registry data. Diagnosis of IPAH was made in 1,191 (85.4%), and this included 69 patients with a significant history of drugs and toxin use, 91 (6.5%) had a clinical diagnosis of HPAH. The median age at diagnosis was 50 [35;65] years. This cohort also included paediatric cases, 36 individuals

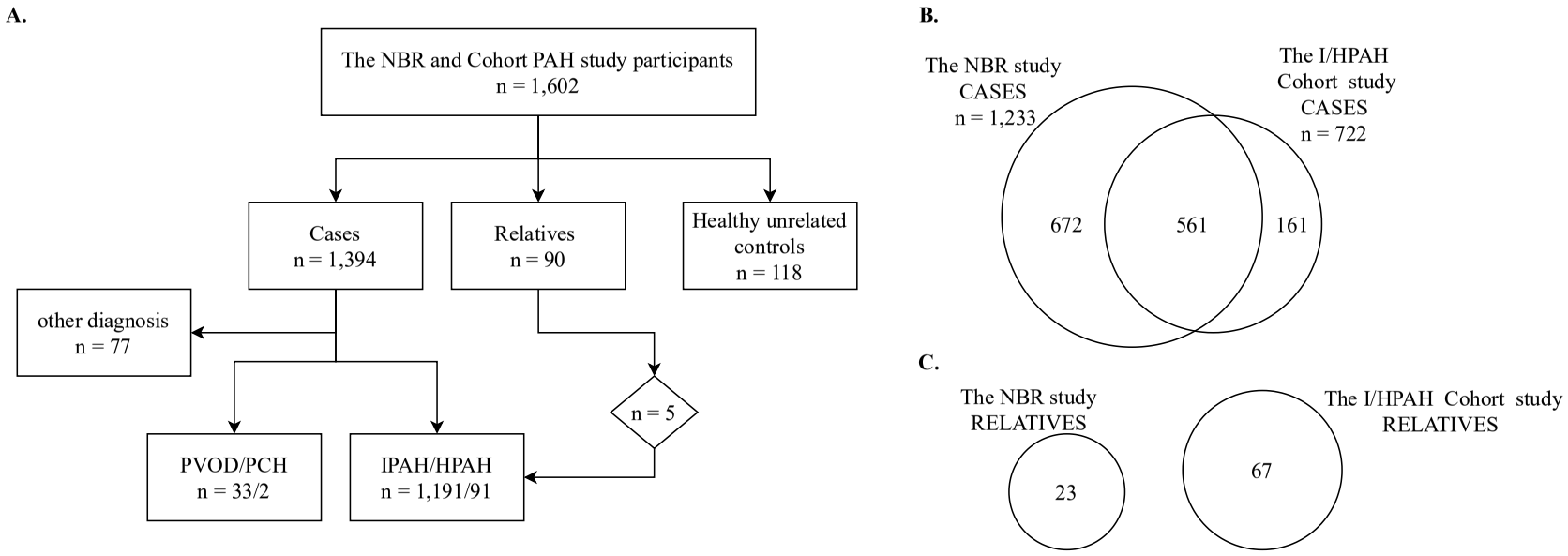


Fig. 3.1 Consort diagram of the NBR and Cohort studies.
(A) The NBR and PAH Cohort studies consort diagram, (B) Overlap of cases participating in both studies, (C) Relatives participating in both studies.

were diagnosed before the age of 16 (a threshold for adulthood in the study); 15 of these were recruited at Great Ormond Street Hospital, whereas the rest were prevalent cases recruited at adult centres. When patients were stratified by age, significant differences were seen in most clinical characteristics and survival (Table 3.2, Fig. 3.3 A). There was a typical female predominance, which is characteristic of the I/HPAH population. The majority of the cohort (75.9%) was of European descent. Most (65%) patients were diagnosed in functional class III. A strong family history of PAH was reported in 163 patients ranging from 1 to 7 family members affected. Additionally, 74 index cases had a suspicion of FPAH, 18 relatives of those patients died possibly due to PAH. In total 132 lost pregnancies were reported of which 13 (9 abortions and 4 miscarriages) happened after the diagnosis. Initial treatment in most cases was an upfront combination therapy followed by monotherapy with Sildenafil. In accordance with recent registries, the cohort was characterised by the high prevalence of comorbidities which were reported in 79.2% of cases. The five most common were: diseases of the cardiovascular system, endocrine disorders, disorders of the respiratory system, gastrointestinal disease and diseases of the blood and blood-forming organs. To assess the potential impact of comorbidities on survival, Charlson and Elixhouser scores were calculated (Table B.6). This analysis revealed that 31% of patients scored ≥ 1 in either risk score indicative of bearing comorbidity that affects survival. Figure 3.3 B shows the impact of Charlson scores on survival. The most common congenital abnormality reported was PFO/ASD (n = 96), followed by congenital abnormalities of great arteries (n = 7). Interestingly a small proportion of patients had been diagnosed with syndromes of known genetic background (Noonan, Edward, Turner syndrome).

Over the median follow up of 6.9 years, 329 died, and 56 underwent lung transplantation. Overall survival in the cohort was 97.5, 90.6, 82.3% at 1, 3 and 5 years, and transplant-free survival was 96.9, 89.2, 80.3%, respectively. Both NBR and the I/HPAH Cohort studies allowed (similarly to many registries and observational studies) recruitment of both prevalent and incident cases. This approach, although allowing for the capture of an accurate snapshot of clinical heterogeneity related to PAH, is also a source of bias due to the exclusion of prevalent cases that did not survive until the study started and therefore leads to a study population biased towards favourable survival. This situation is depicted in Fig. 3.3 C and D, showing a significantly better outcome in prevalent cases (overall 5-year survival 70.6% and 86.2% and transplant-free survival 68% and 84% in the incident and prevalent cases respectively, p-value < 0.0001).

Table 3.1 Study cohort characteristics (The NBR and PAH Cohort studies)

	[ALL] N=1394	**N**
Diagnosis:		1394
IPAH/HPAH	1191 (85.4)/91 (6.5)	
PVOD/PCH	35 (2.5)	
APAH	35 (2.5)	
PH-LHD	9 (0.6)	
PH-LD	17 (1.2)	
CTEPH	8 (0.6)	
Multifactorial PH	7 (0.5)	
HHT (no PH)	1 (0.1)	
Age [years]	50 [35;65]	1379
Sex: female, n (%)	944 (68)	1386
Ethnicity: European, n (%)	1058 (75.9)	1394
Incident case, n (%)	450 (32.7)	1375
Drug exposure, n (%)	69 (4.95)	1394
BMI [kg/m^2]	27.0 [23.1;31.5]	1258
WHO FC: I/II/III/IV, %	2/20/65/13	1339
6MWD [m]	335 [220;419]	895
Shuttle walk distance [m]	320 [190;410]	180
SpO_2 pre [%]	96.0 [93.0;97.0]	1089
SpO_2 post [%]	91.0 [85.0;95.0]	1028
mRAP [mmHg]	8 [5;12]	1234
mPAP [mmHg]	52 [44;61]	1308
mPAWP [mmHg]	9 [7;12]	1180
CO [L/min]	3.9 [3.1;4.9]	1252
SvO_2 [%]	64.0 [58.0;70.0]	1035
FEV1 [% pred.]	86.0 [73.0;97.0]	1063
FVC [% pred.]	94.3 [81.0;106]	1045
TLC [% pred.]	94.0 [84.7;104]	809
KCO [% pred.]	71 [50;86]	806
Emphysema: n (%)		752
minimal/mild	48 (6)	
moderate	22 (3)	
severe	7 (1)	
Fibrosis: n (%)		755
minimal/mild	35 (5)	
moderate	3 (0)	
severe	1 (0)	
Thromboembolic disease, n (%)	13 (2)	769
Smoking: past/current, n (%)	529 (51.0)	1037
NTproBNP [ng/l]	1116 [280;2742]	380
BNP [ng/l]	192 [69.0;427]	357
CRP [mg/l]	4.10 [2.00;8.15]	818
Creatinine [μmol/l]	85.0 [70.0;102]	1061
Initial therapy: n (%)		805
CCB	90 (11.2)	
combination therapy	390 (48.4)	
ERA	88 (10.9)	

Table 3.1 Study cohort characteristics (The NBR and PAH Cohort studies)

	[ALL] N=1394	**N**
PA	45 (5.59)	
PDE5	191 (23.7)	
sGC	1 (0.12)	

Abbreviations: I/HPAH: idiopathic/hereditary pulmonary arterial hypertension; PVOD/PCH: pulmonary vaso-occlusive disease/pulmonary capillary hemangiomatosis; APAH: associated pulmonary arterial hypertension; PH-LHD: pulmonary hypertension due to left heart disease; PH-LD: pulmonary hypertension due to lung disease and/or hypoxia; CTEPH: chronic thromboembolic pulmonary hypertension; HHT: hereditary hemorrhagic telangiectasia; BMI: body mass index; WHO FC: World Health Organization Functional Class; 6MWD: 6-minute walk distance; SpO_2: peripheral capillary oxygen saturation; mRAP: mean right atrial pressure; mPAP: mean pulmonary artery pressure; PAWP: Pulmonary Artery Wedge Pressure; CO: cardiac output, SvO_2: Mixed venous oxygen saturation; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; TLC: total lung capacity; KCO: transfer coefficient for carbon monoxide; NTproBNP: N-terminal pro B-Type Natriuretic Peptide; BNP: B-Type Natriuretic Peptide; CRP: C-Reactive Protein Protein; CCB: calcium channel blockers; ERA: endothelin receptor antagonists; PA: prostacyclin analogues; PDE5: phosphodiesterase type 5; sGC: Soluble Guanylate Cyclase Stimulator

Table 3.2 Clinical characteristics by age strata (The NBR and Cohort studies).

	Lower age tertile N=460	**Middle age tertile N=460**	**Higher age tertile N=459**	***p*-value**	**N**
Diagnosis: n (%)				<0.001	1379
IPAH	387 (84.1)	398 (86.5)	400 (87.1)		
HPAH	49 (10.7)	30 (6.52)	12 (2.61)		
PVOD/PCH	9 (1.96)	7 (1.52)	19 (4.14)		
APAH	12 (2.61)	16 (3.48)	5 (1.09)		
PH LHD	0 (0.00)	0 (0.00)	7 (1.53)		
PH LD	1 (0.22)	4 (0.87)	11 (2.40)		
CTEPH	1 (0.22)	3 (0.65)	2 (0.44)		
Multifactorial PH	1 (0.22)	2 (0.43)	3 (0.65)		
Age [years]	30 [24;35]	50 [45;54]	69 [65;74]	**<0.001**	1379
Sex: female, n (%)	344 (75)	329 (72)	265 (58)	**<0.001**	1379
Incident cases, n (%)	123 (26.9)	122 (26.6)	205 (44.7)	**<0.001**	1375
WHO FC: I/II/III/IV, %	4/26/55/15	1/22/67/10	1/12/74/14	**<0.001**	1337
6MWD [m]	385 [315;460]	355 [266;433]	235 [124;335]	**<0.001**	895
Shuttle walk distance [m]	360 [260;455]	290 [185;383]	210 [60.0;360]	**<0.001**	180
SpO_2 pre [%]	97[95;98]	96[93;97]	94 [90;96]	**<0.001**	1089
SpO_2 post [%]	94 [88;97]	92 [86;96]	88 [82;93]	**<0.001**	1028
mRAP [mmHg]	7 [5;11]	9 [5;12]	8 [5;12]	0.027	1233
mPAP [mmHg]	55 [47;66]	55 [46;61]	47 [40;55]	**<0.001**	1307
mPAWP [mmHg]	9 [6;11]	9 [7;12]	10 [7;12]	**0.001**	1179
CO [L/min]	4.0 [3.1;5.0]	3.8 [3.1;4.9]	3.8 [3.2;4.7]	0.659	1252
SvO_2 [%]	67[60;72]	64[58;70]	63 [57;68]	**<0.001**	1035
FEV1 [% pred.]	87[77;97]	85[71;97]	85 [70;98]	0.305	1063
FVC [% pred.]	91 [81;102]	97[82;109]	96 [81;108]	0.005	1045
TLC [% pred.]	95[85;103]	95 [86;106]	93 [82;102]	0.009	809
KCO [% pred.]	77 [66;90]	76 [58;89]	56 [39;76]	**<0.001**	806
Emphysema: n (%)				<0.001	751
minimal/mild	1 (0)	14 (6)	33 (12)		
moderate	0 (0)	6 (2)	16 (6)		
severe	0 (0)	0 (0)	7 (3)		
Fibrosis: n (%)				**<0.001**	754
minimal/mild	3 (1)	3 (1)	29 (11)		
moderate	0 (0)	1 (0)	2 (1)		
severe	0 (0)	1 (0)	0 (0)		
Smoking: past/current, n (%)	122 (35.9)	183 (54.0)	224 (63.1)	**<0.001**	1034
NTproBNP [ng/l]	566 [154;2129]	866 [205;1573]	2050 [619;3848]	**<0.001**	380
BNP [ng/l]	113 [30.0;384]	181 [87.8;401]	255 [103;504]	<0.001	356
CRP [mg/l]	4.00 [2.00;7.00]	4.00 [2.00;8.00]	5.00 [2.60;9.20]	0.04	817
Creatinine [μmol/l]	77.0 [65.0;91.0]	82.0 [69.0;96.0]	96.0 [80.0;120]	**<0.001**	1060
DM type 2, n (%)	7 (1.52)	45 (9.78)	117 (25.5)	**<0.001**	1379
COPD, n (%)	2 (0.43)	25 (5.43)	54 (11.8)	**<0.001**	1379
OSA, n (%)	5 (1.09)	24 (5.22)	40 (8.71)	**<0.001**	1379
CAD, n (%)	0 (0.00)	12 (2.61)	41 (8.93)	**<0.001**	1379
HTN, n (%)	24 (5.22)	97 (21.1)	197 (42.9)	**<0.001**	1379

Legend: lower age tertile range [< 1 to 40], middle age tertile range [> 40 to 58], higher age tertile [> 58 to 88]

Abbreviations: I/HPAH: idiopathic/hereditary pulmonary arterial hypertension; PVOD/PCH: pulmonary vaso-occlusive disease/pulmonary capillary hemangiomatosis; APAH: associated pulmonary arterial hypertension; PH-LHD: pulmonary hypertension due to left heart disease; PH-LD: pulmonary hypertension due to lung disease and/or hypoxia; CTEPH: chronic thromboembolic pulmonary hyper-

tension; WHO FC: World Health Organization Functional Class; 6MWD: 6-minute walk distance; SpO_2: peripheral capillary oxygen saturation; mRAP: mean right atrial pressure; mPAP: mean pulmonary artery pressure; PAWP: Pulmonary Artery Wedge Pressure; CO: cardiac output; SvO_2: Mixed venous oxygen saturation; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; TLC: total lung capacity; KCO: transfer coefficient of carbon monoxide; NTproBNP: N-terminal pro B-Type Natriuretic Peptide; BNP: B-Type Natriuretic Peptide; CRP: C-Reactive Protein Protein; DM: diabetes mellitus; COPD: chronic obstructive pulmonary disease; OSA: obstructive sleep apnea; CAD: coronary artery disease; HTN: systemic hypertension

Table 3.3 Clinical characteristics of poor- and super-survivors (The NBR and Cohort studies).

	Non-poor-survivors N=1079	**Poor-survivors** N=296	***p*-value**	**Non-super-survivors** N=941	**Super-survivors** N=434	***p*-value**	**N**
Diagnosis: n (%)			**<0.001**			**<0.001**	1375
IPAH	953 (88.3)	229 (77.4)		797 (84.7)	385 (88.7)		
HPAH	76 (7.04)	16 (5.41)		56 (5.95)	36 (8.29)		
PVOD/PCH	13 (1.20)	22 (7.43)		35 (3.72)	0 (0.00)		
APAH	20 (1.85)	13 (4.39)		21 (2.23)	12 (2.76)		
PH LHD	4 (0.37)	3 (1.01)		6 (0.64)	1 (0.23)		
PH LD	5 (0.46)	9 (3.04)		14 (1.49)	0 (0.00)		
CTEPH	5 (0.46)	1 (0.34)		6 (0.64)	0 (0.00)		
Multifactorial PH	3 (0.28)	3 (1.01)		6 (0.64)	0 (0.00)		
Age [years]	48 [34;62]	57 [42;71]	**<0.001**	54 [39;68]	42 [32;54]	**<0.001**	1375
Sex: female, n (%)	764 (71)	172 (58)	**<0.001**	612 (65)	324 (75)	**<0.001**	1375
Incident case, n (%)	249 (23.1)	201 (67.9)	**<0.001**	450 (47.8)	0 (0.00)	**<0.001**	1375
WHO FC:I/II/III/IV, %	2/22/63/12	1/10/74/16	**<0.001**	2/15/70/14	2/31/55/11	**<0.001**	1333
6MWD [m]	343 [240;423]	300 [137;378]	**<0.001**	308 [180;400]	377 [302;444]	**<0.001**	893
Shuttle walk distance [m]	341 [240;420]	180 [40.0;275]	**<0.001**	270 [150;382]	365 [268;455]	**<0.001**	180
SpO_2 pre [%]	96.0 [93.0;97.0]	95.0 [92.0;97.0]	**0.002**	95.0 [92.0;97.0]	96.0 [94.0;98.0]	**<0.001**	1087
SpO_2 post [%]	91.0 [85.0;96.0]	89.0 [83.0;95.0]	**0.004**	90.0 [84.0;95.0]	92.0 [87.0;96.0]	**0.01**	1026
mRAP [mmHg]	8 [5;12]	8 [6;13]	0.084	8 [6;12]	7 [5;11]	**0.002**	1229
mPAP [mmHg]	53 [44;61]	50 [44;59]	0.063	52 [43;60]	54 [46;62]	**0.005**	1303
mPAWP [mmHg]	9 [7;12]	9 [7;12]	0.772	9 [7;12]	9 [6;11]	**0.034**	1176
CO [L/min]	4.0 [3.2;5.0]	3.6 [2.9;4.6]	**<0.001**	3.8 [3.1;4.8]	4.0 [3.2;5.0]	0.051	1249
SvO_2 [%]	65.0 [59.0;70.4]	62.1 [55.3;67.6]	**<0.001**	63.2 [57.5;69.0]	68.0 [61.0;72.0]	**<0.001**	1031
FEV1 [% pred.]	86.0 [74.0;97.0]	85.4 [71.0;98.0]	0.689	86.0 [72.3;97.2]	86.0 [74.0;97.0]	0.834	1062
FVC [% pred.]	95.0 [82.0;106]	94.0 [78.8;106]	0.425	95.0 [81.0;107]	93.0 [82.0;106]	0.526	1044
TLC [% pred.]	95.4 [86.0;104]	90.0 [80.0;102]	**0.001**	93.0 [83.2;104]	97.5 [87.0;104]	**0.022**	809
KCO [% pred.]	73 [55;87]	60 [40;78]	**<0.001**	67 [45;82]	80 [68;94]	**<0.001**	806
Emphysema:n (%)			**0.014**			**<0.001**	750
minimal/mild	29 (5)	19 (10)		47 (8)	1 (1)		
moderate	13 (2)	9 (5)		20 (4)	2 (1)		
severe	5 (1)	2 (1)		7 (1)	0 (0)		
Fibrosis: n (%)			**0.003**			0.072	753
minimal/mild	20 (4)	15 (8)		32 (6)	3 (2)		
moderate	1 (0)	2 (1)		3 (1)	0 (0)		
severe	0 (0)	1 (1)		1 (0)	0 (0)		
Smoking: past/current, n (%)	400 (49.4)	129 (58.1)	**0.027**	399 (54.0)	130 (44.5)	**0.008**	1031
NTproBNP [ng/l]	909 [224;2484]	1573 [584;3264]	**0.003**	1220 [325;2888]	667 [150;2036]	**0.012**	379
BNP [ng/l]	174 [60.2;402]	264 [95.8;546]	**0.022**	214 [82.0;462]	113 [27.5;282]	**0.001**	356
CRP [mg/l]	4.00 [2.00;8.00]	5.00 [2.00;9.35]	0.204	4.10 [2.00;9.00]	4.45 [2.00;7.00]	0.692	817
Creatinine [μmol/l]	84.0 [69.0;99.2]	90.0 [73.8;111]	**<0.001**	86.0 [70.0;103]	83.0 [68.5;98.0]	**0.003**	1060
DM type 2, n (%)	118 (10.9)	51 (17.2)	**0.005**	145 (15.4)	24 (5.53)	**<0.001**	1375
COPD, n (%)	61 (5.65)	20 (6.76)	0.565	62 (6.59)	19 (4.38)	0.135	1375
OSA, n (%)	50 (4.63)	19 (6.42)	0.273	59 (6.27)	10 (2.30)	**0.003**	1375
CAD, n (%)	36 (3.34)	17 (5.74)	0.083	49 (5.21)	4 (0.92)	**<0.001**	1375
HTN, n (%)	243 (22.5)	75 (25.3)	0.347	243 (25.8)	75 (17.3)	**0.001**	1375

Abbreviations: I/HPAH: idiopathic/hereditary pulmonary arterial hypertension; PVOD/PCH: pulmonary vaso-occlusive disease/pulmonary capillary hemangiomatosis; APAH: associated pulmonary arterial hypertension; PH-LHD: pulmonary hypertension due to left heart disease; PH-LD: pulmonary hypertension due to lung disease and/or hypoxia; CTEPH: chronic thromboembolic pulmonary hypertension; WHO FC: World Health Organization Functional Class; 6MWD: 6-minute walk distance; SpO_2: peripheral capillary oxygen saturation; mRAP: mean right atrial pressure; mPAP: mean pulmonary artery pressure; PAWP: Pulmonary Artery Wedge Pressure; CO: cardiac output; SvO_2: Mixed venous oxygen saturation; FEV1: forced expiratory capacity in 1 second; FVC: forced vital capacity; TLC: Total Lung Capacity; KCO: transfer coefficient for carbon monoxide; NTproBNP: N-terminal pro B-Type Natriuretic Peptide; BNP: B-Type Natriuretic Peptide; CRP: C-Reactive Protein Protein; DM: diabetes mellitus; COPD: chronic obstructive pulmonary disease; OSA: obstructive sleep apnea; CAD: coronary artery disease; HTN: systemic hypertension

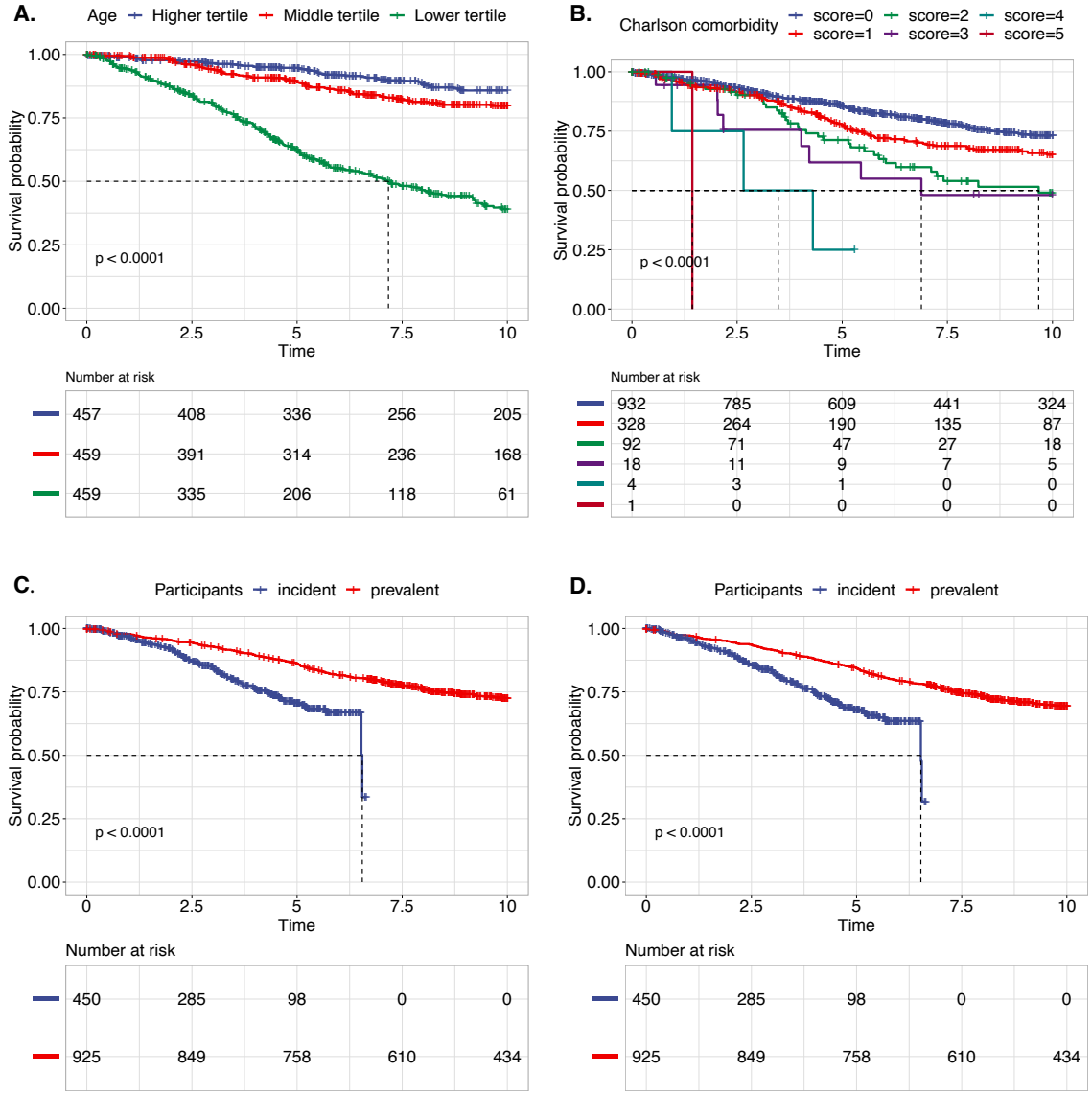


Fig. 3.3 Kaplan-Meier curves by age tertiles (A), Charlson comorbidity score (B), incident or prevalent label (C) overall survival, (D) transplant free survival

Conversely, the inclusion of prevalent cases with favourable survival allowed to subset a group of super-survivors, whose identification would require long observation time in case only incident cases were allowed. Inclusion of incident cases on the other hand allowed the identification of patients with particularly aggressive disease and short survival. The cut-offs for poor-survivors and super-survivors were chosen based on the domain knowledge at 3 and 10 years respectively. Clinical characteristics of poor- and super-survivors are presented in Table 3.3. Commonly used risk scores, such as abbreviated ESC/ERS risk score and Reveal risk score (Table 3.4), accurately stratified patients but the risk profiles were different (more favourable) than those reported in the US and other European cohorts [43, 219] (Fig. 3.4).

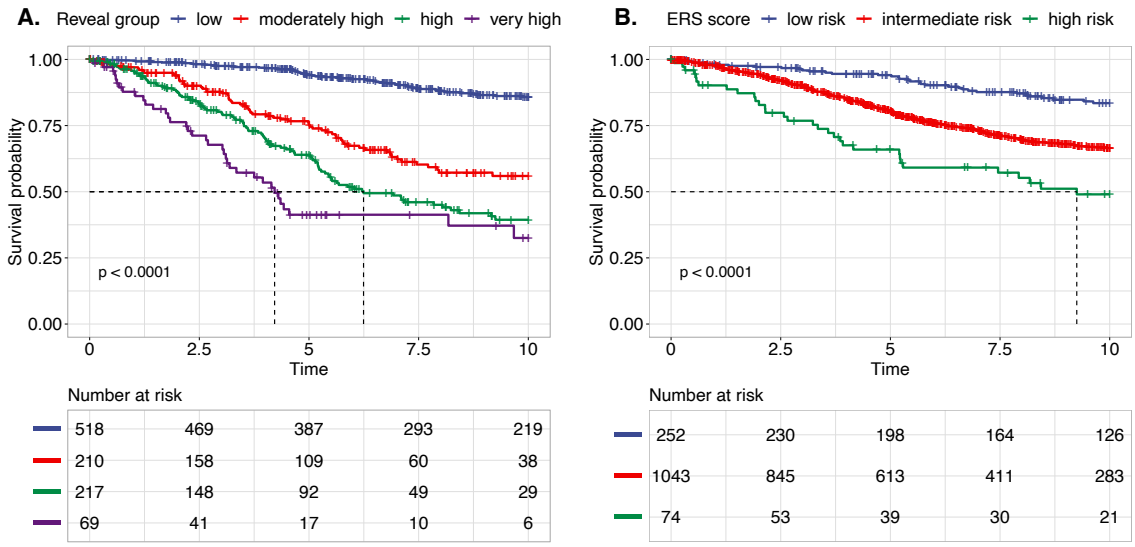


Fig. 3.4 Kaplan-Meier curves by Reveal risk group (A) and ERS risk group (B).

Out of 90 relatives, five individuals developed PAH. Time from the inclusion into the study to disease development was 13, 1, 5, 26 and 2 months respectively. One relative who developed the disease died during the follow-up.

3.1.2 The PAH USBB study

In this chapter, I summarise the results of the harmonisation effort between the NBR study and the US PAHBB. The harmonisation process was performed as described in the methods section (Chapter 2.5.8).

Results of data harmonisation

Two studies applied different inclusion criteria and definitions of diagnostic groups. The NBR/Cohort studies aimed at recruiting patients with I/HPAH and PVOD/PCH cases whereas the US PAHBB recruited individuals from various subgroups of group 1: I/H-PAH, PVOD/PCH, drug and toxin associated PAH, PAH associated with other diseases (congenital heart disease, connective tissue disease, portal hypertension) and persistent pulmonary hypertension of newborn. After the diagnosis verification procedure, 77 patients from the NBR/Cohort studies had their initial diagnosis changed. In the US PAHBB cohort, 11 patients had a discrepancy between disease type and subtype and were subsequently reclassified; 21 patients labelled as APAH did not have enough information to define disease subtype, an additional assessment of comorbidities undermined the diagnosis of APAH altogether, and for two patients, no clinical data were available (Fig. 3.5).

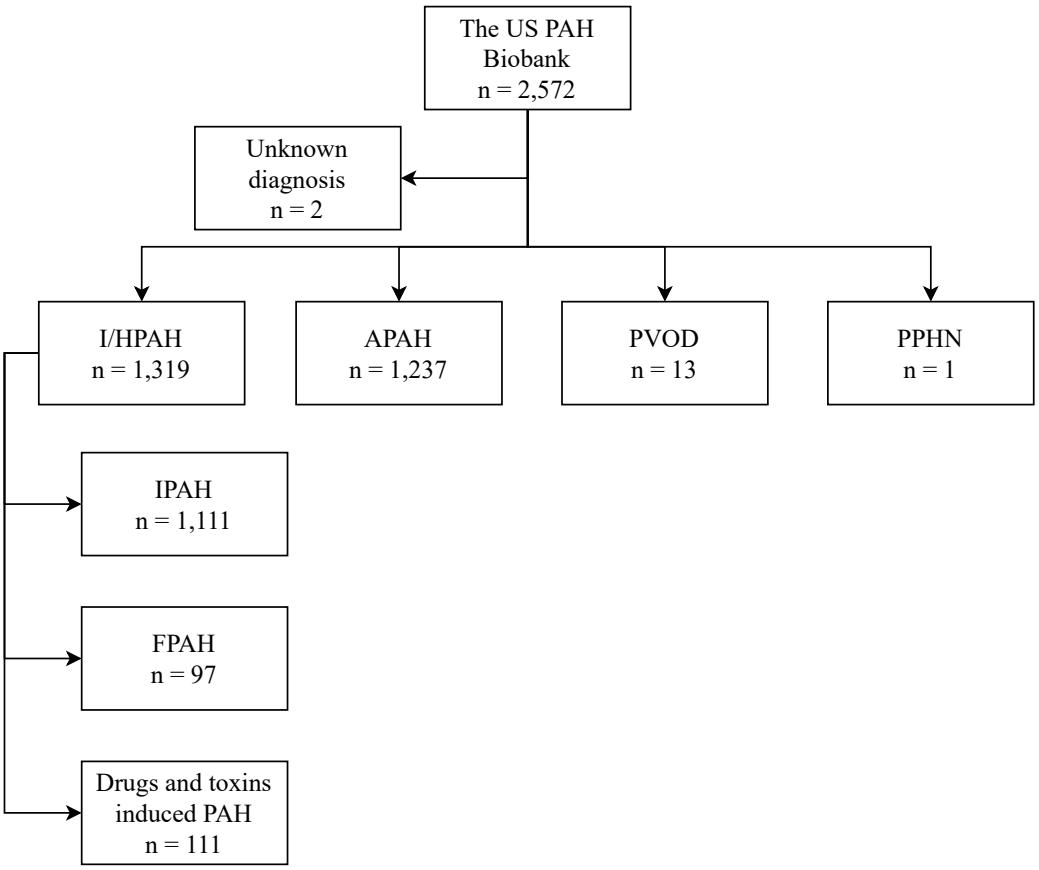


Fig. 3.5 The US PAH Biobank study participants who underwent whole-exome sequencing (WES).

While both studies allowed patients with a history of drugs and toxins exposure, the phenotype definitions varied between the two studies. In the European cohort those patients were labelled as IPAH, whereas in the US PAHBB cohort patients with a history of drugs and toxins exposure with the definite and possible association with PAH had a diagnosis of drugs and toxins induced PAH. In NBR/Cohort studies, out of 96 patients with a history of drug exposure, 69 (4.9%) were adjudicated as drugs and toxins induced PAH and 32 (2.3%) were exposed to more than one agent. In both cohorts, anorexigen exposure was the leading cause of drugs and toxins induced PAH. Fig. 3.6. shows various drugs overlap by study and by diagnosis.

Both studies provided information on self-reported race but the categorisation methods differed making head to head comparisons difficult. Nevertheless, both cohorts included a similar number of patients of European descent (NBR 75.9%, US PAHBB 74.5%). The US PAHBB diagnosed with I/HPAH/PVOD/PCH population was slightly younger at diagnosis which was a result of the larger paediatric population included in the study (Fig. 3.7), when patients above the age of 16 were extracted from both studies, median ages were 50 [35;65] and 49 [38;60] p-value = 0.14 for the NBR study and the US PAHBB study respectively.

Both studies showed similar population structure trends over time (Table 3.5), but differed in respect of sex composition. The female to male ratios were 2.2:1 and 3.4:1 (p-value < 0.001) in the NBR and the US PAHBB cohorts respectively when I/HPAH and PVOD/PCH cohorts were evaluated. It is known from registry data that ethnic minorities have higher female predominance [148]. The analysis of ethnic groups in both cohorts confirmed a higher female to male ratio in ethnic minorities when compared to European populations (77.85 vs 71.4%, p-value = 0.001). Further analysis revealed differences in disease onset between ethnic minorities and patients of European descent (42.0 [29.1;55.9] vs 50.7 [37.0;63.0], p-value < 0.001). The study of the functional status, exercise capacity and haemodynamics revealed a milder phenotype in the US PAHBB cohort. The fraction of patients with positive acute NO challenge was similar between the two studies (6.6 vs 7.4%, p-value = 0.426) (Table 3.6). Comorbidity analysis of patients diagnosed with I/HPAH and PVOD/PCH showed no differences between the two cohorts Fig. 3.8.

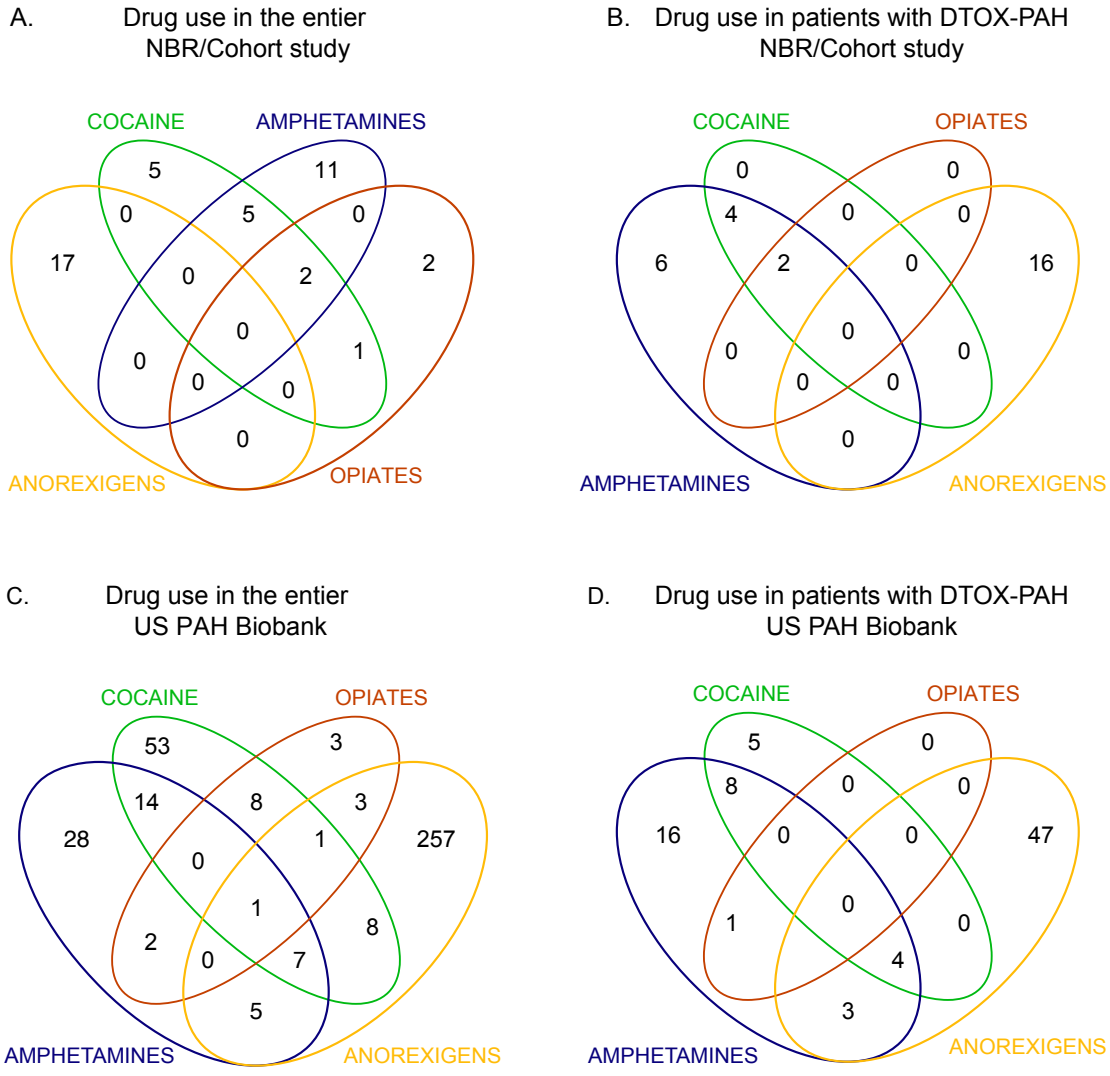


Fig. 3.6 Use of drugs associated with PAH in the entire NBR/Cohort study (A) and the entire US PAHBB cohort (C), and among patients with drugs and toxins induced PAH from the NBR/Cohort study (B) and the US PAHBB (D).

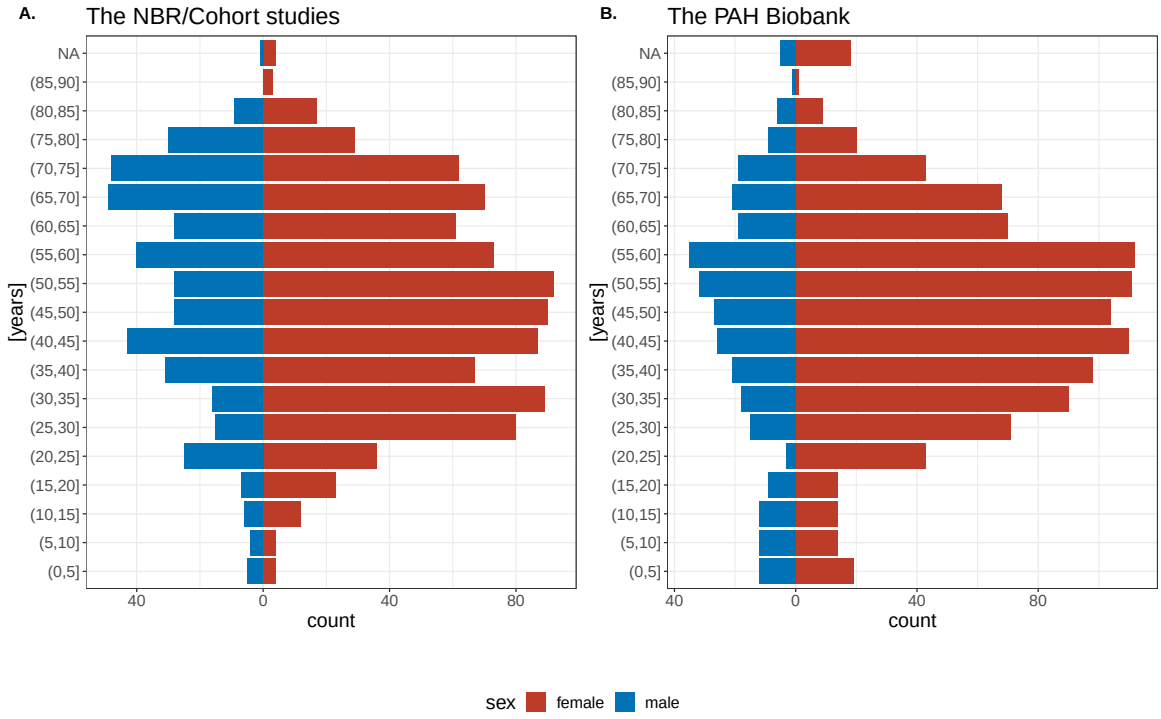


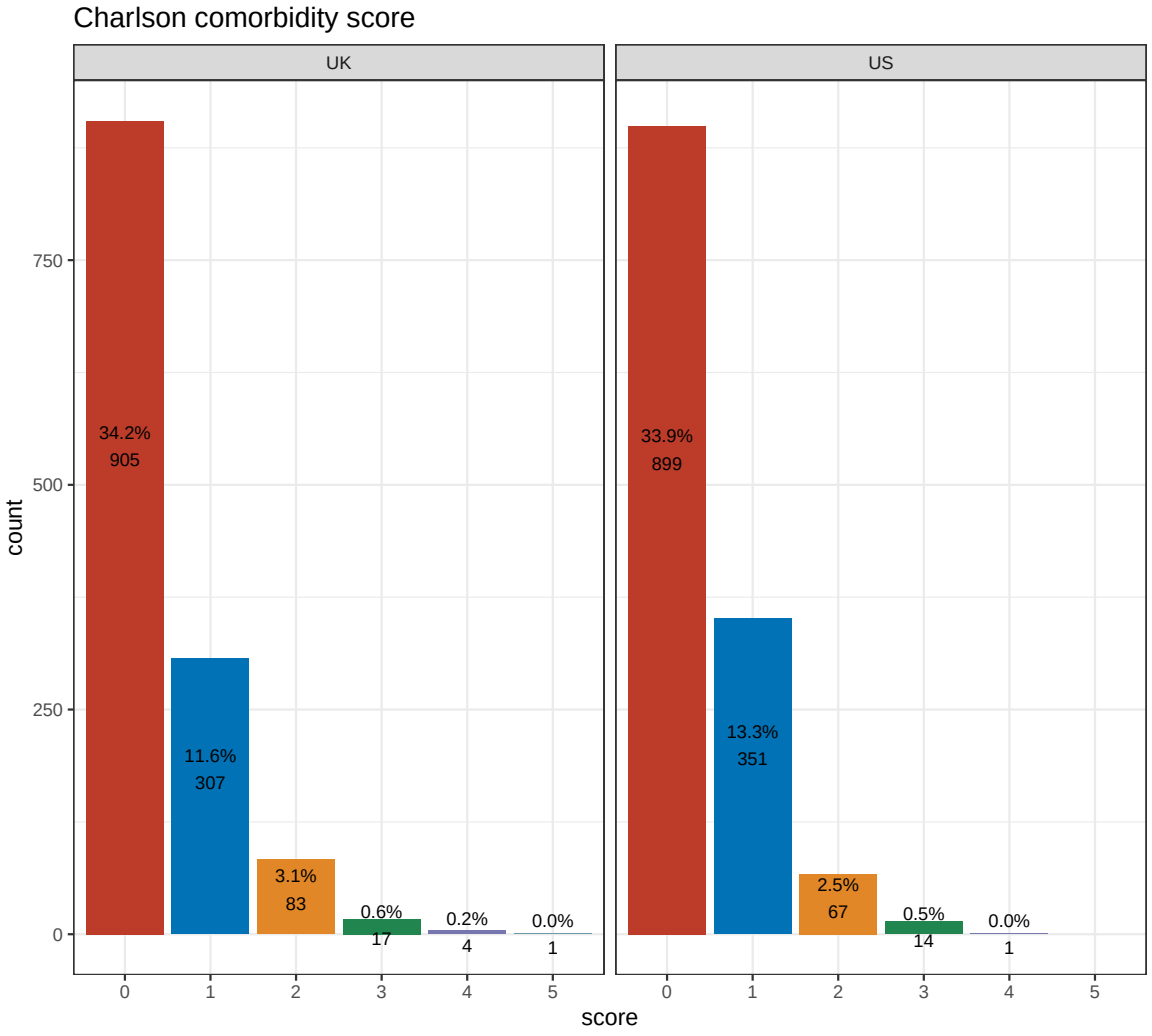
Fig. 3.7 Population structure over time (I/HPAH and PVOD/PCH) in (A) the NBR/Cohort studies, (B) the US PAH Biobank.

Table 3.4 Risk stratification by ESC/ERS and REVEAL risk scores.

	[ALL] N=1375	N
ERS risk score:		1375
low risk	254 (18.5%)	
intermediate risk	1046 (76.1%)	
high risk	75 (5.45%)	
REVEAL risk score:		1016
low risk	519 (51.1%)	
moderately high risk	211 (20.8%)	
high risk	217 (21.4%)	
very high risk	69 (6.79%)	

3.1.3 Discussion

The NBR/Cohort study is the to-date largest cohort of deeply phenotyped PAH patients who were also whole-genome sequenced. It is also the largest PAH repository of biological samples for "omic" studies. While the NBR study allowed a small proportion of samples from European centres, the Cohort study included only the UK patients. This national effort with central coordination aimed at recruiting all eligible patients in the country, is thus an accurate snapshot of the contemporary PAH population. Additionally, the longitudinal design of the PAH Cohort study allows for the study of disease trajectory and facilitates recall by genotype studies. Both studies collect information on all clinically relevant variables obtained to make the diagnosis as well as variables important for follow-up, reflective of current health care practices in the UK. In contrast to the traditional registries, all the clinical information available in medical records was accommodated allowing investigation of new disease patterns and phenotypes. Both structured and unstructured data was collected allowing for free text search. This unprecedented depth and breadth of the data, as well as acceptance of legacy samples, resulted in relatively high missing rates, although this could be mitigated by studying smaller but better-phenotyped subgroups. The current study included mainly patients of European descent, however, the analysis of the ethnic minorities corroborates the data from smaller studies on other populations by showing differences in disease demographics and characteristics. Whether these differences translate into the differential genetic background, response to treatment and overall outcomes remains to be investigated. Both the NBR/Cohort study and the US PAHBB show typical features of the contemporary PAH population, which are female predominance, progressively later disease onset (Table 3.5, Table 3.6, Table B.11) and significant comorbidity burden (Fig. 3.8). In our study, in all cases, the treatment was commenced at the tertiary referral centre and in 48% of



Colors indicate unweighted Charlson comorbidity score.

Fig. 3.8 Charlson comorbidity score in patients diagnosed with IPAH/HPAH/PVOD/PCH (A) The NBR/Cohort studies, (B) The US PAHBB.

cases with combination therapy, indicating that patients are still referred to these centres late in the disease.

Table 3.5 Age at diagnosis (median and IQR) by time period under investigation for the NBR/Cohort study and the US PAH Biobank.

Study	Time period	Age at diagnosis	*p*-value
The NBR/Cohort study	1985 - 1999	37 [28;47]	<0.001
	2000 - 2009	45 [33;58]	
	2000 - 2017	54 [39;68]	
The US PAH Biobank	1963 – 1999	34 [27;45]	<0.001
	2000 – 2009	45 [33;56]	
	2010 – 2017	52 [37;63]	

Table 3.6 Comparison of baseline clinical characteristics and missingness rates between the NBR/Cohort study and The US PAH Biobank.

	The NBR/ Cohort studies n = 1316	**The PAH USBB study n = 1,332**	***p*-value**
Sex: female, n (%)	770 (69.6)	1029 (77.3)	<0.001
missing, n (%)	0 (0.0)	1 (0.0)	0.5
Age at diagnosis [years]	50 [35;64]	48.0 [35-59]	0.011
missing, n (%)	6 (0.46)	20 (1.5)	0.018
WHO FC: I/II/III/IV, %	2/20/65/13	7/30/54/9	<0.001
missing, n (%)	44(3)	374 (28)	<0.001
mRAP [mmHg]	8 [5;12]	9 [6;13]	0.013
missing, n (%)	145 (12)	28 (3)	<0.001
mPAP [mmHg]	53 [44;61]	52 [42;61]	0.124
missing, n (%)	74 (6)	1 (0)	<0.001
CO [L/min]	3.9 [3.1;4.9]	4.3 [3.4;5.3]	<0.001
missing, n (%)	124 (9.4)	415 (40.6)	<0.001
Acute NO challenge*, n (%) (vr)	87 (6.6)	98 (7.4)	0.426
missing, n (%)	777(59)	524 (51.3)	<0.001
6MWD [m]	332 [216;416]	347 [268;413]	0.007
missing, n (%)	210 (15.9)	972(73)	<0.001

I/HPAH and PVOD/PCH cases only. *Lenient vasoresponse criteria were used. Abbreviations: mRAP: mean right atrial pressure; mPAP: mean pulmonary artery pressure; CO: cardiac output; 6MWD: 6-minute walked distance; WHO FC: World Health Organisation Functional Class

Survival in the NBR/Cohort study was better than previously reported (Table B.11). There are a couple of potential explanations for that. Firstly, we present a "clean" I/H-PAH/PVOD/PCH sample which has been recruited to the study to further dissect the underlying genetic architecture of PAH and may therefore be different to "all comers" cohorts reported in national and international registries. Secondly, our cohort is characterised by strong incident/prevalent case bias for which I have accounted in the survival analysis. Finally, this cohort is more recently diagnosed than those previously reported with $> 80\%$ diagnosed

in the last 10 years, therefore improvements in therapy, availability of PAH medication and upfront combination therapy might have impacted survival. Similarly, the poor predictive capacity of the abbreviated ERS risk score might be reflective of a biased population. On the other hand, our patients comparably, to historical cohorts present in functional class III, have a similar degree of haemodynamic derangements, this may indicate that the natural history of the disease has changed and argues against lead-time bias due to earlier diagnosis, it also argues against less severely diseased patients. Interestingly, the differences in survival between the NBR/Cohort study and most recent registries were most notable at 1 and 3 years with convergence at 5 years, which may indicate that the disease course is mostly affected in the first 3 years after diagnosis.

3.1.4 Conclusions

The NBR and PAH Cohort studies population is a representative sample of PAH patients diagnosed in the UK. While the depth of clinical phenotyping of this cohort was unprecedented, the study suffered from many bias typical for multicentre observational studies including high missingness and inclusion of prevalent cases. Similarly, to previously reported studies there were significant differences in survival between prevalent and incident cases as well upward shift in age at diagnosis and number of comorbidities. Whilst the inclusion criteria to the NBR and PAH Cohort studies and the PAH US BB study differed the adult I/HPAH population shared the same clinical characteristics such as female predominance, progressively later age of onset, median mPAP and proportion of patients with positive NO challenge test.

In Chapter 3.1 I addressed aim number 2 of my Thesis.

3.2 Forward phenotyping

Parts of this chapter have been published previously: Swietlik EM, et.al, Bayesian Inference Associates Rare *KDR* Variants with Specific Phenotypes in Pulmonary Arterial Hypertension, Circ Genom Precis Med. 2020 Dec 15;14(1):e003155. PMID: 33320693

3.2.1 Clinical phenotypes

Abnormalities in lung function test in idiopathic and hereditary pulmonary arterial hypertension

Results

These results are based on the 15^{th} July 2020 *OpenClinica* data release and consider patients diagnosed with IPAH or HPAH only. From 2002 to 2019, 817 PAH patients with IPAH (n = 757) or HPAH (n = 60) were recruited to the study from 8 UK PH expert centres (Fig. 3.9).

The data set was analysed for missing rate and pattern. Data was found to have a multivariate (many variables had missingness) missing pattern. In total 20% of data relevant for the analysis were missing. Most of the patients had PFTs recorded at the time of diagnostic RHC. Proportions of missing PFTs and missingness patterns are shown in Fig. 3.10; in brief, percentages of missing data points were 13, 15, 37, 21% for FEV1, FVC, TLC and KCO, respectively.

No significant systematic differences between patients with available and missing PFT values were identified, except that the included subjects were more likely to be incident patients and were slightly older (Table B.12). Clinical follow-up data at target 3 to 12 months post-diagnosis was available in 17% (87/496) of cases with complete PFTs and in 15% (123/823) in the entire population. The reasons for high missingness in follow-up data were twofold, firstly 24% of the study population did not enter the longitudinal study, and secondly, many who entered the study had the first follow-up visits beyond the target period. Additionally, 42 patients died before follow up and therefore were excluded from the analysis (and from the imputation of follow-up data). For the purpose of the imputation procedure, the missing data were assumed to be missing at random [48]. Most subjects were in WHO functional class III with IPAH (Table 3.7). FEV1, FVC, TLC and KCO were $< 80\%$ of predicted in 39%, 22%, 17% and 67% of patients, respectively, whereas 24% had an FEV1/FVC ratio < 0.70 (Fig. 3.11).

Table 3.7 Characteristics of the study population - abnormalities in lung function tests.

	[ALL] N = 817	**N**
Sex: female, n (%)	552 (68)	817
Prevalent case, n (%)	479 (59)	817
Diagnosis verified: n (%)		817
HPAH	60 (7)	
IPAH	757 (93)	
Ethnicity: European, n (%)	677 (83)	817
BMI [kg/m^2]	28.0 [24.2;32.7]	785
Age [years]	51 [37;67]	817
WHO FC: I/II/II/IV, %	1/15/70/14	796
mRAP [mmHg]	9 [6;12]	759
mPAP [mmHg]	53 [44;61]	786
mPAWP [mmHg]	10 [7;12]	709
CO [L/min]	3.8 [3.0;4.8]	768
CI [L/min/m^2]	2.0 [1.7;2.6]	747
PVR [WU]	11.2 [7.9;15.7]	699
SvO_2 [%]	64 [58;70]	720
FEV1 [% pred.]	86.0 [74.0;97.0]	709
FVC [% pred.]	96.0 [82.0;107]	695
FEV1/FVC ratio	0.76 [0.70;0.81]	712
TLC [% pred.]	94.0 [84.0;104]	504
KCO [% pred.]	72.0 [51.7;87.0]	638
Fibrosis: n (%)		567
none	540 (95.2)	
minimal/mild	25 (4.4)	
moderate	2 (0.4)	
Emphysema: n (%)		565
none	504 (89.2)	
minimal/mild	41 (7.3)	
moderate	20 (3.5)	
Smoking: past/current, n (%)	412 (53.2)	774
COPD: n (%)	32 (3.9)	811
OSA: n (%)	43 (5.3)	817
Asthma: n (%)	74 (9.1)	817
Initial therapy: n (%)		630
CCB	54 (8.6)	
combination therapy	323 (51.3)	
ERA	72 (11.4)	
PA	21 (3.3)	
PDE5	159 (25.2)	
sGC	1 (0.2)	

Data are presented as median [interquartile range] unless otherwise specified.*38/61 patients with an initial diagnosis of HPAH had mutations in known PAH genes. 124/638 patients with an initial diagnosis of IPAH had mutations in known PAH genes. Imaging also revealed ascites in 21/812 (2.6%) and pericardial effusion in 116/691 (16.8%) patients. These findings were not significantly different between tertiles of FEV1, FVC, TLC and between patients with (FEV1/FVC < 0.7) and without obstruction. Abbreviations: COPD: chronic obstructive pulmonary disease; I/HPAH: idiopathic/hereditary pulmonary arterial hypertension; KCO: transfer coefficient for carbon monoxide; L/min: liters per minute; mPAP: mean pulmonary arterial pressure; mPAWP: mean pulmonary artery wedge pressure; mRAP: mean right atrial pressure; PVR: pulmonary vascular resistance; WU: Wood units; WHO FC: World Health Organisation Functional Class; CO: cardiac output; CI: cardiac index; SvO_2: mixed venous oxygen saturation; OSA: obstructive sleep apnea; BMI: body mass index; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; TLC: total lung capacity; CCB: calcium channel blockers; ERA: endothelin receptor antagonist; PA: prostacyclin analogue; PDE5: phosphodiesterase type 5; sGC: Soluble Guanylate Cyclase Stimulator

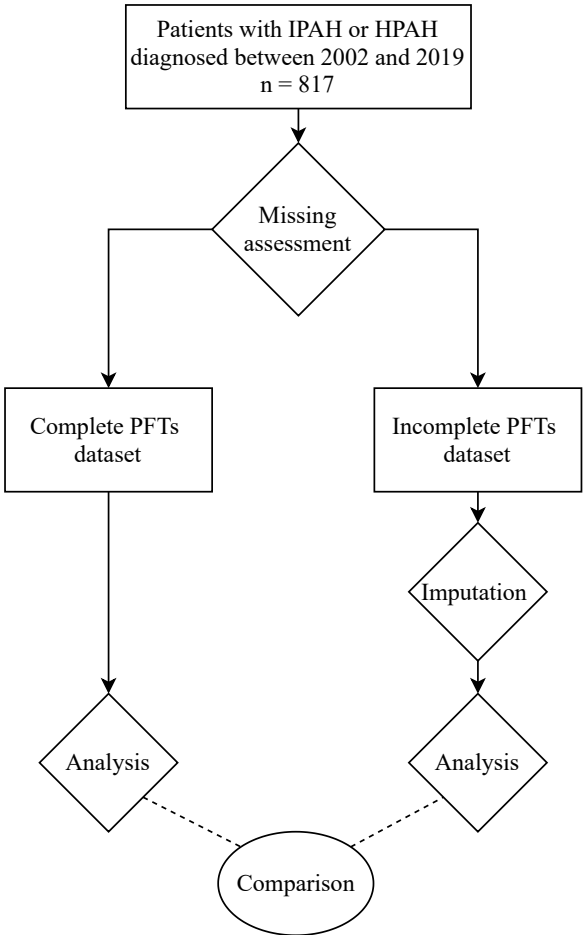


Fig. 3.9 Analysis flow chart for abnormalities in pulmonary function tests in I/HPAH patients.

Similar proportions (39%, 22%, 20%, 65% and 24%) were observed when the few patients with parenchymal abnormalities on CT were excluded. Tables 3.8, 3.9 and 3.10 also describe patient characteristics according to TLC severity and FEV1/FVC ratio and KCO severity, whereas patients' characteristics according to FEV1 and FVC severity are available in Tables B.13 and B.14. Patients with TLC below 80% predicted had worse exercise tolerance and greater desaturation during 6MWT, whereas patients with FEV1/FVC ratio below 0.7 threshold had worse functional class, were more likely to be older and to have a positive smoking history with emphysema on CT (Table 3.9 and Tables B.16 and B.15). Unlike abnormalities in pulmonary volumes, derangements in KCO were more pronounced in males. Patients with reduced KCO were older, had worse functional status and exercise tolerance, were more likely to be smokers and have either fibrotic or emphysematous changes on CT (Table 3.10, Tables B.17 and B.18). Lung function tests were not associated with a specific pattern of pulmonary haemodynamics, except for slightly higher mRAP and PAWP

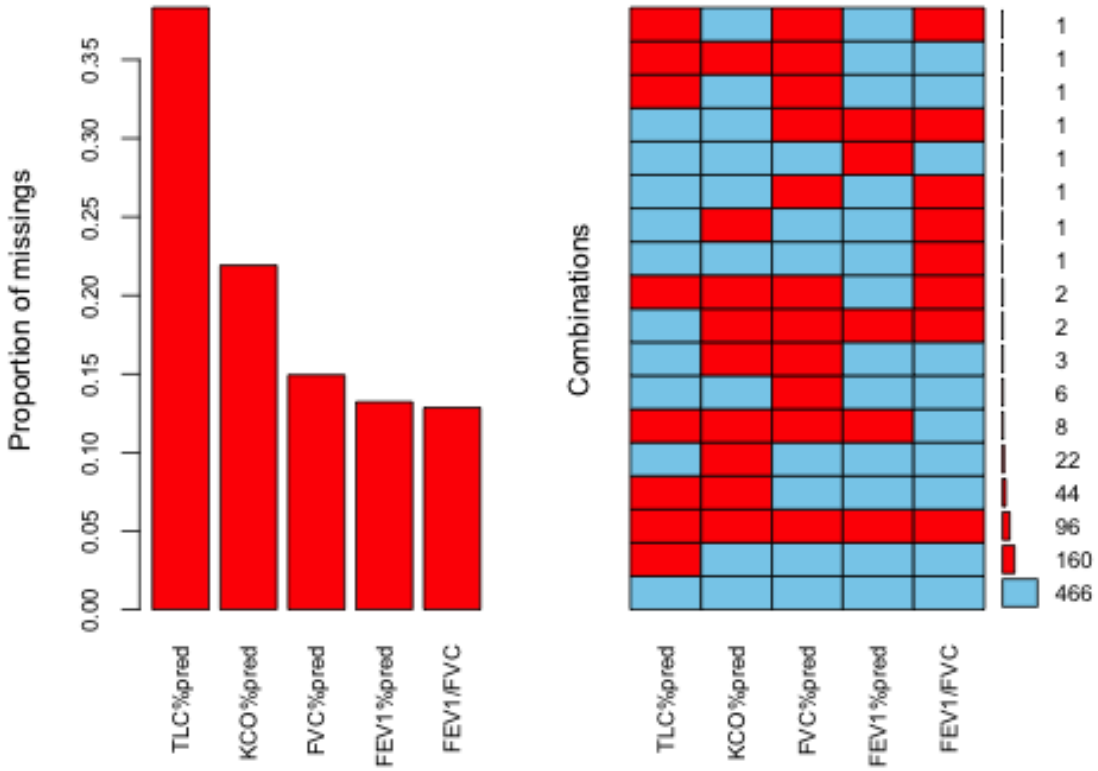


Fig. 3.10 Missingness rate and pattern in pulmonary function tests. Missing data shown in red, existing data shown in blue.

in patients with the lowest FEV1 and FVC, slightly lower mPAP and PVR in patients with decreased KCO. Half of the patients were started on upfront combination therapy (Table 3.7).

During a median follow-up of 5.6 [3-9.9] years, 257 (31.5%) patients died, resulting in a 1, 3 and 5-year overall survival of 97%, 92%, 83% for prevalent, and 96%, 86%, 70% for incident patients, respectively. In univariate analysis KCO, FEV1/FVC ratio and TLC were statistically significant predictors of long-term mortality, whereas FEV1 and FVC did not correlate with outcomes (Tables 3.11, 3.12). Long-term survival according to TLC and KCO severity and FEV1/FVC ratio is shown in Fig. 3.12 A, C and E. ROC curve analyses suggested that cut-off values of 100.5% and 62% of predicted for TLC and KCO, respectively, and 0.76 for FEV1/FVC ratio were the most discriminative for long-term survival (Fig. 3.12 B, D, F). However, neither TLC nor FEV1/FVC ratio maintained predictive capacity in the multivariate analysis after adjusting for age, sex, 6MWD and whether the case was an incident or prevalent (Tables 3.11, 3.12). Consistently, TLC and FEV1/FVC had similar predictive capacity compared to the abbreviated ESC/ERS risk score. Conversely, KCO

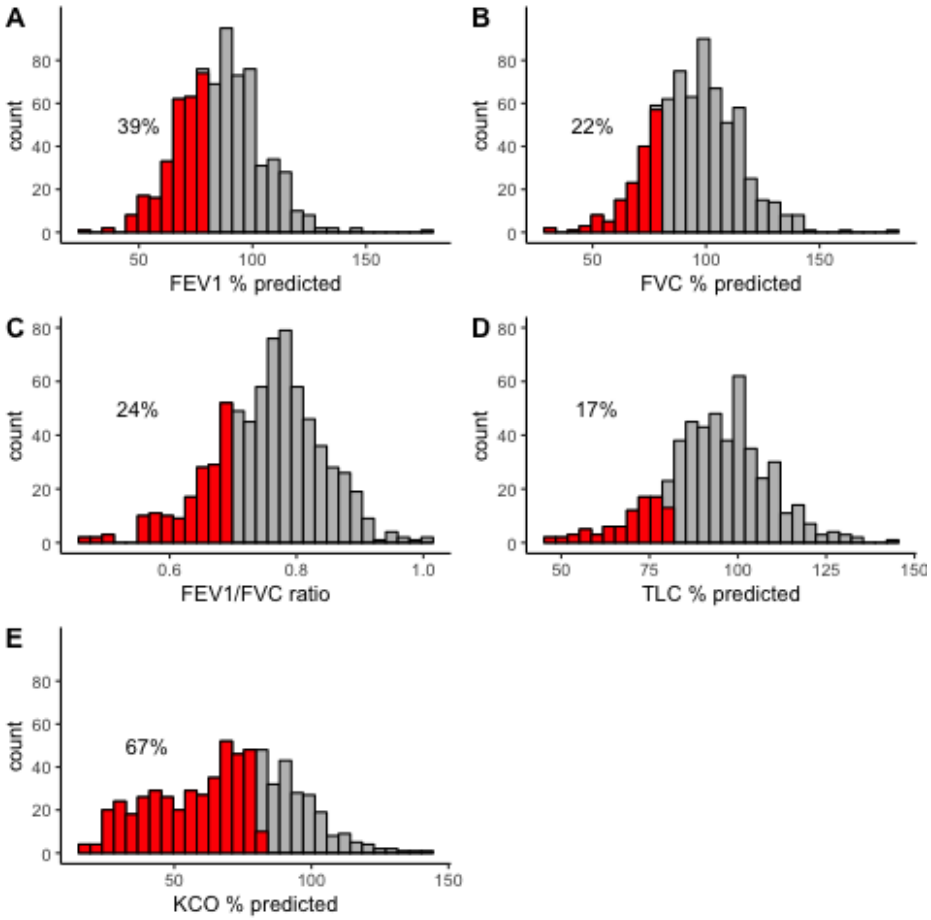


Fig. 3.11 Distribution of % predicted values of pulmonary function tests and FEV1/FVC ratio. (A) FEV1 % pred., (B) FVC % pred., (C) FEV1/FVC ratio, (D) TLC % pred., (E) KCO % pred. In red are the patients with values below 80% predicted for (A) FEV1 (39%), (B) FVC (22%), (D) TLC (17%), (E) KCO (67%) and (C) below 0.7 for FEV1/FVC ratio (24%).

showed better predictive capacity than the abbreviated ESC/ERS risk score. Incorporation of KCO but not lung volumes into the ESC/ERS risk score improved risk stratification (Fig. A.1, Table 3.13).

Post-baseline data available between 3 to 12-months after first-line therapy was initiated were available for 15% (120/817) of cases of the entire study cohort. Patients for whom the follow up data were available were more likely to be incident, receiving initial combination therapy, and were characterised by higher WHO functional class (Table B.19). The median walking distance increased from 311 [165;398] to 366 [265;442] meters and median NT-proBNP levels dropped from 1200 [503;2934] to 339 [128;1111] [ng/l]. Overall, baseline PFT results were not associated with differences in exercise capacity or natriuretic peptides

changes with treatment. The results from the complete cases were replicated in imputed data analysis (Table 3.14).

While the frequency of the effect variants in genes associated lung volumes and diffusion capacity among I/HPAH and PVOD/PCH patients was similar to disease controls (the NBR control cohort), patients with deleterious variants in *BMPR2* and *ATP13A3* had significantly higher KCO (84.9 [75.0;97.0] and 86.0 [79.5;91.3] respectively) when compared to patients with other mutations (64.0 [42.5;86.5]) or to those without mutations (69.0 [48.0;83.9]) in known PAH risk genes (p-value < 0.001).

Table 3.8 Clinical characteristics of study population by total lung capacity threshold.

	TLC $\geq$ 80 N=418	**TLC $<$ 80 N=86**	***p*-value**	**N**
Diagnosis verified: n (%)			0.768	504
HPAH	31 (7%)	5 (6%)		
IPAH	387 (93%)	81 (94%)		
Sex: female, n (%)	295 (71%)	46 (53%)	**0.003**	504
Age [years]	50 [39;66]	61 [45;71]	**0.001**	504
WHO FC: I/II/III/IV, %	1/16/70/13	2/12/65/21	0.076	498
6MWD [m]	320 [192;400]	240 [122;336]	**0.001**	382
SpO_2 pre [%]	96 [93;98]	94 [92;97]	**0.023**	447
SpO_2 post [%]	93 [85;96]	91 [86;95]	0.209	453
mRAP [mmHg]	8 [6;11]	10 [7;14]	**0.002**	479
mPAP [mmHg]	52 [43;60]	48 [42;57]	**0.035**	493
mPAWP [mmHg]	9 [7;11]	10 [7;12]	0.235	445
CO [L/min]	3.7 [3.0;4.7]	3.5 [2.9;4.5]	0.506	487
PVR [WU]	11.2 [7.9;15.8]	11.5 [7.7;15.0]	0.482	448
FEV1 [% pred.]	88 [78;99]	70 [61;80]	**<0.001**	500
FVC [% pred.]	99 [87;109]	75 [66;86]	**<0.001**	491
FEV1/FVC ratio	0.8 [0.7;0.8]	0.8 [0.7;0.8]	**0.003**	498
TLC [% pred.]	98 [89;105]	72 [67;77]	**<0.001**	504
KCO [% pred.]	72.2 [51.8;87.0]	71.5 [58.5;92.2]	0.392	476
Fibrosis: n (%)			0.353	341
none	263 (96.3%)	64 (94.1%)		
minimal/mild	9 (3.3%)	3 (4.4%)		
moderate	1 (0.4%)	1 (1.5%)		
Emphysema: n (%)			0.343	340
none	241 (88.6%)	60 (88.2%)		
minimal/mild	18 (6.6%)	7 (10.3%)		
moderate	13 (4.8%)	1 (1.5%)		
Smoking: past/current	228 (57.3%)	35 (43.8%)	**0.036**	478
COPD, n (%)	20 (4.8%)	2 (2.3%)	0.398	502
Asthma, n (%)	40 (9.6%)	11 (12.8%)	0.48	504
Initial therapy, n (%)			0.623	386
CCB	25 (7.8%)	2 (3.1%)		
combination therapy	162 (50.3%)	32 (50.0%)		
ERA	38 (11.8%)	7 (10.9%)		
PA	5 (1.6%)	2 (3.1%)		
PDE5	91 (28.3%)	21 (32.8%)		
sGC	1 (0.3%)	0 (0.0%)		

Data are presented as median [interquartile range] unless otherwise specified. Abbreviations: ERA: endothelin receptor antagonists; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; I/HPAH: idiopathic/hereditary pulmonary arterial hypertension; KCO: transfer coefficient for carbon monoxide; L/min: litres per minute; mPAP: mean pulmonary arterial pressure; mPAWP: mean pulmonary artery wedge pressure; mRAP: mean right atrial pressure; PDE5: phosphodiesterase type 5; PVR: pulmonary vascular resistance; sGC: Soluble Guanylate Cyclase Stimulator; TLC: total lung capacity; WHO FC: World Health Organization functional class; WU: Wood units; 6MWD: six-minute walk distance; SpO_2: arterial oxygen saturation; CO: cardiac output; COPD: chronic obstructive pulmonary disease; CCB: calcium channel blockers; PA: prostacyclin analogue

Table 3.9 Clinical characteristics of study population according to FEV1/FVC ratio threshold at 0.7.

	FEV1/FVC $\geq$ 0.7 N=539	**FEV1/FVC $<$ 0.7 N=173**	***p*-value**	**N**
Diagnosis verified: n (%)			**0.027**	712
HPAH	44 (8)	5 (3)		
IPAH	495 (92)	168 (97)		
Sex: female, n (%)	365 (68)	109 (63)	0.293	712
Age [years]	47 [34;62]	65 [55;72]	**<0.001**	712
WHO FC: I/II/III/IV, %	1/14/69/16	1/15/75/8	**0.045**	697
6MWD [m]	312 [180;396]	280 [166;376]	0.134	460
SpO_2 pre [%]	96 [93;98]	94 [91;96]	**<0.001**	636
SpO_2 post [%]	92 [85;96]	88 [81;94]	**<0.001**	580
mRAP [mmHg]	9 [6;12]	9 [6;12]	0.716	679
mPAP [mmHg]	53 [45;62]	52 [42;60]	0.149	694
mPAWP [mmHg]	9 [7;12]	10 [8;12]	0.195	631
CO [L/min]	3.8 [3.1;4.8]	3.7 [3.0;4.6]	0.714	686
PVR [WU]	11.4 [7.9;16.1]	10.8 [7.5;13.9]	0.124	623
FEV1 [% pred.]	88 [76;99]	80 [69;91]	**<0.001**	703
FVC [% pred.]	93 [80;106]	100 [90;112]	**<0.001**	693
FEV1/FVC ratio	1 [1;1]	1 [1;1]	**<0.001**	712
TLC [% pred.]	92 (15)	99 (15)	**<0.001**	498
KCO [% pred.]	75.0 [58.0;89.0]	59.0 [40.0;78.0]	**<0.001**	634
Fibrosis: n (%)			0.685	503
none	363 (95.5)	116 (94.3)		
minimal/mild	15 (3.9)	7 (5.7)		
moderate	2 (0.5)	0 (0.0)		
Emphysema: n (%)			**<0.001**	501
none	346 (91.3)	98 (80.3)		
minimal/mild	27 (7.1)	12 (9.8)		
moderate	6 (1.6)	12 (9.8)		
Smoking: past/current, n (%)	250 (49.1)	119 (70.4)	**<0.001**	678
COPD, n (%)	0 (0.0)	32 (18.7)	**<0.001**	710
Asthma, n (%)	52 (9.6)	14 (8.1)	0.643	712
Initial therapy, n (%)			**0.008**	554
CCB	45 (10.6)	4 (3.1)		
combination therapy	221 (52.2)	61 (46.6)		
ERA	46 (10.9)	15 (11.5)		
PA	11 (2.6)	3 (2.3)		
PDE5	99 (23.4)	48 (36.6)		
sGC	1 (0.2)	0 (0.0)		

Data are presented as median [interquartile range] unless otherwise specified. Abbreviations: ERA: endothelin receptor antagonists; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; I/HPAH: idiopathic/hereditary pulmonary arterial hypertension; KCO: transfer coefficient for carbon monoxide; L/min: liters per minute; mPAP: mean pulmonary arterial pressure; mPAWP: mean pulmonary artery wedge pressure; mRAP: mean right atrial pressure; SpO_2: arterial oxygen saturation; CO: cardiac output; COPD: chronic obstructive pulmonary disease; CCB: calcium channel blockers; PA: prostacyclin analogue; PDE5: phosphodiesterase type 5; PVR: pulmonary vascular resistance; sGC: Soluble Guanylate Cyclase Stimulator; TLC: total lung capacity; WHO FC: World Health Organization functional class; WU: Wood units; 6MWD: six-minute walk distance.

Table 3.10 Clinical characteristics of study population according to severity of reduction of KCO

	KCO ≥ 60 N=432	**KCO < 60 N=206**	***p*-value**	**N**
Diagnosis verified: n (%)			**0.031**	638
HPAH	36 (8)	7 (3)		
IPAH	396 (92)	199 (97)		
Sex: female, n (%)	311 (72)	110 (53)	**<0.001**	638
Age [years]	46 [35;60]	66 [53;73]	**<0.001**	638
WHO FC: I/II/III/IV, %	1/17/67/15	0/7/79/14	**0.001**	625
6MWD [m]	330 [221;402]	218 [115;340]	**<0.001**	423
SpO_2 pre [%]	96 [94;98]	93 [90;96]	**<0.001**	571
SpO_2 post [%]	94 [89;97]	85 [78;91]	**<0.001**	525
mRAP [mmHg]	9 [6;13]	8 [5;12]	0.061	605
mPAP [mmHg]	54 [46;64]	49 [42;56]	**<0.001**	623
mPAWP [mmHg]	10 [7;12]	10 [7;12]	0.873	564
CO [L/min]	3.8 [3.0;4.8]	3.7 [3.1;4.6]	0.728	616
PVR [WU]	12.0 [8.0;16.7]	10.4 [7.4;13.6]	**0.001**	564
FEV1 [% pred.]	85 [73;97]	90 [78;99]	**0.002**	636
FVC [% pred.]	93 [80;106]	101 [89;112]	**<0.001**	628
TLC [% pred.]	93 (15)	96 (15)	0.1	476
KCO [% pred.]	81.0 [71.1;92.1]	42.0 [32.2;50.9]	**<0.001**	638
Fibrosis: n (%)			**<0.001**	446
none	290 (98.3)	133 (88.1)		
minimal/mild	5 (1.7)	16 (10.6)		
moderate	0 (0.0)	2 (1.3)		
Emphysema: n (%)			**<0.001**	**445**
none	283 (96.6)	110 (72.4)		
minimal/mild	9 (3.1)	28 (18.4)		
moderate	1 (0.3)	14 (9.2)		
Smoking: past/current, n (%)	198 (47.8)	143 (73.0)	**<0.001**	**610**
COPD	12 (2.8)	17 (8.3)	**<0.001**	**638**
Asthma	54 (12.5)	9 (4.4)	**0.002**	**638**
Initial therapy: n (%)			0.07	503
CCB	29 (8.6)	12 (7.2)		
combination therapy	183 (54.3)	84 (50.6)		
ERA	37 (11.0)	13 (7.8)		
PA	10 (3.0)	1 (0.6)		
PDE5	77 (22.8)	56 (33.7)		
sGC	1 (0.3)	0 (0.0)		

Data are presented as median [interquartile range] unless otherwise specified. Abbreviations: ERA: endothelin receptor antagonists; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; I/HPAH: idiopathic/hereditary pulmonary arterial hypertension; KCO: transfer coefficient for carbon monoxide; L/min: liters per minute; mPAP: mean pulmonary arterial pressure; mPAWP: mean pulmonary artery wedge pressure; mRAP: mean right atrial pressure; SpO_2: arterial oxygen saturation; CO: cardiac output; COPD: chronic obstructive pulmonary disease; CCB: calcium channel blockers; PA: prostacyclin analogue; PDE5: phosphodiesterase type 5; PVR: pulmonary vascular resistance; sGC: Soluble Guanylate Cyclase Stimulator; TLC: total lung capacity; WHO FC: World Health Organization functional class; WU: Wood units; 6MWD: six-minute walk distance.

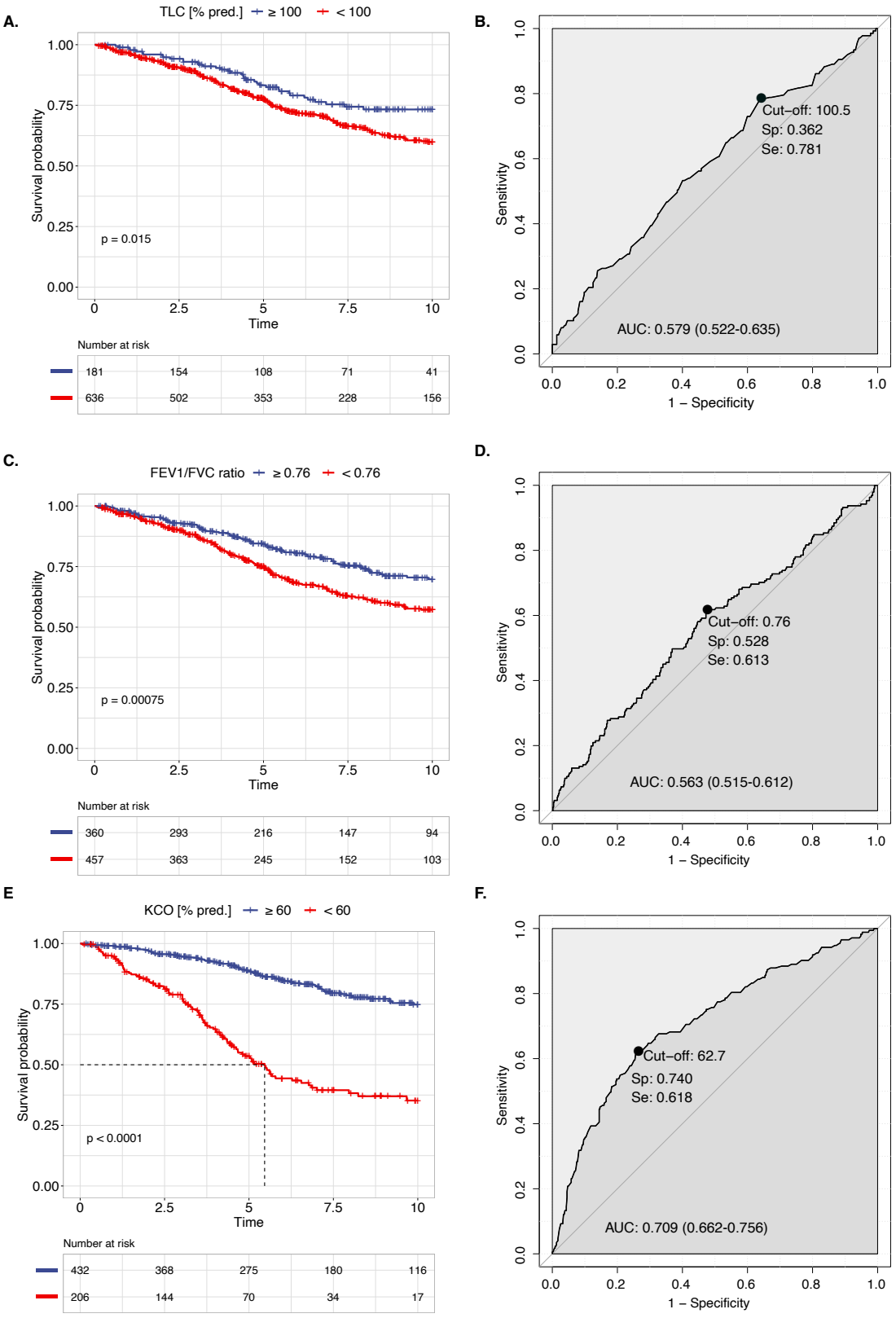


Fig. 3.12 Kaplan-Meier curves showing survival according to TLC (A), FEV1/FVC ratio (C), and KCO (E) threshold, and ROC curves showing the trade-off between sensitivity and specificity for TLC (B), FEV1/FVC ratio (D) and KCO (F).

Table 3.11 Result of Cox regression analysis relating overall survival to individually selected variables at baseline. Complete cases analysis.

	COMPLETE CASES ANALYSIS									
	Univariate		Multivariate							
			Model 1 (TLC)		Model 1 (TLC)		Model 2 (FEV1/FVC ratio)		Model 3 (KCO)	
	HR [95 % CI]	***p*-value**	**HR [95 % CI]**	***p*-value**	**HR [95 % CI]**	***p*-value**	**HR [95 % CI]**	***p*-value**	**HR [95 % CI]**	***p*-value**
Diagnosis: IPAH (HPAH ref)	1.69 [0.92;3.10]	0.086								
Sex: male	1.93 [1.49;2.52]	**<0.001**	1.57 [1.1;2.26]	**0.014**	2.11 [1.42;3.12]	**<0.001**	1.95 [1.45;2.63]	**<0.001**	1.9 [1.37;2.63]	**<0.001**
Age [years]	1.05 [1.04;1.06]	**<0.001**	1.05 [1.04;1.07]	**<0.001**	1.04 [1.02;1.05]	**<0.001**	1.06 [1.05;1.07]	**<0.001**	1.05 [1.04;1.06]	**<0.001**
Case: prevalent	0.54 [0.40;0.74]	**<0.001**	0.4 [0.27;0.61]	**<0.001**	0.46 [0.29;0.71]	**<0.001**	0.62 [0.44;0.87]	**0.006**	0.66 [0.46;0.95]	**0.024**
6MWD [m] for 10 m	0.60 [0.54;0.66]	**<0.001**			0.72 [0.61;0.85]	**<0.001**				
mRAP [mmHg]	1.04 [1.02;1.07]	**<0.001**	1.04 [1.01;1.08]	**0.008**	1.02 [0.99;1.06]	0.189			1.08 [1.05;1.1]	**<0.001**
mPAP [mmHg]	0.99 [0.98;1.00]	**0.005**					1.01 [1;1.02]	0.107	1 [0.99;1.02]	0.866
CI [L/min/m^2]*	0.80 [0.66;0.98]	**0.032**								
PVR [WU]	0.99 [0.96;1.01]	0.263								
FEV1 [% pred.]	1.00 [0.99;1.01]	0.901								
FVC [% pred.]	1.00 [0.99;1.01]	0.636								
FEV1/FVC ratio	0.11 [0.02;0.51]	**0.005**					5.79 [0.93;36.26]	0.06		
TLC [% pred.]	0.98 [0.97;0.99]	**0.001**	0.99 [0.98;1]	0.091	1 [0.98;1.01]	0.554				
KCO [% pred.]	0.97 [0.96;0.97]	**<0.001**							0.98 [0.97;0.98]	**<0.001**
Smoking	1.67 [1.27;2.20]	**<0.001**								

Abbreviations: CI: Confidence interval; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; HR: hazard ratio; KCO: transfer coefficient of carbon monoxide; mPAP: mean pulmonary arterial pressure; mRAP: mean right atrial pressure; PVR: pulmonary vascular resistance; TLC: total lung capacity; WU: Wood units; 6MWD: six-minute walk distance.*CI: cardiac index

Table 3.12 Result of Cox regression analysis relating overall survival to individually selected variables at baseline. Imputed data analysis.

	IMPUTED DATA ANALYSIS									
	Univariate		Multivariate							
			Model 1 (TLC)		Model 2 (TLC)		Model 3 (FEV1/FVC ratio)		Model 4 (KCO)	
	HR [95 % CI]	***p*-value**	**HR [95 % CI]**	***p*-value**	**HR [95 % CI]**	***p*-value**	**HR [95 % CI]**	***p*-value**	**HR [95 % CI]**	***p*-value**
Diagnosis: IPAH (HPAH ref)	1.69[0.92;3.1]	0.09								
Sex: male	1.93 [1.49;2.52]	**<0.001**	1.65 [1.26;2.16]	**<0.001**	2.1 [1.58;2.8]	**<0.001**	1.77 [1.35;2.32]	**<0.001**	1.71 [1.3;2.25]	**<0.001**
Age [years]	1.05 [1.04;1.06]	**<0.001**	1.05 [1.04;1.06]	**<0.001**	1.03 [1.02;1.04]	**<0.001**	1.05 [1.04;1.06]	**<0.001**	1.04 [1.03;1.05]	**<0.001**
Case: prevalent	0.54 [0.4;0.74]	**<0.001**	0.67 [0.49;0.92]	**0.013**	0.72 [0.52;1]	**0.048**	0.73 [0.53;1]	**0.048**	0.76 [0.55;1.05]	0.093
6MWD [m] for 10 m	0.95 [0.94;0.96]	**<0.001**			0.96 [0.95;0.98]	**<0.001**				
mRAP [mmHg]	1.04 [1.02;1.06]	**<0.001**	1.05 [1.02;1.07]	**<0.001**	1.02 [1;1.05]	**0.078**			1.06 [1.04;1.09]	**<0.001**
mPAP [mmHg]	0.99 [0.98;1]	**0.004**					1 [0.99;1.02]	0.416	1 [0.99;1.01]	0.968
CI [L/min/m^2]*	0.8 [0.65;0.97]	**0.026**								
PVR [WU]	0.99 [0.96;1.01]	0.205								
FEV1 [% pred.]	1 [0.99;1.01]	0.791								
FVC [% pred.]	1 [0.99;1.01]	0.591								
FEV1/FVC ratio	0.11 [0.02;0.56]	**0.008**					5.87 [0.97;35.51]	0.054		
TLC [% pred.]	0.98 [0.97;0.99]	**<0.001**	0.99 [0.98;1]	**0.017**	0.99 [0.98;1]	**0.167**				
KCO [% pred.]	0.97 [0.96;0.98]	**<0.001**							0.98 [0.97;0.98]	**<0.001**
Smoking	1.67 [1.27;2.2]	**<0.001**								

Abbreviations: CI: Confidence interval; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; HR: hazard ratio; KCO: transfer coefficient of carbon monoxide; mPAP: mean pulmonary arterial pressure; mRAP: mean right atrial pressure; PVR: pulmonary vascular resistance; TLC: total lung capacity; WU: Wood units; 6MWD: six-minute walk distance.*CI: cardiac index

Table 3.13 ROC analysis on complete and imputed data.

	Complete cases analysis					Imputed data analysis				
	Best cut off	**Specificity**	**Sensitivity**	**AUC (95%CI)**	**SE**	**Best cut off**	**Specificity**	**Sensitivity**	**AUC (95%CI)**	**SE**
FEV1/FVC ratio	0.8	0.53	0.61	0.56 (0.52;0.61)	0.02	0.76	0.56	0.55	0.55 (0.51;0.6)	0.26
TLC % pred.	100.5	0.36	0.78	0.58 (0.52;0.64)	0.03	95.3	0.48	0.66	0.59 (0.54;0.63)	0.25
KCO % pred.	62.7	0.74	0.618	0.7 (0.66;0.76)	0.03	63.93	0.71	0.6	0.69 (0.65;0.73)	0.22
ESC/ERS risk score	1.9	0.56	0.64	0.63 (0.59;0.67)	0.03	1.9	0.56	0.63	0.63 (0.59;0.67)	0.24
ESC/ERS risk score with FEV1/FVC ratio	1.8	0.46	0.72	0.62 (0.58;0.67)	0.03	2.19	0.67	0.54	0.62 (0.58;0.67)	0.24
ESC/ERS risk score with TLC % pred.	2.1	0.79	0.42	0.64 (0.59;0.68)	0.02	2.19	0.71	0.5	0.63 (0.59;0.68)	0.24
ESC/ERS risk score with KCO % pred.	1.92	0.56	0.69	0.67 (0.63;071)	0.03	1.92	0.62	0.6	0.64 (0.6;0.69)	0.24

Abbreviations: FEV1: forced expiratory volume in 1 sec; FVC: forced vital capacity; TLC: total lung capacity; KCO: transfer factor for carbon monoxide; ESC: European Society of Cardiology; ERS: European Respiratory Society. SE in imputed data analysis pooled using Rubin's rules.

Table 3.14 Correlation between baseline TLC and FEV1/FVC ratio and the change in 6MWD and NTproBNP.

Correlation		**Complete cases analysis n=120**		**Imputed data n = 775***	
Variable 1	**Variable 2**	**Rho [95%CI]**	***p*-value**	**Rho [95% CI]**	***p*-value**
Delta 6MWD [m]	TLC % pred.	0.1 [-0.18;0.38]	0.49	0.012 [-0.149;0.172]	0.884
Delta NTproBNP [ng/l]		-0.04 [-0.33;0.26]	0.79	-0.025 [-0.253;0.206]	0.835
Delta 6MWD [m]	FEV1/FVC ratio	-0.21 [-0.41;0.01]	0.08	-0.029 [-0.201;0.145]	0.749
Delta NTproBNP [ng/l]		0.04 [-0.25;0.33]	0.81	-0.022 [-0.23;0.188]	0.839
Delta 6MWD [m]	KCO % pred.	-0.22 [-0.45;0.02]	0.08	-0.028 [-0.213;0.159]	0.774
Delta NTproBNP [ng/l]		-0.08 [-0.4;0.24]	0.58	-0.035 [-0.273;0.207]	0.78

Complete cases analysis and imputed data analysis. *42 patients died; Abbreviations: 6MWD: six-minute walk distance; NTproBNP: N-terminal pro B-type natriuretic peptide; TLC: total lung capacity; FEV1: forced expiratory volume in 1 sec; FVC: forced vital capacity; KCO: transfer factor for carbon monoxide

Discussion

The present analysis confirmed that the most common pulmonary function abnormality in patients with IPAH/HPAH is a decrease in KCO followed by mild-to-moderate obstructive, restrictive or mixed respiratory impairment. KCO reduction was more common in males and correlated with age, exercise capacity, smoking and emphysema on CT scan, and was an independent predictor of survival. The results of this analysis showed that a significant proportion of patients with I/HPAH exhibit mild-to-moderate obstructive, restrictive or mixed respiratory impairment even in the absence of significant bronchial or parenchymal lung disease. These abnormalities were largely unrelated to smoking history, pulmonary haemodynamic severity, the presence of pericardial and pleural effusion or ascites, or concomitant fibrosis and emphysema on CT, except for obstructive lung disease that was more common in older patients with a smoking history and signs of emphysema on CT. Conversely, lung volume abnormalities were consistently associated with poor exercise performance. Ultimately, neither low TLC nor FEV1/FVC ratio were associated with long-term survival after adjusting for confounders, and they did not impact on response to treatment or improve the discriminative capacity of the abbreviated ESC/ERS risk score.

