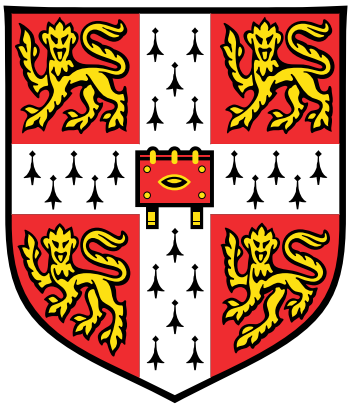
Characterisation of the Tumour Microenvironment in Ovarian Cancer



Alejandro Jiménez Sánchez

Supervisor: Dr Martin L. Miller

Cancer Research UK Cambridge Institute
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

St Catherine's College

October 2018

Para HM

Declaration

I hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in whole or in part for consideration for any other degree or qualification in this, or any other university. This dissertation is my own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and acknowledgements. This dissertation contains fewer than 60,000 words including appendices, bibliography, footnotes, tables and equations and has fewer than 150 figures.

Alejandro Jiménez Sánchez
October 2018

Acknowledgements

Words cannot express my gratitude to each one of you who made this work possible,

literally

Layman's abstract

Metastasis, the process of cancer dissemination, is responsible for at least 90% of cancer deaths. Treating or preventing metastasis remains a long-standing aim in cancer research, however tumours are a complex mixture of cancer cells which precludes the success of therapy. To understand what factors contribute to metastasis and treatment response, it is critical to characterise the tumour microenvironment (non-cancerous cells in the tumour tissue). Ovarian cancer offers a unique opportunity to study the tumour microenvironment because diagnosis often occurs at advanced stages when metastasis has already taken place. This thesis presents a comprehensive analysis of an long-term (>10 years) survivor of metastatic ovarian cancer (5 tumours) who exhibited differential progression/regression of different tumour lesions. Regressing tumours showed a higher expression of genes related to immune cell infiltration and activation, while progressing tumours showed high expression of genes associated with blocking immune infiltration. This demonstrates co-existence of tumour microenvironments within a patient. Then a follow-up study is presented where different tumour microenvironments within patients with metastatic ovarian cancer at diagnosis are characterised (8 patients and 38 tumours). Co-existence of different tumour microenvironments within patients was corroborated, and genes associated with immune exclusion appeared highly expressed in some tumours. Evaluation of the effect of neoadjuvant chemotherapy was assessed in an independent cohort with matched and unmatched pre/post therapy samples (40 patients and 80 tumours). This study showed that chemotherapy induces immune activation, and paradoxically, that low levels of immune activation pre-chemotherapy correlate with immune activation upon chemotherapy treatment. Importantly, these observations were only detected when intra-patient tumour microenvironment differences were considered. Finally, computational tools that predict cell types in tumours using cell-type specific gene sets were benchmarked and a consensus approach (Consensus^{TME}) was developed. Consensus^{TME} outperforms other methods in estimating immune infiltration levels in ovarian cancer and in most tumour types. This study provides a comprehensive examination of the tumour microenvironment in ovarian cancer at diagnosis and in response to chemotherapy, and generated a resource for a consensus computational approach for estimation of immune cells in tumours using gene expression profiles.

Abstract

The tumour microenvironment comprises the non-cancerous cells present in the tumour mass (fibroblasts, endothelial, and immune cells), as well as signalling molecules and extracellular matrix. Tumour growth, invasion, metastasis, and response to therapy are influenced by the tumour microenvironment. Therefore, characterising the cellular and molecular components of the tumour microenvironment, and understanding how they influence tumour progression, represent a crucial aim for the success of cancer therapies. High-grade serous ovarian cancer provides an excellent opportunity to systematically study the tumour microenvironment due to its clinical presentation of advanced disseminated disease and debulking surgery being standard of care.

This thesis first presents a case report of a long-term survivor (>10 years) of metastatic high-grade serous ovarian cancer who exhibited concomitant regression/progression of the metastatic lesions (5 samples). We found that progressing metastases were characterized by immune cell exclusion, whereas regressing metastases were infiltrated by $CD8^+$ and $CD4^+$ T cells. Through a T cell - neoepitope challenge assay we demonstrated that predicted neoepitopes were recognised by the $CD8^+$ T cells obtained from blood drawn from the patient, suggesting that regressing tumours were subjected to immune attack. Immune excluded tumours presented a higher expression of immunosuppressive Wnt signalling, while infiltrated tumours showed a higher expression of the T cell chemoattractant *CXCL9* and evidence of immunoediting. These findings suggest that multiple distinct tumour immune microenvironments can co-exist within a single individual and may explain in part the heterogeneous fates of metastatic lesions often observed in the clinic post-therapy.

Second, this thesis explores the prevalence of intra-patient tumour microenvironment heterogeneity in high-grade serous ovarian cancer at diagnosis (38 samples from 8 patients), as well as the effect of chemotherapy on the tumour microenvironment (80 paired samples from 40 patients). Whole transcriptome analysis and image-based quantification of T cells from treatment-naive tumours revealed highly prevalent variability in immune signalling and distinct immune microenvironments co-existing within the same individuals at diagnosis.

Consensus^{TME}, a method that generates consensus immune and stromal cell gene signatures by intersecting state-of-the-art deconvolution methods that predict immune cell populations using bulk RNA data was developed. Consensus^{TME} improved accuracy and sensitivity of T cell and leukocyte deconvolutions in ovarian cancer samples. As previously observed in the case report, Wnt signalling expression positively correlated with immune cell exclusion. To evaluate the effect of chemotherapy on the tumour microenvironment, we compared site-matched and site-unmatched tumours before and after neoadjuvant chemotherapy. Site-matched samples showed increased cytotoxic immune activation and oligoclonal expansion of T cells after chemotherapy, unlike site-unmatched samples where heterogeneity could not be accounted for. In addition, low levels of immune activation pre-chemotherapy were found to be correlated with immune activation upon chemotherapy treatment. These results corroborate that the tumour-immune interface in advanced high-grade serous ovarian cancer is intrinsically heterogeneous, and that chemotherapy induces an immunogenic effect mediated by cytotoxic cells.

Finally, the different deconvolution methods were benchmarked along with Consensus^{TME} in a pan-cancer setting by comparing deconvolution scores to DNA-based purity scores, leukocyte methylation data, and tumour infiltrating lymphocyte counts from image analysis. In so far as it has been benchmarked, unlike the other methods, Consensus^{TME} performs consistently among the top three methods across cancer-related benchmarks. Additionally, Consensus^{TME} provides a dynamic and evolvable framework that can integrate newer deconvolution tools and benchmark their performance against itself, thus generating an ever updated version.

Overall, this thesis presents a systematic characterisation of the tumour microenvironment of high grade serous ovarian cancer in treatment-naive and chemotherapy treated samples, and puts forward the development of an integrative computational method for the systematic analysis of the tumour microenvironment of different tumour types using bulk RNA data.