Although previously described in IPAH, the association between abnormalities in pulmonary mechanics and outcomes has been largely overlooked. Early studies documented that 20 - 40% of PAH patients exhibit lung restriction (defined as TLC $< 80\%$ of predicted) [63, 224, 249, 444], while a similar proportion of patients have airway obstruction (defined as FEV1/FVC ratio < 0.70) [63, 485, 98, 133]. Small airway disease was first documented during post-mortem studies and lung biopsy specimens of patients with PAH [133]. Subsequently, experimental models of PH induced by chronic hypoxia or monocrotaline documented that structural changes in the pulmonary vasculature extend to the airways, resulting in increased airways resistance [285, 239]. More recently, studies showed that bosentan prevented endothelin-induced bronchoconstriction [346], whereas endothelin-1 circulating levels significantly correlated with bronchial obstruction in patients with PAH associated with congenital heart disease [319]. Similarly, impaired production of prostacyclin [386] and nitric oxide [41] have bronchoconstrictive properties, mimicking several features of asthma [206]. Changes in lung compliance have also been reported in animal models with altered pulmonary haemodynamics, presumably resulting from altered elastic recoil and compliance due to pulmonary vascular rigidity and mechanical coupling of the vascular and air spaces [168, 205]. Restrictive ventilatory impairments may also result from extra-parenchymal/vascular abnormalities, including enlarged proximal pulmonary vessels, cardiomegaly, pericardial effusion, pleural effusion and marked ascites. Importantly, PFT

abnormalities did not [63, 498, 249], or minimally [352], correlate with haemodynamics severity in previous cohorts, suggesting that air-flow limitation and ventilatory restriction may occur independently of the severity of PAH. It is possible that small airways and lung parenchyma are affected as innocent bystanders in PAH, although ventilatory impairment may also be related to potential confounders, such as previous smoking, coincidental obstructive or parenchymal disease [462], or congenital heart defects that may impair lung development [63, 249, 326]. Consistent with previous studies, 18% and 26% of patients in this contemporary cohort of patients with I/HPAH exhibited restrictive and obstructive ventilatory impairment, respectively, and lung volumes were generally not associated with differences in pulmonary haemodynamics, evidence of extraparenchymal restriction or comorbidities, except for a non-clinically relevant increase in mRAP and more frequent smoking history and emphysema on CT amongst patients with an FEV1/FVC ratio < 0.70. Airway obstruction may thus be coincidental and attributed to the relatively high prevalence of current or previous smoking history documented at baseline. It is noteworthy, however, that obstructive ventilatory impairment was twice as prevalent as compared to the general population with a similar smoking history [370].

Importantly, while TLC of 100.5% predicted and FEV1/FVC ratio of 0.76 best discriminated I/HPAH patients in terms of survival, pivotal randomised clinical trials evaluating currently available PAH therapies included patients with FEV1/FVC ratio and TLC as low as 50% and 60 - 70% of predicted, respectively (Table B.1). Similarly, the presence of minor co-existing lung disease on CT scan was recently associated with poorer long-term prognosis [393, 305], whereas PAH studies have not previously utilized chest imaging to exclude patients with lung disease. It is thus possible that a number of patients with lung volumes above these inclusion thresholds might have an underappreciated burden of parenchymal lung disease, and indeed constitute a distinct phenogroup [393, 305]. In the current analysis, initial improvements of PAH patients with mild-to-moderate lung volume abnormalities with available therapies were of similar magnitude to patients with normal PFTs. These findings thus support the eligibility criteria frequently used in pivotal trials, although the effects of current therapies on clinical worsening and survival remain unknown in the absence of subgroup analyses from these studies, highlighting the importance of prespecified subgroup analyses and data sharing for delineating clinical phenotypes predicting treatment responders, and advancing precision medicine in PAH.

While reduced lung volumes seem to be a bystander phenotype, a reduction of diffusion capacity is almost a universal feature in patients with pulmonary hypertension irrespective of the aetiology, with the lowest values observed in patients with PVOD/PCH, which can be

explained by significant capillary involvement in this group. Thus it is not a surprise that the only reported, statistically significant SNP associated with KCO (but not DLCO) is in the gene *ADGRG6* (rs17280293), which plays an important role in angiogenesis through regulation of VEGFR2 [516]. Despite a broad spectrum of factors influencing KCO (obstructive and restrictive ventilatory impairment, anaemia, heart failure and smoking), this feature remained a significant predictor of survival after adjusting for confounders. Thus, I conclude that stratification by KCO as a primary measurement might be a useful clustering method for genetic analysis.

Pulmonary arterial hypertension in long term responders to calcium channel blockers (CCB)

These results are based on the 15^{th} July 2020 *OpenClinica* data release and consider either all patients or only those diagnosed with I/HPAH and PVOD/PCH as indicated in the analysis from both the NBR and the PAH Cohort studies.

Acute vasoresponse criteria and missingness assessment

I initially used three sets of diagnostic criteria to define positive NO challenge: strict haemodynamic criteria (decrease of mPAP of $\geq$ 10 mmHg to reach an absolute value of mPAP $\leq$ 40 mmHg with an increased or unchanged cardiac output (CO) = vasoresponder - stringent; see Chapter 1.2.1), lenient haemodynamic criteria (only mPAP criteria met to accommodate instances where CO data was missing = vasoresponder lenient), and additionally, drug information was checked to assert which patients were treated with CCB (either stringent or lenient haemodynamic criteria met and the patient treated with CCB alone or in combination = vasoresponder complex). The overlap between diagnostic categories is shown in Fig. 3.13. Because of the high missing rate (42%) of initial treatment modality and the high proportion of patients with a positive acute NO challenge being treated with medication other than CCB (17.3%), I dropped the CBB treatment criterium and in further analysis defined acute vasoresponders as those meeting either lenient or stringent haemodynamic criteria. Five UK cases had acute NO challenge repeated during follow up; all of them showed sustained vasoresponse.

Results

Of all 1393 patients (one patient with HHT and no PH excluded) included in the study, 550 (39.5%) had the acute vasoreactivity test performed at diagnosis. The breakdown of diagnosis of all study participants by the vasoreactivity test is summarised in Table B.20 and showed that although some patients with a diagnosis other than I/HPAH or PVOD/PCH had acute NO test done, none of these patients fulfilled vasoreactivity criteria. Subsequently, a subset of patients diagnosed with I/HPAH and PCH/PVOD were taken forward for the analysis. Those who had the test done were significantly younger (45 [32;58] vs 53 [39;67], p-value < 0.001), had better exercise capacity (370 [285;450] vs 306 [180;396], p-value < 0.001), higher cardiac output (4 [2;7] vs 3.7 [3;4.7], p-value < 0.001), lower NTproBNP (776 [239;2487] vs 1304 [324;3329], p-value $= 0.012$) and were less likely to have obstructive lung physiology (19 (3.52%) vs 53 (6.82%), p-value $= 0.014$). There were no differences between the groups in regards to parenchymal abnormalities and the frequency of other comorbidities. In keeping

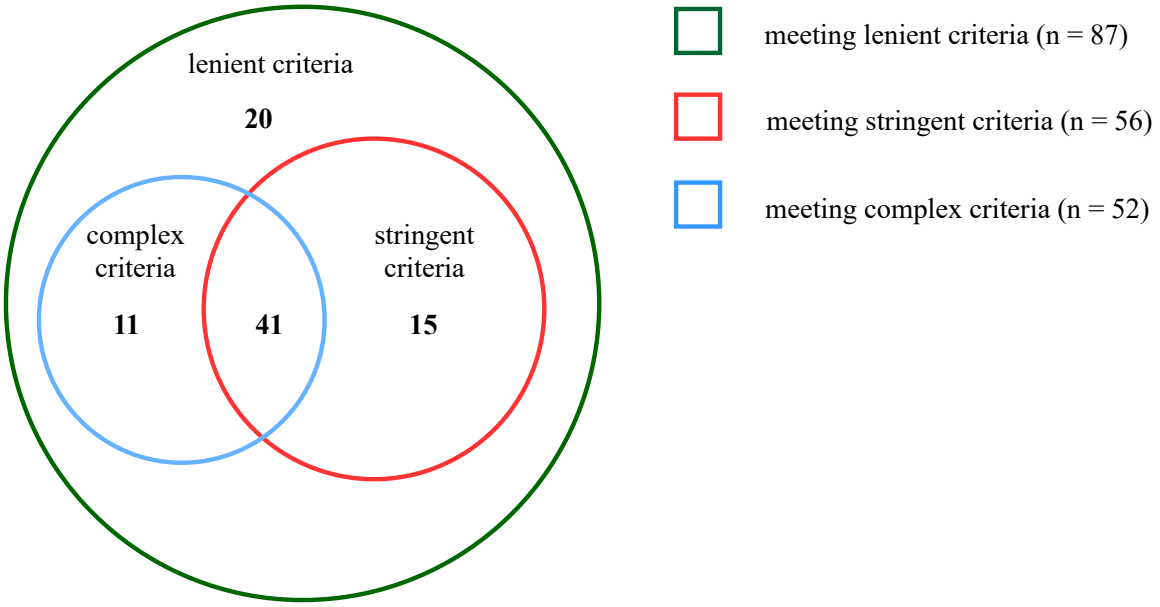


Fig. 3.13 Overlap between patients fulfilling various diagnostic criteria used to define acute vasoreactivity

with the above differences, disease severity scores were lower in patients who had the test done (ESC/ERS risk score p-value = 0.001 and Reveal risk score p-value < 0.001). As there were no differences between patients with a negative test and those who did not have the test done, missing data were not imputed under the assumption that those without the data would be negative if the test was done.

Patients having a positive acute NO challenge test (lenient criteria met) were more likely to be diagnosed with IPAH than HPAH, were younger at diagnosis, had less severe haemodynamics, lower NT-proBNP or BNP, and CRP levels. Although there were no differences in 6MWD, resting and post exercise SpO_2 were higher in vasoresponders (Table 3.15). Non-responders were more often in the high and very high-risk category by REVEAL risk score (p-value = 0.001). The same associations were seen when stringent criteria were applied. As expected, acute vasoresponders (defined by lenient criteria) had better survival than non-responders (p-value = 0.0037) (Fig. 3.14) also after adjusting for age and sex; this finding was not replicated when stringent criteria were applied. There were no differences in survival between non-responders and patients who did not have the test done (p-value = 0.2).

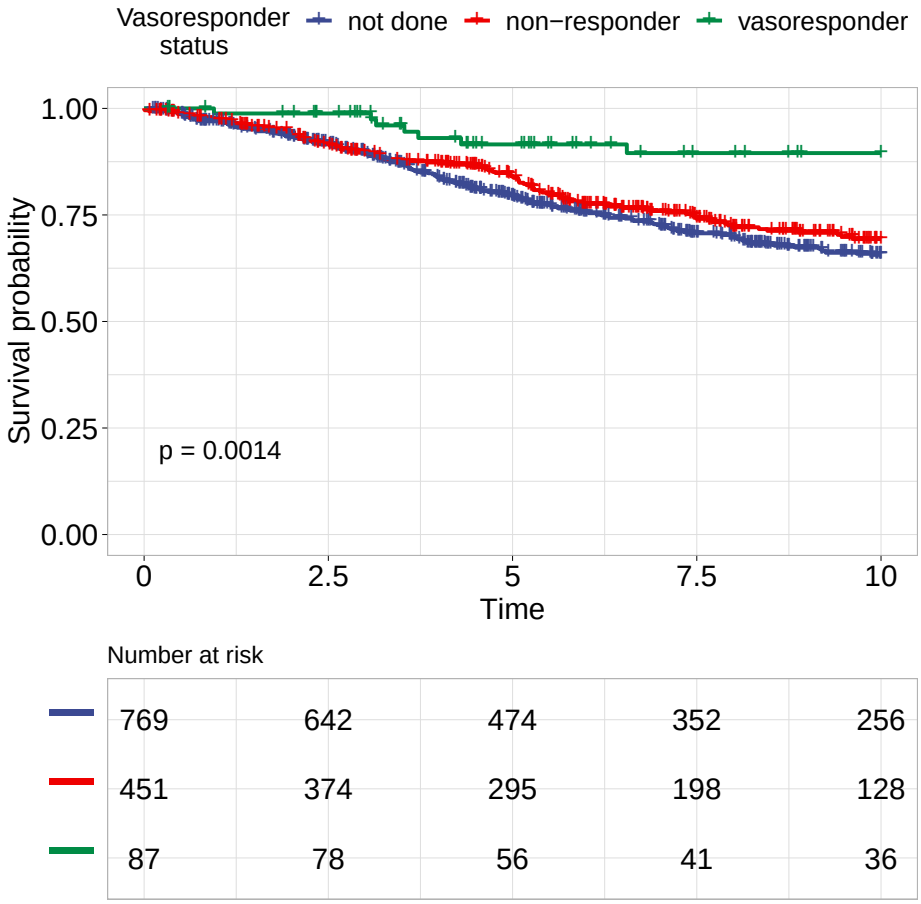


Fig. 3.14 Long-term survival according to the vasoresponse status. Two patients in non-responder group were excluded from the survival analysis due to missing date of diagnosis.

Table 3.15 Comparison of clinical characteristics between vasoresponders and non-responders.

	non-responder N=453	vasoresponder N=87	*p*-value	N
Diagnosis: n (%)			**0.036**	540
IPAH	400 (88.3)	84 (96.6)		
HPAH	45 (9.93)	2 (2.30)		
PVOD/PCH	8 (1.77)	1 (1.15)		
Age [years]	45 [33;59]	42 [27;52]	**0.005**	540
Sex: female, n (%)	298 (66)	67 (77)	0.054	540
Incident case, n (%)	140 (31.0)	41 (47.1)	**0.005**	538
WHO FC: I/II/III/IV, %	2/21/67/10	2/30/53/14	0.079	528
6MWD [m]	368 [285;441]	390 [299;488]	0.152	354
Shuttle walk distance [m]	340 [210;432]	370 [340;482]	0.127	83
SpO_2 pre [%]	96.0 [93.0;98.0]	97.0 [96.0;98.0]	**<0.001**	464
SpO_2 post [%]	92.0 [85.0;96.0]	94.0 [86.0;97.0]	**0.033**	413
mRAP [mmHg]	8 [5;12]	6 [5;9]	**0.003**	522
mPAP [mmHg]	53 [45;62]	50 [44;55]	**0.015**	539
mPAWP [mmHg]	9 [6;11]	8 [5;11]	0.072	514
CO [L/min]	4.0 [3.2;4.9]	4.7 [3.8;5.7]	**<0.001**	532
SvO_2 [%]	63.9 [57.3;69.0]	71.0 [66.5;75.0]	**<0.001**	475
FEV1 [% pred.]	88.1 (18.3)	89.7 (15.4)	0.437	438
FVC [% pred.]	94.8 (19.6)	97.7 (15.9)	0.19	433
TLC [% pred.]	94.8 (15.0)	94.2 (13.1)	0.761	318
KCO [% pred.]	71 [53;86]	72 [63;84]	0.524	329
Emphysema: n (%)			**0.022**	288
minimal/mild	15 (6)	4 (7)		
moderate	2 (1)	4 (7)		
Fibrosis: minimal/mild, n (%)	12 (5)	3 (5)	1	288
Smoking: past/current, n (%)	171 (53.3)	36 (45.0)	0.23	401
NTproBNP [ng/l]	1093 [244;2841]	518 [126;1151]	**0.016**	152
BNP [ng/l]	233 [86.0;466]	77.5 [26.9;98.0]	**0.015**	138
CRP [mg/l]	4.00 [2.00;7.80]	2.45 [1.77;4.78]	**0.004**	271
Creatinine [μmol/l]	86.0 [72.0;102]	75.0 [66.0;89.0]	**<0.001**	373
Initial therapy: n (%)			**<0.001**	341
CCB	14 (5.3)	41 (54.7)		
combination therapy	150 (56.4)	21 (28.0)		
ERA	24 (9.0)	4 (5.3)		
PA	15 (5.6)	1 (1.3)		
PDE5	62 (23.3)	8 (10.7)		
sGC	1 (0.4)	0 (0.0)		

Abbreviations: ERA: endothelin receptor antagonists; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; I/HPAH: idiopathic/hereditary pulmonary arterial hypertension; PVOD/PCH: pulmonary vasooclusive disease/pulmonary capillary hemangiomatosis; KCO: transfer coefficient for carbon monoxide; L/min: litres per minute; mPAP: mean pulmonary arterial pressure; mPAWP: mean pulmonary artery wedge pressure; mRAP: mean right atrial pressure; PDE5: phosphodiesterase type 5; sGC: Soluble Guanylate Cyclase Stimulator; TLC: total lung capacity; WHO FC: World Health Organization functional class; WU: Wood units; 6MWD: six-minute walk distance; SpO_2: arterial oxygen saturation; CO: cardiac output; CCB: calcium channel blockers; CRP: C-reactive protein; NTproBNP: N-terminal pro B-type natriuretic peptide; BNP: B-type natriuretic peptide; PA: prostacyclin analogue; SvO_2: mixed venous oxygen saturation

Discussion

PAH in long-term responders to CCB is now a well established PAH subphenotype recognised in the recent PH classification. Responders to CCB are characterised by younger age of onset, and clear survival benefit. The phenotype can be discerned based on the results of acute NO challenge [473]. The reported proportion of patients meeting acute vasoresponse criteria varies from as high as 32.2% in Danish registry [275] to as low as 3.5% in Chinese registry [591], but usually stays below 10% (5.8% in the US cohort diagnosed in between 2002 and 2006 [519], 10.3% of IPAH cases in French registry [229]). The purpose of the current analysis was to test missingness rate and causes and discern the best criteria for positive vasoresponse testing in view of missingness. My analysis showed that patients who did not have acute vasoresponse tests done clustered together with those with negative tests. Because of that, I decided against imputation of acute NO challenge results assuming that patients with missing results would test negative. Subsequently, based on the available data I devised three sets of criteria to define acute vasoresponders. Lenient vasoresponse criteria captured the biggest group 87 (6.2%), whose clinical characteristics and outcomes aligned the best with reported characteristics of patients with long-term response to CCB. While pathway-based analysis of WES data revealed multiple rare variants that differed between vasoresponders and non-responders [208], GWAS in this subgroup did not yield any positive results [436]. Interestingly, data so far indicates that patients with deleterious mutations in known PAH risk genes usually do not show acute response to NO [124]. This might be due to the polygenic contribution to the expression of this trait or suggest that the analyses preformed so far were underpowered to detect the association. Strong survival benefit and response to particular class of mediation makes this phenotype an interesting candidate for genotype-phenotype analysis.

Pulmonary arterial hypertension with autoimmune features

These results are based on the 15^{th} July 2020 *OpenClinica* data release and consider either all patients or only those diagnosed with I/HPAH and PVOD/PCH as indicated in the analysis from both the NBR and the I/HPAH Cohort studies.

Results

Out of 1,394 patients in the study, the autoimmune screen, which included anti-ANA, anti-dsDNA, anti-ENA, anti-SSA/Ro, anti-sc70, anti-centromere, ANCA and anti-cardiolipin was missing in 424 (30%) individuals (Fig. A.2). Patients who did not have an autoimmune screen reported were younger (48 [33;61] vs 51 [37;66], p-value = 0.002), were more often prevalent (303 (81%) vs 592 (61.3%), p-value < 0.001), had better exercise capacity (352 [240;432] vs 327 [213;412], p-value = 0.031) and were less likely to be current or past smokers (434 (54%) vs 95 (42%), p-value = 0.003). Significant variability within and between centres in both screening and reporting was noticed. Of patients who had an autoimmune screen performed (n = 970), 928 (96%) were diagnosed with I/HPAH/PVOD/PCH, while 8 (0.8%) were diagnosed with PAH-CTD. Among patients with I/HPAH/PVOD/PCH 174 (18.8%) had at least one autoantibody detected, the most common was anti-ANA antibody positivity recorded in 130 (14.8%) patients, followed by ANCA antibodies detected in 26 (6%) patients (Table B.21). I/HPAH/PVOD/PCH patients who met the criterion for the presence of autoimmune phenomena were older (55 [43;69] vs 49 [35;65], p-value < 0.001), had lower KCO (62 [42;79] vs 71 [53;87] % pred., p-value = 0.002) and more mild and moderate fibrotic changes on HRCT (9% vs 4%, p-value = 0.01) (Table 3.16). Survival was significantly worse in patients with autoantibody positivity when compared to those without autoantibodies (p-value = 0.03 for the entire cohort, p-value = 0.027 for I/HPAH/PVOD/PCH), but the significance was lost after adjusting for age and sex. The only single antibody that predicted survival both in the entire cohort and in I/HPAH/PVOD/PCH was anti-ANCA antibody (p-value = 0.00011 and p-value = 0.00013 respectively) and this remained significant after adjustment for confounders. Because of high missingness and inconsistent reporting, antibody measurement was repeated on a representative sample of the PAH Cohort study. A panel of autoantibodies was measured by immunofluorescence in 473 patients (Table B.22.). The percentage of patients with any autoantibody positivity or anti-ANA autoantibody positivity was significantly higher than reported (26.8% vs 19% p-value = 0.003 any antibody, 21.8% vs 15.5% p-value = 0.005 ant-ANA antibody; comparisons between patients diagnosed with either IPAH or HPAH), also contrary to the previous analysis, there was no difference in age, KCO or lung parenchymal changes or comorbidities other than thyroid disease between

patients with at least one autoantibody positivity and those without (Table 3.17). When the results of autoantibody testing were compared between reported and measured there was a very poor agreement (Kappa = -0.152 for anti-ANA antibody, Kappa = -0.00162). There was no survival difference after adjustment for confounders.

Discussion

In this chapter I explored the prevalence of autoimmunity among patients with PAH and hypothesised that patients with autoimmune phenomena constitute a distinct phenotype. Based on the finding that two thirds of the known autoimmune loci are shared between autoimmune diseases [130, 563] I defined the autoimmune phenotype as a presence of any autoantibody either reported in clinical records or measured in house using the immunofluorescence assay. I found that frequency of positive ANA antibodies in the current cohort was lower than in other studies: 15.5% reported vs 21.8% in house measured in the NBR and the Cohort studies vs 29% in the NIH registry [438] vs 33% in the Chicago registry [519]. Furthermore there was a poor agreement between reported and in house measured antibodies. The reasons for poor agreement within the same study can be at least two fold, firstly as the measurements were taken at different time points one can suspect a certain degree of variability in seropositivity over time contributed to differences, a phenomenon which has been previously reported in many autoimmune conditions [510, 556]. Secondly, significant irregularities in reporting within and between centres may reflect the variability of the historical results caused by differences in local protocols over varying periods of time. Finally, diagnostic thresholds for positivity might have contributed to differences, for example a titter of > 1:80 used to define positivity in NIH registry while currently a titter of 1:160 or above is commonly considered a positive test result.

Whilst autoantibody positivity, particularly ANA antibody positivity, is a frequent finding among patients with I/HPAH and vascular changes develop early in autoimmune diseases (such as SSc) and determine outcomes, there are also other important similarities between autoimmune diseases and PAH such as female predominance and worse outcomes and more aggressive disease in man. Surprisingly, grouping patients by autoantibody positivity, did not show any significant sex bias or differences in survival. In accordance with studies in vasculitis [143], ANCA antibody positivity was associated with significantly worse survival.

In summary, a high missing rate and lack of common protocols across the centres and overall disagreement between in house and reported tests indicate that although investigation of the genetic basis of autoimmune phenomena in PAH is a valid research question, the quality of the available data preclude such analysis.

Table 3.16 Differences between patients with reported positive and negative autoimmune screen. The NBR and Cohort Studies)

	autoimmune screen negative N=779	autoimmune screen positive N=191	*p*-value	N
Diagnosis: n (%)			**0.03**	970
IPAH	678 (87.0)	157 (82.2)		
HPAH	57 (7.32)	11 (5.76)		
PVOD/PCH	19 (2.44)	6 (3.14)		
APAH	12 (1.54)	7 (3.66)		
PH LHD	3 (0.39)	1 (0.52)		
PH LD	6 (0.77)	5 (2.62)		
CTEPH	2 (0.26)	2 (1.05)		
Multifactorial PH	2 (0.26)	2 (1.05)		
Age [years]	49 [35;65]	55 [43;69]	**<0.001**	968
Sex: female, n (%)	526 (68)	124 (65)	0.549	970
Incident cases, n (%)	299 (38.5)	75 (39.5)	0.876	966
WHO FC: I/II/III/IV, %	2/19/6613	2/15/67/16	0.468	955
6MWD [m]	322 [214;413]	335 [212;409]	0.752	645
Shuttle walk distance [m]	332 [198;410]	255 [180;341]	0.112	139
SpO_2 pre [%]	96.0 [93.0;97.0]	96.0 [92.0;97.0]	0.134	814
SpO_2 post [%]	91.5 [85.0;96.0]	90.0 [84.0;95.0]	0.198	760
mRAP [mmHg]	8 [6;12]	8 [5;12]	0.132	907
mPAP [mmHg]	53 [44;61]	52 [43;61]	0.592	935
mPAWP [mmHg]	10 [7;12]	9 [6;11]	**0.043**	863
CO [L/min]	3.8 [3.1;4.8]	3.7 [3.2;4.6]	0.626	920
SvO_2 [%]	64.0 [58.2;70.0]	62.2 [56.0;70.0]	0.158	789
FEV1 [% pred.]	86.0 [74.0;97.4]	85.0 [71.0;100]	0.877	819
FVC [% pred.]	94.8 (19.2)	93.6 (20.9)	0.528	802
TLC [% pred.]	93.3 [84.0;103]	97.0 [85.0;105]	0.348	619
KCO [% pred.]	71 [52;87]	62 [42;79]	**0.002**	655
Emphysema: n (%)			0.265	629
minimal/mild	32 (6)	11 (9)		
moderate	11 (2)	5 (4)		
severe	6 (1)	0 (0)		
Fibrosis: n (%)			**0.012**	630
minimal/mild	20 (4)	9 (7)		
moderate	1 (0)	2 (2)		
severe	0 (0)	1 (1)		
Smoking: past/current n (%)	348 (52.0)	86 (60.6)	0.078	811
NTproBNP [ng/l]	1119 [279;2742]	1429 [497;3019]	0.19	317
BNP [ng/l]	206 [74.1;425]	210 [123;618]	0.23	235
CRP [mg/l]	4.80 [2.00;8.05]	4.00 [2.00;9.00]	0.716	628
Creatinine [μmol/l]	86.0 [71.0;102]	90.5 [72.0;109]	0.051	818
DM type 2	96 (12.3)	31 (16.2)	0.189	970
COPD, n (%)	44 (5.65)	13 (6.81)	0.661	970
OSA, n (%)	35 (4.49)	15 (7.85)	0.089	970
CAD, n (%)	36 (4.62)	6 (3.14)	0.483	970
HTN, n (%)	184 (23.6)	51 (26.7)	0.426	970
Hypothyroidism, n (%)	80 (10.3)	24 (12.6)	0.43	970
Hyperthyroidism, n (%)	10 (1.28)	6 (3.14)	0.104	970
SLE, n (%)	1 (0.13)	1 (0.52)	0.355	970
Systemic sclerosis, n (%)	1 (0.13)	0 (0.00)	1	970
UCTD, n (%)	1 (0.13)	1 (0.52)	0.355	970
Psoriasis, n (%)	12 (1.54)	5 (2.62)	0.352	970
Sjogren syndrome, n (%)	0 (0.00)	2 (1.05)	0.039	970

positivity defined as reported positive on OpenClinica. Abbreviations: I/HPAH: idiopathic/hereditary pulmonary arterial hypertension; PVOD/PCH: pulmonary vasooclusive disease/pulmonary capillary hemangiomatosis; APAH: associated pulmonary arterial hypertension; PH-LHD: pulmonary hypertension due to left heart disease, PH-LD: pulmonary hypertension due to lung disease and/or hypoxia; CTEPH: chronic thromboembolic pulmonary hypertension; KCO: transfer coefficient for carbon monoxide; L/min: litres per minute; mPAP: mean pulmonary arterial pressure; mPAWP: mean pulmonary artery wedge pressure; mRAP: mean right atrial pressure; TLC: total lung capacity; WHO FC: World Health Organization functional class; WU: Wood units; 6MWD: six-minute walk distance; SpO_2: arterial oxygen saturation; CO: cardiac output; COPD: chronic obstructive pulmonary disease; OSA: obstructive sleep apnea; CAD: coronary artery disease; HTN: systemic hypertension; SLE: systemic lupus erythematosus; UCTD: undifferentiated connective tissue disease; DM: diabetes mellitus; CRP: C-reactive protein; NTproBNP: N-terminal pro B-type natriuretic peptide; BNP: B-type natriuretic peptide; SvO_2: mixed venous oxygen saturation; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity

Table 3.17 Comparison between patients with positive and negative in house autoimmune screen. The NBR and Cohort studies.

	negative autoimmune screen N=346	positive autoimmune screen N=127	*p*-value	N
Diagnosis: n (%)			0.292	473
IPAH	295 (85.3)	117 (92.1)		
HPAH	39 (11.3)	7 (5.51)		
PVOD/PCH	9 (2.60)	2 (1.57)		
PH LHD	1 (0.29)	0 (0.00)		
PH LD	2 (0.58)	1 (0.79)		
Age [years]	45 [34;61]	48 [37;59]	0.564	470
Sex: female, n (%)	240 (69)	92 (72)	0.593	473
Incident case, n (%)	109 (31.5)	25 (20.2)	0.022	470
WHO FC: I/II/III/IV, %	1/18/67/14	3/19/66/13	0.828	452
6MWD [m]	320 [204;405]	328 [222;418]	0.516	296
Shuttle walk distance [m]	370 [270;480]	300 [240;400]	0.139	68
SpO_2 pre [%]	96.0 [93.0;97.0]	96.0 [93.0;98.0]	0.492	408
SpO_2 post [%]	92.5 [85.0;96.0]	93.0 [85.0;96.0]	0.681	373
mRAP [mmHg]	8 [5;12]	8 [6;10]	0.261	432
mPAP [mmHg]	53 [44;60]	54 [44;62]	0.525	452
mPAWP [mmHg]	9 [7;12]	9 [7;11]	0.566	402
CO [L/min]	3.7 [3.0;4.8]	4.0 [3.2;4.9]	0.149	435
SvO_2 [%]	64.0 [59.1;70.0]	66.8 [59.1;71.6]	0.062	407
FEV1 [% pred.]	86.0 [75.0;97.3]	87.3 [73.0;97.9]	0.92	411
FVC [% pred.]	95.9 [82.0;107]	95.0 [84.0;107]	0.976	404
TLC [% pred.]	94.1 (13.8)	94.9 (14.9)	0.678	299
KCO [% pred.]	72 [56;89]	78 [65;88]	0.258	360
Emphysema: n (%)			0.603	303
minimal/mild	17 (8)	3 (4)		
moderate	3 (1)	1 (1)		
severe	3 (1)	1 (1)		
Fibrosis: n (%)			0.092	304
minimal/mild	5 (2)	6 (7)		
moderate	1 (0)	1 (1)		
Smoking: past/current, n (%)	173 (52.0)	60 (48.4)	0.567	457
NTproBNP [ng/l]	880 [266;2626]	1400 [320;2246]	0.674	146
BNP [ng/l]	127 [48.1;377]	161 [51.0;422]	0.663	110
CRP [mg/l]	4.00 [2.00;7.00]	4.50 [2.00;7.00]	0.848	333
Creatinine [μmol/l]	86.0 [72.5;101]	85.0 [72.0;101]	0.68	438
DM type 2, n (%)	47 (13.6)	17 (13.4)	1	473
COPD, n (%)	19 (5.49)	6 (4.72)	0.922	473
OSA, n (%)	17 (4.91)	7 (5.51)	0.979	473
CAD, n (%)	8 (2.31)	1 (0.79)	0.456	473
HTN, n (%)	75 (21.7)	27 (21.3)	1	473
Hypothyroidism, n (%)	39 (11.3)	26 (20.5)	**0.015**	473
Hyperthyroidism, n (%)	2 (0.58)	10 (7.87)	**<0.001**	473
SLE, n (%)	0 (0.00)	2 (1.57)	0.072	473
Systemic sclerosis, n (%)	0 (0.00)	0 (0.00)	-	473
UCTD, n (%)	1 (0.29)	0 (0.00)	1	473
Psoriasis, n (%)	9 (2.60)	2 (1.57)	0.735	473
Sjogren syndrome, n (%)	0 (0.00)	1 (0.79)	0.268	473

Abbreviations: I/HPAH: idiopathic/hereditary pulmonary arterial hypertension; PVOD/PCH: pulmonary vasooclusive disease/pulmonary capillary hemangiomatosis; PH-LHD: pulmonary hypertension due to left heart disease, PH-LD: pulmonary hypertension due to lung disease and/or hypoxia; CTEPH: chronic thromboembolic pulmonary hypertension; KCO: transfer coefficient for carbon monoxide; L/min: litres per minute; mPAP: mean pulmonary arterial pressure; mPAWP: mean pulmonary artery wedge pressure; mRAP: mean right atrial pressure; PDE5: phosphodiesterase type 5; PVR: pulmonary vascular resistance; sGC: Soluble Guanylate Cyclase Stimulator; TLC: total lung capacity; WHO FC: World Health Organization functional class; WU: Wood units; 6MWD: six-minute walk distance; SpO_2: arterial oxygen saturation; CO: cardiac output; COPD: chronic obstructive pulmonary disease; OSA: obstructive sleep apnea; CAD: coronary artery disease; HTN: systemic hypertension; SLE: systemic lupus erythematosus; UCTD: undifferentiated connective tissue disease; DM: diabetes mellitus; CRP: C-reactive protein; NTproBNP: N-terminal pro B-type natriuretic peptide; BNP: B-type natriuretic peptide

Pulmonary arterial hypertension in patients with small cardiac defects

These results are based on the 15^{th} July 2020 *OpenClinica* data release and consider either all patients or only those diagnosed with I/HPAH and PVOD/PCH as indicated in the analysis from both the NBR and the PAH Cohort studies.

Results

The presence of cardiac defects was assessed in all but one patient (diagnosed with HHT without PH; n = 1393) as a part of a diagnostic work-up and confirmed through diagnosis verification procedures (see Chapter 2.5.2). Ninety-one (6.5%) individuals were found to have small cardiac defects of which 16 were diagnosed with APAH, one with PH-LD and one with CTEPH, the remaining 72 (5.2%) patients had a diagnosis of IPAH/HPAH/PVOD/PCH (Table 3.18). Patients with PFOs were younger at diagnosis, more often prevalent, showed a greater degree of desaturation post-exercise, and were more likely to have COPD. Survival was significantly better in patients with cardiac defects (Fig. 3.15), also after adjusting for age, sex and whether the case was prevalent or incident, this was also reflected in the lower risk profile defined by the Reveal risk score (Table 3.18). The assessment of other congenital abnormalities revealed single cases of Noonan syndrome (harbouring deleterious missense variant in *RAF1*), Klippel-Feil syndrome (KFS) (no known risk gene found) and HHT (harbouring rare deleterious variant in *ENG*) among patients with small interatrial defects, and single cases of Edward syndrome, situs inversus and congenital fusion of radial bones in the group without PFO/small ASD.

Discussion

Congenital cardiac defects are the most common birth defects, affecting around 1% of infants. Both environmental and genetic causes have been identified. While molecular genetics studies found that single-gene mutations were responsible for 3 to 5%, large chromosomal abnormalities/aneuploidy in 8 to 10%, and pathogenic CNVs in 3% to 25% part of a syndrome, and in 3% to 10% among those with an isolated congenital heart defect (HD), the environmental causes are identifiable only in 2% of cases. The unexplained remainder of congenital HD is presumed to be multifactorial [91, 408]. In our cohort, among patients with small interatrial defects, there were single cases of Noonan and KFS. Noonan syndrome presents with congenital heart disease in 75% of cases. The most commonly associated HD are dysplastic pulmonary valve stenosis, ASD, tetralogy of Fallot, atrioventricular septal defect, hypertrophic cardiomyopathy, ventricular septal defect, and patent ductus arteriosus. Other clinical features include short stature, hypertelorism, down slanting palpebral

Table 3.18 Differences between the patients with and without small cardiac defects. The NBR and Cohort studies

	No cardiac defect N=1245	Cardiac defect present N=72	*p*-value	N
Diagnosis: n (%)			1	1317
IPAH	1125 (90.4)	66 (91.7)		
HPAH	86 (6.91)	5 (6.94)		
PVOD/PCH	34 (2.73)	1 (1.39)		
Age [years]	50 [35;65]	44 [35;54]	**0.013**	1311
Sex: female, n (%)	852 (68)	51 (71)	0.775	1316
Incident case, n (%)	420 (34.0)	14 (19.7)	**0.019**	1307
WHO FC: I/II/III/IV, %	2/20/64/13	1/18/68/13	0.982	1273
6MWD [m]	336 [225;420]	330 [207;470]	0.569	846
Shuttle walk distance [m]	325 [188;410]	260 [190;262]	0.213	173
SpO_2 pre [%]	96.0 [93.0;97.8]	96.0 [94.0;97.0]	0.614	1036
SpO_2 post [%]	91.0 [85.0;96.0]	87.0 [80.0;92.5]	0.001	977
mRAP [mmHg]	8 [5;12]	6 [4;11]	0.051	1172
mPAP [mmHg]	53 [44;61]	52 [45;60]	0.981	1243
mPAWP [mmHg]	9 [7;12]	9 [6;12]	0.918	1128
CO [L/min]	3.9 [3.1;4.9]	4.2 [3.3;5.2]	0.335	1193
SvO_2 [%]	64.0 [58.0;70.0]	66.3 [61.3;69.3]	0.058	977
FEV1 [% pred.]	86.4 [74.0;97.8]	85.5 [69.8;97.2]	0.361	1009
FVC [% pred.]	95.0 [82.0;107]	93.0 [82.3;106]	0.7	991
TLC [% pred.]	94.0 [85.0;104]	94.0 [86.2;104]	0.775	763
KCO [% pred.]	71 [50;86]	75 [52;88]	0.549	757
Emphysema: n (%)			0.462	711
minimal/mild	45 (7)	1 (3)		
moderate	17 (3)	2 (6)		
severe	4 (1)	0 (0)		
Fibrosis: n (%)			0.708	714
minimal/mild	30 (4)	2 (6)		
moderate	3 (0)	0 (0)		
Smoking: past/current, n (%)	462 (50.3)	32 (55.2)	0.556	977
NTproBNP [ng/l]	1102 [280;2861]	1613 [813;2579]	0.482	360
BNP [ng/l]	200 [78.2;452]	78.0 [40.0;151]	**0.029**	335
CRP [mg/l]	4.10 [2.00;8.00]	3.95 [2.00;8.50]	0.54	773
Creatinine [μmol/l]	86.0 [70.0;102]	83.0 [70.5;97.0]	0.214	1003
COPD, n (%)	63 (5.06)	9 (12.5)	**0.014**	1317
REVEAL risk score: n (%)			**0.031**	958
low risk	456 (50.2)	31 (63.3)		
moderately high risk	187 (20.6)	13 (26.5)		
high risk	201 (22.1)	4 (8.16)		
very high risk	65 (7.15)	1 (2.04)		
ERS risk score: n (%)			0.497	1307
low risk	228 (18.4)	17 (23.9)		
intermediate risk	939 (76.0)	51 (71.8)		
high risk	69 (5.58)	3 (4.23)		

Abbreviations: I/HPAH: idiopathic/hereditary pulmonary arterial hypertension; PVOD/PCH: pulmonary vasooclusive disease/pulmonary capillary hemangiomatosis; KCO: transfer coefficient for carbon monoxide; mPAP: mean pulmonary arterial pressure; mPAWP: mean pulmonary artery wedge pressure; mRAP: mean right atrial pressure; TLC: total lung capacity; WHO FC: World Health Organization functional class; WU: Wood units; 6MWD: six-minute walk distance; SpO_2: arterial oxygen saturation; CO: cardiac output; COPD: chronic obstructive pulmonary disease; CRP: C-reactive protein; NTproBNP: N-terminal pro B-type natriuretic peptide; BNP: B-type natriuretic peptide; SvO_2: mixed venous oxygen saturation; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; ERS: European Respiratory Society

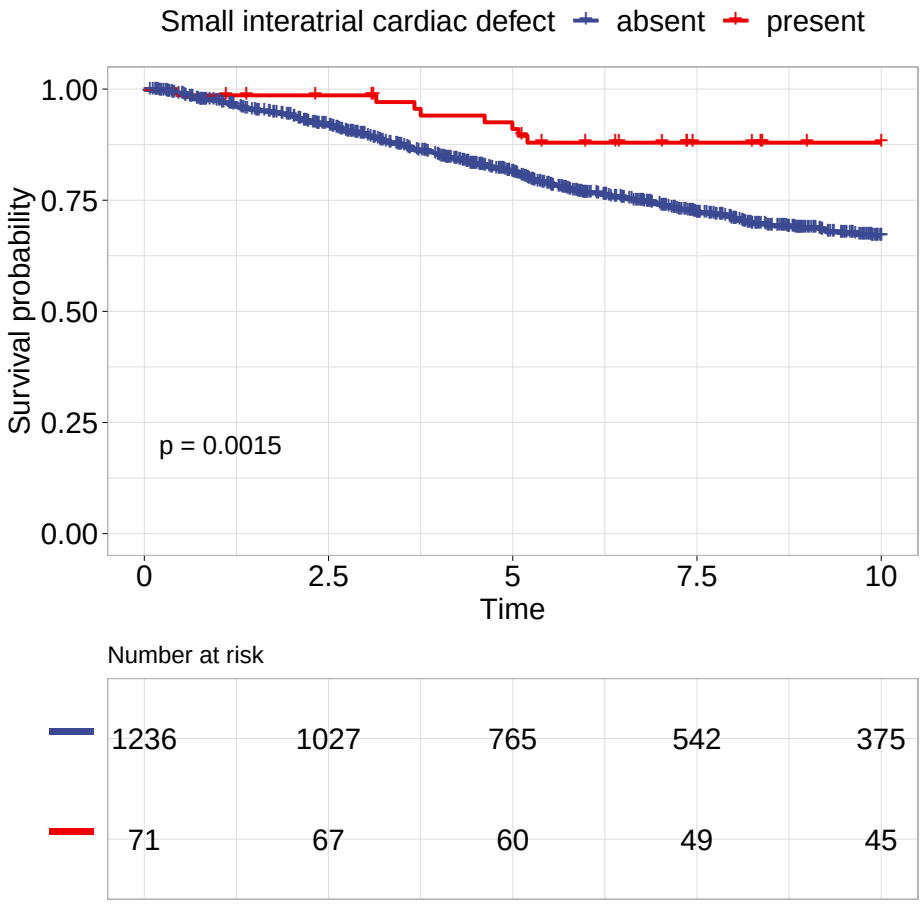


Fig. 3.15 Long-term survival according to the presence or absence of small cardiac atrial defects.

fissures, ptosis, low, posterior hairline, pectus deformity, bleeding disorder, chylothorax, and cryptorchidism. Several loci have been associated with this syndrome *PTPN11*, *SOS1*, *RAF1*, *KRAS*, *NRAS*, *RIT1*, *SHOC2*, *SOS2*, *BRAF* [443], but not with pulmonary hypertension. KFS is another complex entity characterised by cervical vertebral fusion and female predominance. KFS may occur as an isolated abnormality (Klippel-Feil anomaly) or as a syndrome with associated anomalies such as scoliosis, spina bifida occulta, Sprengel's deformity, congenital HD (aortic insufficiency, mitral valve prolapse, mitral insufficiency, coarctation of aorta, total anomalous venous return, secundum ASD), hearing impairment, malformations of the craniofacial area, renal agenesis or horseshoe kidney, webbing of the digits and digital hypoplasia. The most commonly associated KFS driver genes are *GDF3*, *GDF6*, *MEOX1* and *MYO18B* [147]. Our patient with KFS had small ASD, paroxysmal ventricular tachycardia and facial abnormalities requiring nasal bridge reconstruction; no

kidney abnormalities were found on abdominal ultrasound. The genetic analysis did not reveal any variants in known genes associated with KFS. While both Noonan syndrome and KFS are characterised by the presence of congenital HD including PFO/small ASD neither of the syndromes are enriched for patients with pulmonary hypertension when compared to other syndromic or non-syndromic CHD.

One unanticipated finding in the current analysis was that the proportion of patients with small interatrial defects was lower than previously reported in the general population [195, 350] and in patients with I/HPAH [385, 157, 464]. As I/HPAH is the diagnosis of exclusion all patients have detailed imaging studies to exclude HD as well as assessment of arterial oxygen saturation during the RHC in search of saturation step up, hence the low proportion of picked up small interatrial defects is surprising. A possible explanation for these results may be underreporting of benign findings. Similarly to patients with PAH-CHD, our population of patients with small interatrial defects had earlier disease onset and better survival than the remainder of the cohort. These findings suggest that while small interatrial defects may be causally unrelated to PAH they may share the same genetic background.

3.2.2 Computational phenotypes

Clinical phenotype and HPO coding

All available clinical information from both data sets (the NBR/Cohort studies and the US PAHBB) has been annotated with HPO terms. For the terms where modifiers (severity modifiers such as mild, moderate, severe or clinical course modifiers such as acute, subacute, chronic) were required, sham terms were created as descendants of existing terms to accommodate additional clinical information. The median number of HPO terms annotated to cases in the NBR/Cohort studies was 21 [15;27], while patients in the US PAHBB were characterised by a median of 11 [9;14] terms; this difference was statistically significant (p-value < 0.001) (Fig. 3.16).

The most common terms were those describing abnormalities in the cardiovascular system, followed by abnormalities of respiratory, metabolism, musculoskeletal, neurologic, blood and blood-forming organs, endocrine and immune systems; abnormalities in genitourinary, gastrointestinal, ear, eye, and breast were sporadic (Fig. 3.17).

After merging both data sets, the information content (IC) was calculated per term to compare phenotyping depth. The NBR/Cohort studies were characterised by significantly higher additive IC (42.3 [26.2;60.4] vs 16.7 [11.5;24.9], p-value $= 0.001$) as well as higher mean IC (2.02 [1.69;2.35] 1.53 [1.23;1.88], p-value < 0.001) per patient when compared the US PAHBB. Figure 3.17 shows the number of HPO terms per case in A. the NBR/Cohort studies and B. the US PAHBB.

Similarity and clustering strength

The validity of the ontological phenotypes was ascertained by measuring the strength of ontological similarity between cases with known genetic syndromes such as HHT (Fig. A.3). The feasibility of using the ontological similarity for defining new phenotypes for genotype-phenotype association analysis was based on the finding that mutations in some PAH risk genes defined phenotypically similar groups, which were statistically significant (Table 3.19).

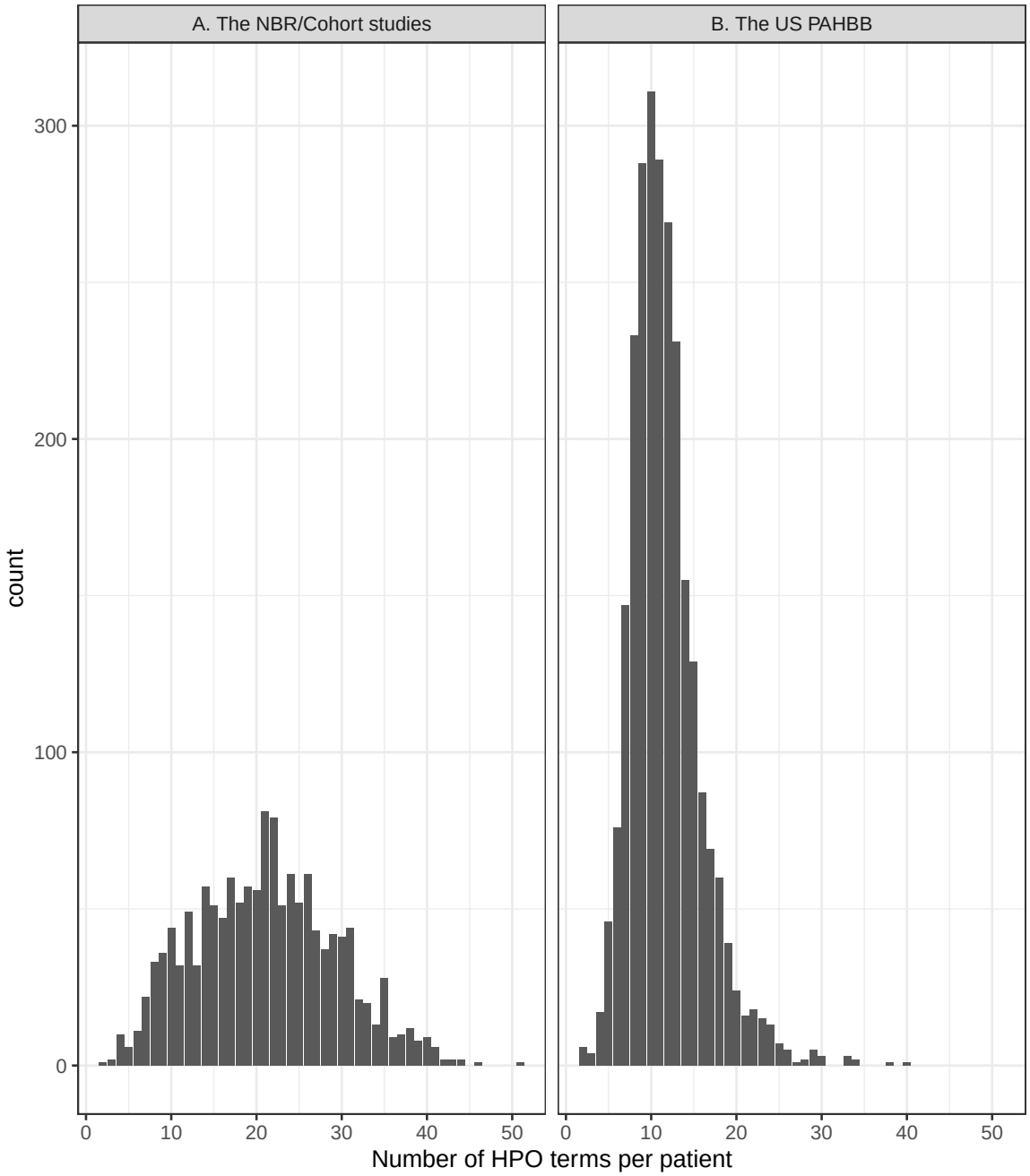
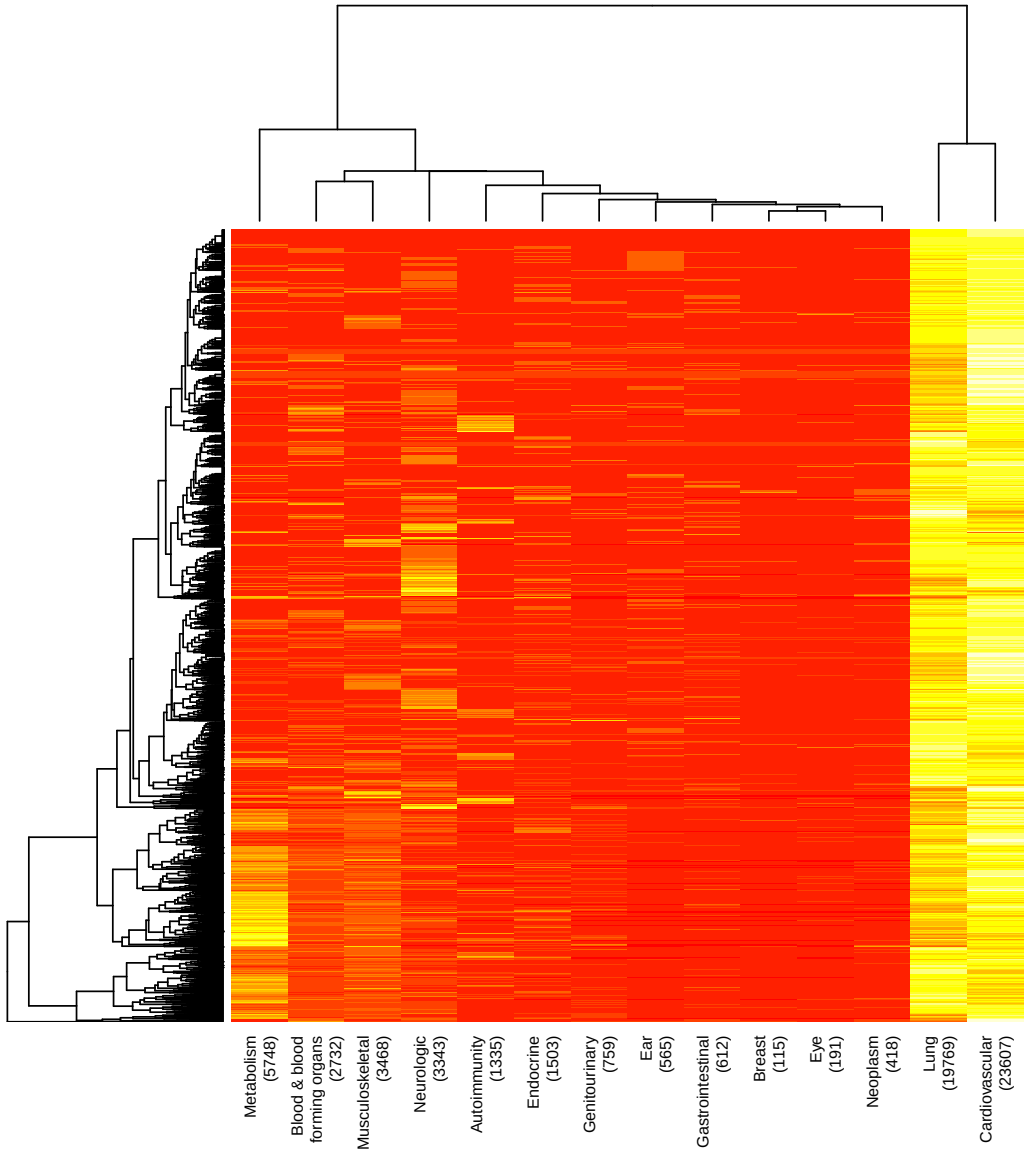


Fig. 3.16 Bar plot indicating the number of human phenotype ontology (HPO) terms annotated to the cases in (A) the NBR/Cohort studies, (B) the US PAHBB study.



Yellow indicates the presence of the term.

Fig. 3.17 Heat map showing cumulative number of term occurrences per body system.

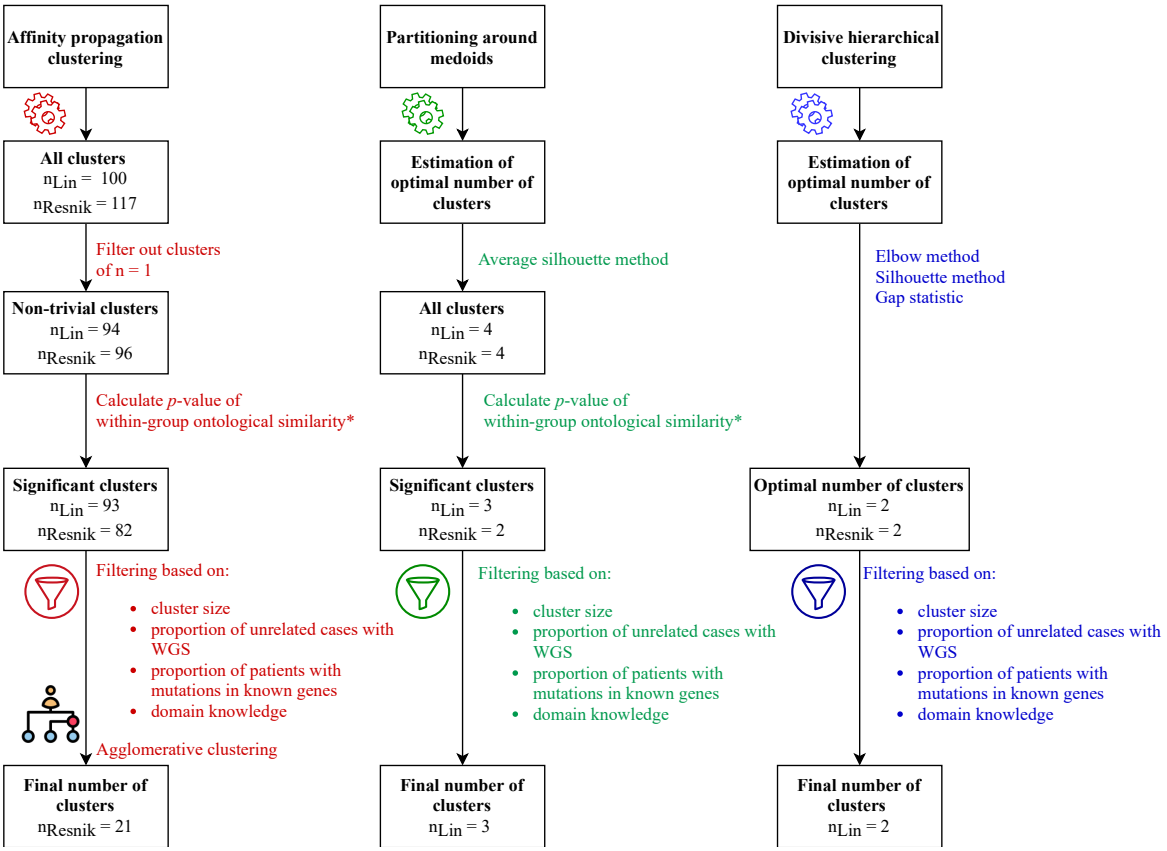
Table 3.19 Strength of ontological similarity in clusters defined by mutation in PAH risk genes or presence of a clinical syndrome with known genetic background (HHT). The NBR and Cohort study.

Cluster name	***p*-value (Lin)**	***p*-value (Resnik)**
Cluster *BMPR2*	**0.0001**	0.5524
Cluster *EIF2AK4*	**0.0095**	**0.0005**
Cluster *SMAD9*	0.1838	0.0511
Cluster *EIF2AK4* comp.het	0.1878	**0.0013**
Cluster *GDF2*	0.3397	0.7493
Cluster *SOX17*	0.3676	0.9161
Cluster *ENG*	0.5345	0.7293
Cluster *AQP1*	0.5774	0.8521
Cluster *ACVRL1*	0.8082	**<0.0001**
Cluster *KCNK3*	0.8501	0.6543
Cluster *ATP13A3*	0.8911	0.5285
Cluster *TBX4*	0.9201	0.2917
Cluster HHT	0.0566	**<0.0001**

Abbreviations: *BMPR2*: Bone Morphogenetic Protein Receptor Type 2; *EIF2AK4*: Eukaryotic Translation Initiation Factor 2 Alpha Kinase 4; *SMAD9*: SMAD Family Member 9; *GDF2*: Growth Differentiation Factor 2; *SOX17*: SRY-Box Transcription Factor 17; ENG: Endoglin; *AQP1*: Aquaporin 1 (Colton Blood Group); *ACVRL1*: Activin A Receptor Like Type 1; *KCNK3*: Potassium Two Pore Domain Channel Subfamily K Member 3; *ATP13A3*: ATPase 13A3; *TBX4*: T-Box Transcription Factor 4; HHT: Hereditary Hemorrhagic Telangiectasia

Clustering of cases using HPO

Three different machine learning methods were employed to devise clusters in the NBR/Cohort studies and the combined data set. The selected clusters from the NBR/Cohort studies were then taken forward for phenotype-genotype association analysis, while clusters devised from the combined data set were planned to be used for the combined UK-US analysis. I selected clusters based on the following criteria: significance of within-group ontological similarity, cluster size, the proportion of unrelated cases who were whole-genome sequenced, and proportion of patients with mutations in known genes in the given cluster (as these patients would not be tested for mutations in other genes). The final decision regarding cluster inclusion rested on the prior probability of cluster characteristics resulting from a single gene mutation (Fig. 3.18).



*significant *p*-value indicates that individuals within the cluster were more similar to each other than 99.99% of random groups of the same size; Non-trivial clusters - clusters that consist of more than one case. Average silhouette method - a high average silhouette width indicates good clustering, which means that the optimal number of clusters k is the one that maximises the average silhouette over a range of possible values for k [260].

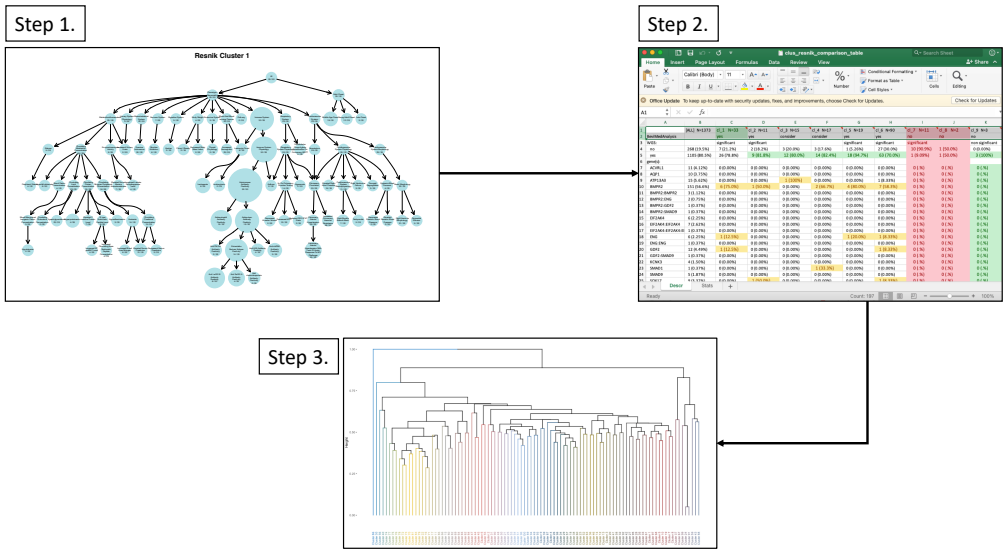
Fig. 3.18 Machine learning methods used to discern ontological phenotypes in the NBR/Cohort studies.

The NBR/Cohort studies

The computational phenotyping pipeline and the results of computational clustering analysis are presented in Fig. 3.18. Three machine learning algorithms were deployed, and the analysis was performed on both Resnik and Lin similarity matrices.

Affinity propagation clustering

When affinity propagation clustering was performed on a Lin similarity matrix it produced 100 clusters of which there were 94 non-trivial. Of those all but one were statistically significant (see Chapter 2.8, Fig. 3.18). Additionally, exemplar based agglomerative clustering was performed to combine small phenotypically similar clusters. The same analysis on a Resnik similarity matrix produced 117 clusters of which 96 were non-trivial and 82 were statistically significant. After careful filtering and visual inspection of clusters, 21 clusters based on Resnik similarity were selected for further analysis (Fig. 3.19). Interestingly, while many of the clusters selected were driven by features that have already been of interest in clinical phenotyping such as autoimmunity (Clusters 1 and 27), small interatrial defects (Cluster 33), age of onset (Clusters 61), and positive acute NO challenge test (Cluster 50) there were also other less studied features such as stimulant addiction (Cluster 86) (Fig. A.4) or blood neoplasms (Cluster 24). Clusters' characteristics are shown in Table B.23.



Step 1. Visual assessment of clusters, Step 2. Filtering based on significance, number of unrelated cases with WGS data and other metrics, Step 3. Exemplar based agglomerative clustering to increase size of the clusters.

Fig. 3.19 Affinity propagation clustering and cluster selection.

Partitioning around medoids (PAM)

The PAM algorithm showed the best results when the cohort was partitioned into four clusters using Lin semantic similarity (Fig. 3.20 A). Of these clusters three proved to be statistically significant (PAM 1 p-value < 0.001, PAM 2 p-value < 0.001, PAM 3 p-value $= 1$, PAM4 p-value < 0.001) and were included in the genotype-phenotype association study. Cluster characteristics are shown in Table 3.20. In brief, all clusters were represented by at least 76% of individuals who were whole-genome sequenced. While clusters 1 and 4 consisted of over 23% of patients with mutations in known PAH risk genes, only 10% of patients in cluster 2 had a genetic diagnosis. Demographics differed significantly between the clusters, whilst cluster 2 had the lowest proportion of females (59%) and the highest median age 68 [63;74], cluster 4 had the highest proportion of ethnic minorities (53.5%). Moreover, cluster 2 was characterised by the worst functional status and exercise performance and significant desaturation on exertion. While patients in clusters 2 and 4 had significantly reduced KCO, only patients in cluster 2 had parenchymal changes on CT imaging. Patients in cluster 2 had significantly lower survival when compared to other clusters (Fig. 3.20 B).

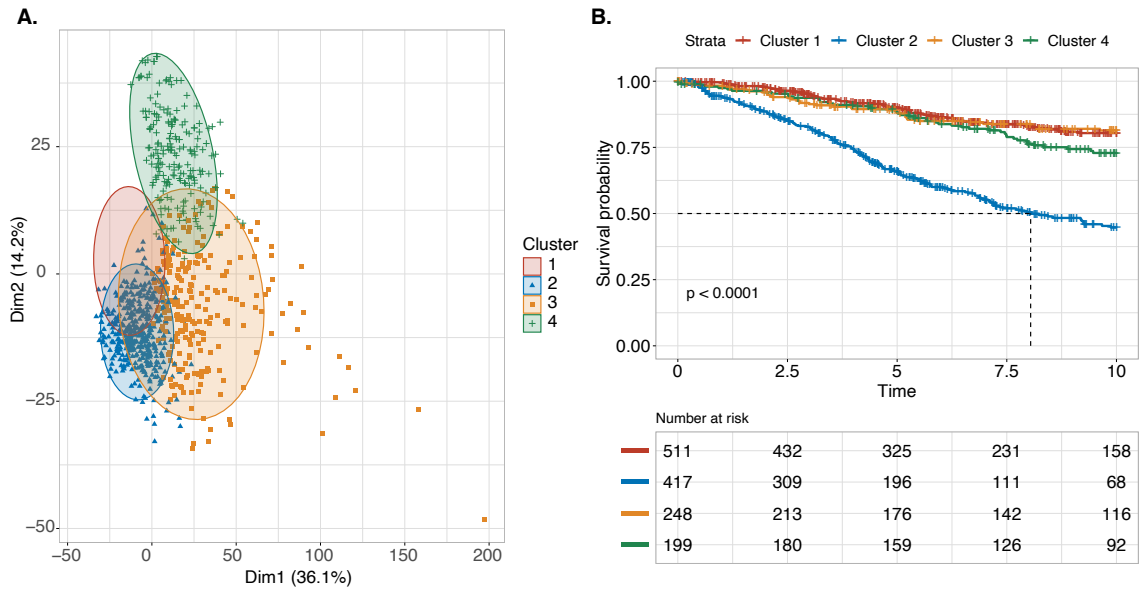


Fig. 3.20 Results of the partitioning around medoids clustering on the similarity matrix derived using Lin method. (A) Two-dimensional projection of four clusters, of which three were statistically significant (cluster 1, cluster 2 and cluster 4), (B) Kaplan-Meier survival curves by cluster.

Table 3.20 Clinical characteristics of patients in four clusters devised by partitioning around medoids algorithm (PAM). The NBR and Cohort studies.

	PAM 1 N=514	**PAM 2 N=415**	**PAM 3* N=263**	**PAM 4 N=202**	***p*-value**	**N**
WGS, n (%)	405 (78.8)	317 (76.4)	223 (84.8)	177 (87.6)	**0.002**	1394
Mutations in PAH risk genes, n (%)	125 (24.3)	42 (10.1)	55 (20.9)	47 (23.3)	**<0.001**	1394
Diagnosis: n (%)					**0.001**	1394
IPAH	450 (87.5)	362 (87.2)	199 (75.7)	180 (89.1)		
HPAH	42 (8.2)	13 (3.1)	20 (7.6)	16 (7.9)		
PVOD/PCH	6 (1.2)	22 (5.3)	6 (2.3)	1 (0.5)		
APAH	13 (2.5)	5 (1.2)	15 (5.7)	2 (1.0)		
PH LHD	0 (0.0)	1 (0.2)	8 (3.0)	0 (0.0)		
PH LD	1 (0.2)	10 (2.4)	5 (1.9)	1 (0.5)		
CTEPH	2 (0.4)	1 (0.2)	3 (1.1)	2 (1.0)		
Multifactorial PH	0 (0.0)	1 (0.2)	6 (2.3)	0 (0.0)		
HHT	0 (0.0)	0 (0.0)	1 (0.4)	0 (0.0)		
Age[years]	42 [33;51]	68 [63;74]	44 [30;55]	41 [31;56]	**<0.001**	1379
Sex: female, n (%)	371 (72)	245 (59)	188 (74)	140 (69)	**<0.001**	1386
Ethnicity: European, n (%)	410 (79.8)	356 (85.8)	198 (75.3)	94 (46.5)	**<0.001**	1394
Drug exposure: n (%)	22 (4.28)	13 (3.13)	11 (4.18)	23 (11.4)	**<0.001**	1394
BMI [kg/m²]	27.4 [23.2;32.3]	28.5 [25.0;32.6]	25.7 [22.1;30.0]	24.1 [21.6;28.6]	**<0.001**	1258
WHO FC: I/II/III/IV, %	1/21/65/13	0/11/74/15	6/24/58/12	2/31/56/11	**<0.001**	1339
6MWD [m]	360 [276;430]	235 [124;335]	374 [262;470]	380 [310;450]	**<0.001**	895
Shuttle walk distance [m]	340 [205;420]	250 [120;376]	305 [218;426]	386 [240;508]	0.068	180
SpO_2 pre [%]	96.0 [94.0;98.0]	94.0 [90.0;96.0]	96.0 [94.0;98.0]	96.0 [94.0;98.0]	**<0.001**	1089
SpO_2 post [%]	93.0 [87.0;96.0]	88.0 [83.0;93.0]	91.0 [84.0;96.0]	91.0 [88.0;95.0]	**<0.001**	1028
mRAP [mmHg]	9 [6;12]	8 [5;12]	7 [4;10]	7 [4;11]	**<0.001**	1234
mPAP [mmHg]	55 [47;63]	48 [40;56]	52 [42;59]	54 [47;64]	**<0.001**	1308
mPAWP [mmHg]	9 [7;12]	10 [7;12]	9 [5;13]	8 [6;11]	**0.003**	1180
CO [L/min]	3.8 [3.0;4.9]	3.8 [3.1;4.7]	4.0 [3.2;5.0]	4.2 [3.5;5.0]	**0.004**	1252
SvO_2 [%]	65.0 [58.5;71.0]	62.7 [57.5;68.0]	67.0 [60.2;71.2]	64.0 [56.5;70.0]	**<0.001**	1035
FEV1 [% pred.]	86.0 [76.0;97.0]	83.0 [69.6;96.5]	86.0 [71.0;97.0]	89.0 [77.5;100]	**0.041**	1063
FVC [% pred.]	95.0 [82.4;107]	94.7 [80.0;106]	95.8 [79.3;106]	91.0 [78.0;103]	0.268	1045
TLC [% pred.]	95.0 [86.0;105]	93.0 [83.1;102]	97.0 [83.5;106]	95.0 [85.0;104]	0.06	809
KCO [% pred.]	77 [64;90]	57 [39;76]	78 [58;93]	58 [45;69]	**<0.001**	806
Emphysema: n (%)					**<0.001**	752
minimal/mild	13 (3.98)	31 (12.0)	4 (3.15)	0 (0.00)		
moderate	3 (0.92)	14 (5.41)	5 (3.94)	0 (0.00)		
severe	0 (0.00)	7 (2.70)	0 (0.00)	0 (0.00)		
Fibrosis: n (%)					**<0.001**	755
minimal/mild	5 (1.52)	26 (10.0)	4 (3.12)	0 (0.00)		
moderate	1 (0.30)	2 (0.77)	0 (0.00)	0 (0.00)		
severe	0 (0.00)	0 (0.00)	1 (0.78)	0 (0.00)		
NTproBNP [ng/l]	830 [280;1810]	2049 [488;4092]	467 [123;1233]	343 [204;392]	**<0.001**	380
BNP [ng/l]	162 [51.5;415]	249 [92.4;526]	120 [40.0;222]	205 [92.0;448]	**0.003**	357
Uric acid [mmol/l]	0.39 [0.30;0.48]	0.46 [0.35;0.56]	0.33 [0.19;0.44]	0.43 [0.28;0.50]	**<0.001**	442
Creatinine [μmol/l]	82.0 [69.0;95.0]	96.0 [78.5;118]	76.0 [65.0;90.5]	82.0 [65.0;89.0]	**<0.001**	1061

*non significant, Abbreviations: I/HPAH: idiopathic/hereditary pulmonary arterial hypertension; PVOD/PCH: pulmonary vaso-occlusive disease/pulmonary capillary hemangiomatosis; PH-LHD: pulmonary hypertension due to left heart disease; PH-LD: pulmonary hypertension due to lung disease or hypoxia; CTEPH: chronic thromboembolic pulmonary hypertension; HHT: hereditary hemorrhagic telangiectasia; KCO: transfer coefficient for carbon monoxide; mPAP: mean pulmonary arterial pressure; mPAWP: mean pulmonary artery wedge pressure; mRAP: mean right atrial pressure; TLC: total lung capacity; WHO FC: World Health Organization functional class; WU: Wood units; 6MWD: six-minute walk distance; SpO_2: arterial oxygen saturation; CO: cardiac output; NTproBNP: N-terminal pro B-type natriuretic peptide; BNP: B-type natriuretic peptide; WGS: whole-exome sequencing; APAH: associated pulmonary arterial hypertension; BMI: body mass index; SvO_2: mixed venous oxygen saturation; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity

Hierarchical clustering

Similarly to PAM clustering, hierarchical clustering on the Lin semantic similarity matrix yielded more clinically distinct clusters, which are presented here. For agglomerative hierarchical clustering, I used four different linkage functions ("single", "average", "complete", and "Ward's method") and compared the results. The cophenetic correlation coefficient was the highest (0.43) for the "average" method, while agglomerative coefficients were the highest for "single" (0.77) and "complete" (0.61) methods (Table B.24). The optimal number of clusters assessed by direct and statistical methods pointed towards 2 clusters (Fig. A.5), but cutting the tree at two clusters level produced one singleton and a second large cluster, which were not informative (this was irrespective of clustering method). I compared the dendrograms produced by various methods and found a very poor correlation between them (Fig. A.6). I then performed divisive hierarchical clustering. The divisive coefficient was 0.55, and cutting the tree at two, the level of two clusters produced groups of 887 and 507 cases with distinct diagnosis profiles, functional status, KCO measurements and small but statistically significant difference in mPAP (Table 3.21). Moreover, patients in cluster 2 had significantly worse survival than those in cluster 1 (p-value = 0.03), this difference remained significant also after adjusting for sex and whether the case was prevalent or incident (Fig. 3.21).

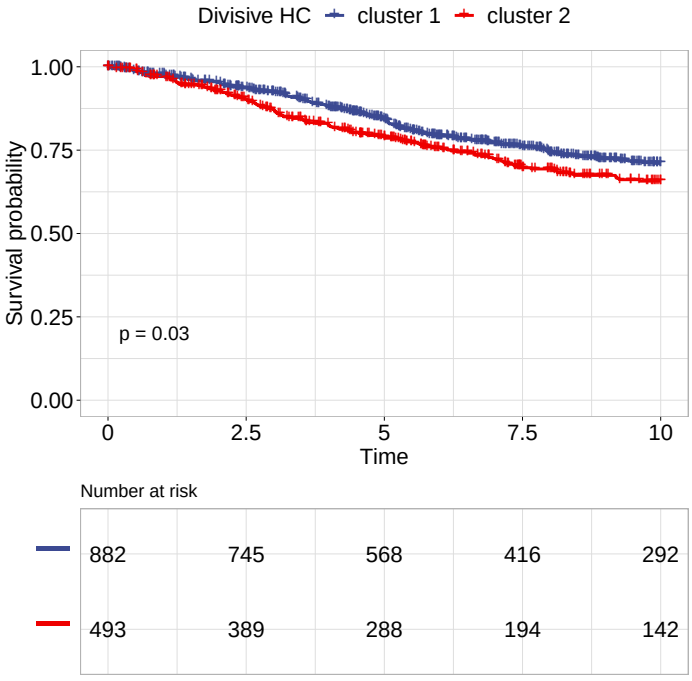


Fig. 3.21 Kaplan-Meier survival curves by clusters devised by divisive hierarchical clustering.

Table 3.21 Clinical characteristics of clusters produced by divisive hierarchical clustering on Lin similarity matrix. The NBR and Cohort studies.

	Divisive HC 1. N=887	**Divisive HC 2. N=507**	***p*-value**	**N**
Diagnosis: n (%)			**<0.001**	1394
IPAH	767 (86.5)	424 (83.6)		
HPAH	68 (7.7)	23 (4.5)		
PVOD/PCH	7 (0.8)	28 (5.5)		
APAH	25 (2.8)	10 (2.0)		
PH-LHD	1 (0.1)	8 (1.6)		
PH-LD	13 (1.5)	4 (0.8)		
CTEPH	6 (0.7)	2 (0.4)		
Multifactorial PH	0 (0.0)	7 (1.4)		
HHT	0 (0.0)	1 (0.2)		
Age [years]	50 [35;63]	49 [35;66]	0.634	1379
Sex: female, n (%)	598 (68)	346 (69)	0.608	1386
Ethnicity: European, n (%)	691 (77.9)	367 (72.4)	**0.024**	1394
Drug exposure: n (%)	51 (5.75)	18 (3.55)	0.09	1394
BMI [kg/m^2]	27.2 [23.4;31.8]	26.5 [22.7;31.2]	0.058	1258
WHO FC: I/II/III/IV, %	1/18/69/13	4/24/59/13	**<0.001**	1339
6MWD [m]	332 [220;416]	340 [221;420]	0.804	895
Shuttle walk distance [m]	325 [192;400]	300 [125;421]	0.685	180
SpO_2 pre [%]	96.0 [93.0;97.0]	96.0 [93.0;98.0]	0.211	1089
SpO_2 post [%]	91.0 [85.0;95.0]	91.0 [85.0;96.0]	0.627	1028
mRAP [mmHg]	8 [5;12]	8 [5;12]	0.592	1234
mPAP [mmHg]	53 [44;62]	51 [42;60]	**0.003**	1308
mPAWP [mmHg]	9 [7;12]	9 [7;12]	0.376	1180
CO [L/min]	3.9 [3.1;4.9]	3.9 [3.1;5.0]	0.643	1252
SvO_2 [%]	64.0 [58.0;70.0]	64.2 [58.2;70.0]	0.544	1035
FEV1 [% pred.]	86.8 [74.0;97.0]	85.0 [71.0;97.2]	0.196	1063
FVC [% pred.]	95.0 [81.0;107]	94.0 [81.0;104]	0.299	1045
TLC [% pred.]	95.0 [85.0;104]	93.0 [83.6;103]	0.243	809
KCO [% pred.]	72 [53;87]	69 [44;82]	**0.011**	806
Emphysema: n (%)			0.598	752
minimal/mild	34 (6.92)	14 (5.36)		
moderate	12 (2.44)	10 (3.83)		
severe	5 (1.02)	2 (0.77)		
Fibrosis: n (%)			0.144	755
minimal/mild	20 (4.04)	15 (5.77)		
moderate	1 (0.20)	2 (0.77)		
severe	0 (0.00)	1 (0.38)		
Thromboembolic disease: n (%)	9 (2)	4 (2)	0.944	769
NTproBNP [ng/l]	1048 [281;2495]	1251 [280;3210]	0.159	380
BNP [ng/l]	191 [73.1;454]	195 [65.0;420]	0.693	357
Uric acid [mmol/l]	0.41 [0.31;0.52]	0.40 [0.30;0.52]	0.992	442
CRP [mg/l]	4.00 [2.00;8.00]	4.50 [2.00;9.00]	0.151	818

Abbreviations: I/HPAH: idiopathic/hereditary pulmonary arterial hypertension; PVOD/PCH: pulmonary vasooclusive disease/pulmonary capillary hemangiomatosis; APAH: associated pulmonary arterial hypertension; PH-LHD: pulmonary hypertension due to left heart disease, PH-LD: pulmonary hypertension due to lung disease and/or hypoxia; CTEPH: chronic thromboembolic pulmonary hypertension; HHT: hereditary hemorrhagic telangiectasia; KCO: transfer coefficient for carbon monoxide; L/min: litres per minute; mPAP: mean pulmonary arterial pressure; mPAWP: mean pulmonary artery wedge pressure; mRAP: mean right atrial pressure; TLC: total lung capacity; WHO FC: World Health Organization functional class; WU: Wood units; 6MWD: six-minute walk distance; SpO_2: arterial oxygen saturation; CO: cardiac output; CRP: C-reactive protein; NTproBNP: N-terminal pro B-type natriuretic peptide; BNP: B-type natriuretic peptide; BMI: body mass index; SvO_2: mixed venous oxygen saturation; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity

Combined analysis

The harmonisation effort proved that combined analysis of the NBR/Cohort studies and the US PAHBB might face problems due to differences in the depth and breadth of phenotypic data collected for both studies. These problems included ascertainment of diagnosis, lack of pulmonary function testing and autoimmune screening results, unavailability of survival data, and the depth of phenotype description (see Chapters 3.1.2 and 3.2.2). For these reasons, I turned to computational phenotyping as a method to discern clusters using all available phenotype information in an unsupervised fashion. Similarly to the previous analysis of the NBR/Cohort study data, I used three clustering methods. The clustering showed a tendency for patients to be grouped by a study rather than by phenotypic clusters, highlighting the problem of variable data availability between the two data sets (Fig. 3.22). After limiting the analysis to HPO terms present in both data sets, the data did not show any clustering tendency.

Study	Cluster 1	Cluster 2	Cluster 3	Cluster 4
The NBR/Cohort	0	31	3	1360
The US PAHBB	649	789	1134	0

Fig. 3.22 The results of partitioning around medoids algorithm performed on a combined data set (The NBR/Cohort studies and the US PAHBB).

3.2.3 Discussion

Pulmonary arterial hypertension is an individually rare but *en masse* common disease characterised by genetic and phenotypic heterogeneity. The clinical complexity, combined with a large number of plausible candidate genes, has been hampering gene discovery. Computational phenotypes derived from biobank data, EHR or harvested from commercial drug studies offer fast and accurate alternatives to clinical phenotyping. Standardised descriptions of phenotypes are at the heart of computational phenotyping, providing means to compare, analyse and exchange clinical data between physicians, sites and studies. The success of such endeavours depends on multiple factors. Firstly, defining variables of interest and the extent of phenotypic information gathered prior to genetic testing (forward phenotyping) and afterwards (reverse phenotyping), and level of detail required from standardised descriptors are of utmost importance. Secondly, improving data capturing mechanisms (using standardised proformas, wearable devices, and continuous invasive monitoring) and sharing phenotypes allows the creation of composite cohorts, crucial in the research of rare diseases. Data fragmentation and harmonisation problems are visible in a combined data set where differences in phenotypic data depth and breadth varied substantially between studies, which significantly limited the analysis.

I used two methods to calculate semantic similarity, namely Resnik and Lin. While Resnik weights rarer terms more strongly and is sensitive to specific pathognomonic terms, Lin is more robust to the background noise. This played a role in the differences in the strength of ontological similarity within the cluster formed by patients harbouring mutations in *ACVRL1*, where Resnik similarity was driven by rare terms describing arteriovenous malformations and telangiectasia and proved to be significant. At the same time, the Lin method did not yield positive results. Whilst well known syndromic diseases such as the HHT showed significant ontological similarity, mutations in only four genes defined significant clusters indicative of high genetic heterogeneity and the shortcomings inherent in the phenotyping of these patients.

All three machine learning algorithms partitioning around medoids, affinity propagation clustering and divisive hierarchical clustering, yielded interesting results.

The PAM algorithm produced three statistically significant clusters with distinct clinical characteristics and outcomes. The clusters also differed in respect to the percentage of cases with mutations in known PAH risk genes. This could have significant implications for the subsequent genotype-phenotype analysis, whereby selecting a cohort (Cluster 2) with largely unknown genetic underpinnings of the disease could increase the power to detect new risk genes. Interestingly, the cluster in question (Cluster 2) overlapped phenotypically with two

clusters devised on a clinical basis, namely age in higher tertile and KCO in lower tertile, indicative of convergence of the results of clinical and computational phenotyping.

Similarly to PAM, affinity propagation clustering identified clusters driven by clinical characteristics of interest in clinical phenotyping such as strong autoimmunity signal, congenital heart abnormalities, age of onset, and new less studied features. The first scenario indicates convergence of clinical and computational approaches and adds value to creating smaller, more compact clusters than those devised by stratification based on a single characteristic. The second scenario points towards less studied features such as abnormalities of blood and blood-forming organs and the cross-talk between environmental factors such as substance abuse and PAH. Finally, hierarchical clustering produced two phenogroups characterised by different diagnosis, functional status and outcomes.

Apart from using computational phenotyping to cluster unrelated patients for genotype-phenotype association analysis in the future, I plan to compute phenotype closeness within the pedigree groups without a genetic diagnosis, who are expected to share a causal gene variant.

3.2.4 Conclusions

My phenotype optimisation approach was aimed at finding homogenous groups of patients that are likely to share the same genetic architecture. To achieve this goal, I used three strategies. Firstly, I ascertained the diagnosis in all patients as described in Chapter 2.5.2 and "tagged" patients with a diagnostic group and subgroup labels. Secondly, I used an extreme phenotype strategy that focuses on identifying rare variants with large effect sizes through the recruitment of patients with traits at either end of the phenotypic spectrum. Through this approach, I chose the following phenotypes: vasoresponders, poor- and super-survivors, positive family history, age at diagnosis and KCO strata. Finally, to account for complex phenotypes, I devised computational phenotypes using unsupervised clustering of ontological phenotypes and chose 26 clusters produced by affinity propagation clustering, PAM and divisive hierarchical clustering. I tagged patients with the above-mentioned phenotypes as cases in genotype-phenotype association analysis described in Chapter 3.3.

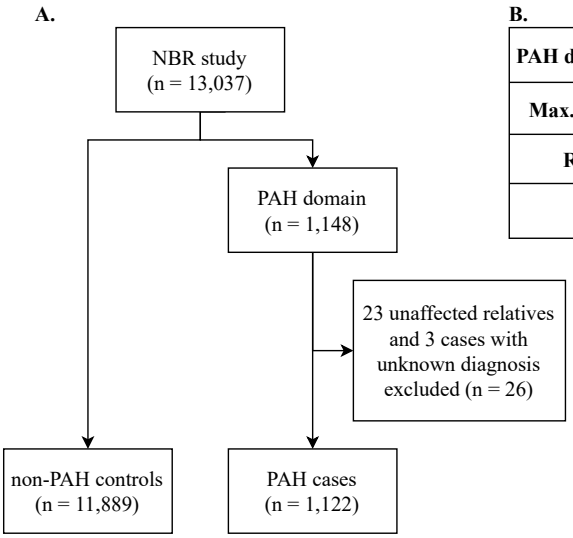
In Chapter 3.2 I addressed aim number 3 of my Thesis.

3.3 Forward genetics

Parts of this chapter have been previously published: Swietlik EM, et.al, Nichols WC, Morrell NW, Gräf S. Bayesian Inference Associates Rare *KDR* Variants with Specific Phenotypes in Pulmonary Arterial Hypertension, Circ Genom Precis Med. 2020 Dec 15;14(1):e003155. PMID: 33320693

3.3.1 Characterisation of study cohorts and tag definitions

WGS was performed in 13,037 participants of the NBR study, of which 1,148 were recruited to the PAH domain [538]. The PAH domain included 23 unaffected parents and three cases with an unknown phenotype, which were removed from the analysis (Fig. 3.23, Table 3.22).



B.

PAH domain samples	Cases	Cases excluded	Unaffected relatives	Total
Max. unrelated set	1,102	3	18	**1,123**
Related set	20	-	5	25
Total	**1,122**	3	23	**1,148**

Fig. 3.23 Consort diagram of the genotype-phenotype study. (A) Case-control study flow chart, (B) Number of PAH domain samples in the analysis broken down by maximal unrelated set and related set, and cases (included and excluded) and unaffected relatives.

Of the remaining 1,122 participants, 972 (86.6%) had a clinical diagnosis of IPAH, 73 (6.5%) of HPAH, and 20 (1.8%) were diagnosed with PVOD/PCH. Diagnosis verification revealed that 57 participants (5%) had a diagnosis other than IPAH, HPAH or PVOD/PCH. These cases were subsequently relabelled and moved to the respective tag group for analysis. A breakdown of the number of individuals by tag is presented in Table B.9. For example, in the case of the PH tag, 1,112 of the 1,122 cases were included in the association testing after excluding nine affected relatives and one case without the tag.

The comprehensive clinical characterisation of the study cohort is shown in Table 3.22. In summary, the median age at diagnosis was 49[35;63] years with a female predominance of 68%. Europeans constituted 84% of the study cohort. Overall survival in the studied population was 97% at one year, 91% at three years and 84% at five years. As expected, there was a significant difference in survival between prevalent and incident cases. In prevalent cases, survival at one, three, and five years was 98%, 93%, and 87%, whereas in incident cases it was 97%, 84%, and 72%, respectively. Median transfer coefficient for carbon monoxide (KCO) in the entire studied population was 71[52;86]% predicted. Based on the previous analysis, I stratified patients by KCO tertiles; the cut off for the lower tertile was the same as the best KCO threshold in survival analysis (60% pred.), and by KCO threshold at 50% as the KCO histogram showed a bimodal distribution with the dip at 50%, which was also in keeping with thresholds reported in the literature [505]. Cases in the lower tertile or below the KCO threshold of 50% predicted were more commonly male, older at diagnosis, had a current or past history of cigarette smoking and an increased number of cardiorespiratory comorbidities (Tables B.25, B.26, and B.27). Survival in these groups was significantly worse than in those with preserved or mildly reduced KCO (Fig. A.7 A-D). After adjusting for confounding factors (age, sex, comorbidities, smoking status and whether the case was prevalent or incident), KCO remained an independent predictor of survival (Table B.28). Age at diagnosis was calculated as age at the time of diagnostic right heart catheterisation and was available in all but 10 cases. Patients in the higher age tertile showed more functional impairment despite milder hemodynamics, lower FEV1/FVC ratio and KCO [% predicted], as well as mild emphysematous and fibrotic changes on CT scans (Table B.29). Patients in higher age tertile had the worst survival (Fig. A.7 E and F). Likewise, clinical characteristics of super- and poor-survivors, vasoresponders, patients with small interatrial defects and computationally derived phenotypes taken forward for genetic analysis were in keeping with the results of the overall analysis (see Chapter 3.2).

3.3.2 Rare variants in previously established genes

We identified single nucleotide variants (SNVs), small insertions and deletions (indels) and larger deletions (Fig. A.8) in previously established genes (namely, *BMPR2*, *ACVRL1*, *ENG*, *SMAD1*, *SMAD4*, *SMAD9*, *KCNK3*, *TBX4*, *EIF2AK4*, *AQP1*, *ATP13A3*, *GDF2*, *SOX17*) in 271 (24.2%) of the 1,122 cases recruited to the PAH domain and interpreted them based on the ACMG standards and guidelines [440]. The majority of these variants have already been described by Gräf et al. [179]. We identified 124 (45.8%) and 110 (40.6%) subjects

Table 3.22 Clinical characterization of the NBR study population.