Table of contents

List of figures **xvii**

List of tables **xix**

Abbreviations **xxi**

Publications **xxvii**

1 Introduction **1**

1.1 The tumour microenvironment 2

1.1.1 Cellular composition of the tumour microenvironment 3

1.1.2 Cancer and tumour microenvironment cellular interactions 4

1.1.3 Tumour heterogeneity at a multi-scale level 13

1.1.4 Therapy targeting the tumour microenvironment 14

1.1.5 Immunogenomics of the tumour microenvironment 16

1.2 High grade serous ovarian cancer 16

1.2.1 Epidemiology 17

1.2.2 Molecular features 17

1.2.3 Treatment 18

1.3 Computational approaches for the analysis of tumour-stroma interactions . 19

1.3.1 Genomics 20

1.3.2 ImmunoGenomics 23

2 Co-existence of tumour-immune microenvironments in an ovarian cancer patient **29**

2.1 Graphical abstract 30

2.2 Summary 31

2.3 Introduction 32

2.4 Methods 33

2.5 Results . . . 45
2.5.1 Case report . . . 45
2.5.2 Phylogenetic analysis of somatic alterations in tumours . . . 45
2.5.3 Immune-related pathways are over-expressed in regressing tumours 47
2.5.4 Heterogeneous immune infiltration in growing and regressing lesions 49
2.5.5 T cell oligoclonal expansion detected in regressing metastases . . . 52
2.5.6 Peripheral blood CD8+ T cells react against predicted neoepitopes . 55
2.6 Discussion . . . 57
2.7 Contributions . . . 59

3 Unravelling tumour-immune heterogeneity in advanced ovarian cancer uncovers immunogenic effect of chemotherapy 63
3.1 Graphical abstract . . . 64
3.2 Summary . . . 65
3.3 Introduction . . . 66
3.4 Methods . . . 67
3.5 Results . . . 80
3.5.1 Intra-patient transcriptomic heterogeneity is largely explained by immune signalling . . . 80
3.5.2 Co-existence of distinct tumour-immune microenvironments in treatment-naive HGSOC . . . 82
3.5.3 Consensus tumour microenvironment gene sets improve cell estimation from bulk tumour mRNA . . . 84
3.5.4 Tumour microenvironment cell decomposition and molecular comparison of high and low purity treatment-naive HGSOC . . . 88
3.5.5 Chemotherapy induces immune activation in HGSOC . . . 89
3.5.6 Tumour-immune microenvironment intra-patient heterogeneity masks chemotherapy induced immune activation effect . . . 90
3.6 Discussion . . . 95
3.7 Contributions . . . 96

4 Comprehensive benchmarking and integration of tumour microenvironment cell estimation methods 99
4.1 Graphical abstract . . . 100
4.2 Summary . . . 101
4.3 Introduction . . . 102
4.4 Methods . . . 103

4.5 Results . . . 108
4.5.1 Consensus tumour-type specific superset of tumour microenvironment cell populations . . . 108
4.5.2 Pan-cancer stroma, leukocyte, and lymphocyte benchmarks . . . 108
4.5.3 Independent methods benchmarks . . . 111
4.6 Discussion . . . 114
4.7 Contributions . . . 115

5 Discussion **117**
5.1 HGSOC case study . . . 118
5.2 Treatment-naive HGSOC . . . 120
5.3 Pre/post neoadjuvant chemotherapy HGSOC . . . 122
5.4 TME cell estimation from bulk tumour expression data . . . 124

6 Conclusion **127**

References **129**

7 Appendix A **157**

8 Appendix B **165**

9 Appendix C **175**

List of figures

1.1 Cells of the tumour microenvironment 4
1.2 HLA nomenclature system 11
1.3 Endogenous HLA-I epitope presentation pathway 12
1.4 Neoepitope prediction pipeline 26

2.1 Metastatic tumours exhibit heterogeneous growth and somatic mutation patterns after multi-line chemotherapy 46
2.2 Differential expression of immune-related pathways in heterogeneously growing tumours 49
2.3 Immune infiltration status shows heterogeneous microenvironments across tumour samples 51
2.4 Higher HLA expression and T cell oligoclonal expansion detected in regressing tumours 52
2.5 Predicted immunogenicity of HLA class I neoepitopes 54
2.6 Predicted neoepitopes with higher mutant than wild-type HLA-I binding affinity elicit a T cell response 56
2.7 Overall associations between tumour fates and tumour-immune microenvironmental features 58

3.1 Immune signalling contributes to the majority of the transcriptional variance observed across multiple tumour samples from treatment-naive HGSOC patients 81
3.2 T cell infiltrate variation across patients, within patients, and within tumours 83
3.3 Consensus tumour microenvironment cell enrichment method improves estimation of tumour infiltrating T cells and leukocytes 85
3.4 Unbiased analysis of tumour microenvironment heterogeneity in treatment-naive HGSOC tumours 87

3.5 Unbiased signalling pathway and tumour microenvironment cell decomposition analysis of chemotherapy treated HGSOC site-matched and unmatched tumour samples 91
3.6 Immune activation induced by neoadjuvant chemotherapy is evident in site-matched but not site-unmatched sample analysis 93

4.1 Benchmark of methods for estimating TME cell components using purity and leukocyte data from TCGA data 110
4.2 Side-by-side benchmark using available datasets published by the individual methods 112

7.1 Non-silent somatic mutations and copy-number alterations 159
7.2 Gene set enrichment analysis of transcript abundance and somatic alteration patterns across samples 160
7.3 Complete slide haematoxylin and eosin and immunofluorescent staining . . 161
7.4 Neoepitope distributions and HLA-I neoepitope depletion analysis 162
7.5 PBMCs sample time line and T cell-neoepitope recognition assay 163

8.1 Principal component hallmark gene sets' loadings of treatment-naive HGSOC samples 167
8.2 Total counts of T cell subsets in treatment-naive HGSOC samples 168
8.3 Flowcharts for consensus tumour microenvironment (TME) gene set generation and The Cancer Genome Atlas (TCGA) OV leukocyte methylation benchmarks 169
8.4 Principal component TME cell gene sets' loadings of treatment-naive HGSOC samples and gene sets' overlaps 170
8.5 Normalised enrichment scores of estimated TME cells in TCGA ovarian cancer and the treatment-naive high-grade serous ovarian cancer (HGSOC) cohort 171
8.6 Principal component hallmark gene sets' loadings of pre/post-neoadjuvant chemotherapy (NACT) matched and unmatched HGSOC samples 172
8.7 Principal component Consensus^{TME} gene sets' loadings of pre/post-NACT matched and unmatched HGSOC samples 173

List of tables

Tables can be accessed through:
https://github.com/cansysbio/immunogenomics/tree/master/AJSPhDThesis/

Abbreviations

Roman Symbols

$Bobs$	Observed number of neoepitopes
$Bpred$	Expected number of neoepitopes
Bs	Expected number of predicted neoepitopes per missense mutation
D	Diffusion coefficient
F	Volume fraction of blood flowing through micro-vessels
K^{trans}	Volume transfer constant between the blood plasma and the extravascular extracellular space
$\hat{x}$	Median
$\hat{z}_x$	Modified z-score
$Nobs$	Observed number of missense mutations
$Npred$	Expected number of missense mutations
Ns	Expected number of missense mutations per silent mutation
R	Ratio of observed versus expected neoepitopes per tumour sample
Ss	Number of silent mutations per codon change
V	Volume

Greek Symbols

α	Image cluster with lowest intra-cluster distance
β	Image cluster with highest intra-cluster distance

γ	Image cluster with intermediate intra-cluster distance
π	$\simeq 3.14\ldots$

Acronyms / Abbreviations

ACC	Adrenocortical Carcinoma
acctran	Accelerated Transformation Function
AIC	Akaike Information Criterion
ANN	Artificial Neural Network
APC	Antigen Presenting Cells
BCR	B Cell Receptor
BH	Benjamini and Hochberg method
BIC	Bayesian Information Criterion
BLCA	Bladder Urothelial Carcinoma
BRCA	Breast Invasive Carcinoma
CEF	Cytomegalovirus, Epstein-Barr virus, Influenza virus
CESC	Cervical Squamous Cell Carcinoma and Endocervical Adenocarcinoma
CHOL	Cholangiocarcinoma
CNA	Copy Number Alterations
COAD	Colon Adenocarcinoma
CT	Computarised Tomography
CyTOF	Time of Flight Cytometry
DF	Degrees of Freedom
DLBC	Lymphoid Neoplasm Diffuse Large B-cell Lymphoma
DMSO	Dimethyl Sulfoxide
EMT	Epithelial Mesenchymal Transition

ER	Endoplasmic Reticulum
ESCA	Esophageal Carcinoma
ES	Enrichment Score
FDR	False Discovery Rate
FFPE	Formalin-Fixed Paraffin-Embedded
GBM	Glioblastoma Multiforme
GLM	Generalised Linear Model
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis
H&E	Haematoxylin and Eosin
HGSOC	High-Grade Serous Ovarian Cancer
HLA	Human Leukocyte Antigen
HNSC	Head and Neck Squamous Cell Carcinoma
ICS	Intra-Cellular Staining
IEDB	Immune Epitope Database and Analysis Resource
IF	Immunofluorescence
IFN	Interferon
IGV	Integrative Genomics Viewer
IHC	Immunohistochemistry
ILC	Innate Lymphoid Cells
ITH	Intra-Tumour Heterogeneity
KDE	Kernel Density Estimate
KEGG	Kyoto Encyclopedia of Genes and Genomes
KICH	Kidney Chromophobe

KIRC	Kidney Renal Clear Cell Carcinoma
KIRP	Kidney Renal Papillary Cell Carcinoma
LASSO	Least Absolute Shrinkage and Selection Operator
LGG	Brain Lower Grade Glioma
LIHC	Liver Hepatocellular Carcinoma
LOESS	Local Regression
LUAD	Lung Adenocarcinoma
LUSC	Lung Squamous Cell Carcinoma
MAD	Median Absolute Deviation
MESO	Mesothelioma
MHC	Major Histocompatibility Complex
MRI	Magnetic Resonance Imaging
MR	Magnetic Resonance
MSKCC	Memorial Sloan Kettering Cancer Center
NACT	Neoadjuvant Chemotherapy
NES	Normalised Enrichment Score
NK	Natural Killer Cells
OV	Ovarian Serous Cystadenocarcinoma
PAAD	Pancreatic Adenocarcinoma
PBMC	Peripheral Blood Mononuclear Cell
PCA	Principal Component Analysis
PCPG	Pheochromocytoma and Paraganglioma
PCR	Polymerase Chain Reaction
PET	Positron Emission Tomography

PRAD	Prostate Adenocarcinoma
READ	Rectum Adenocarcinoma
RNA-seq	RNA Sequencing
RUQ	Right Upper Quadrant
SARC	Sarcoma
SCNA	Somatic Copy Number Alterations
scRNA-seq	Single Cell RNA Sequencing
SKCM	Skin Cutaneous Melanoma
SNV	Single Nucleotide Variant
ssGSEA	Single Sample Gene Set Enrichment Analysis
STAD	Stomach Adenocarcinoma
SUV	Standarised Uptake Value
t-SNE	t-distributed Stochastic Neighbor Embedding
TAP	Transporter Associated with Antigen Processing Molecules
TCGA	The Cancer Genome Atlas
TCR-seq	T Cell Receptor Sequencing
TCR	T Cell Receptor
TGCT	Testicular Germ Cell Tumors
THCA	Thyroid Carcinoma
THYM	Thymoma
TIL	Tumour Infiltrating Lymphocyte
TME	Tumour Microenvironment
Tregs	Regulatory T Cells
UCEC	Uterine Corpus Endometrial Carcinoma

UCS	Uterine Carcinosarcoma
UPGMA	Unweighted Pair Group Method with Arithmetic Mean
UVM	Uveal Melanoma
VAF	Variant Allele Frequency
WES	Whole Exome Sequencing

Publications

Lead author publications

* Equal contribution. ^ Correspondence.

1) **Jiménez-Sánchez A**, Memon D, Pourpe S, Veeraghavan H, Li Y, Vargas HA, Gill MB, Park KJ, Zivanovic O, Konner J, Ricca J, Zamarin D, Walther T, Aghajanian C, Wolchok JD, Sala E, Merghoub T, Snyder A^ , Miller ML^ . Heterogeneous tumor-immune microenvironments among differentially growing metastases in an ovarian cancer patient. *Cell*; 170: 927-938 (2017). doi:10.1016/j.cell.2017.07.025

2) **Jiménez-Sánchez A***, Cybulska P*, Lavigne K*, Walther T, Nikolovski I, Mazaheri Y, Weigelt B, Chi DS, Park KJ, Hollmann T, Couturier D-L, Vargas A, Brenton JD, Sala E, Snyder A, Miller ML^ . Unraveling tumor-immune heterogeneity in advanced ovarian cancer uncovers immunogenic effect of chemotherapy. *In revision in Nature Genetics*.

3) **Jiménez-Sánchez A***^ , Cast O*, Miller ML^ . Comprehensive benchmarking and integration of tumour microenvironment cell estimation methods. *In revision in Cancer Research*.

Collaborations

Not included in this thesis

4) Shahida D*^ , Wong K*, Mueller M, Oniscu A, Hewinson J, Black C, Miller ML, **Jiménez-Sánchez A**, Rabbie R, Rashid M, Santsangi J, Adams D, Arends M^ . Hypermutation and mutational signatures in inflammatory bowel disease-associated colorectal cancers. *Clinical Cancer Research*; In press. (2018). doi:10.1158/1078-0432.CCR-17-3713.