	ALL N = 1122	N	I/HPAH and PVOD/PCH N = 1065	N
		Demographics and functional status		
Sex: female, n (%)	760 (68)	1116	732 (69)	1064
Ethnicity: European, n (%)	945 (84)	1122	898 (85)	
Age [years]	49 [35;63]	1112	49 [35;63]	1061
BMI [kg/m^2]	27 [23;32]	1015	27 [23;31]	970
WHO FC: I/II/III/IV, %	2/20/65/13	1079	2/20/64/13	1029
6MWD [m]	335 [220;415]	953	336 [220;415]	906
		Haemodynamics		
mRAP [mmHg]	8 [5;12]	985	8 [5;12]	939
mPAP [mmHg]	53 [44;61]	1052	53 [44;61]	1004
CO [L/min]	3.9 [3.1;4.9]	1003	3.9 [3.1;4.9]	960
FEV1 [% pred.]	85 [73;97]	849	86 [74;97]	811
FVC [% pred.]	94 [81;106]	831	95 [82;106]	793
KCO [% pred.]	71 [52;86]	644	71 [52;86]	610
		Clinical blood tests		
Hb [g/l]	151 [138;165]	847	152 [138;164]	805
RDW [%]	14 [14;16]	413	14 [14;16]	392
WBC [x10e9/l]	8.2 [6.8;9.8]	839	8.2 [6.8;9.8]	797
Platelets [x10e9/l]	224 [182;272]	836	225 [183;274]	795
Creatinine [μmol/l]	86 [70;102]	832	86 [70;102]	790
NTproBNP [ng/l]	926 [215;2637]	276	963 [217;2672]	265
BNP [ng/l]	195 [72;432]	271	197 [74.6;454]	252
CRP [mg/l]	4 [2;8]	639	4 [2;8]	604
		Comorbidities		
HTN, n (%)	265 (24)	1122	256 (24)	1065
DM type 1, n (%)	20 (2)	1122	19 (2)	1065
DM type 2, n (%)	138 (12)	1122	132 (12)	1065
CAD, n (%)	45 (4)	1122	42 (4)	1065
CVA, n (%)	17 (2)	1122	15 (1)	1065
Hypothyroidism, n (%)	135 (12)	1122	130 (12)	1065
COPD, n (%)	66 (6)	1122	57 (5)	1065
Asthma, n (%)	78 (7)	1122	74 (7)	1065
Cancer, n (%)	4 (0.4)	1122	3 (0.3)	1065
		Medication		
Initial therapy: n (%)		631		604
CCB	73 (12)		72 (12)	
combination therapy	279 (44)		269 (45)	
ERA	86 (14)		82 (14)	
PA	42 (7)		41 (7)	
PDE5 inhibitor	150 (24)		139 (23)	
sGC stimulator	1 (0)		1 (0)	

Entire cohort (n=1122) was composed of IPAH (n = 972), HPAH (n = 73), PVOD/PCH (n = 20), PH associated with left heart disease (n=7), PH associated with lung disease (n = 8), chronic thromboembolic pulmonary hypertension (n = 6), multifactorial PH (n = 6), hereditary hemorrhagic telangiectasia (n=1). Abbreviations: BMI: body mass index, WHO FC: World Health Organization Functional Class, 6MWD: 6-minute walk distance, mRAP: mean right atrial pressure, mPAP: mean pulmonary artery pressure, CO: cardiac output, FEV1: forced expiratory capacity in 1 second, FVC: forced vital capacity, KCO: transfer factor coefficient for carbon monoxide, Hb: haemoglobin, RDW: red cell distribution width, WBC: white blood cell count, NTproBNP: N-terminal pro B-type natriuretic peptide, BNP: B-type natriuretic peptide, CRP: C-Reactive Protein Protein, HTN: hypertension, DM: diabetes mellitus, CAD: coronary artery disease, CVA: cerebrovascular accident, COPD: chronic obstructive pulmonary disease, CCB: calcium channel blocker, ERA: endothelin receptor antagonists, PA: prostacyclin analogues, PED5: phosphodiesterase type 5, sGC: soluble guanylate cyclase.

carrying pathogenic and likely pathogenic variants, respectively, and 37 individuals (13.6%) with variants of uncertain significance. Of these, 161 (59.4% of explained, 14.3% of all cases) carried pathogenic mutations in *BMPR2*, 96 subjects (59.6%) carried loss-of-function variants (including stop gained, frameshift and splice donor/acceptor variants), 27 (16.8%) possessed larger deletions and 38 cases (23.6%) presented with deleterious, missense variants. Twenty-one cases had rare variants in *TBX4* (7.7% of explained, 1.9% of all cases) comprising of either loss-of-function mutations (10 subjects (3.7%)), larger deletions (two cases (0.7%)) or deleterious missense variants (nine individuals (3.3%)). Mutations in *ATP13A3* (15 cases; 5.5% of explained and 1.3% of all cases) comprised loss-of-function mutations (six cases (2.2%)) and missense mutations (nine cases (3.3%), including one homozygous mutation in a child already reported in Barozzi et al. [35]). We identified 14 patients (5.2% of explained, 1.2% of all cases) with biallelic *EIF2AK4* mutations, six homozygous (2.2%) and eight compound heterozygous (3%) individuals. A further 14 cases (5.2% of explained, 1.2% of all cases) carried mutations in *GDF2*, the majority exhibited deleterious missense mutations (10 individuals (3.7%)), two cases (0.7%) harboured large deletions and another two cases (0.7%) had loss-of-function variants which co-occurred with a large deletion in *BMPR2* and a heterozygous missense variant in *SMAD9* respectively. *ACVRL1* mutations were detected in 11 cases (4.1% of explained, 1% of all cases), with the majority being heterozygous missense variants (10 cases (3.7%)). Only one individual (0.4%) had a loss-of-function variant in *ACVRL1*. Heterozygous missense variants in *AQP1* were revealed in nine subjects (3.3% of explained, 0.8% of all cases), of which one was accompanied with a heterozygous missense variant in *BMPR2*. An additional case presented with a loss-of-function variant in *AQP1*. In *ENG* we predominantly identified heterozygous missense variants in eight cases (3% of explained, 0.7% of all cases), of which two co-occurred with a large deletion in *BMPR2*, and a loss-of-function variant in one case (0.4% of explained, 0.1% of all cases). Another nine cases (3.3% of explained, 0.8% of all cases) harboured variants in *SOX17*, three with loss-of-function and six with heterozygous missense variants. *SMAD9* variants were found in eight cases (3% of explained, 0.2% of all cases), six had heterozygous missense variants, of which two co-occurred with loss-of-function variants in *BMPR2* and *GDF2*, and two possessed heterozygous missense mutations, of which one was accompanied by a heterozygous missense variant. A further four patients (1.5% of explained, 0.4% of all cases) exhibited deleterious missense (three cases (1.1%)) and loss-of-function (one case (0.4%)) variants in *KCNK3*. Finally, two cases (0.7% of explained, 0.2% of all cases) harboured a loss-of-function variant in *SMAD1*, of which one co-existed with a loss-of-function variant in *BMPR2*.

3.3.3 Rare variant association testing

I used BeviMed to consolidate previously reported PAH genes and to discover novel genotype-phenotype associations. Of note, cases explained by rare deleterious variants in previously established genes were only included for the association testing with the respective disease gene (see Chapter 2.9.3). This analysis identified 59 significant gene-tag associations with posterior probability (PP) above 0.75 (when prior set to 0.001; Table 3.23 and Fig. 3.24 A and B). *BMPR2*, *TBX4*, *EIF2AK4*, *ACVRL1* and *AQP1* showed the highest association ($PP \geq 0.99$); I also confirmed significant associations in the majority of other previously identified genes. Individuals with rare variants in *BMPR2*, *TBX4* (high impact), *EIF2AK4* (biallelic) and *SOX17* had a significantly younger age of disease onset (tag: young age). I also confirmed the association of rare variants in *AQP1* with FPAH ($\log(BF) = 10.023$, $PP = 0.958$). The refined phenotype approach corroborated the association between high impact variants in *BMPR2* and preserved KCO (KCO higher tertile, $\log(BF) = 99.923$, $PP = 1$) together with an association of biallelic *EIF2AK4* mutations with significantly reduced KCO (KCO < 50% predicted, $\log(BF) = 29.741$, $PP = 1$). Phenotypes "poor" and "super survivors" did not yield any new genetic discoveries, but confirmed poor outcomes in *EIF2AK4* mutation carriers ($\log(BF) = 16.494$, $PP = 0.999$) and long survival among patients harbouring mutations in *TBX4* ($\log(BF) = 9.877$, $PP = 0.951$). No significant associations were found with groups characterised by positive NO challenge response or presence of small interatrial defects.

Under an autosomal dominant mode of inheritance, high impact variants in the Kinase Insert Domain Receptor (*KDR*) were associated with a significantly reduced KCO (tag: KCO lower tertile, $\log(BF) = 11.362$, $PP = 0.989$), older age at diagnosis (tag: old age, $\log(BF) = 9.249$, $PP = 0.912$) and PAM 2 (Lin) cluster (tag: PAM2, $\log(BF) = 8.048$, $PP = 0.758$). Of note, when lower KCO tertile was constructed based on imputed KCO data the association remained significant (tag: imputed KCO lower tertile, $\log(BF) = 8.613$, $PP = 0.846$). Moderate and high impact variants in *COL6A5* were strongly associated with divisive hierarchical clustering cluster 2 (divisive HC 2) (tag: Divisive HC 2, $\log(BF) = 10.627$, $PP = 0.976$) under autosomal dominant mode of inheritance. Computationally derived phenotypes resulted in the discovery of three additional new candidate genes (*OR6T1* - Olfactory Receptor Family 6 Subfamily T Member 1, *EYS* - Eyes Shut Homolog, *HPSE2* - Heparanase 2 (Inactive)), albeit at lower significance levels (Table 3.23)

Table 3.23 BeviMed analysis result comparison.

Gene	Phenotype	log(BF)	PP, $\pi = 0.01$	PP, $\pi = 0.001$	PP, $\pi = 0.0003$	Variant impact	MOI
			Clinical phenotypes				
BMPR2	I/HPAH	265.762	1	1	1	High	dominant
BMPR2	PAH	265.639	1	1	1	High	dominant
BMPR2	I/HPAH/PVOD/PCH	263.481	1	1	1	High	dominant
BMPR2	PH	262.625	1	1	1	High	dominant
BMPR2	young age	149.576	1	1	1	Moderate and high	dominant
BMPR2	HPAH	149.091	1	1	1	Moderate and high	dominant
BMPR2	FPAH	147.822	1	1	1	Moderate and high	dominant
BMPR2	IPAH	144.582	1	1	1	High	dominant
BMPR2	KCO higher tertile	99.923	1	1	1	High	dominant
BMPR2	middle age	63.119	1	1	1	Moderate and high	dominant
BMPR2	KCO middle tertile	52.706	1	1	1	Moderate and high	dominant
EIF2AK4	low KCO	29.741	1	1	1	Moderate and high	recessive
EIF2AK4	KCO lower tertile	26.247	1	1	1	Moderate and high	recessive
TBX4	I/HPAH	23.783	1	1	1	High	dominant
TBX4	I/HPAH/PVOD/PCH	23.549	1	1	1	High	dominant
TBX4	PAH	23.141	1	1	1	High	dominant
TBX4	PH	22.877	1	1	1	High	dominant
EIF2AK4	young age	20.547	1	1	1	Moderate and high	recessive
TBX4	IPAH	19.99	1	1	1	High	dominant
EIF2AK4	I/HPAH/PVOD/PCH	15.718	1	1	1	Moderate and high	recessive
ACVRL1	HPAH	15.501	1	1	0.999	Moderate and high	dominant
EIF2AK4	PAH	15.407	1	1	0.999	Moderate and high	recessive
EIF2AK4	PH	15.071	1	1	0.999	Moderate and high	recessive
EIF2AK4	PVOD/PCH	14.441	1	0.999	0.998	Moderate and high	recessive
AQP1	HPAH	12.075	0.999	0.994	0.983	Moderate	dominant
EIF2AK4	FPAH	11.858	0.999	0.993	0.979	High	recessive
TBX4	young age	11.5	0.999	0.99	0.97	High	dominant
AQP1	I/HPAH	11.466	0.999	0.99	0.969	Moderate and high	dominant
KDR	KCO lower tertile	11.362	0.999	0.989	0.966	High	dominant
AQP1	I/HPAH/PVOD/PCH	11.291	0.999	0.988	0.964	Moderate and high	dominant
AQP1	PAH	11.047	0.998	0.984	0.954	Moderate and high	dominant
AQP1	PH	10.791	0.998	0.98	0.941	Moderate and high	dominant
AQP1	FPAH	10.023	0.996	0.958	0.881	Moderate	dominant
KDR	old age	9.249	0.991	0.912	0.774	High	dominant
GDF2	I/HPAH	9.091	0.989	0.899	0.746	Moderate and high	dominant
BMPR2	old age	8.913	0.987	0.881	0.71	High	dominant
GDF2	I/HPAH/PVOD/PCH	8.775	0.985	0.866	0.681	Moderate and high	dominant
SOX17	young age	8.554	0.981	0.839	0.631	Moderate and high	dominant
GDF2	PAH	8.478	0.98	0.828	0.614	Moderate and high	dominant
ATP13A3	KCO higher tertile	8.035	0.969	0.755	0.505	High	dominant
			Computational phenotypes				
BMPR2	Divisive HC 1 (Lin)	160.425	1	1	1	Moderate and high	dominant
BMPR2	PAM 1 (Lin)	138.955	1	1	1	High	dominant
BMPR2	Divisive HC 2 (Lin)	72.823	1	1	1	High	dominant
BMPR2	PAM 4 (Lin)	44.612	1	1	1	Moderate and high	dominant
BMPR2	PAM 2 (Lin)	40.098	1	1	1	High	dominant
BMPR2	AP 61 (Resnik)	20.787	1	1	1	Moderate and high	dominant
BMPR2	AP 42 (Resnik)	19.731	1	1	1	Moderate and high	dominant

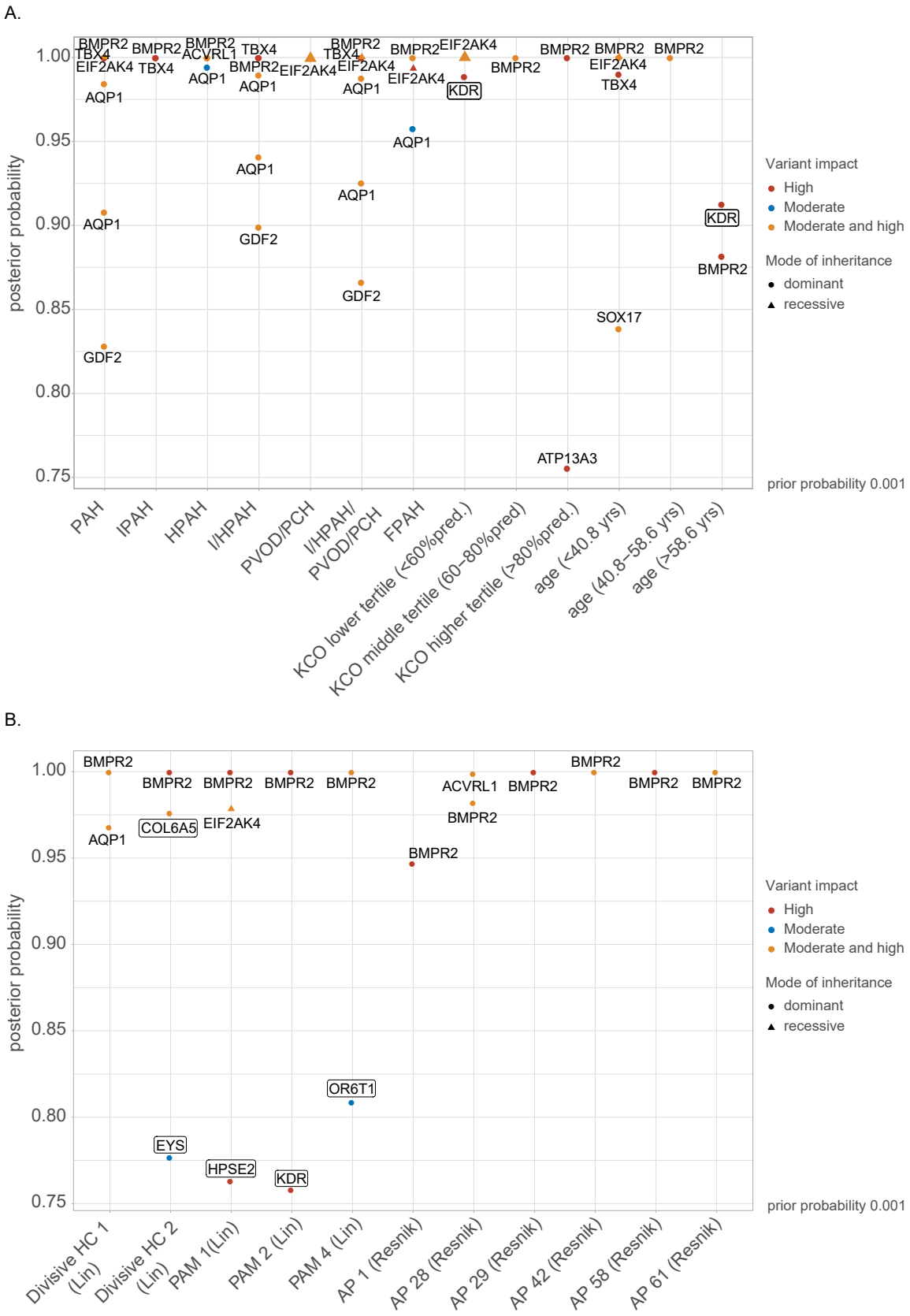
Table 3.23 BeviMed analysis result comparison.

Gene	Phenotype	log(BF)	PP, $\pi = 0.01$	PP, $\pi = 0.001$	PP, $\pi = 0.0003$	Variant impact	MOI
BMPR2	AP 58 (Resnik)	17.098	1	1	1	High	dominant
BMPR2	AP 29 (Resnik)	14.542	1	1	0.999	High	dominant
ACVRL1	AP 28 (Resnik)	14.466	1	0.999	0.998	Moderate and high	dominant
BMPR2	AP 28 (Resnik)	10.926	0.998	0.982	0.948	Moderate and high	dominant
EIF2AK4	PAM 1 (Lin)	10.686	0.998	0.978	0.935	Moderate and high	recessive
COL6A5	Divisive HC 2 (Lin)	10.627	0.998	0.976	0.932	Moderate and high	dominant
AQP1	Divisive HC 1 (Lin)	10.308	0.997	0.968	0.908	Moderate and high	dominant
BMPR2	AP 1 (Resnik)	9.782	0.994	0.947	0.854	High	dominant
OR6T1	PAM 4 (Lin)	8.351	0.977	0.809	0.583	Moderate	dominant
EYS	Divisive HC 2 (Lin)	8.154	0.972	0.777	0.534	Moderate	dominant
HPSE2	PAM 1 (Lin)	8.078	0.97	0.763	0.515	High	dominant
KDR	PAM 2 (Lin)	8.048	0.969	0.758	0.508	High	dominant

Shown are Bayes factors and posterior probabilities for three different priors.
Abbreviations: *BMPR2*: Bone Morphogenetic Protein Receptor Type 2; *EIF2AK4*: Eukaryotic Translation Initiation Factor 2 Alpha Kinase 4; *TBX4*: T-Box Transcription Factor 4; *ACVRL1*:Activin A Receptor Like Type 1; *AQP1*: Aquaporin 1 (Colton Blood Group); *KDR*: Kinase Insert Domain Receptor; *GDF2*: Growth Differentiation Factor 2; *SOX17*: SRY-Box Transcription Factor 17; *ATP13A3*: ATPase 13A3; *COL6A5*: Collagen Type VI Alpha 5 Chain; *RPS11*: Ribosomal Protein S11; *DAPL1*: Death Associated Protein Like 1; *EYS*: Eyes Shut Homolog; *OR6T1*: Olfactory Receptor Family 6 Subfamily T Member 1

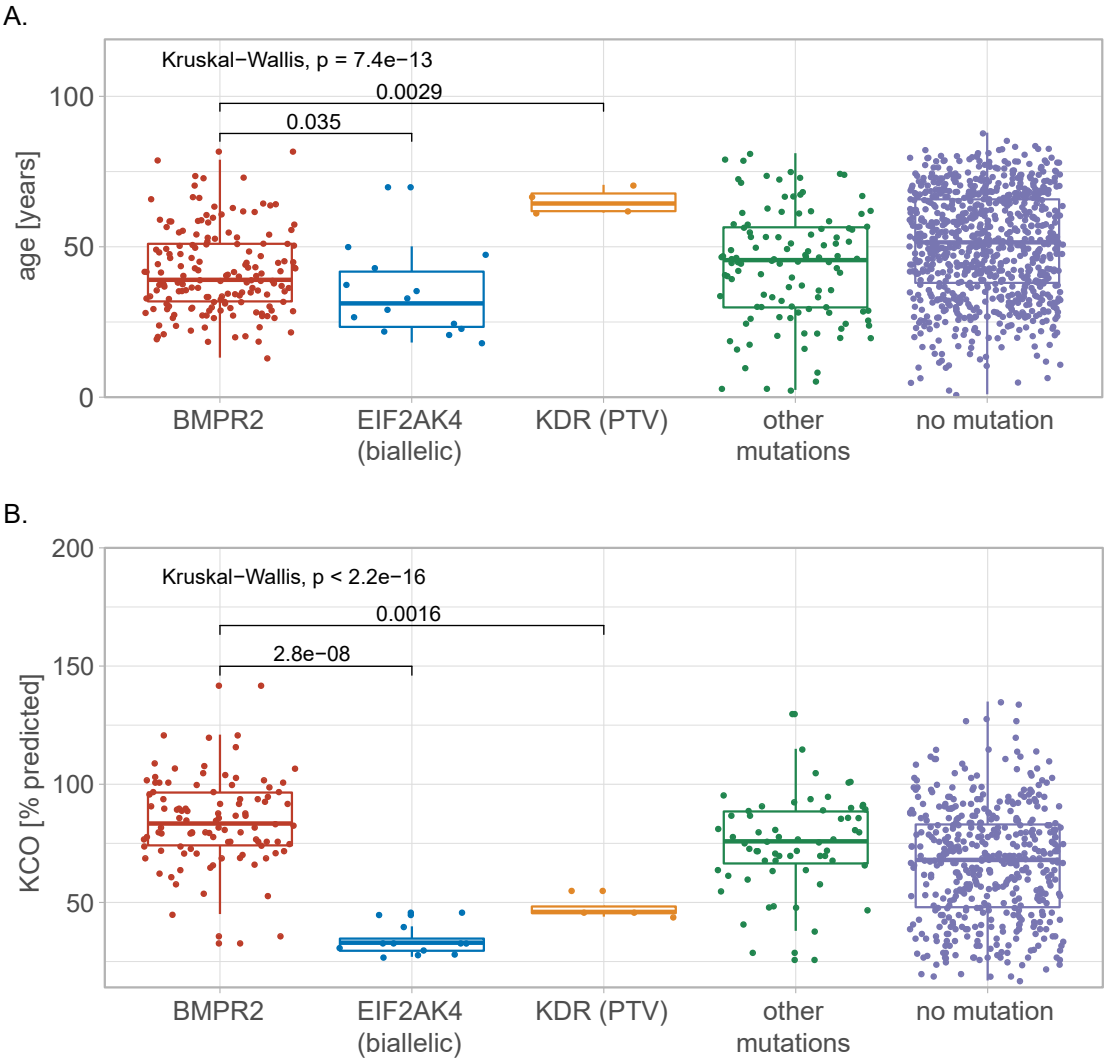
3.3.4 Rare high impact variants in the new PAH candidate gene *KDR*

I identified five ultra-rare high impact variants in *KDR* in the study cohort. Ultra-rare variants exist in the general population only at a frequency of less than 1 in 10,000 (0.01%). Four of these were in PAH cases: one frameshift variant in exon 3 of 30 (c.183del, p.Tryp61CysfsTer*16), two nonsense variants, one in exon 3 (c.183G>A, p.Trp61Ter) and one in exon 22 (c.3064C>T, p.Arg1022Ter) and one splice acceptor variant in intron 4 of 29 (c.490-1G>A). In addition, one nonsense variant was identified in exon 27 (p.Glu1206Ter) in a non-PAH control subject (Table 3.24). This latter nonsense variant appears late in the amino acid sequence, in exon 27 of 30, and hence is likely to escape nonsense-mediated decay, but this remains to be studied functionally. Furthermore, 13 PAH cases (1%) and 108 non-PAH controls (0.9%) harboured rare, predicted-deleterious *KDR* missense variants of moderate impact (Fig. 3.26). The missense variant carriers, however, did not exhibit a reduced KCO or older age at diagnosis. Instead, these patients show the opposite trend in KCO (Fig. 3.25 A and B). Importantly, seven of the 13 *KDR* missense variants seen in PAH cases were also detected in several non-PAH controls and thus, are of unknown significance. Furthermore, three of these missense variants co-occurred with a predicted-deleterious variant in an established PAH risk gene (two patients carried also a variant in *BMPR2* and one a variant in *AQP1*).



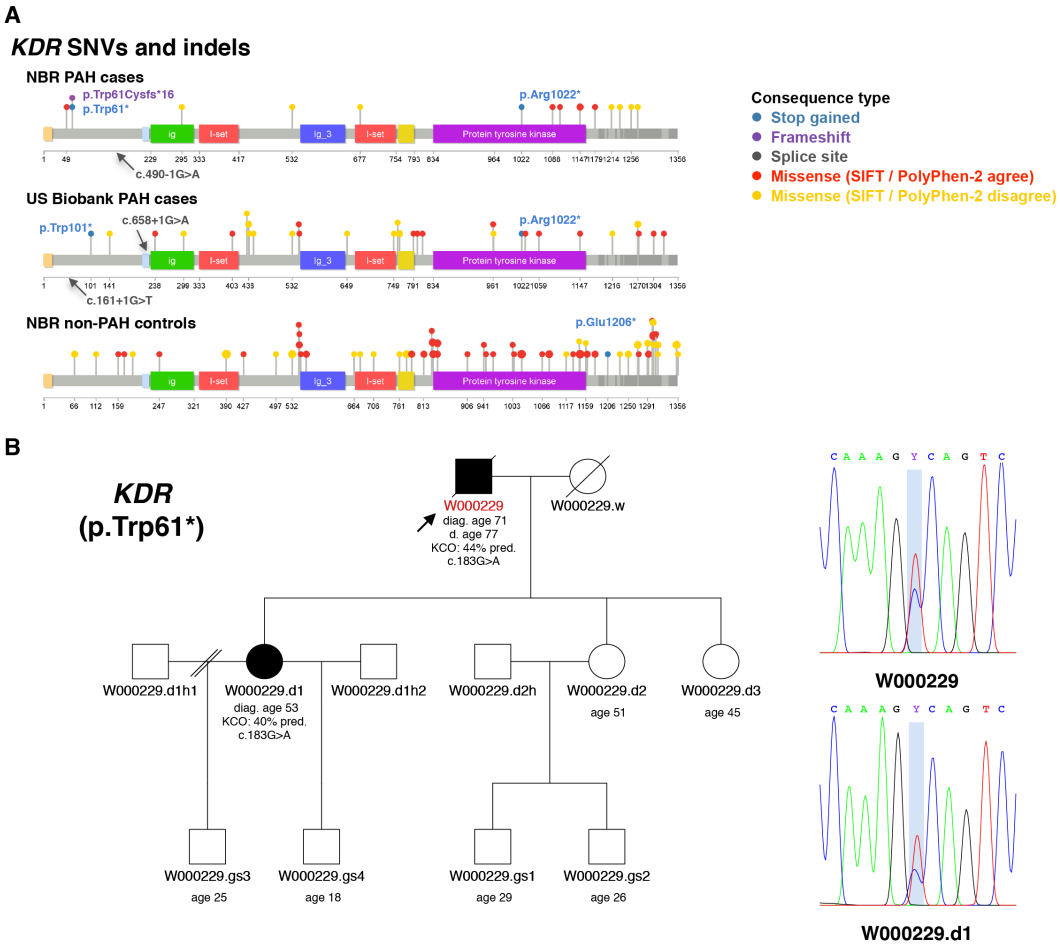
Figures (A) and (B) show phenotype tags on the x-axis and corresponding posterior probability of genotype-phenotype association on the y-axis, as calculated by BeviMed. Shape and colour of points indicate the mode of inheritance and impact/consequence type of variants driving the association.

Fig. 3.24 Rare variant association study results revealing established and new genotype-phenotype links.



Box-and-whisker plots show the distribution of (A) the transfer coefficient for carbon monoxide (KCO) and (B) the age at diagnosis stratified by genotype across the PAH domain. The two-tailed Wilcoxon signed-rank test was used to determine differences in the medians of the distributions, which are indicated by the bars at the top of the figures providing the respective *p*-values. Abbreviations: bial. - biallelic, lof - loss-of-function, mis. - missense.

Fig. 3.25 Established and new genotype-phenotype links



(A) Only rare predicted deleterious variants in *KDR* are shown (MAF $<$ 1/10,000 and CADD $\geq$ 10). SNVs and indels are represented by colored lollipops on top of the protein sequence. The domain annotations were retrieved from Uniprot (accession number P35968). Lollipop colors indicate the consequence type and sizes represent the variant frequency within a cohort. Missense variants that are predicted to be deleterious (SIFT) and damaging (PolyPhen-2) are colored in red, otherwise in yellow (i.e. SIFT and PolyPhen-2 disagree). High impact variants are labelled with the respective HGVS notation. The number of variants by predicted consequence type and cohort is provided in the table. (B) Familial segregation of *KDR* nonsense variant c.183G>A (p.Trp61*) with PAH (i.e. reduced KCO and late onset) from father (W000229) to daughter (W000229.d). Sanger sequencing results are shown in the chromatograms.

Fig. 3.26 Summary of rare single nucleotide variants (SNVs) and small insertions and deletions (indels) identified in the novel PAH candidate gene *KDR*.

Table 3.24 Gene changes for IPAH patients harboring protein-truncating variants (PTV) in the *KDR* gene.

	KDR								
Cohort	**UK**			**US**					
ID	**W000229**	**W000229.d**	**E003448**	**W000274**	**E001392**	**CUMC-JM161**	**CCHMC-12-190**	**CCHMC-19-023**	**CCHMC-27-015**
Exon	3/30	3/30		22/30	3/30	2/30	3/30	5/30	22/30
HGVSc	c.183G>A	c.183G>A	c.490-1G>A	c.3064C>T	c.183del	c.161+1G>T	c.303C>A	c.658+1G>A	c.3064C>T
HGVSp	p.Trp61Ter	p.Trp61Ter	-	p.Arg1022Ter	p.Trp61CysfsTer16		p.Tyr101Ter		p.Arg1022Ter
Consequence type	stop gained	stop gained	splice acceptor variant	stop gained	frameshift variant	splice donor variant	stop gained	stop gained	stop gained
Shared	PAH(1)	PAH(1)	PAH(1)	PAH(1)	PAH(1)	No	No	No	No
gnomAD	NA	NA	NA	NA	NA	NA	NA	NA	NA
CADD PHRED v1.3	38	38	26	37	35	26.4	38	24.3	37
GerpN	5.93	5.93	5.75	5.95	5.93	5.83	5.48	5.8	5.95
Ancestry	European	European	European	European	European	East-Asian	European	European	European
Sex	male	female	female	male	female	female	male	female	female
Diagnosis	IPAH	IPAH	IPAH	IPAH	IPAH	APAH-CHD#	IPAH	IPAH	IPAH
Age at diagnosis [years]	71	53	62	67	61	4	72	65	42
WHO FC	II	III	III	III	III	II	NA	NA	NA
6MWD [m]	472	202	422	660	180	NA	380	NA	245
SpO_2 pre [%]	95	96	97	98	97	NA	NA	NA	NA
SpO_2 post [%]	86	87	86	NA	91	NA	NA	NA	NA
FEV1 [% pred.]	116	70	90	83	67.3	85	NA	77	NA
FVC [% pred.]	115	76	94	91	72.8	92	NA	83	NA
TLC [% pred.]	NA	69	NA	NA	NA	NA	NA	65	NA
KCO [% pred.]	44	40	46	46	55.2	NA	NA	35*	NA
Smoking history	Never	Never	Never	Ex-smoker	Never	Never	Never	Ex-smoker	Never
mRAP [mmHg]	5	13	8	8	3	NA	5	29	14
mPAP [mmHg]	62	45	57	41	44	NA	49	66	60
PAWP [mmHg]	4	5	15	12	9	NA	5	16	15
CO [L/min]	3.6	3.3	4.58	5.97	5.23	NA	4.33	1.8	4.6
PVR [WU]	16.11	12.12	9.17	4.86	6.69	NA	NA	27.9	9.8
Comorbidities	hyperlipidemia, HTN, DM type 2	DM type 2, OSA, pulmonary fibrosis	HTN, hypothyroidism	DM type 2	CAD, DM type 2	No	HTN, hyperlipidemia,	HTN, hypothyroidism, OA	Obesity, CAD, DM type 2, hypothyroidism
Family history	Yes, daughter	Yes, father	No	No	No	No	No	No	No
Status	dead	alive	alive	alive	dead	alive	alive	alive	alive

None of the *KDR* variants has been previously reported in gnomAD, ExAC or internal controls. *KDR* HGVSc notations are based on the transcript sequence ENST00000263923.4. HGVSp notations are based on the amino acid sequence ENSP00000263923.4. None of the patients harboring PTV in *KDR* had capillary hemangioma. W000229.d is an affected daughter of W000229. Protein truncating variants were defined as stop gained; splice acceptor variants or frameshift variants. Abbreviations: *KDR*: Kinase insert domain receptor; WHO FC: World Health Organization functional class; 6MWD: 6-minute walk distance; SpO_2: arterial oxygen saturation; mRAP: mean right atrial pressure; mPAP: mean pulmonary artery pressure; mPAWP: mean pulmonary artery wedge pressure; CO: cardiac output; PVR: pulmonary vascular resistance; FEV1: forced expiratory volume in 1 sec; FVC: forced vital capacity; KCO: transfer factor coefficient for carbon monoxide; TLC: total lung capacity; HTN: systemic hypertension; CAD: coronary artery disease; OA: osteoarthritis; DM: diabetes mellitus; OSA: obstructive sleep apnea; * DLCO% predicted; #secondary to double outlet RV

3.3.5 Clinical characterisation of *KDR* mutation carriers (reverse phenotyping)

Patients with high impact variants in *KDR* were older and exhibited significantly reduced KCO similar to biallelic *EIF2AK4* mutation carriers and in contrast to *KDR* missense variant and *BMPR2* mutation carriers (Fig. 3.24 A and B, Table 3.25). Three of the four cases did not have a history of smoking. CT scans for all four patients showed a range of mild lung parenchymal changes (Fig. 3.27). W000229 had evidence of mild mainly subpleural interstitial lung disease (ILD), mild emphysema, and air trapping. W000274 had signs of ILD with traction bronchiectasis in the lower zones, mild air trapping, and mild diffuse ground-glass opacities (GGO) and neovascularity. E001392 showed mild centrilobular GGO in addition to moderate pleural effusion and a trace of air trapping, but no ILD. In these cases, it seemed likely that the observed parenchymal changes contributed to the low KCO. In contrast, E003448 had a low KCO despite only a trace of central nonspecific GGO on the CT images. Comparisons of clinical characteristics and CT findings between patients harbouring deleterious mutations in *BMPR2*, *EIF2AK4*, *KDR*, other PAH risk genes, and patients without mutations are presented in Tables B.30 and B.31. Clinical similarity between patients harbouring rare deleterious variants in *KDR* was also reflected in significant ontological similarity (p-value = 0.014 (Lin), p-value = 0.002 (Resnik)). Survival analysis could not be conducted due to the small number of mutation carriers, as well as only two events occurring in this group. Following the death of W000229, his daughter, aged 53, was diagnosed with PAH and had a reduced KCO at 40% predicted. On the CT scan, mild interstitial fibrosis was observed (Fig. A.9), additionally, wedge lung biopsy carried out before PAH diagnosis revealed histological features consistent with bronchocentric fibrotic process. Sanger sequencing confirmed that the father and daughter carried the same deleterious *KDR* nonsense variant p.Trp61Ter (Fig. 3.26).

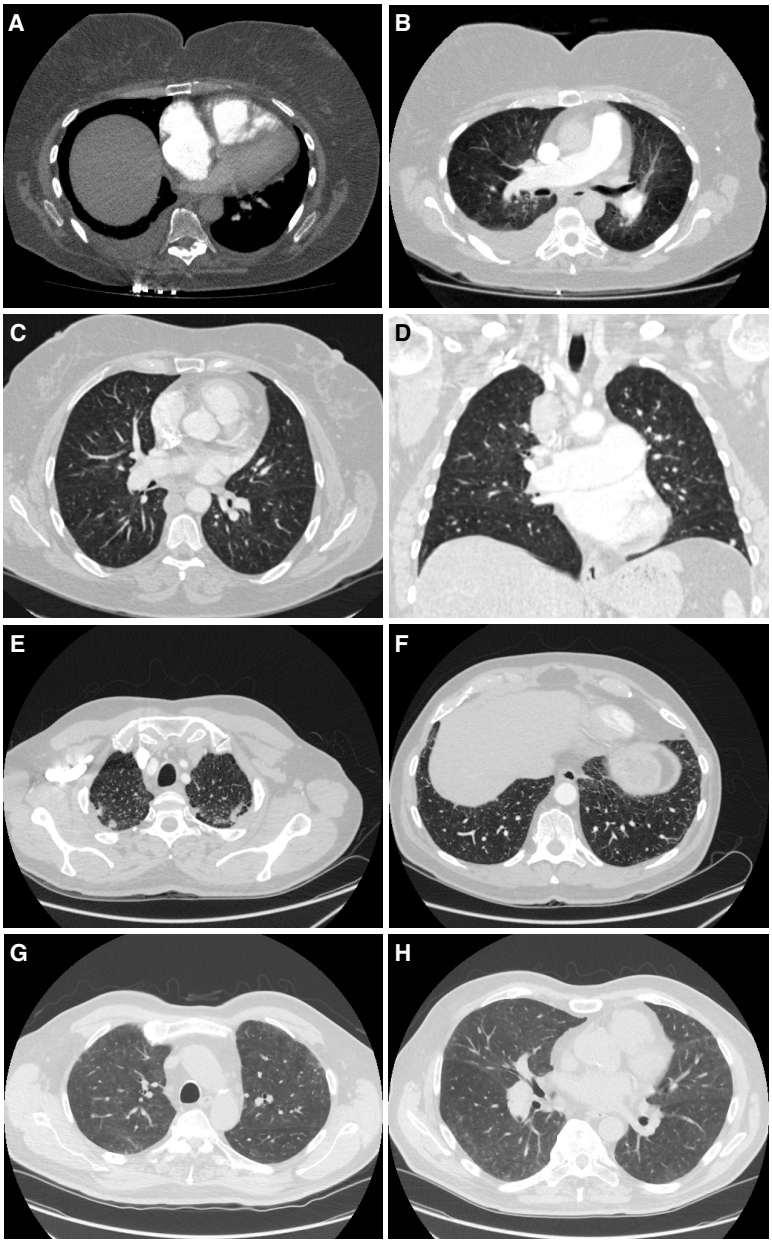
Table 3.25 Differences between patients harbouring missense and PTV *KDR* variants

	KDR missense N = 13	KDR PTV N = 4	*p*-value	N
Diagnosis verified: IPAH	13 (100%)	4 (100%)	-	17
Age [years]	46 [36;59]	64 [62;68]	0.113	17
Sex: female	9 (69%)	2 (50%)	0.584	17
BMI [kg/m^2]	29 [24;32]	26 [26;30]	1	13
WHO FC: II/III/IV [%]	23.1/9.2/7.7	25/75/0	1	17
6MWD [m]	312 [150;355]	301 [240;362]	0.814	11
SpO_2 pre [%]	95 [93;97]	97 [96;97]	0.335	11
SpO_2 post [%]	90 [80;96]	86 [86;88]	0.926	12
mRAP [mmHg]	8 [6;13]	6 [4;8]	0.431	14
mPAP [mmHg]	53 [42;62]	50 [43;58]	0.896	15
PAWP [mmHg]	10 [8;13]	10 [8;13]	0.642	13
CO [L/min]	4.0 [3.0;5.5]	4.9 [4.3;5.4]	0.514	15
PVR [WU]	10.2 [4.56;14.3]	7.93 [6.23;10.9]	1	13
Acute NO challenge: vasoresponder	1 (33.3%)	1 (25.0%)	1	7
FEV1[% pred.]	84 [65;94]	86 [79;96]	0.48	14
FVC [% pred.]	86 [72;97]	92 [86;99]	0.723	14
FEV1/FVC ratio	0.78 [0.75;0.87]	0.78 [0.76;0.79]	0.671	14
KCO [% pred.]	89 [74;93]	46 [46;48]	**0.008**	11
Smoking history	6 (54.5%)	1 (25.0%)	0.677	15
COPD	1 (7.69%)	0 (0.00%)	1	17
Pulmonary fibrosis	0 (0.00%)	2 (50.0%)	**0.044**	17
CAD	1 (7.69%)	1 (25.0%)	0.426	17
HTN	5 (38.5%)	2 (50.0%)	1	17
CKD	1 (7.69%)	0 (0.00%)	1	17

Abbreviations: BMI: body mass index, WHO FC: World Health Organization Functional Class, 6MWD: 6-minute walk distance, SpO_2: oxygen saturation; mRAP: mean right atrial pressure, mPAP: mean pulmonary artery pressure, CO: cardiac output, PAWP: pulmonary artery wedge pressure; PVR: pulmonary vascular resistance; NO: nitric oxide; FEV1: forced expiratory capacity in 1 second, FVC: forced vital capacity, KCO: transfer factor coefficient for carbon monoxide, HTN: hypertension, CAD: coronary artery disease, COPD: chronic obstructive pulmonary disease, CKD: chronic kidney disease

3.3.6 Additional *KDR* cases in the US PAH cohorts

To seek further evidence for *KDR* as a new candidate gene for PAH, I analysed subjects recruited to the US PAHBB [598] to identify additional patients carrying predicted pathogenic rare variants. Four additional individuals harbouring rare high impact *KDR* variants were identified. These comprised, two nonsense variants, one in exon 3 (c.303C>A, p.Tyr101Ter) and one in exon 22 (c.3064C>T, p.Arg1022Ter) and two splice donor variants, one in intron 2 of 29 (c.161+1G>T) and one in intron 5 (c.658+1G>A). Interestingly, the nonsense variant p.Arg1022Ter appeared in both cohorts (Fig. 3.26). Patient-level data for these individuals are summarised in Table 3.24. Three of the four patients were diagnosed with idiopathic PAH at 72, 65 and 42 years respectively, whereas one patient was diagnosed at age four with PAH associated with double outlet right ventricle. The diffusing capacity of carbon monoxide was



(A) Axial image of CT pulmonary angiogram at the level of the right ventricle (RV) moderator band, showing flattening of interventricular septum, leftwards bowing of the interatrial septum and the enlargement of the right atrium (RA) and RV, indicative of RV strain; bilateral pleural effusion, larger on the right side. (B) Axial image of a pulmonary CT angiogram demonstrating enlarged pulmonary artery and mild central lung ground-glass opacity (GGO). (C) Axial high-resolution CT slice of the chest in the lung window showing a trace of non-specific GGO with a central distribution. (D) Coronal image showing the trace of central GGO and enlarged central pulmonary arteries. Axial high-resolution CT slice of the chest in the lung window showing apical subpleural fibrosis (E), and very minor subpleural fibrosis at the lung bases (F). Axial high-resolution CT slice of the chest in the lung window showing subpleural GGO at apical level (G), and mild GGO at mid-thoracic level (H). Patients: E001392 (A, B), E003448 (C, D), W000229 (E, F), W000274 (G,H).

Fig. 3.27 Chest computerised tomography (CT) scans of patients carrying high impact *KDR* mutations.

available for one patient and was decreased at 35% predicted, with minor pleural scarring in the left upper lobe found on CT imaging. Two out of four patients (50%) harbouring a high impact variant in *KDR* had been diagnosed with systemic hypertension.

3.3.7 Other putative genes

Divisive hierarchical clustering cluster 2 was associated with two new genes, namely *COL6A5* and *EYS*, under a dominant mode of inheritance. Strong association with *COL6A5* was driven by 15 moderate and high impact variants in 16 patients Table B.32. Notably, none of the variants driving the association were found in controls. There was one patient with a positive family history (maternal aunt, who died of PAH at the age of 39). Clinical characteristics of patients harbouring deleterious variants in *COL6A5* are shown in Table 3.26. Of note, none of the patients presented with dermatological phenotypes, muscular dystrophy or lung cancer, phenotypes previously reported to be linked to mutations in collagen VI. While the divisive HC cluster 2 was characterised by worse outcomes when compared to divisive HC cluster 1, the survival analysis among patients harbouring *COL6A5* was not possible due to the small number of events (two deaths). Similarly, the *EYS* signal was driven by 15 moderate impact variants in 16 individuals Table B.32, including one patient who was carrying deleterious variants in both *COL6A5* and *EYS* (W000042). When the individual with variants in both *COL6A5* and *EYS* was removed from the association testing the strength of the association between Divisive HC cluster 2 and *EYS* decreased ($\log(BF) = 7.72$, $PP = 0.693$). Deleterious variants in PAM clusters 1 and 4 highlighted associations with high impact variants in *HPSE2* (signal driven by stop gain, splice acceptor variant and deletion) and five moderate impact variants (missense variants) in *OR6T1*, respectively.

Table 3.26 Clinical characteristics of patients harbouring deleterious variants in *COL6A5*.

	[ALL] N=16	N
Diagnosis: IPAH	16 (100)	16
Sex: female, n (%)	13 (81)	16
Age [years]	41 [30;67]	16
WHO FC: I/II/III	20/13/67	15
6MWD [m]	353 (206)	11
mRAP [mmHg]	7 (4)	12
mPAP [mmHg]	41 (14)	13
PAWP [mmHg]	8 (3)	13
CI [L/min/m^2]	2.7 (0.9)	9
PVR [WU]	8 (4)	12
FEV1 [% pred.]	84.9 (12.0)	12
FVC [% pred.]	93.8 (15.2)	12
TLC [% pred.]	97.4 (11.7)	10
KCO [% pred.]	86.0 (19.0)	9
Emphysema: none, n (%)	5 (100)	5
Fibrosis: minimal/mild, n (%)	1 (20.0)	5
Smoking: past/current, n (%)	5 (41.7)	12
NTproBNP [ng/l]	2332.5	2
BNP [ng/l]	75.0 (56.8)	7
CRP [mg/l]	4.30 [2.08;9.93]	12
Creatinine [μmol/l]	70.5 [68.8;76.0]	12
DM type 2, n (%)	2 (12.5)	16
COPD, n (%)	1 (6.25)	16
OSA, n (%)	1 (6.25)	16
CAD, n (%)	1 (6.25)	16
HTN, n (%)	6 (37.5)	16

Abbreviations: IPAH: idiopathic pulmonary arterial hypertension; WHO FC: World Health Organisation Functional Class; 6MWD: six minute walk distance; mRAP: mean right atrial pressure, mPAP: mean pulmonary artery pressure; PAWP: pulmonary artery wedge pressure; CI: cardiac index; PVR: pulmonary vascular resistance; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; TLC: total lung capacity; KCO: transfer factor for carbon monoxide; NTproBNP: N-terminal pro B-type natriuretic peptide; BNP: B-type natriuretic peptide; CRP: C-reactive protein; DM: diabetes mellitus; COPD: chronic obstructive pulmonary disease; OSA: obstructive sleep apnea; CAD: coronary artery disease; HTN: hypertension

Besides *KDR*, the KCO stratification also highlighted an association between isocitrate dehydrogenase (NAD(+)) 3 non-catalytic subunit gamma (*IDH3G*) and moderately reduced KCO in patients with pulmonary hypertension (log(BF) = 9.34, PP = 0.91). In total 5 patients harbouring *IDH3G* variants have been identified (3 with missense variants and 2 deletions). Detailed analysis of the deletions revealed that the region has been poorly covered and that the deletions were most likely falsely called. Upon removal of 2 deletions, the association between *IDH3G* weaken (log(BF) = 7.523, PP = 0.649). Additionally, three individuals carrying rare missense variants in *IDH3G* were identified in the US cohort: a male neonate diagnosed with Scimitar syndrome, hypoplastic right lung and atrial septal defect (ASD) (c.1091C>T, p.Pro364Leu), a 55-year-old female with large ASD (c.217G>C, p.Val73Leu) and a two-year-old female with CHD-PAH (c.5C>T, p.Ala2Val); these phenotypes were

not in keeping with those found in the NBR cohort (Table 3.27). To further elucidate the role of *IDH3G*, I started a collaboration with a structural biology group in Shanghai, China, which has a special interest and expertise in IDH3 structure. Missense mutations were all found to be located at the large domain of the γ subunit, and are positioned far away from the active site, the allosteric site, and the N-terminal of the γ subunit (which participates in the assembly of the IDH3 holoenzyme). Thus, these mutations appeared to have no direct effects on the activity of the holoenzyme. Furthermore the enzymatic activity assay performed by the Chinese group showed that the mutant holoenzymes exhibit comparable activity and kinetic properties as the wild-type IDH3 holoenzyme (Fig. A.10).

Table 3.27 Gene changes for IPAH patients harboring PTV and missense variants in the *IDH3G* gene.

Gene	***IDH3G***						
Cohort	**UK**				**US**		
ID	**E004149**	**E004194**	**E001063**	**W000031**	**CCHMC_22-105**	**JM1415**	**CCHMC_10-074**
Exon	1/13	1/13	12/13	12/13	13/13		4/13
HGVSc	c.74C>T	c.74C>T	c.1067T>C	c.1037C>T	c.1091C>T	c.161+1G>T	c.217G>C
HGVSp	p.Pro25Leu	p.Pro25Leu	p.Met356Thr	p.Thr346Ile	p.Pro364Leu		p.Val73Leu
Consequence type	missense variant	missense variant	missense variant	missense variant	missense variant	splice donor variant	missense variant
Shared	PAH(2)	PAH(2)	PAH(1)	PAH(1)	NA	NA	NA
gnomAD	NA	NA	1.15E-05	NA	5.47E-06	2.62E-05	1.09E-05
CADD PHRED v1.3	23.3	23.3	25.6	12.02	23.3	26.4	22.9
SIFT	deleterious(0.04)	deleterious(0.04)	deleterious(0)	deleterious(0.04)	deleterious(0.04)		deleterious(0.02)
PolyPhen	benign(0.006)	benign(0.006)	benign(0.297)	benign(0.193)	benign(0.104)		bening(0.022)
GerpN	5.22	5.22	5.46	5.46	5.18	5.22	4.72
Ancestry	European	European	European	European	European	European	European
Sex	female	female	female	female	female	female	male
Diagnosis	IPAH	IPAH	IPAH	IPAH	CHD-PAH	CHD-PAH	CHD-PAH
Age at diagnosis [years]	27	34	51	68	0	19	55
WHO FC	III	IV	IV	II	III	II	III
6MWD [m]	NA	414	NA	414	316	364.2	316
SpO_2 pre [%]	96	95	98	96	NA	98	NA
SpO_2 post [%]		99	96	95	NA	99	NA
FEV1 [% pred.]	87	104	95	99.1	NA	59	NA
FVC [% pred.]	90	109	95.8	96.3	NA	67	NA
TLC [% pred.]	NA	105	76	98	NA	85	NA
KCO [% pred.]	71	64	78	73	NA	NA	NA
Smoking history	Ex-smoker	Never	Never	Never	Never	Never	NA
mRAP [mmHg]	14	8	12	6	3	2	7
mPAP [mmHg]	64	49	50	62	46	37	69
PAWP [mmHg]	8	10	12	7	NA	5	10
CO [L/min]	3.23	NA	3.29	4.1	4.4	6.32	
PVR [WU]	17.3	NA	11.6	13.4	NA	5.2	
Comorbidities	No	No	PFO	No	Scimitar syndrome, hypoplastic right lung, ASD with spontaneous closure	VSD	Large ASD
Family history	No	No	No	No	No	No	NA
Status	alive	alive	dead	alive	alive	alive	alive

Gene changes for IPAH patients harboring PTV and missense variants in the IDH3G gene. IDH3G HGVSc notations are based on the transcript sequence ENST00000217901.5, HGVSp notations are based on the amino acid sequence ENSP00000217901.5. Protein truncating variants were defined as stop gained, splice acceptor variants or frameshift variants. Abbreviations: IDH3G: Isocitrate dehydrogenase (NAD(+)) 3 non-catalytic subunit gamma; IPAH: idiopathic pulmonary arterial hypertension; CHD: congenital heart disease; WHO FC: World Health Organization functional class; 6MWD: 6-minute walk distance; SpO2: arterial oxygen saturation; mRAP: mean right atrial pressure; mPAP: mean pulmonary artery pressure; mPAWP: mean pulmonary artery wedge pressure; CO: cardiac output; PVR: pulmonary vascular resistance; FEV1: forced expiratory volume in 1 sec; FVC: forced vital capacity; TLC: total lung capacity; KCO: transfer factor coefficient for carbon monoxide; DM: diabetes mellitus; ASD: atrial septal defect; PFO: patent foramen ovale; VSD: ventricular septal defect; WU: Wood Units; * DLCO% predicted.

3.3.8 Discussion

One of the critical steps in identifying novel, causative genes in rare disorders is the discovery of genotype-phenotype associations to inform patient care and outcomes. A pragmatic focus on deeply phenotyped individuals and "smart" experimental design provides additional leverage to identify novel risk variants [140]. To deploy this approach in PAH, I brought together phenotypic and genetic data using BeviMed. The Bayesian framework allows the inclusion of prior information regarding the hypothesis being tested in a flexible manner and compares a range of possible genetic models in a single analysis. To generate case-control labels, I tagged PAH cases with diagnostic labels and predefined phenotypes. Analyses were then performed to identify associations between tags and ultra-rare gene variants under dominant and recessive modes of inheritance and different variant impact categories.

The BeviMed analysis provided strong statistical evidence of an association between ultra-rare, high impact variants in *KDR* and PAH with significantly reduced KCO and older age at diagnosis under a dominant mode of inheritance. Strikingly, likely loss-of-function variants in *KDR* exist in the general population with a frequency of only 4-7 per 100,000 (Table 3.28). In contrast, I identified four PAH cases in the NBR cohort which equates to almost 2 in 1,000. Additionally, the statistical constraint metrics provided by gnomAD [254] strongly suggest that loss-of-function variants in *KDR* are not tolerated (pLI = 1; o/e = 0.15 (0.09 - 0.25)). Besides the statistical evidence, I also identified one additional case with a family history, which together with a recently published case report of two families in which loss-of-function variants in *KDR* segregated with PAH and significantly reduced KCO [126], amounts to three reported familial cases with a distinct phenotype.

Table 3.28 Comparison of high impact likely loss-of-function variants in *KDR* in the Human large-scale sequencing reference populations gnomAD and TOPMed with the NBR non-PAH controls and PAH cases.

Large-scale sequencing population	High impact LoF variants in *KDR*	Individuals	Alleles	Frequency
gnomAD (v2.1)*	20	141.456	282.912	0.000071
gnomAD (v3)	10	71.702	143.404	0.00007
TOPMed	5	62.784	125.568	0.00004
NBR non-PAH controls	1	11.889	23.778	0.000042
NBR PAH cases	4	1.122	2.244	0.001783

**KDR* constraint metrics: pLI = 1; o/e = 0.15 (0.09 - 0.25); exp(LoF) = 73; obs(LoF) = 11

Abbreviations: gnomAD: The Genome Aggregation Database; TOPMed: The Trans-Omics for Precision Medicine program; *KDR*: kinase insert domain receptor; NBR: NIHR BioResource - Rare Diseases; LoF: loss of function.

VEGFR2, which is encoded by *KDR* [515], binds VEGFA, a critical growth factor for physiological and pathological angiogenesis in vascular endothelial cells. In mice, even though *VegfA* haploinsufficiency is embryonically lethal [134], heterozygosity of its receptor,

Vegfr2, is compatible with life and unperturbed vascular development [387]. The role of VEGF signalling in the pathogenesis of PAH has been an area of intense interest since increased expression of VEGF, VEGFR1 and VEGFR2 were reported in rat lung tissue in response to acute and chronic hypoxia [536]. An increase in lung VEGF has also been reported in rats with PH following monocrotaline exposure [78]. In humans, VEGF-A is highly expressed in plexiform lesions in patients with IPAH [535]. In addition, inhibition of VEGF signalling by SU5416 (sugen) combined with chronic hypoxia triggers severe angioproliferative PH [513]. SU5416, a small-molecule inhibitor of the tyrosine kinase segment of VEGF receptors, inhibits VEGFR1 [245] and VEGFR2 [535] causing endothelial cell apoptosis, loss of lung capillaries and emphysema [256]. Similarly, female conditional *Kdr* knockout mice were found to have increased RVSP and decreased RV ejection fraction as well as pulmonary vasculopathy similar to that seen in patients with IPAH. Surprisingly, no emphysema, systemic hypertension, left ventricular dysfunction or changes in vessels of venous origin were reported [574]. Further evidence supporting the role of VEGF inhibition in the pathobiology of PAH comes from reports of PH in patients treated with bevacizumab [161] and the multi-tyrosine kinase inhibitors [356, 122]. Mutations in *KDR* have also been linked to congenital heart diseases. Bleyl et al. reported that *KDR* might be a candidate for familial total anomalous pulmonary venous return [49]. In addition, haploinsufficiency at the *KDR* locus has also been associated with tetralogy of Fallot [434]. I identified one patient in the CUMC cohort with PAH associated with congenital heart disease harbouring a *KDR* likely protein-truncating splice donor variant (c.161+1G>T).

In the present study, I highlight that deep clinical phenotyping, in combination with genotype data, can improve the identification of novel disease risk genes and disease subtypes. *KDR* was already identified as a possible candidate gene, which did not achieve genome-wide significance, in our previous rare variant association study [179]. In combination with deep phenotyping data, *KDR* reached, in the present study, a significance level comparable to the most commonly affected genes in PAH. Reduced KCO, which reflects impairment of alveolar-capillary membrane function, has been noted in the analysis of early PAH registry data [517] to be an independent predictor of survival. Decreased KCO was also found in patients with PVOD/PCH with or without biallelic *EIF2AK4* mutations [194]. Although some reduction in KCO is one of the typical features of pulmonary vascular disease, PVOD patients show the lowest KCO values when compared to IPAH or CTEPH. In contrast, KCO is relatively preserved in *BMPR2* and *ATP13A3* mutation carriers. Strong association with survival and a link with other causative mutations makes the KCO phenotype particularly attractive for stratification in genetic studies.

As lung disease should always be taken under consideration as a cause of low KCO, I applied the World Symposium on PH criteria [372] to exclude lung disease as a cause of PH: TLC $\geq$ 70% pred., FVC $\geq$ 70% pred., FEV1 $\geq$ 60% pred., and no severe fibrosis and/or emphysema on chest CT. None of the cases carrying a high impact variant in *KDR* met these criteria, although two of the four patients did show evidence of early ILD. Another potential reason for low KCO in the PAH population is the diagnosis of PVOD/PCH [358]. Careful analysis of CT scans and clinical data did not reveal convincing evidence for this diagnosis in *KDR* mutation carriers. Cigarette smoking is a well-known factor leading to the decrease of KCO. Only one of the four *KDR* high impact variant carriers had a significant 15 pack-years smoking history, but with no signs of emphysema on CT. These findings suggest that loss-of-function variants in *KDR* are associated with a form of PAH characterised by a range of lung parenchymal abnormalities, including small airways disease, emphysema and ILD, as two of the four patients harbouring a high impact variant in *KDR* had mild fibrotic lung changes. Notably, patients with mutations in other PAH risk genes, or those without the identified genetic mutation, showed less than 10% incidence of fibrotic changes on CT imaging. Additionally, the affected daughter of W000229 was found to have histological features consistent with bronchocentric fibrotic process on lung wedge biopsy. Further larger studies are needed to determine the full range of lung parenchymal abnormalities in PAH cases with deleterious variants in *KDR* and their progression over time.

The second new association I found was between *KDR* and late disease onset (tag: old age). This finding is of paramount importance in the context of age-dependent impairment of angiogenesis, whereby age-related endothelial health decline coupled with germline likely-loss-of-function variants in *KDR* may explain the development of the disease later in life. Age is a major cause of endothelial dysfunction [543]. Several mechanisms have been implicated in vascular ageing, such as loss of NO vasoprotective and cardioprotective effects [543], increased production of reactive oxygen species (ROS) [201, 93], endothelial apoptosis [19, 94] and senescence [566]. Besides, age-related changes to circulating and local factors contribute to impaired angiogenesis. For example studies in animals and humans showed an age-dependent decline in circulating growth hormone [263], insulin-like growth factor-I [512], oestrogens [284] and VEGF [456], similarly, age was shown to impact on pro-angiogenic gene expression signature in heart and skeletal muscles. The impact of age on *KDR* expression is debatable, with some studies showing no effect [557, 448], while others report significant correlation and point to *KDR* as a potential therapeutic target [248].

Finally, *KDR* was associated with computationally derived composite phenotype, PAM 2 (Lin), whose characteristics overlapped with two clusters devised on a clinical basis, namely

age in higher tertile and KCO in lower tertile. This convergence of clinical and computational phenotypes coupled with a new significant genotype-phenotype association validates the use of computational phenotypes in RVAS.

Candidate genes, pitfalls of the current analysis

Other candidate genes identified in this study, although some functionally plausible, do not have enough supporting evidence to be considered drivers of PAH at present.

COL6A5

The strongest new association revealed by computational phenotypes was found between the divisive HC cluster 2 and *COL6A5*, encoding Collagen Type VI Alpha 5 Chain. Collagen Type VI is encoded by six different genes (*COL6A1*, *COL6A2*, *COL6A3*, *COL6A4*, *COL6A5* and *COL6A6*), although each triple-helix monomer is composed of only three chains α1(VI), α2(VI), and αX(VI), wherein αX(VI) can be either an α3, α4, α5 or α6 chain [160]. Collagen VI plays an important cytoprotective role in reducing apoptosis and oxidative damage, regulating autophagy and cell differentiation, promoting tumour growth and progression and contributing to stemness in various tissues [70]. Complex structure and variety of domains allow collagen binding to diverse components of the extracellular matrix. Of interest is the interaction of collagen VI with fibulin-2 (FBLN-2) [454], an extracellular matrix protein involved in elastic fibre formation and regulation of cell motility, proliferation and angiogenesis and also recently associated with PAH [596]. Given high expression in skin and lung, it is not surprising that polymorphisms in *COL6A5* have been linked to dermal phenotypes [493] and susceptibility to lung cancer [115]. Strong association of *COL6A5* with the cluster characterised by different PAH groups broadly supports the association results from the combined UK-US analysis, which showed a significant association with PAH phenotype (Group 1) but not with IPAH [596]. Interestingly, a polymorphism in another member of the collagen family, Collagen Type XVIII α1 Chain (*Col18a1*), was reported at increased frequency in patients with IPAH and PAH-CTD when compared to controls. Col18a1 is cleaved into two chains, endostatin and non-collagenous domain 1. Patients with single minor allele A (genotype AG), had lower serum endostatin levels than those who carried WT (genotype GG) and showed better survival even after adjusting for confounders [101]. Endostatin is a potent inhibitor of endothelial cell proliferation and angiogenesis [488]. It was found to suppress VEGF-induced endothelial cell proliferation and migration by impeding the interaction of VEGFA to its receptor VEGFR2. It also interacts with FBLN1 and FBLN2 and may negatively regulate the activity of homotrimeric non-collagenous domain 1

[278]. Taken together, strong statistical evidence and biological plausibility, along with *in silico* predictions of the impact of identified mutations on collagen VI function and possible familiar segregation of the variant with the phenotype, make *COL6A5* a strong candidate gene that warrants further research.

IDH3G

Metabolic disturbances in pulmonary arteries and the right ventricle [450], also reflected in plasma metabolite levels [437], are well-documented drivers of the PAH process. *IDH3G* encodes the gamma subunit of isocitrate dehydrogenase 3 (IDH3). IDH3 is one of the three highly conserved IDHs [347]. IDH3 catalyses the irreversible conversion of isocitrate to alpha-ketoglutarate as a part of the tricarboxylic acid (TCA) cycle [227]. Functional studies, using human endothelial cells, indicated that IDH3 activity remained comparable between wild type lines and those harbouring disease-causing *BMPR2* variants, however, mutants displayed a significant increase in IDH1 and IDH2 activity. Moreover, IDH1 and IDH2 activity was significantly higher in the serum of patients with H/IPAH irrespective of mutation status, when compared to healthy controls, and correlated with disease severity [135]. I have identified three deleterious, heterozygous missense variants in *IDH3G* in four IPAH patients with moderately decreased KCO (non-extreme phenotype) at PP below the discovery threshold. Additional patients with deleterious variants in *IDH3G* from the US PAHBB did not share clinical phenotype. Functional analysis showed that identified mutations did not have direct effects on IDH3 activity. Taken together this evidence suggests that the discovered variants were false positives, alternatively non-enzymatic properties of IDH3 could be responsible for the effect of these mutations.

The fact that mutations in many established PAH genes do not result in distinct phenotypes suggests a convergence of many molecular pathways in PAH. For example, while VEGF has been shown to upregulate *SOX17*, the Notch pathways play a critical role in downregulating *Sox17* in the endothelium [266]. Likewise, *ADGRG6*, in which variation is associated with KCO, plays an important role in angiogenesis through the regulation of VEGFR2 [516]. Conversely, many computationally derived phenotypes, despite strong ontological similarity, did not yield positive results in genotype-phenotype association analysis. This apparent inconsistency may be attributed to variables unrelated but correlated to the phenotype of interest, which inflate prediction accuracy.

In this analysis, I have assumed that PAH is a monogenic condition, which is caused by either deleterious heterozygous or biallelic variants in a single gene. This assumption,

although widely supported by the literature, may not be entirely accurate. Alternatively, some cases of PAH might represent an oligogenic inheritance involving two or more genes. Although not statistically explored in the current analysis, I found a total of 22 PAH cases carrying deleterious variants in more than one PAH gene. These variants could contribute as genetic modifiers, impacting penetrance and/or expressivity. In this analysis, I have explored only a limited number of clinical phenotypes. Further studies with larger numbers of phenotypic tags derived from clinical and molecular data will increase the power to detect new associations. Finally, to mitigate potential selection bias introduced by missingness in KCO measurements I performed the analysis on complete and imputed KCO data and obtained similar results.

3.3.9 Conclusions

In summary, this analysis shows that deep phenotyping enables patient clustering into subgroups with shared pathobiology and ultimately increases power to detect new genotype-phenotype associations. I provide statistical evidence for an association between high impact, likely loss-of-function variants in *KDR* and significantly decreased KCO and later disease onset, further supported by familial segregation, and propose *COL6A5* as a new PAH candidate gene.

In Chapter 3.3 I addressed aim number 3 of my Thesis.

3.4 Reverse phenotyping

Parts of this chapter have been previously published: Hodgson J, Swietlik EM, et al. Characterization of *GDF2* Mutations and Levels of BMP9 and BMP10 in Pulmonary Arterial Hypertension Am J Respir Crit Care Med. 2020 Mar 1;201(5):575-585. PMID: 31661308

3.4.1 Clinical characterisation of *GDF2* carriers

In this chapter I assess the impact of *GDF2* mutations on clinical characteristics of affected mutation carriers and evaluate plasma BMP9 and pBMP10 levels in a large cohort of PAH patients. I hypothesised that patients harbouring mutations in *BMPR2* and *GDF2* would show distinct phenotypes from patients without mutations in previously established disease genes.

Our previous analysis [179] identified seven likely pathogenic missense variants, and one frameshift variant in *GDF2*. Additional analysis found two patients with large deletions, not previously described, encompassing the *GDF2* locus and several neighbouring genes (Table 3.29, Fig. 3.28 A). The p.Tyr351His variant, found in two unrelated patients, was not previously reported and was predicted to disrupt the hydrophobic core of BMP9 (Fig. 3.28 B). Two patients with *GDF2* missense variants (p.Met89Val and p.Tyr351His) had a positive family history for PAH, but only one variant carrier (p.Met89Val) had enough information to draw a pedigree, indicating familial clustering of the phenotype (Fig. 3.28 C).

Comparison of baseline characteristics between the *GDF2* mutation carriers (n = 12), patients harbouring heterozygous *BMPR2* mutations (n = 159) and patients with no identified mutation (n = 759) is shown in Table 3.30. *GDF2* mutation carriers were similar to PAH patients without mutations and showed no features of HHT or vascular anomaly syndromes. *GDF2* mutation carriers were significantly older and had less severe haemodynamics than *BMPR2* mutation carriers. Transplant-free survival was similar between groups after adjustment for age, sex, and whether cases were incident or prevalent. The two patients with large deletions exhibited earlier disease onset at 19 and 30 years of age (Table 3.29). The 19-year-old patient underwent transplantation three years after diagnosis. Despite the deletions affecting a number of additional genes (Fig. 3.28 A), these individuals did not have other reported co-morbidities and mutations in these other genes cause autosomal recessive disorders. Unexpectedly, patients harbouring deleterious variants in *GDF2* not only did not share the phenotype with *BMPR2* mutation carriers, but also were not phenotypically similar to each other (Table 3.19). Two rare and deleterious missense variants in *BMP10* (p.Ala361Glu and p.Arg353Cys) were also identified in PAH patients (Table 3.31).

Table 3.29 Gene changes, demographics and pulmonary haemodynamic data for PAH patients harbouring *BMP9* mutations.

ID	**E013254**	**E005056**	**E006312**	**E010660**	**E000837**	**E010643**	**E012223**	**E001147**	**E004186**	**W000134**	**E011142**	**W000048**
Nucleotide change	c.137_150 delGTGGGCT-GCCTGAG	c.265A>G	c.328C>T	c.347A>T	c.427G>A	c.958A>T	c.1040C>T	c.1051T>C	c.1051T>C	c.1238C>A	N/A	N/A
Protein change	p.Gly46Ala fsTer21	p.Met89Val (M89V)	p.Arg110Trp (R110W)	p.Asp116Val (D116V)	p.Glu143Lys (E143K)	p.Ser320Cys (S320C)	p.Ala347Val (A347V)	p.Tyr351His (Y351H)	p.Tyr351His (Y351H)	p.Thr413Asn (T413N)	N/A	N/A
Impact	Frameshift variant	Missense variant	Missense variant	Missense variant	Missense variant	Missense variant	Missense variant	Missense variant	Missense variant	Missense variant	Deletion	Deletion
Ethnicity	European	European	European	European	European	South-Asian	European	European	European	Other	European	South-Asian
Age at diagnosis	45	53	46	52	45	56	30	73	45	54	19	30
Sex	male	female	female	male	male	male	female	female	female	female	male	female
WHO FC	III	III	II	III	n.d	III	III	III	II	III	IV	II
Exercise capacity	470 (6WMT)	432 (6MWT)	588 (6MWT)	395 (6MWT)	n.d	403 (6MWT)	220 (Shuttle)	160 (6MWT)	420 (6MWT)	374 (6MWT)	280 (Shuttle)	405 (6MWT)
mPAP [mmHg]	46	34	38	55	52	56	53	59	73	50	n.d.	55
PAWP [mmHg]	16	8	5	9	10	7	n.d	13	9	7	n.d.	8
PVR [WU]	5.8	8.7	6.9	15	8.4	9.4	n.d	11	15	15	n.d.	4.66
CO [L/min]	5	3	4.8	3	5	5.2	4.87	4.1	4.2	2.8	n.d.	10
Family history	No	Yes	No	No	No	No	No	Yes*	No	No	No	No

Abbreviations: WHO FC: World Health organization Functional Class; mPAP: mean pulmonary artery pressure; PVR: pulmonary vascular resistance; PAWP: pulmonary artery wedge pressure; CO: cardiac output; n.d.: not done. *A relative in the parental generation was affected, but there is not enough information to draw a pedigree.

Table 3.30 Clinical characteristics of *BMPR2* and *GDF2* mutation carriers and patients without these mutations.

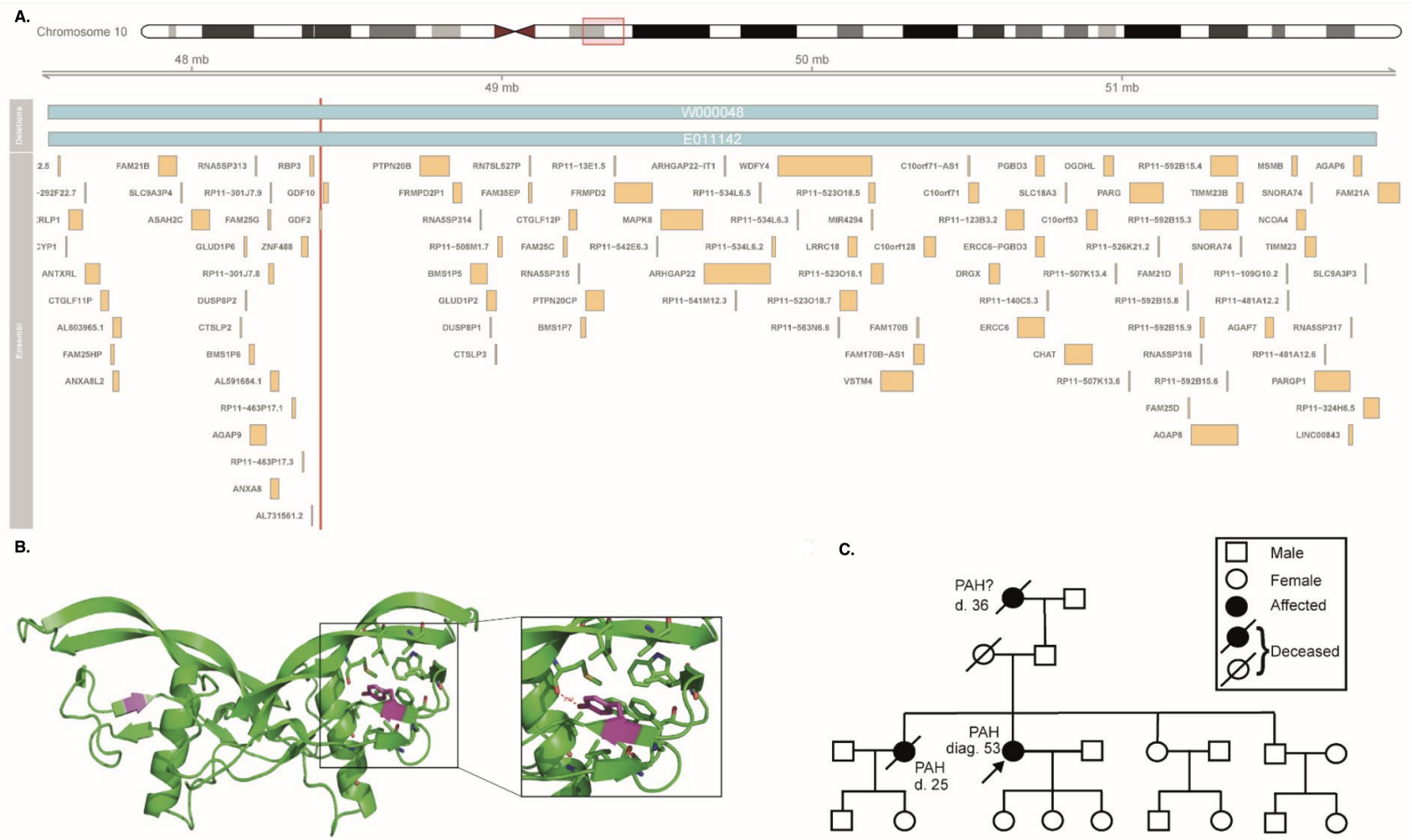
	***BMPR2* N=159**	***GDF2* N=12**	**no mutation N=750**	***p*-value**	**N**
Sex: female, n (%)	105 (66)	7 (58)	530 (71)	0.357	921
Age [years]	39 [32;51]	46 [41;54]	52 [39;66]	<0.001	921
Diagnosis: n (%)				<0.001	921
HPAH	51 (32.1)	1 (8.33)	7 (0.93)		
IPAH	108 (67.9)	11 (91.7)	743 (99.1)		
WHO FC: I/II/III/IV, %	1/20/60/19	0/27/64/9	2/19/66/12	0.294	895
Ethnicity: n (%)				0.328	921
African	2 (1.26)	0 (0.00)	20 (2.67)		
East-Asian	1 (0.63)	0 (0.00)	7 (0.93)		
European	136 (85.5)	9 (75.0)	638 (85.1)		
Finnish-European	0 (0.00)	0 (0.00)	1 (0.13)		
South-Asian	6 (3.77)	2 (16.7)	46 (6.13)		
Other	14 (8.81)	1 (8.33)	38 (5.07)		
mPAP [mmHg]	58 [52;68]	53 [48;56]	52 [43;61]	<0.001	877
mPAWP [mmHg]	9 [7;12]	8 [7;10]	9 [7;12]	0.931	780
PVR [WU]	14 [11;20]	10 [9;15]	10 [7;14]	<0.001	754
CO [L/min]	3.3 [2.7;4.0]	4.7 [3.5;4.9]	4.0 [3.3;5.1]	<0.001	842
Acute NO challenge (vr)	0 (0)	0 (0)	46 (16)	<0.001	373
KCO [% pred.]	82 [74;94]	68 [67;71]	69 [49;84]	<0.001	546
Hb [g/l]	162 [151;173]	140 [132;153]	150 [136;163]	<0.001	702
Hct [l/l]	0.5 [0.5;0.5]	0.4 [0.4;0.4]	0.5 [0.4;0.5]	<0.001	575
WBC [x10e9/l]	9 [7;10]	7 [7;8]	8 [7;10]	0.013	696
Plt [x10e9/l]	211 [170;251]	225 [208;256]	228 [185;276]	0.079	693
ALP [IU/l]	75 [62;104]	78 [70;90]	87 [69;113]	0.01	675
Bilirubin [µmol/l]	18 [12;26]	8 [8;13]	14 [10;22]	0.001	671

None of the patients suffered from hereditary hemorrhagic telangiectasia. Data are presented as either median [interquartile range] or count (%).
Abbreviations: I/HPAH: idiopathic/hereditary pulmonary arterial hypertension; WHO FC: World Health Organization Functional Class; mPAP: mean pulmonary artery pressure; mPAWP: mean pulmonary artery wedge pressure; PVR: pulmonary vascular resistance; CO: cardiac output; NO challenge: Nitric Oxide challenge; KCO: transfer coefficient for carbon monoxide; Hb: hemoglobin; Hct: hematocrit; WBC: white blood cells; Plt: platelets; ALP: alkaline phosphatase; vr: vasoresponder; BMPR2: Bone Morphogenetic Protein Receptor Type 2; GDF2: Growth Differentiation Factor 2

3.4.2 Plasma BMP9 and pBMP10 levels in PAH patients

It was shown by Josh Hodgson, Department of Medicine, University of Cambridge, that both plasma BMP9 and pBMP10 levels are reduced in patients harbouring *GDF2* mutations. Subsequently Dr Paul Upton, Department of Medicine, University of Cambridge, measured BMP9 and PBMP10 levels in 120 control samples and 260 cases of heritable or idiopathic PAH. Clinical characteristics of patients and controls are shown in Table B.33.

Levels of BMP9 and pBMP10 were significantly higher in females than males in both control and PAH groups (Fig. 3.29). Levels of ligands were not associated with age (Fig. A.11). No significant differences were observed in the plasma levels of BMP9 between PAH patients and controls of the same sex (Fig. 3.29 A). Although no difference in pBMP10 levels was observed between control and PAH males, pBMP10 levels in females with PAH were significantly lower than control females (Fig. 3.29 B). A proportion of PAH patients exhibited



Legend: (A) Large deletions leading to loss of the *GDF2* allele and surrounding genes in two patients with PAH. The region of Chromosome 10 that is deleted is indicated by a red lined box. The turquoise boxes show the coverage of the two deletions and the red vertical line indicates the position of the *GDF2* gene. (B) Mapping the position of Tyrosine-351 on the BMP9 structure (violet). This tyrosine forms a hydrogen bond with methionine-421, an interaction which if disrupted by substitution in p.Tyr351His mutant may lead to destabilisation of the hydrophobic core. (C) Pedigree tree of proband harbouring missense variant in *GDF2* (p.Met89Val). WGS was performed only in the proband.

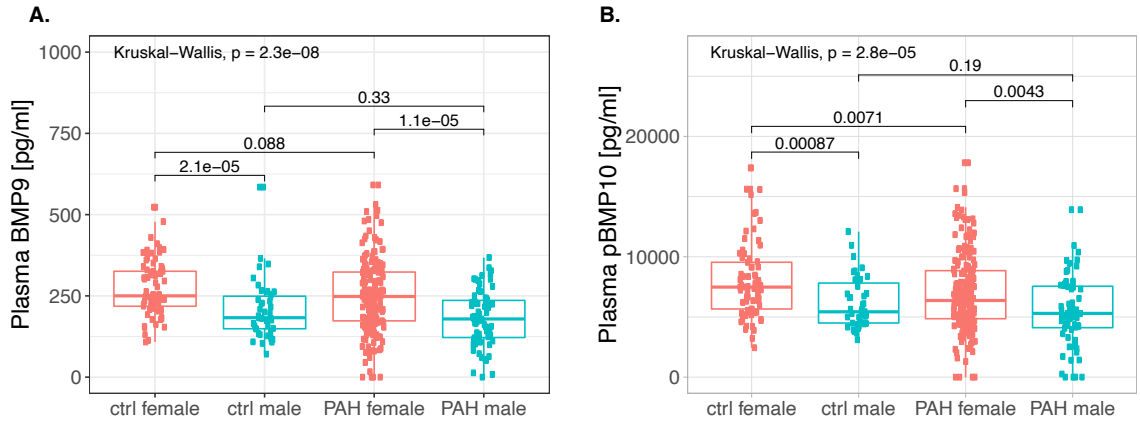
Fig. 3.28 Mutations in *GDF2*.

Table 3.31 Clinical phenotypes of BMP10 mutation carriers.

Variable	E001250	E010850
	Gene changes	
Nucleotide change	c.1082C>A	c.1057C>T
Protein change	p.Ala361Glu	p.Arg353Cys
Consequence	Missense variant	Missense variant
CADD	25.7	28.2
SIFT	deleterious	deleterious
PolyPhen-2	Probably damaging	Probably damaging
GERP	06.07	06.07
	Demographics and functional status	
Age at diagnosis [years]	72	28
Sex	male	female
Ethnic category	White and Black Caribbean	Other White
Genetic race	African	Other
WHO FC	III	IV
6MWD [m]	340	215
	Haemodynamics	
mRAP [mmHg]	5	6
mPAP [mmHg]	30	71
PAWP [mmHg]	8	8
PVR [mmHg]	16.5	18
CO [L/min]	1.33	3.5
	Clinical blood tests	
Hb [g/l]	106	
RDW [%]	17.6	
WBC x10e9/l	2.7	
Platelets x10e9/l	168	
Creatinine [μmol/l]	129	
BNP [ng/l]	171.8	220
ALP [IU/l]	92	
Albumin [g/l]	40	
ALT [IU/l]	26	
AST [IU/l]	27	
Bilirubin [µmol/l]	10	
Total protein [g/l]	74	
Cholesterol [mmol/l]	4.6	
HDL [mmol/l]	1.16	
LDL [mmol/l]	2.93	
TG [mmol/l]	1.12	
	Lung function tests	
FEV1 [% pred.]	82	67
FVC [% pred.]	82	65
FEV1/FVC	0.76	
KCO [% pred.]	92	
	Comorbidities	
	Systemic hypertension with LV hypertrophy, obstructive sleep apnoea, mild chronic kidney disease with renal cysts, cataract, benign prostatic enlargement	Diabetes mellitus type 2, systemic hypertension
	Family history	
Family members with PAH	No	No

Abbreviations: WHO FC: World Health Organization functional class; 6MWT: 6 minute walk test; mRAP: mean right atrial pressure; mPAP: mean pulmonary arterial pressure; PAWP : pulmonary artery wedge pressure; PVR: pulmonary vascular resistance; CO: cardiac output; Hb: hemoglobin, RDW: red cell distribution width; WBC: white blood cell count; BNP: Brain Natriuretic Peptide; ALP: alkaline phosphatase; ALT: alanine aminotransferase; AST: aspartate aminotransferase; HDL: high density lipoprotein; LDL: low density lipoprotein; TG: triglycerides; FEV1: forced expiratory volume; FVC: forced vital capacity; KCO: transfer coefficient for carbon monoxide.

reduced plasma levels of BMP9 (10% of males and 3.7% of females) or pBMP10 (14.9% of males and 4.81% of females) which were lower than the lowest values measured in healthy individuals.



Legend: Plasma BMP9 (bone morphogenetic protein type 9) and pBMP10 levels are not reduced in pulmonary arterial hypertension (PAH), but a subset of patients with PAH exhibit reduced plasma BMP9 and pBMP10 levels.

Fig. 3.29 Plasma samples collected from control subjects and patients with PAH were assayed for (A) BMP9 and (B) pBMP10.

There was a significant correlation between plasma BMP9 and pBMP10 in control subjects (Spearman r = 0.72, 95% CI [0.61;0.80] $p < 0.0001$) and in PAH patients (Spearman r = 0.75, 95% CI [0.68;0.80], p < 0.0001). This was also observed in an analysis of the combined samples (r = 0.72, 95% CI [0.67;0.76], $p < 0.001$) (Fig. 3.30).

Associations between BMP9 and BMP10 levels and clinical parameters

In PAH patients, BMP9 and pBMP10 levels were not associated with exercise capacity measured by 6MWT (r = -0.045, 95% CI [-0.22;0.13], *p*-value = 0.634 and r = 0.15, 95% CI [-0.032;0.32], *p*-value = 0.106 respectively). Haemodynamics contemporary to the time of sampling were available for 38 patients but did not show any correlation with either BMPs. Of note, BMP9 and pBMP10 levels negatively correlated with BMI in PAH cases but not in controls (Fig. 3.31).

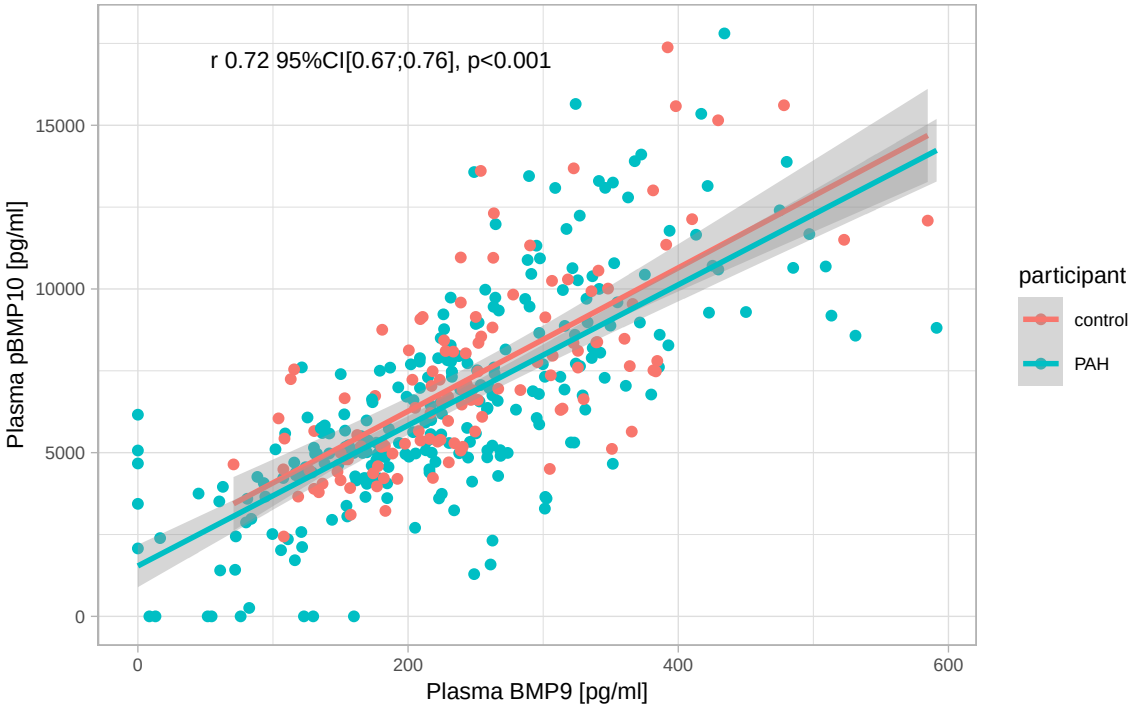


Fig. 3.30 Correlation between plasma pBMP10 and plasma BMP9 levels.

Analysis of clinical blood tests in the PAH cohort revealed that pBMP10 negatively correlated with red cell distribution width (RDW) (r = -0.21, 95% CI[-0.35;-0.06], p-value = 0.006) and alkaline phosphatase activity (r = -0.16, 95% CI[-0.28;-0.03], p-value = 0.014) and positively with albumin concentration (r = 0.32, 95% CI [0.20;0.43], p-value < 0.001), whereas BMP9 levels correlated positively with platelet count (r = 0.18, 95% CI[0.05;0.30], p-value = 0.006). Both BMP9 and pBMP10 correlated negatively with CRP (r=-0.32, 95% CI [-0.45;-0.18], p-value < 0.001; r = -0.35, 95% CI[-0.47;-0.20], p-value < 0.001 respectively) (also see Tables B.34 and B.35). Binary logistic regression showed that plasma pBMP10 concentrations were significantly negatively associated with systemic hypertension even after controlling for BMI and sex in the PAH cohort. When pBMP10 levels were in the lower tertile the log odds ratio of suffering from systemic hypertension were -1.30 95% CI [-2.21;-0.49]. Low pBMP10 concentrations were also predictive of diabetes mellitus type 2, but this association disappeared after controlling for BMI and sex. Receiver operating characteristics (ROC) curves (Fig. A.12) suggested that the cut-off values of 6784.7pg/ml (specificity 46.3%, sensitivity 79.2%, AUC 63.8%) and 6180.1pg/ml (specificity 51.9%, sensitivity 78.9%, AUC 66%) (data for PAH) were the most discriminative to predict systemic hypertension and diabetes mellitus respectively. Circulating BMP9 levels were not associated with either diabetes mellitus or systemic hypertension.

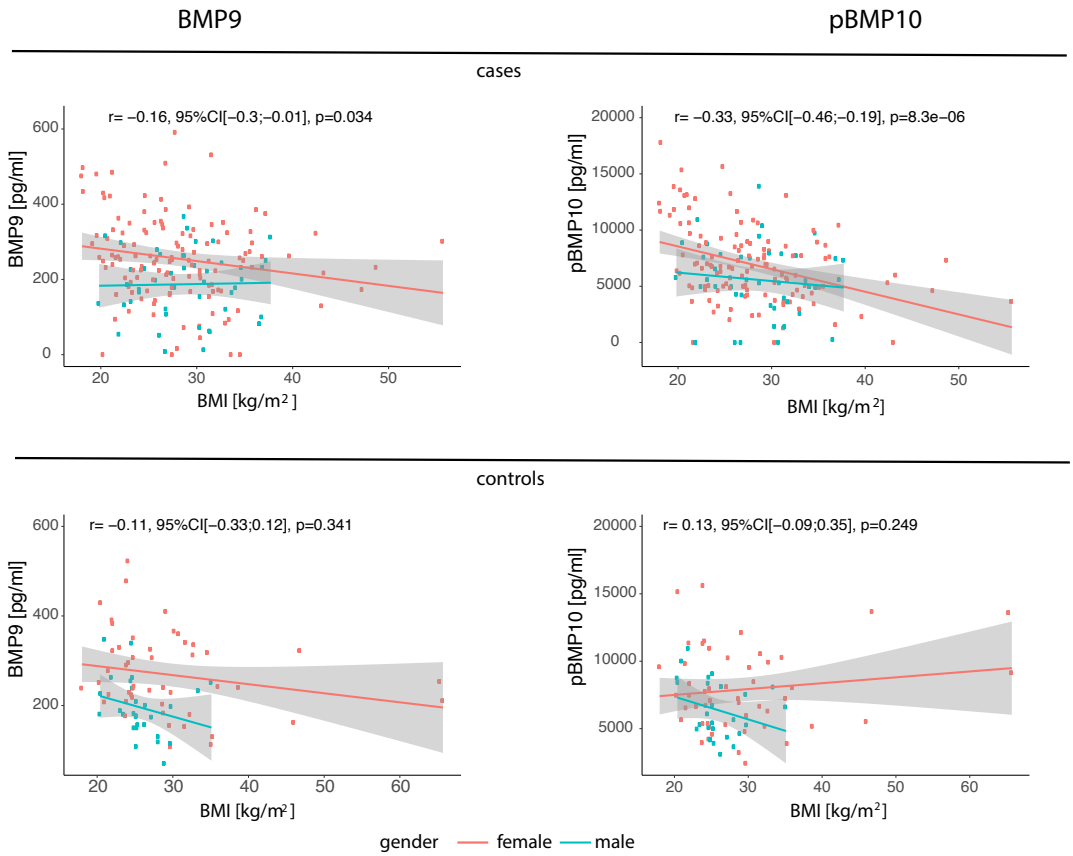


Fig. 3.31 Correlations between plasma BMP9 and pBMP10 levels and BMI in cases and controls.

During the median follow-up of 4.4 years since sampling, 48 (19%) PAH patients died, resulting in a 1- and 3-years overall survival of 95% and 86%, respectively. Neither BMP9 nor pBMP10 concentrations were predictive of mortality (Fig. 3.32).

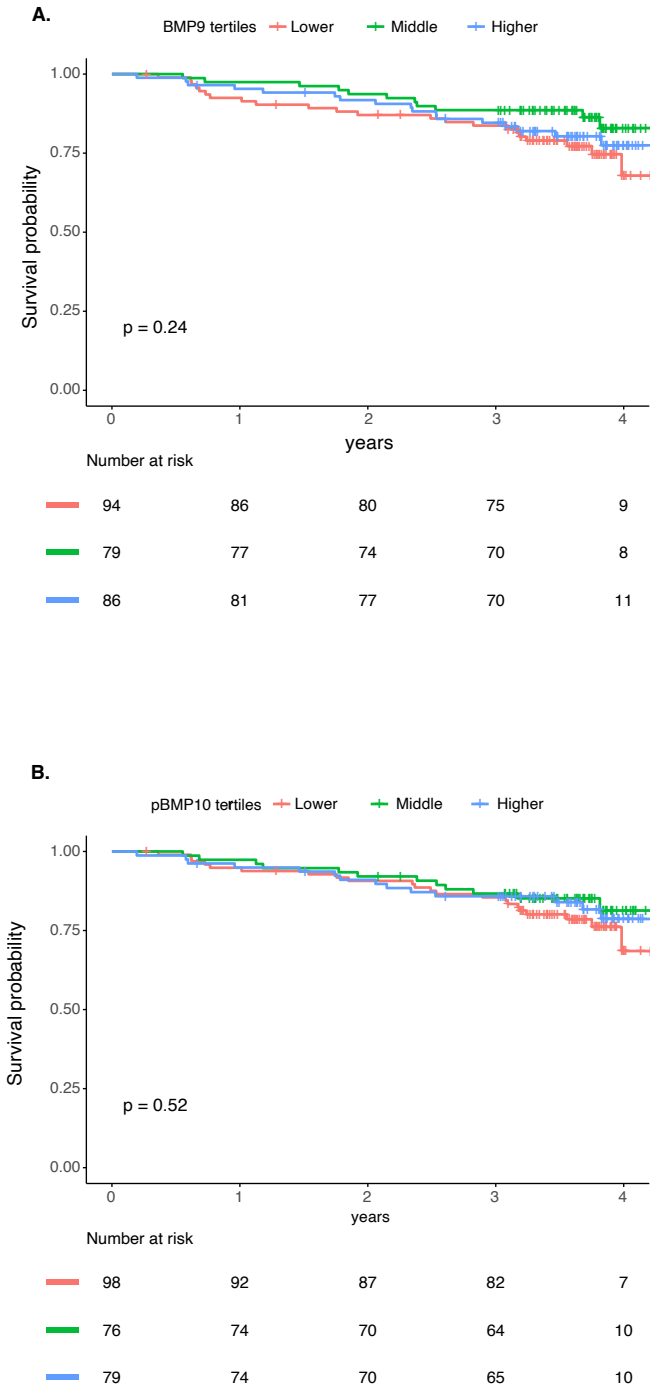
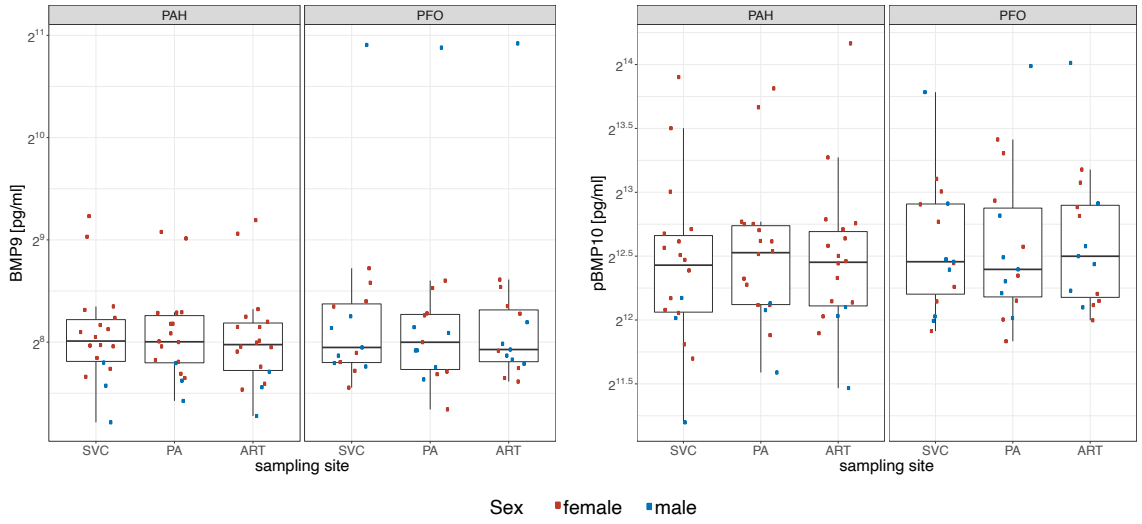


Fig. 3.32 Kapplan-Meier survival curves for plasma BMP9 (A) and pBMP10 levels (B).

Superior vena cava-pulmonary artery and pulmonary artery-radial artery gradients of BMP9 and pBMP10.

A separate group, consisting of 18 IPAH and 15 PFO patients was sampled during right heart catheterisation from three anatomical locations, namely superior vena cava (SVC), pulmonary artery (PA) and radial artery (ART), to check for differences in BMPs concentrations between sampling sites in pulmonary hypertension and patients with normal pulmonary pressures. The mutation status was unknown in this population. The comparison between IPAH and PFO patients did not reveal any differences in concentration or activity of BMPs irrespective of the sampling site (Table 3.32). Similarly to the previous analysis, there were significant differences in BMPs concentrations between males and females in the IPAH cohort but surprisingly, not in the comparator group (Table 3.31, Fig. 3.33).



Legend: Plasma BMP9 (bone morphogenetic protein type 9) (A) and pBMP10 (B) levels are lower in males with IPAH across sampling sites than in females, but there is no difference in either BMP levels between IPAH and PFO cohorts.

Fig. 3.33 Plasma gradients of BMP9 and pBMP10.

Table 3.32 Differences in BMP9 and pBMP10 levels and activity at various anatomical locations.

	IPAH N=18	**PFO N=15**	***p*-value**
Age at sampling [years]	44.0 [32.5;59.0]	45.0 [41.0;51.5]	0.664
BMI [kg/m²]	28.4 (4.96)	28.1 (4.32)	0.853
SVC BMP9 [pg/ml]	258 [225;298]	247 [223;332]	0.8
PA BMP9 [pg/ml]	257 [223;307]	256 [213;309]	0.942
ART BMP9 [pg/ml]	252 [211;291]	243 [224;319]	0.638
SVC pBMP10 [pg/ml]	5518 [4275;6479]	5620 [4716;7688]	0.47
PA pBMP10 [pg/ml]	5903 [4454;6834]	5391 [4644;7524]	0.857
ART pBMP10 [pg/ml]	5604 [4422;6621]	5790 [4632;7635]	0.406
SVC-PA BMP9 gradient	4.24 (18.7)	13.8 (19.1)	0.157
PA-ART BMP9 gradient	7.43 [-5.91;15.6]	-2.76 [-15.80;6.78]	0.096
SVC-PA pBMP10 gradient	-280.26 (724)	-246.31 (1265)	0.927
PA-ART pBMMP10 gradient	112 [-57.04;282]	-154.02 [-467.53;317]	0.112
SVC BMP activity	2104 [1592;2356]	3657 [1852;5596]	0.055
PA BMP activity	2355 [1700;3003]	2658 [1858;4524]	0.406
ART BMP activity	1796 [1495;2574]	2381 [1486;3877]	0.47
SVC-PA BMP activity gradient	-461.12 [-1227.21;555]	-104.01 [-343.17;727]	0.104
PA-ART BMP activity gradient	7.43 [-5.91;15.6]	-2.76 [-15.80;6.78]	0.096

Abbreviations: IPAH: idiopathic pulmonary arterial hypertension; PFO: patent foramen ovale; SVC: superior vena cava; PA: pulmonary artery; ART: radial artery; BMI: body mass index

3.4.3 Discussion

Our original finding of the enrichment of PAH cohort for rare, heterozygous, predicted deleterious variants in *GDF2* has been independently validated in other cohorts [564, 128, 549]. Additionally, Wang et al. published a case report of childhood-onset PAH in an individual with homozygous protein-truncating *GDF2* variant (c.76C>T; p.Gln26Ter) [561]. In the Chinese PAH cohort, the authors measured plasma BMP9 by ELISA in 19 PAH patients harbouring *GDF2* mutations and age- and sex-matched patients with IPAH and healthy control subjects [564]. They found that median plasma BMP9 levels were significantly reduced in patients with idiopathic PAH and were lowest in patients carrying *GDF2* mutations. On the contrary, our analysis on much larger cohort of control subjects and patients with idiopathic PAH, did not recapitulate these findings, although, we observed that some patients

with idiopathic PAH had very low BMP9 levels, unexplained by the presence of mutations. Interestingly, a recent study found that patients with portopulmonary hypertension have profound reductions in plasma BMP9 and that low levels predict the PAH presence in patients with cirrhosis [382]. Of note, presence of liver disease was excluded in the patients and controls recruited to our cohort. In summary, the genetic and nongenetic evidence in humans suggests that reduction of BMP9 levels, or BMP9 signalling, is relevant in the pathogenesis of PAH. The observed difference between plasma levels of BMP9 and pBMP10 in males and females is of interest, and the underlying mechanism for this difference warrants further study. Although higher levels in females superficially conflict with the observation that idiopathic and heritable PAH more commonly affects females, the outcomes of PAH are worse in male patients.

This is the first study to measure plasma levels of pBMP10 in patients with PAH, or indeed in a large cohort of control subjects. Overall, levels of pBMP10 were significantly lower in female patients with IPAH compared with control subjects. Remarkably, there was a close correlation between the levels of BMP9 and pBMP10 in human plasma, suggesting a degree of co-regulation that might be explained, at least in part, by the presence of circulating heterodimers of these ligands [528]. This is important because *GDF2* mutations leading to reduced secretion of ligand would be predicted to impact the circulating levels of BMP10, as well as BMP9, as observed by Hodgson J et al. [215]. While the fact that BMP9 and pBMP10 form a complex was confirmed by immunoprecipitation, much higher levels of pBMP10 than BMP9 suggests that this association is not 1:1. Additionally, only about half of the plasma BMP9 and pBMP10 were in the complex, implying that BMP9:10 heterodimers could not represent the only active form [215]. The potentially important role of BMP10 in PAH was further supported by finding two individuals with likely pathogenic missense mutations in BMP10, providing independent validation of the previous report implicating BMP10 mutations in pathogenesis of PAH [128]. Unfortunately, plasma was not available from these individuals for further analysis.

While low levels of circulating BMP9 were previously found to be associated with markers of metabolic syndrome [581] and coronary artery disease [315], we found a negative correlation between BMP9 and pBMP10 levels and BMI and CRP. Additionally, pBMP10 level was an independent predictor of systemic hypertension. These findings are indicative of the role of BMP9 and pBMP10 in cardiovascular health beyond their role in pulmonary hypertension.

The human data presented here are important as they are in striking contrast to a recent study in rodents, suggesting that inhibition of BMP9 signalling might be protective against

the development of pulmonary hypertension [534]. In the study by Tu et al. *Gdf2* knockout mice not only did not develop spontaneous PH but were protected from the development of hypoxia-induced PH. This was further corroborated by an experiment showing that the inhibition of Bmp9 signalling with a neutralising anti-BMP9 antibody could partially prevent the development of PH during chronic hypoxia. This outcome is contrary to that of another study showing that inhibition of BMP9 and BMP10 with Fc-ALK1 exacerbated pulmonary hypertension during chronic hypoxia [382]. Taken together, the human genetic evidence [179, 564, 561, 128] and the data from patients with cirrhosis [382] support the view that BMP9 and BMP10 are protective to the pulmonary vascular endothelium, consistent with the beneficial effects of therapeutic supplementation of BMP9 in preclinical models.

As BMP9 is mainly expressed in the liver and pBMP10 primarily in the right atrium, I set out to analyse whether gradients of these ligands existed across three sampling sites, namely the SVC, PA and radial artery. Interestingly, no differences in ligand levels were observed, neither in cases nor in the comparator group.

Heterozygous *GDF2* mutations have been previously implicated in the pathogenesis of vascular anomaly syndrome resembling HHT [575, 210, 529] in small number of patients. Recently, an additional two cases of homozygous *GDF2* variant were described, one in association with childhood-onset HHT [314] and another with lymphatic dysplasia, hydrothorax and nonimmune hydrops fetalis [22]. The mutations in patients with HHT were distinct from those reported in PAH but were also predicted be deleterious. The finding of mutations in *GDF2* in both HHT-like syndromes and PAH is in keeping with the pathogenic role of the mutations in the endothelial receptor for BMP9, *ALK1* in HHT and PAH [531]. Of note, none of the patients carrying *GDF2* mutations in the present study had clinical features of HHT. Unexpectedly, patients harbouring *GDF2* variants did not share clinical phenotype with *BMPR2* mutation carriers conversely, they were older and had milder haemodynamics and were more alike patients with no mutations in PAH risk genes. These findings are in keeping with *GDF2* carrier characteristics reported by Wang et al. [564]. Moreover, measurement of the strength of semantic similarity revealed that *GDF2* carriers do not carry a distinct phenotype. There are several potential explanations for this finding. Firstly, the relevant phenotypes may not have been captured, and a more in-depth search of phenotypes close to genetic underpinnings of the disease is needed. Secondly, it is likely that while patients harbouring heterozygous variants do not present with distinct phenotype, those with homozygous variants are characterised by specific features such as, for example, "HHT-like" syndrome [214]. Alternatively, a cross-talk between various pathways implicated in the

pathogenesis of PAH and cardiovascular health in general [511] may be responsible for the lack of noticeable phenotype.

3.4.4 Conclusions

The present analysis demonstrates that rare heterozygous mutations in *GDF2* are loss of function and likely causal in the pathobiology of PAH. *GDF2* mutation carriers are clinically different from *BMPR2* mutation carriers but do not carry a distinct phenotype. Furthermore, *GDF2* mutations lead to reduced circulating levels of BMP10, as well as BMP9, supporting a degree of co-regulation of these ligands. Circulating levels of BMP9 and pBMP10 in IPAH and controls are significantly lower in males than in females but only pBMP10 levels are lower female patients than in control subjects. Both ligands show negative correlation with BMI and CRP, while pBMP10 levels are an independent predictor of systemic hypertension, suggestive of the role of BMPs in cardiovascular health. Taken together, these findings further support the central role of the BMP9/BMPR2/ALK1 axis in PAH and cardiovascular health in general.

In Chapter 3.4 I addressed aim number 4 of my Thesis.

Chapter 4

Final remarks

In this chapter I summarise my conclusions and comment on limitations and challenges of this and similar genetic studies. Finally, I outline my future research plans.

4.1 Conclusions

In my PhD project I set out to collect, harmonise and verify deep phenotype data of patients with PAH recruited to the NBR and the PAH Cohort studies as well as the US PAHBB study; these data were indispensable for my research project [503, 215], but also for multiple projects I co-authored [541, 542, 478, 436]. The second aim of this thesis was to enrich the Human Phenotype Ontology database for terms relevant to PAH and associated diseases, which would enable discovery of ontology-based phenotypes, and improve ranking differential diagnosis of rare pulmonary diseases based on phenotypes. I have added and annotated over 250 HPO terms, which appeared in the latest HPO ontology release (February 2021), the manuscript demonstrating the improved ranking of possible diagnoses is in preparation.

The current study's main goal was to determine clinical and computational phenotypes of patients with PAH that would carry diagnostic, therapeutic and prognostic information as well as allow genotype-phenotype association analysis in enriched disease cohorts. To this end, I established clinical and computational phenotypes meeting the above-mentioned criteria, and showed that unsupervised computational phenotyping might be as accurate and significant as clinically derived phenotypes, of which extraction is labour intensive and requires domain knowledge. As hypothesised, new homogeneous clinical and computational PAH clusters proved to be enriched for rare deleterious disease-causing variants, led to the discovery of *KDR* as a new PAH risk gene [503, 179], and highlighted a new PAH candidate gene (*COL6A5*). Further reverse phenotyping analysis of patients harbouring protein-truncating

variants in *KDR* showed presence of parenchymal abnormalities in all index cases and in an affected relative. This finding was subsequently corroborated by histological analysis of wedge lung biopsy in an affected relative. I now lead the study into the disease trajectory of *KDR* mutation carriers, which should help to improve predictions of the impact of protein-truncating variants in *KDR* on components of the disease phenotype. Further research in the causal role of *COL6A5* and other new candidate genes will be a focus of my postdoctoral studies.

This thesis adds to the rapidly expanding field of genetics and genomics in PAH by contributing a wealth of phenotypic and genomic data, endorsing new tools to analyse and interpret them and proposing new PAH risk genes.

4.2 Limitations, challenges and future directions

The primary limitations in PAH genetic studies originate from either their study design or the genetic methods used. The former is mainly limited by sample size and population structure and features in GWAS and RVAS alike, and the latter is impacted by sequencing technology, genome assembly and analysis pipelines. Additionally, a variable quality of reporting may obscure the validity of the results. Several approaches may accelerate causal variant discovery. Firstly, international efforts to set up platforms and governance, thereby reducing the barriers to genotype and phenotype data harmonisation, across studies and countries, along with regulatory standards, compatible with the international legislative landscape, aid variant discovery. To this end, I verified and harmonised clinical data from the two largest PAH cohorts that have also been sequenced, which are now used for combined analysis [596] and I am participating in the International Consortium for Genetic Studies in PAH (PAH-ICON) Task Force on Phenotyping which aims at phenotypic data collection and harmonisation for genetic studies across the world as well as works on White Paper on phenotyping patients with pulmonary hypertension. Secondly, the application of computational and/or intermediate phenotypes may further increase power to detect new risk genes, contribution of digenic and oligogenic inheritance, and epigenetic changes. Likewise, researching isolated populations resulting from recent bottlenecks (i.e., Icelanders, Ashkenazi Jews) have shown promising results in other rare diseases. As for genetic methods, long-read technologies will improve significantly, and with improvements to the error rate, long-read alignment to a pan-genome graph may identify additional genetic variants. Expansion of allele frequency estimation resources (i.e., gnomAD), encompassing diverse ethnicities and variant classes, including structural variants, will improve variant interpretation. Studying gene sets that are likely to

be enriched for disease-associated loci (i.e., implicated by GWAS, or expressed in disease-relevant tissues) and extending RVAS to non-coding regions (a natural step given GWAS findings) is likely to increase discovery yield further.

Reverse genetics techniques are vital components of PAH research, without which forward genetics discoveries cannot be validated. Nonetheless, such studies must be interrogated for any limitations that may affect their results. A robust experimental design and protocol, which includes power calculations for animal studies, minimising bias and reporting all measurements in publications are crucial for successful study. Cellular models are potent *in vitro* tools for understanding specific PAH processes; however, focusing on certain cell types may produce more pronounced effects than those seen *in vivo*. Hence understanding the widespread impact of PAH mutations are best captured using animal models. Still, it is essential to remember that all currently used animal models fail to recapitulate the disease fully and that any toxic side effects of pharmacological models may confound the results. Studies must choose models which are most appropriate for the hypothesis or use multiple where necessary. In the coming years new tools may increase the speed and depth of information derived using reverse genetics; these include high-throughput technologies for mutagenesis and spatial gene expression analysis [553, 554]. Such developments offer exciting prospects for future research in PAH.

Reverse phenotyping has the potential to shed light on genotype–phenotype associations and accelerate the recruitment of homogenous, well-defined groups to clinical trials, thereby increasing the chances of effective treatment and decreasing the risk of side effects. Its success depends on the quality of forward phenotyping, and forward and reverse genetic studies accurately reporting the phenotype of interest (use of standardised vocabulary to describe phenotypic abnormalities (i.e., HPO terms), clinical measurements (i.e., Logical Observation Identifiers (LOINC) [494]) and systematic approaches to data sharing. With that in mind, I set out to examine disease trajectory in patients harbouring deleterious PTV in *KDR* and I am co-authoring a project on reverse genetics and phenotyping of *TBX4* mutation carriers.

In the future, national and international efforts should be directed at optimising recruitment of large homogeneous cohorts for forward genetics studies and, simultaneously, preempting any potential ethico-legal considerations related to subsequent recall by genotype studies. Similarly, recall by genotype studies will face challenges related to specific study design, i.e., when recruiting newborns or children who might not express the full phenotype by the time of diagnosis/inclusion.

References

[1] (1995). American thoracic society. single-breath carbon monoxide diffusing capacity (transfer factor). recommendations for a standard technique–1995 update. *American Journal of Respiratory and Critical Care Medicine*, 152(6):2185–2198.

[2] Abraham, W. T., Raynolds, M. V., Badesch, D. B., Wynne, K. M., Groves, B. M., Roden, R. L., Robertson, A. D., Lowes, B. D., Zisman, L. S., Voelkel, N. F., Bristow, M. R., and Perryman, M. B. (2003). Angiotensin-converting enzyme DD genotype in patients with primary pulmonary hypertension: increased frequency and association with preserved haemodynamics. *Journal of the Renin-Angiotensin-Aldosterone System*, 4(1):27–30.

[3] Adir, Y., Elia, D., and Harari, S. (2015). Pulmonary hypertension in patients with chronic myeloproliferative disorders. *European Respiratory Review*, 24(137):400–410.

[4] Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., Kondrashov, A. S., and Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nat. Methods*, 7(4):248–249.

[5] Aggarwal, R., Lucas, M., Fertig, N., Oddis, C. V., and Medsger, T. A. (2009). Anti-u3 RNP autoantibodies in systemic sclerosis. *Arthritis & Rheumatism*, 60(4):1112–1118.

[6] Akazawa, Y., Okumura, K., Ishii, R., Slorach, C., Hui, W., Ide, H., Honjo, O., Sun, M., Kabir, G., Connelly, K., and Friedberg, M. K. (2020). Pulmonary artery banding is a relevant model to study the right ventricular remodeling and dysfunction that occurs in pulmonary arterial hypertension. *J. Appl. Physiol.*, 129(2):238–246.

[7] Akhavein, F., St-Michel, E. J., Seifert, E., and Rohlicek, C. V. (2007). Decreased left ventricular function, myocarditis, and coronary arteriolar medial thickening following monocrotaline administration in adult rats. *J. Appl. Physiol.*, 103(1):287–295.

[8] Alaa, A. M., Bolton, T., Di Angelantonio, E., Rudd, J. H. F., and van der Schaar, M. (2019). Cardiovascular disease risk prediction using automated machine learning: A prospective study of 423,604 UK biobank participants. *PLoS One*, 14(5):e0213653.

[9] Aldred, M. A., Comhair, S. A., Varella-Garcia, M., Asosingh, K., Xu, W., Noon, G. P., Thistlethwaite, P. A., Tuder, R. M., Erzurum, S. C., Geraci, M. W., and Coldren, C. D. (2010). Somatic chromosome abnormalities in the lungs of patients with pulmonary arterial hypertension. *American Journal of Respiratory and Critical Care Medicine*, 182(9):1153–1160.

[10] Allcock, R. J. N. (2012). The major histocompatibility complex: a paradigm for studies of the human genome. *Methods Mol. Biol.*, 882:1–7.

[11] Alperin, J. M., Ortiz-Fernández, L., and Sawalha, A. H. (2018). Monogenic lupus: A developing paradigm of disease. *Front. Immunol.*, 9:2496.

[12] Alves, J. L., Gavilanes, F., Jardim, C., dos Santos Fernandes, C. J. C., Morinaga, L. T. K., Dias, B., Hoette, S., Humbert, M., and Souza, R. (2015). Pulmonary arterial hypertension in the southern hemisphere. *Chest*, 147(2):495–501.

[13] Amanat, S., Requena, T., and Lopez-Escamez, J. A. (2020). A systematic review of extreme phenotype strategies to search for rare variants in genetic studies of complex disorders. *Genes*, 11(9):987.

[14] Anaya, J.-M., Corena, R., Castiblanco, J., Rojas-Villarraga, A., and Shoenfeld, Y. (2007). The kaleidoscope of autoimmunity: multiple autoimmune syndromes and familial autoimmunity. *Expert Rev. Clin. Immunol.*, 3(4):623–635.

[15] Anaya, J.-M., Rojas-Villarraga, A., and García-Carrasco, M. (2012). The autoimmune tautology: from polyautoimmunity and familial autoimmunity to the autoimmune genes. *Autoimmune Dis.*, 2012:297193.

[16] Anthony, T. G., McDaniel, B. J., Byerley, R. L., McGrath, B. C., Cavener, D. R., McNurlan, M. A., and Wek, R. C. (2004). Preservation of liver protein synthesis during dietary leucine deprivation occurs at the expense of skeletal muscle mass in mice deleted for eIF2 kinase GCN2. *J. Biol. Chem.*, 279(35):36553–36561.

[17] Archer, S. L., Marsboom, G., Kim, G. H., Zhang, H. J., Toth, P. T., Svensson, E. C., Dyck, J. R., Gomberg-Maitland, M., Thebaud, B., Husain, A. N., Cipriani, N., and Rehman, J. (2010). Epigenetic attenuation of mitochondrial superoxide dismutase 2 in pulmonary arterial hypertension. *Circulation*, 121(24):2661–2671.

[18] Arora, R., Metzger, R. J., and Papaioannou, V. E. (2012). Multiple roles and interactions of tbx4 and tbx5 in development of the respiratory system. *PLoS Genet.*, 8(8):e1002866.

[19] Asai, K., Kudej, R. K., Shen, Y. T., Yang, G. P., Takagi, G., Kudej, A. B., Geng, Y. J., Sato, N., Nazareno, J. B., Vatner, D. E., Natividad, F., Bishop, S. P., and Vatner, S. F. (2000). Peripheral vascular endothelial dysfunction and apoptosis in old monkeys. *Arterioscler. Thromb. Vasc. Biol.*, 20(6):1493–1499.

[20] Asimit, J. L., Day-Williams, A. G., Morris, A. P., and Zeggini, E. (2012). ARIEL and AMELIA: testing for an accumulation of rare variants using next-generation sequencing data. *Hum. Hered.*, 73(2):84–94.

[21] Atkinson, C., Stewart, S., Upton, P. D., Machado, R., Thomson, J. R., Trembath, R. C., and Morrell, N. W. (2002). Primary pulmonary hypertension is associated with reduced pulmonary vascular expression of type II bone morphogenetic protein receptor. *Circulation*, 105(14):1672–1678.

[22] Aukema, S. M., Ten Brinke, G. A., Timens, W., Vos, Y. J., Accord, R. E., Kraft, K. E., Santing, M. J., Morssink, L. P., Streefland, E., van Diemen, C. C., Vrijlandt, E. J., Hulzebos, C. V., and Kerstjens-Frederikse, W. S. (2020). A homozygous variant in growth and differentiation factor 2 (GDF2) may cause lymphatic dysplasia with hydrothorax and nonimmune hydrops fetalis. *Am. J. Med. Genet. A*, 182(9):2152–2160.

[23] Austin, E. D., Cogan, J. D., West, J. D., Hedges, L. K., Hamid, R., Dawson, E. P., Wheeler, L. A., Parl, F. F., Loyd, J. E., and Phillips, 3rd, J. A. (2009a). Alterations in oestrogen metabolism: implications for higher penetrance of familial pulmonary arterial hypertension in females. *Eur. Respir. J.*, 34(5):1093–1099.

[24] Austin, E. D., Hamid, R., Hemnes, A. R., Loyd, J. E., Blackwell, T., Yu, C., Phillips, Iii, J. A., Gaddipati, R., Gladson, S., Gu, E., West, J., and Lane, K. B. (2012a). BMPR2 expression is suppressed by signaling through the estrogen receptor. *Biol. Sex Differ.*, 3(1):6.

[25] Austin, E. D., Ma, L., LeDuc, C., Berman Rosenzweig, E., Borczuk, A., Phillips, 3rd, J. A., Palomero, T., Sumazin, P., Kim, H. R., Talati, M. H., West, J., Loyd, J. E., and Chung, W. K. (2012b). Whole exome sequencing to identify a novel gene (caveolin-1) associated with human pulmonary arterial hypertension. *Circ. Cardiovasc. Genet.*, 5(3):336–343.

[26] Austin, E. D., Phillips, J. A., Cogan, J. D., Hamid, R., Yu, C., Stanton, K. C., Phillips, C. A., Wheeler, L. A., Robbins, I. M., Newman, J. H., and Loyd, J. E. (2009b). Truncating and missense BMPR2 mutations differentially affect the severity of heritable pulmonary arterial hypertension. *Respiratory Research*, 10(1).

[27] Awad, M. and Khanna, R. (2015). Machine learning. In *Efficient Learning Machines*, pages 1–18. Apress, Berkeley, CA.

[28] Babenko, A. P., Polak, M., Cavé, H., Busiah, K., Czernichow, P., Scharfmann, R., Bryan, J., Aguilar-Bryan, L., Vaxillaire, M., and Froguel, P. (2006). Activating mutations in the ABCC8 gene in neonatal diabetes mellitus. *N. Engl. J. Med.*, 355(5):456–466.

[29] Badesch, D., Gibbs, S., Gomberg-Maitland, M., Humbert, M., Mclaughlin, V., Preston, I., Souza, R., Waxman, A., De Oliveira Pena, J., Barnes, J., Zhang, X., and Zeldin, R. K. (2019). PULSAR: A phase 2, randomized, double-blind, placebo-controlled study to assess the efficacy and safety of sotatercept (ACE-011) when added to standard of care (SOC) for treatment of pulmonary arterial hypertension (PAH). In *Pulmonary hypertension*, page PA4750. European Respiratory Society.

[30] Ballif, B. C., Theisen, A., Rosenfeld, J. A., Traylor, R. N., Gastier-Foster, J., Thrush, D. L., Astbury, C., Bartholomew, D., McBride, K. L., Pyatt, R. E., Shane, K., Smith, W. E., Banks, V., Gallentine, W. B., Brock, P., Rudd, M. K., Adam, M. P., Keene, J. A., Phillips, 3rd, J. A., Pfotenhauer, J. P., Gowans, G. C., Stankiewicz, P., Bejjani, B. A., and Shaffer, L. G. (2010). Identification of a recurrent microdeletion at 17q23.1q23.2 flanked by segmental duplications associated with heart defects and limb abnormalities. *Am. J. Hum. Genet.*, 86(3):454–461.

[31] Ballouz, S., Dobin, A., and Gillis, J. A. (2019). Is it time to change the reference genome? *Genome Biol.*, 20(1):159.

[32] Baloira Villar, A., Valverde, D., and Lago, M. (2019). Functional study of polymorphisms in the promoter region of the endothelin-1 gene in pulmonary arterial hypertension. In *Pulmonary hypertension*, page PA5043. European Respiratory Society.

[33] Barker, D. (1986). INFANT MORTALITY, CHILDHOOD NUTRITION, AND ISCHAEMIC HEART DISEASE IN ENGLAND AND WALES. *Lancet*, 327(8489):1077–1081.

[34] Barker, D. J. P., Osmond, C., Winter, P. D., Margetts, B., and Simmonds, S. J. (1989). WEIGHT IN INFANCY AND DEATH FROM ISCHAEMIC HEART DISEASE. *Lancet*, 334(8663):577–580.

[35] Barozzi, C., Galletti, M., Tomasi, L., De Fanti, S., Palazzini, M., Manes, A., Sazzini, M., and Galiè, N. (2019). A combined targeted and whole exome sequencing approach identified novel candidate genes involved in heritable pulmonary arterial hypertension. *Sci. Rep.*, 9(1):753.

[36] Basson, C. T., Bachinsky, D. R., Lin, R. C., Levi, T., Elkins, J. A., Soults, J., Grayzel, D., Kroumpouzou, E., Traill, T. A., Leblanc-Straceski, J., Renault, B., Kucherlapati, R., Seidman, J. G., and Seidman, C. E. (1997). Mutations in human TBX5 [corrected] cause limb and cardiac malformation in Holt-Oram syndrome. *Nat. Genet.*, 15(1):30–35.

[37] Bauer, S., Köhler, S., Schulz, M. H., and Robinson, P. N. (2012). Bayesian ontology querying for accurate and noise-tolerant semantic searches. *Bioinformatics*, 28(19):2502–2508.

[38] Bean, L. J. H., on behalf of the ACMG Laboratory Quality Assurance Committee, Funke, B., Carlston, C. M., Gannon, J. L., Kantarci, S., Krock, B. L., Zhang, S., and Bayrak-Toydemir, P. (2020). Diagnostic gene sequencing panels: from design to report—a technical standard of the american college of medical genetics and genomics (ACMG).

[39] Becker, M. O., Kill, A., Kutsche, M., Guenther, J., Rose, A., Tabeling, C., Witzenrath, M., Kühl, A. A., Heidecke, H., Ghofrani, H. A., Tiede, H., Schermuly, R. T., Nickel, N., Hoeper, M. M., Lukitsch, I., Gollasch, M., Kuebler, W. M., Bock, S., Burmester, G. R., Dragun, D., and Riemekasten, G. (2014). Vascular receptor autoantibodies in pulmonary arterial hypertension associated with systemic sclerosis. *Am. J. Respir. Crit. Care Med.*, 190(7):808–817.

[40] Bell, J. T., Pai, A. A., Pickrell, J. K., Gaffney, D. J., Pique-Regi, R., Degner, J. F., Gilad, Y., and Pritchard, J. K. (2011). DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. *Genome Biol.*, 12(1):R10.

[41] Belvisi, M. G., Stretton, C. D., Yacoub, M., and Barnes, P. J. (1992). Nitric oxide is the endogenous neurotransmitter of bronchodilator nerves in humans. *Eur. J. Pharmacol.*, 210(2):221–222.

[42] Benza, R. L., Gomberg-Maitland, M., Demarco, T., Frost, A. E., Torbicki, A., Langleben, D., Pulido, T., Correa-Jaque, P., Passineau, M. J., Wiener, H. W., Tamari, M., Hirota, T., Kubo, M., and Tiwari, H. K. (2015). Endothelin-1 pathway polymorphisms and outcomes in pulmonary arterial hypertension. *American Journal of Respiratory and Critical Care Medicine*, 192(11):1345–1354.

[43] Benza, R. L., Gomberg-Maitland, M., Miller, D. P., Frost, A., Frantz, R. P., Foreman, A. J., Badesch, D. B., and McGoon, M. D. (2012). The REVEAL registry risk score calculator in patients newly diagnosed with pulmonary arterial hypertension. *Chest*, 141(2):354–362.

[44] Beppu, H., Ichinose, F., Kawai, N., Jones, R. C., Yu, P. B., Zapol, W. M., Miyazono, K., Li, E., and Bloch, K. D. (2004). BMPR-II heterozygous mice have mild pulmonary hypertension and an impaired pulmonary vascular remodeling response to prolonged hypoxia. *American Journal of Physiology-Lung Cellular and Molecular Physiology*, 287(6):L1241–L1247.

[45] Berry, C. E., Billheimer, D., Jenkins, I. C., Lu, Z. J., Stern, D. A., Gerald, L. B., Carr, T. F., Guerra, S., Morgan, W. J., Wright, A. L., and Martinez, F. D. (2016). A distinct low lung function trajectory from childhood to the fourth decade of life. *Am. J. Respir. Crit. Care Med.*, 194(5):607–612.

[46] Best, D. H., Sumner, K. L., Smith, B. P., Damjanovich-Colmenares, K., Nakayama, I., Brown, L. M., Ha, Y., Paul, E., Morris, A., Jama, M. A., Dodson, M. W., Bayrak-Toydemir, P., and Elliott, C. G. (2017). EIF2AK4 mutations in patients diagnosed with pulmonary arterial hypertension. *Chest*, 151(4):821–828.

[47] Best, D. H., Vaughn, C., McDonald, J., Damjanovich, K., Runo, J. R., Chibuk, J. M., and Bayrak-Toydemir, P. (2011). Mosaic ACVRL1 and ENG mutations in hereditary haemorrhagic telangiectasia patients. *Journal of Medical Genetics*, 48(5):358–360.

[48] Bhaskaran, K. and Smeeth, L. (2014). What is the difference between missing completely at random and missing at random? *Int. J. Epidemiol.*, 43(4):1336–1339.

[49] Bleyl, S., Nelson, L., Odelberg, S. J., Ruttenberg, H. D., Otterud, B., Leppert, M., and Ward, K. (1995). A gene for familial total anomalous pulmonary venous return maps to chromosome 4p13-q12. *Am. J. Hum. Genet.*, 56(2):408–415.

[50] Bodenhofer, U., Kothmeier, A., and Hochreiter, S. (2011). Apcluster: an r package for affinity propagation clustering. *Bioinformatics*, 27:2463–2464.

[51] Bohnen, M. S., Ma, L., Zhu, N., Qi, H., McClenaghan, C., Gonzaga-Jauregui, C., Dewey, F. E., Overton, J. D., Reid, J. G., Shuldiner, A. R., Baras, A., Sampson, K. J., Bleda, M., Hadinnapola, C., Haimel, M., Bogaard, H. J., Church, C., Coghlan, G., Corris, P. A., Eyries, M., Gibbs, J. S. R., Girerd, B., Houweling, A. C., Humbert, M., Guignabert, C., Kiely, D. G., Lawrie, A., MacKenzie Ross, R. V., Martin, J. M., Montani, D., Peacock, A. J., Pepke-Zaba, J., Soubrier, F., Suntharalingam, J., Toshner, M., Treacy, C. M., Trembath, R. C., Vonk Noordegraaf, A., Wharton, J., Wilkins, M. R., Wort, S. J., Yates, K., Gräf, S., Morrell, N. W., Krishnan, U., Rosenzweig, E. B., Shen, Y., Nichols, C. G., Kass, R. S., and Chung, W. K. (2018). Loss-of-Function ABCC8 mutations in pulmonary arterial hypertension. *Circ Genom Precis Med*, 11(10):e002087.

[52] Bohnen, M. S., Roman-Campos, D., Terrenoire, C., Jnani, J., Sampson, K. J., Chung, W. K., and Kass, R. S. (2017). The impact of heterozygous KCNK3 mutations associated with pulmonary arterial hypertension on channel function and pharmacological recovery. *Journal of the American Heart Association*, 6(9).

[53] Bongers, E. M. H. F., Duijf, P. H. G., van Beersum, S. E. M., Schoots, J., Van Kampen, A., Burckhardt, A., Hamel, B. C. J., Losan, F., Hoefsloot, L. H., Yntema, H. G., Knoers, N. V. A. M., and van Bokhoven, H. (2004). Mutations in the human TBX4 gene cause small patella syndrome. *Am. J. Hum. Genet.*, 74(6):1239–1248.

[54] Bonnet, S., Michelakis, E. D., Porter, C. J., Andrade-Navarro, M. A., The baud, B., Bonnet, S., Haromy, A., Harry, G., Moudgil, R., McMurtry, M. S., Weir, E. K., and Archer, S. L. (2006). An abnormal mitochondrial–hypoxia inducible factor-1–kv channel pathway disrupts oxygen sensing and triggers pulmonary arterial hypertension in fawn hooded rats. *Circulation*, 113(22):2630–2641.

[55] Borozdin, W., Bravo Ferrer Acosta, A. M., Bamshad, M. J., Botzenhart, E. M., Froster, U. G., Lemke, J., Schinzel, A., Spranger, S., McGaughran, J., Wand, D., Chrzanowska, K. H., and Kohlhase, J. (2006). Expanding the spectrum of TBX5 mutations in Holt-Oram syndrome: detection of two intragenic deletions by quantitative real time PCR, and report of eight novel point mutations. *Hum. Mutat.*, 27(9):975–976.

[56] Boycott, K. M., Rath, A., Chong, J. X., Hartley, T., Alkuraya, F. S., Baynam, G., Brookes, A. J., Brudno, M., Carracedo, A., den Dunnen, J. T., Dyke, S. O. M., Estivill, X., Goldblatt, J., Gonthier, C., Groft, S. C., Gut, I., Hamosh, A., Hieter, P., Höhn, S., Hurles, M. E., Kaufmann, P., Knoppers, B. M., Krischer, J. P., Macek, Jr, M., Matthijs, G., Olry, A., Parker, S., Paschall, J., Philippakis, A. A., Rehm, H. L., Robinson, P. N., Sham, P.-C., Stefanov, R., Taruscio, D., Unni, D., Vanstone, M. R., Zhang, F., Brunner, H., Bamshad, M. J., and Lochmüller, H. (2017). International cooperation to enable the diagnosis of all rare genetic diseases. *Am. J. Hum. Genet.*, 100(5):695–705.

[57] Bravo-Gil, N., González-Del Pozo, M., Martín-Sánchez, M., Méndez-Vidal, C., Rodríguez-de la Rúa, E., Borrego, S., and Antiñolo, G. (2017). Unravelling the genetic basis of simplex retinitis pigmentosa cases. *Sci. Rep.*, 7:41937.

[58] Brewis, M. J., Church, A. C., Johnson, M. K., and Peacock, A. J. (2015). Severe pulmonary hypertension in lung disease: phenotypes and response to treatment. *Eur. Respir. J.*, 46(5):1378–1389.

[59] Bris, C., Goudenege, D., Desquiret-Dumas, V., Charif, M., Colin, E., Bonneau, D., Amati-Bonneau, P., Lenaers, G., Reynier, P., and Procaccio, V. (2018). Bioinformatics tools and databases to assess the pathogenicity of mitochondrial DNA variants in the field of next generation sequencing. *Front. Genet.*, 9:632.

[60] Brock, M., Samillan, V. J., Trenkmann, M., Schwarzwald, C., Ulrich, S., Gay, R. E., Gassmann, M., Ostergaard, L., Gay, S., Speich, R., and Huber, L. C. (2014). AntagomiR directed against mir-20a restores functional BMPR2 signalling and prevents vascular remodelling in hypoxia-induced pulmonary hypertension. *Eur. Heart J.*, 35(45):3203–3211.

[61] Brown, W. M., Beck, S. R., Lange, E. M., Davis, C. C., Kay, C. M., Langefeld, C. D., Rich, S. S., and Framingham Heart Study (2003). Age-stratified heritability estimation in the framingham heart study families. *BMC Genet.*, 4 Suppl 1:S32.

[62] Burg, E. D., Remillard, C. V., and Yuan, J. X.-J. (2008). Potassium channels in the regulation of pulmonary artery smooth muscle cell proliferation and apoptosis: pharmacotherapeutic implications. *Br. J. Pharmacol.*, 153 Suppl 1:S99–S111.

[63] Burke, C. M., Glanville, A. R., Morris, A. J., Rubin, D., Harvey, J. A., Theodore, J., and Robin, E. D. (1987). Pulmonary function in advanced pulmonary hypertension. *Thorax*, 42(2):131–135.

[64] Buyske, S., Yang, G., Matise, T. C., and Gordon, D. (2009). When a case is not a case: Effects of phenotype misclassification on power and sample size requirements for the transmission disequilibrium test with affected child trios. *Human Heredity*, 67(4):287–292.

[65] Cacheiro, P., Muñoz-Fuentes, V., Murray, S. A., Dickinson, M. E., Bucan, M., Nutter, L. M. J., Peterson, K. A., Haselimashhadi, H., Flenniken, A. M., Morgan, H., Westerberg, H., Konopka, T., Hsu, C.-W., Christiansen, A., Lanza, D. G., Beaudet, A. L., Heaney, J. D., Fuchs, H., Gailus-Durner, V., Sorg, T., Prochazka, J., Novosadova, V., Lelliott, C. J., Wardle-Jones, H., Wells, S., Teboul, L., Cater, H., Stewart, M., Hough, T., Wurst, W., Sedlacek, R., Adams, D. J., Seavitt, J. R., Tocchini-Valentini, G., Mammano, F., Braun, R. E., McKerlie, C., Herault, Y., de Angelis, M. H., Mallon, A.-M., Lloyd, K. C. K., Brown, S. D. M., Parkinson, H., Meehan, T. F., Smedley, D., Genomics England Research Consortium, and International Mouse Phenotyping Consortium (2020). Human and mouse essentiality screens as a resource for disease gene discovery. *Nat. Commun.*, 11(1):655.

[66] Campisi, J., Kim, S. H., Lim, C. S., and Rubio, M. (2001). Cellular senescence, cancer and aging: the telomere connection. *Exp. Gerontol.*, 36(10):1619–1637.

[67] Caruso, P., Dunmore, B. J., Schlosser, K., Schoors, S., Dos Santos, C., Perez-Iratxeta, C., Lavoie, J. R., Zhang, H., Long, L., Flockton, A. R., Frid, M. G., Upton, P. D., D'Alessandro, A., Hadinnapola, C., Kiskin, F. N., Taha, M., Hurst, L. A., Ormiston, M. L., Hata, A., Stenmark, K. R., Carmeliet, P., Stewart, D. J., and Morrell, N. W. (2017). Identification of MicroRNA-124 as a major regulator of enhanced endothelial cell glycolysis in pulmonary arterial hypertension via PTBP1 (polypyrimidine tract binding protein) and pyruvate kinase M2. *Circulation*, 136(25):2451–2467.

[68] Caruso, P., MacLean, M. R., Khanin, R., McClure, J., Soon, E., Southgate, M., MacDonald, R. A., Greig, J. A., Robertson, K. E., Masson, R., Denby, L., Dempsie, Y., Long, L., Morrell, N. W., and Baker, A. H. (2010). Dynamic changes in lung MicroRNA profiles during the development of pulmonary hypertension due to chronic hypoxia and monocrotaline. *Arteriosclerosis, Thrombosis, and Vascular Biology*, 30(4):716–723.

[69] Castel, S. E., Cervera, A., Mohammadi, P., Aguet, F., Reverter, F., Wolman, A., Guigo, R., Iossifov, I., Vasileva, A., and Lappalainen, T. (2018). Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk. *Nat. Genet.*, 50(9):1327–1334.

[70] Cescon, M., Gattazzo, F., Chen, P., and Bonaldo, P. (2015). Collagen VI at a glance. *J. Cell Sci.*, 128(19):3525–3531.

[71] Chaouat, A. (2004). Endoglin germline mutation in a patient with hereditary haemorrhagic telangiectasia and dexfenfluramine associated pulmonary arterial hypertension. *Thorax*, 59(5):446–448.

[72] Charlson, M. E., Pompei, P., Ales, K. L., and MacKenzie, C. R. (1987). A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J. Chronic Dis.*, 40(5):373–383.

[73] Chen, J., Guo, J., Cui, X., Dai, Y., Tang, Z., Qu, J., Raj, J. U., Hu, Q., and Gou, D. (2018). The long noncoding RNA LnRPT is regulated by PDGF-BB and modulates the

proliferation of pulmonary artery smooth muscle cells. *American Journal of Respiratory Cell and Molecular Biology*, 58(2):181–193.

[74] Chen, N.-Y., D Collum, S., Luo, F., Weng, T., Le, T.-T., M Hernandez, A., Philip, K., Molina, J. G., Garcia-Morales, L. J., Cao, Y., Ko, T. C., Amione-Guerra, J., Al-Jabbari, O., Bunge, R. R., Youker, K., Bruckner, B. A., Hamid, R., Davies, J., Sinha, N., and Karmouty-Quintana, H. (2016). Macrophage bone morphogenic protein receptor 2 depletion in idiopathic pulmonary fibrosis and group III pulmonary hypertension. *Am. J. Physiol. Lung Cell. Mol. Physiol.*, 311(2):L238–54.

[75] Cheung, W. A., Shao, X., Morin, A., Siroux, V., Kwan, T., Ge, B., Aïssi, D., Chen, L., Vasquez, L., Allum, F., Guénard, F., Bouzigon, E., Simon, M.-M., Boulier, E., Redensek, A., Watt, S., Datta, A., Clarke, L., Flicek, P., Mead, D., Paul, D. S., Beck, S., Bourque, G., Lathrop, M., Tchernof, A., Vohl, M.-C., Demenais, F., Pin, I., Downes, K., Stunnenberg, H. G., Soranzo, N., Pastinen, T., and Grundberg, E. (2017). Functional variation in allelic methylomes underscores a strong genetic contribution and reveals novel epigenetic alterations in the human epigenome. *Genome Biology*, 18(1).

[76] Chinnappan, M., Mohan, A., Agarwal, S., Dalvi, P., and Dhillon, N. K. (2018). Network of MicroRNAs mediate translational repression of bone morphogenetic protein receptor-2: Involvement in HIV-associated pulmonary vascular remodeling. *Journal of the American Heart Association*, 7(5).

[77] Chiurazzi, P. and Pirozzi, F. (2016). Advances in understanding - genetic basis of intellectual disability. *F1000Res.*, 5.

[78] Cho, Y. J., Han, J. Y., Lee, S. G., Jeon, B. T., Choi, W. S., Hwang, Y. S., Roh, G. S., and Lee, J. D. (2009). Temporal changes of angiopoietins and tie2 expression in rat lungs after monocrotaline-induced pulmonary hypertension. *Comp. Med.*, 59(4):350–356.

[79] Chung, W. K., Austin, E. D., Best, D. H., Brown, L. M., and Elliott, C. G. (2015). When to offer genetic testing for pulmonary arterial hypertension. *Can. J. Cardiol.*, 31(4):544–547.

[80] Chung, W. K., Deng, L., Carroll, J. S., Mallory, N., Diamond, B., Rosenzweig, E. B., Barst, R. J., and Morse, J. H. (2009). Polymorphism in the angiotensin II type 1 receptor (AGTR1) is associated with age at diagnosis in pulmonary arterial hypertension. *J. Heart Lung Transplant.*, 28(4):373–379.

[81] Church, D. M., Schneider, V. A., Steinberg, K. M., Schatz, M. C., Quinlan, A. R., Chin, C.-S., Kitts, P. A., Aken, B., Marth, G. T., Hoffman, M. M., Herrero, J., Mendoza, M. L. Z., Durbin, R., and Flicek, P. (2015). Extending reference assembly models. *Genome Biol.*, 16:13.

[82] Ciuclan, L., Sheppard, K., Dong, L., Sutton, D., Duggan, N., Hussey, M., Simmons, J., Morrell, N. W., Jarai, G., Edwards, M., Dubois, G., Thomas, M., Van Heeke, G., and England, K. (2013). Treatment with anti-gremlin 1 antibody ameliorates chronic hypoxia/SU5416-induced pulmonary arterial hypertension in mice. *Am. J. Pathol.*, 183(5):1461–1473.

[83] Clapp, C. and Martínez de la Escalera, G. (2006). Aquaporin-1: a novel promoter of tumor angiogenesis. *Trends Endocrinol. Metab.*, 17(1):1–2.

[84] Consortium, I. H. G. S. and International Human Genome Sequencing Consortium (2001). Initial sequencing and analysis of the human genome.

[85] Cooper, G. M., Stone, E. A., Asimenos, G., NISC Comparative Sequencing Program, Green, E. D., Batzoglou, S., and Sidow, A. (2005). Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.*, 15(7):901–913.

[86] Copeland, C. A., Han, B., Tiwari, A., Austin, E. D., Loyd, J. E., West, J. D., and Kenworthy, A. K. (2017). A disease-associated frameshift mutation in caveolin-1 disrupts caveolae formation and function through introduction of a de novo ER retention signal. *Mol. Biol. Cell*, 28(22):3095–3111.

[87] Corbin, L. J., Tan, V. Y., Hughes, D. A., Wade, K. H., Paul, D. S., Tansey, K. E., Butcher, F., Dudbridge, F., Howson, J. M., Jallow, M. W., John, C., Kingston, N., Lindgren, C. M., O'Donavan, M., O'Rahilly, S., Owen, M. J., Palmer, C. N. A., Pearson, E. R., Scott, R. A., van Heel, D. A., Whittaker, J., Frayling, T., Tobin, M. D., Wain, L. V., Smith, G. D., Evans, D. M., Karpe, F., McCarthy, M. I., Danesh, J., Franks, P. W., and Timpson, N. J. (2018). Formalising recall by genotype as an efficient approach to detailed phenotyping and causal inference. *Nat. Commun.*, 9(1):711.

[88] Cordell, H. J., Töpf, A., Mamasoula, C., Postma, A. V., Bentham, J., Zelenika, D., Heath, S., Blue, G., Cosgrove, C., Granados Riveron, J., Darlay, R., Soemedi, R., Wilson, I. J., Ayers, K. L., Rahman, T. J., Hall, D., Mulder, B. J. M., Zwinderman, A. H., van Engelen, K., Brook, J. D., Setchfield, K., Bu'Lock, F. A., Thornborough, C., O'Sullivan, J., Stuart, A. G., Parsons, J., Bhattacharya, S., Winlaw, D., Mital, S., Gewillig, M., Breckpot, J., Devriendt, K., Moorman, A. F. M., Rauch, A., Lathrop, G. M., Keavney, B. D., and Goodship, J. A. (2013). Genome-wide association study identifies loci on 12q24 and 13q32 associated with tetralogy of fallot. *Hum. Mol. Genet.*, 22(7):1473–1481.

[89] Corris, P. A. (2015). The UK national pulmonary hypertension service, registry and research collaboration. *Global Cardiology Science and Practice*, 2015(3):37.

[90] Cotes, J. E., Chinn, D. J., Quanjer, P. H., Roca, J., and Yernault, J. C. (1993). Standardization of the measurement of transfer factor (diffusing capacity). report working party standardization of lung function tests, european community for steel and coal. official statement of the european respiratory society. *Eur. Respir. J. Suppl.*, 16:41–52.

[91] Cowan, J. R. and Ware, S. M. (2015). Genetics and genetic testing in congenital heart disease. *Clinics in Perinatology*, 42(2):373–393.

[92] Crosby, A., Toshner, M. R., Southwood, M. R., Soon, E., Dunmore, B. J., Groves, E., Moore, S., Wright, P., Ottersbach, K., Bennett, C., Guerrero, J., Ghevaert, C., and Morrell, N. W. (2018). Hematopoietic stem cell transplantation alters susceptibility to pulmonary hypertension in bmpr2-deficient mice. *Pulm. Circ.*, 8(4):2045894018801642.

[93] Csiszar, A., Ungvari, Z., Edwards, J. G., Kaminski, P., Wolin, M. S., Koller, A., and Kaley, G. (2002). Aging-induced phenotypic changes and oxidative stress impair coronary arteriolar function. *Circ. Res.*, 90(11):1159–1166.

[94] Csiszar, A., Ungvari, Z., Koller, A., Edwards, J. G., and Kaley, G. (2004). Proinflammatory phenotype of coronary arteries promotes endothelial apoptosis in aging. *Physiol. Genomics*, 17(1):21–30.

[95] Cui, H., Wang, Y., Huang, H., Yu, W., Bai, M., Zhang, L., Bryan, B. A., Wang, Y., Luo, J., Li, D., Ma, Y., and Liu, M. (2014). GPR126 protein regulates developmental and pathological angiogenesis through modulation of VEGFR2 receptor signaling. *J. Biol. Chem.*, 289(50):34871–34885.

[96] Culley, M. K. and Chan, S. Y. (2018). Mitochondrial metabolism in pulmonary hypertension: beyond mountains there are mountains. *J. Clin. Invest.*, 128(9):3704–3715.

[97] D'Alonzo, G. E., Barst, R. J., Ayres, S. M., Bergofsky, E. H., Brundage, B. H., Detre, K. M., Fishman, A. P., Goldring, R. M., Groves, B. M., and Kernis, J. T. (1991). Survival in patients with primary pulmonary hypertension. results from a national prospective registry. *Ann. Intern. Med.*, 115(5):343–349.

[98] D'Alonzo, G. E., Bower, J. S., and Dantzker, D. R. (1984). Differentiation of patients with primary and thromboembolic pulmonary hypertension. *Chest*, 85(4):457–461.

[99] Dalton, L. E., Healey, E., Irving, J., and Marciniak, S. J. (2012). Phosphoproteins in stress-induced disease. *Prog. Mol. Biol. Transl. Sci.*, 106:189–221.

[100] Dalvi, P., O'Brien-Ladner, A., and Dhillon, N. K. (2013). Downregulation of bone morphogenetic protein receptor axis during HIV-1 and cocaine-mediated pulmonary smooth muscle hyperplasia: implications for HIV-related pulmonary arterial hypertension. *Arterioscler. Thromb. Vasc. Biol.*, 33(11):2585–2595.

[101] Damico, R., Kolb, T. M., Valera, L., Wang, L., Housten, T., Tedford, R. J., Kass, D. A., Rafaels, N., Gao, L., Barnes, K. C., Benza, R. L., Rand, J. L., Hamid, R., Loyd, J. E., Robbins, I. M., Hemnes, A. R., Chung, W. K., Austin, E. D., Drummond, M. B., Mathai, S. C., and Hassoun, P. M. (2015). Serum endostatin is a genetically determined predictor of survival in pulmonary arterial hypertension. *Am. J. Respir. Crit. Care Med.*, 191(2):208–218.

[102] D'Amico, R. W., Faley, S., Shim, H.-N., Prosser, J. R., Agrawal, V., Bellan, L. M., and West, J. D. (2018). Pulmonary vascular platform models the effects of flow and pressure on endothelial dysfunction in BMPR2 associated pulmonary arterial hypertension. *Int. J. Mol. Sci.*, 19(9).

[103] Dannewitz Prosseda, S., Tian, X., Kuramoto, K., Boehm, M., Sudheendra, D., Miyagawa, K., Zhang, F., Solow-Cordero, D., Saldivar, J. C., Austin, E. D., Loyd, J. E., Wheeler, L., Andruska, A., Donato, M., Wang, L., Huebner, K., Metzger, R. J., Khatri, P., and Spiekerkoetter, E. (2019). FHIT, a novel modifier gene in pulmonary arterial hypertension. *Am. J. Respir. Crit. Care Med.*, 199(1):83–98.

[104] Davies, P. and Reid, L. (1982). Pulmonary veno-occlusive disease in siblings: case reports and morphometric study. *Hum. Pathol.*, 13(10):911–915.

[105] de Jesus Perez, V. A., Yuan, K., Lyuksyutova, M. A., Dewey, F., Orcholski, M. E., Shuffle, E. M., Mathur, M., Yancy, Jr, L., Rojas, V., Li, C. G., Cao, A., Alastalo, T.-P., Khazeni, N., Cimprich, K. A., Butte, A. J., Ashley, E., and Zamanian, R. T. (2014). Whole-exome sequencing reveals TopBP1 as a novel gene in idiopathic pulmonary arterial hypertension. *Am. J. Respir. Crit. Care Med.*, 189(10):1260–1272.

[106] De Vilder, E. Y. G., Debacker, J., and Vanakker, O. M. (2017). GGCX-Associated phenotypes: An overview in search of Genotype-Phenotype correlations. *Int. J. Mol. Sci.*, 18(2).

[107] Del Greco M, F., Pattaro, C., Luchner, A., Pichler, I., Winkler, T., Hicks, A. A., Fuchsberger, C., Franke, A., Melville, S. A., Peters, A., Wichmann, H. E., Schreiber, S., Heid, I. M., Krawczak, M., Minelli, C., Wiedermann, C. J., and Pramstaller, P. P. (2011). Genome-wide association analysis and fine mapping of NT-proBNP level provide novel insight into the role of the MTHFR-CLCN6-NPPA-NPPB gene cluster. *Hum. Mol. Genet.*, 20(8):1660–1671.

[108] Deng, Z., Morse, J. H., Slager, S. L., Cuervo, N., Moore, K. J., Venetos, G., Kalachikov, S., Cayanis, E., Fischer, S. G., Barst, R. J., Hodge, S. E., and Knowles, J. A. (2000). Familial primary pulmonary hypertension (gene PPH1) is caused by mutations in the bone morphogenetic protein receptor-II gene. *Am. J. Hum. Genet.*, 67(3):737–744.

[109] Dolk, H., Loane, M., and and, E. G. (2011). Congenital heart defects in europe. *Circulation*, 123(8):841–849.

[110] Donaldson, J. W., McKeever, T. M., Hall, I. P., Hubbard, R. B., and Fogarty, A. W. (2014). The UK prevalence of hereditary haemorrhagic telangiectasia and its association with sex, socioeconomic status and region of residence: a population-based study. *Thorax*, 69(2):161–167.

[111] Drake, K. M., Comhair, S. A., Erzurum, S. C., Tuder, R. M., and Aldred, M. A. (2015). Endothelial chromosome 13 deletion in congenital heart disease-associated pulmonary arterial hypertension dysregulates SMAD9 signaling. *Am. J. Respir. Crit. Care Med.*, 191(7):850–854.

[112] Drake, K. M., Dunmore, B. J., McNelly, L. N., Morrell, N. W., and Aldred, M. A. (2013). Correction of nonsense BMPR2 and SMAD9 mutations by ataluren in pulmonary arterial hypertension. *Am. J. Respir. Cell Mol. Biol.*, 49(3):403–409.

[113] Dresdale, D. T., Michtom, R. J., and Schultz, M. (1954). Recent studies in primary pulmonary hypertension, including pharmacodynamic observations on pulmonary vascular resistance. *Bull. N. Y. Acad. Med.*, 30(3):195–207.

[114] Dromparis, P., Paulin, R., Sutendra, G., Qi, A. C., Bonnet, S., and Michelakis, E. D. (2013). Uncoupling protein 2 deficiency mimics the effects of hypoxia and endoplasmic reticulum stress on mitochondria and triggers pseudohypoxic pulmonary vascular remodeling and pulmonary hypertension. *Circ. Res.*, 113(2):126–136.

[115] Duan, Y., Liu, G., Sun, Y., Wu, J., Xiong, Z., Jin, T., and Chen, M. (2020). Collagen type VI $\alpha 5$ gene variations may predict the risk of lung cancer development in chinese han population. *Sci. Rep.*, 10(1):5010.

[116] Dunmore, B. J., Drake, K. M., Upton, P. D., Toshner, M. R., Aldred, M. A., and Morrell, N. W. (2013). The lysosomal inhibitor, chloroquine, increases cell surface BMPR-II levels and restores BMP9 signalling in endothelial cells harbouring BMPR-II mutations. *Hum. Mol. Genet.*, 22(18):3667–3679.

[117] Durrington, H. J., Upton, P. D., Hoer, S., Boname, J., Dunmore, B. J., Yang, J., Crilley, T. K., Butler, L. M., Blackbourn, D. J., Nash, G. B., Lehner, P. J., and Morrell, N. W. (2010). Identification of a lysosomal pathway regulating degradation of the bone morphogenetic protein receptor type II. *J. Biol. Chem.*, 285(48):37641–37649.

[118] Dweik, R. A., Rounds, S., Erzurum, S. C., Archer, S., Fagan, K., Hassoun, P. M., Hill, N. S., Humbert, M., Kawut, S. M., Krowka, M., Michelakis, E., Morrell, N. W., Stenmark, K., Tuder, R. M., Newman, J., and ATS Committee on Pulmonary Hypertension Phenotypes (2014). An official american thoracic society statement: pulmonary hypertension phenotypes. *Am. J. Respir. Crit. Care Med.*, 189(3):345–355.

[119] Eichstaedt, C. A., Song, J., Benjamin, N., Harutyunova, S., Fischer, C., Grünig, E., and Hinderhofer, K. (2016). EIF2AK4 mutation as "second hit" in hereditary pulmonary arterial hypertension. *Respir. Res.*, 17(1):141.

[120] Eilbeck, K., Quinlan, A., and Yandell, M. (2017). Settling the score: variant prioritization and mendelian disease. *Nat. Rev. Genet.*, 18(10):599–612.

[121] El-Bizri, N., Wang, L., Merklinger, S. L., Guignabert, C., Desai, T., Urashima, T., Sheikh, A. Y., Knutsen, R. H., Mecham, R. P., Mishina, Y., and Rabinovitch, M. (2008). Smooth muscle protein 22alpha-mediated patchy deletion of bmpr1a impairs cardiac contractility but protects against pulmonary vascular remodeling. *Circ. Res.*, 102(3):380–388.

[122] El-Dabh, A. and Acharya, D. (2019). Pulmonary hypertension with dasatinib and other tyrosine kinase inhibitors. *Pulmonary Circulation*, 9(3):204589401986570.

[123] Escribano-Subias, P., Blanco, I., López-Meseguer, M., Lopez-Guarch, C. J., Roman, A., Morales, P., Castillo-Palma, M. J., Segovia, J., Gómez-Sanchez, M. A., Barberà, J. A., and REHAP investigators (2012). Survival in pulmonary hypertension in spain: insights from the spanish registry. *Eur. Respir. J.*, 40(3):596–603.

[124] Evans, J. D. W., Girerd, B., Montani, D., Wang, X.-J., Galiè, N., Austin, E. D., Elliott, G., Asano, K., Grünig, E., Yan, Y., Jing, Z.-C., Manes, A., Palazzini, M., Wheeler, L. A., Nakayama, I., Satoh, T., Eichstaedt, C., Hinderhofer, K., Wolf, M., Rosenzweig, E. B., Chung, W. K., Soubrier, F., Simonneau, G., Sitbon, O., Gräf, S., Kaptoge, S., Di Angelantonio, E., Humbert, M., and Morrell, N. W. (2016). BMPR2 mutations and survival in pulmonary arterial hypertension: an individual participant data meta-analysis. *Lancet Respir Med*, 4(2):129–137.

[125] Eyries, M., Coulet, F., Girerd, B., Montani, D., Humbert, M., Lacombe, P., Chinet, T., Gouya, L., Roume, J., Axford, M. M., Pearson, C. E., and Soubrier, F. (2012). ACVRL1 germinal mosaic with two mutant alleles in hereditary hemorrhagic telangiectasia associated with pulmonary arterial hypertension. *Clin. Genet.*, 82(2):173–179.

[126] Eyries, M., Montani, D., Girerd, B., Favrolt, N., Riou, M., Faivre, L., Manaud, G., Perros, F., Gräf, S., Morrell, N. W., Humbert, M., and Soubrier, F. (2020). Familial pulmonary arterial hypertension by KDR heterozygous loss of function. *European Respiratory Journal*, 55(4):1902165.

[127] Eyries, M., Montani, D., Girerd, B., Perret, C., Leroy, A., Lonjou, C., Chelghoum, N., Coulet, F., Bonnet, D., Dorfmüller, P., Fadel, E., Sitbon, O., Simonneau, G., Tregouët, D.-A., Humbert, M., and Soubrier, F. (2014). EIF2AK4 mutations cause pulmonary veno-occlusive disease, a recessive form of pulmonary hypertension. *Nat. Genet.*, 46(1):65–69.

[128] Eyries, M., Montani, D., Nadaud, S., Girerd, B., Levy, M., Bourdin, A., Trésorier, R., Chaouat, A., Cottin, V., Sanfiorenzo, C., Prevot, G., Reynaud-Gaubert, M., Dromer, C., Houeijeh, A., Nguyen, K., Coulet, F., Bonnet, D., Humbert, M., and Soubrier, F. (2019). Widening the landscape of heritable pulmonary hypertension mutations in paediatric and adult cases. *Eur. Respir. J.*, 53(3).

[129] Farber, H. W., Miller, D. P., Poms, A. D., Badesch, D. B., Frost, A. E., Rouzic, E. M.-L., Romero, A. J., Benton, W. W., Elliott, C. G., McGoon, M. D., and Benza, R. L. (2015). Five-year outcomes of patients enrolled in the REVEAL registry. *Chest*, 148(4):1043–1054.

[130] Farh, K. K.-H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W. J., Beik, S., Shoresh, N., Whitton, H., Ryan, R. J. H., Shishkin, A. A., Hatan, M., Carrasco-Alfonso, M. J., Mayer, D., Luckey, C. J., Patsopoulos, N. A., De Jager, P. L., Kuchroo, V. K., Epstein, C. B., Daly, M. J., Hafler, D. A., and Bernstein, B. E. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. *Nature*, 518(7539):337–343.

[131] Farha, S., Hu, B., Comhair, S., Zein, J., Dweik, R., Erzurum, S. C., and Aldred, M. A. (2016). Mitochondrial haplogroups and risk of pulmonary arterial hypertension. *PLoS One*, 11(5):e0156042.

[132] Fauvel, C., Raitière, O., Belkacem, N. S., Dominique, S., Artaud-Macari, E., Viacroze, C., Schleifer, D., and Bauer, F. (2020). Prognostic importance of kidney, heart and interstitial lung diseases (KHI triad) in PH: A machine learning study. *Arch. Cardiovasc. Dis.*, 113(10):630–641.

[133] Fernandez-Bonetti, P., Lupi-Herrera, E., Martinez-Guerra, M. L., Barrios, R., Seoane, M., and Sandoval, J. (1983). Peripheral airways obstruction in idiopathic pulmonary artery hypertension (primary). *Chest*, 83(5):732–738.

[134] Ferrara, N., Carver-Moore, K., Chen, H., Dowd, M., Lu, L., O'Shea, K. S., Powell-Braxton, L., Hillan, K. J., and Moore, M. W. (1996). Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene. *Nature*, 380(6573):439–442.

[135] Fessel, J. P., Hamid, R., Wittmann, B. M., Robinson, L. J., Blackwell, T., Tada, Y., Tanabe, N., Tatsumi, K., Hemnes, A. R., and West, J. D. (2012). Metabolomic analysis of bone morphogenetic protein receptor type 2 mutations in human pulmonary endothelium reveals widespread metabolic reprogramming. *Pulm. Circ.*, 2(2):201–213.

[136] Finer, S., Martin, H. C., Khan, A., Hunt, K. A., MacLaughlin, B., Ahmed, Z., Ashcroft, R., Durham, C., MacArthur, D. G., McCarthy, M. I., Robson, J., Trivedi, B., Griffiths, C., Wright, J., Trembath, R. C., and van Heel, D. A. (2020). Cohort profile: East london genes & health (ELGH), a community-based population genomics and health study in british bangladeshi and british pakistani people. *Int. J. Epidemiol.*, 49(1):20–21i.

[137] Firth, A. L., Remillard, C. V., Platoshyn, O., Fantozzi, I., Ko, E. A., and Yuan, J. X.-J. (2011). Functional ion channels in human pulmonary artery smooth muscle cells: Voltage-dependent cation channels. *Pulm. Circ.*, 1(1):48–71.

[138] Firth, H. V., Richards, S. M., Bevan, A. P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R. M., and Carter, N. P. (2009). DECIPHER: Database of chromosomal imbalance and phenotype in humans using ensembl resources. *Am. J. Hum. Genet.*, 84(4):524–533.

[139] Fisher, R. A. (1919). XV.—the correlation between relatives on the supposition of mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52(2):399–433.

[140] FitzGerald, G., Botstein, D., Califf, R., Collins, R., Peters, K., Bruggen, N. V., and Rader, D. (2018). The future of humans as model organisms. *Science*, 361(6402):552–553.

[141] Flanagan, S. E., Clauin, S., Bellanné-Chantelot, C., de Lonlay, P., Harries, L. W., Gloyn, A. L., and Ellard, S. (2008). Update of mutations in the genes encoding the pancreatic beta-cell KATPchannel subunits kir6.2 (KCNJ11) and sulfonylurea receptor 1 (ABCC8) in diabetes mellitus and hyperinsulinism. *Human Mutation*, 30(2):170–180.

[142] Flannick, J., Beer, N. L., Bick, A. G., Agarwala, V., Molnes, J., Gupta, N., Burtt, N. P., Florez, J. C., Meigs, J. B., Taylor, H., Lyssenko, V., Irgens, H., Fox, E., Burslem, F., Johansson, S., Brosnan, M. J., Trimmer, J. K., Newton-Cheh, C., Tuomi, T., Molven, A., Wilson, J. G., O'Donnell, C. J., Kathiresan, S., Hirschhorn, J. N., Njølstad, P. R., Rolph, T., Seidman, J. G., Gabriel, S., Cox, D. R., Seidman, C. E., Groop, L., and Altshuler, D. (2013). Assessing the phenotypic effects in the general population of rare variants in genes for a dominant mendelian form of diabetes. *Nat. Genet.*, 45(11):1380–1385.

[143] Flossmann, O., Berden, A., de Groot, K., Hagen, C., Harper, L., Heijl, C., Höglund, P., Jayne, D., Luqmani, R., Mahr, A., Mukhtyar, C., Pusey, C., Rasmussen, N., Stegeman, C., Walsh, M., Westman, K., and European Vasculitis Study Group (2011). Long-term patient survival in ANCA-associated vasculitis. *Ann. Rheum. Dis.*, 70(3):488–494.

[144] Francois, M., Koopman, P., and Beltrame, M. (2010). SoxF genes: Key players in the development of the cardio-vascular system. *Int. J. Biochem. Cell Biol.*, 42(3):445–448.

[145] Freeman, P. J., Hart, R. K., Gretton, L. J., Brookes, A. J., and Dalgleish, R. (2018). VariantValidator: Accurate validation, mapping, and formatting of sequence variation descriptions. *Hum. Mutat.*, 39(1):61–68.

[146] Frey, B. J. and Dueck, D. (2007). Clustering by passing messages between data points. *Science*, 315(5814):972–976.

[147] Frikha, R. (2020). Klippel-Feil syndrome: a review of the literature. *Clin. Dysmorphol.*, 29(1):35–37.

[148] Frost, A. E., Badesch, D. B., Barst, R. J., Benza, R. L., Elliott, C. G., Farber, H. W., Krichman, A., Liou, T. G., Raskob, G. E., Wason, P., Feldkircher, K., Turner, M., and McGoon, M. D. (2011). The changing picture of patients with pulmonary arterial hypertension in the united states: how REVEAL differs from historic and non-US contemporary registries. *Chest*, 139(1):128–137.

[149] Fujiwara, M., Yagi, H., Matsuoka, R., Akimoto, K., Furutani, M., Imamura, S.-I., Uehara, R., Nakayama, T., Takao, A., Nakazawa, M., and Saji, T. (2008). Implications of mutations of activin receptor-like kinase 1 gene (ALK1) in addition to bone morphogenetic protein receptor II gene (BMPR2) in children with pulmonary arterial hypertension. *Circ. J.*, 72(1):127–133.

[150] Gabler, N. B., French, B., Strom, B. L., Liu, Z., Palevsky, H. I., Taichman, D. B., Kawut, S. M., and Halpern, S. D. (2012). Race and sex differences in response to endothelin receptor antagonists for pulmonary arterial hypertension. *Chest*, 141(1):20–26.

[151] Galambos, C., Mullen, M. P., Shieh, J. T., Schwerk, N., Kielt, M. J., Ullmann, N., Boldrini, R., Stucin-Gantar, I., Haass, C., Bansal, M., Agrawal, P. B., Johnson, J., Peca, D., Surace, C., Cutrera, R., Pauciulo, M. W., Nichols, W. C., Griese, M., Ivy, D., Abman, S. H., Austin, E. D., and Danhaive, O. (2019). Phenotype characterisation of TBX4 mutation and deletion carriers with neonatal and paediatric pulmonary hypertension. *European Respiratory Journal*, 54(2):1801965.

[152] Galiè, N., Humbert, M., Vachiery, J.-L., Gibbs, S., Lang, I., Torbicki, A., Simonneau, G., Peacock, A., Noordegraaf, A. V., Beghetti, M., Ghofrani, A., Sanchez, M. A. G., Hansmann, G., Klepetko, W., Lancellotti, P., Matucci, M., McDonagh, T., Pierard, L. A., Trindade, P. T., Zompatori, M., and Hoeper, M. (2015). 2015 ESC/ERS guidelines for the diagnosis and treatment of pulmonary hypertension. *European Heart Journal*, 37(1):67–119.

[153] Galiè, N., Humbert, M., Vachiery, J.-L., Gibbs, S., Lang, I., Torbicki, A., Simonneau, G., Peacock, A., Vonk Noordegraaf, A., Beghetti, M., Ghofrani, A., Gomez Sanchez, M. A., Hansmann, G., Klepetko, W., Lancellotti, P., Matucci, M., McDonagh, T., Pierard, L. A., Trindade, P. T., Zompatori, M., and Hoeper, M. (2015). 2015 ESC/ERS guidelines for the diagnosis and treatment of pulmonary hypertension: The joint task force for the diagnosis and treatment of pulmonary hypertension of the european society of cardiology (ESC) and the european respiratory society (ERS): Endorsed by: Association for european paediatric and congenital cardiology (AEPC), international society for heart and lung transplantation (ISHLT). *Eur. Respir. J.*, 46(4):903–975.

[154] Galili, T. (2015). dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. *Bioinformatics*, 31(22):3718–3720.

[155] Gall, H., Felix, J. F., Schneck, F. K., Milger, K., Sommer, N., Voswinckel, R., Franco, O. H., Hofman, A., Schermuly, R. T., Weissmann, N., Grimminger, F., Seeger, W., and Ghofrani, H. A. (2017). The giessen pulmonary hypertension registry: Survival in pulmonary hypertension subgroups. *The Journal of Heart and Lung Transplantation*, 36(9):957–967.

[156] Gallanti, A., Tonelli, A., Cardin, V., Bussone, G., Bresolin, N., and Bassi, M. T. (2008). A novel de novo nonsense mutation in ATP1a2 associated with sporadic hemiplegic migraine and epileptic seizures. *Journal of the Neurological Sciences*, 273(1-2):123–126.

[157] Gallo de Moraes, A., Vakil, A., and Moua, T. (2016). Patent foramen ovale in idiopathic pulmonary arterial hypertension: Long-term risk and morbidity. *Respir. Med.*, 118:53–57.

[158] Gamer, M., Lemon, J., and <puspendra.pusp22@gmail.com>, I. F. P. S. (2019). *irr: Various Coefficients of Interrater Reliability and Agreement*. R package version 0.84.1.

[159] Gangopahyay, A., Oran, M., Bauer, E. M., Wertz, J. W., Comhair, S. A., Erzurum, S. C., and Bauer, P. M. (2011). Bone morphogenetic protein receptor II is a novel mediator of endothelial nitric-oxide synthase activation. *J. Biol. Chem.*, 286(38):33134–33140.

[160] Gara, S. K., Grumati, P., Urciuolo, A., Bonaldo, P., Kobbe, B., Koch, M., Paulsson, M., and Wagener, R. (2008). Three novel collagen VI chains with high homology to the alpha3 chain. *J. Biol. Chem.*, 283(16):10658–10670.

[161] Garcia, A. A., Hirte, H., Fleming, G., Yang, D., Tsao-Wei, D. D., Roman, L., Groshen, S., Swenson, S., Markland, F., Gandara, D., Scudder, S., Morgan, R., Chen, H., Lenz, H.-J., and Oza, A. M. (2008). Phase II clinical trial of bevacizumab and low-dose metronomic oral cyclophosphamide in recurrent ovarian cancer: A trial of the california, chicago, and princess margaret hospital phase II consortia. *Journal of Clinical Oncology*, 26(1):76–82.

[162] Gazetopoulos, N., Salonikides, N., and Davies, H. (1974). Cardiopulmonary function in patients with pulmonary hypertension. *Br. Heart J.*, 36(1):19–28.

[163] Gazzo, A., Raimondi, D., Daneels, D., Moreau, Y., Smits, G., Van Dooren, S., and Lenaerts, T. (2017). Understanding mutational effects in digenic diseases. *Nucleic Acids Res.*, 45(15):e140.

[164] Génin, E. (2020). Missing heritability of complex diseases: case solved? *Hum. Genet.*, 139(1):103–113.

[165] Germain, M., Eyries, M., Montani, D., Poirier, O., Girerd, B., Dorfmüller, P., Coulet, F., Nadaud, S., Maugenre, S., Guignabert, C., Carpentier, W., Vonk-Noordegraaf, A., Lévy, M., Chaouat, A., Lambert, J.-C., Bertrand, M., Dupuy, A.-M., Letenneur, L., Lathrop, M., Amouyel, P., de Ravel, T. J. L., Delcroix, M., Austin, E. D., Robbins, I. M., Hemnes, A. R., Loyd, J. E., Berman-Rosenzweig, E., Barst, R. J., Chung, W. K., Simonneau, G., Trégouët, D. A., Humbert, M., and Soubrier, F. (2013). Genome-wide association analysis identifies a susceptibility locus for pulmonary arterial hypertension. *Nat. Genet.*, 45(5):518–521.

[166] German, K., Deutsch, G. H., Freed, A. S., Dipple, K. M., Chabra, S., and Bennett, J. T. (2019). Identification of a deletion containing TBX4 in a neonate with acinar dysplasia by rapid exome sequencing. *American Journal of Medical Genetics Part A*, 179(5):842–845.

[167] Ghigna, M.-R., Guignabert, C., Montani, D., Girerd, B., Jaïs, X., Savale, L., Hervé, P., Thomas de Montpréville, V., Mercier, O., Sitbon, O., Soubrier, F., Fadel, E., Simonneau, G., Humbert, M., and Dorfmüller, P. (2016). BMPR2 mutation status influences bronchial vascular changes in pulmonary arterial hypertension. *Eur. Respir. J.*, 48(6):1668–1681.

[168] Giannelli, Jr, S., Ayres, S. M., and Buehler, M. E. (1967). Effect of pulmonary blood flow upon lung mechanics. *J. Clin. Invest.*, 46(10):1625–1642.

[169] Girerd, B., Coulet, F., Jaïs, X., Eyries, M., Van Der Bruggen, C., De Man, F., Houweling, A., Dorfmüller, P., Savale, L., Sitbon, O., Vonk-Noordegraaf, A., Soubrier, F., Simonneau, G., Humbert, M., and Montani, D. (2015a). Characteristics of pulmonary arterial hypertension in affected carriers of a mutation located in the cytoplasmic tail of bone morphogenetic protein receptor type 2. *Chest*, 147(5):1385–1394.

[170] Girerd, B., Montani, D., Coulet, F., Sztrymf, B., Yaici, A., Jaïs, X., Tregouet, D., Reis, A., Drouin-Garraud, V., Fraisse, A., Sitbon, O., O'Callaghan, D. S., Simonneau, G., Soubrier, F., and Humbert, M. (2010). Clinical outcomes of pulmonary arterial hypertension in patients carrying an ACVRL1 (ALK1) mutation. *Am. J. Respir. Crit. Care Med.*, 181(8):851–861.

[171] Girerd, B., Montani, D., Jaïs, X., Eyries, M., Yaici, A., Sztrymf, B., Savale, L., Parent, F., Coulet, F., Godinas, L., Lau, E. M., Tamura, Y., Sitbon, O., Soubrier, F., Simonneau, G., and Humbert, M. (2015b). Genetic counselling in a national referral centre for pulmonary hypertension. *European Respiratory Journal*, 47(2):541–552.

[172] Glessner, J. T., Bick, A. G., Ito, K., Homsy, J., Rodriguez-Murillo, L., Fromer, M., Mazaika, E., Vardarajan, B., Italia, M., Leipzig, J., DePalma, S. R., Golhar, R., Sanders, S. J., Yamrom, B., Ronemus, M., Iossifov, I., Willsey, A. J., State, M. W., Kaltman, J. R., White, P. S., Shen, Y., Warburton, D., Brueckner, M., Seidman, C., Goldmuntz, E., Gelb, B. D., Lifton, R., Seidman, J., Hakonarson, H., and Chung, W. K. (2014). Increased frequency of de novo copy number variants in congenital heart disease by integrative analysis of single nucleotide polymorphism array and exome sequence data. *Circ. Res.*, 115(10):884–896.

[173] Godinas, L., Amar, D., Montani, D., Lau, E. M., Jaïs, X., Savale, L., Jevnikar, M., Sitbon, O., Simonneau, G., Humbert, M., Laveneziana, P., and Garcia, G. (2016). Lung capillary blood volume and membrane diffusion in precapillary pulmonary hypertension. *J. Heart Lung Transplant.*, 35(5):647–656.

[174] Gogarten, S. M., Sofer, T., Chen, H., Yu, C., Brody, J. A., Thornton, T. A., Rice, K. M., and Conomos, M. P. (2019). Genetic association testing using the GENESIS r/bioconductor package. *Bioinformatics*, 35(24):5346–5348.

[175] Gomez-Arroyo, J. G., Farkas, L., Alhussaini, A. A., Farkas, D., Kraskauskas, D., Voelkel, N. F., and Bogaard, H. J. (2012). The monocrotaline model of pulmonary hypertension in perspective. *Am. J. Physiol. Lung Cell. Mol. Physiol.*, 302(4):L363–9.

[176] Gong, J., Chen, Z., Chen, Y., Lv, H., Lu, H., Yan, F., Li, L., Zhang, W., and Shi, J. (2019). Long non-coding RNA CASC2 suppresses pulmonary artery smooth muscle cell proliferation and phenotypic switch in hypoxia-induced pulmonary hypertension. *Respir. Res.*, 20(1):53.

[177] Goswami, R. S. and Harada, S. (2020). An overview of molecular genetic diagnosis techniques. *Current Protocols in Human Genetics*, 105(1).

[178] Gottlieb, A., Stein, G. Y., Ruppin, E., and Sharan, R. (2011). PREDICT: a method for inferring novel drug indications with application to personalized medicine. *Mol. Syst. Biol.*, 7:496.

[179] Gräf, S., Haimel, M., Bleda, M., Hadinnapola, C., Southgate, L., Li, W., Hodgson, J., Liu, B., Salmon, R. M., Southwood, M., Machado, R. D., Martin, J. M., Treacy, C. M., Yates, K., Daugherty, L. C., Shamardina, O., Whitehorn, D., Holden, S., Aldred, M., Bogaard, H. J., Church, C., Coghlan, G., Condliffe, R., Corris, P. A., Danesino, C., Eyries, M., Gall, H., Ghio, S., Ghofrani, H.-A., Gibbs, J. S. R., Girerd, B., Houweling, A. C., Howard, L., Humbert, M., Kiely, D. G., Kovacs, G., MacKenzie Ross, R. V., Moledina, S., Montani, D., Newnham, M., Olschewski, A., Olschewski, H., Peacock, A. J., Pepke-Zaba, J., Prokopenko, I., Rhodes, C. J., Scelsi, L., Seeger, W., Soubrier, F., Stein, D. F., Suntharalingam, J., Swietlik, E. M., Toshner, M. R., van Heel, D. A., Vonk Noordegraaf, A., Waisfisz, Q., Wharton, J., Wort, S. J., Ouwehand, W. H., Soranzo, N., Lawrie, A., Upton, P. D., Wilkins, M. R., Trembath, R. C., and Morrell, N. W. (2018). Identification of rare sequence variation underlying heritable pulmonary arterial hypertension. *Nat. Commun.*, 9(1):1416.

[180] Green, D. E., Sutliff, R. L., and Hart, C. M. (2011). Is peroxisome proliferator-activated receptor gamma (PPAR) a therapeutic target for the treatment of pulmonary hypertension? *Pulmonary Circulation*, 1(1):33–47.

[181] Greene, D. (2016). *ontologyPlot: Functions for Visualising Sets of Ontological Terms*. R package version 1.4.

[182] Greene, D. (2019). *ontologySimilarity: Functions for Calculating Ontological Similarities*. R package version 2.3.

[183] Greene, D., NIHR BioResource, Richardson, S., and Turro, E. (2017a). A fast association test for identifying pathogenic variants involved in rare diseases. *Am. J. Hum. Genet.*, 101(1):104–114.

[184] Greene, D., Richardson, S., and Turro, E. (2017b). ontologyx: a suite of R packages for working with ontological data. *Bioinformatics*, 33(7):1104–1106.

[185] Gregersen, P. K. (1989). HLA class II polymorphism: implications for genetic susceptibility to autoimmune disease. *Lab. Invest.*, 61(1):5–19.

[186] Grigoriadis, S., Vonderporten, E. H., Mamisashvili, L., Tomlinson, G., Dennis, C.-L., Koren, G., Steiner, M., Mousmanis, P., Cheung, A., and Ross, L. E. (2014). Prenatal exposure to antidepressants and persistent pulmonary hypertension of the newborn: systematic review and meta-analysis. *BMJ*, 348:f6932.

[187] Grubert, F., Zaugg, J. B., Kasowski, M., Ursu, O., Spacek, D. V., Martin, A. R., Greenside, P., Srivas, R., Phanstiel, D. H., Pekowska, A., Heidari, N., Euskirchen, G., Huber, W., Pritchard, J. K., Bustamante, C. D., Steinmetz, L. M., Kundaje, A., and Snyder, M. (2015). Genetic control of chromatin states in humans involves local and distal chromosomal interactions. *Cell*, 162(5):1051–1065.

[188] Grunewald, J. and Eklund, A. (2007). Sex-specific manifestations of löfgren's syndrome. *American Journal of Respiratory and Critical Care Medicine*, 175(1):40–44.

[189] Grutters, J., Berg, J. T., Zeijden, J. V. D., Jaarsma, W., Ernst, J., and Westermann, C. (2001). Patent foramen ovale causing position-dependent shunting in a patient, when laying down her corset. *European Respiratory Journal*, 18(4):731–733.

[190] Guerra, S. G., Vyse, T. J., and Cunninghame Graham, D. S. (2012). The genetics of lupus: a functional perspective. *Arthritis Res. Ther.*, 14(3):211.

[191] Guerreso, K. and Conner, E. A. (2016). Possible role of anti-SSA/Ro antibodies in the pathogenesis of pulmonary hypertension. *Respir Med Case Rep*, 17:47–49.

[192] Guo, L., Li, M., Chen, Y., Wang, Q., Tian, Z., Pan, S., Zeng, X., and Ye, S. (2015). Anti-Endothelin receptor type a autoantibodies in systemic lupus Erythematosus-Associated pulmonary arterial hypertension. *Arthritis Rheumatol*, 67(9):2394–2402.

[193] Guo, Y., Dai, Y., Yu, H., Zhao, S., Samuels, D. C., and Shyr, Y. (2017). Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis. *Genomics*, 109(2):83–90.

[194] Hadinnapola, C., Bleda, M., Haimel, M., Screaton, N., Swift, A., Dorfmüller, P., Preston, S. D., Southwood, M., Hernandez-Sanchez, J., Martin, J., Treacy, C., Yates, K., Bogaard, H., Church, C., Coghlan, G., Condliffe, R., Corris, P. A., Gibbs, S., Girerd, B., Holden, S., Humbert, M., Kiely, D. G., Lawrie, A., Machado, R., Ross, R. M., Moledina, S., Montani, D., Newnham, M., Peacock, A., Pepke-Zaba, J., Rayner-Matthews, P., Shamardina, O., Soubrier, F., Southgate, L., Suntharalingam, J., Toshner, M., Trembath, R., Noordegraaf, A. V., Wilkins, M. R., Wort, S. J., Wharton, J., Gräf, S., Morrell, N. W., Aitman, T., Bennett, D., Caulfield, M., Chinnery, P., Gale, D., Koziell, A., Kuijpers, T. W., Laffan, M. A., Maher, E., Markus, H. S., Ouwehand, W. H., Perry, D., Raymond, F. L., Roberts, I., Smith, K., Thrasher, A., Watkins, H., Williamson, C., Woods, G., Ashford, S., Bradley, J. R., Fletcher, D., Hammerton, T., James, R., Kingston, N., Ouwehand, W. H., Penkett, C. J., Raymond, F. L., Stirrups, K., Veltman, M., Young, T., Ashford, S., Brown, M., Clements-Brod, N., Davis, J., Dewhurst, E., Erwood, M., Frary, A., Linger, R., Papadia, S., Rehnstrom, K., Stark, H., Allsup, D., Austin, S., Bakchoul, T., Bariana, T. K., Bolton-Maggs, P., Chalmers, E., Collins, P., Erber, W. N., Everington, T., Favier, R., Freson, K., Furie, B., Gattens, M., Gomez, K., Greene, D., Greinacher, A., Hart, D., Heemskerk, J. W., Henskens, Y., Kazmi, R., Keeling, D., Kelly, A. M., Laffan, M. A., Lambert, M. P., Lentaigne, C., Liesner, R., Mangles, S., Mathias, M., Millar, C. M., Mumford, A., Nurden, P., Ouwehand, W. H., Papadia, S., Payne, J., Pasi, J., Perry, D. J., Peerlinck, K., Richards, M., Rondina, M., Roughley, C., Schulman, S., Schulze, H., Scully, M., Sivapalaratnam, S., Tait, R. C., Talks, K., Thachil, J., Turro, E., Toh, C.-H., Geet, C. V., Vries, M. D., Warner, T. Q., Westbury, S., Furnell, A., Mapeta, R., Simeoni, I., Staines, S., Stephens, J., Stirrups, K., Whitehorn, D., Watt, C., Attwood, A., Daugherty, L., Deevi, S. V., Halmagyi, C., Hu, F., James, R., Matser, V., Meacham, S., Megy, K., Penkett, C. J., Stirrups, K., Titterton, C., Tuna, S., Yu, P., von Ziegenweldt, J., Astle, W., Carss, K., Greene, D., Lango-Allen, H., Turro, E., Astle, W., Greene, D., Richardson, S., Turro, E., Calleja, P., Rankin, S., Turek, W., Bryson, C., Anderson, J., Fletcher, D., McJannet, C., Stock, S., Young, T., Wassmer, E., Sohal, A., Santra, S., Vogt, J., Chitre, M., Krishnakumar, D., Ambegaonkar, G., Maw, A., Armstrong, R., Park, S.-M., Mehta, S., Paterson, J., Carmichael, J., Allen, L., Hensiek, A., Firth, H., Stein, P., Deegan, P., Doffinger, R., Parker, A., Bitner-Glindzicz, M., Scott, R., Hurst, J., Rosser, E., Lees, M., Clement, E., Henderson, R., Thompson, D., Gardham, A., Gissen,

P., Josifova, D., Thomas, E., Patch, C., Deshpande, C., Flinter, F., Holder, M., Canham, N., Wakeling, E., Holder, S., Ghali, N., Brady, A., Clowes, V., MacLaren, R., Webster, A., Moore, A., Arno, G., Michaelides, M., Rankin, J., Kurian, M., Murphy, E., Carss, K., Sanchis-Juan, A., Erwood, M., Dewhurst, E., Grozeva, D., Raymond, F. L., Reid, E., Woods, G., Tischkowitz, M., Sandford, R., Ali, S., Creaser-Myers, A., Cookson, V., DaCosta, R., Dormand, N., Ghataorhe, P. K., Greenhalgh, A., in't Veld, A. H., Kennedy, F., Ross, R. M., Masati, L., Meehan, S., Othman, S., Pollock, V., Polwarth, G., Rhodes, C. J., Rue-Albrecht, K., Schotte, G., Shipley, D., Tan, Y., Wanjiku, I., Wort, J., Smith, K., Kuijpers, T., Thrasher, A., Thaventhiran, J., Brown, M., Allen, H. L., Simeoni, I., Staples, E., Samarghitean, C., Alachkar, H., Antrobus, R., Arumugakani, G., Bacchelli, C., Baxendale, H., Bethune, C., Bibi, S., Booth, C., Browning, M., Burns, S., Chandra, A., Cooper, N., Davies, S., Devlin, L., Doffinger, R., Drewe, E., Edgar, D., Egner, W., Ghurye, R., Gilmour, K., Goddard, S., Gordins, P., Grigoriadou, S., Hackett, S., Hague, R., Hayman, G., Herwadkar, A., Huissoon, A., Jolles, S., Kelleher, P., Kumararatne, D., Lear, S., Longhurst, H., Lorenzo, L., Maimaris, J., Manson, A., McDermott, E., Murng, S., Nejentsev, S., Noorani, S., Oksenhendler, E., Ponsford, M., Qasim, W., Quinti, I., Richter, A., Sargur, R., Savic, S., Seneviratne, S., Sewell, C., Stauss, H., Thomas, M., Welch, S., Willcocks, L., Yeatman, N., and and, P. Y. (2017). Phenotypic characterization of EIF2ak4 mutation carriers in a large cohort of patients diagnosed clinically with pulmonary arterial hypertension. *Circulation*, 136(21):2022–2033.

[195] Hagen, P. T., Scholz, D. G., and Edwards, W. D. (1984). Incidence and size of patent foramen ovale during the first 10 decades of life: an autopsy study of 965 normal hearts. *Mayo Clin. Proc.*, 59(1):17–20.

[196] Hakim, M., Broza, Y. Y., Barash, O., Peled, N., Phillips, M., Amann, A., and Haick, H. (2012). Volatile organic compounds of lung cancer and possible biochemical pathways. *Chem. Rev.*, 112(11):5949–5966.

[197] Hales, C. N., Barker, D. J., Clark, P. M., Cox, L. J., Fall, C., Osmond, C., and Winter, P. D. (1991). Fetal and infant growth and impaired glucose tolerance at age 64. *BMJ*, 303(6809):1019–1022.