5) Wolf Y, Patkar S*, Eli BG*, Sang-Lee J, Karathia H, **Jiménez-Sánchez A**, Barnea

E, Chagit, Day CP, Levy R, Zilberberg A, Perez Guijarro E, Miller ML, Efroni S, Merlino G, Admon A, Ruppin E^ , Samuels Y^ . UVB-induced tumor heterogeneity, rather than mutational load, dictate the anti-melanoma immune response. *In review in Cell*.

6) Van den Eynden J^ , **Jiménez-Sánchez A**, Miller ML, Larsson E^ . Lack of detectable neo-antigen depletion in cancer when taking mutational signatures into account. *In submission*.

7) Veeraghavan H*, Vargas HA*, **Jiménez-Sánchez A**, Miccó M, Mema E, Capanu M, Zheng J, Lakhman Y, Huang E, Levine DA, Deasy Jo, Snyder A, Miller ML, Brenton JD^ , Sala E^ . Computed tomography measures of inter-site tumour heterogeneity for classifying outcomes in high-grade serous ovarian carcinoma: a retrospective study. *In submission*.

Chapter 1

Introduction

Cancer is more than one disease, more than a disease of the genome, and more than a group of cells that proliferate uncontrollably. Instead, the word *cancer* as a concept, comprises a number of different tumour types that relate to their tissue and cell of origin. It also includes de-regulation of multiple molecular systems in the cells besides the genome, including the epigenome, the transcriptome, the proteome, the secretome, etc. Finally, tumours are complex semi-organised tissues that interact with their surrounding microenvironment and, therefore, consist of a network of heterotypic interactions between cancerous and non-cancerous cells. Altogether, cancer cannot and should not be understood, conceived, or thought *only* as a disease of the genome any more.

Studying the tumour niche and how the non-cancerous cells interact with the tumour has become a central paradigm in cancer research. As such, integrating different data types beyond genome sequencing through computational systematic approaches is a central endeavour towards gaining a comprehensive understanding of cancer. Additionally, due to the complexity of tumours, it is foreseeable that accounting for the different layers of heterogeneity and differences across tumour types, patients with the same tumour type, and even between tumours within the same patient, would become common practice both for cancer researchers and clinicians.

By employing computational and statistical analyses of tumours from patients with ovarian cancer, this doctoral thesis explores and illustrates how the TME interacts and shapes tumour development, and contributes to the field of immunogenomics of cancer by trying to improve stromal and immune cell estimations using bulk RNA data in a pan-cancer setting.

1.1 The tumour microenvironment

It has been proposed that cancer development can be rationalised as a multi-step process by which cancer cells acquire several traits that allow them to proliferate indefinitely in an uncontrolled manner (Hanahan and Weinberg, 2011). Following this rationale, for cancer cells to succeed at establishing a consolidated tumour, the following traits have been proposed to be sufficient:

1. Constitutive proliferative signalling
2. Bypass growth suppressor signalling
3. Epithelial-mesenchymal transition
4. Replicative immortality
5. Angiogenesis induction
6. Cell death resistance
7. Cellular metabolism adaptation
8. Immune evasion

Furthermore, two traits have been proposed to be consequential of neoplastic development but also facilitators of the carcinogenic process, thus constituting feedback loop traits, these are:

1. Genome instability and mutation acquisition
2. Innate immune tumour-promoting inflammation

These 10 traits together comprise what has been termed "The Hallmarks of Cancer" (Hanahan and Weinberg, 2011). To note, angiogenesis induction, immune evasion, and innate immune tumour-promoting inflammation involve other cells other than the cancer cells themselves, since tumour progression requires the interaction with extrinsic elements as well. In this sense, the collection of the non-cancerous cellular and non-cellular components present in the tumour niche are defined as the TME.

1.1.1 Cellular composition of the tumour microenvironment

As previously noted, tumours are considered as tissue-like entities that, in terms of complexity, closely resemble healthy tissues and organs composition. Therefore, to comprehensively understand tumour progression it is necessary to study the different cell types present in the tumour mass and the TME. The TME consists of signalling molecules and extracellular matrix, but most importantly the non-cancerous cells which can have passive or actively roles in tumour progression as described in the next section.

In the interest of clarity, throughout the text the terms *tumour microenvironment*, *niche*, and *stroma* refer to the non-cancerous elements of the tumour. In addition, despite signalling molecules and extracellular matrix being elements of the TME, they are not directly discussed in this thesis.

A great variety of stromal cell types populate tumours, as they do in healthy organs. These can be classified according to their developmental origin and their functional interactions (**Figure 1.1**):

1. Fibroblasts and adipocytes
2. Neuroendocrine cells
3. Endothelial cells and pericytes
4. Innate immune cells
5. Innate/adaptive immune cells
6. Adaptive immune cells

Each of the cell types and differentiation states that constitute the categories listed above perform distinct functions and promote, maintain, restrict, or counteract tumour progression differently [reviewed in Wang et al. 2017]. The listed TME cell-type categories are critical for various "Hallmarks of Cancer" through heterotypic interactions between cancer cells and cells of the TME, as well as different cell types of the TME as explained in the following section.

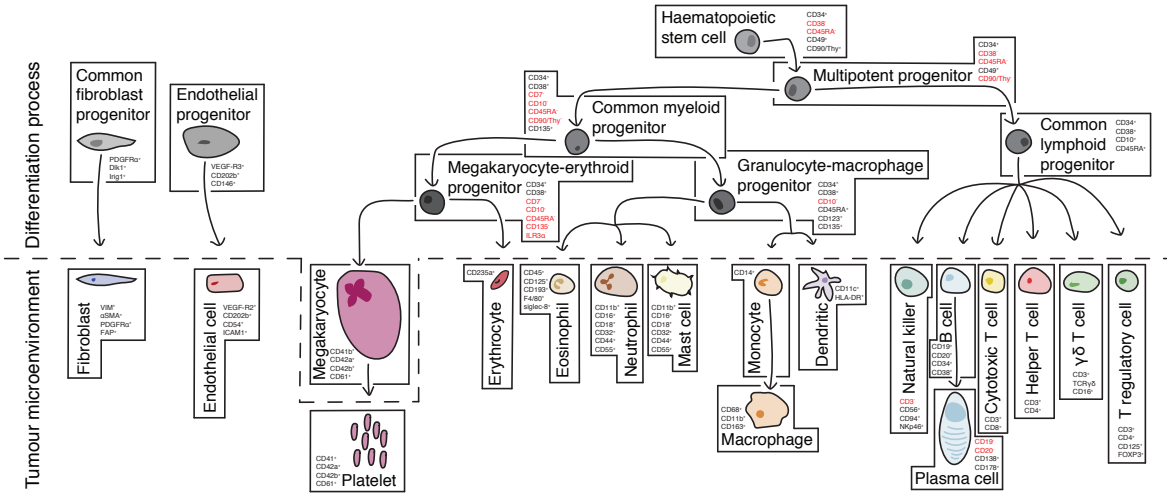


Fig. 1.1 Cells of the tumour microenvironment. Summary of cellular origins and differentiation processes of cells present in the tumour microenvironment with gene classical gene markers. Markers in black are expressed, while markers in red are not.