[198] Hamid, R., Cogan, J. D., Hedges, L. K., Austin, E., Phillips, 3rd, J. A., Newman, J. H., and Loyd, J. E. (2009). Penetrance of pulmonary arterial hypertension is modulated by the expression of normal BMPR2 allele. *Hum. Mutat.*, 30(4):649–654.

[199] Han, C., Hong, K.-H., Kim, Y. H., Kim, M.-J., Song, C., Kim, M. J., Kim, S.-J., Raizada, M. K., and Oh, S. P. (2013). SMAD1 deficiency in either endothelial or smooth muscle cells can predispose mice to pulmonary hypertension. *Hypertension*, 61(5):1044–1052.

[200] Hancock, D. B., Eijgelsheim, M., Wilk, J. B., Gharib, S. A., Loehr, L. R., Marciante, K. D., Franceschini, N., van Durme, Y. M. T. A., Chen, T.-H., Barr, R. G., Schabath, M. B., Couper, D. J., Brusselle, G. G., Psaty, B. M., van Duijn, C. M., Rotter, J. I., Uitterlinden, A. G., Hofman, A., Punjabi, N. M., Rivadeneira, F., Morrison, A. C., Enright, P. L., North, K. E., Heckbert, S. R., Lumley, T., Stricker, B. H. C., O'Connor, G. T., and London, S. J. (2010). Meta-analyses of genome-wide association studies identify multiple loci associated with pulmonary function. *Nat. Genet.*, 42(1):45–52.

[201] Harman, D. (1956). Aging: a theory based on free radical and radiation chemistry. *J. Gerontol.*, 11(3):298–300.

[202] Harrison, R. E., Berger, R., Haworth, S. G., Tulloh, R., Mache, C. J., Morrell, N. W., Aldred, M. A., and Trembath, R. C. (2005). Transforming growth factor-beta receptor mutations and pulmonary arterial hypertension in childhood. *Circulation*, 111(4):435–441.

[203] Harrison, R. E., Flanagan, J. A., Sankelo, M., Abdalla, S. A., Rowell, J., Machado, R. D., Elliott, C. G., Robbins, I. M., Olschewski, H., McLaughlin, V., Gruenig, E., Kermeen, F., Halme, M., Räisänen-Sokolowski, A., Laitinen, T., Morrell, N. W., and Trembath, R. C. (2003). Molecular and functional analysis identifies ALK-1 as the predominant cause of pulmonary hypertension related to hereditary haemorrhagic telangiectasia. *J. Med. Genet.*, 40(12):865–871.

[204] Hatano, S. (1975). *Primary pulmonary hypertension : report on a WHO meeting, Geneva, 15-17 October 1973*. World Health Organization Distributed by Q Corp, Geneva Albany, N.Y.

[205] Hauge, A., Bø, G., and Waaler, B. A. (1975). Interrelations between pulmonary liquid volumes and lung compliance. *J. Appl. Physiol.*, 38(4):608–614.

[206] Hay, D. W. (1999). Putative mediator role of endothelin-1 in asthma and other lung diseases. *Clin. Exp. Pharmacol. Physiol.*, 26(2):168–171.

[207] Hemnes, A. R., Brittain, E. L., Trammell, A. W., Fessel, J. P., Austin, E. D., Penner, N., Maynard, K. B., Gleaves, L., Talati, M., Absi, T., Disalvo, T., and West, J. (2014). Evidence for right ventricular lipotoxicity in heritable pulmonary arterial hypertension. *Am. J. Respir. Crit. Care Med.*, 189(3):325–334.

[208] Hemnes, A. R., Zhao, M., West, J., Newman, J. H., Rich, S., Archer, S. L., Robbins, I. M., Blackwell, T. S., Cogan, J., Loyd, J. E., Zhao, Z., Gaskill, C., Jetter, C., Kropski, J. A., Majka, S. M., and Austin, E. D. (2016). Critical genomic networks and vasoreactive variants in idiopathic pulmonary arterial hypertension. *Am. J. Respir. Crit. Care Med.*, 194(4):464–475.

[209] Hennig, C. (2020). *fpc: Flexible Procedures for Clustering*. R package version 2.2-9.

[210] Hernandez, F., Huether, R., Carter, L., Johnston, T., Thompson, J., Gossage, J. R., Chao, E., and Elliott, A. M. (2015). Mutations in RASA1 and GDF2 identified in patients with clinical features of hereditary hemorrhagic telangiectasia. *Hum Genome Var*, 2:15040.

[211] Hinchcliff, M., Khanna, S., Hsu, V. M., Lee, J., Almagor, O., Chang, R. W., Steen, V., and Chung, L. (2015). Survival in systemic sclerosis–pulmonary arterial hypertension by serum autoantibody status in the pulmonary hypertension assessment and recognition of outcomes in scleroderma (PHAROS) registry. *Seminars in Arthritis and Rheumatism*, 45(3):309–314.

[212] Hiraide, T., Kataoka, M., Suzuki, H., Aimi, Y., Chiba, T., Kanekura, K., Satoh, T., Fukuda, K., Gamou, S., and Kosaki, K. (2018). SOX17 mutations in japanese patients with pulmonary arterial hypertension. *Am. J. Respir. Crit. Care Med.*, 198(9):1231–1233.

[213] Hirschey, M. D., Shimazu, T., Jing, E., Grueter, C. A., Collins, A. M., Aouizerat, B., Stančáková, A., Goetzman, E., Lam, M. M., Schwer, B., Stevens, R. D., Muehlbauer, M. J., Kakar, S., Bass, N. M., Kuusisto, J., Laakso, M., Alt, F. W., Newgard, C. B., Farese, Jr, R. V., Kahn, C. R., and Verdin, E. (2011). SIRT3 deficiency and mitochondrial protein hyperacetylation accelerate the development of the metabolic syndrome. *Mol. Cell*, 44(2):177–190.

[214] Hodgson, J., Ruiz-Llorente, L., McDonald, J., Quarrell, O., Ugonna, K., Bentham, J., Mason, R., Martin, J., Moore, D., Bergstrom, K., Bayrak-Toydemir, P., Wooderchak-Donahue, W., Morrell, N. W., Condliffe, R., Bernabeu, C., and Upton, P. D. (2021). Homozygous GDF2 nonsense mutations result in a loss of circulating BMP9 and BMP10 and are associated with either PAH or an "HHT-like" syndrome in children. *Mol Genet Genomic Med*, page e1685.

[215] Hodgson, J., Swietlik, E. M., Salmon, R. M., Hadinnapola, C., Nikolic, I., Wharton, J., Guo, J., Liley, J., Haimel, M., Bleda, M., Southgate, L., Machado, R. D., Martin, J. M., Treacy, C. M., Yates, K., Daugherty, L. C., Shamardina, O., Whitehorn, D., Holden, S., Bogaard, H. J., Church, C., Coghlan, G., Condliffe, R., Corris, P. A., Danesino, C., Eyries, M., Gall, H., Ghio, S., Ghofrani, H.-A., Gibbs, J. S. R., Girerd, B., Houweling, A. C., Howard, L., Humbert, M., Kiely, D. G., Kovacs, G., Lawrie, A., MacKenzie Ross, R. V., Moledina, S., Montani, D., Olschewski, A., Olschewski, H., Ouwehand, W. H., Peacock, A. J., Pepke-Zaba, J., Prokopenko, I., Rhodes, C. J., Scelsi, L., Seeger, W., Soubrier, F., Suntharalingam, J., Toshner, M. R., Trembath, R. C., Vonk Noordegraaf, A., Wort, S. J., Wilkins, M. R., Yu, P. B., Li, W., Gräf, S., Upton, P. D., and Morrell, N. W. (2020). Characterization of mutations and levels of BMP9 and BMP10 in pulmonary arterial hypertension. *Am. J. Respir. Crit. Care Med.*, 201(5):575–585.

[216] Hoeper, M. M., Behr, J., Held, M., Grunig, E., Vizza, C. D., Vonk-Noordegraaf, A., Lange, T. J., Claussen, M., Grohé, C., Klose, H., Olsson, K. M., Zelniker, T., Neurohr, C., Distler, O., Wirtz, H., Opitz, C., Huscher, D., Pittrow, D., and Gibbs, J. S. R. (2015). Pulmonary hypertension in patients with chronic fibrosing idiopathic interstitial pneumonias. *PLOS ONE*, 10(12):e0141911.

[217] Hoeper, M. M., Humbert, M., Souza, R., Idrees, M., Kawut, S. M., Sliwa-Hahnle, K., Jing, Z.-C., and Gibbs, J. S. R. (2016). A global view of pulmonary hypertension. *Lancet Respir Med*, 4(4):306–322.

[218] Hoeper, M. M., Huscher, D., Ghofrani, H. A., Delcroix, M., Distler, O., Schweiger, C., Grunig, E., Staehler, G., Rosenkranz, S., Halank, M., Held, M., Grohé, C., Lange, T. J., Behr, J., Klose, H., Wilkens, H., Filusch, A., Germann, M., Ewert, R., Seyfarth, H.-J., Olsson, K. M., Opitz, C. F., Gaine, S. P., Vizza, C. D., Vonk-Noordegraaf, A., Kaemmerer, H., Gibbs, J. S. R., and Pittrow, D. (2013). Elderly patients diagnosed with idiopathic pulmonary arterial hypertension: results from the COMPERA registry. *Int. J. Cardiol.*, 168(2):871–880.

[219] Hoeper, M. M., Kramer, T., Pan, Z., Eichstaedt, C. A., Spiesshoefer, J., Benjamin, N., Olsson, K. M., Meyer, K., Vizza, C. D., Vonk-Noordegraaf, A., Distler, O., Opitz, C., Gibbs, J. S. R., Delcroix, M., Ghofrani, H. A., Huscher, D., Pittrow, D., Rosenkranz, S., and Grünig, E. (2017). Mortality in pulmonary arterial hypertension: prediction by the

2015 european pulmonary hypertension guidelines risk stratification model. *Eur. Respir. J.*, 50(2).

[220] Hoeper, M. M., Pausch, C., Grünig, E., Klose, H., Staehler, G., Huscher, D., Pittrow, D., Olsson, K. M., Vizza, C. D., Gall, H., Benjamin, N., Distler, O., Opitz, C., Gibbs, J. S. R., Delcroix, M., Ghofrani, H. A., Rosenkranz, S., Ewert, R., Kaemmerer, H., Lange, T. J., Kabitz, H.-J., Skowasch, D., Skride, A., Jureviciene, E., Paleviciute, E., Miliauskas, S., Claussen, M., Behr, J., Milger, K., Halank, M., Wilkens, H., Wirtz, H., Pfeuffer-Jovic, E., Harbaum, L., Scholtz, W., Dumitrescu, D., Bruch, L., Coghlan, G., Neurohr, C., Tsangaris, I., Gorenflo, M., Scelsi, L., Vonk-Noordegraaf, A., Ulrich, S., and Held, M. (2020). Idiopathic pulmonary arterial hypertension phenotypes determined by cluster analysis from the COMPERA registry. *The Journal of Heart and Lung Transplantation*, 39(12):1435–1444.

[221] Hon, C.-C., Ramilowski, J. A., Harshbarger, J., Bertin, N., Rackham, O. J. L., Gough, J., Denisenko, E., Schmeier, S., Poulsen, T. M., Severin, J., Lizio, M., Kawaji, H., Kasukawa, T., Itoh, M., Burroughs, A. M., Noma, S., Djebali, S., Alam, T., Medvedeva, Y. A., Testa, A. C., Lipovich, L., Yip, C.-W., Abugessaisa, I., Mendez, M., Hasegawa, A., Tang, D., Lassmann, T., Heutink, P., Babina, M., Wells, C. A., Kojima, S., Nakamura, Y., Suzuki, H., Daub, C. O., de Hoon, M. J. L., Arner, E., Hayashizaki, Y., Carninci, P., and Forrest, A. R. R. (2017). An atlas of human long non-coding RNAs with accurate 5' ends. *Nature*, 543(7644):199–204.

[222] Honaker, J., King, G., and Blackwell, M. (2011). Amelia II: A program for missing data. *Journal of Statistical Software*, 45(7):1–47.

[223] Hong, K.-H., Lee, Y. J., Lee, E., Park, S. O., Han, C., Beppu, H., Li, E., Raizada, M. K., Bloch, K. D., and Oh, S. P. (2008). Genetic ablation of the BMPR2 gene in pulmonary endothelium is sufficient to predispose to pulmonary arterial hypertension. *Circulation*, 118(7):722–730.

[224] Horn, M., Ries, A., Neveu, C., and Moser, K. (1983). Restrictive ventilatory pattern in precapillary pulmonary hypertension. *Am. Rev. Respir. Dis.*, 128(1):163–165.

[225] Huang, Z., Wang, D., Ihida-Stansbury, K., Jones, P. L., and Martin, J. F. (2009). Defective pulmonary vascular remodeling in smad8 mutant mice. *Hum. Mol. Genet.*, 18(15):2791–2801.

[226] Hughes, J. (2009). Interpreting pulmonary function tests. *Breathe*, pages 102–110.

[227] Huh, T.-L., Kim, Y.-O., Oh, I.-U., Song, B. J., and Inazawa, J. (1996). Assignment of the human mitochondrial NAD-specific isocitrate dehydrogenase subunit (IDH3a) gene to 15q25.1 $\rightarrow$ q25.2 byin SituHybridization. *Genomics*, 32(2):295–296.

[228] Humbert, M., Guignabert, C., Bonnet, S., Dorfmüller, P., Klinger, J. R., Nicolls, M. R., Olschewski, A. J., Pullamsetti, S. S., Schermuly, R. T., Stenmark, K. R., and Rabinovitch, M. (2019). Pathology and pathobiology of pulmonary hypertension: state of the art and research perspectives. *Eur. Respir. J.*, 53(1).

[229] Humbert, M., Sitbon, O., Chaouat, A., Bertocchi, M., Habib, G., Gressin, V., Yaici, A., Weitzenblum, E., Cordier, J.-F., Chabot, F., Dromer, C., Pison, C., Reynaud-Gaubert, M., Haloun, A., Laurent, M., Hachulla, E., and Simonneau, G. (2006). Pulmonary arterial hypertension in france: results from a national registry. *Am. J. Respir. Crit. Care Med.*, 173(9):1023–1030.

[230] Hunt, K. A., Mistry, V., Bockett, N. A., Ahmad, T., Ban, M., Barker, J. N., Barrett, J. C., Blackburn, H., Brand, O., Burren, O., Capon, F., Compston, A., Gough, S. C. L., Jostins, L., Kong, Y., Lee, J. C., Lek, M., MacArthur, D. G., Mansfield, J. C., Mathew, C. G., Mein, C. A., Mirza, M., Nutland, S., Onengut-Gumuscu, S., Papouli, E., Parkes, M., Rich, S. S., Sawcer, S., Satsangi, J., Simmonds, M. J., Trembath, R. C., Walker, N. M., Wozniak, E., Todd, J. A., Simpson, M. A., Plagnol, V., and van Heel, D. A. (2013). Negligible impact of rare autoimmune-locus coding-region variants on missing heritability. *Nature*, 498(7453):232–235.

[231] Hurdman, J., Condliffe, R., Elliot, C., Davies, C., Hill, C., Wild, J., Capener, D., Sephton, P., Hamilton, N., Armstrong, I., Billings, C., Lawrie, A., Sabroe, I., Akil, M., O'Toole, L., and Kiely, D. (2011). ASPIRE registry: Assessing the spectrum of pulmonary hypertension identified at a REferral centre. *European Respiratory Journal*, 39(4):945–955.

[232] Hurdman, J., Condliffe, R., Elliot, C. A., Davies, C., Hill, C., Wild, J. M., Capener, D., Sephton, P., Hamilton, N., Armstrong, I. J., Billings, C., Lawrie, A., Sabroe, I., Akil, M., O'Toole, L., and Kiely, D. G. (2012). ASPIRE registry: assessing the spectrum of pulmonary hypertension identified at a REferral centre. *Eur. Respir. J.*, 39(4):945–955.

[233] Hurdman, J., Condliffe, R., Elliot, C. A., Swift, A., Rajaram, S., Davies, C., Hill, C., Hamilton, N., Armstrong, I. J., Billings, C., Pollard, L., Wild, J. M., Lawrie, A., Lawson, R., Sabroe, I., and Kiely, D. G. (2013). Pulmonary hypertension in COPD: results from the ASPIRE registry. *Eur. Respir. J.*, 41(6):1292–1301.

[234] Hurst, L. A., Dunmore, B. J., Long, L., Crosby, A., Al-Lamki, R., Deighton, J., Southwood, M., Yang, X., Nikolic, M. Z., Herrera, B., Inman, G. J., Bradley, J. R., Rana, A. A., Upton, P. D., and Morrell, N. W. (2017). TNFα drives pulmonary arterial hypertension by suppressing the BMP type-II receptor and altering NOTCH signalling. *Nat. Commun.*, 8:14079.

[235] Husson, T., Duboc, J.-B., Quenez, O., Charbonnier, C., Rotharmel, M., Cuenca, M., Jegouzo, X., Richard, A.-C., Frebourg, T., Deleuze, J.-F., Boland, A., Genin, E., Debette, S., Tzourio, C., Campion, D., Nicolas, G., Guillin, O., and FREX Consortium (2018). Identification of potential genetic risk factors for bipolar disorder by whole-exome sequencing. *Transl. Psychiatry*, 8(1):268.

[236] Hyun, K., Jeon, J., Park, K., and Kim, J. (2017). Writing, erasing and reading histone lysine methylations. *Exp. Mol. Med.*, 49(4):e324.

[237] Iannuzzi, M. C. and Baughman, R. P. (2007). Reverse phenotyping in sarcoidosis. *Am. J. Respir. Crit. Care Med.*, 175(1):4–5.

[238] Idrees, M., Alnajashi, K., Abdulhameed, J., Khan, A., Batubara, E., Alotay, A., Fayed, A., Aldammas, S., Alseif, M., Alawwad, H., Abusabaa, Y., Almobrad, M., Kashour,

T., and SAPH, T. R. T. (2015). Saudi experience in the management of pulmonary arterial hypertension the outcome of PAH therapy with the exclusion of chronic parenteral prostacyclin. *Annals of Thoracic Medicine*, 10(3):204.

[239] Inscore, S. C., Stenmark, K. R., Orton, C., and Irvin, C. G. (1991). Neonatal calves develop airflow limitation due to chronic hypobaric hypoxia. *J. Appl. Physiol.*, 70(1):384–390.

[240] International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. *Nature*, 431(7011):931–945.

[241] International PPH Consortium, Lane, K. B., Machado, R. D., Pauciulo, M. W., Thomson, J. R., Phillips, 3rd, J. A., Loyd, J. E., Nichols, W. C., and Trembath, R. C. (2000). Heterozygous germline mutations in BMPR2, encoding a TGF-beta receptor, cause familial primary pulmonary hypertension. *Nat. Genet.*, 26(1):81–84.

[242] Ioannidis, N. M., Rothstein, J. H., Pejaver, V., Middha, S., McDonnell, S. K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., Cannon-Albright, L. A., Teerlink, C. C., Stanford, J. L., Isaacs, W. B., Xu, J., Cooney, K. A., Lange, E. M., Schleutker, J., Carpten, J. D., Powell, I. J., Cussenot, O., Cancel-Tassin, G., Giles, G. G., MacInnis, R. J., Maier, C., Hsieh, C.-L., Wiklund, F., Catalona, W. J., Foulkes, W. D., Mandal, D., Eeles, R. A., Kote-Jarai, Z., Bustamante, C. D., Schaid, D. J., Hastie, T., Ostrander, E. A., Bailey-Wilson, J. E., Radivojac, P., Thibodeau, S. N., Whittemore, A. S., and Sieh, W. (2016). REVEL: An ensemble method for predicting the pathogenicity of rare missense variants. *Am. J. Hum. Genet.*, 99(4):877–885.

[243] Ionita-Laza, I., Buxbaum, J. D., Laird, N. M., and Lange, C. (2011). A new testing strategy to identify rare variants with either risk or protective effect on disease. *PLoS Genetics*, 7(2):e1001289.

[244] Ionita-Laza, I., Lee, S., Makarov, V., Buxbaum, J. D., and Lin, X. (2013). Family-based association tests for sequence data, and comparisons with population-based association tests. *Eur. J. Hum. Genet.*, 21(10):1158–1162.

[245] Itokawa, T., Nokihara, H., Nishioka, Y., Sone, S., Iwamoto, Y., Yamada, Y., Cherrington, J., McMahon, G., Shibuya, M., Kuwano, M., and Ono, M. (2002). Antiangiogenic effect by SU5416 is partly attributable to inhibition of flt-1 receptor signaling. *Mol. Cancer Ther.*, 1(5):295–302.

[246] Jacher, J. E., Martin, L. J., Chung, W. K., Loyd, J. E., and Nichols, W. C. (2017). Pulmonary arterial hypertension: Specialists' knowledge, practices, and attitudes of genetic counseling and genetic testing in the USA. *Pulmonary Circulation*, 7(2):372–383.

[247] Jandl, K., Thekkekara Puthenparampil, H., Marsh, L. M., Hoffmann, J., Wilhelm, J., Veith, C., Sinn, K., Klepetko, W., Olschewski, H., Olschewski, A., Brock, M., and Kwapiszewska, G. (2019). Long non-coding RNAs influence the transcriptome in pulmonary arterial hypertension: the role of PAXIP1-AS1. *J. Pathol.*, 247(3):357–370.

[248] Jiang, Z., Lu, Z., Kou, S., Feng, T., Wei, Y., Gao, Z., Deng, D., Meng, J., Lin, C.-P., Zhou, B., and Zhang, H. (2021). Overexpression of kdr in adult endocardium induces endocardial neovascularization and improves heart function after myocardial infarction. *Cell Res.*, 31(4):485–487.

[249] Jing, Z.-C., Xu, X.-Q., Badesch, D. B., Jiang, X., Wu, Y., Liu, J.-M., Wang, Y., Pan, L., Li, H.-P., Pu, J.-L., Zhang, Z.-L., and Yang, Y.-J. (2009). Pulmonary function testing in patients with pulmonary arterial hypertension. *Respir. Med.*, 103(8):1136–1142.

[250] Jing, Z.-C., Xu, X.-Q., Han, Z.-Y., Wu, Y., Deng, K.-W., Wang, H., Wang, Z.-W., Cheng, X.-S., Xu, B., Hu, S.-S., Hui, R.-T., and Yang, Y.-J. (2007). Registry and survival study in chinese patients with idiopathic and familial pulmonary arterial hypertension. *Chest*, 132(2):373–379.

[251] Joung, J. K. and Sander, J. D. (2012). TALENs: a widely applicable technology for targeted genome editing. *Nature Reviews Molecular Cell Biology*, 14(1):49–55.

[252] Kallunki, T., Barisic, M., Jäättelä, M., and Liu, B. (2019). How to choose the right inducible gene expression system for mammalian studies? *Cells*, 8(8).

[253] Kang, H., Louie, J., Weisman, A., Sheu-Gruttadauria, J., Davis-Dusenbery, B. N., Lagna, G., and Hata, A. (2012). Inhibition of microRNA-302 (mir-302) by bone morphogenetic protein 4 (BMP4) facilitates the BMP signaling pathway. *J. Biol. Chem.*, 287(46):38656–38664.

[254] Karczewski, K. J., Francioli, L. C., Tiao, G., Cummings, B. B., Alföldi, J., Wang, Q., Collins, R. L., Laricchia, K. M., Ganna, A., Birnbaum, D. P., Gauthier, L. D., Brand, H., Solomonson, M., Watts, N. A., Rhodes, D., Singer-Berk, M., England, E. M., Seaby, E. G., Kosmicki, J. A., Walters, R. K., Tashman, K., Farjoun, Y., Banks, E., Poterba, T., Wang, A., Seed, C., Whiffin, N., Chong, J. X., Samocha, K. E., Pierce-Hoffman, E., Zappala, Z., O'Donnell-Luria, A. H., Minikel, E. V., Weisburd, B., Lek, M., Ware, J. S., Vittal, C., Armean, I. M., Bergelson, L., Cibulskis, K., Connolly, K. M., Covarrubias, M., Donnelly, S., Ferriera, S., Gabriel, S., Gentry, J., Gupta, N., Jeandet, T., Kaplan, D., Llanwarne, C., Munshi, R., Novod, S., Petrillo, N., Roazen, D., Ruano-Rubio, V., Saltzman, A., Schleicher, M., Soto, J., Tibbetts, K., Tolonen, C., Wade, G., Talkowski, M. E., Genome Aggregation Database Consortium, Neale, B. M., Daly, M. J., and MacArthur, D. G. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature*, 581(7809):434–443.

[255] Karolak, J. A., Vincent, M., Deutsch, G., Gambin, T., Cogné, B., Pichon, O., Vetrini, F., Mefford, H. C., Dines, J. N., Golden-Grant, K., Dipple, K., Freed, A. S., Leppig, K. A., Dishop, M., Mowat, D., Bennetts, B., Gifford, A. J., Weber, M. A., Lee, A. F., Boerkoel, C. F., Bartell, T. M., Ward-Melver, C., Besnard, T., Petit, F., Bache, I., Tümer, Z., Denis-Musquer, M., Joubert, M., Martinovic, J., Bénéteau, C., Molin, A., Carles, D., André, G., Bieth, E., Chassaing, N., Devisme, L., Chalabreysse, L., Pasquier, L., Secq, V., Don, M., Orsaria, M., Missirian, C., Mortreux, J., Sanlaville, D., Pons, L., Küry, S., Bézieau, S., Liet, J.-M., Joram, N., Bihouée, T., Scott, D. A., Brown, C. W., Scaglia, F., Tsai, A. C.-H., Grange, D. K., Phillips, 3rd, J. A., Pfotenhauer, J. P., Jhangiani, S. N., Gonzaga-Jauregui, C. G., Chung, W. K., Schauer, G. M., Lipson, M. H., Mercer, C. L., van Haeringen, A., Liu, Q., Popek, E., Coban Akdemir, Z. H., Lupski, J. R., Szafranski, P., Isidor, B., Le Caignec, C., and Stankiewicz, P. (2019). Complex compound inheritance of lethal lung developmental disorders due to disruption of the TBX-FGF pathway. *Am. J. Hum. Genet.*, 104(2):213–228.

[256] Kasahara, Y., Tuder, R. M., Taraseviciene-Stewart, L., Le Cras, T. D., Abman, S., Hirth, P. K., Waltenberger, J., and Voelkel, N. F. (2000). Inhibition of VEGF receptors causes lung cell apoptosis and emphysema. *J. Clin. Invest.*, 106(11):1311–1319.

[257] Kassambara, A., Kosinski, M., and Biecek, P. (2019). *survminer: Drawing Survival Curves using 'ggplot2'*. R package version 0.4.6.

[258] Kassambara, A. and Mundt, F. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. R package version 1.0.7.

[259] Kato, G. J., Piel, F. B., Reid, C. D., Gaston, M. H., Ohene-Frempong, K., Krishnamurti, L., Smith, W. R., Panepinto, J. A., Weatherall, D. J., Costa, F. F., and Vichinsky, E. P. (2018). Sickle cell disease. *Nat Rev Dis Primers*, 4:18010.

[260] Kaufman, L. and Rousseeuw, P. J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis*. John Wiley & Sons.

[261] Kaunisto, M. A., Harno, H., Vanmolkot, K. R. J., Gargus, J. J., Sun, G., Hämäläinen, E., Liukkonen, E., Kallela, M., van den Maagdenberg, A. M. J. M., Frants, R. R., Färkkilä, M., Palotie, A., and Wessman, M. (2004). A novel missense ATP1a2 mutation in a finnish family with familial hemiplegic migraine type 2. *Neurogenetics*, 5(2):141–146.

[262] Kerstjens-Frederikse, W. S., Bongers, E. M. H. F., Roofthooft, M. T. R., Leter, E. M., Douwes, J. M., Van Dijk, A., Vonk-Noordegraaf, A., Dijk-Bos, K. K., Hoefsloot, L. H., Hoendermis, E. S., Gille, J. J. P., Sikkema-Raddatz, B., Hofstra, R. M. W., and Berger, R. M. F. (2013). TBX4 mutations (small patella syndrome) are associated with childhood-onset pulmonary arterial hypertension. *J. Med. Genet.*, 50(8):500–506.

[263] Khan, A. S., Lynch, C. D., Sane, D. C., Willingham, M. C., and Sonntag, W. E. (2001). Growth hormone increases regional coronary blood flow and capillary density in aged rats. *J. Gerontol. A Biol. Sci. Med. Sci.*, 56(8):B364–71.

[264] Kim, D. H. and Rossi, J. J. (2008). RNAi mechanisms and applications. *BioTechniques*, 44(5):613–616.

[265] Kim, J., Weber, J. A., Jho, S., Jang, J., Jun, J., Cho, Y. S., Kim, H.-M., Kim, H., Kim, Y., Chung, O., Kim, C. G., Lee, H., Kim, B. C., Han, K., Koh, I., Chae, K. S., Lee, S., Edwards, J. S., and Bhak, J. (2018). KoVariome: Korean national standard reference variome database of whole genomes with comprehensive SNV, indel, CNV, and SV analyses. *Sci. Rep.*, 8(1):5677.

[266] Kim, K., Kim, I.-K., Yang, J. M., Lee, E., Koh, B. I., Song, S., Park, J., Lee, S., Choi, C., Kim, J. W., Kubota, Y., Koh, G. Y., and Kim, I. (2016). SoxF transcription factors are positive feedback regulators of VEGF signaling. *Circ. Res.*, 119(7):839–852.

[267] Kircher, M., Witten, D. M., Jain, P., O'Roak, B. J., Cooper, G. M., and Shendure, J. (2014). A general framework for estimating the relative pathogenicity of human genetic variants. *Nat. Genet.*, 46(3):310–315.

[268] Köhler, S., Carmody, L., Vasilevsky, N., Jacobsen, J. O. B., Danis, D., Gourdine, J.-P., Gargano, M., Harris, N. L., Matentzoglu, N., McMurry, J. A., Osumi-Sutherland, D., Cipriani, V., Balhoff, J. P., Conlin, T., Blau, H., Baynam, G., Palmer, R., Gratian, D., Dawkins, H., Segal, M., Jansen, A. C., Muaz, A., Chang, W. H., Bergerson, J., Lauledeerkind, S. J. F., Yüksel, Z., Beltran, S., Freeman, A. F., Sergouniotis, P. I., Durkin, D., Storm, A. L., Hanauer, M., Brudno, M., Bello, S. M., Sincan, M., Rageth, K., Wheeler, M. T., Oegema, R., Lourghi, H., Della Rocca, M. G., Thompson, R., Castellanos, F., Priest, J., Cunningham-Rundles, C., Hegde, A., Lovering, R. C., Hajek, C., Olry, A., Notarangelo, L., Similuk, M., Zhang, X. A., Gómez-Andrés, D., Lochmüller, H., Dollfus, H., Rosenzweig, S., Marwaha, S., Rath, A., Sullivan, K., Smith, C., Milner, J. D., Leroux, D., Boerkoel, C. F., Klion, A., Carter, M. C., Groza, T., Smedley, D., Haendel, M. A., Mungall, C., and Robinson, P. N. (2019). Expansion of the human phenotype ontology (HPO) knowledge base and resources. *Nucleic Acids Res.*, 47(D1):D1018–D1027.

[269] Köhler, S., Doelken, S. C., Mungall, C. J., Bauer, S., Firth, H. V., Bailleul-Forestier, I., Black, G. C. M., Brown, D. L., Brudno, M., Campbell, J., FitzPatrick, D. R., Eppig, J. T., Jackson, A. P., Freson, K., Girdea, M., Helbig, I., Hurst, J. A., Jähn, J., Jackson, L. G., Kelly, A. M., Ledbetter, D. H., Mansour, S., Martin, C. L., Moss, C., Mumford, A., Ouwehand, W. H., Park, S.-M., Riggs, E. R., Scott, R. H., Sisodiya, S., Van Vooren, S., Wapner, R. J., Wilkie, A. O. M., Wright, C. F., Vulto-van Silfhout, A. T., de Leeuw, N., de Vries, B. B. A., Washingthon, N. L., Smith, C. L., Westerfield, M., Schofield, P., Ruef, B. J., Gkoutos, G. V., Haendel, M., Smedley, D., Lewis, S. E., and Robinson, P. N. (2014). The human phenotype ontology project: linking molecular biology and disease through phenotype data. *Nucleic Acids Res.*, 42(Database issue):D966–74.

[270] Köhler, S., Schulz, M. H., Krawitz, P., Bauer, S., Dölken, S., Ott, C. E., Mundlos, C., Horn, D., Mundlos, S., and Robinson, P. N. (2009). Clinical diagnostics in human genetics with semantic similarity searches in ontologies. *Am. J. Hum. Genet.*, 85(4):457–464.

[271] Kojonazarov, B., Hadzic, S., Ghofrani, H. A., Grimminger, F., Seeger, W., Weissmann, N., and Schermuly, R. T. (2019). Severe emphysema in the SU5416/Hypoxia rat model of pulmonary hypertension. *Am. J. Respir. Crit. Care Med.*, 200(4):515–518.

[272] Konstantinides, S., Kasper, W., Geibel, A., Hofmann, T., Köster, W., and Just, H. (1993). Detection of left-to-right shunt in atrial septal defect by negative contrast echocardiography: a comparison of transthoracic and transesophageal approach. *Am. Heart J.*, 126(4):909–917.

[273] Koolen, D. A., Kramer, J. M., Neveling, K., Nillesen, W. M., Moore-Barton, H. L., Elmslie, F. V., Toutain, A., Amiel, J., Malan, V., Tsai, A. C.-H., Cheung, S. W., Gilissen, C., Verwiel, E. T. P., Martens, S., Feuth, T., Bongers, E. M. H. F., de Vries, P., Scheffer, H., Vissers, L. E. L. M., de Brouwer, A. P. M., Brunner, H. G., Veltman, J. A., Schenck, A., Yntema, H. G., and de Vries, B. B. A. (2012). Mutations in the chromatin modifier gene KANSL1 cause the 17q21.31 microdeletion syndrome. *Nat. Genet.*, 44(6):639–641.

[274] Kopanos, C., Tsiolkas, V., Kouris, A., Chapple, C. E., Albarca Aguilera, M., Meyer, R., and Massouras, A. (2019). VarSome: the human genomic variant search engine. *Bioinformatics*, 35(11):1978–1980.

[275] Korsholm, K., Andersen, A., Kirkfeldt, R. E., Hansen, K. N., Mellemkjær, S., and Nielsen-Kudsk, J. E. (2015). Survival in an incident cohort of patients with pulmonary arterial hypertension in denmark. *Pulm. Circ.*, 5(2):364–369.

[276] Koutroulou, I., Tsivgoulis, G., Tsalikakis, D., Karacostas, D., Grigoriadis, N., and Karapanayiotides, T. (2020). Epidemiology of patent foramen ovale in general population and in stroke patients: A narrative review. *Front. Neurol.*, 11:281.

[277] Kowarik, A. and Templ, M. (2016). Imputation with the R package VIM. *Journal of Statistical Software*, 74(7):1–16.

[278] Kuo, C. J., LaMontagne, Jr, K. R., Garcia-Cardeña, G., Ackley, B. D., Kalman, D., Park, S., Christofferson, R., Kamihara, J., Ding, Y. H., Lo, K. M., Gillies, S., Folkman, J., Mulligan, R. C., and Javaherian, K. (2001). Oligomerization-dependent regulation of motility and morphogenesis by the collagen XVIII NC1/endostatin domain. *J. Cell Biol.*, 152(6):1233–1246.

[279] Kurzyna, M., Dabrowski, M., Bielecki, D., Fijalkowska, A., Pruszczyk, P., Opolski, G., Burakowski, J., Florczyk, M., Tomkowski, W. Z., Wawrzynska, L., Szturmowicz, M., and Torbicki, A. (2007). Atrial septostomy in treatment of end-stage right heart failure in patients with pulmonary hypertension. *Chest*, 131(4):977–983.

[280] Kwasnicka-Crawford, D., Carson, A., Roberts, W., Summers, A., Rehnstrom, K., Jarvela, I., and Scherer, S. (2005). Characterization of a novel cation transporter ATPase gene (ATP13a4) interrupted by 3q25–q29 inversion in an individual with language delay. *Genomics*, 86(2):182–194.

[281] Kylhammar, D., Kjellström, B., Hjalmarsson, C., Jansson, K., Nisell, M., Söderberg, S., Wikström, G., and Rådegran, G. (2017). A comprehensive risk stratification at early follow-up determines prognosis in pulmonary arterial hypertension. *European Heart Journal*, 39(47):4175–4181.

[282] Lago-Docampo, M., Tenorio, J., Hernández-González, I., Pérez-Olivares, C., Escribano-Subías, P., Pousada, G., Baloira, A., Arenas, M., Lapunzina, P., and Valverde, D. (2020). Characterization of rare ABCC8 variants identified in spanish pulmonary arterial hypertension patients. *Scientific Reports*, 10(1).

[283] Lahm, H., Jia, M., Dreßen, M., Wirth, F., Puluca, N., Gilsbach, R., Keavney, B. D., Cleuziou, J., Beck, N., Bondareva, O., Dzilic, E., Burri, M., König, K. C., Ziegelmüller, J. A., Abou-Ajram, C., Neb, I., Zhang, Z., Doppler, S. A., Mastantuono, E., Lichtner, P., Eckstein, G., Hörer, J., Ewert, P., Priest, J. R., Hein, L., Lange, R., Meitinger, T., Cordell, H. J., Müller-Myhsok, B., and Krane, M. (2021). Congenital heart disease risk loci identified by genome-wide association study in european patients. *Journal of Clinical Investigation*, 131(2).

[284] Lähteenvuo, J. and Rosenzweig, A. (2012). Effects of aging on angiogenesis. *Circ. Res.*, 110(9):1252–1264.

[285] Lai, Y. L., Olson, J. W., and Gillespie, M. N. (1991). Ventilatory dysfunction precedes pulmonary vascular changes in monocrotaline-treated rats. *J. Appl. Physiol.*, 70(2):561–566.

[286] Lambert, M., Capuano, V., Boet, A., Tesson, L., Bertero, T., Nakhleh, M. K., Remy, S., Anegon, I., Pechoux, C., Hautefort, A., Rucker-Martin, C., Manoury, B., Domergue, V., Mercier, O., Girerd, B., Montani, D., Perros, F., Humbert, M., and Antigny, F. (2019). Characterization of kcnk3 -mutated rat, a novel model of pulmonary hypertension. *Circulation Research*, 125(7):678–695.

[287] Lambert, M., Capuano, V., Olschewski, A., Sabourin, J., Nagaraj, C., Girerd, B., Weatherald, J., Humbert, M., and Antigny, F. (2018). Ion channels in pulmonary hypertension: A therapeutic interest? *Int. J. Mol. Sci.*, 19(10).

[288] Landini, S., Mazzinghi, B., Becherucci, F., Allinovi, M., Provenzano, A., Palazzo, V., Ravaglia, F., Artuso, R., Bosi, E., Stagi, S., Sansavini, G., Guzzi, F., Cirillo, L., Vaglio, A., Murer, L., Peruzzi, L., Pasini, A., Materassi, M., Roperto, R. M., Anders, H.-J., Rotondi, M., Giglio, S. R., and Romagnani, P. (2020). Reverse phenotyping after Whole-Exome sequencing in Steroid-Resistant nephrotic syndrome. *Clin. J. Am. Soc. Nephrol.*, 15(1):89–100.

[289] Lange, A. W., Haitchi, H. M., LeCras, T. D., Sridharan, A., Xu, Y., Wert, S. E., James, J., Udell, N., Thurner, P. J., and Whitsett, J. A. (2014). Sox17 is required for normal pulmonary vascular morphogenesis. *Dev. Biol.*, 387(1):109–120.

[290] Lantuéjoul, S., Sheppard, M. N., Corrin, B., Burke, M. M., and Nicholson, A. G. (2006). Pulmonary veno-occlusive disease and pulmonary capillary hemangiomatosis: a clinicopathologic study of 35 cases. *Am. J. Surg. Pathol.*, 30(7):850–857.

[291] Lappalainen, T., Scott, A. J., Brandt, M., and Hall, I. M. (2019). Genomic analysis in the age of human genome sequencing. *Cell*, 177(1):70–84.

[292] Larkin, E. K., Newman, J. H., Austin, E. D., Hemnes, A. R., Wheeler, L., Robbins, I. M., West, J. D., Phillips, 3rd, J. A., Hamid, R., and Loyd, J. E. (2012). Longitudinal analysis casts doubt on the presence of genetic anticipation in heritable pulmonary arterial hypertension. *Am. J. Respir. Crit. Care Med.*, 186(9):892–896.

[293] Laveneziana, P., Montani, D., Dorfmüller, P., Girerd, B., Sitbon, O., Jaïs, X., Savale, L., Eyries, M., Soubrier, F., Similowski, T., Simonneau, G., Humbert, M., and Garcia, G. (2014). Mechanisms of exertional dyspnoea in pulmonary veno-occlusive disease with EIF2ak4 mutations. *European Respiratory Journal*, 44(4):1069–1072.

[294] Lavoie, J. R., Ormiston, M. L., Perez-Iratxeta, C., Courtman, D. W., Jiang, B., Ferrer, E., Caruso, P., Southwood, M., Foster, W. S., Morrell, N. W., and Stewart, D. J. (2014). Proteomic analysis implicates translationally controlled tumor protein as a novel mediator of occlusive vascular remodeling in pulmonary arterial hypertension. *Circulation*, 129(21):2125–2135.

[295] Lawrie, A. (2014). A report on the use of animal models and phenotyping methods in pulmonary hypertension research. *Pulm. Circ.*, 4(1):2–9.

[296] Lee, H.-W., Adachi, T., Park, S., Kowalski, P., Anderson, D., Simons, M., Jin, S. W., and Chun, H. J. (2018). Bmpr1a and its role in endomt in pulmonary arterial hypertension. *Circulation*, 138(Suppl_1):A17070–A17070.

[297] Lee, S., Emond, M. J., Bamshad, M. J., Barnes, K. C., Rieder, M. J., Nickerson, D. A., NHLBI GO Exome Sequencing Project—ESP Lung Project Team, Christiani, D. C., Wurfel, M. M., and Lin, X. (2012). Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am. J. Hum. Genet.*, 91(2):224–237.

[298] Lee Heon-Woo, Adachi Takaomi, Park Saejeong, Pak Boryeong, Kowalski Piotr, Choi Woosoung, Clapham Katharine, Lee Ahram, Anderson Daniel, Simons Michael, Jin Sukwon, and Chun Hyung J (2019). Abstract 14616: Loss of endothelial BMPR1A results in pulmonary hypertension through endothelial to mesenchymal transition. *Circulation*, 140(Suppl_1):A14616–A14616.

[299] Legchenko, E., Liu, B., West, J., Vangheluwe, P., Upton, P., and Morrell, N. (2020). Protein truncating mutations in ATP13a3 promote pulmonary arterial hypertension. In *Pulmonary hypertension*. European Respiratory Society.

[300] Lepic, E., Burger, D., Lu, X., Song, W., and Feng, Q. (2006). Lack of endothelial nitric oxide synthase decreases cardiomyocyte proliferation and delays cardiac maturation. *American Journal of Physiology-Cell Physiology*, 291(6):C1240–C1246.

[301] Leter, E. M., Boonstra, A. B., Postma, F. B., Gille, J. J. P., Meijers-Heijboer, E. J., and Noordegraaf, A. V. (2011). Genetic counselling for pulmonary arterial hypertension: a matter of variable variability. *Netherlands Heart Journal*, 19(2):89–92.

[302] Leung, A., Trac, C., Jin, W., Lanting, L., Akbany, A., Sætrom, P., Schones, D. E., and Natarajan, R. (2013). Novel long noncoding RNAs are regulated by angiotensin II in vascular smooth muscle cells. *Circ. Res.*, 113(3):266–278.

[303] Levy, M., Eyries, M., Szezepanski, I., Ladouceur, M., Nadaud, S., Bonnet, D., and Soubrier, F. (2016). Genetic analyses in a cohort of children with pulmonary hypertension. *European Respiratory Journal*, 48(4):1118–1126.

[304] Lewandoski, M. (2001). Conditional control of gene expression in the mouse. *Nat. Rev. Genet.*, 2(10):743–755.

[305] Lewis, R. A., Thompson, A. A. R., Billings, C. G., Charalampopoulos, A., Elliot, C. A., Hamilton, N., Hill, C., Hurdman, J., Rajaram, S., Sabroe, I., Swift, A. J., Kiely, D. G., and Condliffe, R. (2020). Mild parenchymal lung disease and/or low diffusion capacity impacts survival and treatment response in patients diagnosed with idiopathic pulmonary arterial hypertension. *Eur. Respir. J.*

[306] Li, B. and Leal, S. M. (2008). Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. *Am. J. Hum. Genet.*, 83(3):311–321.

[307] Li, C., Stoma, S., Lotta, L. A., Warner, S., Albrecht, E., Allione, A., Arp, P. P., Broer, L., Buxton, J. L., Da Silva Couto Alves, A., Deelen, J., Fedko, I. O., Gordon, S. D., Jiang, T., Karlsson, R., Kerrison, N., Loe, T. K., Mangino, M., Milaneschi, Y., Miraglio, B., Pervjakova, N., Russo, A., Surakka, I., van der Spek, A., Verhoeven, J. E., Amin, N., Beekman, M., Blakemore, A. I., Canzian, F., Hamby, S. E., Hottenga, J.-J., Jones, P. D., Jousilahti, P., Mägi, R., Medland, S. E., Montgomery, G. W., Nyholt, D. R., Perola,

M., Pietiläinen, K. H., Salomaa, V., Sillanpää, E., Suchiman, H. E., van Heemst, D., Willemsen, G., Agudo, A., Boeing, H., Boomsma, D. I., Chirlaque, M.-D., Fagherazzi, G., Ferrari, P., Franks, P., Gieger, C., Eriksson, J. G., Gunter, M., Hägg, S., Hovatta, I., Imaz, L., Kaprio, J., Kaaks, R., Key, T., Krogh, V., Martin, N. G., Melander, O., Metspalu, A., Moreno, C., Onland-Moret, N. C., Nilsson, P., Ong, K. K., Overvad, K., Palli, D., Panico, S., Pedersen, N. L., Penninx, B. W. J. H., Quirós, J. R., Jarvelin, M. R., Rodríguez-Barranco, M., Scott, R. A., Severi, G., Slagboom, P. E., Spector, T. D., Tjonneland, A., Trichopoulou, A., Tumino, R., Uitterlinden, A. G., van der Schouw, Y. T., van Duijn, C. M., Weiderpass, E., Denchi, E. L., Matullo, G., Butterworth, A. S., Danesh, J., Samani, N. J., Wareham, N. J., Nelson, C. P., Langenberg, C., and Codd, V. (2020). Genome-wide association analysis in humans links nucleotide metabolism to leukocyte telomere length. *Am. J. Hum. Genet.*, 106(3):389–404.

[308] Li, D., Lewinger, J. P., Gauderman, W. J., Murcray, C. E., and Conti, D. (2011). Using extreme phenotype sampling to identify the rare causal variants of quantitative traits in association studies. *Genet. Epidemiol.*, 35(8):790–799.

[309] Li, R., Li, Y., Zheng, H., Luo, R., Zhu, H., Li, Q., Qian, W., Ren, Y., Tian, G., Li, J., Zhou, G., Zhu, X., Wu, H., Qin, J., Jin, X., Li, D., Cao, H., Hu, X., Blanche, H., Cann, H., Zhang, X., Li, S., Bolund, L., Kristiansen, K., Yang, H., Wang, J., and Wang, J. (2010). Building the sequence map of the human pan-genome. *Nat. Biotechnol.*, 28(1):57–63.

[310] Li, Y., Wang, Y., Zhang, M., Ma, Y., Wang, K., Liu, X., Hao, M., Sun, C., Wang, X., Chen, X., Zhang, Y., Duan, W., Kang, L., Qiao, B., Wang, J., and Jin, L. (2018). Using composite phenotypes to reveal hidden physiological heterogeneity and model oxygen saturation variation of high altitude acclimatization in a chinese han longitudinal cohort.

[311] Lin, D. (1998). An information-theoretic definition of similarity. In *In Proceedings of the 15th International Conference on Machine Learning*, pages 296–304. Morgan Kaufmann.

[312] Ling, Y., Johnson, M. K., Kiely, D. G., Condliffe, R., Elliot, C. A., Gibbs, J. S. R., Howard, L. S., Pepke-Zaba, J., Sheares, K. K. K., Corris, P. A., Fisher, A. J., Lordan, J. L., Gaine, S., Coghlan, J. G., Wort, S. J., Gatzoulis, M. A., and Peacock, A. J. (2012). Changing demographics, epidemiology, and survival of incident pulmonary arterial hypertension: results from the pulmonary hypertension registry of the united kingdom and ireland. *Am. J. Respir. Crit. Care Med.*, 186(8):790–796.

[313] Liu, B., Haimel, M., Bleda, M., Li, W., Gräf, S., Upton, P., and Morrell, N. (2018a). S42 characterizing atp13a3 loss of function in pulmonary arterial hypertension (pah). *Thorax*, 73(Suppl 4):A26–A27.

[314] Liu, J., Yang, J., Tang, X., Li, H., Shen, Y., Gu, W., and Zhao, S. (2020). Homozygous -related hereditary hemorrhagic telangiectasia in a chinese family. *Pediatrics*, 146(2).

[315] Liu, R., Hu, W., Li, X., Pu, D., Yang, G., Liu, H., Tan, M., and Zhu, D. (2019). Association of circulating BMP9 with coronary heart disease and hypertension in chinese populations. *BMC Cardiovascular Disorders*, 19(1).

[316] Liu, Y., Sun, Z., Zhu, J., Xiao, B., Dong, J., and Li, X. (2018b). LncRNA-TCONS_00034812 in cell proliferation and apoptosis of pulmonary artery smooth muscle cells and its mechanism. *Journal of Cellular Physiology*, 233(6):4801–4814.

[317] Long, L., MacLean, M. R., Jeffery, T. K., Morecroft, I., Yang, X., Rudarakanchana, N., Southwood, M., James, V., Trembath, R. C., and Morrell, N. W. (2006). Serotonin increases susceptibility to pulmonary hypertension in BMPR2-deficient mice. *Circ. Res.*, 98(6):818–827.

[318] Long, L., Ormiston, M. L., Yang, X., Southwood, M., Gräf, S., Machado, R. D., Mueller, M., Kinzel, B., Yung, L. M., Wilkinson, J. M., Moore, S. D., Drake, K. M., Aldred, M. A., Yu, P. B., Upton, P. D., and Morrell, N. W. (2015). Selective enhancement of endothelial BMPR-II with BMP9 reverses pulmonary arterial hypertension. *Nat. Med.*, 21(7):777–785.

[319] Low, A., George, S., Howard, L., Bell, N., Millar, A., and Tulloh, R. M. R. (2018). Lung function, inflammation, and endothelin-1 in congenital heart disease–associated pulmonary arterial hypertension. *Journal of the American Heart Association*, 7(4).

[320] Loyd, J. E., Primm, R. K., and Newman, J. H. (1984). Familial primary pulmonary hypertension: clinical patterns. *Am. Rev. Respir. Dis.*, 129(1):194–197.

[321] Luo, F., Wang, X., Luo, X., Li, B., Zhu, D., Sun, H., and Tang, Y. (2019). Invasive hemodynamic assessment for the right ventricular system and Hypoxia-Induced pulmonary arterial hypertension in mice. *J. Vis. Exp.*, (152).

[322] Ma, L., Roman-Campos, D., Austin, E. D., Eyries, M., Sampson, K. S., Soubrier, F., Germain, M., Trégouët, D.-A., Borczuk, A., Rosenzweig, E. B., Girerd, B., Montani, D., Humbert, M., Loyd, J. E., Kass, R. S., and Chung, W. K. (2013a). A novel channelopathy in pulmonary arterial hypertension. *N. Engl. J. Med.*, 369(4):351–361.

[323] Ma, L., Roman-Campos, D., Austin, E. D., Eyries, M., Sampson, K. S., Soubrier, F., Germain, M., Trégouët, D.-A., Borczuk, A., Rosenzweig, E. B., Girerd, B., Montani, D., Humbert, M., Loyd, J. E., Kass, R. S., and Chung, W. K. (2013b). A novel channelopathy in pulmonary arterial hypertension. *N. Engl. J. Med.*, 369(4):351–361.

[324] Ma, T., Yang, B., Gillespie, A., Carlson, E. J., Epstein, C. J., and Verkman, A. S. (1998). Severely impaired urinary concentrating ability in transgenic mice lacking aquaporin-1 water channels. *J. Biol. Chem.*, 273(8):4296–4299.

[325] Ma, Z., Mao, L., and Rajagopal, S. (2016). Hemodynamic characterization of rodent models of pulmonary arterial hypertension. *J. Vis. Exp.*, (110).

[326] MacArthur, C. G., Hunter, D., and Gibson, G. J. (1979). Ventilatory function in the eisenmenger syndrome. *Thorax*, 34(3):348–353.

[327] Machado, R. D., Aldred, M. A., James, V., Harrison, R. E., Patel, B., Schwalbe, E. C., Gruenig, E., Janssen, B., Koehler, R., Seeger, W., Eickelberg, O., Olschewski, H., Elliott, C. G., Glissmeyer, E., Carlquist, J., Kim, M., Torbicki, A., Fijalkowska, A., Szewczyk, G., Parma, J., Abramowicz, M. J., Galie, N., Morisaki, H., Kyotani, S., Nakanishi, N., Morisaki, T., Humbert, M., Simonneau, G., Sitbon, O., Soubrier, F., Coulet, F., Morrell,

N. W., and Trembath, R. C. (2006). Mutations of the TGF-beta type II receptor BMPR2 in pulmonary arterial hypertension. *Hum. Mutat.*, 27(2):121–132.

[328] Machado, R. D., Eickelberg, O., Elliott, C. G., Geraci, M. W., Hanaoka, M., Loyd, J. E., Newman, J. H., Phillips, 3rd, J. A., Soubrier, F., Trembath, R. C., and Chung, W. K. (2009). Genetics and genomics of pulmonary arterial hypertension. *J. Am. Coll. Cardiol.*, 54(1 Suppl):S32–42.

[329] Machado, R. D., James, V., Southwood, M., Harrison, R. E., Atkinson, C., Stewart, S., Morrell, N. W., Trembath, R. C., and Aldred, M. A. (2005). Investigation of second genetic hits at the BMPR2 locus as a modulator of disease progression in familial pulmonary arterial hypertension. *Circulation*, 111(5):607–613.

[330] Machado, R. D., Southgate, L., Eichstaedt, C. A., Aldred, M. A., Austin, E. D., Best, D. H., Chung, W. K., Benjamin, N., Elliott, C. G., Eyries, M., Fischer, C., Gräf, S., Hinderhofer, K., Humbert, M., Keiles, S. B., Loyd, J. E., Morrell, N. W., Newman, J. H., Soubrier, F., Trembath, R. C., Viales, R. R., and Grünig, E. (2015). Pulmonary arterial hypertension: A current perspective on established and emerging molecular genetic defects. *Hum. Mutat.*, 36(12):1113–1127.

[331] MacIntyre, N. (2005). Standardisation of the single-breath determination of carbon monoxide uptake in the lung. *European Respiratory Journal*, 26(4):720–735.

[332] MacNee, W., Rabinovich, R. A., and Choudhury, G. (2014). Ageing and the border between health and disease. *European Respiratory Journal*, 44(5):1332–1352.

[333] Madan, M., Patel, A., Skruber, K., Geerts, D., Altomare, D. A., and Iv, O. P. (2016). ATP13A3 and caveolin-1 as potential biomarkers for difluoromethylornithine-based therapies in pancreatic cancers. *Am. J. Cancer Res.*, 6(6):1231–1252.

[334] Madsen, B. E. and Browning, S. R. (2009). A groupwise association test for rare mutations using a weighted sum statistic. *PLoS Genet.*, 5(2):e1000384.

[335] Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., and Hornik, K. (2019). *cluster: Cluster Analysis Basics and Extensions*. R package version 2.1.0 — For new features, see the 'Changelog' file (in the package source).

[336] Mahajan, A., Taliun, D., Thurner, M., Robertson, N. R., Torres, J. M., Rayner, N. W., Payne, A. J., Steinthorsdottir, V., Scott, R. A., Grarup, N., Cook, J. P., Schmidt, E. M., Wuttke, M., Sarnowski, C., Mägi, R., Nano, J., Gieger, C., Trompet, S., Lecoeur, C., Preuss, M. H., Prins, B. P., Guo, X., Bielak, L. F., Below, J. E., Bowden, D. W., Chambers, J. C., Kim, Y. J., Ng, M. C. Y., Petty, L. E., Sim, X., Zhang, W., Bennett, A. J., Bork-Jensen, J., Brummett, C. M., Canouil, M., Ec Kardt, K.-U., Fischer, K., Kardia, S. L. R., Kronenberg, F., Läll, K., Liu, C.-T., Locke, A. E., Luan, J., Ntalla, I., Nylander, V., Schönherr, S., Schurmann, C., Yengo, L., Bottinger, E. P., Brandslund, I., Christensen, C., Dedoussis, G., Florez, J. C., Ford, I., Franco, O. H., Frayling, T. M., Giedraitis, V., Hackinger, S., Hattersley, A. T., Herder, C., Ikram, M. A., Ingelsson, M., Jørgensen, M. E., Jørgensen, T., Kriebel, J., Kuusisto, J., Ligthart, S., Lindgren, C. M., Linneberg, A., Lyssenko, V., Mamakou, V., Meitinger, T., Mohlke, K. L., Morris, A. D., Nadkarni, G., Pankow, J. S., Peters, A., Sattar, N., Stančáková, A., Strauch, K., Taylor, K. D., Thorand, B., Thorleifsson, G., Thorsteinsdottir, U., Tuomilehto, J., Witte, D. R., Dupuis, J., Peyser,

P. A., Zeggini, E., Loos, R. J. F., Froguel, P., Ingelsson, E., Lind, L., Groop, L., Laakso, M., Collins, F. S., Jukema, J. W., Palmer, C. N. A., Grallert, H., Metspalu, A., Dehghan, A., Köttgen, A., Abecasis, G. R., Meigs, J. B., Rotter, J. I., Marchini, J., Pedersen, O., Hansen, T., Langenberg, C., Wareham, N. J., Stefansson, K., Gloyn, A. L., Morris, A. P., Boehnke, M., and McCarthy, M. I. (2018). Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nat. Genet.*, 50(11):1505–1513.

[337] Maher, B. (2008). Personal genomes: The case of the missing heritability.

[338] Makino, A., Firth, A. L., and Yuan, J. X.-J. (2011). Endothelial and smooth muscle cell ion channels in pulmonary vasoconstriction and vascular remodeling. *Compr. Physiol.*, 1(3):1555–1602.

[339] Malenfant, S., Neyron, A.-S., Paulin, R., Potus, F., Meloche, J., Provencher, S., and Bonnet, S. (2013). Signal transduction in the development of pulmonary arterial hypertension. *Pulmonary Circulation*, 3(2):278–293.

[340] Mancuso, M. R., Davis, R., Norberg, S. M., O'Brien, S., Sennino, B., Nakahara, T., Yao, V. J., Inai, T., Brooks, P., Freimark, B., Shalinsky, D. R., Hu-Lowe, D. D., and McDonald, D. M. (2006). Rapid vascular regrowth in tumors after reversal of VEGF inhibition. *J. Clin. Invest.*, 116(10):2610–2621.

[341] Mann, S. L., Wadsworth, M. E., and Colley, J. R. (1992). Accumulation of factors influencing respiratory illness in members of a national birth cohort and their offspring. *J. Epidemiol. Community Health*, 46(3):286–292.

[342] Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorff, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A., Cho, J. H., Guttmacher, A. E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C. N., Slatkin, M., Valle, D., Whittemore, A. S., Boehnke, M., Clark, A. G., Eichler, E. E., Gibson, G., Haines, J. L., Mackay, T. F. C., McCarroll, S. A., and Visscher, P. M. (2009). Finding the missing heritability of complex diseases. *Nature*, 461(7265):747–753.

[343] Mansoor, J. K., Schelegle, E. S., Davis, C. E., Walby, W. F., Zhao, W., Aksenov, A. A., Pasamontes, A., Figueroa, J., and Allen, R. (2014). Analysis of volatile compounds in exhaled breath condensate in patients with severe pulmonary arterial hypertension. *PLoS One*, 9(4):e95331.

[344] Marchini, J., Donnelly, P., and Cardon, L. R. (2005). Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat. Genet.*, 37(4):413–417.

[345] Marques, F. Z. (2018). Missing heritability of hypertension and our microbiome. *Circulation*, 138(14):1381–1383.

[346] Martin, C., Held, H. D., and Uhlig, S. (2000). Differential effects of the mixed ET(A)/ET(B)-receptor antagonist bosentan on endothelin-induced bronchoconstriction, vasoconstriction and prostacyclin release. *Naunyn. Schmiedebergs. Arch. Pharmacol.*, 362(2):128–136.

[347] McAlister-Henn, L. (2012). Ligand binding and structural changes associated with allostery in yeast NAD(+)-specific isocitrate dehydrogenase. *Arch. Biochem. Biophys.*, 519(2):112–117.

[348] McDonald, J., Gedge, F., Burdette, A., Carlisle, J., Bukjiok, C. J., Fox, M., and Bayrak-Toydemir, P. (2009). Multiple sequence variants in hereditary hemorrhagic telangiectasia cases: illustration of complexity in molecular diagnostic interpretation. *J. Mol. Diagn.*, 11(6):569–575.

[349] McGoon, M., Gutterman, D., Steen, V., Barst, R., McCrory, D. C., Fortin, T. A., and Loyd, J. E. (2004). Screening, early detection, and diagnosis of pulmonary arterial hypertension. *Chest*, 126(1):14S–34S.

[350] Meissner, I., Whisnant, J. P., Khandheria, B. K., Spittell, P. C., Michael O'Fallon, W., Pascoe, R. D., Enriquez-Sarano, M., Seward, J. B., Covalt, J. L., Sicks, J. D., and Wiebers, D. O. (1999). Prevalence of potential risk factors for stroke assessed by transesophageal echocardiography and carotid ultrasonography: The SPARC study.

[351] Menendez, M. E., Neuhaus, V., van Dijk, C. N., and Ring, D. (2014). The elixhauser comorbidity method outperforms the charlson index in predicting inpatient death after orthopaedic surgery. *Clin. Orthop. Relat. Res.*, 472(9):2878–2886.

[352] Meyer, F. J., Ewert, R., Hoeper, M. M., Olschewski, H., Behr, J., Winkler, J., Wilkens, H., Breuer, C., Kübler, W., Borst, M. M., and German PPH Study Group (2002). Peripheral airway obstruction in primary pulmonary hypertension. *Thorax*, 57(6):473–476.

[353] Michelakis, E. D., Gurtu, V., Webster, L., Barnes, G., Watson, G., Howard, L., Cupitt, J., Paterson, I., Thompson, R. B., Chow, K., O'Regan, D. P., Zhao, L., Wharton, J., Kiely, D. G., Kinnaird, A., Boukouris, A. E., White, C., Nagendran, J., Freed, D. H., Wort, S. J., Gibbs, J. S. R., and Wilkins, M. R. (2017). Inhibition of pyruvate dehydrogenase kinase improves pulmonary arterial hypertension in genetically susceptible patients. *Sci. Transl. Med.*, 9(413).

[354] Miller, N. A., Farrow, E. G., Gibson, M., Willig, L. K., Twist, G., Yoo, B., Marrs, T., Corder, S., Krivohlavek, L., Walter, A., Petrikin, J. E., Saunders, C. J., Thiffault, I., Soden, S. E., Smith, L. D., Dinwiddie, D. L., Herd, S., Cakici, J. A., Catreux, S., Ruehle, M., and Kingsmore, S. F. (2015). A 26-hour system of highly sensitive whole genome sequencing for emergency management of genetic diseases. *Genome Med.*, 7:100.

[355] Mitri, G. M., Lucas, M., Fertig, N., Steen, V. D., and Medsger, T. A. (2003). A comparison between anti-th/to– and anticentromere antibody–positive systemic sclerosis patients with limited cutaneous involvement. *Arthritis & Rheumatism*, 48(1):203–209.

[356] Montani, D., Bergot, E., Günther, S., Savale, L., Bergeron, A., Bourdin, A., Bouvaist, H., Canuet, M., Pison, C., Macro, M., Poubeau, P., Girerd, B., Natali, D., Guignabert, C., Perros, F., O'Callaghan, D. S., Jaïs, X., Tubert-Bitter, P., Zalcman, G., Sitbon, O., Simonneau, G., and Humbert, M. (2012). Pulmonary arterial hypertension in patients treated by dasatinib. *Circulation*, 125(17):2128–2137.

[357] Montani, D., Dorfmüller, P., Girerd, B., Pavec, J. L., Fadel, E., Simonneau, G., Sitbon, O., and Humbert, M. (2018). Natural history over 8 years of pulmonary vascular disease in a patient carrying biallelic EIF2ak4 mutations. *American Journal of Respiratory and Critical Care Medicine*, 198(4):537–541.

[358] Montani, D., Dorfmuller, P., Maitre, S., Jaïs, X., Sitbon, O., Simonneau, G., and Humbert, M. (2010a). [pulmonary veno-occlusive disease and pulmonary capillary hemangiomatosis]. *Presse Med.*, 39(1):134–143.

[359] Montani, D., Girerd, B., Günther, S., Riant, F., Tournier-Lasserve, E., Magy, L., Maazi, N., Guignabert, C., Savale, L., Sitbon, O., Simonneau, G., Soubrier, F., and Humbert, M. (2014). Pulmonary arterial hypertension in familial hemiplegic migraine with ATP1A2 channelopathy. *Eur. Respir. J.*, 43(2):641–643.

[360] Montani, D., Girerd, B., Jaïs, X., Levy, M., Amar, D., Savale, L., Dorfmüller, P., Seferian, A., Lau, E. M., Eyries, M., Le Pavec, J., Parent, F., Bonnet, D., Soubrier, F., Fadel, E., Sitbon, O., Simonneau, G., and Humbert, M. (2017). Clinical phenotypes and outcomes of heritable and sporadic pulmonary veno-occlusive disease: a population-based study. *Lancet Respir Med*, 5(2):125–134.

[361] Montani, D., Lau, E. M., Descatha, A., Jaïs, X., Savale, L., Andujar, P., Bensefa-Colas, L., Girerd, B., Zendah, I., Le Pavec, J., Seferian, A., Perros, F., Dorfmüller, P., Fadel, E., Soubrier, F., Sitbon, O., Simonneau, G., and Humbert, M. (2015). Occupational exposure to organic solvents: a risk factor for pulmonary veno-occlusive disease. *Eur. Respir. J.*, 46(6):1721–1731.

[362] Montani, D., Savale, L., Natali, D., Jaïs, X., Herve, P., Garcia, G., Humbert, M., Simonneau, G., and Sitbon, O. (2010b). Long-term response to calcium-channel blockers in non-idiopathic pulmonary arterial hypertension. *Eur. Heart J.*, 31(15):1898–1907.

[363] Montserrat Moliner, A. and Waligóra, J. (2013). The european union policy in the field of rare diseases. *Public Health Genomics*, 16(6):268–277.

[364] Moore, B. J., White, S., Washington, R., Coenen, N., and Elixhauser, A. (2017). Identifying increased risk of readmission and in-hospital mortality using hospital administrative data: The AHRQ elixhauser comorbidity index. *Med. Care*, 55(7):698–705.

[365] Morrell, N. W., Adnot, S., Archer, S. L., Dupuis, J., Jones, P. L., MacLean, M. R., McMurtry, I. F., Stenmark, K. R., Thistlethwaite, P. A., Weissmann, N., Yuan, J. X.-J., and Weir, E. K. (2009). Cellular and molecular basis of pulmonary arterial hypertension. *J. Am. Coll. Cardiol.*, 54(1 Suppl):S20–31.

[366] Morrell, N. W., Aldred, M. A., Chung, W. K., Elliott, C. G., Nichols, W. C., Soubrier, F., Trembath, R. C., and Loyd, J. E. (2019). Genetics and genomics of pulmonary arterial hypertension. *Eur. Respir. J.*, 53(1).

[367] Morris, A. P. and Zeggini, E. (2010). An evaluation of statistical approaches to rare variant analysis in genetic association studies. *Genet. Epidemiol.*, 34(2):188–193.

[368] Morse, J. H., Jones, A. C., Barst, R. J., Hodge, S. E., Wilhelmsen, K. C., and Nygaard, T. G. (1997). Mapping of familial primary pulmonary hypertension locus (PPH1) to chromosome 2q31-q32. *Circulation*, 95(12):2603–2606.

[369] Mouraret, N., Houssaïni, A., Abid, S., Quarck, R., Marcos, E., Parpaleix, A., Gary-Bobo, G., Dubois-Randé, J.-L., Derumeaux, G., Boczkowski, J., Delcroix, M., Blasco, M. A., Lipskaia, L., Amsellem, V., and Adnot, S. (2015). Role for telomerase in pulmonary hypertension. *Circulation*, 131(8):742–755.

[370] Murtagh, E., Heaney, L., Gingles, J., Shepherd, R., Kee, F., Patterson, C., and MacMahon, J. (2005). The prevalence of obstructive lung disease in a general population sample: The NICECOPD study.

[371] Nasim, M. T., Ogo, T., Ahmed, M., Randall, R., Chowdhury, H. M., Snape, K. M., Bradshaw, T. Y., Southgate, L., Lee, G. J., Jackson, I., Lord, G. M., Gibbs, J. S. R., Wilkins, M. R., Ohta-Ogo, K., Nakamura, K., Girerd, B., Coulet, F., Soubrier, F., Humbert, M., Morrell, N. W., Trembath, R. C., and Machado, R. D. (2011). Molecular genetic characterization of SMAD signaling molecules in pulmonary arterial hypertension. *Hum. Mutat.*, 32(12):1385–1389.

[372] Nathan, S. D., Barbera, J. A., Gaine, S. P., Harari, S., Martinez, F. J., Olschewski, H., Olsson, K. M., Peacock, A. J., Pepke-Zaba, J., Provencher, S., Weissmann, N., and Seeger, W. (2019). Pulmonary hypertension in chronic lung disease and hypoxia. *Eur. Respir. J.*, 53(1).

[373] Navas Tejedor, P., Tenorio Castaño, J., Palomino Doza, J., Arias Lajara, P., Gordo Trujillo, G., López Meseguer, M., Román Broto, A., Lapunzina Abadía, P., and Escribano Subía, P. (2017). An homozygous mutation in KCNK3 is associated with an aggressive form of hereditary pulmonary arterial hypertension. *Clin. Genet.*, 91(3):453–457.

[374] Nebel, A., Kleindorp, R., Caliebe, A., Nothnagel, M., Blanché, H., Junge, O., Wittig, M., Ellinghaus, D., Flachsbart, F., Wichmann, H.-E., Meitinger, T., Nikolaus, S., Franke, A., Krawczak, M., Lathrop, M., and Schreiber, S. (2011). A genome-wide association study confirms APOE as the major gene influencing survival in long-lived individuals. *Mech. Ageing Dev.*, 132(6-7):324–330.

[375] Negi, V. S., Tripathy, N. K., Misra, R., and Nityanand, S. (1998). Antiendothelial cell antibodies in scleroderma correlate with severe digital ischemia and pulmonary arterial hypertension. *J. Rheumatol.*, 25(3):462–466.

[376] Nelson, C. P., Goel, A., Butterworth, A. S., Kanoni, S., Webb, T. R., Marouli, E., Zeng, L., Ntalla, I., Lai, F. Y., Hopewell, J. C., Giannakopoulou, O., Jiang, T., Hamby, S. E., Di Angelantonio, E., Assimes, T. L., Bottinger, E. P., Chambers, J. C., Clarke, R., Palmer, C. N. A., Cubbon, R. M., Ellinor, P., Ermel, R., Evangelou, E., Franks, P. W., Grace, C., Gu, D., Hingorani, A. D., Howson, J. M. M., Ingelsson, E., Kastrati, A., Kessler, T., Kyriakou, T., Lehtimäki, T., Lu, X., Lu, Y., März, W., McPherson, R., Metspalu, A., Pujades-Rodriguez, M., Ruusalepp, A., Schadt, E. E., Schmidt, A. F., Sweeting, M. J., Zalloua, P. A., AlGhalayini, K., Keavney, B. D., Kooner, J. S., Loos, R. J. F., Patel, R. S., Rutter, M. K., Tomaszewski, M., Tzoulaki, I., Zeggini, E., Erdmann, J., Dedoussis, G., Björkegren, J. L. M., EPIC-CVD Consortium, CARDIoGRAMplusC4D, UK Biobank CardioMetabolic Consortium CHD working group, Schunkert, H., Farrall, M., Danesh, J., Samani, N. J., Watkins, H., and Deloukas, P. (2017). Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat. Genet.*, 49(9):1385–1391.

[377] Nelson, M. R., Tipney, H., Painter, J. L., Shen, J., Nicoletti, P., Shen, Y., Floratos, A., Sham, P. C., Li, M. J., Wang, J., Cardon, L. R., Whittaker, J. C., and Sanseau, P. (2015). The support of human genetic evidence for approved drug indications. *Nat. Genet.*, 47(8):856–860.

[378] Neumann, P., Jaé, N., Knau, A., Glaser, S. F., Fouani, Y., Rossbach, O., Krüger, M., John, D., Bindereif, A., Grote, P., Boon, R. A., and Dimmeler, S. (2018). The lncRNA GATA6-AS epigenetically regulates endothelial gene expression via interaction with LOXL2. *Nat. Commun.*, 9(1):237.

[379] Nguengang Wakap, S., Lambert, D. M., Olry, A., Rodwell, C., Gueydan, C., Lanneau, V., Murphy, D., Le Cam, Y., and Rath, A. (2020). Estimating cumulative point prevalence of rare diseases: analysis of the orphanet database. *Eur. J. Hum. Genet.*, 28(2):165–173.

[380] Nichols, W. C., Koller, D. L., Slovis, B., Foroud, T., Terry, V. H., Arnold, N. D., Siemieniak, D. R., Wheeler, L., Phillips, J. A., Newman, J. H., Conneally, P. M., Ginsburg, D., and Loyd, J. E. (1997). Localization of the gene for familial primary pulmonary hypertension to chromosome 2q31–32. *Nature Genetics*, 15(3):277–280.

[381] Nickel, N. P., Spiekerkoetter, E., Gu, M., Li, C. G., Li, H., Kaschwich, M., Diebold, I., Hennigs, J. K., Kim, K.-Y., Miyagawa, K., Wang, L., Cao, A., Sa, S., Jiang, X., Stockstill, R. W., Nicolls, M. R., Zamanian, R. T., Bland, R. D., and Rabinovitch, M. (2015). Elafin reverses pulmonary hypertension via Caveolin-1-Dependent bone morphogenetic protein signaling. *Am. J. Respir. Crit. Care Med.*, 191(11):1273–1286.

[382] Nikolic, I., Yung, L.-M., Yang, P., Malhotra, R., Paskin-Flerlage, S. D., Dinter, T., Bocobo, G. A., Tumelty, K. E., Faugno, A. J., Troncone, L., McNeil, M. E., Huang, X., Coser, K. R., Lai, C. S. C., Upton, P. D., Goumans, M. J., Zamanian, R. T., Elliott, C. G., Lee, A., Zheng, W., Berasi, S. P., Huard, C., Morrell, N. W., Chung, R. T., Channick, R. W., Roberts, K. E., and Yu, P. B. (2019). Bone morphogenetic protein 9 is a mechanistic biomarker of portopulmonary hypertension. *Am. J. Respir. Crit. Care Med.*, 199(7):891–902.

[383] Nimmakayalu, M., Major, H., Sheffield, V., Solomon, D. H., Smith, R. J., Patil, S. R., and Shchelochkov, O. A. (2011). Microdeletion of 17q22q23.2 encompassing TBX2 and TBX4 in a patient with congenital microcephaly, thyroid duct cyst, sensorineural hearing loss, and pulmonary hypertension. *American Journal of Medical Genetics Part A*, 155(2):418–423.

[384] Nogueira-Ferreira, R., Faria-Costa, G., Ferreira, R., and Henriques-Coelho, T. (2016). Animal models for the study of pulmonary hypertension: potential and limitations. *Cardiol. Cardiovasc. Med*, 1:1–22.

[385] Nootens, M. T., Berarducci, L. A., Kaufmann, E., Devries, S., and Rich, S. (1993). The prevalence and significance of a patent foramen ovale in pulmonary hypertension. *Chest*, 104(6):1673–1675.

[386] Norel, X., Walch, L., Labat, C., Gascard, J. P., Dulmet, E., and Brink, C. (1999). Prostanoid receptors involved in the relaxation of human bronchial preparations. *Br. J. Pharmacol.*, 126(4):867–872.

[387] Oladipupo, S. S., Smith, C., Santeford, A., Park, C., Sene, A., Wiley, L. A., Osei-Owusu, P., Hsu, J., Zapata, N., Liu, F., Nakamura, R., Lavine, K. J., Blumer, K. J., Choi, K., Apte, R. S., and Ornitz, D. M. (2014). Endothelial cell FGF signaling is required for injury response but not for vascular homeostasis. *Proceedings of the National Academy of Sciences*, 111(37):13379–13384.

[388] Ozdemir, A. O., Tamayo, A., Munoz, C., Dias, B., and Spence, J. D. (2008). Cryptogenic stroke and patent foramen ovale: clinical clues to paradoxical embolism. *J. Neurol. Sci.*, 275(1-2):121–127.

[389] Pan, B., Kusko, R., Xiao, W., Zheng, Y., Liu, Z., Xiao, C., Sakkiah, S., Guo, W., Gong, P., Zhang, C., Ge, W., Shi, L., Tong, W., and Hong, H. (2019). Similarities and differences between variants called with human reference genome HG19 or HG38. *BMC Bioinformatics*, 20(S2).

[390] Pang, C., Sollie, A., Sijtsma, A., Hendriksen, D., Charbon, B., de Haan, M., de Boer, T., Kelpin, F., Jetten, J., van der Velde, J. K., Smidt, N., Sijmons, R., Hillege, H., and Swertz, M. A. (2015). SORTA: a system for ontology-based re-coding and technical annotation of biomedical phenotype data. *Database*, 2015:bav089.

[391] Paten, B., Novak, A. M., Eizenga, J. M., and Garrison, E. (2017). Genome graphs and the evolution of genome inference. *Genome Res.*, 27(5):665–676.

[392] Paulin, R., Dromparis, P., Sutendra, G., Gurtu, V., Zervopoulos, S., Bowers, L., Haromy, A., Webster, L., Provencher, S., Bonnet, S., and Michelakis, E. D. (2014). Sirtuin 3 deficiency is associated with inhibited mitochondrial function and pulmonary arterial hypertension in rodents and humans. *Cell Metab.*, 20(5):827–839.

[393] Peacock, A. J., Ling, Y., Johnson, M. K., Kiely, D. G., Condliffe, R., Elliot, C. A., Gibbs, J. S. R., Howard, L. S., Pepke-Zaba, J., Sheares, K. K., Corris, P. A., Fisher, A. J., Lordan, J. L., Gaine, S., Coghlan, J. G., Wort, S. J., and Gatzoulis, M. A. (2020). Idiopathic pulmonary arterial hypertension and co-existing lung disease: is this a new phenotype? *Pulmonary Circulation*, 10(1):204589402091485.

[394] Peacock, A. J., Murphy, N. F., McMurray, J. J. V., Caballero, L., and Stewart, S. (2007). An epidemiological study of pulmonary arterial hypertension. *European Respiratory Journal*, 30(1):104–109.

[395] Peacock, A. J. and Vonk Noordegraaf, A. (2013). Cardiac magnetic resonance imaging in pulmonary arterial hypertension. *Eur. Respir. Rev.*, 22(130):526–534.

[396] Pereira, N. L., Sargent, D. J., Farkouh, M. E., and Rihal, C. S. (2015). Genotype-based clinical trials in cardiovascular disease. *Nat. Rev. Cardiol.*, 12(8):475–487.

[397] Perenthaler, E., Yousefi, S., Niggl, E., and Barakat, T. S. (2019). Beyond the exome: The non-coding genome and enhancers in neurodevelopmental disorders and malformations of cortical development. *Front. Cell. Neurosci.*, 13:352.

[398] Peron, A., Au, K. S., and Northrup, H. (2018). Genetics, genomics, and genotype-phenotype correlations of TSC: Insights for clinical practice. *Am. J. Med. Genet. C Semin. Med. Genet.*, 178(3):281–290.

[399] Perosa, F., Favoino, E., Favia, I. E., Vettori, S., Prete, M., Corrado, A., Cantatore, F. P., and Valentini, G. (2016). Subspecificities of anticentromeric protein a antibodies identify systemic sclerosis patients at higher risk of pulmonary vascular disease. *Medicine*, 95(25):e3931.

[400] Perros, F., Dorfmüller, P., Montani, D., Hammad, H., Waelput, W., Girerd, B., Raymond, N., Mercier, O., Mussot, S., Cohen-Kaminsky, S., Humbert, M., and Lambrecht, B. N. (2012). Pulmonary lymphoid neogenesis in idiopathic pulmonary arterial hypertension. *Am. J. Respir. Crit. Care Med.*, 185(3):311–321.

[401] Perrotta, A., Gambardella, S., Ambrosini, A., Anastasio, M. G., Albano, V., Fornai, F., and Pierelli, F. (2018). A novel ATP1A2 gene variant associated with pure sporadic hemiplegic migraine improved after patent foramen ovale closure: A case report. *Front. Neurol.*, 9:332.

[402] Pesquita, C., Faria, D., Falcão, A. O., Lord, P., and Couto, F. M. (2009). Semantic similarity in biomedical ontologies. *PLoS Comput. Biol.*, 5(7):e1000443.

[403] Peters, J. L., Cnudde, F., and Gerats, T. (2003). Forward genetics and map-based cloning approaches. *Trends Plant Sci.*, 8(10):484–491.

[404] Petrovski, S., Todd, J. L., Durheim, M. T., Wang, Q., Chien, J. W., Kelly, F. L., Frankel, C., Mebane, C. M., Ren, Z., Bridgers, J., Urban, T. J., Malone, C. D., Copeland, A. F., Brinkley, C., Allen, A. S., O'Riordan, T., McHutchison, J. G., Palmer, S. M., and Goldstein, D. B. (2017). An exome sequencing study to assess the role of rare genetic variation in pulmonary fibrosis. *American Journal of Respiratory and Critical Care Medicine*, 196(1):82–93.

[405] Pfarr, N., Fischer, C., Ehlken, N., Becker-Grünig, T., López-González, V., Gorenflo, M., Hager, A., Hinderhofer, K., Miera, O., Nagel, C., Schranz, D., and Grünig, E. (2013). Hemodynamic and genetic analysis in children with idiopathic, heritable, and congenital heart disease associated pulmonary arterial hypertension. *Respir. Res.*, 14:3.

[406] Philippakis, A. A., Azzariti, D. R., Beltran, S., Brookes, A. J., Brownstein, C. A., Brudno, M., Brunner, H. G., Buske, O. J., Carey, K., Doll, C., Dumitriu, S., Dyke, S. O. M., den Dunnen, J. T., Firth, H. V., Gibbs, R. A., Girdea, M., Gonzalez, M., Haendel, M. A., Hamosh, A., Holm, I. A., Huang, L., Hurles, M. E., Hutton, B., Krier, J. B., Misyura, A., Mungall, C. J., Paschall, J., Paten, B., Robinson, P. N., Schiettecatte, F., Sobreira, N. L., Swaminathan, G. J., Taschner, P. E., Terry, S. F., Washington, N. L., Züchner, S., Boycott, K. M., and Rehm, H. L. (2015). The matchmaker exchange: a platform for rare disease gene discovery. *Hum. Mutat.*, 36(10):915–921.

[407] Phillips, 3rd, J. A., Poling, J. S., Phillips, C. A., Stanton, K. C., Austin, E. D., Cogan, J. D., Wheeler, L., Yu, C., Newman, J. H., Dietz, H. C., and Loyd, J. E. (2008). Synergistic heterozygosity for TGFbeta1 SNPs and BMPR2 mutations modulates the age at diagnosis and penetrance of familial pulmonary arterial hypertension. *Genet. Med.*, 10(5):359–365.

[408] Pierpont, M. E., Brueckner, M., Chung, W. K., Garg, V., Lacro, R. V., McGuire, A. L., Mital, S., Priest, J. R., Pu, W. T., Roberts, A., Ware, S. M., Gelb, B. D., Russell, M. W., and American Heart Association Council on Cardiovascular Disease in the Young; Council on Cardiovascular and Stroke Nursing; and Council on Genomic and Precision Medicine

(2018). Genetic basis for congenital heart disease: Revisited: A scientific statement from the american heart association. *Circulation*, 138(21):e653–e711.

[409] Playford, D., Bordin, E., Talbot, L., Mohamad, R., Anderson, B., and Strange, G. (2018). Analysis of aortic stenosis using artificial intelligence. *Heart, Lung and Circulation*, 27:S216.

[410] Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R., and Siepel, A. (2010). Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.*, 20(1):110–121.

[411] Polubriaginof, F. C. G., Vanguri, R., Quinnies, K., Belbin, G. M., Yahi, A., Salmasian, H., Lorberbaum, T., Nwankwo, V., Li, L., Shervey, M. M., Glowe, P., Ionita-Laza, I., Simmerling, M., Hripcsak, G., Bakken, S., Goldstein, D., Kiryluk, K., Kenny, E. E., Dudley, J., Vawdrey, D. K., and Tatonetti, N. P. (2018). Disease heritability inferred from familial relationships reported in medical records. *Cell*, 173(7):1692–1704.e11.

[412] Poms, A. D., Turner, M., Farber, H. W., Meltzer, L. A., and McGoon, M. D. (2013). Comorbid conditions and outcomes in patients with pulmonary arterial hypertension: a REVEAL registry analysis. *Chest*, 144(1):169–176.

[413] Potocki, L., Bi, W., Treadwell-Deering, D., Carvalho, C. M. B., Eifert, A., Friedman, E. M., Glaze, D., Krull, K., Lee, J. A., Lewis, R. A., Mendoza-Londono, R., Robbins-Furman, P., Shaw, C., Shi, X., Weissenberger, G., Withers, M., Yatsenko, S. A., Zackai, E. H., Stankiewicz, P., and Lupski, J. R. (2007). Characterization of Potocki-Lupski syndrome (dup(17)(p11.2p11.2)) and delineation of a dosage-sensitive critical interval that can convey an autism phenotype. *Am. J. Hum. Genet.*, 80(4):633–649.

[414] Potocki, L., Chen, K.-S., Park, S.-S., Osterholm, D. E., Withers, M. A., Kimonis, V., Summers, A. M., Meschino, W. S., Anyane-Yeboa, K., Kashork, C. D., Shaffer, L. G., and Lupski, J. R. (2000). Molecular mechanism for duplication 17p11.2— the homologous recombination reciprocal of the smith-magenis microdeletion. *Nature Genetics*, 24(1):84–87.

[415] Potus, F., Pauciulo, M. W., Cook, E. K., Zhu, N., Hsieh, A., Welch, C. L., Shen, Y., Tian, L., Lima, P., Mewburn, J., D'Arsigny, C. L., Lutz, K. A., Coleman, A. W., Damico, R., Snetsinger, B., Martin, A. Y., Hassoun, P. M., Nichols, W. C., Chung, W. K., Rauh, M. J., and Archer, S. L. (2020). Novel mutations and decreased expression of the epigenetic regulator in pulmonary arterial hypertension. *Circulation*, 141(24):1986–2000.

[416] Pousada, G., Baloira, A., and Valverde, D. (2016). Pulmonary arterial hypertension associated with several mutations. In *4.3 Pulmonary Circulation and Pulmonary Vascular Diseases*, page PA2478. European Respiratory Society.

[417] Povysil, G., Petrovski, S., Hostyk, J., Aggarwal, V., Allen, A. S., and Goldstein, D. B. (2019). Rare-variant collapsing analyses for complex traits: guidelines and applications. *Nat. Rev. Genet.*, 20(12):747–759.

[418] Provencher, S., Archer, S. L., Ramirez, F. D., Hibbert, B., Paulin, R., Boucherat, O., Lacasse, Y., and Bonnet, S. (2018). Standards and methodological rigor in pulmonary arterial hypertension preclinical and translational research. *Circ. Res.*, 122(7):1021–1032.

[419] Przewlocka-Kosmala, M., Marwick, T. H., Dabrowski, A., and Kosmala, W. (2019). Contribution of cardiovascular reserve to prognostic categories of heart failure with preserved ejection fraction: A classification based on machine learning. *Journal of the American Society of Echocardiography*, 32(5):604–615.e6.

[420] Pullabhatla, V., Roberts, A. L., Lewis, M. J., Mauro, D., Morris, D. L., Odhams, C. A., Tombleson, P., Liljedahl, U., Vyse, S., Simpson, M. A., Sauer, S., de Rinaldis, E., Syvänen, A.-C., and Vyse, T. J. (2018). De novo mutations implicate novel genes in systemic lupus erythematosus. *Hum. Mol. Genet.*, 27(3):421–429.

[421] Pullamsetti, S. S., Doebele, C., Fischer, A., Savai, R., Kojonazarov, B., Dahal, B. K., Ghofrani, H. A., Weissmann, N., Grimminger, F., Bonauer, A., Seeger, W., Zeiher, A. M., Dimmeler, S., and Schermuly, R. T. (2012). Inhibition of microRNA-17 improves lung and heart function in experimental pulmonary hypertension. *Am. J. Respir. Crit. Care Med.*, 185(4):409–419.

[422] Quan, H., Sundararajan, V., Halfon, P., Fong, A., Burnand, B., Luthi, J.-C., Saunders, L. D., Beck, C. A., Feasby, T. E., and Ghali, W. A. (2005). Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. *Med. Care*, 43(11):1130–1139.

[423] Quanjer, P. H., Tammeling, G. J., Cotes, J. E., Pedersen, O. F., Peslin, R., and Yernault, J.-C. (1993). Lung volumes and forced ventilatory flows. *European Respiratory Journal*, 6(Suppl 16):5–40.

[424] R Core Team (2017). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[425] Rabbani, B., Tekin, M., and Mahdieh, N. (2013). The promise of whole-exome sequencing in medical genetics. *Journal of Human Genetics*, 59(1):5–15.

[426] Rakocevic, G., Semenyuk, V., Lee, W.-P., Spencer, J., Browning, J., Johnson, I. J., Arsenijevic, V., Nadj, J., Ghose, K., Suciu, M. C., Ji, S.-G., Demir, G., Li, L., Toptaş, B. Ç., Dolgoborodov, A., Pollex, B., Spulber, I., Glotova, I., Kómár, P., Stachyra, A. L., Li, Y., Popovic, M., Källberg, M., Jain, A., and Kural, D. (2019). Fast and accurate genomic analyses using genome graphs. *Nat. Genet.*, 51(2):354–362.

[427] Ran, F. A., Hsu, P. D., Wright, J., Agarwala, V., Scott, D. A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.*, 8(11):2281–2308.

[428] Ranchoux, B., Antigny, F., Rucker-Martin, C., Hautefort, A., Péchoux, C., Bogaard, H. J., Dorfmüller, P., Remy, S., Lecerf, F., Planté, S., Chat, S., Fadel, E., Houssaini, A., Anegon, I., Adnot, S., Simonneau, G., Humbert, M., Cohen-Kaminsky, S., and Perros, F. (2015). Endothelial-to-mesenchymal transition in pulmonary hypertension. *Circulation*, 131(11):1006–1018.

[429] Ranchoux, B., Bigorgne, A., Hautefort, A., Girerd, B., Sitbon, O., Montani, D., Humbert, M., Tcherakian, C., and Perros, F. (2017). Gut-Lung connection in pulmonary arterial hypertension. *Am. J. Respir. Cell Mol. Biol.*, 56(3):402–405.

[430] Ravenscroft, G., Nolent, F., Rajagopalan, S., Meireles, A. M., Paavola, K. J., Gaillard, D., Alanio, E., Buckland, M., Arbuckle, S., Krivanek, M., Maluenda, J., Pannell, S., Gooding, R., Ong, R. W., Allcock, R. J., Carvalho, E. D. F., Carvalho, M. D. F., Kok, F., Talbot, W. S., Melki, J., and Laing, N. G. (2015). Mutations of GPR126 are responsible for severe arthrogryposis multiplex congenita. *Am. J. Hum. Genet.*, 96(6):955–961.

[431] Rehm, H. L., Berg, J. S., Brooks, L. D., Bustamante, C. D., Evans, J. P., Landrum, M. J., Ledbetter, D. H., Maglott, D. R., Martin, C. L., Nussbaum, R. L., Plon, S. E., Ramos, E. M., Sherry, S. T., and Watson, M. S. (2015). ClinGen — the clinical genome resource.

[432] Resnik, P. (1999). Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *Journal of Artificial Intelligence Research*, 11:95–130.

[433] Retterer, K., Juusola, J., Cho, M. T., Vitazka, P., Millan, F., Gibellini, F., Vertino-Bell, A., Smaoui, N., Neidich, J., Monaghan, K. G., McKnight, D., Bai, R., Suchy, S., Friedman, B., Tahiliani, J., Pineda-Alvarez, D., Richard, G., Brandt, T., Haverfield, E., Chung, W. K., and Bale, S. (2016). Clinical application of whole-exome sequencing across clinical indications. *Genet. Med.*, 18(7):696–704.

[434] Reuter, M. S., Jobling, R., Chaturvedi, R. R., Manshaei, R., Costain, G., Heung, T., Curtis, M., Hosseini, S. M., Liston, E., Lowther, C., Oechslin, E., Sticht, H., Thiruvahindrapuram, B., van Mil, S., Wald, R. M., Walker, S., Marshall, C. R., Silversides, C. K., Scherer, S. W., Kim, R. H., and Bassett, A. S. (2019). Haploinsufficiency of vascular endothelial growth factor related signaling genes is associated with tetralogy of fallot. *Genet. Med.*, 21(4):1001–1007.

[435] Reyes-Palomares, A., Gu, M., Grubert, F., Berest, I., Sa, S., Kasowski, M., Arnold, C., Shuai, M., Srivas, R., Miao, S., Li, D., Snyder, M. P., Rabinovitch, M., and Zaugg, J. B. (2020). Remodeling of active endothelial enhancers is associated with aberrant gene-regulatory networks in pulmonary arterial hypertension. *Nat. Commun.*, 11(1):1673.

[436] Rhodes, C. J., Batai, K., Bleda, M., Haimel, M., Southgate, L., Germain, M., Pauciulo, M. W., Hadinnapola, C., Aman, J., Girerd, B., Arora, A., Knight, J., Hanscombe, K. B., Karnes, J. H., Kaakinen, M., Gall, H., Ulrich, A., Harbaum, L., Cebola, I., Ferrer, J., Lutz, K., Swietlik, E. M., Ahmad, F., Amouyel, P., Archer, S. L., Argula, R., Austin, E. D., Badesch, D., Bakshi, S., Barnett, C., Benza, R., Bhatt, N., Bogaard, H. J., Burger, C. D., Chakinala, M., Church, C., Coghlan, J. G., Condliffe, R., Corris, P. A., Danesino, C., Debette, S., Elliott, C. G., Elwing, J., Eyries, M., Fortin, T., Franke, A., Frantz, R. P., Frost, A., Garcia, J. G. N., Ghio, S., Ghofrani, H.-A., Gibbs, J. S. R., Harley, J., He, H., Hill, N. S., Hirsch, R., Houweling, A. C., Howard, L. S., Ivy, D., Kiely, D. G., Klinger, J., Kovacs, G., Lahm, T., Laudes, M., Machado, R. D., MacKenzie Ross, R. V., Marsolo, K., Martin, L. J., Moledina, S., Montani, D., Nathan, S. D., Newnham, M., Olschewski, A., Olschewski, H., Oudiz, R. J., Ouwehand, W. H., Peacock, A. J., Pepke-Zaba, J., Rehman, Z., Robbins, I., Roden, D. M., Rosenzweig, E. B., Saydain, G., Scelsi, L., Schilz, R., Seeger, W., Shaffer, C. M., Simms, R. W., Simon, M., Sitbon, O., Suntharalingam, J., Tang, H., Tchourbanov, A. Y., Thenappan, T., Torres, F., Toshner, M. R., Treacy, C. M., Vonk Noordegraaf, A., Waisfisz, Q., Walsworth, A. K., Walter, R. E., Wharton, J., White, R. J., Wilt, J., Wort, S. J., Yung, D., Lawrie, A., Humbert, M., Soubrier, F.,

Trégouët, D.-A., Prokopenko, I., Kittles, R., Gräf, S., Nichols, W. C., Trembath, R. C., Desai, A. A., Morrell, N. W., Wilkins, M. R., UK NIHR BioResource Rare Diseases Consortium, UK PAH Cohort Study Consortium, and US PAH Biobank Consortium (2019). Genetic determinants of risk in pulmonary arterial hypertension: international genome-wide association studies and meta-analysis. *Lancet Respir Med*, 7(3):227–238.

[437] Rhodes, C. J., Ghataorhe, P., Wharton, J., Rue-Albrecht, K. C., Hadinnapola, C., Watson, G., Bleda, M., Haimel, M., Coghlan, G., Corris, P. A., Howard, L. S., Kiely, D. G., Peacock, A. J., Pepke-Zaba, J., Toshner, M. R., Wort, S. J., Gibbs, J. S. R., Lawrie, A., Gräf, S., Morrell, N. W., and Wilkins, M. R. (2017). Plasma metabolomics implicates modified transfer RNAs and altered bioenergetics in the outcomes of pulmonary arterial hypertension. *Circulation*, 135(5):460–475.

[438] Rich, S., Dantzker, D. R., Ayres, S. M., Bergofsky, E. H., Brundage, B. H., Detre, K. M., Fishman, A. P., Goldring, R. M., Groves, B. M., and Koerner, S. K. (1987). Primary pulmonary hypertension. a national prospective study. *Ann. Intern. Med.*, 107(2):216–223.

[439] Rich, S., Kaufmann, E., and Levy, P. S. (1992). The effect of high doses of calcium-channel blockers on survival in primary pulmonary hypertension. *N. Engl. J. Med.*, 327(2):76–81.

[440] Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W. W., Hegde, M., Lyon, E., Spector, E., Voelkerding, K., Rehm, H. L., and ACMG Laboratory Quality Assurance Committee (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the american college of medical genetics and genomics and the association for molecular pathology. *Genet. Med.*, 17(5):405–424.

[441] Risch, N. and Merikangas, K. (1996). The future of genetic studies of complex human diseases. *Science*, 273(5281):1516–1517.

[442] Roberts, C. M., MacRae, K. D., and Seed, W. A. (1990). Multi-breath and single breath helium dilution lung volumes as a test of airway obstruction. *Eur. Respir. J.*, 3(5):515–520.

[443] Romano, A. A., Allanson, J. E., Dahlgren, J., Gelb, B. D., Hall, B., Pierpont, M. E., Roberts, A. E., Robinson, W., Takemoto, C. M., and Noonan, J. A. (2010). Noonan syndrome: Clinical features, diagnosis, and management guidelines. *PEDIATRICS*, 126(4):746–759.

[444] Romano, A. M., Tomaselli, S., Gualtieri, G., Zoia, M. C., Fanfulla, F., Berrayah, L., and Cerveri, I. (1993). Respiratory function in precapillary pulmonary hypertension. *Monaldi Arch. Chest Dis.*, 48(3):201–204.

[445] Roobol, A., Roobol, J., Bastide, A., Knight, J. R. P., Willis, A. E., and Smales, C. M. (2015). p58IPK is an inhibitor of the eIF2α kinase GCN2 and its localization and expression underpin protein synthesis and ER processing capacity. *Biochem. J*, 465(2):213–225.

[446] Rubin, D. B. (2009). *Multiple Imputation for Nonresponse in Surveys*. John Wiley & Sons.

[447] Runo, J. R., Vnencak-Jones, C. L., Prince, M., Loyd, J. E., Wheeler, L., Robbins, I. M., Lane, K. B., Newman, J. H., Johnson, J., Nichols, W. C., and Phillips, 3rd, J. A. (2003). Pulmonary veno-occlusive disease caused by an inherited mutation in bone morphogenetic protein receptor II. *Am. J. Respir. Crit. Care Med.*, 167(6):889–894.

[448] Ryan, N. A., Zwetsloot, K. A., Westerkamp, L. M., Hickner, R. C., Pofahl, W. E., and Gavin, T. P. (2006). Lower skeletal muscle capillarization and VEGF expression in aged vs. young men. *Journal of Applied Physiology*, 100(1):178–185.

[449] Saadoun, S., Papadopoulos, M. C., Hara-Chikuma, M., and Verkman, A. S. (2005). Impairment of angiogenesis and cell migration by targeted aquaporin-1 gene disruption. *Nature*, 434(7034):786–792.

[450] Sakao, S., Miyauchi, H., Voelkel, N. F., Sugiura, T., Tanabe, N., Kobayashi, Y., and Tatsumi, K. (2015). Increased right ventricular fatty acid accumulation in chronic thromboembolic pulmonary hypertension. *Ann. Am. Thorac. Soc.*, 12(10):1465–1472.

[451] Sakornsakolpat, P., McCormack, M., Bakke, P., Gulsvik, A., Make, B. J., Crapo, J. D., Cho, M. H., and Silverman, E. K. (2019). Genome-wide association analysis of single-breath DlCO. *American Journal of Respiratory Cell and Molecular Biology*, 60(5):523–531.

[452] Sandoval, J., Gaspar, J., Pena, H., Santos, L. E., Cordova, J., del Valle, K., Rodriguez, A., and Pulido, T. (2011). Effect of atrial septostomy on the survival of patients with severe pulmonary arterial hypertension. *European Respiratory Journal*, 38(6):1343–1348.

[453] Sandoval, J., Gaspar, J., Pulido, T., Bautista, E., Martínez-Guerra, M. L., Zeballos, M., Palomar, A., and Gómez, A. (1998). Graded balloon dilation atrial septostomy in severe primary pulmonary hypertension. a therapeutic alternative for patients nonresponsive to vasodilator treatment. *J. Am. Coll. Cardiol.*, 32(2):297–304.

[454] Sasaki, T., Göhring, W., Pan, T. C., Chu, M. L., and Timpl, R. (1995). Binding of mouse and human fibulin-2 to extracellular matrix ligands. *J. Mol. Biol.*, 254(5):892–899.

[455] Satoh, T., Kimura, K., Okano, Y., Hirakata, M., Kawakami, Y., and Kuwana, M. (2005). Lack of circulating autoantibodies to bone morphogenetic protein receptor-II or activin receptor-like kinase 1 in mixed connective tissue disease patients with pulmonary arterial hypertension. *Rheumatology*, 44(2):192–196.

[456] Scheubel, R. J., Zorn, H., Silber, R.-E., Kuss, O., Morawietz, H., Holtz, J., and Simm, A. (2003). Age-dependent depression in circulating endothelial progenitor cells in patients undergoing coronary artery bypass grafting. *J. Am. Coll. Cardiol.*, 42(12):2073–2080.

[457] Schneider, V. A., Graves-Lindsay, T., Howe, K., Bouk, N., Chen, H.-C., Kitts, P. A., Murphy, T. D., Pruitt, K. D., Thibaud-Nissen, F., Albracht, D., Fulton, R. S., Kremitzki, M., Magrini, V., Markovic, C., McGrath, S., Steinberg, K. M., Auger, K., Chow, W., Collins, J., Harden, G., Hubbard, T., Pelan, S., Simpson, J. T., Threadgold, G., Torrance, J., Wood, J. M., Clarke, L., Koren, S., Boitano, M., Peluso, P., Li, H., Chin, C.-S., Phillippy, A. M., Durbin, R., Wilson, R. K., Flicek, P., Eichler, E. E., and Church, D. M. (2017). Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. *Genome Res.*, 27(5):849–864.

[458] Schulze, T. G. and McMahon, F. J. (2004). Defining the phenotype in human genetic studies: forward genetics and reverse phenotyping. *Hum. Hered.*, 58(3-4):131–138.

[459] Schuoler, C., Haider, T. J., Leuenberger, C., Vogel, J., Ostergaard, L., Kwapiszewska, G., Kohler, M., Gassmann, M., Huber, L. C., and Brock, M. (2017). Aquaporin 1 controls the functional phenotype of pulmonary smooth muscle cells in hypoxia-induced pulmonary hypertension. *Basic Res. Cardiol.*, 112(3):30.

[460] Schwarze, K., Buchanan, J., Fermont, J. M., Dreau, H., Tilley, M. W., Taylor, J. M., Antoniou, P., Knight, S. J. L., Camps, C., Pentony, M. M., Kvikstad, E. M., Harris, S., Popitsch, N., Pagnamenta, A. T., Schuh, A., Taylor, J. C., and Wordsworth, S. (2020). The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the united kingdom. *Genet. Med.*, 22(1):85–94.

[461] Seah, J. C. Y., Tang, J. S. N., Kitchen, A., Gaillard, F., and Dixon, A. F. (2019). Chest radiographs in congestive heart failure: Visualizing neural network learning. *Radiology*, 290(2):514–522.

[462] Seeger, W., Adir, Y., Barberà, J. A., Champion, H., Coghlan, J. G., Cottin, V., De Marco, T., Galiè, N., Ghio, S., Gibbs, S., Martinez, F. J., Semigran, M. J., Simonneau, G., Wells, A. U., and Vachiéry, J.-L. (2013). Pulmonary hypertension in chronic lung diseases. *J. Am. Coll. Cardiol.*, 62(25 Suppl):D109–16.

[463] Sgarra, L., Bortone, A. S., Potenza, M. A., Nacci, C., De Salvia, M. A., Acquaviva, T., De Cillis, E., Ciccone, M. M., Grimaldi, M., and Montagnani, M. (2020). Endothelial dysfunction may link interatrial septal abnormalities and MTHFR-Inherited defects to cryptogenic stroke predisposition. *Biomolecules*, 10(6).

[464] Sharan, L., Stackhouse, K., Awerbach, J. D., Bashore, T. M., and Krasuski, R. A. (2018). Effect of patent foramen ovale in patients with pulmonary hypertension. *Am. J. Cardiol.*, 122(3):505–510.

[465] Shatat, M. A., Tian, H., Zhang, R., Tandon, G., Hale, A., Fritz, J. S., Zhou, G., Martínez-González, J., Rodríguez, C., Champion, H. C., Jain, M. K., and Hamik, A. (2014). Endothelial krüppel-like factor 4 modulates pulmonary arterial hypertension. *American Journal of Respiratory Cell and Molecular Biology*, 50(3):647–653.

[466] Sheeba, C. J. and Logan, M. P. O. (2017). The roles of T-Box genes in vertebrate limb development. *Curr. Top. Dev. Biol.*, 122:355–381.

[467] Shintani, M., Yagi, H., Nakayama, T., Saji, T., and Matsuoka, R. (2009). A new nonsense mutation of SMAD8 associated with pulmonary arterial hypertension. *J. Med. Genet.*, 46(5):331–337.

[468] Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L. W., Richards, S., Weinstock, G. M., Wilson, R. K., Gibbs, R. A., Kent, W. J., Miller, W., and Haussler, D. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.*, 15(8):1034–1050.

[469] Sifrim, A., Popovic, D., Tranchevent, L.-C., Ardeshirdavani, A., Sakai, R., Konings, P., Vermeesch, J. R., Aerts, J., De Moor, B., and Moreau, Y. (2013). extasy: variant prioritization by genomic data fusion. *Nat. Methods*, 10(11):1083–1084.

[470] Sim, N.-L., Kumar, P., Hu, J., Henikoff, S., Schneider, G., and Ng, P. C. (2012). SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res.*, 40(Web Server issue):W452–7.

[471] Simonneau, G., Montani, D., Celermajer, D. S., Denton, C. P., Gatzoulis, M. A., Krowka, M., Williams, P. G., and Souza, R. (2019). Haemodynamic definitions and updated clinical classification of pulmonary hypertension. *Eur. Respir. J.*, 53(1).

[472] Simonneau, G., Robbins, I. M., Beghetti, M., Channick, R. N., Delcroix, M., Denton, C. P., Elliott, C. G., Gaine, S. P., Gladwin, M. T., Jing, Z.-C., Krowka, M. J., Langleben, D., Nakanishi, N., and Souza, R. (2009). Updated clinical classification of pulmonary hypertension. *J. Am. Coll. Cardiol.*, 54(1 Suppl):S43–54.

[473] Sitbon, O., Humbert, M., Jaïs, X., Ioos, V., Hamid, A. M., Provencher, S., Garcia, G., Parent, F., Hervé, P., and Simonneau, G. (2005). Long-term response to calcium channel blockers in idiopathic pulmonary arterial hypertension. *Circulation*, 111(23):3105–3111.

[474] Snijder, R. J. R., Luermans, J. G. L. M., de Heij, A. H., Thijs, V., Schonewille, W. J., Van De Bruaene, A., Swaans, M. J., Budts, W. I. H. L., and Post, M. C. (2016). Patent foramen ovale with atrial septal aneurysm is strongly associated with migraine with aura: A large observational study. *J. Am. Heart Assoc.*, 5(12).

[475] Sobanski, V., Giovannelli, J., Lynch, B. M., Schreiber, B. E., Nihtyanova, S. I., Harvey, J., Handler, C. E., Denton, C. P., and Coghlan, J. G. (2016). Characteristics and survival of Anti-U1 RNP Antibody-Positive patients with connective tissue Disease-Associated pulmonary arterial hypertension. *Arthritis Rheumatol*, 68(2):484–493.

[476] Soden, S. E., Saunders, C. J., Willig, L. K., Farrow, E. G., Smith, L. D., Petrikin, J. E., LePichon, J.-B., Miller, N. A., Thiffault, I., Dinwiddie, D. L., Twist, G., Noll, A., Heese, B. A., Zellmer, L., Atherton, A. M., Abdelmoity, A. T., Safina, N., Nyp, S. S., Zuccarelli, B., Larson, I. A., Modrcin, A., Herd, S., Creed, M., Ye, Z., Yuan, X., Brodsky, R. A., and Kingsmore, S. F. (2014). Effectiveness of exome and genome sequencing guided by acuity of illness for diagnosis of neurodevelopmental disorders. *Sci. Transl. Med.*, 6(265):265ra168.

[477] Soemedi, R., Topf, A., Wilson, I. J., Darlay, R., Rahman, T., Glen, E., Hall, D., Huang, N., Bentham, J., Bhattacharya, S., Cosgrove, C., Brook, J. D., Granados-Riveron, J., Setchfield, K., Bu'lock, F., Thornborough, C., Devriendt, K., Breckpot, J., Hofbeck, M., Lathrop, M., Rauch, A., Blue, G. M., Winlaw, D. S., Hurles, M., Santibanez-Koref, M., Cordell, H. J., Goodship, J. A., and Keavney, B. D. (2012). Phenotype-specific effect of chromosome 1q21.1 rearrangements and GJA5 duplications in 2436 congenital heart disease patients and 6760 controls. *Hum. Mol. Genet.*, 21(7):1513–1520.

[478] Sofianopoulou, E., Kaptoge, S., Gräf, S., Hadinnapola, C., Treacy, C. M., Church, C., Coghlan, G., Gibbs, J. S. R., Haimel, M., Howard, L. S., Johnson, M., Kiely, D. G., Lawrie, A., Lordan, J., MacKenzie Ross, R. V., Martin, J. M., Moledina, S., Newnham, M., Peacock, A. J., Price, L. C., Rhodes, C. J., Suntharalingam, J., Swietlik, E. M., Toshner, M. R., Wharton, J., Wilkins, M. R., Wort, S. J., Pepke-Zaba, J., Condliffe, R., Corris, P. A., Di Angelantonio, E., Provencher, S., and Morrell, N. W. (2019). Traffic exposures, air pollution and outcomes in pulmonary arterial hypertension: a UK cohort study analysis. *Eur. Respir. J.*, 53(5).

[479] Solari, V. and Puri, P. (2004). Genetic polymorphisms of angiotensin system genes in congenital diaphragmatic hernia associated with persistent pulmonary hypertension. *J. Pediatr. Surg.*, 39(3):302–6; discussion 302–6.

[480] Soler Artigas, M., Loth, D. W., Wain, L. V., Gharib, S. A., Obeidat, M., Tang, W., Zhai, G., Zhao, J. H., Smith, A. V., Huffman, J. E., Albrecht, E., Jackson, C. M., Evans, D. M., Cadby, G., Fornage, M., Manichaikul, A., Lopez, L. M., Johnson, T., Aldrich, M. C., Aspelund, T., Barroso, I., Campbell, H., Cassano, P. A., Couper, D. J., Eiriksdottir, G., Franceschini, N., Garcia, M., Gieger, C., Gislason, G. K., Grkovic, I., Hammond, C. J., Hancock, D. B., Harris, T. B., Ramasamy, A., Heckbert, S. R., Heliövaara, M., Homuth, G., Hysi, P. G., James, A. L., Jankovic, S., Joubert, B. R., Karrasch, S., Klopp, N., Koch, B., Kritchevsky, S. B., Launer, L. J., Liu, Y., Loehr, L. R., Lohman, K., Loos, R. J. F., Lumley, T., Al Balushi, K. A., Ang, W. Q., Barr, R. G., Beilby, J., Blakey, J. D., Boban, M., Boraska, V., Brisman, J., Britton, J. R., Brusselle, G. G., Cooper, C., Curjuric, I., Dahgam, S., Deary, I. J., Ebrahim, S., Eijgelsheim, M., Francks, C., Gaysina, D., Granell, R., Gu, X., Hankinson, J. L., Hardy, R., Harris, S. E., Henderson, J., Henry, A., Hingorani, A. D., Hofman, A., Holt, P. G., Hui, J., Hunter, M. L., Imboden, M., Jameson, K. A., Kerr, S. M., Kolcic, I., Kronenberg, F., Liu, J. Z., Marchini, J., McKeever, T., Morris, A. D., Olin, A.-C., Porteous, D. J., Postma, D. S., Rich, S. S., Ring, S. M., Rivadeneira, F., Rochat, T., Sayer, A. A., Sayers, I., Sly, P. D., Smith, G. D., Sood, A., Starr, J. M., Uitterlinden, A. G., Vonk, J. M., Wannamethee, S. G., Whincup, P. H., Wijmenga, C., Williams, O. D., Wong, A., Mangino, M., Marciante, K. D., McArdle, W. L., Meibohm, B., Morrison, A. C., North, K. E., Omenaas, E., Palmer, L. J., Pietiläinen, K. H., Pin, I., Pola Sbreve Ek, O., Pouta, A., Psaty, B. M., Hartikainen, A.-L., Rantanen, T., Ripatti, S., Rotter, J. I., Rudan, I., Rudnicka, A. R., Schulz, H., Shin, S.-Y., Spector, T. D., Surakka, I., Vitart, V., Völzke, H., Wareham, N. J., Warrington, N. M., Wichmann, H.-E., Wild, S. H., Wilk, J. B., Wjst, M., Wright, A. F., Zgaga, L., Zemunik, T., Pennell, C. E., Nyberg, F., Kuh, D., Holloway, J. W., Boezen, H. M., Lawlor, D. A., Morris, R. W., Probst-Hensch, N., International Lung Cancer Consortium, GIANT consortium, Kaprio, J., Wilson, J. F., Hayward, C., Kähönen, M., Heinrich, J., Musk, A. W., Jarvis, D. L., Gläser, S., Järvelin, M.-R., Ch Stricker, B. H., Elliott, P., O'Connor, G. T., Strachan, D. P., London, S. J., Hall, I. P., Gudnason, V., and Tobin, M. D. (2011). Genome-wide association and large-scale follow up identifies 16 new loci influencing lung function. *Nat. Genet.*, 43(11):1082–1090.

[481] Song, J., Eichstaedt, C. A., Viales, R. R., Benjamin, N., Harutyunova, S., Fischer, C., Grünig, E., and Hinderhofer, K. (2016). Identification of genetic defects in pulmonary arterial hypertension by a new gene panel diagnostic tool. *Clin. Sci.*, 130(22):2043–2052.

[482] Song, Y., Jones, J. E., Beppu, H., Keaney, Jr, J. F., Loscalzo, J., and Zhang, Y.-Y. (2005). Increased susceptibility to pulmonary hypertension in heterozygous BMPR2-mutant mice. *Circulation*, 112(4):553–562.

[483] Soon, E., Crosby, A., Southwood, M., Yang, P., Tajsic, T., Toshner, M., Appleby, S., Shanahan, C. M., Bloch, K. D., Pepke-Zaba, J., Upton, P., and Morrell, N. W. (2015). Bone morphogenetic protein receptor type II deficiency and increased inflammatory cytokine production. a gateway to pulmonary arterial hypertension. *Am. J. Respir. Crit. Care Med.*, 192(7):859–872.

[484] Southgate, L., Machado, R. D., Gräf, S., and Morrell, N. W. (2020). Molecular genetic framework underlying pulmonary arterial hypertension. *Nat. Rev. Cardiol.*, 17(2):85–95.

[485] Spiekerkoetter, E., Fabel, H., and Hoeper, M. M. (2002). Effects of inhaled salbutamol in primary pulmonary hypertension. *Eur. Respir. J.*, 20(3):524–528.

[486] Spiekerkoetter, E., Sung, Y. K., Sudheendra, D., Scott, V., Del Rosario, P., Bill, M., Haddad, F., Long-Boyle, J., Hedlin, H., and Zamanian, R. T. (2017). Randomised placebo-controlled safety and tolerability trial of FK506 (tacrolimus) for pulmonary arterial hypertension. *Eur. Respir. J.*, 50(3).

[487] St Laurent, G., Wahlestedt, C., and Kapranov, P. (2015). The landscape of long noncoding RNA classification. *Trends Genet.*, 31(5):239–251.

[488] Ständker, L., Schrader, M., Kanse, S. M., Jürgens, M., Forssmann, W. G., and Preissner, K. T. (1997). Isolation and characterization of the circulating form of human endostatin. *FEBS Lett.*, 420(2-3):129–133.

[489] Stanojevic, S., Wade, A., and Stocks, J. (2010). Reference values for lung function: past, present and future. *Eur. Respir. J.*, 36(1):12–19.

[490] Steen, V. D., Conte, C., Owens, G. R., and Medsger, Jr, T. A. (1994). Severe restrictive lung disease in systemic sclerosis. *Arthritis Rheum.*, 37(9):1283–1289.

[491] Stenmark, K. R., Meyrick, B., Galie, N., Mooi, W. J., and McMurtry, I. F. (2009). Animal models of pulmonary arterial hypertension: the hope for etiological discovery and pharmacological cure. *Am. J. Physiol. Lung Cell. Mol. Physiol.*, 297(6):L1013–32.

[492] Stern, D. A., Morgan, W. J., Wright, A. L., Guerra, S., and Martinez, F. D. (2007). Poor airway function in early infancy and lung function by age 22 years: a non-selective longitudinal cohort study. *Lancet*, 370(9589):758–764.

[493] Strafella, C., Caputo, V., Minozzi, G., Milano, F., Arcangeli, M., Sobhy, N., Abdelmaksood, R., Hashad, D., Vakirlis, E., Novelli, G., Cascella, R., and Giardina, E. (2019). Atopic eczema: Genetic analysis of COL6a5, COL8a1, and COL10a1 in mediterranean populations. *BioMed Research International*, 2019:1–7.

[494] Stram, M., Gigliotti, T., Hartman, D., Pitkus, A., Huff, S. M., Riben, M., Henricks, W. H., Farahani, N., and Pantanowitz, L. (2020). Logical observation identifiers names and codes for laboratorians. *Arch. Pathol. Lab. Med.*, 144(2):229–239.

[495] Su, H., Xu, X., Yan, C., Shi, Y., Hu, Y., Dong, L., Ying, S., Ying, K., and Zhang, R. (2018). LncRNA h19 promotes the proliferation of pulmonary artery smooth muscle cells through AT1r via sponging let-7b in monocrotaline-induced pulmonary arterial hypertension. *RESPIRATORY RESEARCH*, 19(1).

[496] Suhrie, K., Pajor, N. M., Ahlfeld, S. K., Dawson, D. B., Dufendach, K. R., Kitzmiller, J. A., Leino, D., Lombardo, R. C., Smolarek, T. A., Rathbun, P. A., Whitsett, J. A., Towe, C., and Wikenheiser-Brokamp, K. A. (2019). Neonatal lung disease associated with TBX4 mutations. *J. Pediatr.*, 206:286–292.e1.

[497] Sui, H., Han, B. G., Lee, J. K., Walian, P., and Jap, B. K. (2001). Structural basis of water-specific transport through the AQP1 water channel. *Nature*, 414(6866):872–878.

[498] Sun, X.-G., Hansen, J. E., Oudiz, R. J., and Wasserman, K. (2003). Pulmonary function in primary pulmonary hypertension. *J. Am. Coll. Cardiol.*, 41(6):1028–1035.

[499] Sun, Z., Nie, X., Sun, S., Dong, S., Yuan, C., Li, Y., Xiao, B., Jie, D., and Liu, Y. (2017). Long Non-Coding RNA MEG3 downregulation triggers human pulmonary artery smooth muscle cell proliferation and migration via the p53 signaling pathway. *Cell. Physiol. Biochem.*, 42(6):2569–2581.

[500] Sutendra, G. and Michelakis, E. D. (2014). The metabolic basis of pulmonary arterial hypertension. *Cell Metab.*, 19(4):558–573.

[501] Sutliff, R. L., Kang, B.-Y., and Hart, C. M. (2010). PPAR as a potential therapeutic target in pulmonary hypertension. *Therapeutic Advances in Respiratory Disease*, 4(3):143–160.

[502] Sweatt, A. J., Hedlin, H. K., Balasubramanian, V., Hsi, A., Blum, L. K., Robinson, W. H., Haddad, F., Hickey, P. M., Condliffe, R., Lawrie, A., Nicolls, M. R., Rabinovitch, M., Khatri, P., and Zamanian, R. T. (2019). Discovery of distinct immune phenotypes using machine learning in pulmonary arterial hypertension. *Circ. Res.*, 124(6):904–919.

[503] Swietlik, E. M., Greene, D., Zhu, N., Megy, K., Cogliano, M., Rajaram, S., Pandya, D., Tilly, T., Lutz, K. A., Welch, C. C. L., Pauciulo, M. W., Southgate, L., Martin, J. M., Treacy, C. M., Penkett, C. J., Stephens, J. C., Bogaard, H. J., Church, C., Coghlan, G., Coleman, A. W., Condliffe, R., Eichstaedt, C. A., Eyries, M., Gall, H., Ghio, S., Girerd, B., Grünig, E., Holden, S., Howard, L., Humbert, M., Kiely, D. G., Kovacs, G., Lordan, J., Machado, R. D., Mackenzie Ross, R. V., McCabe, C., Moledina, S., Montani, D., Olschewski, H., Pepke-Zaba, J., Price, L., Rhodes, C. J., Seeger, W., Soubrier, F., Suntharalingam, J., Toshner, M. R., Vonk Noordegraaf, A., Wharton, J., Wild, J. M., Wort, S. J., Lawrie, A., Wilkins, M. R., Trembath, R. C., Shen, Y., Chung, W. K., Swift, A. J., Nichols, W. C., Morrell, N. W., and Gräf, S. (2020). Bayesian inference associates rare variants with specific phenotypes in pulmonary arterial hypertension. *Circ Genom Precis Med*.

[504] Sztrymf, B., Coulet, F., Girerd, B., Yaici, A., Jais, X., Sitbon, O., Montani, D., Souza, R., Simonneau, G., Soubrier, F., and Humbert, M. (2008). Clinical outcomes of pulmonary arterial hypertension in carriers of BMPR2 mutation. *Am. J. Respir. Crit. Care Med.*, 177(12):1377–1383.

[505] Szturmowicz, M., Kacprzak, A., Franczuk, M., Burakowska, B., Kurzyna, M., Fijałkowska, A., Skoczylas, A., Wesołowski, S., Kuś, J., and Torbicki, A. (2016). Low DLCO in idiopathic pulmonary arterial hypertension - clinical correlates and prognostic significance. *Pneumonol. Alergol. Pol.*, 84(2):87–94.

[506] Tabima, D. M., Hacker, T. A., and Chesler, N. C. (2010). Measuring right ventricular function in the normal and hypertensive mouse hearts using admittance-derived pressure-volume loops. *Am. J. Physiol. Heart Circ. Physiol.*, 299(6):H2069–75.

[507] Taliun, D., Harris, D. N., Kessler, M. D., Carlson, J., Szpiech, Z. A., Torres, R., Taliun, S. A. G., Corvelo, A., Gogarten, S. M., Kang, H. M., Pitsillides, A. N., LeFaive, J., Lee, S.-B., Tian, X., Browning, B. L., Das, S., Emde, A.-K., Clarke, W. E., Loesch, D. P., Shetty, A. C., Blackwell, T. W., Smith, A. V., Wong, Q., Liu, X., Conomos, M. P., Bobo,

D. M., Aguet, F., Albert, C., Alonso, A., Ardlie, K. G., Arking, D. E., Aslibekyan, S., Auer, P. L., Barnard, J., Barr, R. G., Barwick, L., Becker, L. C., Beer, R. L., Benjamin, E. J., Bielak, L. F., Blangero, J., Boehnke, M., Bowden, D. W., Brody, J. A., Burchard, E. G., Cade, B. E., Casella, J. F., Chalazan, B., Chasman, D. I., Chen, Y.-D. I., Cho, M. H., Choi, S. H., Chung, M. K., Clish, C. B., Correa, A., Curran, J. E., Custer, B., Darbar, D., Daya, M., de Andrade, M., DeMeo, D. L., Dutcher, S. K., Ellinor, P. T., Emery, L. S., Eng, C., Fatkin, D., Fingerlin, T., Forer, L., Fornage, M., Franceschini, N., Fuchsberger, C., Fullerton, S. M., Germer, S., Gladwin, M. T., Gottlieb, D. J., Guo, X., Hall, M. E., He, J., Heard-Costa, N. L., Heckbert, S. R., Irvin, M. R., Johnsen, J. M., Johnson, A. D., Kaplan, R., Kardia, S. L. R., Kelly, T., Kelly, S., Kenny, E. E., Kiel, D. P., Klemmer, R., Konkle, B. A., Kooperberg, C., Köttgen, A., Lange, L. A., Lasky-Su, J., Levy, D., Lin, X., Lin, K.-H., Liu, C., Loos, R. J. F., Garman, L., Gerszten, R., Lubitz, S. A., Lunetta, K. L., Mak, A. C. Y., Manichaikul, A., Manning, A. K., Mathias, R. A., McManus, D. D., McGarvey, S. T., Meigs, J. B., Meyers, D. A., Mikulla, J. L., Minear, M. A., Mitchell, B. D., Mohanty, S., Montasser, M. E., Montgomery, C., Morrison, A. C., Murabito, J. M., Natale, A., Natarajan, P., Nelson, S. C., North, K. E., O'Connell, J. R., Palmer, N. D., Pankratz, N., Peloso, G. M., Peyser, P. A., Pleiness, J., Post, W. S., Psaty, B. M., Rao, D. C., Redline, S., Reiner, A. P., Roden, D., Rotter, J. I., Ruczinski, I., Sarnowski, C., Schoenherr, S., Schwartz, D. A., Seo, J.-S., Seshadri, S., Sheehan, V. A., Sheu, W. H., Shoemaker, M. B., Smith, N. L., Smith, J. A., Sotoodehnia, N., Stilp, A. M., Tang, W., Taylor, K. D., Telen, M., Thornton, T. A., Tracy, R. P., Van Den Berg, D. J., Vasan, R. S., Viaud-Martinez, K. A., Vrieze, S., Weeks, D. E., Weir, B. S., Weiss, S. T., Weng, L.-C., Willer, C. J., Zhang, Y., Zhao, X., Arnett, D. K., Ashley-Koch, A. E., Barnes, K. C., Boerwinkle, E., Gabriel, S., Gibbs, R., Rice, K. M., Rich, S. S., Silverman, E. K., Qasba, P., Gan, W., NHLBI Trans-Omics for Precision Medicine (TOPMed) Consortium, Papanicolaou, G. J., Nickerson, D. A., Browning, S. R., Zody, M. C., Zöllner, S., Wilson, J. G., Cupples, L. A., Laurie, C. C., Jaquish, C. E., Hernandez, R. D., O'Connor, T. D., and Abecasis, G. R. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed program. *Nature*, 590(7845):290–299.

[508] Tamby, M. C. (2005). Anti-endothelial cell antibodies in idiopathic and systemic sclerosis associated pulmonary arterial hypertension. *Thorax*, 60(9):765–772.

[509] Tamby, M. C., Humbert, M., Guilpain, P., Servettaz, A., Dupin, N., Christner, J. J., Simonneau, G., Fermanian, J., Weill, B., Guillevin, L., and Mouthon, L. (2006). Antibodies to fibroblasts in idiopathic and scleroderma-associated pulmonary hypertension. *Eur. Respir. J.*, 28(4):799–807.

[510] Tana, M. M., Shums, Z., Milo, J., Norman, G. L., Leung, P. S., Gershwin, M. E., Noureddin, M., Kleiner, D. E., Zhao, X., Heller, T., and Hoofnagle, J. H. (2015). The significance of autoantibody changes over time in primary biliary cirrhosis. *Am. J. Clin. Pathol.*, 144(4):601–606.

[511] Tao, B., Kraehling, J. R., Ghaffari, S., Ramirez, C. M., Lee, S., Fowler, J. W., Lee, W. L., Fernandez-Hernando, C., Eichmann, A., and Sessa, W. C. (2020). BMP-9 and LDL crosstalk regulates ALK-1 endocytosis and LDL transcytosis in endothelial cells. *J. Biol. Chem.*, 295(52):18179–18188.

[512] Tarantini, S., Tucsek, Z., Valcarcel-Ares, M. N., Toth, P., Gautam, T., Giles, C. B., Ballabh, P., Wei, J. Y., Wren, J. D., Ashpole, N. M., Sonntag, W. E., Ungvari, Z.,

and Csiszar, A. (2016). Circulating IGF-1 deficiency exacerbates hypertension-induced microvascular rarefaction in the mouse hippocampus and retrosplenial cortex: implications for cerebromicrovascular and brain aging. *Age*, 38(4):273–289.

[513] Taraseviciene-Stewart, L., Kasahara, Y., Alger, L., Hirth, P., Mc Mahon, G., Waltenberger, J., Voelkel, N. F., and Tuder, R. M. (2001). Inhibition of the VEGF receptor 2 combined with chronic hypoxia causes cell death-dependent pulmonary endothelial cell proliferation and severe pulmonary hypertension. *FASEB J.*, 15(2):427–438.

[514] Tenesa, A. and Haley, C. S. (2013). The heritability of human disease: estimation, uses and abuses. *Nat. Rev. Genet.*, 14(2):139–149.

[515] Terman, B. I., Carrion, M. E., Kovacs, E., Rasmussen, B. A., Eddy, R. L., and Shows, T. B. (1991). Identification of a new endothelial cell growth factor receptor tyrosine kinase. *Oncogene*, 6(9):1677–1683.

[516] Terzikhan, N., Sun, F., Verhamme, F. M., Adams, H. H. H., Loth, D., Bracke, K. R., Stricker, B. H. C., Lahousse, L., Dupuis, J., Brusselle, G. G., and O'Connor, G. T. (2018). Heritability and genome-wide association study of diffusing capacity of the lung. *Eur. Respir. J.*, 52(3).

[517] Test, V., Farber, H., McGoon, M., Parsons, L., and Channick, R. (2009). Pulmonary arterial hypertension in the elderly: Baseline characteristics and evaluation of therapeutics. an examination of the reveal registry. In *B27. FROM ALPHA TO OMEGA: ASSESSMENT AND OUTCOMES IN PULMONARY HYPERTENSION*. American Thoracic Society.

[518] Thaventhiran, J. E. D., Lango Allen, H., Burren, O. S., Rae, W., Greene, D., Staples, E., Zhang, Z., Farmery, J. H. R., Simeoni, I., Rivers, E., Maimaris, J., Penkett, C. J., Stephens, J., Deevi, S. V. V., Sanchis-Juan, A., Gleadall, N. S., Thomas, M. J., Sargur, R. B., Gordins, P., Baxendale, H. E., Brown, M., Tuijnenburg, P., Worth, A., Hanson, S., Linger, R. J., Buckland, M. S., Rayner-Matthews, P. J., Gilmour, K. C., Samarghitean, C., Seneviratne, S. L., Sansom, D. M., Lynch, A. G., Megy, K., Ellinghaus, E., Ellinghaus, D., Jorgensen, S. F., Karlsen, T. H., Stirrups, K. E., Cutler, A. J., Kumararatne, D. S., Chandra, A., Edgar, J. D. M., Herwadkar, A., Cooper, N., Grigoriadou, S., Huissoon, A. P., Goddard, S., Jolles, S., Schuetz, C., Boschann, F., Primary Immunodeficiency Consortium for the NIHR Bioresource, Lyons, P. A., Hurles, M. E., Savic, S., Burns, S. O., Kuijpers, T. W., Turro, E., Ouwehand, W. H., Thrasher, A. J., and Smith, K. G. C. (2020). Whole-genome sequencing of a sporadic primary immunodeficiency cohort. *Nature*, 583(7814):90–95.

[519] Thenappan, T., Shah, S. J., Rich, S., and Gomberg-Maitland, M. (2007). A USA-based registry for pulmonary arterial hypertension: 1982-2006. *Eur. Respir. J.*, 30(6):1103–1110.

[520] Therneau, T. M. (2021). *A Package for Survival Analysis in R*. R package version 3.2-11.

[521] Thomson, J. R. (2000). Sporadic primary pulmonary hypertension is associated with germline mutations of the gene encoding BMPR-II, a receptor member of the TGF-beta family. *Journal of Medical Genetics*, 37(10):741–745.

[522] Thoré, P., Girerd, B., Jaïs, X., Savale, L., Ghigna, M.-R., Eyries, M., Levy, M., Ovaert, C., Servettaz, A., Guillaumot, A., Dauphin, C., Chabanne, C., Boiffard, E., Cottin, V., Perros, F., Simonneau, G., Sitbon, O., Soubrier, F., Bonnet, D., Remy-Jardin, M., Chaouat, A., Humbert, M., and Montani, D. (2020). Phenotype and outcome of pulmonary arterial hypertension patients carrying a TBX4 mutation. *European Respiratory Journal*, 55(5):1902340.

[523] Thum, T. and Condorelli, G. (2015). Long noncoding RNAs and MicroRNAs in cardiovascular pathophysiology. *Circulation Research*, 116(4):751–762.

[524] Tian, L., Gao, J., Garcia, I. M., Chen, H. J., Castaldi, A., and Chen, Y.-W. (2020). Human pluripotent stem cell-derived lung organoids: Potential applications in development and disease modeling. *Wiley Interdiscip. Rev. Dev. Biol.*, page e399.

[525] Tian, W., Jiang, X., Sung, Y. K., Shuffle, E., Wu, T.-H., Kao, P. N., Tu, A. B., Dorfmüller, P., Cao, A., Wang, L., Peng, G., Kim, Y., Zhang, P., Chappell, J., Pasupneti, S., Dahms, P., Maguire, P., Chaib, H., Zamanian, R., Peters-Golden, M., Snyder, M. P., Voelkel, N. F., Humbert, M., Rabinovitch, M., and Nicolls, M. R. (2019). Phenotypically silent bone morphogenetic protein receptor 2 mutations predispose rats to Inflammation-Induced pulmonary arterial hypertension by enhancing the risk for neointimal transformation. *Circulation*, 140(17):1409–1425.

[526] Tibshirani, R., Walther, G., and Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2):411–423.

[527] Tie, L., Wang, D., Shi, Y., and Li, X. (2017). Aquaporins in cardiovascular system. *Adv. Exp. Med. Biol.*, 969:105–113.

[528] Tillet, E., Ouarné, M., Desroches-Castan, A., Mallet, C., Subileau, M., Didier, R., Lioutsko, A., Belthier, G., Feige, J.-J., and Bailly, S. (2018). A heterodimer formed by bone morphogenetic protein 9 (BMP9) and BMP10 provides most BMP biological activity in plasma. *Journal of Biological Chemistry*, 293(28):10963–10974.

[529] Topiwala, K. K., Patel, S. D., Nouh, A. M., and Alberts, M. J. (2020). Novel GDF2 gene mutation associated with pulmonary arteriovenous malformation. *J. Stroke Cerebrovasc. Dis.*, 29(12):105301.

[530] Toshner, M., Voswinckel, R., Southwood, M., Al-Lamki, R., Howard, L. S. G., Marchesan, D., Yang, J., Suntharalingam, J., Soon, E., Exley, A., Stewart, S., Hecker, M., Zhu, Z., Gehling, U., Seeger, W., Pepke-Zaba, J., and Morrell, N. W. (2009). Evidence of dysfunction of endothelial progenitors in pulmonary arterial hypertension. *Am. J. Respir. Crit. Care Med.*, 180(8):780–787.

[531] Trembath, R. C., Thomson, J. R., Machado, R. D., Morgan, N. V., Atkinson, C., Winship, I., Simonneau, G., Galie, N., Loyd, J. E., Humbert, M., Nichols, W. C., Morrell, N. W., Berg, J., Manes, A., McGaughran, J., Pauciulo, M., and Wheeler, L. (2001). Clinical and molecular genetic features of pulmonary hypertension in patients with hereditary hemorrhagic telangiectasia. *N. Engl. J. Med.*, 345(5):325–334.

[532] Trip, P., Girerd, B., Bogaard, H.-J., de Man, F. S., Boonstra, A., Garcia, G., Humbert, M., Montani, D., and Vonk-Noordegraaf, A. (2014). Diffusion capacity and BMPR2 mutations in pulmonary arterial hypertension. *Eur. Respir. J.*, 43(4):1195–1198.

[533] Trip, P., Nossent, E. J., de Man, F. S., van den Berk, I. A. H., Boonstra, A., Groepenhoff, H., Leter, E. M., Westerhof, N., Grünberg, K., Bogaard, H.-J., and Vonk-Noordegraaf, A. (2013). Severely reduced diffusion capacity in idiopathic pulmonary arterial hypertension: patient characteristics and treatment responses. *Eur. Respir. J.*, 42(6):1575–1585.

[534] Tu, L., Desroches-Castan, A., Mallet, C., Guyon, L., Cumont, A., Phan, C., Robert, F., Thuillet, R., Bordenave, J., Sekine, A., Huertas, A., Ritvos, O., Savale, L., Feige, J.-J., Humbert, M., Bailly, S., and Guignabert, C. (2019). Selective BMP-9 inhibition partially protects against experimental pulmonary hypertension. *Circ. Res.*, 124(6):846–855.

[535] Tuder, R. M., Chacon, M., Alger, L., Wang, J., Taraseviciene-Stewart, L., Kasahara, Y., Cool, C. D., Bishop, A. E., Geraci, M., Semenza, G. L., Yacoub, M., Polak, J. M., and Voelkel, N. F. (2001). Expression of angiogenesis-related molecules in plexiform lesions in severe pulmonary hypertension: evidence for a process of disordered angiogenesis. *The Journal of Pathology*, 195(3):367–374.

[536] Tuder, R. M., Flook, B. E., and Voelkel, N. F. (1995). Increased gene expression for VEGF and the VEGF receptors KDR/Flk and flt in lungs exposed to acute or to chronic hypoxia. modulation of gene expression by nitric oxide. *J. Clin. Invest.*, 95(4):1798–1807.

[537] Turner, A. W., Wong, D., Khan, M. D., Dreisbach, C. N., Palmore, M., and Miller, C. L. (2019). Multi-omics approaches to study long non-coding RNA function in atherosclerosis. *Frontiers in Cardiovascular Medicine*, 6.

[538] Turro, E., Astle, W. J., Megy, K., Gräf, S., Greene, D., Shamardina, O., Allen, H. L., Sanchis-Juan, A., Frontini, M., Thys, C., Stephens, J., Mapeta, R., Burren, O. S., Downes, K., Haimel, M., Tuna, S., Deevi, S. V. V., Aitman, T. J., Bennett, D. L., Calleja, P., Carss, K., Caulfield, M. J., Chinnery, P. F., Dixon, P. H., Gale, D. P., James, R., Koziell, A., Laffan, M. A., Levine, A. P., Maher, E. R., Markus, H. S., Morales, J., Morrell, N. W., Mumford, A. D., Ormondroyd, E., Rankin, S., Rendon, A., Richardson, S., Roberts, I., Roy, N. B. A., Saleem, M. A., Smith, K. G. C., Stark, H., Tan, R. Y. Y., Themistocleous, A. C., Thrasher, A. J., Watkins, H., Webster, A. R., Wilkins, M. R., Williamson, C., Whitworth, J., Humphray, S., Bentley, D. R., NIHR BioResource for the 100,000 Genomes Project, Kingston, N., Walker, N., Bradley, J. R., Ashford, S., Penkett, C. J., Freson, K., Stirrups, K. E., Raymond, F. L., and Ouwehand, W. H. (2020). Whole-genome sequencing of patients with rare diseases in a national health system. *Nature*, 583(7814):96–102.

[539] Uchanska-Ziegler, B., Loll, B., Fabian, H., Hee, C. S., Saenger, W., and Ziegler, A. (2012). HLA class i-associated diseases with a suspected autoimmune etiology: HLA-B27 subtypes as a model system. *Eur. J. Cell Biol.*, 91(4):274–286.

[540] Uchida, S. and Dimmeler, S. (2015). Long noncoding RNAs in cardiovascular diseases. *Circulation Research*, 116(4):737–750.

[541] Ulrich, A., Otero-Núñez, P., Wharton, J., Swietlik, E. M., Gräf, S., Morrell, N. W., Wang, D., Lawrie, A., Wilkins, M. R., Prokopenko, I., Rhodes, C. J., Consortium, O. B. O.

T. N. B.-R. D., and Consortium, U. P. C. S. (2020a). Expression quantitative trait locus mapping in pulmonary arterial hypertension. *Genes*, 11(11).

[542] Ulrich, A., Wharton, J., Thayer, T. E., Swietlik, E. M., Assad, T. R., Desai, A. A., Gräf, S., Harbaum, L., Humbert, M., Morrell, N. W., Nichols, W. C., Soubrier, F., Southgate, L., Trégouët, D.-A., Trembath, R. C., Brittain, E. L., Wilkins, M. R., Prokopenko, I., Rhodes, C. J., NIHR BioResource – Rare Diseases Consortium, UK PAH Cohort Study Consortium, and US PAH Biobank Consortium (2020b). Mendelian randomisation analysis of red cell distribution width in pulmonary arterial hypertension. *Eur. Respir. J.*, 55(2).

[543] Ungvari, Z., Kaley, G., de Cabo, R., Sonntag, W. E., and Csiszar, A. (2010). Mechanisms of vascular aging: new perspectives. *J. Gerontol. A Biol. Sci. Med. Sci.*, 65(10):1028–1041.

[544] Upton, P. D., Davies, R. J., Trembath, R. C., and Morrell, N. W. (2009). Bone morphogenetic protein (BMP) and activin type II receptors balance BMP9 signals mediated by activin receptor-like kinase-1 in human pulmonary artery endothelial cells. *J. Biol. Chem.*, 284(23):15794–15804.

[545] Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S., and Gregory, P. D. (2010). Genome editing with engineered zinc finger nucleases. *Nature Reviews Genetics*, 11(9):636–646.

[546] Vaillancourt, M., Ruffenach, G., Meloche, J., and Bonnet, S. (2015). Adaptation and remodelling of the pulmonary circulation in pulmonary hypertension. *Can. J. Cardiol.*, 31(4):407–415.

[547] van Boxtel, R. and Cuppen, E. (2010). Rat traps: filling the toolbox for manipulating the rat genome. *Genome Biol.*, 11(9):217.

[548] van Buuren, S. and Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations inR. *Journal of Statistical Software*, 45(3).

[549] van den Heuvel, L. M., Jansen, S. M. A., Alsters, S. I. M., Post, M. C., van der Smagt, J. J., Handoko-De Man, F. S., van Tintelen, J. P., Gille, H., Christiaans, I., Vonk Noordegraaf, A., Bogaard, H., and Houweling, A. C. (2020). Genetic evaluation in a cohort of 126 dutch pulmonary arterial hypertension patients. *Genes*, 11(10).

[550] van der Bruggen, C. E., Spruijt, O. A., Nossent, E. J., Trip, P., Marcus, J. T., de Man, F. S., Jan Bogaard, H., and Noordegraaf, A. V. (2017). Treatment response in patients with idiopathic pulmonary arterial hypertension and a severely reduced diffusion capacity. *Pulm. Circ.*, 7(1):137–144.

[551] van der Weyden, L., White, J. K., Adams, D. J., and Logan, D. W. (2011). The mouse genetics toolkit: revealing function and mechanism. *Genome Biol.*, 12(6):224.

[552] van Walraven, C., Austin, P. C., Jennings, A., Quan, H., and Forster, A. J. (2009). A modification of the elixhauser comorbidity measures into a point system for hospital death using administrative data. *Med. Care*, 47(6):626–633.

[553] Varshney, G. K., Carrington, B., Pei, W., Bishop, K., Chen, Z., Fan, C., Xu, L., Jones, M., LaFave, M. C., Ledin, J., Sood, R., and Burgess, S. M. (2016). A high-throughput functional genomics workflow based on CRISPR/Cas9-mediated targeted mutagenesis in zebrafish. *Nat. Protoc.*, 11(12):2357–2375.

[554] Vickovic, S., Eraslan, G., Salmén, F., Klughammer, J., Stenbeck, L., Schapiro, D., Äijö, T., Bonneau, R., Bergenstråhle, L., Navarro, J. F., Gould, J., Griffin, G. K., Borg, Å., Ronaghi, M., Frisén, J., Lundeberg, J., Regev, A., and Ståhl, P. L. (2019). High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods*, 16(10):987–990.

[555] Voordes, C. G., Kuipers, J. R., and Elema, J. D. (1977). Familial pulmonary veno-occlusive disease: a case report. *Thorax*, 32(6):763–766.

[556] Vordenbäumen, S., Brinks, R., Hoyer, A., Fischer-Betz, R., Pongratz, G., Lowin, T., Zucht, H.-D., Budde, P., Bleck, E., Schulz-Knappe, P., and Schneider, M. (2019). Comprehensive longitudinal surveillance of the IgG autoantibody repertoire in established systemic lupus erythematosus. *Arthritis Rheumatol*, 71(5):736–743.

[557] Wagatsuma, A. (2006). Effect of aging on expression of angiogenesis-related factors in mouse skeletal muscle. *Exp. Gerontol.*, 41(1):49–54.

[558] Wain, L. V., Shrine, N., Artigas, M. S., Erzurumluoglu, A. M., Noyvert, B., Bossini-Castillo, L., Obeidat, M., Henry, A. P., Portelli, M. A., Hall, R. J., Billington, C. K., Rimington, T. L., Fenech, A. G., John, C., Blake, T., Jackson, V. E., Allen, R. J., Prins, B. P., Understanding Society Scientific Group, Campbell, A., Porteous, D. J., Jarvelin, M.-R., Wielscher, M., James, A. L., Hui, J., Wareham, N. J., Zhao, J. H., Wilson, J. F., Joshi, P. K., Stubbe, B., Rawal, R., Schulz, H., Imboden, M., Probst-Hensch, N. M., Karrasch, S., Gieger, C., Deary, I. J., Harris, S. E., Marten, J., Rudan, I., Enroth, S., Gyllensten, U., Kerr, S. M., Polasek, O., Kähönen, M., Surakka, I., Vitart, V., Hayward, C., Lehtimäki, T., Raitakari, O. T., Evans, D. M., Henderson, A. J., Pennell, C. E., Wang, C. A., Sly, P. D., Wan, E. S., Busch, R., Hobbs, B. D., Litonjua, A. A., Sparrow, D. W., Gulsvik, A., Bakke, P. S., Crapo, J. D., Beaty, T. H., Hansel, N. N., Mathias, R. A., Ruczinski, I., Barnes, K. C., Bossé, Y., Joubert, P., van den Berge, M., Brandsma, C.-A., Paré, P. D., Sin, D. D., Nickle, D. C., Hao, K., Gottesman, O., Dewey, F. E., Bruse, S. E., Carey, D. J., Kirchner, H. L., Geisinger-Regeneron DiscovEHR Collaboration, Jonsson, S., Thorleifsson, G., Jonsdottir, I., Gislason, T., Stefansson, K., Schurmann, C., Nadkarni, G., Bottinger, E. P., Loos, R. J. F., Walters, R. G., Chen, Z., Millwood, I. Y., Vaucher, J., Kurmi, O. P., Li, L., Hansell, A. L., Brightling, C., Zeggini, E., Cho, M. H., Silverman, E. K., Sayers, I., Trynka, G., Morris, A. P., Strachan, D. P., Hall, I. P., and Tobin, M. D. (2017). Genome-wide association analyses for lung function and chronic obstructive pulmonary disease identify new loci and potential druggable targets. *Nat. Genet.*, 49(3):416–425.

[559] Waller-Evans, H., Prömel, S., Langenhan, T., Dixon, J., Zahn, D., Colledge, W. H., Doran, J., Carlton, M. B. L., Davies, B., Aparicio, S. A. J. R., Grosse, J., and Russ, A. P. (2010). The orphan adhesion-GPCR GPR126 is required for embryonic development in the mouse. *PLoS One*, 5(11):e14047.

[560] Wang, D., Xu, H., Wu, B., Jiang, S., Pan, H., Wang, R., and Chen, J. (2019). Long non-coding RNA MALAT1 sponges miR-124-3p.1/KLF5 to promote pulmonary vascular remodeling and cell cycle progression of pulmonary artery hypertension. *Int. J. Mol. Med.*, 44(3):871–884.

[561] Wang, G., Fan, R., Ji, R., Zou, W., Penny, D. J., Varghese, N. P., and Fan, Y. (2016). Novel homozygous BMP9 nonsense mutation causes pulmonary arterial hypertension: a case report. *BMC Pulm. Med.*, 16:17.

[562] Wang, G., Knight, L., Ji, R., Lawrence, P., Kanaan, U., Li, L., Das, A., Cui, B., Zou, W., Penny, D. J., and Fan, Y. (2014). Early onset severe pulmonary arterial hypertension with 'two-hit' digenic mutations in both BMPR2 and KCNA5 genes.

[563] Wang, S., Wiley, G. B., Kelly, J. A., and Gaffney, P. M. (2015). Disease mechanisms in rheumatology–tools and pathways: defining functional genetic variants in autoimmune diseases. *Arthritis Rheumatol*, 67(1):1–10.

[564] Wang, X.-J., Lian, T.-Y., Jiang, X., Liu, S.-F., Li, S.-Q., Jiang, R., Wu, W.-H., Ye, J., Cheng, C.-Y., Du, Y., Xu, X.-Q., Wu, Y., Peng, F.-H., Sun, K., Mao, Y.-M., Yu, H., Liang, C., Shyy, J. Y.-J., Zhang, S.-Y., Zhang, X., and Jing, Z.-C. (2018). Germline BMP9 mutation causes idiopathic pulmonary arterial hypertension. *European Respiratory Journal*, 53(3):1801609.

[565] Wanger, J., Clausen, J. L., Coates, A., Pedersen, O. F., Brusasco, V., Burgos, F., Casaburi, R., Crapo, R., Enright, P., van der Grinten, C. P. M., Gustafsson, P., Hankinson, J., Jensen, R., Johnson, D., MacIntyre, N., McKay, R., Miller, M. R., Navajas, D., Pellegrino, R., and Viegi, G. (2005). Standardisation of the measurement of lung volumes. *European Respiratory Journal*, 26(3):511–522.

[566] Warrington, J. P., Ashpole, N., Csiszar, A., Lee, Y. W., Ungvari, Z., and Sonntag, W. E. (2013). Whole brain radiation-induced vascular cognitive impairment: Mechanisms and implications. *Journal of Vascular Research*, 50(6):445–457.

[567] Waypa, G. B., Osborne, S. W., Marks, J. D., Berkelhamer, S. K., Kondapalli, J., and Schumacker, P. T. (2013). Sirtuin 3 deficiency does not augment hypoxia-induced pulmonary hypertension. *Am. J. Respir. Cell Mol. Biol.*, 49(6):885–891.

[568] Weismann, C. G. and Gelb, B. D. (2007). The genetics of congenital heart disease: a review of recent developments. *Curr. Opin. Cardiol.*, 22(3):200–206.

[569] West, B. H., Noureddin, N., Mamzhi, Y., Low, C. G., Coluzzi, A. C., Shih, E. J., Gevorgyan Fleming, R., Saver, J. L., Liebeskind, D. S., Charles, A., and Tobis, J. M. (2018). Frequency of patent foramen ovale and migraine in patients with cryptogenic stroke. *Stroke*, 49(5):1123–1128.

[570] West, J., Harral, J., Lane, K., Deng, Y., Ickes, B., Crona, D., Albu, S., Stewart, D., and Fagan, K. (2008). Mice expressing BMPR2R899X transgene in smooth muscle develop pulmonary vascular lesions. *Am. J. Physiol. Lung Cell. Mol. Physiol.*, 295(5):L744–55.

[571] Whicher, D., Philbin, S., and Aronson, N. (2018). An overview of the impact of rare disease characteristics on research methodology. *Orphanet J. Rare Dis.*, 13(1):14.

[572] White, I. R., Royston, P., and Wood, A. M. (2011). Multiple imputation using chained equations: Issues and guidance for practice. *Stat. Med.*, 30(4):377–399.

[573] Wijeratne, D. T., Lajkosz, K., Brogly, S. B., Lougheed, M. D., Jiang, L., Housin, A., Barber, D., Johnson, A., Doliszny, K. M., and Archer, S. L. (2018). Increasing incidence and prevalence of world health organization groups 1 to 4 pulmonary hypertension. *Circulation: Cardiovascular Quality and Outcomes*, 11(2).

[574] Winter, M.-P., Sharma, S., Altmann, J., Seidl, V., Panzenböck, A., Alimohammadi, A., Zelniker, T., Redwan, B., Nagel, F., Santer, D., Stieglbauer, A., Podesser, B., Sibilia, M., Helbich, T., Prager, G., Ilhan-Mutlu, A., Preusser, M., and Lang, I. M. (2020). Interruption of vascular endothelial growth factor receptor 2 signaling induces a proliferative pulmonary vasculopathy and pulmonary hypertension. *Basic Res. Cardiol.*, 115(6):58.

[575] Wooderchak-Donahue, W. L., McDonald, J., O'Fallon, B., Upton, P. D., Li, W., Roman, B. L., Young, S., Plant, P., Fülöp, G. T., Langa, C., Morrell, N. W., Botella, L. M., Bernabeu, C., Stevenson, D. A., Runo, J. R., and Bayrak-Toydemir, P. (2013). BMP9 mutations cause a vascular-anomaly syndrome with phenotypic overlap with hereditary hemorrhagic telangiectasia. *The American Journal of Human Genetics*, 93(3):530–537.

[576] Wright, C. F., FitzPatrick, D. R., and Firth, H. V. (2018). Paediatric genomics: diagnosing rare disease in children. *Nature Reviews Genetics*, 19(5):253–268.

[577] Wright, S. (1920). The relative importance of heredity and environment in determining the piebald pattern of Guinea-Pigs. *Proc. Natl. Acad. Sci. U. S. A.*, 6(6):320–332.

[578] Wu, M. C., Lee, S., Cai, T., Li, Y., Boehnke, M., and Lin, X. (2011). Rare-variant association testing for sequencing data with the sequence kernel association test. *Am. J. Hum. Genet.*, 89(1):82–93.

[579] Xiang, Y., Zhang, Y., Tang, Y., and Li, Q. (2017). MALAT1 modulates TGF-1-induced endothelial-to-mesenchymal transition through downregulation of miR-145. *Cellular Physiology and Biochemistry*, 42(1):357–372.

[580] Xing, Y., Zheng, X., Fu, Y., Qi, J., Li, M., Ma, M., Wang, S., Li, S., and Zhu, D. (2019). Long noncoding RNA-Maternally expressed gene 3 contributes to hypoxic pulmonary hypertension. *Mol. Ther.*, 27(12):2166–2181.

[581] Xu, X., Li, X., Yang, G., Li, L., Hu, W., Zhang, L., Liu, H., Zheng, H., Tan, M., and Zhu, D. (2017). Circulating bone morphogenetic protein-9 in relation to metabolic syndrome and insulin resistance. *Scientific Reports*, 7(1).

[582] Yang, L., Liang, H., Shen, L., Guan, Z., and Meng, X. (2019). LncRNA tug1 involves in the pulmonary vascular remodeling in mice with hypoxic pulmonary hypertension via the microRNA-374c-mediated foxc1. *Life Sci.*, 237:116769.

[583] Yang, X., Long, L., Southwood, M., Rudarakanchana, N., Upton, P. D., Jeffery, T. K., Atkinson, C., Chen, H., Trembath, R. C., and Morrell, N. W. (2005). Dysfunctional smad signaling contributes to abnormal smooth muscle cell proliferation in familial pulmonary arterial hypertension. *Circulation Research*, 96(10):1053–1063.

[584] Yu, Y., Bhangale, T. R., Fagerness, J., Ripke, S., Thorleifsson, G., Tan, P. L., Souied, E. H., Richardson, A. J., Merriam, J. E., Buitendijk, G. H. S., Reynolds, R., Raychaudhuri, S., Chin, K. A., Sobrin, L., Evangelou, E., Lee, P. H., Lee, A. Y., Leveziel, N., Zack,

D. J., Campochiaro, B., Campochiaro, P., Smith, R. T., Barile, G. R., Guymer, R. H., Hogg, R., Chakravarthy, U., Robman, L. D., Gustafsson, O., Sigurdsson, H., Ortmann, W., Behrens, T. W., Stefansson, K., Uitterlinden, A. G., van Duijn, C. M., Vingerling, J. R., Klaver, C. C. W., Allikmets, R., Brantley, Jr, M. A., Baird, P. N., Katsanis, N., Thorsteinsdottir, U., Ioannidis, J. P. A., Daly, M. J., Graham, R. R., and Seddon, J. M. (2011). Common variants near FRK/COL10A1 and VEGFA are associated with advanced age-related macular degeneration. *Hum. Mol. Genet.*, 20(18):3699–3709.

[585] Yun, X., Jiang, H., Lai, N., Wang, J., and Shimoda, L. A. (2017). Aquaporin 1-mediated changes in pulmonary arterial smooth muscle cell migration and proliferation involve β-catenin. *Am. J. Physiol. Lung Cell. Mol. Physiol.*, 313(5):L889–L898.

[586] Zaidi, S., Choi, M., Wakimoto, H., Ma, L., Jiang, J., Overton, J. D., Romano-Adesman, A., Bjornson, R. D., Breitbart, R. E., Brown, K. K., Carriero, N. J., Cheung, Y. H., Deanfield, J., DePalma, S., Fakhro, K. A., Glessner, J., Hakonarson, H., Italia, M. J., Kaltman, J. R., Kaski, J., Kim, R., Kline, J. K., Lee, T., Leipzig, J., Lopez, A., Mane, S. M., Mitchell, L. E., Newburger, J. W., Parfenov, M., Pe'er, I., Porter, G., Roberts, A. E., Sachidanandam, R., Sanders, S. J., Seiden, H. S., State, M. W., Subramanian, S., Tikhonova, I. R., Wang, W., Warburton, D., White, P. S., Williams, I. A., Zhao, H., Seidman, J. G., Brueckner, M., Chung, W. K., Gelb, B. D., Goldmuntz, E., Seidman, C. E., and Lifton, R. P. (2013). De novo mutations in histone-modifying genes in congenital heart disease. *Nature*, 498(7453):220–223.

[587] Zare, F., Dow, M., Monteleone, N., Hosny, A., and Nabavi, S. (2017). An evaluation of copy number variation detection tools for cancer using whole exome sequencing data. *BMC Bioinformatics*, 18(1):286.

[588] Zhang, C. Y., Baffy, G., Perret, P., Krauss, S., Peroni, O., Grujic, D., Hagen, T., Vidal-Puig, A. J., Boss, O., Kim, Y. B., Zheng, X. X., Wheeler, M. B., Shulman, G. I., Chan, C. B., and Lowell, B. B. (2001). Uncoupling protein-2 negatively regulates insulin secretion and is a major link between obesity, beta cell dysfunction, and type 2 diabetes. *Cell*, 105(6):745–755.

[589] Zhang, H., Liu, Y., Yan, L., Wang, S., Zhang, M., Ma, C., Zheng, X., Chen, H., and Zhu, D. (2019). Long noncoding RNA hoxaas3 contributes to hypoxia-induced pulmonary artery smooth muscle cell proliferation. *Cardiovasc. Res.*, 115(3):647–657.

[590] Zhang, P., McGrath, B. C., Reinert, J., Olsen, D. S., Lei, L., Gill, S., Wek, S. A., Vattem, K. M., Wek, R. C., Kimball, S. R., Jefferson, L. S., and Cavener, D. R. (2002). The GCN2 eIF2alpha kinase is required for adaptation to amino acid deprivation in mice. *Mol. Cell. Biol.*, 22(19):6681–6688.

[591] Zhang, R., Dai, L.-Z., Xie, W.-P., Yu, Z.-X., Wu, B.-X., Pan, L., Yuan, P., Jiang, X., He, J., Humbert, M., and Jing, Z.-C. (2011). Survival of chinese patients with pulmonary arterial hypertension in the modern treatment era. *Chest*, 140(2):301–309.

[592] Zhao, L., Chen, C.-N., Hajji, N., Oliver, E., Cotroneo, E., Wharton, J., Wang, D., Li, M., McKinsey, T. A., Stenmark, K. R., and Wilkins, M. R. (2012). Histone deacetylation inhibition in pulmonary hypertension: therapeutic potential of valproic acid and suberoylanilide hydroxamic acid. *Circulation*, 126(4):455–467.

[593] Zhao, L., Li, B., Dian, K., Ying, B., Lu, X., Hu, X., An, Q., Chen, C., Huang, C., Tan, B., and Qin, L. (2015). Association between the european GWAS-identified susceptibility locus at chromosome 4p16 and the risk of atrial septal defect: A case-control study in southwest china and a meta-analysis. *PLOS ONE*, 10(4):e0123959.

[594] Zhao, Y.-Y., Liu, Y., Stan, R.-V., Fan, L., Gu, Y., Dalton, N., Chu, P.-H., Peterson, K., Ross, Jr, J., and Chien, K. R. (2002). Defects in caveolin-1 cause dilated cardiomyopathy and pulmonary hypertension in knockout mice. *Proc. Natl. Acad. Sci. U. S. A.*, 99(17):11375–11380.

[595] Zhu, B., Gong, Y., Yan, G., Wang, D., Qiao, Y., Wang, Q., Liu, B., Hou, J., Li, R., and Tang, C. (2018a). Down-regulation of lncRNA MEG3 promotes hypoxia-induced human pulmonary artery smooth muscle cell proliferation and migration via repressing PTEN by sponging mir-21. *Biochem. Biophys. Res. Commun.*, 495(3):2125–2132.

[596] Zhu, N., , Swietlik, E. M., Welch, C. L., Pauciulo, M. W., Hagen, J. J., Zhou, X., Guo, Y., Karten, J., Pandya, D., Tilly, T., Lutz, K. A., Martin, J. M., Treacy, C. M., Rosenzweig, E. B., Krishnan, U., Coleman, A. W., Gonzaga-Jauregui, C., Lawrie, A., Trembath, R. C., Wilkins, M. R., Morrell, N. W., Shen, Y., Gräf, S., Nichols, W. C., Chung, W. K., , and and (2021). Rare variant analysis of 4241 pulmonary arterial hypertension cases from an international consortium implicates FBLN2, PDGFD, and rare de novo variants in PAH. *Genome Medicine*, 13(1).

[597] Zhu, N., Gonzaga-Jauregui, C., Welch, C. L., Ma, L., Qi, H., King, A. K., Krishnan, U., Rosenzweig, E. B., Ivy, D. D., Austin, E. D., Hamid, R., Nichols, W. C., Pauciulo, M. W., Lutz, K. A., Sawle, A., Reid, J. G., Overton, J. D., Baras, A., Dewey, F., Shen, Y., and Chung, W. K. (2018b). Exome sequencing in children with pulmonary arterial hypertension demonstrates differences compared with adults. *Circ Genom Precis Med*, 11(4):e001887.

[598] Zhu, N., Pauciulo, M. W., Welch, C. L., Lutz, K. A., Coleman, A. W., Gonzaga-Jauregui, C., Wang, J., Grimes, J. M., Martin, L. J., He, H., PAH Biobank Enrolling Centers' Investigators, Shen, Y., Chung, W. K., and Nichols, W. C. (2019a). Novel risk genes and mechanisms implicated by exome sequencing of 2572 individuals with pulmonary arterial hypertension. *Genome Med.*, 11(1):69.

[599] Zhu, N., Welch, C. L., Wang, J., Allen, P. M., Gonzaga-Jauregui, C., Ma, L., King, A. K., Krishnan, U., Rosenzweig, E. B., Ivy, D. D., Austin, E. D., Hamid, R., Pauciulo, M. W., Lutz, K. A., Nichols, W. C., Reid, J. G., Overton, J. D., Baras, A., Dewey, F. E., Shen, Y., and Chung, W. K. (2018c). Rare variants in SOX17 are associated with pulmonary arterial hypertension with congenital heart disease. *Genome Med.*, 10(1):691.

[600] Zhu, T.-T., Sun, R.-L., Yin, Y.-L., Quan, J.-P., Song, P., Xu, J., Zhang, M.-X., and Li, P. (2019b). Long noncoding RNA UCA1 promotes the proliferation of hypoxic human pulmonary artery smooth muscle cells. *Pflugers Arch.*, 471(2):347–355.

Appendix A

Appendix Figures

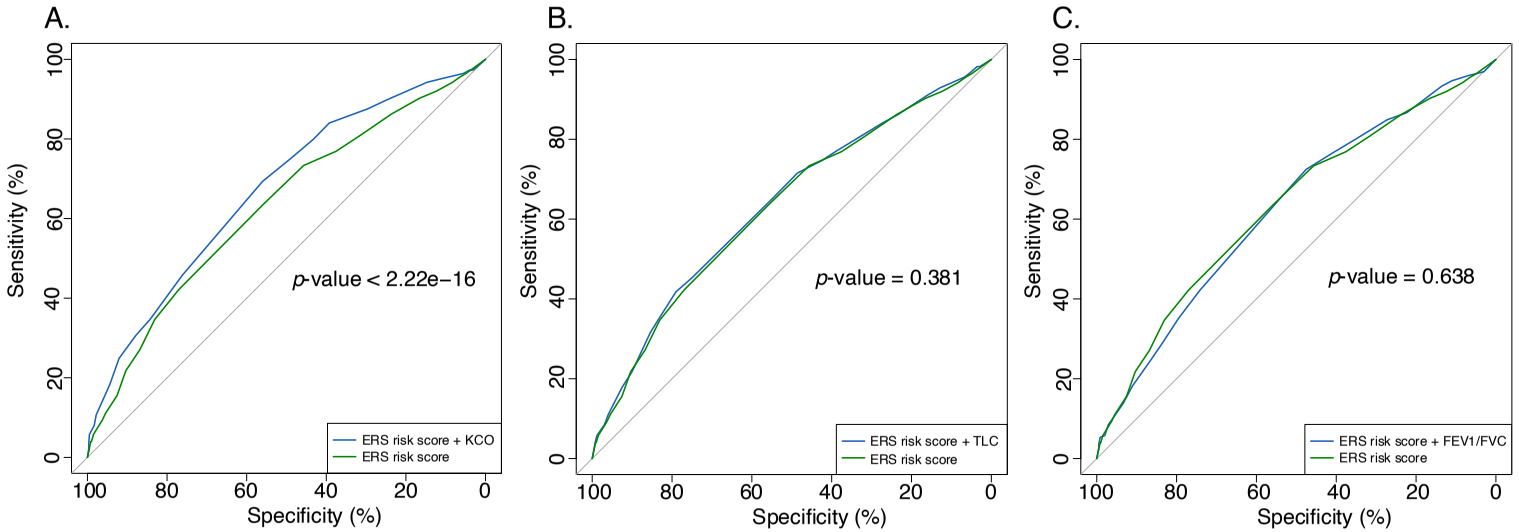


Fig. A.1 Comparison of ROC curves of ERS risk score: (A) ERS risk score + KCO, (B) ERS risk score + TLC, (C) ERS risk score + FEV1/FVC ratio

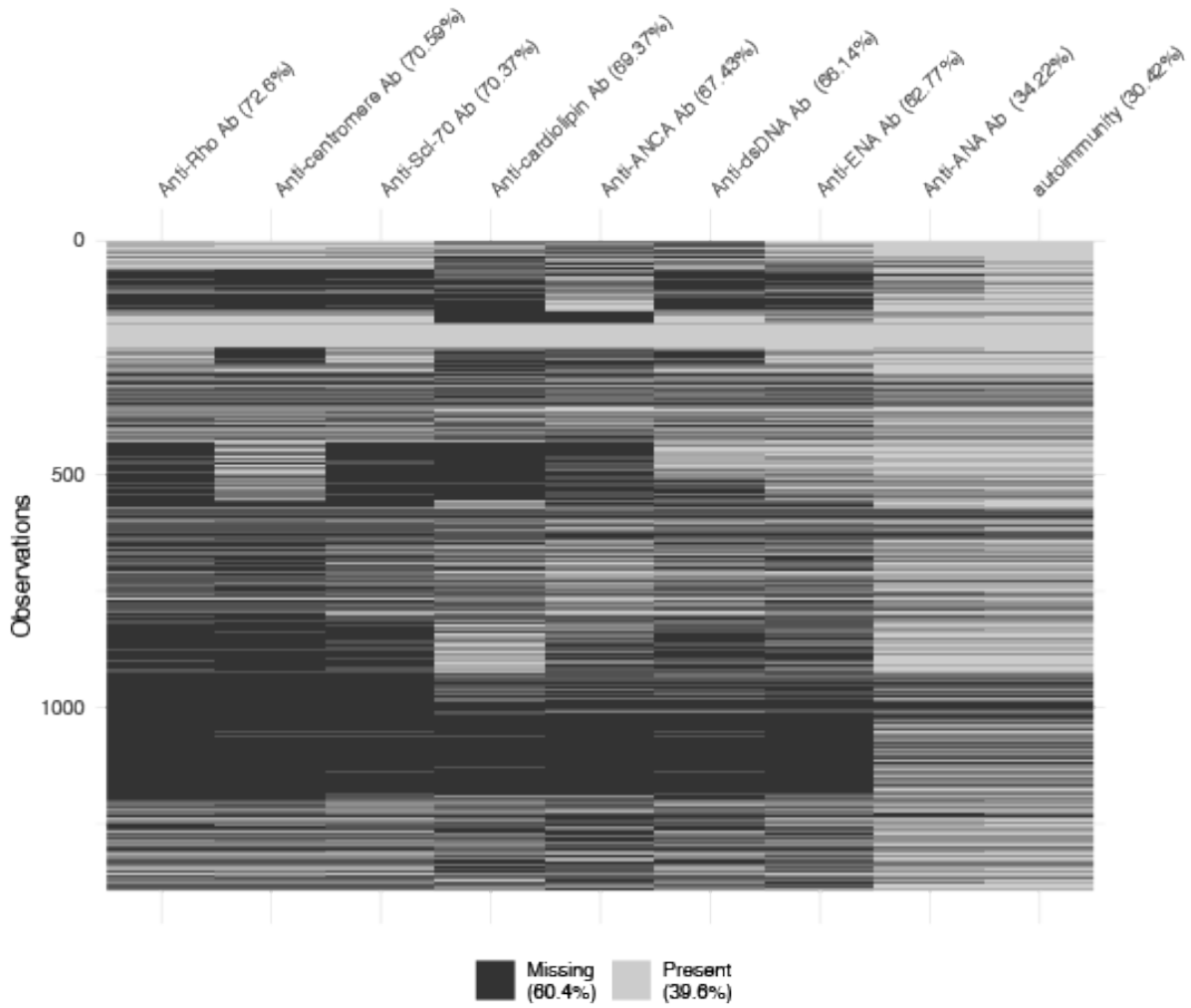


Fig. A.2 Heat map showing missingness in reported autoimmune screen.

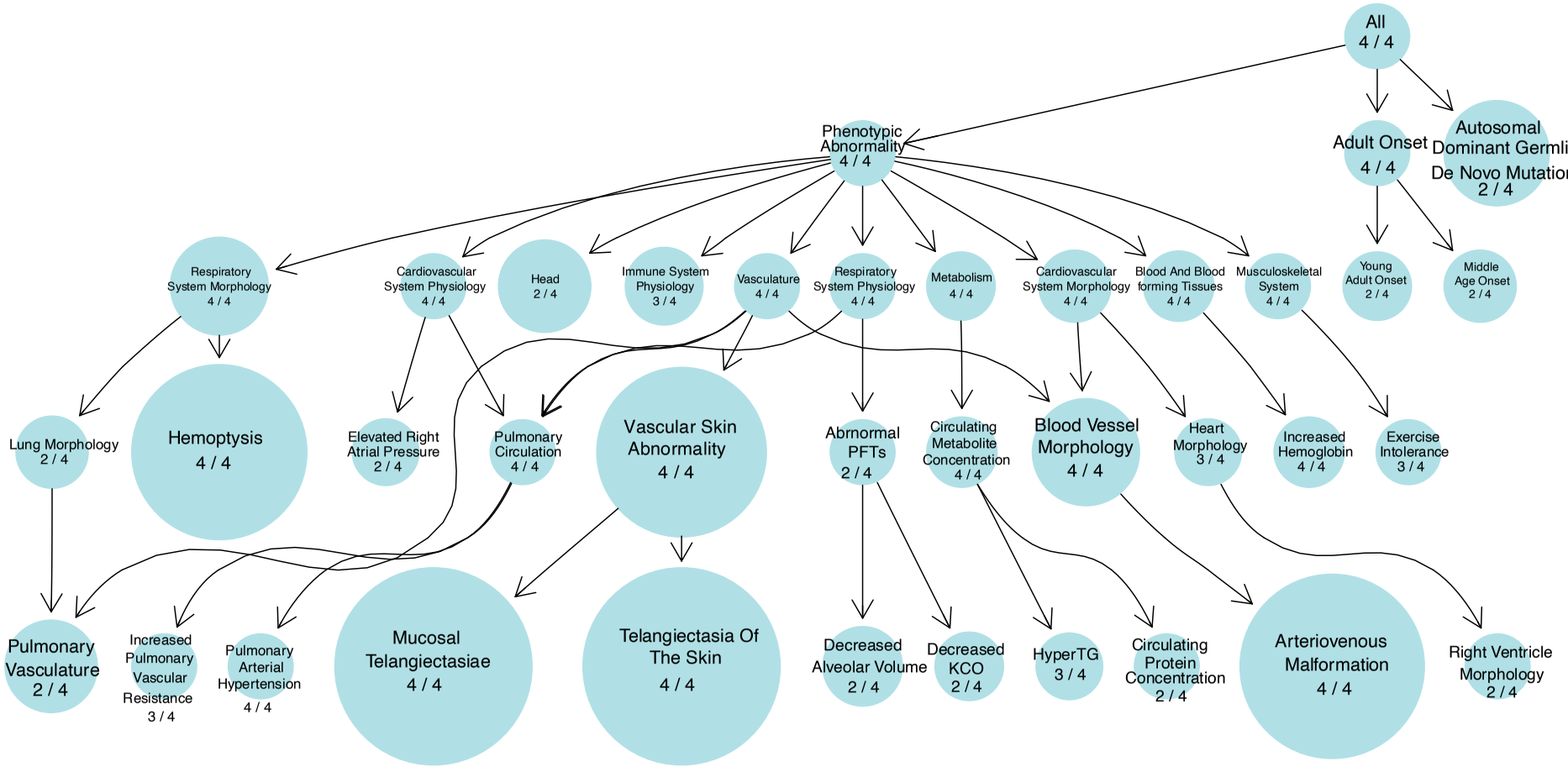


Fig. A.3 Cluster of patients with Hereditary Hemorrhagic Telangiectasia (HHT). Ontological similarity p-value< 0.001.

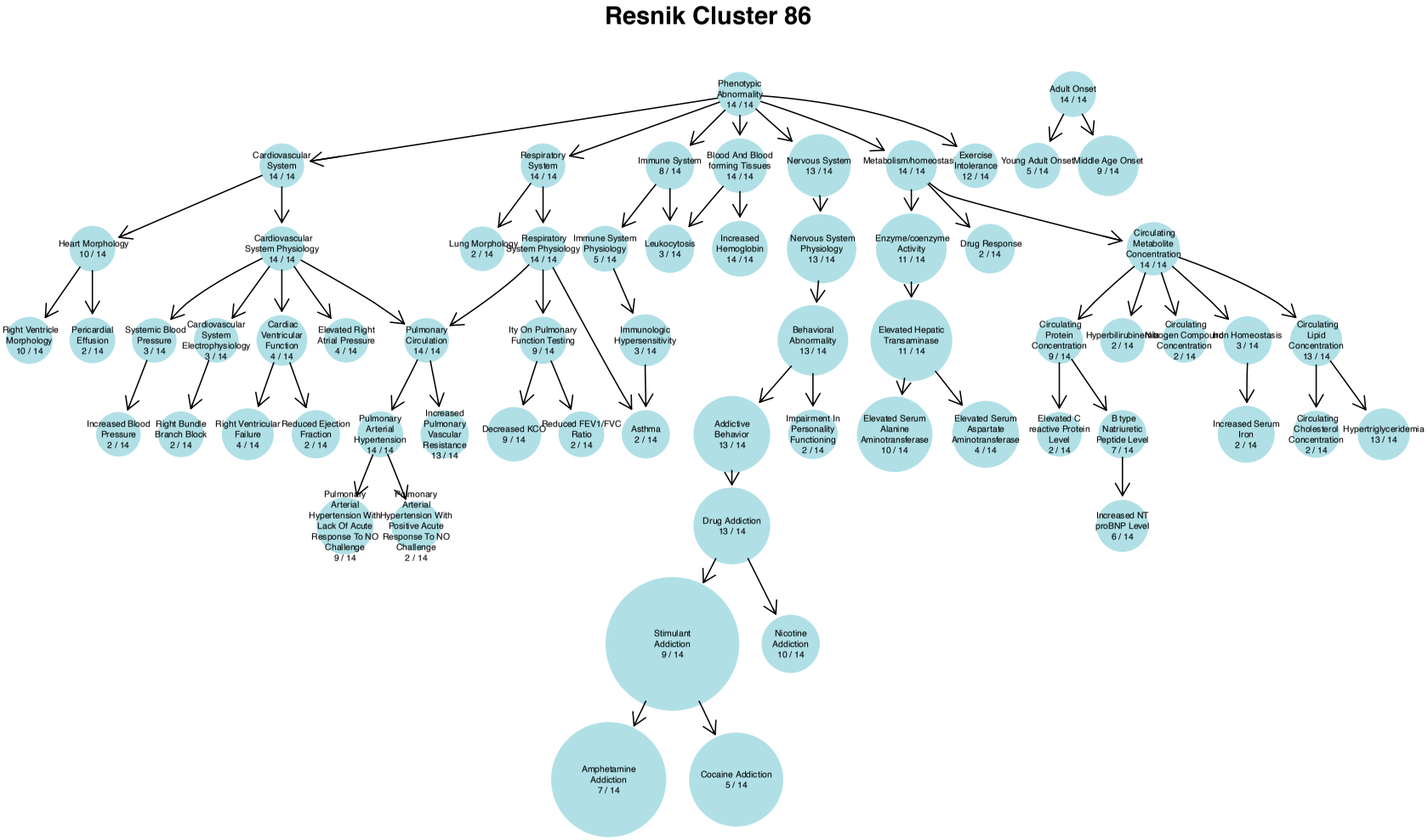


Fig. A.4 An exemplary cluster produced by affinity propagation clustering driven by substance addition phenotype.

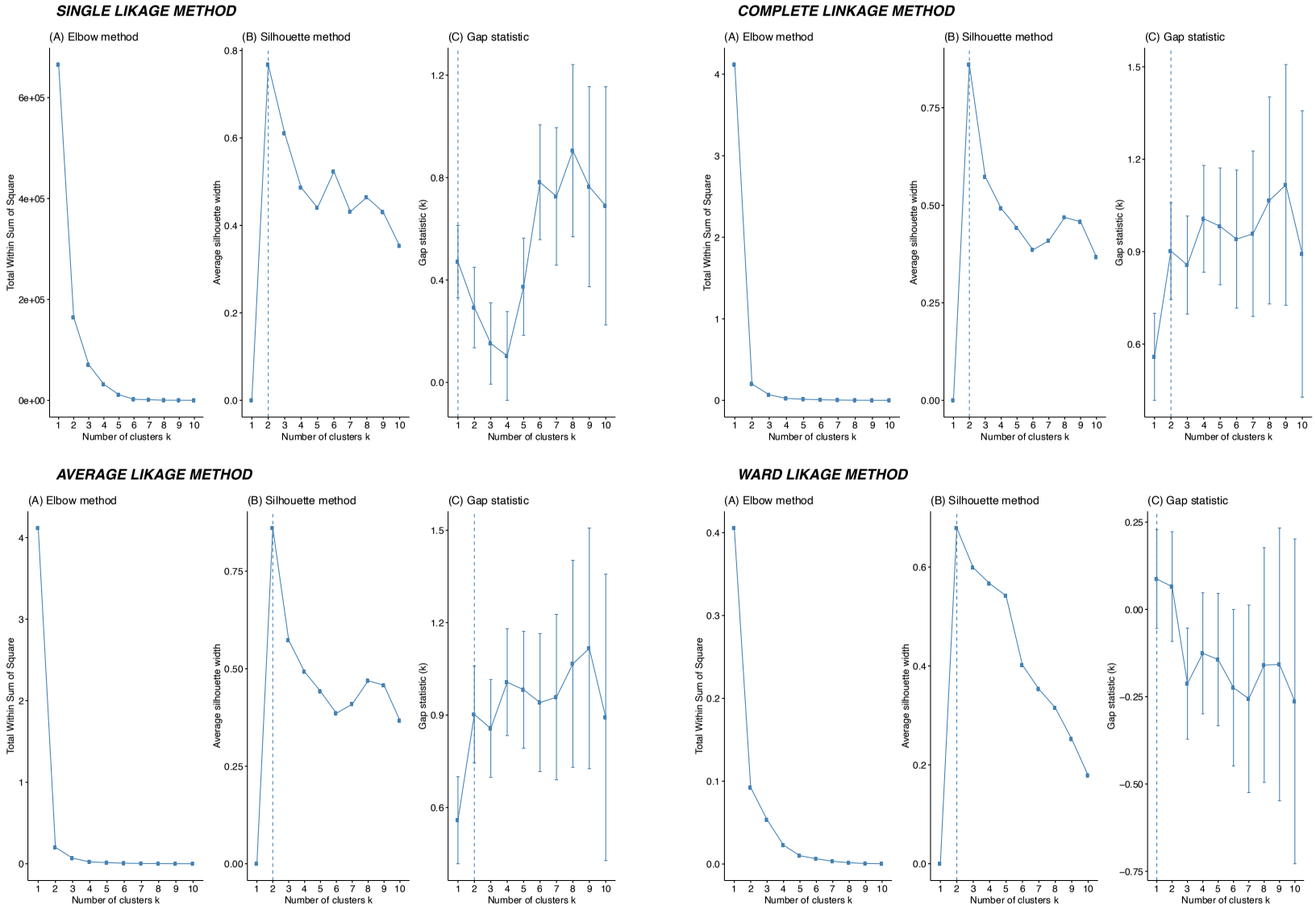
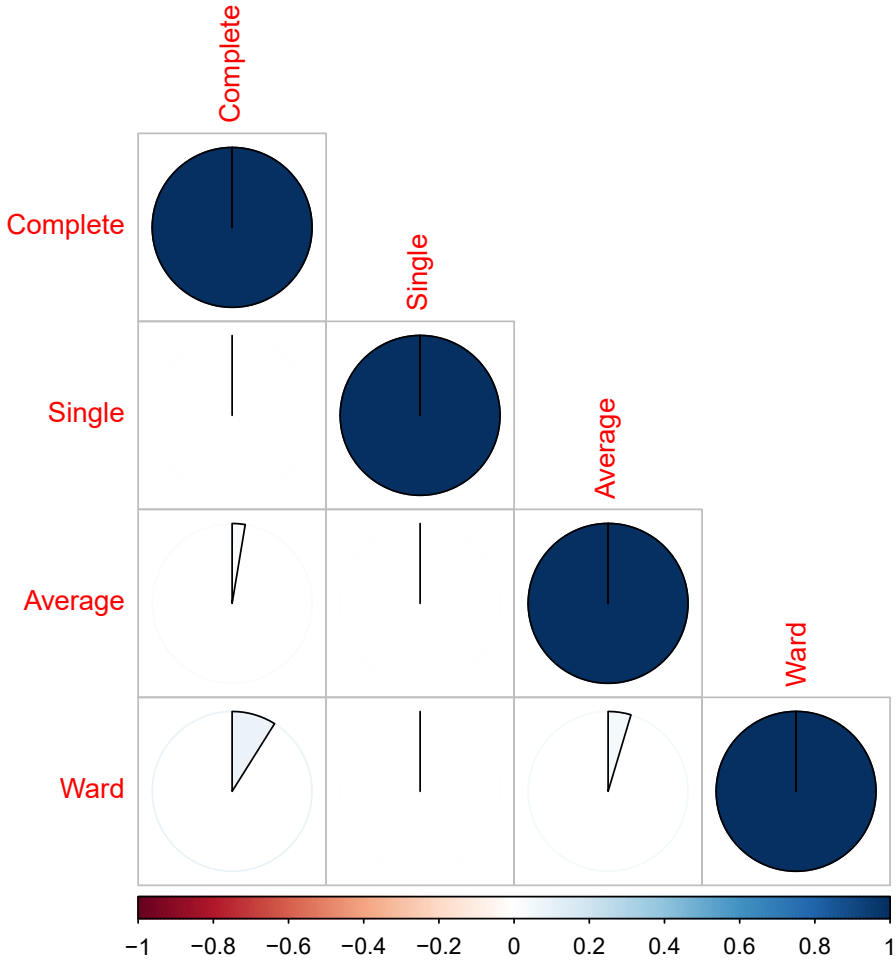


Fig. A.5 Visualisation of the results of direct and statistical methods for choosing the optimal number of clusters. Most methods pointed towards two clusters. Lin semantic similarity used.



	Complete	Single	Average	Ward
Complete	1.00	0.00	0.03	0.09
Single	0.00	1.00	-0.01	-0.01
Average	0.03	-0.01	1.00	0.05
Ward	0.09	-0.01	0.05	1.00

Fig. A.6 Comparison between dendrograms created by different linkage methods on Lin semantic similarity matrix. The plot shows no correlation between dendrograms.