1.1.2 Cancer and tumour microenvironment cellular interactions

The cellular components of the TME play an active role in tumour progression through constant interactions that evolve into complex interactive networks involving both secreted and membrane-bound molecules (Wang et al., 2017). However, the interactions between the cancer cells and cells of each of the TME cell-type categories are different, as the effect of such interactions. The known effects these TME cells produce in tumours in relation to the "Hallmarks of Cancer" are briefly summarised in this following section. The adaptive immune cell compartment, is described in more detail since the adaptive TME compartment resulted the main focus of the research performed in this thesis.

Fibroblasts and adipocytes

Fibroblasts

Fibroblasts can induce angiogenesis, recruit immune cells, stimulate cancer cell proliferation [reviewed in Kalluri and Zeisberg 2006; Wiseman 2002], and induce epithelial mesenchymal transition (EMT) of cancer cells (Eberlein et al., 2014; Jung et al., 2013).

Adipocytes

Adipocytes are involved in the physiology of energy homeostasis by regulating its storage and release (Kajimura, 2017). Epidemiological studies have shown that the risk of developing

cancer is higher in obese individuals (Calle et al., 2003). Multiple lines of evidence also suggest that adipocytes influence cancer initiation and progression. However, most interactions are related to inflammation [reviewed in Wang et al. 2017], except for the observation that adypocites can induce EMT in breast cancer cells (Devarajan et al., 2012; Pinilla et al., 2009; Zimmerlin et al., 2011).

Neuroendocrine cells

Neuroendocrine cells are present in many organs, where their function is to secrete hormones as a response to a neuronal input. Similar to adipocytes, many independent studies have observed tumour induced effect of neuroendocrine cells. However, most of these associations relate to immune cell recruitment and no systematic analyses have been conducted to critically assess the direct effect of neuroendocrine cells on cancer cells [reviewed in Wang et al. 2017]. Therefore, the active role of neuroendocrine cells in the tumour niche seems to be more related to the cross-talk with other TME cells so far.

Endothelial cells and pericytes

Endothelial cells
Endothelial cells are responsible for the vasculature in the body, both blood and lymphatic vessels. In tumours, therefore, they are the main contributors of the angiogenic switch [reviewed in Hanahan and Folkman 1996].

Pericytes
Pericytes also contribute to the formation of blood vessels in tumours providing more stable vascularity [reviewed in Bergers et al. 2005; Gaengel et al. 2009; Pietras and Östman 2010]. The main effect of vascularisation in tumours is the nutrition supply through the blood, which is required by cancer cells to survive and proliferate, as well as for non-cancerous cells in the tumour niche. However, a consequence of vascularisation in tumours is that it can also facilitate metastatasis [reviewed in Gerhardt and Semb 2008; Raza et al. 2010].

Innate immune cells

The fact that the immune system is crucial in the protection against cancer is well supported [reviewed in Schreiber et al. 2011]. For example, experiments have shown that immunodeficient mice have an increased susceptibility to develop cancer compared to immunocompetent mice (Dighe et al., 1994). However, the immune system consists of numerous different cell types and their interaction with cancer cells and with other cells in the TME is different for

each cell type.

Innate immune cells are characterised by mounting a non-specific, memoryless, quick response to pathogens and cancer. Innate immune cells include dendritic cells, monocytes, macrophages, neutrophils, eosinophils, and mast cells [reviewed in Angell and Galon 2013].

Dendritic cells
Dendritic cells are professional antigen presenting cells that are critical for antitumor immunity, as they can de-novo recruit other immune cells (particularly T cells) and induce an adaptive immune response [reviewed in Corrales et al. 2017; Veglia and Gabrilovich 2017], however they can also contribute to an immune suppressive TME (Tesone et al., 2016).

Monocytes and macrophages
Monocytes are usually found in the blood stream and are the direct precursors of macrophages. Monocytes differentiate into macrophages once they are recruited to a tissue, which includes tumours (Bayik et al., 2017). It has been recognised that macrophages can have anti-tumour or pro-tumorigenic effects, and a simplistic conceptualisation of this dichotomy has been rendered as the *M0/M1/M2 macrophage* framework, where M0 represent neutral macrophages, M1 anti-tumour macrophages, and M2 immune suppressive macrophages [reviewed in Martinez and Gordon 2014; Shapouri-Moghaddam et al. 2018]. However, this M1/M2 categorisation could also be an oversimplification, since tumour associated macrophages can have dual roles in tumour progression depending on the context [reviewed in Martinez and Gordon 2014].

Neutrophils
Neutrophils migrate to sites of infection through chemotaxis and amplify inflammation by the release of cytokines in the site of infection. Neutrophils also phagocyte, de-granulate proteins with anti-microbial properties, and generate neutrophil extracellular traps, which are web-like structures of DNA that trap and kill extracellular microorganisms. The role of neutrophils in cancer depends on several different factors and is not completely understood [reviewed in Ocana et al. 2017]. However, it has been shown that neutrophils can induce angiogenesis (Antonio et al., 2015), tumour growth, and metastasis [reviewed in Swierczak et al. 2015].

Eosinophils
Eosinophil functions include de-granulation of anti-microbial proteins, production of reactive oxygen species, secretion of pro-inflammatory cytokines, growth factors, and enzymes at

the site of inflammation. In cancer, eosinophil infiltration has been associated with better prognosis in general, however the mechanisms behind eosinophil infiltration are still largely unknown and seems to be tumour-type dependent [reviewed in Davis et al. 2014].

Mast cells Similarly, mast cells are important in the inflammatory process. Mast cells can either selectively or rapidly de-granulate pro-inflammatory factors. Mast cells have been associated with pro- and anti-tumourigenic effects, or simply bystanders depending on their surrounding microenvironment and the tumour type [reviewed in Varricchi et al. 2017].

Innate/adaptive immune cells

Natural killer cells

Natural Killer cells (NK) have been originally considered to be part of the innate immune system, as they largely comply with the innate immunity functional characteristics (non-specific, memoryless, rapid response). However, NK cells are more closely related to adaptive immune cells in their differentiation pathway (Hanahan and Weinberg, 2011). Indeed, NK cells interactions with cancer cells occur through different mechanisms: one is called antibody-dependent cellular cytotoxicity, where NK cells bind antibodies that have recognised antigens on the cell surface of cancer cells; another well known mechanism depends on the presence of activating ligands and lack of human leukocyte antigen (HLA) molecules which are essential for T cell recognition as below [reviewed in Lowry and Zehring 2017]; lastly expression of "stress" related ligands that are related to HLA molecules are also recognised and activate NK cells, $\gamma\delta$ T cells, and $CD8^+$ T cells (Bauer et al., 1999). In addition, the ability to generate memory has been observed in NK cells against viral infections [reviewed in O'sullivan et al. 2015]. Despite the type of response, NK cell immunity has been mainly associated with anti-tumour response [reviewed in Lowry and Zehring 2017].