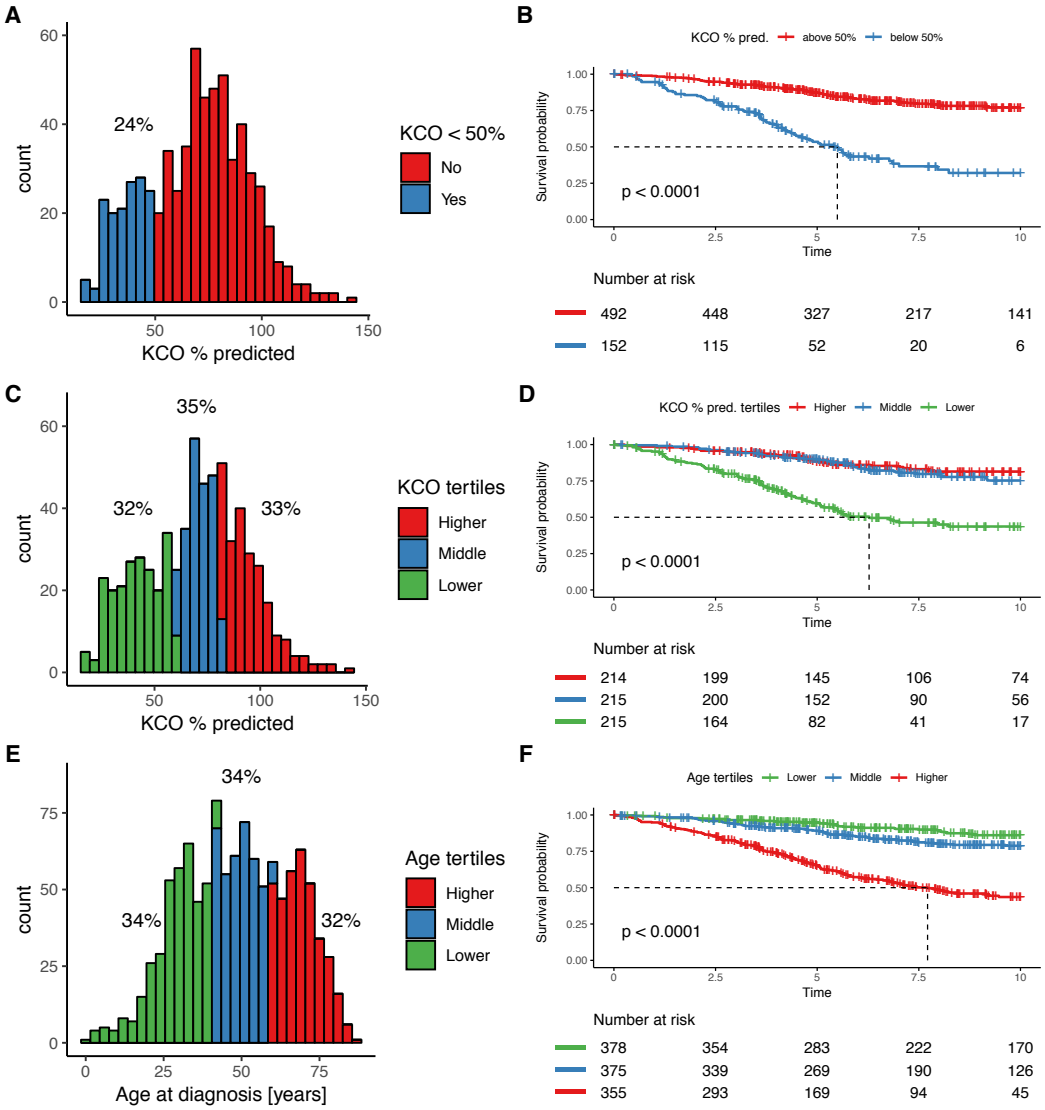


Fig. A.7 Characterisation and survival analysis of the cohort based on transfer coefficient for carbon monoxide (KCO) and age at diagnosis. Distribution of and Kaplan-Meier survival curves for KCO below and above the 50% predicted threshold (A, B), KCO tertiles (C D), and age tertiles (E, F).

Fig. A.8 Summary of large deletions identified in previously established disease genes. Deletions are indicated by light blue boxes. The protein-coding genes, annotated in the displayed region by Ensembl (GRCh37, version 75), are depicted in the bottom panels. The affected genomic regions, with the disease gene locus highlighted in red and the magnified view focusing on the gene loci, are shown for *BMPR2* (A, B), *GDF2* (C, D) and *TBX4* (E, F).

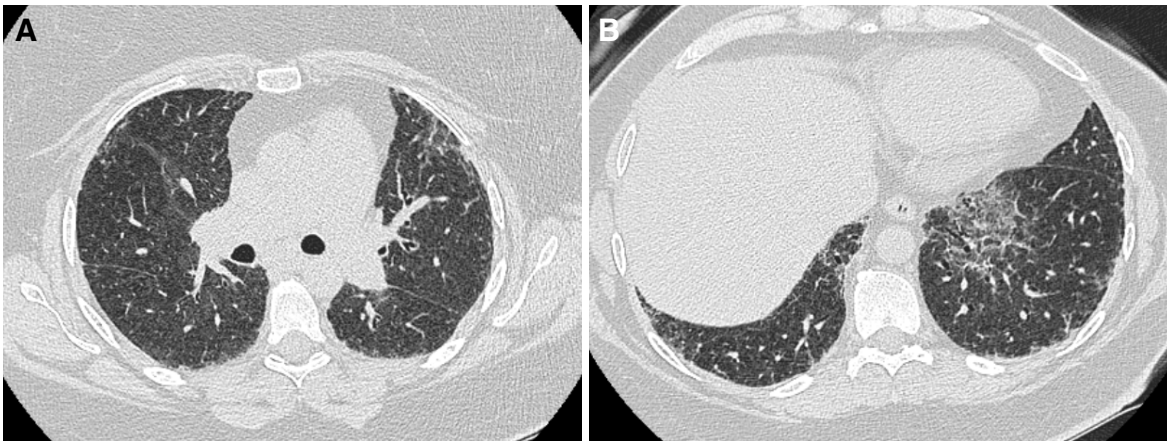
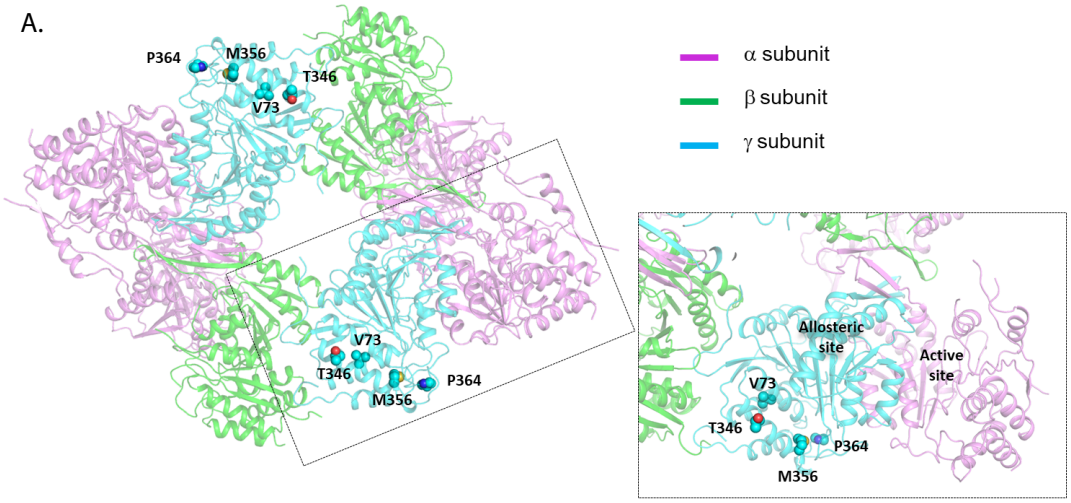
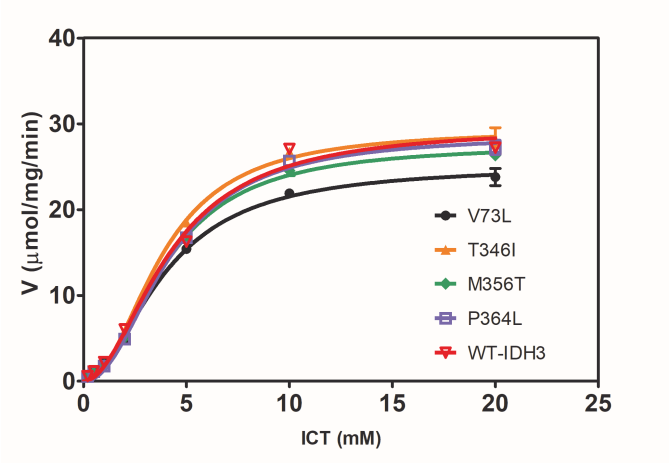


Fig. A.9 Chest computerised tomography (CT) scans of daughter of W000229 carrying the inherited nonsense variant in *KDR*. (A) The scan shows bibasal reticular ground glass changes with mild traction bronchiectasis. (B) Upper left lobe shows sub pleural reticular ground glass changes in keeping with interstitial fibrosis. Further ground glass changes are visible in the right lung.

A.



B.



C.

Enzyme	V73L	T346I	M356T	P364L	Wild-type
V_{max} (μmol/mg/min)	25.2±0.7	29.3±0.8	27.6±0.5	28.8±0.7	29.6±1.2
K_{cat} (s^{-1}) [a]	31.9±0.9	37.1±1.0	35.0±0.6	36.5±0.9	37.4±1.5
K_{cat} (s^{-1}) [b]	127.6±3.6	148.4±4.0	140.0±2.4	146.0±3.6	148.6±6.0
$S_{0.5,ICT}$ (mM)	3.93±0.33	3.81±0.32	3.99±0.27	4.17±0.38	4.16±0.35
Hill coefficient for ICT	1.91±0.14	2.12±0.16	2.08±0.11	2.12±0.14	1.95±0.20

The standard reaction solution (1 ml) consisted of 2 ng/ml enzyme, 33 mM Tris-acetate (pH 7.4), 0-20 mM ICT, 2 mM Mn^{2+}, and 3.2 mM NAD. [a] A molecular mass of 76 kDa was used to calculate the mole of the $\alpha\beta/\alpha\gamma$ heterodimer per mg of protein. [b] A molecular mass of 304 kDa was used to calculate the mole of the IDH3 holoenzyme per mg of protein. Courtesy of Professor Jianping Ding, PhD, Institute of Biochemistry and Cell Biology Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences

Fig. A.10 (A) IDH3G mutations superimposed on the structure of IDH3 holoenzyme, (B) Enzymatic activity assay of mutant and wild type holoenzymes, (C) Enzymatic activity assay results showing that the mutant holoenzymes exhibit comparable activity and kinetic properties as the wild-type IDH3 holoenzyme.

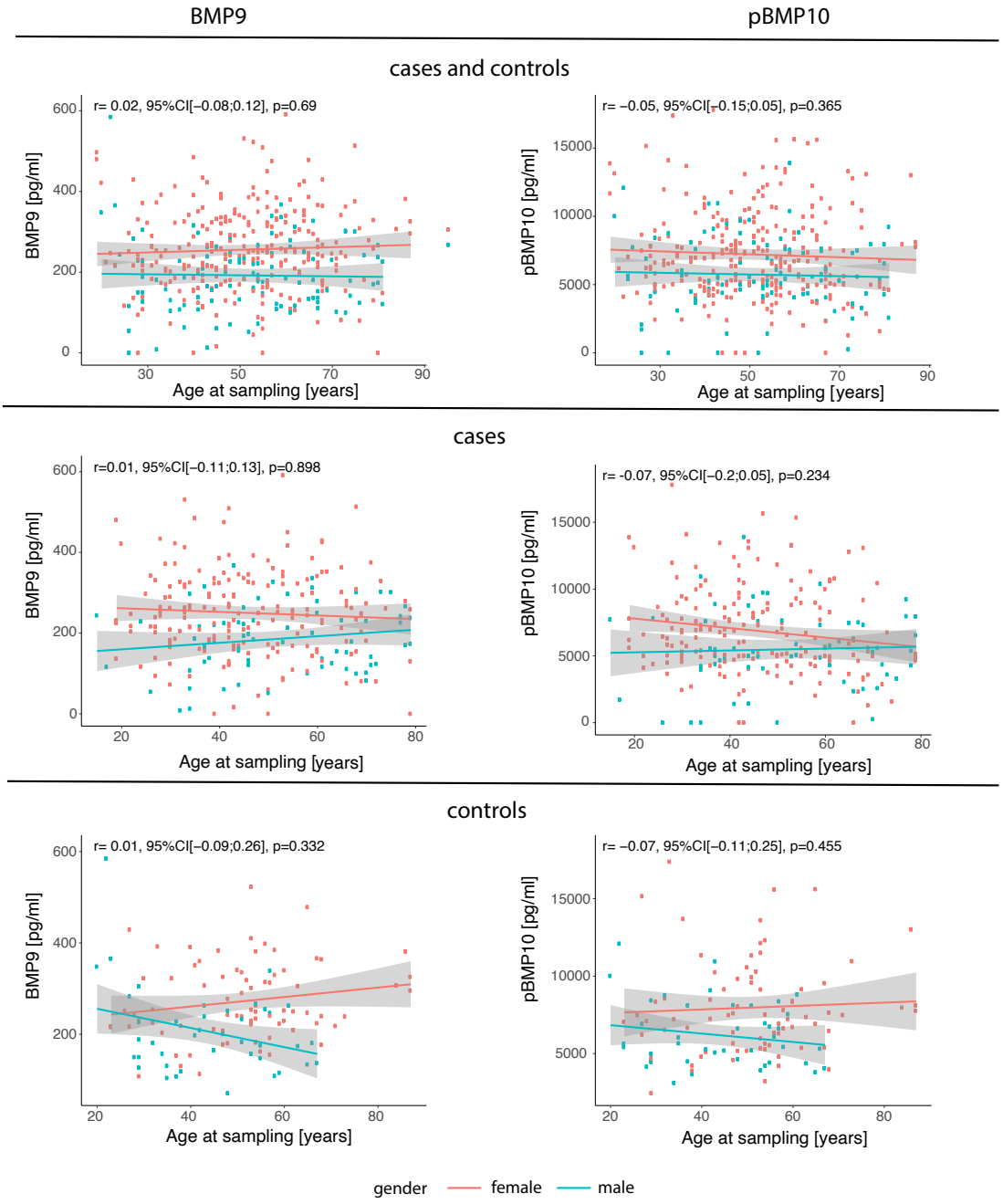


Fig. A.11 BMP9 and pBMP10 levels do not correlate with age for males or females in cases and controls. The levels of BMP9 and pBMP10 were assessed regarding any correlation to age for males or females using a Spearman test for combined cases and controls, cases only or controls only.

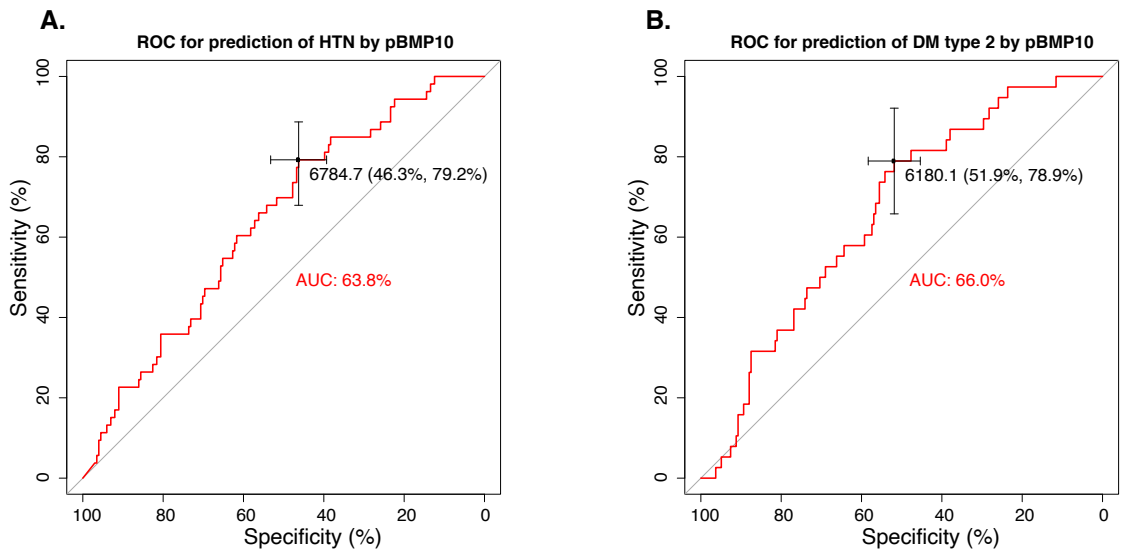


Fig. A.12 ROC curves for pBMP10 measurements as a predictor of systemic hypertension (HTN) and diabetes mellitus type 2 in I/HPAH patients. ROC analysis for pBMP10 levels as a predictor of (A) systemic hypertension (HTN) and (B) diabetes mellitus type 2. The diagonal line indicates the line of no discrimination.

Appendix B

Appendix Tables

Table B.1 Exclusion criteria related to pulmonary function tests in pivotal randomised controlled trials leading to drug approval in PAH.

Study	Medication	FEV1	FVC	FEV1/FVC	TLC	DLCO
Barst, 1996	Epoprostenol	NA	NA	NA	NA	NA
Badesh, 2000	Epoprostenol	Pulmonary function tests or high-resolution computed tomography scanning showing more than mild interstitial lung disease				
AIR, 2002	Iloprost	< 2SD below the mean	< 50%			0
BREATHE-1, 2002	Bosentan	NA***	NA***	NA***	NA***	NA***
Simonneau, 2002	Treprostinil	Significant parenchymal pulmonary disease as evidenced by pulmonary function tests or high-resolution CT scan.				
BREATHE-2, 2004	Bosentan	-	-	-	<60%*	-
STRIDE-1, 2004	Sitaxsentan	Significant parenchymal lung disease				
SUPER-1, 2005	Sildenafil	NA***	NA***	NA***	NA***	NA***
BREATHE-5, 2006	Bosentan	< 70% AND FEV1/FVC < 60%	-	-	< 70%	-
COMBI, 2006	Iloprost	< 50%	-	-	< 70%	-
STEP, 2006	Iloprost	-	-	< 50%	< 60%	-
STRIDE-2, 2006	Sitaxsentan	Significant parenchymal lung disease				
ARIES 1/2, 2008	Ambrisentan	NA***	NA***	NA***	NA***	NA***
EARLY, 2008	Bosentan	NA***	NA***	NA***	NA***	NA***
PACES, 2008	Sildenafil	NA***	NA***	NA***	NA***	NA***
PHIRST, 2009	Tadalafil	NA***	NA***	NA***	NA***	NA***
TRIUMPH, 2010	Treprostinil	NA***	NA***	NA***	NA***	NA***
FREEDOM-C, 2012	Treprostinil	NA***	NA***	NA***	NA***	NA***
FREEDOM-C2, 2013	Treprostinil	-	-	< 50%	< 60%**	0
PATENT, 2013	Riociguat	< 60%	-	-	< 70%	-
SERAPHIN, 2013	Macitentan	< 65% AND FEV1/FVC < 70%	-	-	< 60%	-
AMBITION, 2015	Ambrisentan/ tadalafil	< 55%	-	-	< 60%	-
COMPASS-2, 2015	Bosentan	-	-	< 50%	< 60%	-
GRIPHON, 2015	Selexipag	< 65%	-	< 70%	< 60%	-

Abbreviations: mPAP: mean pulmonary arterial pressure; PAWP: pulmonary arterial wedge pressure; PVR: pulmonary vascular resistance; WU: Wood Units. #: group 1: PAH; group 2: PH due to left heart disease; group 3: PH due to lung diseases and/or hypoxia; group 4:PH due to pulmonary artery obstructions; group 5: PH with unclear and/or multifactorial mechanisms.

* Or high resolution computed tomography scan total score > 2.

** Or TLC 60-70% AND diffuse parenchymal disease on CT scan

*** No details about pulmonary function tests within the main manuscript, supplemental material or http://www.clinicaltrial.gov

Table B.2 Landmark forward genetics studies in Group 1 PAH.

Study	Genes	Study Design	Sample	Ethnicity	Method	Reference Genome
Lane et al. 2000 [241]	*BMPR2*	Case-level data	Cases: n = 8 PPH kindreds for candidate gene mutational analysis	Not stated	TS	H.sapiens mRNA for BMPR-II: Genbank Z48923
Thomson 2000 [521]	*BMPR2*	Case-level data	Cases: n = 50 PPH	Not stated	TS	Not stated
Trembath et al. 2001 [531]	*ACVRL1*	Case-level data	Cases: 5 kindreds plus 1 individual patient with HHT, including n = 10 cases with PH	Not stated	TS	Not stated
Chaouat 2004 [71]	*ENG*	Case-level data	Case: n = 1 HHT, PPH with history of anorexigen use	Not stated	TS	Not stated
Harrison et al. 2005 [202]	*ACVRL1*, ENG	Case-level data	Cases: n = 18 I/APAH	Not stated	TS	Not stated
Shintani et al. 2009 [467]	*SMAD9* (*SMAD8*)	Case-level data	Cases: n = 23 IPAH	Japanese	TS	Not stated
Nasim et al. 2011 [371]	*SMAD1*, *SMAD4*, *SMAD9*	Case-level data	Cases: n = 324 IPAH/APAH/CTEPH; Controls: n = 1,584	European and Japanese	TS	Not stated
Austin et al. 2012 [25]	*CAV1*	Case-level data	Cases: 3-generation family, 6 with PAH; Additional cohort: n = 260 unrelated I/HPAH cases; Controls: n = 1000	European	WES	GRCh37
Ma et al. 2013 [322]	*KCNK3*	Case-level data	Cases: Family in which multiple members had PAH	Not stated	WES	GRCh37
Kerstjens-Frederikse et al. 2013 [262]	*TBX4*	Case-level data	Cases: n = 20 childhood-onset I/HPAH; n = 49 adult-onset I/HPAH; n = 23 SPS	Not stated	TS	Not stated
Eyries et al. 2014 [127]	*EIF2AK4*	Case-level data	Cases: n = 13 PVOD families	Not stated	WES	GRCh37
Best et al. 2017 [46]	*EIF2AK4*	Case-level data	Cases: n = 81 I/HPAH	Not stated	TS	Not stated
Hadinnapola et al. 2017 [194]	*EIF2AK4*	Case-control data	Cases: n = 880 I/FPAH, PVOD/PCH; Controls: n = 7134 non-PAH controls and their relatives recruited to NBR	European: 84.6%	WGS	GRCh37
Gräf et al. 2018 [179]	*GDF2*, *SOX17*, *ATP13A3*, *AQP1*	Case-control data	Cases: n = 1,048 I/F/PAH, PVOD/PCH; Controls: n = 7,979 non-PAH controls and their relatives recruited to NBR	European: 84.6%	WGS	GRCh37
Zhu et al. 2018 [599]	*SOX17*	Case-control data	Cases: n = 256 I/FPAH,CHD-PAH; Additional cohort: n = 413 I/FPAH screened for rare variants in *SOX17*; Controls: n = 7509 gnomAD	Not stated	WES	GRCh37

Table B.2 Landmark forward genetics studies in Group 1 PAH.

Study	Genes	Study Design	Sample	Ethnicity	Method	Reference Genome
Hiraide et al. 2018 [212]	*SOX17*	Case-level data	Cases: n = 12 IPAH and 12 family members; Additional cohort: n = 128 I/HPAH screened for *SOX17* mutations	Japanese: 100%	WES	Not stated
Bohnen et al. 2018 [51]	*ABCC8*	Case-control data	Cases: n = 913; Controls: n = 33,369 European adults from ExAC	n = 49,630 Europeans from the Regeneron-Geisinger DiscovEHR study	WES, WGS	GRCh37
Wang et al. 2019 [560]	*GDF2*	Case-control data	Cases: n = 331 IPAH; Controls: n = 10,508 from available reference data sets	East Asian: 100%	WES, WGS	GRCh37
Eyries et al. 2019 [128]	*BMP10*	Case-level data	Cases: n = 268 I/HPAH, PVOD/PCH	European: > 90%	TS	GRCh37
Hodgson et al. 2020 [215]	*BMP10*	Case-level data	Cases: n = 1,048 I/FPAH, PVOD/PCH	European: 84.6%	WGS	GRCh37
Zhu et al. 2019 [598]	*KLK1*, *GCCX*	Case-control data	Cases: n = 2,572 Group 1 PAH; Controls: n = 12,771	European: 72%	WES	GRCh38
Rhodes et al. 2019 [436]	*HLA-DPA1/DPB1*, *SOX17* enhancer	Case-control data	Cases: n = 2,085; Controls: n = 9,659	European: 100%	WGS	GRCh37
Swietlik et al. 2020 [503]	*KDR*	Case-control data	Cases: n = 1,122 PAH; Controls: n = 11,889 non-PAH NBR	European: 84%	WGS	GRCh37
Eyries et al. 2020 [126]	*KDR*	Case-level data	Cases: n = 311 unrelated PAH	Not stated	TS	Not stated

Abbreviations: *BMPR2*: Bone morphogenic protein receptor type 2; *ENG*: Endoglin; *ACVRL1*: Activin A Receptor Like Type 1; *SMAD*: SMAD Family Member; *CAV1*: Caveolin 1; *KCNK3*: Potassium Two Pore Domain Channel Subfamily K Member 3; *TBX4*: T-Box Transcription Factor 4; *EIF2AK4*: Eukaryotic Translation Initiation Factor 2 α Kinase 4; *GDF2*: Growth Differentiation Factor 2; *SOX17*: SRY-Box Transcription Factor 17; ATP13A3: ATPase 13A3; AQP1: Aquaporin 1; ABCC8: ATP Binding Cassette Subfamily C Member 8; *BMP10*: Bone Morphogenetic Protein 10; *KLK1*: Kallikrein 1; *GCCX*: γ-Glutamyl Carboxylase; *KDR*: Kinase insert domain receptor; *TET2*: Tet Methylcytosine Dioxygenase 2; *FBLN2*: Fibulin 2; *PDGFD*: Platelet-Derived Growth Factor D; TS: targeted sequencing; WES: whole-exome sequencing; HHT: hereditary hemorrhagic telangiectasia; PPH: primary pulmonary hypertension; I/H/F/APAH: idiopathic/hereditary/familial/associated pulmonary arterial hypertension; PVOD: Pulmonary veno-occlusive disease, PCH: Pulmonary capillary haemangiomatosis; CTEPH: chronic thromboembolic pulmonary hypertension; SPS: small patella syndrome; NBR: NIHR BioResource Rare Disease Study; SPARK: Simons Foundation Powering Autism Research for Knowledge; gnomAD: The Genome Aggregation Database

Table B.3 Supporting evidence for the role of risk genes in PAH pathogenesis.

	Forward Genetics				Reverse Genetics	
Gene	**Case-Control Data**	**Case-Level Data**	**Segregation Data**	**Functional Aberration**	**Disease Model**	**Rescue**
			MOI: Autosomal Dominant			
BMPR2	(+) (Gräf et al. 2018; Zhu et al. 2019)	(+) (International PPH Consortium et al. 2000; Eyries et al. 2019)	(+) (International PPH Consortium et al. 2000; Larkin et al. 2012)	(+) (Beppu et al. 2004; Yang et al. 2005; Hong et al. 2008; West et al. 2008; Gangopahyay et al. 2011; Hemnes et al. 2014; Ranchoux et al. 2015; Caruso et al. 2017)	Animal: (+) (Beppu et al. 2004; West et al. 2008; Hemnes et al. 2014; Ranchoux et al. 2015; Caruso et al. 2017; Tian et al. 2019) Cell culture: (+) (Yang et al. 2005; Ranchoux et al. 2015; Caruso et al. 2017)	(+) (Hemnes et al. 2014; Ranchoux et al. 2015; Caruso et al. 2017; Tian et al. 2019) (+)
ACVRL1	(+) (Gräf et al. 2018)	(+) (Harrison et al. 2005; Zhu et al. 2019; Eyries et al. 2019; Trembath et al. 2001)	(+) (Trembath et al. 2001)			
ENG	(-) (Gräf et al. 2018; Zhu et al. 2019)	(+) (Chaouat 2004; Gräf et al. 2018; Zhu et al. 2018)	(+) (Chaouat 2004)		Animal: (-) (Gore et al. 2014)	
SMAD9	(-) (Gräf et al. 2018; Zhu et al. 2019)	(+) (Shintani et al. 2009; Gräf et al. 2018; Zhu et al. 2018; Zhu et al. 2019; Eyries et al. 2019)		(+) (Drake et al. 2011; Nasim et al. 2011; Shintani et al. 2009)	Animal: (+) (Huang et al. 2009) Cell culture: (+) (Nasim et al. 2011; Drake et al. 2011; Shintani et al. 2009)	(+) (Drake et al. 2011; Drake et al. 2013)
SMAD1	(-) (Gräf et al. 2018; Zhu et al. 2019)	(+) (Nasim et al. 2011), (-) (Gräf et al. 2018; Zhu et al. 2019)		(+) (Nasim et al. 2011; Han et al. 2013)	Animal: (+) (Han et al. 2013) Cell culture: (+) (Nasim et al. 2011)	

Table B.3 Supporting evidence for the role of risk genes in PAH pathogenesis.

	Forward Genetics			Reverse Genetics		
Gene	**Case-Control Data**	**Case-Level Data**	**Segregation Data**	**Functional Aberration**	**Disease Model**	**Rescue**
SMAD4	(-) (Gräf et al. 2018; Zhu et al. 2019)	(+) (Nasim et al. 2011; Zhu et al. 2019), (-) (Gräf et al. 2018)		(+) (Nasim et al. 2011), (-) (Nasim et al. 2011)	Cell culture: (+) (Nasim et al. 2011)	
CAV1	(+) (Austin et al. 2012)	(+) (Zhu et al. 2018; Zhu et al. 2019), (-) (Gräf et al. 2018)		(+) (Copeland et al. 2017)	Animal: (+) (Zhao et al. 2002; Drab et al. 2001) Cell culture:(+) (Copeland et al. 2017)	(+) (Copeland et al. 2017)
TBX4	(+) (Gräf et al. 2018; Zhu et al. 2018)	(+) (Kerstjens-Frederikse et al. 2013; Levy et al. 2016; Zhu et al. 2018; Eyries et al. 2019; Thoré et al. 2020)	(+) (Jansen et al. 2020; Hernandez-Gonzalez et al. 2020)			
KCNK3		(+) (Ma et al. 2013; Gräf et al. 2018; Zhu et al. 2019)	(+) (Ma et al. 2013)	(+) (Ma et al. 2013; Bohnen et al. 2017)	Animal: (+) (Lambert et al. 2019) Cell culture: (+) (Ma et al. 2013; Bohnen et al. 2017)	(+) (Ma et al. 2013)
ATP13A3	(+) (Gräf et al. 2018)			(+) (Legchenko et al. 2020; Liu et al. 2018)	Animal: (+) (Legchenko et al. 2020; Liu et al. 2018)	
AQP1	(+) (Gräf et al. 2018)			(+) (Yun et al. 2017; Saadoun et al. 2005)()(Yun et al. 2017)	Animal: (+) (Ma et al. 1998; Yun et al. 2017) Cell culture:(+) (Saadoun et al. 2005)	(+) (Yun et al. 2017)
GDF2	(+) (Gräf et al. 2018; Zhu et al. 2019; Wang et al. 2019)	(+) (Eyries et al. 2019)	(+) (Hodgson et al. 2020)	(+) (Hodgson et al. 2020; Wang et al. 2019)	Animal: (-) (Tu et al. 2019) Cell culture: (+) (Hodgson et al. 2020; Wang et al. 2019)	

Table B.3 Supporting evidence for the role of risk genes in PAH pathogenesis.

	Forward Genetics				Reverse Genetics	
Gene	**Case-Control Data**	**Case-Level Data**	**Segregation Data**	**Functional Aberration**	**Disease Model**	**Rescue**
SOX17	(+) (Zhu et al. 2018; Gräf et al. 2018)	(+) (Hiraide et al. 2018)		(+) (Lange et al. 2014; Shih et al. 2020)	Animal: (+) (Lange et al. 2014) Cell culture: (+) (Shih et al. 2020)	
ABCC8	(+) (Bohnen et al. 2018)		(+) (Bohnen et al. 2018)	(+) (Bohnen et al. 2018)	Cell culture: (+) (Bohnen et al. 2018)	(+) (Bohnen et al. 2018)
BMP10		(+) (Eyries et al. 2019; Hodgson et al. 2020)				
GGCX	(+) (Zhu et al. 2019)					
KLK1	(+) (Zhu et al. 2019)					
KDR	(+) (Swietlik et al. 2020)		(+) (Eyries et al. 2020; Swietlik et al. 2020)	(+) (Winter et al. 2020)	Animal: (+) (Winter et al. 2020)	
FBLN2	(+) (Zhu et al. 2020)					
PDGFD	(+) (Zhu et al. 2020)					
TET2	(+) (Potus et al. 2020)			(+) (Potus et al. 2020)	Animal: (+) (Potus et al. 2020)	(+) (Potus et al. 2020)
BMPR1A		(+) (Zhu et al. 2019; Zhu et al. 2018)		(+) (Lee et al. 2018; Lee Heon-Woo et al. 2019; Du et al. 2003; El-Bizri et al. 2008)	Animal: (+) (Lee et al. 2018; Lee Heon-Woo et al. 2019; El-Bizri et al. 2008)	(+) (Lee et al. 2018)
BMPR1B		(+) (Gräf et al. 2018; Zhu et al. 2018; Zhu et al. 2019)		(+) (Chida et al. 2012)	Cell culture: (+) (Chida et al. 2012)	
TOPBP1		(+) (Gräf et al. 2018)				
THBS1				(+) (Yu et al. 2009)	Cell culture: (+) (Yu et al. 2009)	(+) (Yu et al. 2009)
KCNA5		(+) (Wang et al. 2014)		(+) (Remillard et al. 2007)	Cell culture: (+) (Remillard et al. 2007)	(+) (Remillard et al. 2007)
MOI: Autosomal Recessive						

Table B.3 Supporting evidence for the role of risk genes in PAH pathogenesis.

	Forward Genetics			Reverse Genetics		
Gene	**Case-Control Data**	**Case-Level Data**	**Segregation Data**	**Functional Aberration**	**Disease Model**	**Rescue**
EIF2AK4	(+) (Hadinnapola et al. 2017)	(+) (Eyries et al. 2019)	(+) (Eyries et al. 2014; Best et al. 2017)	(+) (Eyries et al. 2014)	Animal: (+) (Zhang et al. 2002)	
			Common Variation			
enhancer near SOX17	(+) (Rhodes et al. 2019)		(+) (Rhodes et al. 2019)	(+) (Rhodes et al. 2019)		
locus within HLA-DPA1/DPB1	(+) (Rhodes et al. 2019)					
CBLN2	(+) (Germain et al. 2013)					
	(-) (Rhodes et al. 2019)					
SIRT3		(+) (Paulin et al. 2014)	(+) (Paulin et al. 2014)	(+) (Paulin et al. 2014)	Animal: (+) (Michelakis et al. 2017; Paulin et al. 2014)	
UCP2				(+) (Dromparis et al. 2013)	Animal: (+) (Zhang et al. 2001; Michelakis et al. 2017)	
EDN1		(+) (Baloira Villar et al. 2019)	(+) (Giaid et al. 1993)			
AGTR1	(+) (Chung et al. 2009)					
TOPBP1		(+) (de Jesus Perez et al. 2014)		(+) (de Jesus Perez et al. 2014)	Cell culture: (+) (de Jesus Perez et al. 2014)	
Endostatin	(+) (Damico et al. 2015)					
TRPC6				(+) (Maloney et al. 2012)	Cell culture: (+) (Maloney et al. 2012)	

Abbreviations: (+) indicates that the paper provides information in favour of the role of the given gene in the pathogenesis of PAH, () indicates that the paper does not provide support for the role of the given gene in the pathogenesis of PAH. *BMPR2*: Bone morphogenic protein receptor type 2; *ENG*: Endoglin; *ACVRL1*: Activin A Receptor Like Type 1; *SMAD*: SMAD Family Member; *CAV1*: Caveolin 1; *KCNK3*: Potassium Two Pore Domain Channel Subfamily K Member 3; *TBX4*: T-Box Transcription Factor 4; *EIF2AK4*: Eukaryotic Translation Initiation Factor 2 α Kinase 4; *GDF2*: Growth Differentiation Factor 2; *SOX17*: SRY-Box Transcription Factor 17; ATP13A3: ATPase 13A3; AQP1: Aquaporin 1; ABCC8: ATP Binding Cassette Subfamily C Member 8; *BMP10*: Bone Morphogenetic Protein 10; *KLK1*: Kallikrein 1; *GCCX*: γ-Glutamyl Carboxylase; *KDR*: Kinase insert domain receptor; *TET2*: Tet Methylcytosine Dioxygenase 2; *FBLN2*: Fibulin 2; *PDGFD*: Platelet-Derived Growth Factor D; *SIRT3*: Sirtuin 3; *UCP2*: Uncoupling Protein 2; *EDN1*: Endothelin 1; *AGTR1*: Angiotensin II Receptor Type 1; textitTOPBP1: DNA Topoisomerase II Binding Protein 1; TRPC6:Transient Receptor Potential Cation Channel Subfamily C Member 6

Table B.4 Eligibility criteria for A. The NIHR Bioresource Rare Disease (NBR) study PAH domain B. The Cohort PAH study.

	A. The NHIR-BR Rare Disease (NBR) PAH domain study	**B. The Cohort study of I/HPAH**
Inclusion Criteria-Patient (16 and over)	• Participant is willing and able to give informed consent for participation in the study • Male or Female, aged at least 16 years in UK, Male or Female, aged at least 18 years in Ireland • Diagnosed with idiopathic, anorexigen-induced or heritable PAH (PVOD/PCH)	
Exclusion Criteria-Patient (16 and over)	• Patient is unable to give informed consent • Not suffering from idiopathic, anorexigen-induced, or heritable PAH (PVOD/PCH) • Patient is unable to give informed consent • Exclusion Criteria-Patient (16 and over) • Not suffering from idiopathic, anorexigen-induced, or heritable PAH (PVOD/PCH)	
Inclusion criteria - Paediatric patient (1 month-15 years)	• Male or female aged 1 month to 15 years in UK • Diagnosed with idiopathic, anorexigen-induced or heritable PAH (PVOD/PCH) • Patient is willing and able to give assent (6-15 year olds) and/or Guardian is willing and able to give consent for their dependant	
Exclusion criteria - Paediatric patient (1 month-15 years)	• Guardian is unable to give informed consent. • Patient Not suffering from idiopathic, anorexigen-induced, or heritable PAH (PVOD/PCH)	
Inclusion Criteria – Relative (1 month – 80 years)	• Participant is willing and able to give informed consent for participation in the study (16-80 years). • Participant is willing and able to give assent for participation in the study (6-15 years old) and/or guardian willing and able to give consent for their dependant (1 month-15 years old). • Male or Female, aged 1 month -80 years in UK, Male or Female, aged 18-80 years in Ireland • Has a family member diagnosed with idiopathic, anorexigen-induced, or heritable PAH (including PVOD/PCH)	
Exclusion Criteria – Relative (1 month – 80 years)	• Participant and/or guardian is unable to give informed consent.	
Inclusion criteria - Controls		• Participant is willing and able to give informed consent for participation in the study. • Self-reported to be healthy • Age range 16-75 years
Exclusion Criteria - Controls		• Participant is unable to give informed consent • Pregnant • Has taken antibiotics in the last month • Has a history of: Connective tissue disease, Liver disease, Heart disease, Hypertension, Diabetes, Lung disease, Inflammatory bowel disease • On any regular medication for the above conditions • Previous drug history of cocaine or methamphetamine

Table B.5 Summary of electronic case report forms (eCRFs) constructed to capture phenotypic information

ID capture
Demographics
Functional class
Clinical features by history
Clinical features by examination
Risk factors
Haemodynamics
Echocardiography
Electrocardiogram
Lung function
Associated Diseases
Clinical blood tests
Survival
Arterial blood gases
Imaging
Exercise performance
Body system
Drug treatment history (PAH)
Drug treatment history (other)
Family history
Epidemiology questionnaire

Abbreviations: ID: indentificator; PAH: Pulmonary Arterial Hypertension

Table B.6 ICD-10 codes used to calculate Charlson and Elixhauser scores.

The ICD-10 codes used to compute the Charlson comorbidity index:	**The ICD-10 codes used to compute the Elixhauser comorbidity index:**
• Myocardial infarction: I21.x, I22.x, I25.2 • Congestive heart failure: I09.9, I11.0, I13.0, I13.2, I25.5, I42.0, I42.5 - I42.9, I43.x, I50.x, P29.0 • Peripheral vascular disorders: I70.x, I71.x, I73.1, I73.8, I73.9, I77.1, I79.0, I79.2, K55.1, K55.8, K55.9, Z95.8, Z95.9 • Cerebrovascular disease: G45.x, G46.x, H34.0, I60.x - I69.x • Dementia: F00.x - F03.x, F05.1, G30.x, G31.1 • Chronic pulmonary disease: I27.8, I27.9, J40.x - J47.x, J60.x - J67.x, J68.4, J70.1, J70.3 • Rheumatic disease: M05.x, M06.x, M31.5, M32.x - M34.x, M35.1, M35.3, M36.0 • Peptic ulcer disease: K25.x - K28.x • Mild liver disease: B18.x, K70.0 - K70.3, K70.9, K71.3 - K71.5, K71.7, K73.x, K74.x, K76.0, K76.2 - K76.4, K76.8, K76.9, Z94.4 • Diabetes without chronic complication: E10.0, E10.1, E10.6, E10.8, E10.9, E11.0, E11.1, E11.6, E11.8, E11.9, E12.0, E12.1, E12.6, E12.8, E12.9, E13.0, E13.1, E13.6, E13.8, E13.9, E14.0, E14.1, E14.6, E14.8, E14.9 • Diabetes with chronic complication: E10.2 - E10.5, E10.7, E11.2 - E11.5, E11.7, E12.2 - E12.5, E12.7, E13.2 - E13.5, E13.7, E14.2 - E14.5, E14.7 • Hemiplegia or paraplegia: G04.1, G11.4, G80.1, G80.2, G81.x, G82.x, G83.0 - G83.4, G83.9 • Renal disease: I12.0, I13.1, N03.2 - N03.7, N05.2 - N05.7, N18.x, N19.x, N25.0, Z49.0 - Z49.2, Z94.0, Z99.2 • Any malignancy, including lymphoma and leukemia, except malignant neoplasm of skin: C00.x - C26.x, C30.x - C34.x, C37.x - C41.x, C43.x, C45.x - C58.x, C60.x - C76.x, C81.x - C85.x, C88.x, C90.x - C97.x • Moderate or severe liver disease: I85.0, I85.9, I86.4, I98.2, K70.4, K71.1, K72.1, K72.9, K76.5, K76.6, K76.7 • Metastatic solid tumors: C77.x - C80.x • AIDS/HIV: B20.x - B22.x, B24.x	• Congestive heart failure: I09.9, I11.0, I13.0, I13.2, I25.5, I42.0, I42.5 - I42.9, I43.x, I50.x, P29.0 • Cardiac arrhythmias: I44.1 - I44.3, I45.6, I45.9, I47.x - I49.x, R00.0, R00.1, R00.8, T82.1, Z45.0, Z95.0 • Valvular disease: A52.0, I05.x - I08.x, I09.1, I09.8, I34.x - I39.x, Q23.0 - Q23.3, Z95.2 - Z95.4 • Pulmonary circulation disorders: I26.x, I27.x, I28.0, I28.8, I28.9 • Peripheral vascular disorders: I70.x, I71.x, I73.1, I73.8, I73.9, I77.1, I79.0, I79.2, K55.1, K55.8, K55.9, Z95.8, Z95.9 • Hypertension, uncomplicated: I10.x • Hypertension, complicated: I11.x - I13.x, I15.x • Paralysis: G04.1, G11.4, G80.1, G80.2, G81.x, G82.x, G83.0 - G83.4, G83.9 • Other neurological disorders: G10.x - G13.x, G20.x - G22.x, G25.4, G25.5, G31.2, G31.8, G31.9, G32.x, G35.x - G37.x, G40.x, G41.x, G93.1, G93.4, R47.0, R56.x • Chronic pulmonary disease: I27.8, I27.9, J40.x - J47.x, J60.x - J67.x, J68.4, J70.1, J70.3 • Diabetes without chronic complication: E10.0, E10.1, E10.9, E11.0, E11.1, E11.9, E12.0, E12.1, E12.9, E13.0, E13.1, E13.9, E14.0, E14.1, E14.9 • Diabetes,with chronic complication: E10.2 - E10.8, E11.2 - E11.8, E12.2 - E12.8, E13.2 - E13.8, E14.2 - E14.8 • Hypothyroidism: E00.x - E03.x, E89.0 • Renal failure: I12.0, I13.1, N18.x, N19.x, N25.0, Z49.0 - Z49.2, Z94.0, Z99.2 • Liver disease: B18.x, I85.x, I86.4, I98.2, K70.x, K71.1, K71.3 - K71.5, K71.7, K72.x - K74.x, K76.0, K76.2 - K76.9, Z94.4 • Peptic ulcer disease, excluding bleeding: K25.7, K25.9, K26.7, K26.9, K27.7, K27.9, K28.7, K28.9 • AIDS/HIV: B20.x - B22.x, B24.x • Lymphoma: C81.x - C85.x, C88.x, C96.x, C90.0, C90.2 • Metastatic cancer: C77.x - C80.x • Solid tumor without metastasis: C00.x - C26.x, C30.x - C34.x, C37.x - C41.x, C43.x, C45.x - C58.x, C60.x - C76.x, C97.x • Rheumatoid arthritis/collagen vascular diseases: L94.0, L94.1, L94.3, M05.x, M06.x, M08.x, M12.0, M12.3, M30.x, M31.0 - M31.3, M32.x - M35.x, M45.x, M46.1, M46.8, M46.9 • Coagulopathy: D65 - D68.x, D69.1, D69.3 - D69.6 • Obesity: E66.x • Weight loss: E40.x - E46.x, R63.4, R64 • Fluid and electrolyte disorders: E22.2, E86.x, E87.x • Blood loss anemia: D50.0 • Deficiency anemia: D50.8, D50.9, D51.x - D53.x • Alcohol abuse: F10, E52, G62.1, I42.6, K29.2, K70.0, K70.3, K70.9, T51.x, Z50.2, Z71.4, Z72.1 • Drug abuse: F11.x - F16.x, F18.x, F19.x, Z71.5, Z72.2 • Psychoses: F20.x, F22.x - F25.x, F28.x, F29.x, F30.2, F31.2, F31.5 • Depression: F20.4, F31.3 - F31.5, F32.x, F33.x, F34.1, F41.2, F43.2

Table B.7 Inclusion criteria to The US PAH Biobank.

Inclusion criteria	**YES**	**NO**
A. Is the patient newly diagnosed or previously diagnosed with WHO Group I PAH? (*include patients with wedge pressure or LVEDP $\leq$ 18 mm Hg as specified below)		
Documentation of the following hemodynamic parameters by diagnostic right heart catheterization, performed at any time prior to study enrolment:		
B. Specify hemodynamic profile below:		
1. Is mean pulmonary arterial pressure (mPAP) > 25 mmHg at		
rest with a wedge pressure or Left Ventricular End Diastolic Pressure (LVEDP) $\leq$ 15 mmHg?		
2. Is mPAP $>$ 25 mmHg at rest with a wedge pressure or LVEDP $>$ 15 mmHg but wedge pressure or LVEDP $\leq$ 18 mmHg?		
3. Is mPAP $<$ 25 at rest but > 30 mmHg with effort with a wedge pressure or LVEDP $\leq$ 15 mmHg?		
4. Is mPAP $<$ 25 at rest but > 30 mmHg with effort with a wedge pressure or LVEDP > 15 mmHg but $\leq$ 18 mmHg?		
C. Is pulmonary vascular resistance (PVR) $\geq$ 240 dynes.sec.cm^{-5} (i.e., $\geq$ 3.0 Wood units)?		
Exclusion Instructions: Exclusion Criteria must be checked "NO" in order for patient to be eligible for enrolment into the National Biological Sample and Data Repository for PAH (PAH Biobank)		
Exclusion Criteria	**YES**	**NO**
Does patient meet the criteria for inclusion into WHO Groups II, III, IV or V? (*Except with PCWP or LVEDP as above)		
The patient has not had documentation of hemodynamic criteria for PAH by right heart catheterization at some time preceding study entry and following development of symptoms associated with PAH.		
The patient does not meet the required hemodynamic criteria for entry into the study.		

Inclusion Instructions: Inclusion Criteria "a", "b", and "c" must all be checked "YES" in order for patient to be eligible for enrolment into the National Biological Sample and Data Repository for PAH (PAH Biobank). *Patient must meet 1 of 4 hemodynamic profiles for "b".

Table B.8 Reporting proforma for CT scan revision.

Parameter	Response
ID	character
Reader	character
CT scan date	date
Slice thickness	numeric
Number of slices	numeric
CTPA	done/not done
HRCT	done/not done
Expiratory CT	done/not done
Pleural effusion	Nil; Trace; Mild; Moderate; Severe
Subcutaneous oedema	present, absent
Severity of GGO centrilobular pattern	Nil; Trace; Mild; Moderate; Severe
Severity of GGO non-specific mosaic pattern	Nil; Trace; Mild; Moderate; Severe
Distribution of GGO	C-central; U-upper; Z-zonal; D-diffuse
Pulmonary arteriovenous malformations	present, absent
Largest bronchial artery size	numeric [mm]
Mediastinal venous collaterals	present, absent
Intralobular septal thickening	Nil; Trace; Mild; Moderate; Severe
Mediastinal lymphadenopathy	present, absent
Mediastinal lymphadenopathy subcarinal	[mm]
Emphysema	Nil; Mild; Moderate; Severe
Bronchial wall thickening	Nil; Trace; Mild; Moderate; Severe
Fibrosis	Nil; Mild; Moderate; Severe
Air trapping	Nil; Trace; Mild; Moderate; Severe
Subpleural scarring	Nil; Mild; Moderate; Severe

Abbreviations: ID: identifier; CTPA: computerised tomography pulmonary angiogram; HRCT: high resolution computerised tomography; GGO: ground-glass opacity

Table B.9 Definitions or clinical characteristics (in case of computational phenotypes) of labels and the number of unrelated cases and controls included in the genetic association analysis with BeviMed.

Tag	Tag description	Cases	Controls	Excluded relatives
PH	Individuals with mPAP > 25 mmHg	1,112	9,134	2,786
PAH	Patients with one of the following diagnoses: IPAH, HPAH, PVOD, PCH, APAH:CHD-PAH, APAH:CTD-PAH, APAH:HIV-PAH, APAH:PH-PAH	1,085	9,134	2,786
I/HPAH	Patients with a clinical diagnosis of IPAH or HPAH	1,036	9,134	2,786
IPAH	Patients with a clinical diagnosis of IPAH	972	9,134	2,785
HPAH	Patients with a clinical diagnosis of HPAH	67	9,136	2,779
PVOD/PCH	Patients with a clinical diagnosis of PVOD/PCH	20	9,136	2,778
I/HPAH/PVOD/PCH	Patients with one of the following diagnoses: IPAH, HPAH, PVOD, PCH	1,056	9,134	2,786
FPAH	Patients with one of the following diagnoses: IPAH, HPAH, PVOD, PCH and a positive family history	80	9,136	2,781
APAH	Patients with one of the following diagnoses: APAH:CHD-PAH, APAH:CTD-PAH, APAH:HIV-PAH, APAH:PH-PAH	29	9,136	2,778
APAH: CHD-PAH	Patients with PAH associated with congenital heart disease	17	9,136	2,778
APAH: CTD-PAH	Patients with PAH associated with connective tissue disease	10	9,136	2,778
APAH: PoPH	Patients with PAH associated with portopulmonary hypertension	1	9,136	2,778
APAH: HIV-PAH	Patients with PAH associated with HIV	1	9,136	2,778
PH-LHD	Patients with pulmonary hypertension associated with left heart disease (Group 2)	7	9,136	2,778
PH-LD	Patients with pulmonary hypertension associated with lung disease (Group 3)	8	9,136	2,778
CTEPH	Chronic thromboembolic pulmonary hypertension (Group 4)	6	9,136	2,778
PH-multifactorial	Multifactorial pulmonary hypertension (Group 5)	6	9,136	2,778
young age	Lower age tertile (<40.8 years)	378	9,136	2,785
middle age	Middle age tertile (40.8 - 58.6 years)	376	9,134	2,779
old age	Higher age tertile (>58.6 years)	355	9,136	2,778
low KCO	KCO <50% pred.	152	9,136	2,778
KCO lower tertile	KCO <60% pred.	211	9,136	2,778
KCO middle tertile	KCO 60-80% pred.	215	9,136	2,778
KCO higher tertile	KCO >80% pred.	215	9,134	2,779
Super survivor	Survived $\geq$10 years from diagnosis	367	9,883	2,778
Poor survivor	Survived $\leq$3 years from diagnosis	149	9,136	2,778
Vasoresponder	Patients meeting lenient criteria for acute NO challenge	59	9,136	2,778
Small interatrial defect	Patients with patent foramen ovale/small atrial septal defects	78	9,136	2,778
PAM 1 (Lin)	Genetic diagnosis in 24% of cases, 2.5% with diagnosis of APAH, middle age onset, preserved KCO	399	9,063	2,829
PAM 2 (Lin)	Genetic diagnosis in 10% of cases, 5.3% with diagnosis of PVOD/PCH, late disease onset, small female predominance, low KCO, desaturation on exertion, parenchymal changes	315	9,082	2,806
PAM 4 (Lin)	Genetic diagnosis in 23%, highest proportion of patients with I/HPAH, middle age onset, low KCO, desaturation on exertion, no parenchymal changes	177	9,135	2,753
AP 1 (Resnik)	I/HPAH, early disease onset, strong autoimmunity signal	26	9,129	2,759
AP 2 (Resnik)	IPAH, gastrointestinal and metabolic phenotype	9	9,134	2,754

Table B.9 Definitions or clinical characteristics (in case of computational phenotypes) of labels and the number of unrelated cases and controls included in the genetic association analysis with BeviMed.

Tag	Tag description	Cases	Controls	Excluded relatives
AP 5 (Resnik)	I/HPAH, PAH-CTD, keratoconjunctivitis, low NTproBNP	18	9,132	2,756
AP 6 (Resnik)	I/HPAH, PH-LD, multifactorial PH, PVOD, abnormalities of musculoskeletal system and blood and blood forming organs	63	9,126	2,762
AP 24 (Resnik)	IPAH, blood and solid organs neoplasms	13	9,127	2,761
AP 27 (Resnik)	IPAH, early disease onset, strong thyroid and autoimmunity signal	14	9132	2756
AP 28 (Resnik)	I/HPAH, PVOD, vascular skin abnormalities	61	9,128	2,760
AP 29 (Resnik)	I/HPAH, late onset disease, atherosclerosis	34	9,125	2,763
AP 33 (Resnik)	IPAH, small interatrial defects	14	9,133	2,755
AP 34 (Resnik)	I/HPAH, pediatric onset disease	12	9,135	2,753
AP 42 (Resnik)	I/HPAH, early disease onset, high spontaneous abortion rate	21	9,129	2,760
AP 44 (Resnik)	IPAH, high frequency of pulmonary embolism and thyroid disease	10	9,134	2,754
AP 47 (Resnik)	IPAH, PAH-CTD, late disease onset, renal failure, abnormalities of blood and blood forming organs	16	9,132	2,756
AP 50 (Resnik)	I/HPAH, early disease onset, vasoresponders	13	9,131	2,757
AP 58 (Resnik)	I/HPAH, early disease onset, GGO on HRCT, abnormalities in enzymes/coenzymes activity	15	9,133	2,755
AP 61 (Resnik)	I/HPAH, peripheral lung neovascularity	13	9,133	2,755
AP 63 (Resnik)	IPAH, late onset, abnormalities endocrine system and blood vessels morphology	20	9,133	2,755
AP 64 (Resnik)	I/HPAH, PVOD, multifactorial PH, mental and neurological disturbances	19	9,134	2,754
AP 66 (Resnik)	IPAH, PAH-CHD, arrhythmia, PFO	7	9,135	2,753
AP 86 (Resnik)	IPAH, drug use	10	9,135	2,753
AP 92 (Resnik)	IPAH, late disease onset, no mutations in known PAH risk genes	14	9132	2756
Divisive HC 1 (Lin)	Mostly I/HPH and APAH, higher proportion of Europeans, higher proportion of patients in WHO FC III and IV, higher mPAP, higher KCO	709	9,029	2,867
Divisive HC 2 (Lin)	Higher proportion of patients with PVOD/PCH and PH-LHD, lower proportion of Europeans, lower proportion of patients in WHO FC III and IV, lower mPAP and KCO	404	9,085	2,803

Abbreviations: mPAP: mean pulmonary artery pressure, PH: pulmonary hypertension, PAH: pulmonary arterial hypertension, I/H/FPAH: Idiopathic/Hereditary/Familial Pulmonary Arterial Hypertension, PVOD: Pulmonary veno-occlusive disease, PCH: Pulmonary capillary haemangiomatosis, APAH : Associated Pulmonary Arterial Hypertension, CHD: Congenital Heart Disease, CTD: Connective Heart Disease, PoPH: portopulmonary hypertension, LHD: Left Heart Disease, LD: Lung Disease, CTEPH: Chronic Thromboembolic Pulmonary Hypertension, KCO: transfer coefficient of carbon monoxide; PAM: partitioning around medoids; AP: affinity propagation clustering; HC: hierarchical clustering; GGO: ground-glass opacities; WHO FC: World Health Organisation Functional Class; HRCT: high-resolution computerised tomography; PFO: patent foramen ovale; NTproBNP: N-terminal pro-brain natriuretic peptide

Table B.10 Breakdown of cases from the NBR and the PAH Cohort studies per centre.

Centre	Cases, n (%)
Imperial and Hammersmith	319 (22.9)
University of South Paris	166 (11.9)
Papworth	164 (11.8)
Sheffield	133 (9.5)
Glasgow	111 (8)
Royal Brompton	97 (7)
University of Giessen	95 (6.8)
Newcastle Freeman	84 (6)
Pavia	51 (3.7)
Royal Free	50 (3.6)
VU University Medical Center Amsterdam	48 (3.4)
Royal United Hospital Bath	30 (2.2)
Graz	27 (1.9)
Great Ormond Street	19 (1.4)

Table B.11 International registries

Registry	Years	Diagnosis	Incident/Prevalent	N of patients	Age	Sex	WHO FC	NO challenge	ANA	Survival [%]		
										At 1 [y]	At 3 [y]	At 5 [y]
NIH (Rich et al. 1987; D'Alonzo et al. 1991) [438, 97]	1981-1985	PPH	Incident (n=125, 64%)	All 187	All 36(15)	All 59% female	All 75% in WHO class III and IV	NA	All 29% ANA positive, 69% females were ANA positive	68	48	34
Scottish morbidity records* (Peacock et al. 2007) [394]	1986-2001	PAH (IPAH n=175, PAH-CTD n=111, PAH-CHD n=88)	Incident and prevalent	All 374		All 70% female;						
Spanish registry ¶ (Escribano-Subias et al. 2012) [123]	1998-2008	PAH (IPAH 30%, PAH-CTD 15%, PAH-CHD 16%, PPH 6%, PAH-TOS 3.2, PVOD 1.5%)	Incident and prevalent	All 866	All 45(17); IPAH 46(18)	All 71% female; IPAH 73%	All 69% in WHO class III and IV, IPAH 70%WHO class III and IV	12% incident cases; 21% prevalent cases	NA	All 87	All 75	All 65
UK and Ireland registry‖ (Ling et al. 2012)[312]	2001-2009	PAH(I/HPAH, PAH-DTOX)	Incident	All 482	All 50.1(17.1)	All 70% female;	All 84.2% WHO class III and IV			All 92.7	All 73.3	All 61.1
US based registry (single centre; Chicago registry) (Thenappan et al. 2007) [519]	1982-2006	PAH (I/HPAH 48%, PAH-CTD 30%, PAH-CHD 11%, PAH-PPH 7%, PAH-Anorexigens 3%, PAH-HIV 1%)	Incident (n=82), prevalent (n=496)	758	All 48(14) (21-83); I/HPAH 45(14)	All 77% female; I/HPAH 75% female	All 80% in WHO class III and IV	All 4.6%; I/H-PAH 5.4%	202/380 positive (>1:80) I/HPAH 33%, PAH-CHD 34%, PAH-CTD 94%, PAH-Anorexigen 40%, PAH-PoPH 35%			
Danish registry (Korsholm et al. 2015) [275]	2000-2012	PAH (IPAH 32%, FPAH 0.8%, PAH-CTD 23%, PAH-CHD 35.8%, PAH-HIV 2.2%, PPH 2.2%)	Incident	All 134		All 58% female;	All 61.2% in WHO class III and IV	27/83 (32.5%)	NA	All 86.4 [79.3;91.2]	All 72.9 [64.1;79.9]	All 65 [55.8;73.4]

Table B.11 International registries

Registry	Years	Diagnosis	Incident/Prevalent	N of patients	Age	Sex	WHO FC	NO challenge	ANA	Survival [%]		
										At 1 [y]	At 3 [y]	At 5 [y]
German registry (Hoeper et al. 2016) [217]	2014			All 1754		All 62% female;						
UK National Audit 2014 (Corris 2015) [89]	2004-2014	PAH(I/H/drug induced PAH 33%, PAH-CTD 23%, PAH-CHD 34%, PPH 4%, Other 2%, No sub diagnosis 5%	Incident and prevalent	All 2940		All 65% female						
Chinese registry** (Jing et al. 2007) [250]	1999-2004	I/HPAH	Incident	All 72	All 35.9(12.2)	All 71% female	All 85.9% in WHO class III and IV	3/12 (25%) of IPAH cases	NA	All 68	All 38.9	All 20.8
Chinese registry (Zhang et al. 2011) [591]	2007-2009	PAH(IPAH n=173, 62.7%, PAH-CTD n=103, 37.3%)	Prevalent	All 276	All; IPAH 33.4(15.3)	All female; IPAH 69.9%	All; IPAH 52% in WHO class III and IV	IPAH 6(3.5%); PAH-CTD 2(1.9%)	NA	IPAH 92, PAH-CTD 85	IPAH 75, PAH-CTD 54	NA
Brazilian registry (Alves et al. 2015) [12]	2008-2013	IPAH 28.7%, PAH-CTD 25.8%, PAH-Sch 19.7%, Other 10.7%	Incident	All 178	All 46.(15); IPAH 39.8(14.8)	All 77% female; IPAH 77% female	All 81% in WHO class III and IV	NA	NA	All 93; IPAH 97	All 74; IPAH 87	NA
ASPIRE (Hurdman et al. 2012) [231]	2001-2010	PAH(IPAH n=175, PAH-CTD n=188, PAH-CHD n=198, PAH-PPH n=24, PAH-hemolytic anemia n=2)	Incident	All 598	All 54(18); IPAH 55(16)	All 62%; IPAH 67% female	All 78%; IPAH 87% in WHO class III and IV	6% of IPAH cases	NA		IPAH 63	
COMPERA (Hoeper et al. 2013) [218]		PAH(IPAH n=568, 96.8%; HPAH n=7, 1.2%; PAH-drug associated n=12(2%)	Incident	All 587	All 75(16); 18-65 yrs n=209; >65yrs n=378	All 62.5% female	All 90% in WHO class III and IV			All 92	All 74	NA

Table B.11 International registries

Registry	Years	Diagnosis	Incident/Prevalent	N of patients	Age	Sex	WHO FC	NO challenge	ANA	Survival [%]		
										At 1 [y]	At 3 [y]	At 5 [y]
REVEAL§ (Farber et al. 2015) [129]	2006-2009	Incident: IPAH 47.9%; FPAH 2%; APAH-CTD 30.7%; APAH-CHD 5.5%; APAH-PoPH 6.2%; APAH-HIV 1.7%; APAH-drugs and toxins 5.2%; Prevalent: IPAH 46.6%; FPAH 3.1%; APAH-CTD 24.3%; APAH-CHD 11.6%; APAH-PoPH 5.4%; APAH-HIV 1.9%; APAH-drugs and toxins 5.9%	Incident (n=710) and prevalent (n=2039)	All 2749	Incident: 53.1(15.4); Prevalent: 51.7(14.5)	All 79% female	Incident: 73.5%; Prevalent: 53.1% in WHO class III and IV (at enrollment)			Prevalent: 90; Incident 86		Prevalent: 65; Incident 61
French registry‡ (Humbert et al. 2006) [229]	2002-2003	PAH(IPAH n=264, 39.2%; FPAH n=26, 3.9%; anorexigen associated n=64, 9.5%, PAH-CTD n=103, 15.3%; PAH-CHD n=76, 11.3%, PAH-PPH n=70, 10.4%; PAH-HIV n=42, 6.2%; two risk factors n=29, 4.3%)	Incident (n=121) and prevalent (n=553)	All n=674	All 50(15) (18-85yrs)	All 65% of females	All 75% in WHO class III and IV	10.3% of IPAH cases	NA	87	67	NA
Saudi Arabian registry‡ (Idrees et al. 2015) [238]	2009-2012	PAH(IPAH n=55, HPAH n=4, PAH-CHD n=29, PAH-CTD n=16, PAH-PPH n=3)		All 107	All 36(8); IPAH 38(6)	All 63% female; IPAH 65.5%	All 72.9% in WHO class III and IV; IPAH 78.2% in WHO class III and IV	All 5(4.7); IPAH 4(7.3)%		72	57	NA

Table B.11 International registries

Registry	Years	Diagnosis	Incident/Prevalent	N of patients	Age	Sex	WHO FC	NO challenge	ANA	Survival [%]		
										At 1 [y]	At 3 [y]	At 5 [y]
Giessen registry (Gall et al. 2017)[155]	1993-2011	PAH(IPAH 43.1%, PAH-CTD 21.2%, PAH-CHD 13.3%, PAH-PPH 7.4%, PVOD 4.1%, PAH-HIV 3.9%, other 7.2%	Incident and prevalent (IPAH incident n=226, prevalent n=68; APAH incident 325, prevalent n=38)	All 685	All 51	All 65% female	All 81% in WHO class III and IV	NA		All 88; IPAH 90	All 72; IPAH 76	All 59; IPAH 65
The study NBR/Cohort	2002-2020	IPAH 1191(85.4%), HPAH 91(6.5%), PVOD/PCH 35(2.5%), other 77(5.5%)	incident and prevalent	All 1394	All 50 [35;65]; I/H-PAH/PVOD/PCH	All 68%; I/H-PAH/PVOD/PCH	All 78% in WHO class III and IV	I/HPAH/PVOD/PCH 87(16.1%)	Measured 21%, reported 15.5%	98	91	82

*Inclusion age was restricted to 16–65 years. Inclusion age-restricted to those over 16 years. †Inclusion age was restricted to 18–70 year. ‡Inclusion age was 18 years or older. §Inclusion age older than 3 months; only 16% of the patients were incident. Inclusion age in inclusion criteria not stated. ‖Only idiopathic; heritable; and drug-associated pulmonary arterial hypertension; ** RHC data at diagnosis available only for 20 out of 72 patients.

Abbreviations: N of patients: Number of patients; NIH: National Institutes of Health; ASPIRE: assessing the Spectrum of Pulmonary hypertension Identified at a REferral centre; COMPERA: Comparative Prospective Registry of Newly Initiated Therapies for Pulmonary Hypertension; REVEAL: Registry to Evaluate Early And Long-term PAH Disease Management; TOS: toxic oil syndrome; TFS: transplant-free survival; ISWD: incremental shuttle walk test; ANA: antinuclear antibodies; I/H/F/APAH: idiopathic/hereditary/familial pulmonary arterial hypertension; PPH: primary pulmonary hypertension; CHD: congenital heart disease; TOS: toxic oil syndrome; CTD: connective tissue disease; HIV: human immunodeficiency virus; PoPH: portopulmonary hypertension; PVOD/PCH: pulmonary veno-occlusive disease/pulmonary haemangiomatosis; WHO FC: World Health Organisation; PoPH: portopulmonary hypertension

Table B.12 Differences in clinical characteristics between patients with complete and missing pulmonary function tests.

	FEV1			FVC			TLC			KCO		
	missing N=108	**present N=709**	***p*-value**	**missing N=122**	**present N=695**	***p*-value**	**missing N=313**	**present N=504**	***p*-value**	**missing N=179**	**present N=638**	***p*-value**
Diagnosis verified: n (%)			0.158			0.339			0.887			0.277
HPAH	12 (11)	48 (7)		12 (10)	48 (7)		24 (8)	36 (7)		17 (9)	43 (7)	
IPAH	96 (89)	661 (93)		110 (90)	647 (93)		289 (92)	468 (93)		162 (91)	595 (93)	
Sex: female, n (%)	81 (75)	471 (66)	0.097	88 (72)	464 (67)	0.288	211 (67)	341 (68)	1	131 (73)	421 (66)	0.084
Age [years]	44 [31;62]	52 [39;67]	0.002	44 [32;63]	52 [39;67]	0.006	50 [33;65]	52 [40;67]	0.016	46 [31;62]	53 [39;67]	<0.001
Case: incident, n (%)	44 (41)	294 (41)	0.97	48 (39)	290 (42)	0.694	113 (36)	225 (45)	0.019	55 (31)	283 (44)	0.001
WHO FC:			0.183			0.258			0.939			0.078
I	3 (3)	5 (1)		3 (3)	5 (1)		3 (1)	5 (1)		4 (2)	4 (1)	
II	17 (17)	101 (15)		19 (17)	99 (15)		41 (14)	77 (15)		32 (19)	86 (14)	
III	68 (67)	488 (70)		78 (68)	478 (70)		211 (71)	345 (69)		111 (65)	445 (71)	
IV	13 (13)	101 (15)		15 (13)	99 (15)		43 (14)	71 (14)		24 (14)	90 (14)	
6MWD [m]	350 [168;432]	305 [180;390]	0.221	310 [156;430]	308 [180;390]	0.39	296 [152;382]	312 [182;396]	0.321	335 [162;404]	300 [180;392]	0.304
mRAP [mmHg]	8 [5;12]	9 [6;12]	0.508	8 [5;13]	9 [6;12]	0.582	9 [6;14]	8 [6;12]	0.035	8 [5;13]	9 [6;12]	0.728
mPAP [mmHg]	55 [44;60]	53 [44;61]	0.529	55 [43;60]	53 [44;61]	0.722	55 [46;63]	51 [43;60]	0.001	55 [44;61]	53 [44;61]	0.308
mPAWP [mmHg]	10 [7;12]	10 [7;12]	0.795	9 [7;12]	10 [7;12]	0.682	10 [7;13]	9 [7;12]	0.002	10 [7;12]	10 [7;12]	0.956
CO [L/min]	4 [3;5]	4 [3;5]	0.964	4 [3;5]	4 [3;5]	0.784	4 [3;5]	4 [3;5]	0.089	4 [3;5]	4 [3;5]	0.914
PVR [WU]	12 [9;17]	11 [8;16]	0.435	12 [8;16]	11 [8;15]	0.355	11 [8;16]	11 [8;15]	0.932	12 [9;16]	11 [8;15]	0.391

Abbreviations: FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; KCO: transfer coefficient for carbon monoxide; TLC: total lung capacity; I/HPAH: idiopathic/hereditary pulmonary arterial hypertension; mPAP: mean pulmonary arterial pressure; mPAWP: mean pulmonary artery wedge pressure; mRAP: mean right atrial pressure; PVR: pulmonary vascular resistance; WHO FC: World Health Organization functional class; WU: Wood Units; 6MWD: six-minute walk distance; SpO_2: arterial oxygen saturation; CO: cardiac output

Table B.13 Clinical characteristics of study population by FEV1 threshold – complete cases.

	FEV1 $\geq$ 80 N=433	**FEV1<80 N=276**	***p*-value**	**N**
Diagnosis verified: n (%)			0.329	709
HPAH	33 (8)	15 (5)		
IPAH	400 (92)	261 (95)		
Sex: female, n (%)	288 (67)	183 (66)	1	709
Age [years]	52 [38;67]	53 [39;66]	0.851	709
WHO FC: I/II/III/IV, %	0/14/72/14	0/16/68/16	0.729	695
6MWD [m]	319 [182;405]	288 [169;362]	0.027	460
SpO_2 pre [%]	96 [93;98]	95 [92;97]	0.003	632
SpO_2 post [%]	92 [84;96]	91 [85;95]	0.416	578
mRAP [mmHg]	8 [5;12]	10 [7;14]	<0.001	672
mPAP [mmHg]	52 [44;60]	54 [44;62]	0.111	692
mPAWP [mmHg]	9 [7;11]	10 [8;13]	<0.001	628
CO [L/min]	3.8 [3.0;4.7]	3.8 [3.1;4.8]	0.382	682
PVR [WU]	11.2 [8.0;15.9]	11.2 [7.7;15.0]	0.655	622
FEV1 [% pred.]	95 [88;102]	70 [64;76]	<0.001	709
FVC [% pred.]	103 [95;113]	80 [72;88]	<0.001	694
TLC [% pred.]	99 [90;106]	86 [76;94]	<0.001	500
KCO [% pred.]	70.0 [47.8;85.2]	75.0 [58.8;88.6]	0.005	636
Fibrosis: n (%)			0.137	501
none	284 (93.7)	193 (97.5)		
minimal/mild	17 (5.61)	5 (2.53)		
moderate	2 (0.66)	0 (0.00)		
Emphysema: n (%)			0.335	499
none	273 (89.5)	169 (87.1)		
minimal/mild	24 (7.87)	15 (7.73)		
moderate	8 (2.62)	10 (5.15)		
Smoking: past/current, n (%)	227 (55.6)	141 (52.8)	0.521	675
COPD: n (%)	5 (1.2)	27 (9.8)	<0.001	709
Asthma: n (%)	36 (8.3)	30 (10.9)	0.313	709
Initial therapy: n (%)			0.098	550
CCB	36 (10.6)	11 (5.2)		
Combination therapy	169 (49.9)	114 (54.0)		
ERA	42 (12.4)	18 (8.5)		
PA	8 (2.4)	5 (2.4)		
PDE5	83 (24.5)	63 (29.9)		
sGC	1 (0.3)	0 (0.0)		

Data are presented as median [interquartile range] unless otherwise specified. Abbreviations: ERA: endothelin receptor antagonists; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; KCO: transfer coefficient for carbon monoxide; L/min, liters per minute; mPAP: mean pulmonary arterial pressure; mPAWP: mean pulmonary artery wedge pressure; mRAP: mean right atrial pressure; I/HPAH: idiopathic/hereditary pulmonary arterial hypertension; PDE5: phosphodiesterase type 5; PVR: pulmonary vascular resistance; sGC: Soluble Guanylate Cyclase Stimulator; TLC: total lung capacity; WHO FC: World Health Organization functional class; WU: Wood units; 6MWD: six-minute walk distance; SpO_2: arterial oxygen saturation; CO: cardiac output; COPD: chronic obstructive pulmonary disease; CCB: calcium channel blockers; PA: prostacyclin analogue

Table B.14 Clinical characteristics of study population by FVC severity.

	FVC ≥ 80 N=541	**FVC < 80 N=154**	***p*-value**	**N**
Diagnosis verified: n (%)			0.682	695
HPAH	39 (7)	9 (6)		
IPAH	502 (93)	145 (94)		
Sex: female, n (%)	368 (68)	96 (62)	0.221	695
Age [years]	53 [39;67]	49 [37;66]	0.15	695
WHO FC: I/II/III/IV, %	1/14/72/13	1/15/64/20	0.159	681
6MWD [m]	315 [183;399]	280 [156;363]	0.03	453
SpO_2 pre [%]	96 [93;98]	95 [92;97]	0.015	624
SpO_2 post [%]	91 [84;96]	93 [86;96]	0.458	569
mRAP [mmHg]	8 [6;12]	10 [7;14]	<0.001	662
mPAP [mmHg]	52 [44;61]	54 [45;63]	0.113	678
mPAWP [mmHg]	9 [7;12]	11 [8;13]	0.001	617
CO [L/min]	3.8 [3.0;4.7]	3.9 [3.3;4.9]	0.178	669
PVR [WU]	11.2 [7.9;15.8]	10.5 [7.7;15.0]	0.656	611
FEV1 [% pred.]	91 [81;100]	67 [60;74]	<0.001	694
FVC [% pred.]	101 [91;111]	72 [66;77]	<0.001	695
TLC [% pred.]	0.8 [0.7;0.8]	0.8 [0.7;0.8]	0.001	693
KCO [% pred.]	98 (13)	78 (12)	<0.001	491
Fibrosis: n (%)	70.0 [48.0;85.0]	80.0 [64.1;91.2]	<0.001	628
none			0.261	492
minimal/mild	362 (94.3)	106 (98.1)		
moderate	20 (5.21)	2 (1.85)		
Emphysema: n (%)	2 (0.52)	0 (0.00)		
none			0.564	490
minimal/mild	337 (87.8)	97 (91.5)		
moderate	32 (8.33)	7 (6.60)		
Smoking: past/current, n (%)	15 (3.91)	2 (1.89)		
COPD, n (%)	295 (57.2)	68 (46.6)	0.029	662
Asthma, n (%)	25 (4.6)	7 (4.5)	1	695
Initial therapy: n (%)	49 (9.1)	16 (10.4)	0.731	695
CCB	40 (9.3)	7 (6.1)	0.328	544
combination therapy	226 (52.7)	54 (47.0)		
ERA	48 (11.2)	12 (10.4)		
PA	10 (2.3)	2 (1.7)		
PDE5	104 (24.2)	40 (34.8)		
sGC	1 (0.2)	0 (0.0)		

Data are presented as median [interquartile range] unless otherwise specified. Abbreviations: ERA: endothelin receptor antagonists; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; I/HPAH: idiopathic/hereditary pulmonary arterial hypertension; KCO: transfer coefficient for carbon monoxide; mPAP: mean pulmonary arterial pressure; mPAWP: mean pulmonary artery wedge pressure; mRAP: mean right atrial pressure; PDE5: phosphodiesterase type 5; PVR: pulmonary vascular resistance; sGC: Soluble Guanylate Cyclase Stimulator; TLC: total lung capacity; WHO FC: World Health Organization functional class; WU: Wood units; 6MWD: six-minute walk distance; SpO_2: arterial oxygen saturation; CO: cardiac output; COPD: chronic obstructive pulmonary disease; CCB: calcium channel blockers; PA: prostacyclin analogue

Table B.15 Summary of imaging analysis by FEV1/FVC ratio.

	FEV1/FVC $\geq$ 0.7 N=182	**FEV1/FVC $<$ 0.7 N=59**	***p*-value**	**N**
Severity of GGO centrilobular pattern: n (%)			0.176	241
Nil	104 (57.1)	45 (76.3)		
Trace	24 (13.2)	5 (8.5)		
Mild	23 (12.6)	4 (6.8)		
Moderate	19 (10.4)	3 (5.1)		
Severe	12 (6.6)	2 (3.4)		
Severity of GGO non-specific pattern: n (%)			0.866	241
Nil	161 (88.5)	55 (93.2)		
Trace	8 (4.4)	2 (3.4)		
Mild	7 (3.8)	1 (1.7)		
Moderate	6 (3.3)	1 (1.7)		
Distribution of GGO: n (%)			0.393	108
C	8 (8.9)	2 (11.1)		
D	53 (58.9)	14 (77.8)		
U	15 (16.7)	1 (5.6)		
Z	14 (15.6)	1 (5.6)		
PAVM, n (%)	4 (2.4)	0 (0.0)	0.575	222
Mediastinal venous collaterals, n (%)	149 (94.9)	47 (92.2)	0.493	208
Intralobular septal thickening: n (%)			0.121	228
Nil	155 (91.2)	47 (81.0)		
Trace	9 (5.3)	6 (10.3)		
Mild	4 (2.4)	4 (6.9)		
Moderate	2 (1.2)	1 (1.7)		
Mediastinal lymphadenopathy:	153 (84.1)	47 (79.7)	0.56	241
Mediastinal lymphadenopathy [mm]	14.0 [12.0;15.0]	14.0 [11.8;17.2]	0.862	41
Emphysema:			0.007	241
Nil	164 (90.1)	47 (79.7)		
Trace	11 (6.0)	6 (10.2)		
Mild	6 (3.3)	1 (1.7)		
Moderate	1 (0.5)	5 (8.5)		
Bronchial wall thickening: n (%)			0.31	228
Nil	153 (90.0)	48 (82.8)		
Trace	11 (6.5)	6 (10.3)		
Mild	5 (2.9)	4 (6.9)		
Moderate	1 (0.6)	0 (0.0)		
Fibrosis: n (%)			0.226	241
Nil	174 (95.6)	56 (94.9)		
Trace	4 (2.2)	0 (0.0)		
Mild	4 (2.2)	2 (3.4)		
Moderate	0 (0.0)	1 (1.7)		
Pleural effusion: n (%)			0.393	241
Nil	165 (90.7)	52 (88.1)		
Trace	8 (4.4)	1 (1.7)		
Mild	4 (2.2)	3 (5.1)		
Moderate	4 (2.2)	2 (3.4)		
Severe	1 (0.5)	1 (1.7)		
Air trapping: n (%)			0.525	241
Nil	147 (80.8)	46 (78.0)		
Trace	18 (9.9)	6 (10.2)		
Mild	13 (7.1)	4 (6.8)		

Table B.15 Summary of imaging analysis by FEV1/FVC ratio.

	FEV1/FVC $\geq$ 0.7 N=182	**FEV1/FVC $<$ 0.7 N=59**	***p*-value**	**N**
Moderate	1 (0.5)	2 (3.4)		
Severe	3 (1.6)	1 (1.7)		

Intra-rater reliability: GGO centrilobular pattern severity weighted Cohen's Kappa=0.679, p-value < 0.001; GGO distribution unweighted Cohen's Kappa = 1, p-value 0.046; Severity of GGO non-specific pattern - no positive findings; Pulmonary arteriovenous malformations - no positive findings; largest BRA size - no positive findings; Mediastinal venous collaterals: unweighted Cohen's Kappa = 1, p-value < 0.001; Intralobular septal thickening weighted Cohen's Kappa = 1, p-value < 0.001; Mediastinal lymphadenopathy unweighted Cohen's Kappa = 0.83, p-value < 0.001; Mediastinal lymphadenopathy size [mm] intraclass correlation coefficient (ICC) 0.717, p-value = 0.088; Emphysema - not enough positive findings, Bronchial wall thickening - not enough positive findings, Fibrosis - no positive findings; Pleural effusion weighted Cohen's Kappa = 0.826, p-value < 0.001; Air trapping weighted Cohen's Kappa = 0.845, p-value < 0.001; Subpleural scarring - not enough positive findings. Abbreviations: GGO: ground-glass opacities; C: central; U: upper; Z: zonal; D:diffuse; PAVM: pulmonary arteriovenous malformations; FEV1: forced expiratory volume 1; FVC: forced vital capacity

Table B.16 Summary of imaging analysis by TLC threshold.

	TLC $\geq$ 80 N=89	**TLC $<$ 80 N = 20**	***p*-value**	**N**
Severity of GGO centrilobular pattern: n (%)			1	109
Nil	57 (64.0)	13 (65.0)		
Trace	9 (10.1)	2 (10.0)		
Mild	9 (10.1)	2 (10.0)		
Moderate	8 (9.0)	2 (10.0)		
Severe	6 (6.7)	1 (5.0)		
Severity of GGO non-specific pattern: n (%)			0.193	109
Nil	83 (93.3)	18 (90.0)		
Trace	2 (2.2)	0 (0.0)		
Mild	1 (1.1)	2 (10.0)		
Moderate	3 (3.4)	0 (0.0)		
Distribution of GGO: n (%)			0.293	45
C	3 (8.3)	0 (0.0)		
D	24 (66.7)	5 (55.6)		
U	6 (16.7)	1 (11.1)		
Z	3 (8.3)	3 (33.3)		
Neovascularity, n (%)	4 (4.7)	1 (5.3)	1	105
Mediastinal venous collaterals, n (%)	76 (96.2)	17 (94.4)	0.566	97
Intralobular septal thickening: n (%)			0.094	101
Nil	79 (96.3)	16 (84.2)		
Trace	1 (1.2)	2 (10.5)		
Mild	1 (1.2)	1 (5.3)		
Moderate	1 (1.2)	0 (0.0)		
Mediastinal lymphadenopathy, n (%)	75 (84.3)	15 (75.0)	0.336	109
Mediastinal lymphadenopathy [mm]	12.8 (3.8)	17.6 (2.7)	0.012	19
Emphysema: n (%)			0.395	109
Nil	79 (88.8)	16 (80.0)		
Trace	8 (9.0)	4 (20.0)		
Moderate	2 (2.2)	0 (0.0)		
Bronchial wall thickening: n (%)			0.356	101
Nil	69 (84.1)	19 (100.0)		
Trace	8 (9.8)	0 (0.0)		
Mild	5 (6.1)	0 (0.0)		
Fibrosis: n (%)			0.561	109
Nil	86 (96.6)	19 (95.0)		
Trace	1 (1.1)	1 (5.0)		
Mild	1 (1.1)	0 (0.0)		
Moderate	1 (1.1)	0 (0.0)		
Pleural effusion: n (%)			0.005	109
Nil	83 (93.3)	14 (70.0)		
Trace	4 (4.5)	1 (5.0)		
Mild	2 (2.2)	3 (15.0)		
Moderate	0 (0.0)	1 (5.0)		
Severe	0 (0.0)	1 (5.0)		
Air trapping: n (%)			0.735	109
Nil	67 (75.3)	17 (85.0)		
Trace	14 (15.7)	2 (10.0)		
Mild	4 (4.5)	0 (0.0)		
Moderate	2 (2.2)	0 (0.0)		
Severe	2 (2.2)	1 (5.0)		

Intra-rater reliability: GGO centrilobular pattern severity weighted Cohen's Kappa = 0.679, p-value < 0.001; GGO distribution unweighted Cohen's Kappa = 1, p-value 0.046; Severity of GGO non-specific pattern - no positive findings; Pulmonary arteriovenous malformations - no positive findings; largest BRA size - no positive findings; Mediastinal venous collaterals: unweighted Cohen's Kappa = 1, p-value < 0.001; Intralobular septal thickening weighted Cohen's Kappa=1, p-value <0.001; Mediastinal lymphadenopathy unweighted Cohen's Kappa=0.83, p-value<0.001; Mediastinal lymphadenopathy size [mm] intraclass correlation coefficient (ICC) = 0.717, p-value = 0.088; Emphysema - not enough positive findings, Bronchial wall thickening - not enough positive findings, Fibrosis - no positive findings; Pleural effusion weighted Cohen's Kappa = 0.826, p-value < 0.001; Air trapping weighted Cohen's Kappa = 0.845, p-value < 0.001; Subpleural scarring - not enough positive findings. Abbreviations: GGO: ground-glass opacities; C: central; U: upper, Z: zonal; D: diffuse; TLC: total lung capacity

Table B.17 Summary of imaging analysis by KCO threshold.

	[ALL] N=203	KCO $\geq$ 60 N = 139	KCO $<$ 60 N = 64	*p*-value	N
Severity of GGO centrilobular pattern: n (%)				0.005	203
Nil	130 (64.0)	78 (56.1)	52 (81.2)		
Trace	23 (11.3)	17 (12.2)	6 (9.4)		
Mild	22 (10.8)	18 (12.9)	4 (6.2)		
Moderate	17 (8.4)	16 (11.5)	1 (1.6)		
Severe	11 (5.4)	10 (7.2)	1 (1.6)		
Severity of GGO non-specific pattern: n (%)				0.299	203
Nil	186 (91.6)	129 (92.8)	57 (89.1)		
Trace	6 (3.0)	5 (3.6)	1 (1.6)		
Mild	5 (2.5)	2 (1.4)	3 (4.7)		
Moderate	6 (3.0)	3 (2.2)	3 (4.7)		
Distribution of GGO: n (%)				0.004	85
C	9 (10.6)	3 (4.3)	6 (37.5)		
D	55 (64.7)	47 (68.1)	8 (50.0)		
U	13 (15.3)	12 (17.4)	1 (6.2)		
Z	8 (9.4)	7 (10.1)	1 (6.2)		
Neovascularity, n (%)	16 (8.5)	13 (10.0)	3 (5.2)	0.398	188
PAVM, n (%)	4 (2.1)	4 (3.1)	0 (0.0)	0.313	188
Mediastinal venous collaterals, n (%)	166 (95.4)	118 (96.7)	48 (92.3)	0.242	174
Intralobular septal thickening, n (%)				0.002	190
Nil	172 (90.5)	126 (95.5)	46 (79.3)		
Trace	10 (5.3)	3 (2.3)	7 (12.1)		
Mild	6 (3.2)	3 (2.3)	3 (5.2)		
Moderate	2 (1.1)	0 (0.0)	2 (3.4)		
Mediastinal lymphadenopathy, n (%)	169 (83.3)	127 (91.4)	42 (65.6)	<0.001	203
Mediastinal lymphadenopathy [mm]	13.0 [12.0;15.0]	14.0 [12.0;15.0]	13.0 [11.2;16.5]	0.757	34
Emphysema: n (%)				<0.001	203
Nil	176 (86.7)	127 (91.4)	49 (76.6)		
Trace	16 (7.9)	11 (7.9)	5 (7.8)		
Mild	5 (2.5)	0 (0.0)	5 (7.8)		
Moderate	6 (3.0)	1 (0.7)	5 (7.8)		
Bronchial wall thickening: n (%)				0.39	190
Nil	166 (87.4)	118 (89.4)	48 (82.8)		
Trace	15 (7.9)	9 (6.8)	6 (10.3)		
Mild	8 (4.2)	4 (3.0)	4 (6.9)		
Moderate	1 (0.5)	1 (0.8)	0 (0.0)		
Fibrosis: n (%)				0.052	203
Nil	195 (96.1)	136 (97.8)	59 (92.2)		
Trace	2 (1.0)	1 (0.7)	1 (1.6)		
Mild	5 (2.5)	1 (0.7)	4 (6.2)		
Moderate	1 (0.5)	1 (0.7)	0 (0.0)		
Pleural effusion: n (%)				0.121	203
Nil	183 (90.1)	129 (92.8)	54 (84.4)		
Trace	8 (3.9)	5 (3.6)	3 (4.7)		
Mild	5 (2.5)	2 (1.4)	3 (4.7)		
Moderate	5 (2.5)	3 (2.2)	2 (3.1)		
Severe	2 (1.0)	0 (0.0)	2 (3.1)		

Table B.17 Summary of imaging analysis by KCO threshold.

	[ALL] N=203	KCO $\geq$ 60 N = 139	KCO $<$ 60 N = 64	*p*-value	N
Air trapping: n (%)				0.254	203
Nil	166 (81.8)	118 (84.9)	48 (75.0)		
Trace	18 (8.9)	10 (7.2)	8 (12.5)		
Mild	14 (6.9)	7 (5.0)	7 (10.9)		
Moderate	2 (1.0)	2 (1.4)	0 (0.0)		
Severe	3 (1.5)	2 (1.4)	1 (1.6)		
Subpleural scarring: n (%)				0.46	190
Nil	186 (97.9)	130 (98.5)	56 (96.6)		
Trace	1 (0.5)	0 (0.0)	1 (1.7)		
Mild	3 (1.6)	2 (1.5)	1 (1.7)		

Intra-rater reliability: GGO centrilobular pattern severity weighted Cohen's Kappa = 0.679, p-value$<$0.001; GGO distribution unweighted Cohen's Kappa = 1, p-value 0.046; Severity of GGO non-specific pattern - no positive findings; Pulmonary arteriovenous malformations - no positive findings; largest BRA size - no positive findings; Mediastinal venous collaterals: unweighted Cohen's Kappa = 1, p-value$<$0.001; Intralobular septal thickening weighted Cohen's Kappa = 1, p-value$<$0.001; Mediastinal lymphadenopathy unweighted Cohen's Kappa = 0.83, p-value $<$0.001; Mediastinal lymphadenopathy size [mm] intraclass correlation coefficient (ICC) 0.717, p-value = 0.088; Emphysema - not enough positive findings, Bronchial wall thickening - not enough positive findings, Fibrosis - no positive findings; Pleural effusion weighted Cohen's Kappa = 0.826, p-value$<$0.001; Air trapping weighted Cohen's Kappa 0.845, p-value$<$40.001; Subpleural scarring - not enough positive findings. Abbreviations: GGO: ground-glass opacities; PAVM: pulmonary arteriovenous malformations; BA: bronchial artery; C: central; U: upper; Z: zonal; D:diffuse; KCO: transfer factor for carbon monoxide

Table B.18 Correlations between pulmonary function tests predictive of survival in univariate analysis and other relevant clinical parameters.

variable 1	variable 2	Rho [95CI]	*p*-value
FEV1/FVC ratio	Age at diagnosis	-0.46 [-0.52;-0.4]	**< 0.001**
	CI	0.02 [-0.05;0.09]	0.549
	mPAP	0.08 [0;0.15]	**0.048**
	mRAP	-0.02 [-0.1;0.06]	0.613
	PVR	0.09 [0.01;0.17]	**0.027**
	6MWD	0.1 [0.02;0.19]	**0.026**
KCO % pred.	Age at diagnosis	-0.38[-0.45;-0.32]	**< 0.001**
	CI	-0.04 [-0.12;0.04]	0.307
	mPAP	0.24 [0.17;0.31]	**< 0.001**
	mRAP	0.09 [0;0.17]	**0.035**
	PVR	0.15 [0.07;0.22]	**< 0.001**
	6MWD	0.25 [0.15;0.34]	**< 0.001**
TLC % pred.	Age at diagnosis	-0.04[-0.13;0.04]	0.324
	CI	0.06 [-0.03;0.15]	0.196
	mPAP	0.02 [-0.07;0.1]	0.635
	mRAP	-0.19 [-0.28;-0.1]	**< 0.001**
	PVR	0.03 [-0.05;0.13]	0.509
	6MWD	0.2 [0.09;0.29]	**< 0.001**

Confidence intervals of a Spearman's rank correlation coefficients were derived by bootstrapping. Abbreviations: FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; CI: cardiac index; mPAP: mean pulmonary artery pressure; mRAP: mean right atrial pressure; PVR pulmonary vascular resistance; 6MWD: 6-minute walk distance

Table B.19 Comparison of patients with and without post-baseline values.