$\gamma\delta$ T cells

Similarly, $\gamma\delta$ T cells belong to the lymphoid branch of the immune system, express specific rearranged T cell receptor (TCR) genes and develop immunological memory. However, they also express invariant receptors like NK cells, which recognise ligands on stressed epithelial cells without specificity or memory [reviewed in Fleming et al. 2017]. Unlike, NK cells, $\gamma\delta$ T cells have been associated with both anti- and pro-tumourigenic effects, and although still under investigation, the different effect seems to be related to which subset of $\gamma\delta$ T cells are present in the TME [reviewed in Fleming et al. 2017; Wu et al. 2016].

Innate lymphoid cells

Innate Lymphoid Cells (ILC) are a type of lymphoid cells that have been defined more recently, and have been associated with both tumour promoting and suppressing phenotypes [reviewed in Wagner et al. 2017]. Although very similar to NK cells, ILCs have been characterised to be less anti-tumourigenic (Gao et al., 2017). ILCs lack rearrangement of specific receptors [reviewed in Mattner and Wirtz 2017], but secrete cytokines that regulate the immune response, including the regulation of T cell infiltration into tumours (Crome et al., 2017). However, the understanding of the function of ILCs in tumours is still incomplete (Wagner et al., 2017).

Adaptive immune cells

The adaptive immunity is characterised by being specific, creating memory, and provide an enhanced response to specific antigens after an initial encounter. Adaptive immunological memory allows long-lasting protection against specific antigens, and thus vaccination is possible. B and T cell lymphocytes are responsible for the adaptive immune response.

B cells

B cells are lymphocytes express specific B cell receptor (BCR) on their plasma membrane, which when activated is secreted resulting in an antibody. In addition, B cells also present antigens and regulate immunity through cytokine secretion. B cells mature in the bone marrow and are characterised by the expression of their BCR on the membrane. Given the right microenvironment, when a BCR recognises an antigen, B cells become activated and differentiate into plasma cells, which finally produce and secrete large amounts of the antibody that recognises the specific antigen, and become memory B cells [reviewed in Cooper 2015]. The secreted antibodies can then bind the antigens on the cancer cells and then be targeted by macrophages, or used as an anchor for NK cell mediated killing. Despite this clear anti-tumourigenic effect, B cells have been also associated with pro-tumourigenic phenotypes, since they secrete cytokines that regulate infiltration of T cell lymphocytes [reviewed in Yuen et al. 2016].

T cells

T cells can be broadly classified in two groups: $CD4^+$ (helper) and $CD8^+$ (cytotoxic) T cells. $CD4^+$ T cells are specialised in the secretion of cytokines that stimulate and regulate the immune response. Besides, $CD8^+$ T cells are specialised in the secretion of perforin and granzyme proteins, which produce pores in the plasma membrane and/or activates the Fas receptor on the target cell leading to the activation of caspases and apoptosis of the target

cell [reviewed in Smith-Garvin et al. 2009]. Furthermore, many studies have shown that T cells play an important role in the recognition and elimination of cancer [reviewed in van der Woude et al. 2017].

T cell specificity

T cell receptor

The expression of TCR genes, and their translocation to the membrane, is what characterise T cells. The T cell receptors are vastly diverse within an individual (over 10^8 different TCRs) (Arstila et al., 1999) because they are generated by somatic genetic recombination of multiple variable immunoglobulin genes [reviewed in Von Essen et al. 2012]. The molecular process that generates this vast TCR diversity is similar to that for antibodies and B cell antigen receptors. The TCR diversity is generated in the thymus by the genetic recombination of the variable, joining, and sometimes diversity gene segments (V(D)J segments) in individual T cells through RAG1 and RAG2 recombinases. To ensure high affinity of the TCR, unlike B cells which undergo "affinity maturation" through somatic hypermutation of their BCR genes, T cells increase their antigen responsiveness through the sustained expression of genes involved in T cell response once they have encountered an antigen, this process is called "functional avidity maturation" [reviewed in Von Essen et al. 2012]. Each recombined TCR has a unique antigen specificity, which is determined by the tertiary protein structure of the antigen-binding site formed by the α and β chains of $\alpha\beta$ T cells or γ and δ chains of $\gamma\delta$ T cells. Both the α and γ chains are generated by the recombination of VJ segments, while β and δ chains are generated by VDJ recombination. The intersection of the VJ or and VDJ chains correspond to the complementary-determining region 3 (CDR3), the most variable region of the TCR and important for antigen recognition on the Human Leukocyte Antigen (HLA). The unique combination of the V(D)J segments in the CDR3 region, along with the palindromic and random nucleotide additions are responsible for the enormous TCR diversity [reviewed in Alcover et al. 2018].

HLA and epitope determinants

T cells use their TCR which recognise peptides bound to Major Histocompatibility Complex (MHC) molecules, called Human Leukocyte Antigen (HLA) system in humans, on the plasma membrane of other cells. Peptides bound to HLAs are called epitopes, as they can be recognised by a TCR. There are two types of HLA molecules: HLA class I molecules are expressed in every nucleated cell and display intracellular peptides, whereas HLA class II molecules are expressed by Antigen Presenting Cells (APC)s (e.g. dendritic cells and macrophages) and display peptides from extracellular origin after being phagocyted by the

APC. The expression level of HLA-I molecules differs among different cell types, and it is subject to regulation by various cytokines. The process by which a peptide is displayed on the plasma membrane is slightly different for HLA-I and HLA-II molecules [reviewed in Neefjes et al. 2011]. For the HLA-I system, proteins are degraded in the cytoplasm of cells by the proteosome, then the generated peptides are translocated into the endoplasmic reticulum (ER) by transporter associated with antigen processing molecules (TAP) (Reits et al., 2000; Schubert et al., 2000). Once in the ER, peptides form non-covalent interactions with HLA-I molecules, and if the affinity between the peptide and the HLA is strong enough to form a stable complex, the complex leaves the ER through the secretory pathway to reach the cell surface [reviewed in Vyas et al. 2008]. For the HLA-II system, extracellular proteins are phagocyted by APCs, enzymatically processed in the early endosome, and transported trough vesicle traffic into the MHC class II compartment. Finally, stable complexes are transported to the plasma membrane. $CD8^+$ T cells recognise HLA-I, while $CD4^+$ T cells recognise HLA-II molecules. To allow T cells to differentiate non-self from self epitopes, T cells undergo a process of maturation and selection in the thymus where positive (HLA recognition) and negative selection (tolerance against self-proteins) take place [reviewed in Parkin and Cohen 2001]. Therefore, a diverse and ample repertoire of T cells, each one with a different TCR, surveys a sample of the body's proteome. If a T cell recognises an epitope as non-self and co-stimulatory signals take place, the T cell becomes activated and the epitope is considered an antigen (Aptsiauri et al., 2008; Hamaï et al., 2010).