	No follow-up N=697	Follow-up available N=120	*p*-value	N
Diagnosis verified: I/HPAH, n (%)	649 (93)/48(7)	108 (90)/12 (10)	0.309	817
Sex: female, n (%)	472 (68)	80 (67)	0.903	817
Age [years]	51 [37;67]	50 [36;65]	0.539	817
Prevalent case, n (%)	477 (68)	2 (2)	<0.001	817
WHO FC: I/II/III/IV, %	1/16/67/13	0/9/70/21	0.033	796
6MWD [m]	306 [180;394]	311 [165;398]	0.949	510
SpO_2 pre [%]	96 [93;97]	95 [92;98]	0.984	708
SpO_2 post [%]	92 [85;96]	91 [84;96]	0.582	647
mRAP [mmHg]	9 [6;12]	9 [6;13]	0.512	759
mPAP [mmHg]	53 [44;61]	53 [47;63]	0.278	786
mPAWP [mmHg]	10 [7;12]	9 [7;12]	0.591	709
CO [L/min]	3.8 [3.0;4.8]	3.6 [3.1;4.6]	0.743	768
PVR [WU]	11.2 [7.8;15.8]	11.7 [8.6;15.1]	0.396	699
FEV1 [% pred.]	86 [73;97]	86 [76;99]	0.283	709
FVC [% pred.]	95 [82;107]	98 [82;110]	0.349	695
TLC [% pred.]	94 [86;104]	92 [82;102]	0.193	504
Fibrosis: n (%)			0.845	567
none	455 (95.0)	85 (96.6)		
minimal/mild	22 (4.59)	3 (3.41)		
moderate	2 (0.42)	0 (0.00)		
Emphysema: n (%)			0.124	565
none	429 (89.9)	75 (85.2)		
minimal/mild	30 (6.29)	11 (12.5)		
moderate	18 (3.77)	2 (2.27)		
Smoking: past/current, n (%)	347 (52.3)	65 (59.1)	0.22	774
COPD, n (%)	30 (4.3)	2 (1.7)	0.209	813
Asthma, n (%)	60 (8.6)	14 (11.7)	0.365	817
NTproBNP [ng/l]	1176 [342;2720]	1200 [503;2934]	0.404	279
BNP [ng/l]	164 [60.2;412]	307 [113;554]	0.021	199
Initial therapy: n (%)			<0.001	630
CCB	45 (8.6)	9 (8.5)		
combination therapy	241 (46.0)	82 (77.4)		
ERA	71 (13.5)	1 (0.9)		
PA	21 (4.0)	0 (0.0)		
PDE5	145 (27.7)	14 (13.2)		
sGC	1 (0.2)	0 (0.0)		

Data are presented as median [interquartile range] unless otherwise specified. Abbreviations: ERA: endothelin receptor antagonists; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; I/HPAH: idiopathic/hereditary pulmonary arterial hypertension; L/min: liters per minute; mPAP: mean pulmonary arterial pressure; mPAWP: mean pulmonary artery wedge pressure; mRAP: mean right atrial pressure; BNP: brain natriuretic peptide; NTpro-BNP: N-terminal pro-brain natriuretic peptide; PDE5: phosphodiesterase type 5; PVR: pulmonary vascular resistance; sGC: Soluble Guanylate Cyclase Stimulator; TLC: total lung capacity; WHO FC: World Health Organization functional class; WU: Wood units; 6MWD: six-minute walk distance; SpO_2: arterial oxygen saturation; CO: cardiac output; COPD: chronic obstructive pulmonary disease; CCB: calcium channel blockers; PA: prostacyclin analogue

Table B.20 The number of patients who had the acute vasoreactivity test performed by diagnosis and response status.

Diagnosis	NO challenge not done	NO challenge done	Non-responder	Vasoresponder*
IPAH	707	484	400	84
HPAH	44	47	45	2
PVOD/PCH	26	9	8	1
APAH	29	6	6	0
PH-LHD	8	1	1	0
PH-LD	15	2	2	0
CTEPH	7	1	1	0
Multifactorial PH	7	0	0	0
HHT	1	0	0	0

*lenient criteria used; Abbreviations: I/HPAH: idiopathic/hereditary pulmonary arterial hypertension; PVOD/PCH: pulmonary veno-occlusive disease/pulmonary capillary hemangiomatosis; APAH: associated pulmonary arterial hypertension; PH-LHD: pulmonary hypertension due to left heart disease, PH-LD: pulmonary hypertension due to lung disease and/or hypoxia; CTEPH: chronic thromboembolic pulmonary hypertension; HHT: hereditary hemorrhagic telangiectasia

Table B.21 Reported autoantibody screen results.

	All participants		I/HPAH/PVOD/PCH	
Positive autoimmune screen:		970		928
no	779 (80.3)		754 (81.2)	
yes	191 (19.7)		174 (18.8)	
ANA Ab:		917		879
negative	775 (84.5)		749 (85.2)	
positive	142 (15.5)		130 (14.8)	
Anti-ENA Ab:		519		495
negative	506 (97.5)		484 (97.8)	
positive	13 (2.50)		11 (2.22)	
Anti-dsDNA Ab:		472		452
negative	462 (97.9)		443 (98.0)	
positive	10 (2.12)		9 (1.99)	
Anti-Scl70 Ab:		413		392
negative	412 (99.8)		391 (99.7)	
positive	1 (0.24)		1 (0.26)	
Anti-centromere Ab:		410		395
negative	407 (99.3)		392 (99.2)	
positive	3 (0.73)		3 (0.76)	
Anti-Rho Ab:		382		362
negative	368 (96.3)		352 (97.2)	
positive	14 (3.66)		10 (2.76)	
ANCA:		454		431
negative	425 (93.6)		405 (94.0)	
positive	29 (6.39)		26 (6.03)	
Anti-cardiolipin Ab:		427		405
negative	416 (97.4)		395 (97.5)	
positive	11 (2.58)		10 (2.47)	

Data presented as number and percentage. Abbreviations: ANA: antinuclear antibodies; ENA: extractable nuclear antibodies; dsDNA: double stranded DNA; Scl-70: topoisomerase I; ANCA: anti-neutrophil antibodies; I/HPAH: idiopathic/hereditary pulmonary arterial hypertension; PVOD/PCH: pulmonary veno-occlusive disease/pulmonary capillary haemangiomatosis.

Table B.22 Autoimmune screening by in house immunofluorescence.

	[ALL] N = 473
Diagnosis:	
HPAH	46 (9.73)
IPAH	412 (87.1)
PCH	1 (0.21)
PH-LHD	1 (0.21)
PH-LD	3 (0.63)
PVOD	10 (2.11)
Autoimmunity	127 (26.8)
ANA	103 (21.8)
ENA	19 (4.02)
Anti-histone H2 and H4 Ab	0 (0.0)
Anti-histone IIA Ab:	42 (8.88)
Anti-centromere Ab:	11 (2.33)
Anti-Jo-1 Ab:	4 (0.85)
Anti-La/SS-B Ab:	7 (1.48)
Anti-Ro/SS-A Ab:	6 (1.27)
Anti-U1-snRNP Ab:	0 (0.0)
Anti-RNP complex Ab:	29 (6.13)
Anti-Smith (Sm) Ab:	4 (0.85)
Anti-Scl-70 Ab:	2 (0.42)
Anti-Scl34 Ab:	24(5.07)
Anti-cardiolipin Ab:	2 (0.42)
Anti-Mi2 Ab:	2 (0.42)
Anti-myeloperoxidase Ab:	14 (2.96)
Anti-proteinase-3 Ab:	1 (0.21)
Anti-mitochondrial M2 (Anti-pyruvate dehydrogenase Ab): Ab	16 (3.38)
Anti-thyroglobulin Ab:	8 (1.69)
Anti-thyroid peroxidase Ab:	0 (0.0)
Anti-transglutaminase Ab:	2 (0.42)

Data presented as number and percentage. Abbreviations: I/HPAH: idiopathic/hereditary pulmonary arterial hypertension; PVOD/PCH: pulmonary veno-occlusive disease/pulmonary capillary hemangiomatosis; PH-LHD: pulmonary hypertension due to left heart disease, PH-LD: pulmonary hypertension due to lung disease and/or hypoxia; ANA: antinuclear antibodies; ENA: extractable nuclear antibodies; Ab: antibodies

Table B.23 Clinical characteristics of patients in clusters selected for BeviMed analysis.

Cluster*, N	**Cluster 1 N=33**	**Cluster 2 N=11**	**Cluster 5 N=19**	**Cluster 6 N=90**	**Cluster 24 N=18**	**Cluster 27 N=16**	**Cluster 28 N=66**	**Cluster 29 N=36**	**Cluster 33 N=15**	**Cluster 34 N=18**	**Cluster 42 N=23**	**Cluster 44 N=10**	**Cluster 47 N=19**	**Cluster 50 N=17**	**Cluster 58 N=15**	**Cluster 61 N=14**	**Cluster 63 N=28**	**Cluster 64 N=23**	**Cluster 66 N=7**	**Cluster 86 N=14**	**Cluster 92 N=18**
WGS: n (%)	26 (78.8)	9 (81.8)	18 (94.7)	63 (70.0)	13 (72.2)	14 (87.5)	61 (92.4)	34 (94.4)	14 (93.3)	12 (66.7)	22 (95.7)	10 (100)	16 (84.2)	13 (76.5)	15 (100)	13 (92.9)	21 (75.0)	19 (82.6)	7 (100)	10 (71.4)	14 (77.8)
Diagnosis: n (%)																					
HPAH	1 (3.03)	0 (0.00)	2 (10.5)	4 (4.44)	0 (0.00)	0 (0.00)	5 (7.58)	3 (8.33)	0 (0.00)	1 (5.56)	4 (17.4)	0 (0.00)	0 (0.00)	1 (5.88)	2 (13.3)	2 (14.3)	0 (0.00)	4 (17.4)	0 (0.00)	0 (0.00)	0 (0.00)
IPAH	32 (97.0)	11 (100)	16 (84.2)	77 (85.6)	18 (100)	16 (100)	49 (74.2)	32 (88.9)	15 (100)	17 (94.4)	19 (82.6)	10 (100)	18 (94.7)	16 (94.1)	13 (86.7)	12 (85.7)	25 (89.3)	15 (65.2)	6 (85.7)	14 (100)	17 (94.4)
PCH	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
PVOD	0 (0.00)	0 (0.00)	0 (0.00)	7 (7.78)	0 (0.00)	0 (0.00)	4 (6.06)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	3 (10.7)	3 (13.0)	0 (0.00)	0 (0.00)	0 (0.00)
Age [years]	37 [32;50]	49 [41;53]	46 [42;55]	68 [57;74]	67 [56;75]	37 [29;44]	55 [42;65]	68 [64;73]	47 [33;59]	5 [3;12]	33 [29;41]	70 [57;73]	68 [60;73]	37 [27;49]	32 [27;47]	33 [29;36]	64 [57;72]	47 [35;55]	32 [28;44]	37 [30;42]	71 [67;75]
Sex: female, n (%)	23 (70)	9 (82)	15 (79)	54 (60)	7 (39)	14 (88)	37 (57)	19 (53)	11 (73)	13 (72)	22 (96)	7 (70)	12 (63)	13 (76)	9 (60)	8 (57)	21 (75)	19 (83)	4 (57)	6 (43)	8 (44)
Ethnicity: Eur, n (%)	23 (69.7)	10 (90.9)	16 (84.2)	80 (88.9)	17 (94.4)	12 (75.0)	53 (80.3)	33 (91.7)	11 (73.3)	15 (83.3)	17 (73.9)	8 (80.0)	12 (63.2)	13 (76.5)	11 (73.3)	11 (78.6)	21 (75.0)	20 (87.0)	6 (85.7)	14 (100)	16 (88.9)
Drug exposure:	1 (3.03)	0 (0.00)	1 (5.26)	2 (2.22)	3 (16.7)	1 (6.25)	1 (1.52)	0 (0.00)	0 (0.00)	0 (0.00)	4 (17.4)	2 (20.0)	3 (15.8)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	6 (26.1)	0 (0.00)	2 (14.3)	0 (0.00)
BMI [kg/m^2]	24.1 [21.8;26.8]	25.6 [24.8;27.9]	30.0 [25.8;33.9]	28.9 [25.7;32.4]	27.5 [24.3;29.8]	25.6 [22.0;28.6]	27.1 [22.9;32.7]	30.0 [27.4;34.5]	25.6 [24.5;30.1]	16.3 [15.1;17.7]	34.0 [27.6;37.4]	30.3 [24.1;31.8]	31.0 [21.4;37.6]	26.1 [23.3;28.2]	25.6 [22.8;26.5]	26.2 [23.8;28.7]	26.2 [23.9;29.1]	28.5 [24.2;32.2]	22.5 [20.3;29.1]	23.8 [22.0;27.2]	28.9 [26.1;30.1]
WHO FC: I/II/III/IV, %	6/24/64/6	0/11/67/22	0/35/35/29	0/17/72/11	0/11/78/11	0/29/57/14	2/23/65/11	0/0/88/12	0/23/62/15	17/39/33/11	0/14/64/23	0/10/90/0	0/11/74/16	0/35/41/24	0/0/80/20	0/7/79/14	0/4/81/15	0/14/76/10	0/29/57/14	8/31/54/8	0/17/83/0
6MWD [m]	396 [342;460]	373 [297;445]	390 [364;434]	223 [149;308]	374 [116;414]	374 [310;490]	350 [246;401]	268 [222;311]	300 [190;503]	446 [384;482]	285 [192;393]	297 [208;361]	210 [106;355]	400 [348;493]	354 [340;405]	418 [233;548]	199 [96.0;308]	288 [93.0;405]	380 [334;420]	458 [342;509]	282 [149;327]
SpO_2 pre [%]	97.0 [94.0;98.0]	96.0 [93.0;98.0]	98.0 [97.0;98.0]	93.0 [90.0;95.0]	94.5 [91.5;96.2]	96.0 [94.0;97.0]	95.0 [91.0;97.0]	93.5 [88.0;95.2]	96.0 [92.5;97.5]	97.0 [97.0;98.0]	98.0 [93.8;98.0]	95.0 [95.0;98.0]	95.0 [92.5;97.0]	98.0 [97.0;99.0]	98.0 [96.0;98.0]	96.0 [94.2;97.0]	93.5 [89.2;96.0]	96.0 [95.0;97.0]	97.0 [96.0;98.5]	97.0 [97.0;98.0]	94.0 [90.0;95.2]

Table B.23 Clinical characteristics of patients in clusters selected for BeviMed analysis.

Cluster*, N	**Cluster 1 N=33**	**Cluster 2 N=11**	**Cluster 5 N=19**	**Cluster 6 N=90**	**Cluster 24 N=18**	**Cluster 27 N=16**	**Cluster 28 N=66**	**Cluster 29 N=36**	**Cluster 33 N=15**	**Cluster 34 N=18**	**Cluster 42 N=23**	**Cluster 44 N=10**	**Cluster 47 N=19**	**Cluster 50 N=17**	**Cluster 58 N=15**	**Cluster 61 N=14**	**Cluster 63 N=28**	**Cluster 64 N=23**	**Cluster 66 N=7**	**Cluster 86 N=14**	**Cluster 92 N=18**
SpO_2 post [%]	90.5 [85.0;96.0]	91.0 [87.0;94.0]	95.0 [93.0;96.0]	86.0 [81.0;91.8]	89.0 [85.8;93.2]	94.0 [90.2;95.0]	90.0 [82.0;94.0]	87.0 [77.5;94.0]	93.0 [83.5;95.0]	97.0 [93.0;98.0]	96.5 [94.0;99.0]	92.0 [89.0;93.0]	93.0 [84.0;96.0]	96.0 [94.0;98.0]	97.0 [95.5;97.5]	92.0 [87.0;92.0]	87.0 [83.5;91.5]	93.0 [91.0;95.0]	86.5 [83.8;89.8]	95.0 [88.5;97.0]	88.0 [82.0;89.0]
mRAP [mmHg]	8 [3;11]	9 [7;12]	8 [5;12]	9 [6;12]	8 [5;10]	9 [6;13]	8 [6;11]	10 [8;14]	7 [4;9]	6 [5;7]	12 [8;17]	7 [6;9]	10 [7;17]	6 [5;8]	8 [6;10]	12 [8;16]	8 [6;12]	8 [7;11]	5 [4;6]	4 [2;7]	6 [3;8]
mPAP [mmHg]	56 [44;65]	56 [54;58]	49 [42;58]	50 [42;58]	44 [39;52]	58 [41;66]	53 [44;58]	53 [45;64]	55 [49;59]	38 [29;58]	55 [48;66]	48 [37;62]	50 [40;58]	49 [42;54]	56 [51;71]	57 [54;66]	53 [46;58]	44 [40;52]	56 [54;58]	52 [44;56]	46 [38;52]
mPAWP [mmHg]	8 [6;11]	7 [5;10]	8 [7;10]	10 [8;12]	8 [7;10]	10 [8;12]	10 [7;13]	11 [9;14]	8 [6;11]	9 [7;12]	10 [8;13]	8 [6;10]	10 [9;12]	6 [5;6]	8 [6;9]	10 [7;11]	9 [6;12]	9 [8;13]	7 [6;8]	7 [6;9]	10 [7;12]
CO [L/min]	4.1 [3.4;5.4]	4.0 [3.3;5.2]	3.6 [3.1;4.9]	3.9 [3.2;4.8]	3.9 [3.0;4.5]	3.6 [2.9;4.0]	4.2 [3.1;4.8]	4.0 [3.4;5.1]	4.7 [3.2;5.2]	3.0 [1.6;3.6]	3.5 [2.5;4.0]	4.6 [3.2;5.7]	3.8 [2.7;5.1]	4.0 [3.0;4.7]	3.5 [3.0;4.4]	3.2 [2.9;4.6]	3.5 [2.7;4.6]	4.1 [3.4;4.8]	4.3 [3.3;4.9]	5.3 [4.2;6.2]	3.8 [3.2;5.1]
SvO2 [%]	68.0 [58.4;72.0]	65.0 [62.0;67.8]	68.3 [61.6;74.5]	63.0 [57.2;70.0]	61.8 [57.1;68.2]	64.5 [61.5;70.1]	64.0 [61.0;70.5]	60.6 [57.0;66.0]	64.0 [59.1;72.4]	70.0 [64.1;72.7]	63.0 [51.7;66.2]	65.4 [65.0;66.0]	62.0 [57.4;66.9]	70.0 [66.9;72.0]	69.5 [62.4;72.2]	60.3 [51.9;66.7]	62.7 [58.9;68.6]	66.0 [59.4;70.5]	70.0 [69.1;76.2]	70.8 [67.8;73.2]	66.3 [62.2;68.0]
FEV1 [% pred.]	96.8 [78.7;104]	94.5 [84.8;98.8]	83.0 [76.2;90.5]	84.0 [73.0;98.0]	90.0 [88.1;97.0]	86.2 [78.8;92.5]	84.0 [70.0;92.8]	90.0 [74.0;98.0]	81.0 [75.0;101]	91.0 [78.2;97.0]	88.0 [80.0;97.1]	102 [86.0;113]	77.0 [53.8;92.5]	97.7 [83.5;107]	92.5 [84.8;101]	97.2 [78.8;101]	85.4 [79.3;91.2]	88.0 [74.0;97.0]	117 [104;120]	93.8 [88.9;103]	91.0 [80.0;98.1]
FVC [% pred.]	97.0 [88.8;108]	96.0 [83.0;115]	89.0 [81.8;107]	100 [88.0;112]	101 [94.0;103]	96.3 [85.2;102]	93.0 [80.0;100]	96.5 [85.2;103]	85.5 [78.8;95.5]	94.5 [83.8;102]	93.0 [83.5;104]	110 [105;132]	82.8 [65.8;103]	102 [85.7;117]	98.5 [89.2;106]	108 [89.2;113]	102 [96.5;112]	94.0 [79.0;105]	115 [110;118]	104 [96.7;118]	99.0 [89.0;115]
TLC [% pred.]	99.0 [90.6;108]	102 [95.9;104]	96.0 [80.0;102]	94.0 [87.0;101]	96.1 [91.0;106]	101 [90.2;108]	93.0 [86.0;103]	93.6 [90.1;106]	93.0 [87.2;95.8]	117 [117;117]	95.0 [86.5;102]	104 [93.2;112]	90.0 [73.0;93.5]	105 [95.0;112]	97.5 [86.4;108]	109 [97.0;112]	100 [85.8;101]	88.0 [77.0;95.0]	125 [122;128]	103 [98.5;113]	93.0 [81.4;108]
KCO [% pred.]	77 [70;82]	72 [60;84]	90 [85;95]	61 [37;76]	64 [41;79]	72 [70;76]	68 [45;87]	62 [45;92]	92 [80;100]	64 [64;64]	78 [71;85]	82 [59;90]	68 [44;82]	70 [63;76]	80 [72;92]	74 [70;76]	37 [27;47]	72 [62;83]	77 [76;77]	72 [58;77]	49 [41;63]
Emphysema: n (%)																					
minimal/ mild	0 (0.00)	1 (20.0)	1 (8.33)	8 (12.3)	0 (0.00)	0 (0.00)	4 (8.70)	5 (16.7)	1 (11.1)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	1 (7.14)	0 (0.00)	1 (6.67)	1 (5.26)	0 (0.00)	1 (10.0)	3 (25.0)
moderate	0 (0.00)	0 (0.00)	0 (0.00)	6 (9.23)	0 (0.00)	0 (0.00)	2 (4.35)	1 (3.33)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	1 (6.67)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
severe	0 (0.00)	0 (0.00)	0 (0.00)	2 (3.08)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	1 (6.67)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)

Table B.23 Clinical characteristics of patients in clusters selected for BeviMed analysis.

Cluster*, N	**Cluster 1 N=33**	**Cluster 2 N=11**	**Cluster 5 N=19**	**Cluster 6 N=90**	**Cluster 24 N=18**	**Cluster 27 N=16**	**Cluster 28 N=66**	**Cluster 29 N=36**	**Cluster 33 N=15**	**Cluster 34 N=18**	**Cluster 42 N=23**	**Cluster 44 N=10**	**Cluster 47 N=19**	**Cluster 50 N=17**	**Cluster 58 N=15**	**Cluster 61 N=14**	**Cluster 63 N=28**	**Cluster 64 N=23**	**Cluster 66 N=7**	**Cluster 86 N=14**	**Cluster 92 N=18**
Fibrosis: n (%)																					
minimal/ mild	0 (0.00)	0 (0.00)	0 (0.00)	7 (10.9)	0 (0.00)	1 (10.0)	4 (8.70)	5 (16.7)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	1 (8.33)
moderate	0 (0.00)	0 (0.00)	0 (0.00)	1 (1.56)	0 (0.00)	0 (0.00)	1 (2.17)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
severe	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	1 (2.17)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)	0 (0.00)
PE: n (%)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	3 (27)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)	1 (8)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
NTproBNP [ng/l]	1518 [1157;2655]	1269 [1062;2207]	649 [247;2053]	1298 [397;3076]	711 [381;1920]	1467 [396;3487]	883 [310;2343]	1425 [260;3210]	1479 [1343;1615]	430 [168;5020]	3452 [2039;5166]	1828 [329;3446]	6412 [5062;7763]	719 [530;972]	587 [469;783]	1411 [778;2044]	1190 [100;4413]	2184 [646;3091]	10008 [-]	508 [138;943]	1627 [1075;2482]
CRP [mg/l]	2.95 [1.25;4.90]	5.00 [4.50;7.50]	5.00 [5.00;12.3]	5.25 [2.05;11.0]	4.90 [2.00;9.00]	1.50 [1.00;2.88]	4.60 [3.18;7.95]	5.45 [2.45;11.5]	6.00 [4.80;12.3]	5.00 [5.00;7.00]	5.30 [4.25;11.2]	4.30 [2.95;5.75]	6.85 [2.85;9.75]	3.00 [2.00;5.00]	2.00 [1.15;4.25]	4.00 [1.40;11.0]	5.40 [3.00;9.00]	6.00 [2.50;13.0]	4.00 [2.50;16.5]	2.45 [1.75;3.75]	4.00 [2.75;5.00]
Creatinine [μmol/l]	102 [70.5;116]	82.0 [79.0;90.0]	79.5 [67.8;93.8]	97.5 [84.2;111]	107 [93.2;122]	80.0 [70.5;86.0]	94.0 [79.2;106]	98.0 [84.5;120]	84.0 [73.0;91.0]	46.0 [36.0;53.0]	91.0 [67.0;99.0]	90.0 [87.2;92.5]	112 [97.5;128]	74.0 [69.2;85.4]	84.0 [78.0;85.5]	94.5 [79.8;100]	91.5 [76.8;118]	75.0 [69.0;86.0]	88.0 [72.2;104]	74.0 [68.0;100]	90.0 [78.5;112]

Abbreviations: Eur: European; I/HPAH: idiopathic/hereditary pulmonary arterial hypertension; PVOD/PCH: pulmonary veno-occlusive disease/pulmonary capillary hemangiomatosis; KCO: transfer coefficient for carbon monoxide; mPAP: mean pulmonary arterial pressure; mPAWP: mean pulmonary artery wedge pressure; mRAP: mean right atrial pressure; TLC: total lung capacity; WHO FC: World Health Organization functional class; 6MWD: six-minute walk distance; SpO_2: arterial oxygen saturation; CO: cardiac output; CRP: C-reactive protein; NTproBNP: N-terminal pro B-type natriuretic peptide; WGS: whole-exome sequencing; BMI: body mass index; SvO_2: mixed venous oxygen saturation; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; PE: pulmonary embolism;*cluster devised by affinity propagation clustering algorithm

Table B.24 Cophenetic correlation coefficient and agglomerative coefficient for various linkage methods - agglomerative hierarchical clustering.

	Linkage method			
	Average	**Single**	**Complete**	**Ward's**
Cophenetic correlation coefficient*	0.43	-0.01	0.02	0.04
Agglomerative correlation coefficient	0.37	0.77	0.61	0.55

*cophenetic correlation coefficient is a correlation between distance matrix and cophenetic matrix

Table B.25 Clinical differences between patients with present and missing transfer coefficient results.

	[ALL]	KCO [% pred.] missing N = 478	KCO [% pred.] present N = 644	*p*-value	N
	N=1122				
Sex: Female, n (%)	760 (68.1)	326 (69.1)	434 (67.4)	0.597	1116
Prevalent cases, n (%)	852 (75.9)	411 (86.0)	441 (68.5)	<0.001	1122
BMI [kg/m^2]	26.9 [23.1;31.5]	25.6 [22.0;30.1]	27.7 [24.1;32.5]	<0.001	1015
Age [years]	48.8 [34.8;62.7]	44.6 [31.0;58.5]	51.3 [38.1;65.5]	<0.001	1112
WHO FC: I/II/III/IV, %	2/20/65/13	3/26/60/11	1/16/69/14	<0.001	1079
6MWD [m]	340 [230;418]	364 [289;432]	314 [192;405]	<0.001	702
FEV1 [% pred.]	85.0 [73.0;97.0]	83.2 [69.6;97.0]	86.0 [74.0;97.0]	0.2	849
FVC [% pred.]	94.0 [81.4;106]	89.0 [73.0;103]	96.0 [82.6;107]	<0.001	831
TLC [% pred.]	95.0 [85.0;104]	93.2 [83.7;104]	95.0 [86.0;103]	0.625	639
mRAP [mmHg]	8.00 [5.00;12.0]	7.00 [5.00;12.0]	9.00 [6.00;12.0]	<0.001	985
mPAP [mmHg]	53.0 [44.0;61.0]	53.0 [44.0;61.0]	53.0 [45.0;61.0]	0.669	1052
PAWP [mmHg]	9.00 [7.00;12.0]	9.00 [6.00;11.0]	10.0 [7.00;12.0]	0.001	934
CI [L/min/m^2]	2.17 [1.72;2.67]	2.31 [1.81;2.82]	2.07 [1.68;2.59]	<0.001	946
PVR [WU]	11.0 [7.69;15.1]	10.6 [7.60;15.2]	11.1 [7.69;15.1]	0.622	893
SvO_2 [%]	64.0 [58.0;70.0]	64.7 [57.8;70.0]	63.8 [58.2;70.0]	0.732	817
Fibrosis: n (%)				0.445	614
none	586 (95.4)	169 (96.0)	417 (95.2)		
minimal/mild	26 (4.23)	6 (3.41)	20 (4.57)		
moderate	1 (0.16)	0 (0.00)	1 (0.23)		
severe	1 (0.16)	1 (0.57)	0 (0.00)		
Emphysema: n (%)				0.029	612
none	560 (91.5)	169 (96.0)	391 (89.7)		
minimal/mild	33 (5.39)	3 (1.70)	30 (6.88)		
moderate	15 (2.45)	4 (2.27)	11 (2.52)		
severe	4 (0.65)	0 (0.00)	4 (0.92)		
Smoking: past/current, n (%)	435 (50.6)	97 (38.5)	338 (55.7)	<0.001	859
COPD,n (%)	66 (5.88)	24 (5.02)	42 (6.52)	0.353	1122
OSA, n (%)	58 (5.17)	20 (4.18)	38 (5.90)	0.251	1122
Asthma, n (%)	78 (6.95)	15 (3.14)	63 (9.78)	<0.001	1122

Abbreviations: BMI: body mass index; WHO FC: World Health Organization Functional Class; 6MWD: 6-minute walk distance; mRAP: mean right atrial pressure; mPAP: mean pulmonary artery pressure; PAWP: Pulmonary Artery Wedge Pressure; CI: cardiac index; PVR: pulmonary vascular resistance; SvO_2: mixed venous saturation; FEV1: forced expiratory capacity in 1 second; FVC: forced vital capacity; TLC: total lung capacity; COPD: chronic obstructive pulmonary disease; OSA: obstructive sleep apnoea

Table B.26 Clinical characteristics of unrelated individuals used in gene-tag association analysis by KCO threshold.

	[ALL] N = 644	KCO>50% pred N = 492	KCO≤50 % pred. N = 152	*p*-value	N
Age [years]	51 [38;66]	47 [36;60]	66 [54;71]	<0.001	644
Sex: female, n (%)	434 (67)	352 (72)	82 (54)	<0.001	644
Incident cases, n (%)	203 (31.5)	134 (27.2)	69 (45.4)	<0.001	644
BMI [kg/m^2]	27.7 [24.1;32.5]	27.7 [23.7;32.7]	27.7 [24.8;32.3]	0.806	629
WHO FC: I/II/III/IV, %	1/16/69/14	1/18/67/13	1/8/77/17	0.013	628
6MWD [m]	313 [190;404]	334 [229;414]	219 [120;348]	<0.001	599
SpO_2 pre [%]	95.0 [93.0;97.0]	96.0 [93.0;98.0]	92.0 [89.0;95.0]	<0.001	575
SpO_2 post [%]	91.0 [85.0;96.0]	93.0 [88.0;96.0]	83.0 [76.0;88.0]	<0.001	529
mRAP [mmHg]	9 [6;12]	9 [6;13]	9 [6;12]	0.375	601
mPAP [mmHg]	53 [45;61]	54 [46;63]	50 [42;57]	0.001	625
PAWP [mmHg]	10 [7;12]	10 [7;12]	10 [8;12]	0.301	560
CO [L/min]	3.8 [3.1;4.8]	3.9 [3.1;4.9]	3.6 [3.1;4.7]	0.282	614
SvO_2 [%]	64 [58;70]	64 [59;71]	61 [55;67]	<0.001	572
Acute NO challenge: (vr)	43 (17)	38 (18)	5 (10)	0.204	257
FEV1 [% pred.]	86.0 [74.0;97.0]	85.0 [73.4;96.0]	87.0 [77.0;98.0]	0.119	639
FVC [% pred.]	96.0 [82.6;107]	94.0 [81.0;106]	101 [87.0;113]	<0.001	628
FEV1/FVC ratio	0.76 [0.69;0.81]	0.76 [0.71;0.81]	0.70 [0.63;0.77]	<0.001	614
TLC [% pred.]	95.0 [86.0;103]	95.0 [85.0;103]	95.0 [87.5;104]	0.564	485
KCO [%pred.]	71 [52;86]	78 [67;90]	37 [30;44]	<0.001	644
Emphysema, n (%)				<0.001	436
none	391 (90)	307 (95)	84 (74)		
minimal/mild	30 (7)	12 (4)	18 (16)		
moderate	11 (3)	2 (1)	9 (8)		
severe	4 (1)	1 (0)	3 (3)		
Fibrosis, n (%)				<0.001	438
none	417 (95)	319 (98)	98 (87)		
minimal/mild	20 (5)	6 (2)	14 (12)		
moderate	1 (0)	0 (0)	1 (1)		
Smoking: past/current, n (%)	338 (55.7)	233 (50.0)	105 (74.5)	<0.001	607
NTproBNP [ng/l]	980 [248;2673]	866 [257;2590]	1225 [249;2878]	0.359	220
BNP [ng/l]	200 [72.9;432]	198 [69.7;456]	200 [82.2;328]	0.798	143
Uric acid [mmol/l]	0.42 [0.32;0.53]	0.40 [0.30;0.51]	0.48 [0.39;0.56]	0.002	231
CRP [mg/l]	5.00 [2.00;8.60]	5.00 [2.00;8.00]	4.15 [2.00;9.00]	0.953	450
Hb [g/l]	154 [139;166]	153 [138;166]	154 [142;166]	0.658	603
WBC [x10e9/l]	8 [7;10]	8 [7;10]	9 [7;10]	0.011	597
Platelets [x10e9/l]	220 [179;268]	220 [179;270]	220 [182;254]	0.516	595
Sodium [mmol/l]	139 [138;141]	139 [138;141]	140 [138;141]	0.858	597
Potassium [mmol/l]	4.20 [4.00;4.50]	4.20 [3.90;4.50]	4.20 [4.00;4.50]	0.412	591
Urea [mmol/l]	5.80 [4.60;7.70]	5.60 [4.40;7.10]	7.00 [5.30;9.20]	<0.001	594
Creatinine [μmol/l]	88.0 [73.2;105]	87.0 [73.0;102]	95.0 [78.5;114]	0.002	598
COPD, n (%)	42 (7)	23 (5)	19 (12)	0.001	644
Asthma, n (%)	63 (10)	54 (11)	9 (6)	0.093	644
OSA, n (%)	38 (6)	29 (6)	9 (6)	1	644
CAD, n (%)	26 (4)	10 (2)	16 (11)	<0.001	644
CVA, n (%)	10 (2)	5 (1)	5 (3)	0.061	644
PAD, n (%)	2 (0)	0 (0)	2 (1)	0.055	644
HTN, n (%)	170 (26)	116 (24)	54 (36)	0.005	644
DM type 1, n (%)	9 (1)	8 (2)	1 (1)	0.693	644

Table B.26 Clinical characteristics of unrelated individuals used in gene-tag association analysis by KCO threshold.

	[ALL] N = 644	**KCO>50% pred N = 492**	**KCO≤50 % pred. N = 152**	***p*-value**	**N**
DM type 2, n (%)	93 (14)	60 (12)	33 (22)	0.005	644
Hypothyroidism, n (%)	73 (11)	60 (12)	13 (9)	0.275	644
Sjogren syndrome, n (%)	3 (0)	2 (0)	1 (1)	0.555	644

None of the patients had systemic lupus erythematosus, systemic sclerosis, undifferentiated connective tissue disease or ankylosing spondylitis. Abbreviations: BMI: body mass index, WHO FC: World Health Organization Functional Class, 6MWD: 6-minute walk distance, SpO_2: peripheral capillary oxygen saturation, mRAP: mean right atrial pressure, mPAP: mean pulmonary artery pressure, PAWP: Pulmonary Artery Wedge Pressure, CO: cardiac output, SvO_2: Mixed venous oxygen saturation, FEV1: forced expiratory capacity in 1 second, FVC: forced vital capacity, TLC: Total Lung Capacity, KCO: transfer coefficient for carbon monoxide, NTproBNP: N-terminal pro B-Type Natriuretic Peptide, BNP: B-Type Natriuretic Peptide, CRP: C-Reactive Protein Protein, Hb: hemoglobin, WBC: white blood cell count, COPD: chronic obstructive pulmonary disease, OSA: obstructive sleep apnoea, CAD: coronary artery disease, CVA: cerebrovascular accident, PAD: peripheral artery disease, HTN: hypertension, DM: diabetes mellitus, vr: vasoresponder

Table B.27 Clinical characteristics of study individuals unrelated individuals used in gene-tag association analysis by KCO tertiles.

	[ALL] N=644	Higher tertile N=214	Middle tertile N=215	Lower tertile N=215	*p*-value	N
Age [years]	51 [38;66]	44 [37;58]	49 [34;61]	64 [50;71]	<0.001	644
Sex: female, n (%)	434 (67)	149 (70)	160 (74)	125 (58)	0.001	644
Incident cases, n (%)	203 (31.5)	58 (27.1)	54 (25.1)	91 (42.3)	<0.001	644
BMI [kg/m^2]	27.7 [24.1;32.5]	29.1 [25.3;34.5]	26.8 [23.1;30.9]	27.4 [24.2;32.3]	<0.001	629
WHO FC: I/II/III/IV, %	1/16/69/14	2/19/68/11	1/19/65/15	1/10/74/15	0.028	628
6MWD [m]	313 [190;404]	331 [240;420]	340 [230;418]	240 [131;360]	<0.001	599
SpO_2 pre [%]	95.0 [93.0;97.0]	96.0 [94.0;98.0]	96.0 [93.0;98.0]	93.0 [90.0;96.0]	<0.001	575
SpO_2 post [%]	91.0 [85.0;96.0]	94.0 [89.0;96.0]	93.0 [88.0;96.0]	85.0 [79.0;91.0]	<0.001	529
mRAP [mmHg]	9 [6;12]	9 [7;13]	9 [6;12]	8 [5;12]	0.224	601
mPAP [mmHg]	53 [45;61]	55 [47;65]	53 [46;62]	51 [42;57]	<0.001	625
PAWP [mmHg]	10 [7;12]	10 [8;12]	10 [7;12]	10 [7;12]	0.487	560
CO [L/min]	3.8 [3.1;4.8]	3.8 [3.0;5.0]	4.0 [3.1;4.8]	3.8 [3.1;4.8]	0.967	614
SvO_2 [%]	64 [58;70]	64 [60;71]	65 [59;70]	62 [56;68]	0.001	572
Acute NO challenge: (vr)	43 (17)	15 (17)	21 (24)	7 (9)	0.024	257
FEV1 [% pred.]	86.0 [74.0;97.0]	83.8 [72.9;95.0]	86.0 [73.9;97.4]	87.0 [77.0;98.0]	0.08	639
FVC [% pred.]	96.0 [82.6;107]	90.0 [79.8;101]	95.9 [83.0;108]	100 [86.9;112]	<0.001	628
FEV1/FVC ratio	0.76 [0.69;0.81]	0.78 [0.72;0.82]	0.76 [0.69;0.81]	0.71 [0.65;0.78]	<0.001	614
TLC [% pred.]	95.0 [86.0;103]	94.0 [86.2;103]	96.0 [83.0;104]	97.0 [87.0;102]	0.787	485
KCO [% pred.]	71 [52;86]	92 [86;101]	71 [67;76]	42 [33;52]	<0.001	644
Emphysema: n (%)					<0.001	436
none	391 (90)	144 (99)	128 (96)	119 (76)		
minimal/mild	30 (7)	2 (1)	4 (3)	24 (15)		
moderate	11 (3)	0 (0)	0 (0)	11 (7)		
severe	4 (1)	0 (0)	1 (1)	3 (2)		
Fibrosis: n (%)					<0.001	438
none	417 (95)	145 (99)	132 (99)	140 (89)		
minimal/mild	20 (5)	2 (1)	2 (1)	16 (10)		
moderate	1 (0)	0 (0)	0 (0)	1 (1)		
Smoking: past/current, n (%)	338 (55.7)	89 (42.4)	106 (52.7)	143 (73.0)	<0.001	607
NTproBNP [ng/l]	980 [248;2673]	842 [167;2358]	902 [310;2593]	1225 [237;2811]	0.539	220
BNP [ng/l]	200 [72.9;432]	193 [71.7;392]	145 [67.5;427]	214 [82.2;448]	0.659	143
Uric acid [mmol/l]	0.42 [0.32;0.53]	0.40 [0.28;0.47]	0.40 [0.32;0.51]	0.48 [0.38;0.55]	0.011	231
CRP [mg/l]	5.00 [2.00;8.60]	5.00 [2.00;8.00]	5.00 [2.00;9.32]	4.00 [2.00;8.57]	0.496	450
Hb [g/l]	154 [139;166]	160 [145;169]	149 [136;164]	151 [140;165]	<0.001	603
WBC [x10e9/l]	8 [7;10]	8 [7;9]	8 [7;10]	9 [7;10]	0.005	597
Platelets [x10e9/l]	220 [179;268]	224 [183;262]	219 [174;276]	217 [184;262]	0.89	595
Sodium [mmol/l]	139 [138;141]	140 [138;141]	139 [137;141]	140 [138;141]	0.623	597
Potassium [mmol/l]	4.20 [4.00;4.50]	4.30 [4.00;4.50]	4.20 [3.90;4.40]	4.30 [4.00;4.50]	0.13	591
Urea [mmol/l]	5.80 [4.60;7.70]	5.50 [4.43;7.00]	5.70 [4.30;7.10]	6.70 [5.15;8.80]	<0.001	594
Creatinine [μmol/l]	88.0 [73.2;105]	87.0 [73.0;99.5]	86.0 [72.0;102]	92.5 [77.0;111]	0.003	598
COPD, n (%)	42 (7)	5 (2)	12 (6)	25 (12)	<0.001	644
Asthma, n (%)	63 (10)	26 (12)	25 (12)	12 (6)	0.039	644
OSA, n (%)	38 (6)	12 (6)	11 (5)	15 (7)	0.698	644
CAD, n (%)	26 (4)	2 (1)	4 (2)	20 (9)	<0.001	644
CVA, n (%)	10 (2)	1 (0)	3 (1)	6 (3)	0.174	644

Table B.27 Clinical characteristics of study individuals unrelated individuals used in gene-tag association analysis by KCO tertiles.

	[ALL] N=644	Higher tertile N=214	Middle tertile N=215	Lower tertile N=215	*p*-value	N
PAD, n (%)	2 (0)	0 (0)	0 (0)	2 (1)	0.332	644
HTN, n (%)	170 (26)	46 (21)	53 (25)	71 (33)	0.02	644
DM type 1, n (%)	9 (1)	2 (1)	5 (2)	2 (1)	0.526	644
DM type 2, n (%)	93 (14)	22 (10)	22 (10)	49 (23)	<0.001	644
Hypothyroidism, n (%)	73 (11)	25 (12)	30 (14)	18 (8)	0.185	644
Sjogren syndrome, n (%)	3 (0)	2 (1)	0 (0)	1 (0)	0.331	644

None of the patients had systemic lupus erythematosus, systemic sclerosis, undifferentiated connective tissue disease or ankylosing spondylitis Abbreviations: BMI: body mass index, WHO FC: World Health Organisation Functional Class, 6MWD: 6-minute walk distance, SpO_2: peripheral capillary oxygen saturation, mRAP: mean right atrial pressure, mPAP: mean pulmonary artery pressure, PAWP: Pulmonary Artery Wedge Pressure, CO: cardiac output, SvO_2: mixed venous oxygen saturation, NO: nitric oxide, FEV1: forced expiratory capacity in 1 second, FVC: forced vital capacity, TLC: Total Lung Capacity, KCO: transfer coefficient of carbon monoxide, NT-proBNP: N-terminal pro B-type natriuretic peptide, BNP: B-type natriuretic peptide, CRP: C-Reactive Protein Protein, Hb: haemoglobin, WBC: white blood cell count, COPD: chronic obstructive pulmonary disease, OSA: obstructive sleep apnoea, CAD: coronary artery disease, CVA: cerebrovascular accident, PAD: peripheral artery disease, HTN: hypertension, DM: diabetes mellitus; vr: vasoresponder

Table B.28 Results of Cox regression analysis relating overall survival to selected variables at baseline.

			Univariate		Multivariate	
	No event N=481	**Event N=163**	**HR [95%CI]**	***p*-value**	**HR [95%CI]**	***p*-value**
Sex: n (%)				<0.001		<0.001
female	343 (71.3)	91 (55.8)	Ref.		Ref.	
male	138 (28.7)	72 (44.2)	1.98 [1.45;2.70]		2.93 [1.81;4.75]	
Age [years]	4.80 (1.50)	6.17 (1.59)	1.75 [1.56;1.96]	<0.001	1.57 [1.27;1.94]	<0.001
Incident/Prevalent: n (%)				<0.001		0.131
incident	151 (31.4)	52 (31.9)	Ref.		Ref.	
prevalent	330 (68.6)	111 (68.1)	0.40 [0.28;0.58]		0.65 [0.37;1.14]	
6MWD [m]	32.9 (15.0)	21.7 (12.9)	0.95 [0.94;0.96]	<0.001	0.97 [0.95;0.99]	0.002
mRAP [mmHg]	1.84 (1.07)	2.15 (1.13)	1.26 [1.11;1.44]	0.001	1.28 [1.05;1.57]	0.016
mPAP [mmHg]	10.9 (2.75)	10.4 (2.34)	0.94 [0.88;1.00]	0.038	1.08 [0.97;1.2]	0.142
CI [L/min/m^2]	2.28 (0.78)	2.10 (0.68)	0.67 [0.53;0.86]	0.002	0.98 [0.62;1.55]	0.923
PVR [WU]	12.1 (5.96)	11.8 (4.94)	1.00 [0.97;1.03]	0.89		
KCO [%pred]	7.35 (2.22)	5.71 (2.41)	0.71 [0.66;0.77]	<0.001	0.79 [0.7;0.88]	<0.001
Smoking: n (%)				0.002		0.692
no	215 (47.5)	54 (35.1)	Ref.		Ref.	
past/current	238 (52.5)	100 (64.9)	1.67 [1.20;2.33]		1.11 [0.66;1.88]	
CAD: n (%)				0.002		0.081
no	467 (97.1)	151 (92.6)	Ref.		Ref.	
yes	14 (2.91)	12 (7.36)	2.51 [1.39;4.53]		0.38 [0.13;1.13]	
COPD: n (%)				0.141		0.268
no	452 (94.0)	150 (92.0)	Ref.		Ref.	
yes	29 (6.03)	13 (7.98)	1.53 [0.87;2.69]		0.59 [0.23;1.5]	
HTN: n (%)				<0.001		
no	374 (77.8)	100 (61.3)	Ref.		Ref.	
yes	107 (22.2)	63 (38.7)	1.92 [1.40;2.63]		1.12 [0.68;1.84]	
Emphysema: n (%)				0.012		
none	300 (91.5)	91 (84.3)	Ref.		Ref.	
minimal/mild	19 (5.79)	11 (10.2)	2.12 [1.13;3.98]		0.85 [0.38;1.92]	0.696
moderate	6 (1.83)	5 (4.63)	2.83 [1.15;6.98]		0.7 [0.19;2.61]	0.594
severe	3 (0.91)	1 (0.93)	2.36 [0.33;17.0]		0 [0;Inf]	0.996
Fibrosis: n (%)				0.003		
none	321 (97.0)	96 (89.7)	Ref.		Ref.	
minimal/mild	10 (3.02)	10 (9.35)	2.79 [1.45;5.36]		0.83 [0.36;1.93]	0.672
moderate	0 (0.00)	1 (0.93)	3.29 [0.46;23.6]		3.98 [0.47;33.58]	0.204

Abbreviations: CI: Confidence interval; 6MWD: 6-minute walking distance; mPAP: mean pulmonary arterial pressure; mRAP: mean right atrial pressure; PVR: pulmonary vascular resistance; WU: Wood units; KCO: transfer coefficient for carbon monoxide; CAD: coronary artery disease; COPD: chronic obstructive pulmonary disease; HTN: systemic hypertension

Table B.29 Clinical characteristics of all unrelated individuals used in gene-tag association by age tertiles.

	[ALL] N=1112	Lower tertile N=381	Middle tertile N=376	Higher tertile N=355	*p*-value	N
Age [years]	49 [35;63]	30 [25;35]	50 [45;54]	68 [64;73]	<0.001	1112
Sex: female, n (%)	757 (68)	276 (72)	273 (73)	208 (59)	<0.001	1112
Incident cases, n (%)	270 (24.3)	77 (20.2)	74 (19.7)	119 (33.5)	<0.001	1112
BMI [kg/m^2]	26.9 [23.1;31.5]	24.2 [21.0;29.3]	28.0 [24.3;32.5]	28.1 [25.1;32.0]	<0.001	1015
WHO FC: I/II/III/IV, %	2/20/65/13	4/26/56/14	1/21/67/11	1/12/7314	<0.001	1078
6MWD [m]	335 [220;415]	375 [302;460]	343 [250;420]	236 [134;348]	<0.001	953
SpO_2 pre [%]	96.0 [93.0;97.0]	97.0 [95.0;98.0]	96.0 [93.0;97.0]	94.0 [90.0;96.0]	<0.001	890
SpO post [%]	91.0 [85.0;95.0]	94.0 [88.0;97.0]	92.0 [86.0;96.0]	88.0 [82.0;92.8]	<0.001	830
mRAP [mmHg]	8 [5;12]	8 [5;12]	9 [6;13]	8 [5;12]	0.046	984
mPAP [mmHg]	53 [44;61]	55 [47;66]	55 [48;62]	48 [40;57]	<0.001	1051
PAWP [mmHg]	9 [7;12]	9 [6;11]	9 [7;12]	10 [7;13]	0.004	933
CO [L/min]	3.9 [3.1;4.9]	4.0 [3.1;5.0]	3.9 [3.1;4.9]	3.8 [3.2;4.8]	0.651	1003
SvO [%]	64 [58;70]	67 [60;72]	64 [58;70]	62 [57;67]	<0.001	817
Acute NO challenge: vr	59 (14)	31 (17)	21 (14)	7 (7)	0.078	435
FEV1 [% pred.]	85.0 [73.0;97.0]	87.0 [77.0;97.0]	84.0 [71.0;96.0]	85.2 [71.1;97.9]	0.34	849
FVC [% pred.]	94.0 [81.4;106]	90.0 [81.0;102]	96.0 [82.0;108]	96.3 [82.0;108]	0.018	831
FEV1/FVC ratio	0.76 [0.69;0.81]	0.81 [0.77;0.86]	0.75 [0.69;0.80]	0.72 [0.66;0.77]	<0.001	760
TLC [% pred.]	95.0 [85.0;104]	95.0 [85.0;103]	95.0 [87.0;106]	93.6 [83.0;102]	0.082	639
KCO [% pred.]	71 [52;86]	76 [65;91]	76 [62;88]	57 [40;78]	<0.001	644
Emphysema: n (%)					<0.001	611
none	559 (91)	193 (99)	198 (93)	168 (82)		
minimal/mild	33 (5)	1 (1)	10 (5)	22 (11)		
moderate	15 (2)	0 (0)	4 (2)	11 (5)		
severe	4 (1)	0 (0)	0 (0)	4 (2)		
Fibrosis: n (%)					<0.001	613
none	585 (95)	193 (98)	208 (98)	184 (90)		
minimal/mild	26 (4)	3 (2)	2 (1)	21 (10)		
moderate	1 (0)	0 (0)	1 (0)	0 (0)		
severe	1 (0)	0 (0)	1 (0)	0 (0)		
Smoking: past/current	435 (50.8)	107 (36.9)	153 (53.5)	175 (62.3)	<0.001	857
NTproBNP [ng/l]	926 [215;2637]	345 [122;1640]	763 [158;1356]	1996 [501;3706]	<0.001	276
BNP [ng/l]	195 [72.4;432]	117 [30.0;394]	181 [85.1;398]	236 [112;481]	0.005	271
Uric acid [mmol/l]	0.41 [0.31;0.52]	0.37 [0.26;0.46]	0.41 [0.30;0.50]	0.48 [0.36;0.56]	<0.001	358
CRP [mg/l]	4.30 [2.00;8.50]	4.00 [2.00;7.00]	4.15 [2.00;8.50]	5.00 [2.50;9.10]	0.151	639
Hb [g/l]	151 [138;165]	152 [138;164]	154 [141;166]	149 [133;163]	0.007	847
WBC [x10e9/l]	8 [7;10]	8 [6;10]	8 [7;10]	8 [7;10]	0.73	839
Platelets [x10e9/l]	224 [182;272]	231 [186;280]	219 [181;261]	221 [179;272]	0.072	836
Sodium [mmol/l]	140 [138;141]	140 [138;141]	139 [138;141]	140 [137;141]	0.728	835
Potassium [mmol/l]	4.20 [3.90;4.50]	4.20 [3.98;4.40]	4.20 [3.90;4.40]	4.30 [4.00;4.50]	0.03	830
Urea [mmol/l]	5.70 [4.40;7.60]	4.80 [3.88;5.81]	5.50 [4.30;6.70]	7.60 [5.90;10.1]	<0.001	830
Creatinine [mol/l]	85.5 [70.0;102]	79.0 [68.0;93.5]	82.0 [69.0;96.0]	96.0 [80.0;121]	<0.001	832
COPD, n (%)	65 (6)	2 (1)	22 (6)	41 (12)	<0.001	1112
Asthma, n (%)	78 (7)	32 (8)	32 (9)	14 (4)	0.023	1112
OSA, n (%)	57 (5)	5 (1)	21 (6)	31 (9)	<0.001	1112
CAD, n (%)	44 (4)	0 (0)	9 (2)	35 (10)	<0.001	1112
CVA, n (%)	17 (2)	2 (1)	6 (2)	9 (3)	0.084	1112
PAD, n (%)	5 (0)	0 (0)	0 (0)	5 (1)	0.003	1112
HTN, n (%)	264 (24)	19 (5)	80 (21)	165 (46)	<0.001	1112
DM type 1, n (%)	19 (2)	7 (2)	6 (2)	6 (2)	0.967	1112
DM type 2, n (%)	137 (12)	5 (1)	37 (10)	95 (27)	<0.001	1112
Hypothyroidism, n (%)	135 (12)	38 (10)	49 (13)	48 (14)	0.274	1112
SLE, n (%)	1 (0)	0 (0)	0 (0)	1 (0)	0.319	1112
Systemic sclerosis, n (%)	1 (0)	0 (0)	0 (0)	1 (0)	0.319	1112
Ankylosing spondylitis, n (%)	1 (0.09)	0 (0.00)	0 (0.00)	1 (0.28)	0.319	1112
Sjogren syndrome, n (%)	5 (0)	2 (1)	1 (0)	2 (1)	0.871	1112

Abbreviations: BMI: body mass index; WHO FC: World Health Organisation Functional Class; 6MWD: 6-minute walk distance; SpO_2: peripheral capillary oxygen saturation; mRAP: mean right atrial pressure; mPAP: mean pulmonary artery pressure; PAWP: Pulmonary Artery Wedge Pressure; CO: cardiac output; SvO_2: mixed venous oxygen saturation; PVR: pulmonary vascular resistance; WU: Wood units; NO: nitric oxide; FEV1: forced expiratory capacity in 1 second; FVC: forced vital capacity; KCO: transfer coefficient of carbon monoxide; Hb: haemoglobin; WBC: white blood cell count; COPD: chronic obstructive pulmonary disease; OSA: obstructive sleep apnoea; CAD: coronary artery disease; HTN: hypertension; CKD: chronic kidney disease; TSH: thyroid-stimulating hormone

Table B.30 Comparison of clinical characteristics between patients harbouring *BMPR2*, *EIF2AK4*, *KDR* and without mutation.

	***BMPR2* N=162**	**biallelic *EIF2AK4* N=14**	***KDR*.het.ptv N=4**	**no mutation N=829**	***p*-value**	**N**
Age [years]	39 [32;51]	31 [23;42]	64 [62;68]	51 [38;66]	<0.001	1000
Sex: female, n (%)	107 (66)	7 (50)	2 (50)	579 (70)	0.184	1004
BMI [kg/m^2]	27 [23;32]	24 [20;27]	26 [26;30]	27 [23;32]	0.07	915
WHO FC: I/II/III/IV, %	1/20/60/19	0/14/64/21	0/25/75/0	2/20/67/12	0.334	970
6MWD [m]	355 [288;421]	302 [210;466]	301 [240;362]	332 [218;412]	0.27	628
SpO_2 pre [%]	96 [94;98]	92 [90;96]	97 [96;97]	95 [93;97]	0.003	804
SpO_2 post [%]	94 [89;97]	83 [76;86]	86 [86;88]	91 [84;95]	<0.001	746
mRAP [mmHg]	10 [6;14]	8 [6;10]	6 [4;8]	8 [5;12]	0.013	886
mPAP [mmHg]	57 [52;67]	52 [44;59]	50 [43;58]	52 [42;61]	<0.001	950
mPAWP [mmHg]	10 [7;12]	11 [8;12]	10 [8;13]	9 [7;12]	0.864	842
CO [L/min]	3.3 [2.7;3.9]	4.5 [3.0;4.9]	4.9 [4.3;5.4]	4.0 [3.2;5.1]	<0.001	908
PVR [WU]	14.5 [10.8;20.4]	9.56 [8.16;11.1]	7.93 [6.23;10.9]	10.3 [7.13;13.9]	<0.001	808
Acute NO challenge: (vr)	1 (1.28)	0 (0.00)	1 (25.0)	53 (17.5)	<0.001	392
FEV1 [% pred.]	91 [79;100]	93 [84;100]	86 [79;96]	84 [71;95]	<0.001	767
FVC [% pred.]	100 (17)	101 (16)	93 (17)	92 (20)	0.001	751
FEV1/FVC ratio	0.77 [0.73;0.82]	0.79 [0.69;0.81]	0.78 [0.76;0.79]	0.76 [0.69;0.81]	0.063	685
KCO [% pred]	83 [74;96]	33 [30;35]	46 [46;48]	68 [48;83]	<0.001	580
Smoking: current/previous, n (%)	53 (40.8)	4 (30.8)	1 (25.0)	334 (53.2)	0.015	775
COPD, n (%)	6 (3.70)	0 (0.00)	0 (0.00)	49 (5.91)	0.6	1009
Asthma, n (%)	20 (12.3)	4 (28.6)	0 (0.00)	47 (5.67)	0.001	1009
OSA, n (%)	0 (0.00)	0 (0.00)	0 (0.00)	56 (6.76)	0.001	1009
CAD, n (%)	3 (1.85)	1 (7.14)	1 (25.0)	32 (3.86)	0.071	1009
HTN, n (%)	28 (17.3)	0 (0.00)	2 (50.0)	213 (25.7)	0.005	1009
CKD, n (%)	4 (2.47)	0 (0.00)	0 (0.00)	43 (5.19)	0.418	1009
Hypothyroidism, n (%)	14 (8.64)	1 (7.14)	1 (25.0)	108 (13.0)	0.251	1009
Pulmonary embolism, n (%)	2 (1.23)	0 (0.00)	0 (0.00)	27 (3.26)	0.482	1009
Hb [g/l]	162 [152;173]	165 [154;179]	148 [135;152]	149 [135;162]	<0.001	764
WBC [x10e9/l]	8.74 [7.30;10.8]	7.43 [6.50;10.8]	8.80 [8.23;9.55]	8.10 [6.70;9.70]	0.019	757
Platelets [x10e9/l]	210 [174;251]	219 [206;234]	216 [188;251]	228 [181;277]	0.268	753
TSH [mU/l]	2.37 [1.67;3.65]	2.08 [1.09;3.69]	1.76 [1.72;1.84]	2.00 [1.10;3.15]	0.021	589

Abbreviations: BMI: body mass index; WHO FC: World Health Organisation Functional Class; 6MWD: 6-minute walk distance; SpO_2: peripheral capillary oxygen saturation; mRAP: mean right atrial pressure; mPAP: mean pulmonary artery pressure; PAWP: Pulmonary Artery Wedge Pressure; CO: cardiac output; SvO_2: mixed venous oxygen saturation; PVR: pulmonary vascular resistance; WU: Wood units; NO: nitric oxide; FEV1: forced expiratory capacity in 1 second; FVC: forced vital capacity; KCO: transfer coefficient of carbon monoxide; Hb: haemoglobin; WBC: white blood cell count; COPD: chronic obstructive pulmonary disease; OSA: obstructive sleep apnoea; CAD: coronary artery disease; HTN: hypertension; CKD: chronic kidney disease; TSH: thyroid-stimulating hormone

Table B.31 Summary of CT imaging analysis.

	[ALL] N=267	***BMPR2* N=44**	***EIF2AK4* N=6**	***EIF2AK4* biall N=7**	***KDR* mis N=5**	***KDR* ptv N=4**	**no mutation N=183**	**other mutations N=18**	***p*-value**	**N**
Sex: female, n (%)	177 (66.3)	28 (63.6)	5 (83.3)	3 (42.9)	3 (60.0)	2 (50.0)	127 (69.4)	9 (50.0)	0.364	267
Age at diagnosis [years]	50.1 (17.0)	44.2 (13.7)	51.0 (16.7)	28.5 (10.8)	36.6 (13.1)	65.2 (4.3)	52.9 (17.2)	44.4 (14.1)	<0.001	266
Diagnosis verified: n (%)									0.009	267
HPAH	20 (7.5)	13 (29.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	4 (2.2)	3 (16.7)		
IPAH	235 (88.0)	31 (70.5)	5 (83.3)	4 (57.1)	5 (100.0)	4 (100.0)	171 (93.4)	15 (83.3)		
PCH	1 (0.4)	0 (0.0)	0 (0.0)	1 (14.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)		
PVOD	11 (4.1)	0 (0.0)	1 (16.7)	2 (28.6)	0 (0.0)	0 (0.0)	8 (4.4)	0 (0.0)		
GGO centrilobular pattern severity: n (%)									0.04731	
Nil	22 (50.0)	4 (66.7)	2 (28.6)	4 (80.0)	2 (50.0)	119 (65.0)	12 (66.7)	165 (61.8)		
Trace	3 (6.8)	0 (0.0)	1 (14.3)	0 (0.0)	0 (0.0)	23 (12.6)	2 (11.1)	29 (10.9)		
Mild	9 (20.5)	0 (0.0)	1 (14.3)	0 (0.0)	2 (50.0)	16 (8.7)	1 (5.6)	29 (10.9)		
Moderate	3 (6.8)	0 (0.0)	1 (14.3)	1 (20.0)	0 (0.0)	17 (9.3)	2 (11.1)	24 (9.0)		
Severe	7 (15.9)	2 (33.3)	2 (28.6)	0 (0.0)	0 (0.0)	8 (4.4)	1 (5.6)	20 (7.5)		
Severity of GGO non-specific pattern: n (%)									0.723	267
Nil	238 (89.1)	42 (95.5)	3 (50.0)	5 (71.4)	5 (100.0)	2 (50.0)	165 (90.2)	16 (88.9)		
Trace	10 (3.7)	1 (2.3)	2 (33.3)	0 (0.0)	0 (0.0)	1 (25.0)	6 (3.3)	0 (0.0)		
Mild	11 (4.1)	1 (2.3)	0 (0.0)	1 (14.3)	0 (0.0)	1 (25.0)	7 (3.8)	1 (5.6)		
Moderate	7 (2.6)	0 (0.0)	1 (16.7)	1 (14.3)	0 (0.0)	0 (0.0)	4 (2.2)	1 (5.6)		
Severe	1 (0.4)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.5)	0 (0.0)		
Distribution of GGO: n (%)									0.108	122
C	11 (9.0)	0 (0.0)	1 (20.0)	2 (40.0)	0 (0.0)	2 (66.7)	6 (7.7)	0 (0.0)		
D	81 (66.4)	19 (90.5)	3 (60.0)	3 (60.0)	2 (100.0)	1 (33.3)	48 (61.5)	5 (62.5)		
U	14 (11.5)	1 (4.8)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	12 (15.4)	1 (12.5)		
Z	16 (13.1)	1 (4.8)	1 (20.0)	0 (0.0)	0 (0.0)	0 (0.0)	12 (15.4)	2 (25.0)		
Pulmonary arteriovenous malformations, n (%)	4 (1.6)	3 (7.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.6)	0 (0.0)	0.155	244
Largest bronchial artery size [mm]	3.0 (0.6)	3.4 (0.5)	3.0 (.)	2.5 (0.2)	0 (0.0)	0 (0.0)	2.5 (0.4)	4.0 (.)	0.037	12
Mediastinal venous collaterals, n (%)	204 (94.0)	37 (100.0)	1 (100.0)	0 (0.0)	5 (100.0)	3 (100.0)	145 (91.8)	13 (100.0)	0.383	217
Intralobular septal thickening: n (%)									0.113	241
Nil	214 (88.8)	37 (92.5)	0 (0.0)	0 (0.0)	4 (80.0)	4 (100.0)	153 (88.4)	16 (88.9)		
Trace	15 (6.2)	2 (5.0)	0 (0.0)	0 (0.0)	1 (20.0)	0 (0.0)	12 (6.9)	0 (0.0)		
Mild	9 (3.7)	1 (2.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	7 (4.0)	1 (5.6)		
Moderate	3 (1.2)	0 (0.0)	1 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.6)	1 (5.6)		
Mediastinal lymphadenopathy, n (%)	216 (81.2)	40 (90.9)	3 (50.0)	3 (42.9)	2 (50.0)	1 (25.0)	149 (81.4)	18 (100.0)	<0.001	266
Mediastinal lymphoadenopathy size [mm]	14.9 (4.0)	12.5 (1.3)	17.0 (3.6)	14.8 (2.4)	15.5 (0.7)	11.0 (0.0)	15.3 (4.4)	0 (0.0)	0.354	50
Emphysema: n (%)									0.819	267
Nil	233 (87.3)	38 (86.4)	5 (83.3)	7 (100.0)	4 (80.0)	3 (75.0)	159 (86.9)	17 (94.4)		
Trace	20 (7.5)	4 (9.1)	1 (16.7)	0 (0.0)	1 (20.0)	0 (0.0)	13 (7.1)	1 (5.6)		

Table B.31 Summary of CT imaging analysis.

	[ALL] N=267	*BMPR2* N=44	*EIF2AK4* N=6	*EIF2AK4* biall N=7	*KDR* mis N=5	*KDR* ptv N=4	no mutation N=183	other mutations N=18	*p*-value	N
Mild	8 (3.0)	1 (2.3)	0 (0.0)	0 (0.0)	0 (0.0)	1 (25.0)	6 (3.3)	0 (0.0)		
Moderate	6 (2.2)	1 (2.3)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	5 (2.7)	0 (0.0)		
Bronchial wall thickening: n (%)									0.049	241
Nil	210 (87.1)	33 (82.5)	0 (0.0)	0 (0.0)	5 (100.0)	3 (75.0)	156 (90.2)	13 (72.2)		
Trace	18 (7.5)	5 (12.5)	1 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)	10 (5.8)	2 (11.1)		
Mild	11 (4.6)	2 (5.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (25.0)	6 (3.5)	2 (11.1)		
Moderate	2 (0.8)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.6)	1 (5.6)		
Fibrosis: n (%)									0.06	267
Nil	255 (95.5)	43 (97.7)	6 (100.0)	7 (100.0)	5 (100.0)	2 (50.0)	175 (95.6)	17 (94.4)		
Trace	5 (1.9)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	5 (2.7)	0 (0.0)		
Mild	6 (2.2)	1 (2.3)	0 (0.0)	0 (0.0)	0 (0.0)	2 (50.0)	3 (1.6)	0 (0.0)		
Moderate	1 (0.4)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (5.6)		
Pleural effusion: n (%)									0.368	267
Nil	240 (89.9)	41 (93.2)	5 (83.3)	7 (100.0)	4 (80.0)	3 (75.0)	165 (90.2)	15 (83.3)		
Trace	11 (4.1)	1 (2.3)	1 (16.7)	0 (0.0)	0 (0.0)	0 (0.0)	6 (3.3)	3 (16.7)		
Mild	7 (2.6)	2 (4.5)	0 (0.0)	0 (0.0)	1 (20.0)	0 (0.0)	4 (2.2)	0 (0.0)		
Moderate	7 (2.6)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (25.0)	6 (3.3)	0 (0.0)		
Severe	2 (0.7)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2 (1.1)	0 (0.0)		
Air trapping: n (%)									0.085	267
Nil	214 (80.1)	40 (90.9)	5 (83.3)	7 (100.0)	4 (80.0)	1 (25.0)	144 (78.7)	13 (72.2)		
Trace	26 (9.7)	3 (6.8)	1 (16.7)	0 (0.0)	0 (0.0)	1 (25.0)	17 (9.3)	4 (22.2)		
Mild	18 (6.7)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	2 (50.0)	15 (8.2)	1 (5.6)		
Moderate	4 (1.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	4 (2.2)	0 (0.0)		
Severe	5 (1.9)	1 (2.3)	0 (0.0)	0 (0.0)	1 (20.0)	0 (0.0)	3 (1.6)	0 (0.0)		
Subpleural scarring:									0.739	241
Nil	235 (97.5)	39 (97.5)	1 (100.0)	0 (0.0)	5 (100.0)	4 (100.0)	168 (97.1)	18 (100.0)		
Trace	2 (0.8)	1 (2.5)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.6)	0 (0.0)		
Mild	4 (1.7)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	4 (2.3)	0 (0.0)		

Intra-rater reliability: GGO centrilobular pattern severity weighted Cohen's Kappa = 0.679, p-value < 0.001; GGO distribution unweighted Cohen's Kappa = 1, p-value = 0.046; Severity of GGO non-specific pattern - no positive findings; Pulmonary arteriovenous malformations - no positive findings; largest BRA size - no positive findings; Mediastinal venous collaterals: unweighted Cohen's Kappa = 1, p-value < 0.001; Intralobular septal thickening weighted Cohen's Kappa = 1, p-value < 0.001; Mediastinal lymphadenopathy unweighted Cohen's Kappa = 0.83, p-value < 0.001; Mediastinal lymphadenopathy size [mm] intraclass correlation coefficient (ICC) 0.717, p-value = 0.088; Emphysema - not enough positive findings, Bronchial wall thickening - not enough positive findings, Fibrosis - no positive findings; Pleural effusion weighted Cohen's Kappa = 0.826, p-value < 0.001; Air trapping weighted Cohen's Kappa 0.845, p-value < 0.001; Subpleural scarring - not enough positive findings. Abbreviations: IPAH: idiopathic pulmonary arterial hypertension; HPAH: hereditary pulmonary arterial hypertension; PVOD: pulmonary veno-occlusive disease; PCH: Pulmonary capillary haemangiomatosis: GGO: ground-glass opacities; C: central; U: upper; Z: zonal; D: diffuse; BMPR2: Bone Morphogenetic Protein Receptor Type 2; EIF2AK4: Eukaryotic Translation Initiation Factor 2 Alpha Kinase 4; KDR: Kinase Insert Domain Receptor

Table B.32 Variants driving associations with computational phenotypes.

Gene	Biotype	Chr	Exon	Intron	HGVSc	HGVSp	Consequence	SIFT	PolyPhen	CADD Phred	gnomAD
COL6A5	protein_coding	3	-	1 of 41	c.-28-1G>T	-	splice_acceptor_variant	-	-	28.8	-
COL6A5	protein_coding	3	3 of 42	-	c.84T>A	p.Tyr28Ter	stop_gained	-	-	34	-
COL6A5	protein_coding	3	3 of 42	-	c.502G>A	p.Gly168Arg	missense_variant	deleterious(0)	probably_damaging(1)	24.6	1.21E-5
COL6A5	protein_coding	3	4 of 42	-	c.818G>A	p.Arg273Gln	missense_variant	deleterious(0.04)	probably_damaging(1)	22.6	1.95E-5
COL6A5	protein_coding	3	4 of 42	-	c.1060G>A	p.Val354Met	missense_variant	deleterious(0)	probably_damaging(0.914)	23.4	1.953E-5
COL6A5	protein_coding	3	6 of 42	-	c.2087A>T	p.Asp696Val	missense_variant	deleterious(0)	possibly_damaging(0.873)	23.3	-
COL6A5	protein_coding	3	7 of 42	-	c.2731C>T	p.Arg911Cys	missense_variant	deleterious(0.01)	probably_damaging(1)	26	-
COL6A5	protein_coding	3	9 of 42	-	c.3576T>G	p.Phe1192Leu	missense_variant	tolerated(0.12)	benign(0.076)	18.72	-
COL6A5	protein_coding	3	12 of 42	-	c.4139A>C	p.Asn1380Thr	missense_variant	deleterious(0.02)	benign(0.139)	22	-
COL6A5	protein_coding	3	19 of 42	-	c.4597C>A	p.Gln1533Lys	missense_variant	tolerated(0.29)	benign(0.012)	22.7	-
COL6A5	protein_coding	3	24 of 42	-	c.4866A>T	p.Arg1622Ser	missense_variant	tolerated(0.51)	benign(0.007)	16.6	-
COL6A5	protein_coding	3	-	29 of 41	c.5164-2A>G	-	splice_acceptor_variant	-	-	34	-
COL6A5	protein_coding	3	34 of 42	-	c.5822G>A	p.Cys1941Tyr	missense_variant	deleterious(0.01)	probably_damaging(0.997)	28.7	-
COL6A5	protein_coding	3	2 of 42	-	c.35T>C	p.Ile12Thr	missense_variant	deleterious(0.04)	benign(0.007)	17.09	-
COL6A5	protein_coding	3	3 of 42	-	c.350C>T	p.Ala117Val	missense_variant	tolerated(0.44)	benign(0.012)	12.42	
EYS	protein_coding	6	43 of 43	-	c.9131G>T	p.Trp3044Leu	missense_variant	deleterious(0)	probably_damaging(0.919)	25.4	-
EYS	protein_coding	6	43 of 43	-	c.9108T>A	p.Asn3036Lys	missense_variant	tolerated(0.12)	benign(0.033)	44555	6.40E-06
EYS	protein_coding	6	43 of 43	-	c.8465A>G	p.Tyr2822Cys	missense_variant	tolerated(0.13)	probably_damaging(0.994)	23	6.34E-06
EYS	protein_coding	6	41 of 43	-	c.7927T>C	p.Ser2643Pro	missense_variant	deleterious(0.04)	benign(0)	11.75	1.92E-05
EYS	protein_coding	6	41 of 43	-	c.7921A>G	p.Lys2641Glu	missense_variant	tolerated(0.19)	possibly_damaging(0.783)	23.2	-
EYS	protein_coding	6	39 of 43	-	c.7676C>T	p.Thr2559Ile	missense_variant	tolerated(0.14)	benign(0.007)	15.11	-
EYS	protein_coding	6	36 of 43	-	c.7115T>A	p.Phe2372Tyr	missense_variant	tolerated(0.11)	benign(0.091)	20.1	-
EYS	protein_coding	6	35 of 43	-	c.6928C>A	p.Leu2310Ile	missense_variant	deleterious(0.03)	benign(0.092)	9.647	6.37E-06
EYS	protein_coding	6	31 of 43	-	c.6284C>G	p.Pro2095Arg	missense_variant	deleterious(0)	benign(0.178)	13.19	-
EYS	protein_coding	6	26 of 43	-	c.5180A>G	p.Lys1727Arg	missense_variant	tolerated_low_confidence(0.15)	benign(0.053)	44539	-
EYS	protein_coding	6	25 of 43	-	c.3753A>C	p.Gln1251His	missense_variant	tolerated(0.06)	possibly_damaging(0.723)	22.2	-
EYS	protein_coding	6	22 of 43	-	c.3422C>A	p.Pro1141His	missense_variant	tolerated(0.06)	probably_damaging(0.939)	18.13	-
EYS	protein_coding	6	20 of 43	-	c.3046G>A	p.Asp1016Asn	missense_variant	tolerated(0.43)	benign(0.018)	16.09	-
EYS	protein_coding	6	15 of 43	-	c.2381G>A	p.Arg794Gln	missense_variant, splice_region_variant	tolerated(0.29)	benign(0.014)	22.4	5.22E-05
EYS	protein_coding	6	8 of 43	-	c.1249A>G	p.Ile417Val	missense_variant	tolerated(0.48)	benign(0.062)	12.45	4.01E-06
OR6T1	protein_coding	11	1 of 1	-	c.902T>C	p.Leu301Pro	missense_variant	deleterious(0)	probably_damaging(0.984)	22.5	1.19E-05
OR6T1	protein_coding	11	1 of 1	-	c.515A>G	p.Asn172Ser	missense_variant	tolerated(0.06)	possibly_damaging(0.838)	17.03	3.99E-06
OR6T1	protein_coding	11	1 of 1	-	c.511C>T	p.Pro171Ser	missense_variant	tolerated(0.07)	benign(0.05)	16.6	7.98E-06
OR6T1	protein_coding	11	1 of 1	-	c.335T>C	p.Phe112Ser	missense_variant	deleterious(0.02)	possibly_damaging(0.573)	23.7	7.97E-06
OR6T1	protein_coding	11	1 of 1	-	c.183C>A	p.Phe61Leu	missense_variant	deleterious(0.04)	possibly_damaging(0.824)	23.1	-
HPSE2	protein_coding	10	10 of 12	-	c.1414C>T	p.Arg472Ter	stop_gained	-	-	38	7.95E-06
HPSE2	protein_coding	10	-	7 of 11	c.1099-1G>A	-	splice_acceptor_variant	-	-	34	4.81E-05
HPSE2		10	-	-	-	-	del: 100453153-100517252	-	-	-	-

Transcripts: *COL6A5*: ENST00000265379, *EYS*: ENST00000503581, *OR6T1*: ENST00000321252, *HPSE2*: ENST00000370552; Abbreviations: Chr: chromosome; HGVS: Human Genome Variation Society; SIFT: Sorting Intolerant from Tolerant; PolyPhen: Polymorphism Phenotyping; CADD: Combined Annotation Dependent Depletion; gnomAD: The Genome Aggregation Database

Table B.33 Clinical characteristics of patients and controls in whom BMP9 and pBMP10 measurements were undertaken

	[ALL] N=380	**Control N=120**	**PAH** N=260	***p*-value**	**N**
Age at sampling [years]	52 [41;62]	51 [37;57]	53 [42;64]	0.019	380
Diagnosis: n (%)					260
HPAH	27 (10)		27 (10)		
IPAH	233 (90)		233 (90)		
Case: prevalent, n (%)	209 (80)		209 (80)		260
Gender: female, n (%)	267 (70)	77 (64)	190 (73)	0.1	380
BMI [kg/m^2]	28 [24;32]	25 [24;30]	28 [24;32]	0.036	327
Ethnicity: n (%)				<0.001	379
White	293 (77)	66 (55)	227 (88)		
Asian	41 (11)	22 (18)	19 (7)		
Black	10 (3)	6 (5)	4 (2)		
Caribbean	1 (0)	0 (0)	1 (0)		
Other	30 (8)	24 (20)	6 (2)		
Not stated	4 (1)	2 (2)	2 (1)		
Smoking: past/current, n (%)	149 (42)	14 (14)	135 (54)	<0.001	354
HTN, n (%)	74 (20)	20 (20)	54 (21)	0.919	362
CAD, n (%)	15 (4)	7 (7)	8 (3)	0.14	362
DM type 2, n (%)	42 (12)	3 (3)	39 (15)	0.002	362
Dyslipidemia, n (%)	31 (9)	9 (9)	22 (8)	1	362
Hypothyroidism, n (%)	42 (12)	4 (4)	38 (15)	0.007	362
OSA, n (%)	12 (3)	0 (0)	12 (5)	0.023	362

Abbreviations: BMI: body mass index; I/HPAH: idiopathic/hereditary pulmonary arterial hypertension there was no diagnosis of Hereditary Hemorrhagic Telangiectasia in the entire cohort; HTN: systemic hypertension; CAD: coronary artery disease; DM: diabetes mellitus

Table B.34 Clinical characteristics at the time of sampling of I/HPAH cases by BMP9 concentration tertiles.

	[ALL] N=260	Lower tertile N=94	Middle tertile N=80	Higher tertile N=86	*p*-value	N
Demographics and functional status						
Age at sampling	53 [42;64]	54 [42;67]	49 [41;63]	56 [44;64]	0.261	260
Gender: female, n (%)	190 (73)	57 (61)	60 (75)	73 (85)	**0.001**	260
BMI [kg/m^2]	27 [24;32]	28 [25;32]	27 [24;31]	27 [22;31]	0.263	174
WHO FC: I/II/III/IV, %	12/38/45/5	6/41/49/3	16/37/38/9	14/36/46/4	0.178	258
6MWD [m]	389 [335;452]	394 [366;428]	370 [288;481]	390 [336;435]	0.772	116
Shuttle walk test [m]	375 [304;515]	418 [312;625]	330 [275;431]	388 [341;460]	0.558	30
Haemodynamics						
mRAP [mmHg]	10 [4;12]	9 [5;13]	8 [4;12]	11 [8;12]	0.326	38
mPAP [mmHg]	49 [39;63]	47 [32;50]	42 [36;58]	59 [42;66]	0.158	39
PVR [WU]	8.05 [5.72;13.9]	8.05 [6.45;12.8]	7.40 [4.00;10.9]	10.4 [6.20;17.4]	0.43	36
CO [L/min]	4.7 (2.1)	4.6 (2.3)	5.3 (2.0)	4.1 (2.0)	0.289	38
Lung function						
FEV1 [% pred.]	86 (20)	87 (21)	83 (19)	88 (19)	0.763	58
FVC [% pred.]	99 (21)	97 (21)	99 (23)	99 (20)	0.95	59
KCO [% pred.]	77 (16)	81 (14)	67 (17)	82 (11)	**0.016**	36
Clinical blood tests						
Hb [g/l]	141 [130;152]	146 [128;157]	141 [133;152]	139 [128;148]	0.148	239
RDW [%]	14 [14;16]	15 [14;16]	14 [14;16]	14 [14;15]	**0.007**	173
WBC [x10e9/l]	7.0 [5.5;8.6]	7.3 [6.0;9.1]	7.0 [5.3;8.3]	6.5 [5.3;8.4]	0.083	240
Plt [x10e9/l]	212 [170;259]	188 [158;238]	221 [179;264]	224 [175;271]	**0.016**	240
Albumin [g/l]	42 [39;45]	42 [38;44]	43 [40;46]	42 [39;45]	0.072	249
ALT [IU/l]	21 [16;27]	20 [15;28]	22 [16;29]	21 [16;26]	0.532	248
AST [IU/l]	24 [19;28]	22 [18;28]	23 [19;28]	24 [20;28]	0.585	181
Bilirubin [μmol/l]	10 [8;14]	11 [8;16]	9 [8;14]	9 [7;13]	0.069	249
Total protein [g/l]	72 [68;75]	71 [66;76]	72 [68;74]	71 [69;74]	0.925	189
Sodium [mmol/l]	140 [139;141]	140 [138;141]	140 [139;142]	140 [139;142]	0.841	244
Potassium [mmol/l]	4.1 (0.4)	4.2 (0.5)	4.1 (0.4)	4.1 (0.4)	0.198	244
Urea [mmol/l]	5.6 [4.2;6.9]	5.9 [4.5;7.5]	5.5 [4.2;6.5]	5.6 [4.1;6.7]	0.123	244
Creatinine [μmol/l]	79 [68;97]	82 [69;102]	81 [68;98]	76 [66;91]	0.111	242
NTproBNP [ng/l]	222 [61.5;1230]	538 [81.0;1836]	151 [53.3;382]	176 [57.5;1126]	**0.031**	110
BNP [ng/l]	64.8 [30.1;167]	72.4 [29.0;147]	51.6 [28.5;200]	82.0 [36.7;146]	0.993	101
Comorbidities						
HTN, n (%)	54 (21)	24 (26)	18 (22)	12 (14)	0.144	260
DM type 1, n (%)	5 (2)	2 (2)	1 (1)	2 (2)	1	260
DM type 2, n (%)	39 (15)	14 (15)	15 (19)	10 (12)	0.438	260
CAD, n (%)	8 (3)	3 (3)	3 (4)	2 (2)	0.907	260
CVA, n (%)	4 (2)	3 (3)	1 (1)	0 (0)	0.269	260
Dyslipidemia, n (%)	22 (8)	5 (5)	12 (15)	5 (6)	**0.041**	260
AVM, n (%)	2 (1)	1 (1)	1 (1)	0 (0)	0.76	260
Epistaxis, n (%)	3 (1)	1 (1)	1 (1)	1 (1)	1	260
Hepatitis, n (%)	4 (2)	2 (2)	1 (1)	1 (1)	1	260
Hypothyroidism, n (%)	38 (15)	16 (17)	11 (14)	11 (13)	0.7	260
COPD, n (%)	10 (4)	6 (6)	1 (1)	3 (3)	0.258	260
Asthma, n (%)	28 (11)	9 (10)	9 (11)	10 (12)	0.894	260
OSA, n (%)	12 (5)	7 (7)	4 (5)	1 (1)	0.112	260
Heterotaxy, n (%)	1 (0)	1 (1)	0 (0)	0 (0)	1	260
Asplenia, n (%)	1 (0)	1 (1)	0 (0)	0 (0)	1	260

Table B.34 Clinical characteristics at the time of sampling of I/HPAH cases by BMP9 concentration tertiles.

	[ALL] N=260	Lower tertile N=94	Middle tertile N=80	Higher tertile N=86	*p*-value	N
Cancer, n (%)	17 (7)	6 (6)	6 (8)	5 (6)	0.905	260

Tertiles' ranges: lower 0 – 186.7 pg/ml, middle 189.6 – 266.7 pg/ml, 267.1 – 591.2 pg/ml. Data presented as: median [interquartile range], count (%), mean (SD); Abbreviations: IPAH: idiopathic pulmonary arterial hypertension; HPAH: hereditary pulmonary arterial hypertension (diagnosis based on positive family history); BMI: body mass index; WHO class: World Health Organization functional class; 6MWT: 6 minute walk test; mRAP: mean right atrial pressure; mPAP: mean pulmonary arterial pressure; PVR: pulmonary vascular resistance; CO: cardiac output; FEV1: forced expiratory volume in 1 second; FVC: forced vital capacity; KCO: transfer coefficient for carbon monoxide; Hb: hemoglobin, RDW: red cell distribution width; WBC: white blood cell count; Plt: platelets; AST: aspartate aminotransferase; ALT: alanine aminotransferase; BNP: brain natriuretic peptide; NTproBNP: N-terminal pro B-type natriuretic peptide; AVM: arteriovenous malformation (one hepatic AVM and one pulmonary AVM); DM type 1: Diabetes mellitus type 1; DM type 2: diabetes mellitus type 2; OSA : obstructive sleep apnoea; CAD: coronary artery disease; CVA: cardiovascular accident; HTN: systemic hypertension; COPD: chronic obstructive pulmonary disease

Table B.35 Clinical characteristics at the time of sampling of I/HPAH cases by pBMP10 concentration tertiles.

	[ALL] N=254	Lower tertile N=99	Middle tertile N=76	Higher tertile N=79	*p*-value	N
			Demographics and functional status			
Age at sampling	53 [42;64]	55 [43;67]	52 [40;63]	51 [42;61]	0.243	254
Gender: female, n (%)	187 (74)	66 (67)	56 (74)	65 (82)	0.063	254
BMI [kg/m^2]	27 [24;32]	30 [26;33]	27 [24;31]	26 [21;29]	0.004	170
WHO FC: I/II/III/IV	12/38/44/5	3/39/54/4	17/38/37/8	19/37/40/4	0.005	252
6MWD [m]	390 [335;460]	385 [308;420]	380 [315;481]	415 [370;462]	0.305	113
Shuttle walk test [m]	370 [300;500]	340 [300;520]	380 [320;500]	370 [332;436]	0.867	29
			Haemodynamics			
mRAP [mmHg]	10 [4;12]	9 [7;13]	12 [4;12]	9 [5;11]	0.9	38
mPAP [mmHg]	49 [39;63]	40 [38;49]	48 [36;60]	60 [44;65]	0.186	39
PVR [WU]	8.05 [5.72;13.9]	6.10 [3.25;10.7]	7.70 [5.78;10.1]	11.8 [6.80;17.4]	0.138	36
CO [L/min]	4.7 (2.1)	4.5 (2.3)	5.4 (2.2)	4.1 (1.8)	0.315	38
			Lung function			
FEV1 [% pred.]	86 (20)	83 (18)	88 (21)	86 (21)	0.72	56
FVC [% pred.]	99 (21)	92 (19)	105 (23)	99 (21)	0.194	57
KCO [% pred.]	76 (15)	74 (19)	71 (16)	81 (11)	0.233	35
			Clinical blood tests			
Hb [g/l]	140 [129;152]	140 [122;154]	141 [132;152]	140 [130;150]	0.802	233
RDW [%]	14 [14;15]	15 [14;16]	14 [14;16]	14 [13;15]	0.004	168
WBC [x10e9/l]	7.0 [5.5;8.6]	7.4 [5.8;9.0]	7.0 [5.8;8.3]	6.5 [5.1;8.1]	0.041	234
Plt [x10e9/l]	212 [170;260]	196 [161;262]	231 [175;270]	203 [170;246]	0.151	234
Albumin [g/l]	42 [39;45]	41 [37;45]	42 [40;46]	43 [40;46]	0.017	243
ALT [IU/l]	21 [16;27]	21 [16;29]	20 [15;24]	22 [17;28]	0.191	242
AST [IU/l]	23 [19;28]	23 [19;28]	22 [19;26]	24 [20;31]	0.158	177
Bilirubin [μmol/l]	10 [8;14]	10 [8;14]	9 [8;14]	10 [7;14]	0.75	243
Total protein [g/l]	72 [68;75]	72 [67;75]	72 [68;75]	71 [68;74]	0.729	185
Sodium [mmol/l]	140 [139;141]	140 [138;141]	140 [139;142]	140 [139;142]	0.251	238
Potassium [mmol/l]	4.1 (0.4)	4.2 (0.5)	4.1 (0.4)	4.1 (0.4)	0.373	238
Urea [mmol/l]	5.6 [4.2;7.0]	5.8 [4.2;7.5]	5.6 [4.5;6.5]	5.2 [4.0;6.5]	0.179	238
Creatinine [μmol/l]	80 [68;96]	82 [69;100]	79 [68;96]	77 [67;91]	0.181	236
NTproBNP [ng/l]	223 [61.5;1230]	332 [75.7;1849]	167 [59.0;984]	191 [60.0;839]	0.242	106
BNP [ng/l]	61.7 [30.1;148]	59.2 [28.8;143]	93.2 [39.9;156]	58.8 [30.5;169]	0.808	99
			Comorbidities			
HTN, n (%)	53 (21)	29 (29)	16 (21)	8 (10)	0.008	254
DM type 1, n (%)	5 (2)	2 (2)	1 (1)	2 (3)	1	254
DM type 2, n (%)	38 (15)	22 (22)	11 (14)	5 (6)	0.013	254
CAD, n (%)	7 (3)	3 (3)	4 (5)	0 (0)	0.106	254
CVA, n (%)	4 (2)	2 (2)	2 (3)	0 (0)	0.471	254
Dyslipidemia, n (%)	21 (8)	5 (5)	9 (12)	7 (9)	0.263	254
AVM, n (%)	2 (1)	1 (1)	1 (1)	0 (0)	0.757	254
Epistaxis, n (%)	3 (1)	1 (1)	1 (1)	1 (1)	1	254
Hepatitis, n (%)	4 (2)	3 (3)	1 (1)	0 (0)	0.317	254
Hypothyroidism, n (%)	37 (15)	16 (16)	12 (16)	9 (11)	0.627	254
COPD, n (%)	8 (3)	5 (5)	3 (4)	0 (0)	0.101	254
Asthma, n (%)	27 (11)	10 (10)	7 (9)	10 (13)	0.766	254
OSA, n (%)	11 (4)	8 (8)	2 (3)	1 (1)	0.086	254
Heterotaxy, n (%)	1 (0)	1 (1)	0 (0)	0 (0)	1	254
Asplenia, n (%)	1 (0)	1 (1)	0 (0)	0 (0)	1	254
Cancer, n (%)	16 (6)	7 (7)	5 (7)	4 (5)	0.899	254

Tertiles' ranges: lower 0 – 5105 pg/ml, middle 5151 – 7699 pg/ml, 7702 – 17806 pg/ml. Data presented as: median [interquartile range], count (%), mean (SD), Abbreviations: IPAH: idiopathic pulmonary arterial hypertension; HPAH: hereditary pulmonary arterial hypertension (diagnosis based on positive family history); BMI: body mass index; WHO FC: World Health Organization functional class; 6MWD: 6 minute walk distance; mRAP: mean right atrial pressure; mPAP: mean pulmonary arterial pressure; PVR: pulmonary vascular resistance; CO: cardiac output; FEV1: forced expiratory volume; FVC: forced vital capacity; KCO: transfer coefficient for carbon monoxide; Hb: hemoglobin, RDW: red cell distribution width; WBC: white blood cell count; Plt: platelets; AST: aspartate aminotransferase; ALT: alanine aminotransferase; BNP: brain natriuretic peptide; NTproBNP: N-terminal pro B-type natriuretic peptide; AVM: arteriovenous malformation (one hepatic AVM and one pulmonary AVM); DM type 1: Diabetes mellitus type 1; DM type 2: diabetes mellitus type 2; OSA: obstructive sleep apnoea; CAD: coronary artery disease; CVA: cardiovascular accident; HTN: systemic hypertension; COPD: chronic thromboembolic pulmonary disease