HLA variability

HLA molecules are polymorphic in the human population with 13,680 HLA-I and 5,091 HLA-II different alleles identified so far (August 2018) (http://www.ebi.ac.uk/ipd/imgt/hla/stats.html). There are 3 classical HLA-I genes (A, B, C) and 6 main HLA-II (DRA, DRB, DQA1, DQB1, DPA1, DPB1) genes. HLA-I genes form heterodimers with β_2-microglobulin, while HLA-II form homodimers. HLA genes are co-dominantly expressed, thus each person can express up to 6 different HLA-I and 12 HLA-II alleles (**Figure 1.2**) [reviewed in Shiina et al. 2009].

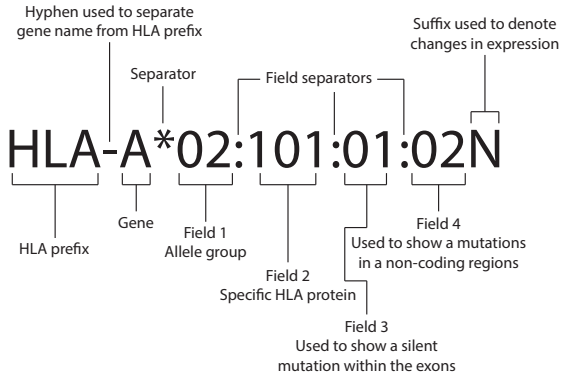


Fig. 1.2 HLA nomenclature system. Each HLA allele has a unique name composed of the HLA prefix, the gene name (A, B, or C) and up to four sets of digits (fields) separated by colons.

Patterns of epitope presentation

HLA molecules are able to present a wide range of epitopes, because only two or three amino acid positions in the peptide interact directly with the HLA, thus all the other amino acids can vary. Each HLA allele binds different sets of peptides with particular characteristics, however general binding patterns have been elucidated.

HLA-I molecules bind peptides of length 8-11 (Eichmann et al., 2014), most often 9 amino acids, while HLA-II molecules bind peptides 14-20 amino acids [reviewed in Mohan and Unanue 2012]. Certain amino acid positions in the peptides interact through non-covalent bonds with the HLA molecule, for example positions 2-3 and positions 8-10 usually interact with HLA-I (Fritsch et al., 2014), while amino acids in the middle arch up and away from the HLA and are more visible to the TCR. Likewise, certain amino acid physicochemical properties that facilitate non-covalent bonds tend to be enriched in such positions. Thus HLA-I molecules are constantly displaying a sample of peptides of all proteins expressed within the cell, and extracellular proteins through HLA-II molecules on APCs.

T cell development

HLA-I restriction and self-tolerance is possible because during their development, T cells are positively (HLA-I restriction) and negatively (self-tolerance) selected in the thymus [reviewed in Parkin and Cohen 2001]. Once the receptor rearrangement has occurred in the bone marrow, T cells move to the thymus where $CD8^+$ and $CD4^+$ T cells that recognise self HLA-I and HLA-II molecules, respectively, are positively selected. Then, T cells with high affinity to self-epitopes die by apoptosis (negative selection). Therefore, T cells that recognise HLA molecules with an epitope but are self-tolerant leave the thymus.

T cell activation

Therefore, every T cell has an antigen specific and clonally restricted receptor, which means that a T cell can become active only when it finds an epitope that is recognised as non-self [reviewed in Huppa and Davis 2003]. Importantly, the TCR is associated with co-receptors needed for a full activation; without these co-signals the T cell becomes anergic or dies by

apoptosis. When a cytotoxic T cell becomes active, it secretes perforin and granzymes to the target cell, and/or activates the fas and TRAIL receptors on the target cell; both strategies lead to the activation of caspases and the death of the target cell by apoptosis [reviewed in Smith-Garvin et al. 2009].

Neoepitopes

T cells react against non-self epitopes, like viral or bacterial epitopes, but also to some self-epitopes which could lead to autoimmunity. In cancer, three classes of self-epitopes have been defined: 1) Proteins that are expressed only during early development before the immune system has matured and therefore are ectopically expressed by cancer cells, 2) over-expressed proteins, and 3) mutant proteins [reviewed in Vigneron 2015]. Particularly, mutant proteins play an important role in the detection of cancer mediated by T cells because mutant proteins arise from somatic mutations in cancer cells and are presented HLA molecules, thus being encountered as new peptide sequences by T cells. Epitopes that originate from non-silent somatic mutations are called neoepitopes, and if a T cell reacts against a neoepitope, then the neoepitope is considered a neoantigen (**Figure 1.3**). Any expressed mutation that produces a change in the protein sequence could, in principle, generate neoepitopes, including missense mutations (point mutation that causes change of one amino acid) which occur frequently in some cancer types (Brown et al., 2014; Gilboa, 1999; Schreiber et al., 2011; Schumacher and Schreiber, 2015).

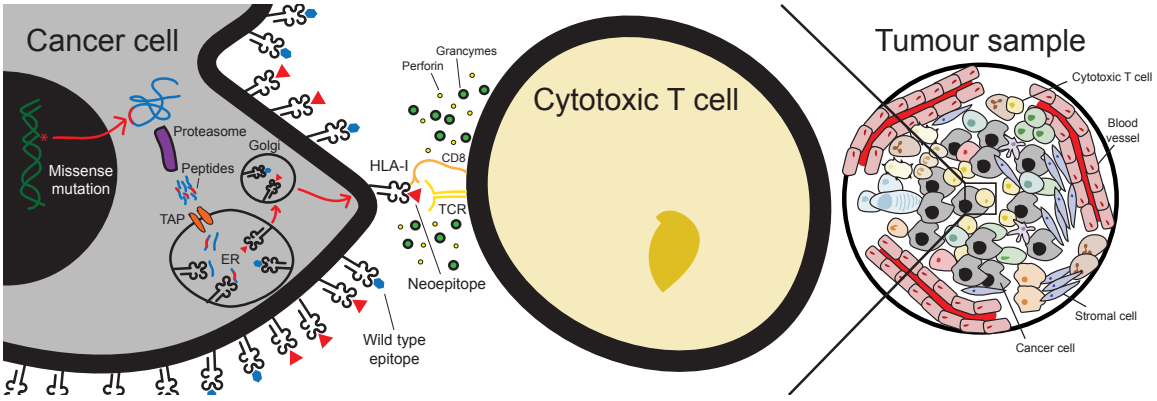


Fig. 1.3 Endogenous HLA-I epitope presentation pathway. Expressed genes, including mutated genes, are degraded by the proteosome in the cytosol as part of the normal protein turn-over. The peptides with and without the mutation are generated, and translocated into the endoplasmic reticulum (ER) by the transporter associated with antigen processing (Reits et al., 2000; Schubert et al., 2000). The HLA-I peptide binding takes place in the lumen of the ER. If the complex is stable, the loaded HLA-I molecule leaves the ER through the secretory pathway to reach the cell surface where the epitope can be recognised by a $CD8^+$ T cell. An analogous process occurs in antigen presenting cells, using HLA-II molecules and presenting both to $CD8^+$ and $CD4^+$ T cells.

Immunoediting and the 3 Es

As explained above, T cells are able to selectively recognise neoepitopes displayed by cancer cells, and these neoepitopes can be predicted. However, this selective recognition imposes a natural selection process where preferential elimination of immunogenic cancer clones leave less immunogenic cancer cells to progress, which may partly explain why cancer occurs in immunocompetent individuals (Matsushita et al., 2012). Moreover, chronic inflammation generated by the constant interaction between cancer and immune cells often lead to immunosuppressive signals from the tumour microenvironment allowing cancer cells to further proliferate and finally evade the immune system [reviewed in Teng et al. 2015]. The process that comprises both tumour suppression and promotion is termed cancer immunoediting (Schreiber et al., 2011), in which neoepitopes play a crucial role as they are one of the key elements of selective recognition mediated by T cells (Mittal et al., 2014; Schreiber et al., 2011). Immunoediting involves three different phases: *Elimination* of cancer cells by the immune system, *equilibrium* of cancer and immune compartments in the TME, and *escape* of cancer cells. Nevertheless, the dynamics of immunoediting throughout tumour progression and the consequence of constant interactions between cancer and immune cells are still largely unknown.

Plenty of evidence supports the role of molecular interactions between cancer cells and the TME on tumour initiation, progression, and metastasis. These interactions shape and constrain tumour evolution, which can lead to resistance to therapy and immune evasion. However, still the role of various cell types is unclear or, so far, seems to be tumour type related. New definitions of cell types or cellular states further complicate the elucidation of their true biological role in different cancer types. In addition, cancer heterogeneity, an intrinsic feature of cancer, poses a remarkable challenge to the study of the effect of each of the components of TME in tumour progression.

1.1.3 Tumour heterogeneity at a multi-scale level

Tumours are not only comprised by various different cell types, cellular states, and cellular interactions, but cancer cells themselves acquire new and different mutations as the tumour develops, leading to a natural state of intra-tumour heterogeneity (ITH), both spatially and temporally [reviewed in McGranahan and Swanton 2017]. Indeed, heterogeneity in cancer spans multiple dimensions: at the molecular level with the presence of genetic ITH (Campbell et al., 2010; Gerlinger et al., 2012; Shah et al., 2009), the cellular level with variability observed in infiltration and recruitment of non-tumour cell populations in the TME

(Natrajan et al., 2016), and the spatial and population levels where variability is observed both within tumours of the same individual with disseminated disease (Jiménez-Sánchez et al., 2017) and between tumours of different patients (McGranahan et al., 2016; Patch et al., 2015). However, the extent of TME heterogeneity and its dynamics with ITH have not been systematically characterised in metastatic disease. The heterogeneous nature of tumours has halted the potential success of cytotoxic chemotherapy and targeted therapies, specially in the more advanced stages of disease. Therefore, studying and understanding the driving forces of TME heterogeneity along with ITH, and accounting for such variations are critical aims to achieve in cancer research both from the basic and clinical point of view.

1.1.4 Therapy targeting the tumour microenvironment

As expected, most traditional cancer therapeutic strategies were originally conceived to target the cancer cells of the tumour. The main aim is to maximise the effect of therapy against the cancer cells, and minimise its effect to normal cells. Therefore, the rationale behind cytotoxic chemotherapy and radiation therapy for instance, was to take advantage of the fact that cancer cells are continuously dividing, and therefore are more susceptible to the DNA damaging effects of radiotherapy and chemotherapy. However, these therapies also damage healthy tissue and side effects become a limiting factor for the increase of dosage or therapy duration. In addition, disease recurrence is common in patients with most cancers after cytotoxic and radiation therapy, since residual disease is difficult to detect. More recently, with the advent of high-throughput sequencing technologies, mutations were thought to provide a source for selective therapeutic target as non-cancer cells do not have such genetic alterations. Particularly, targeting critical proliferation or survival pathways that have been altered in the cancer cells are subject of intense research, clinical trials, and therapies. Indeed, this strategy has improved patient outcomes in various tumour types, however response duration is often limited [reviewed in Fisher et al. 2011]. The reason behind resistance to targeted therapy is thought to be mainly due to cancer heterogeneity. As explained above, tumours are dynamic complex semi-organised tissues that become more heterogeneous during the course of disease. Genetically, ITH results in a non-uniform distribution of distinct cancer sub-clones, sharing clonal mutations but harbouring distinct sub-clonal ones [reviewed in Dagogo-Jack and Shaw 2017]. Therefore, if a sub-clonal mutation is the target of therapy, recurrence is almost inevitable. In addition, most mutations are difficult to target, and knowing all sub-clones in a tumour, is clinically impractical and contradictory, because that would require the resection of the complete tumour mass [reviewed in McGranahan and Swanton 2017]. Overall, the heterogeneous nature of tumours has limited the efficacy of cytotoxic chemotherapy and targeted therapies, especially in the more advanced stages of the disease.

Observations on the effect of different cells of the TME has shifted the attention towards targeting non-cancerous cells that promote tumour progression, inducing the infiltration or enhancing the activation of anti-tumour immune cells. For instance, anti-angiogenic therapy aims to inhibit the formation of blood vessels in the tumour and reduce the blood supply and nutrient distribution to cancer cells through reducing production, inhibiting or blocking the action of pro-angiogenic factors secreted by the tumour cells (e.g. vascular endothelial growth factor). Anti-angiogenic therapies elicit survival benefits in many tumour types, however the clinical response is transient and resistance emerges [reviewed in Bergers and Hanahan 2008].

Another approach when targeting cells of the TME is to enhance the anti-tumour effect of immune cells, in particular T cells. These therapies have been focused on inducing T cell proliferation, survival, and differentiation (e.g. interleukin-2 stimulation) [reviewed in Rosenberg 2014], increasing the number of anti-tumour T cells *ex vivo* and re-infuse them into the patient (e.g. adoptive cell therapy) (Dudley et al., 2002; Rosenberg and Restifo, 2015; Rosenberg et al., 2011), and interfering with molecular mechanisms that inhibit T cell function (e.g. checkpoint-blockade inhibitors) [reviewed in Sharma and Allison 2015]. By far, checkpoint-blockade immunotherapies have showed the greatest success rate, since these therapies have showed objective response rates in $\sim$ 10 different tumour types, ranging from 15% (Head and Neck, Gastroesophageal, and Bladder cancer) to 87% (Hodgkin's disease) objective response rates [reviewed in Ribas and Wolchok 2018]. Most importantly, when successful, checkpoint-blockade immunotherapy has shown durable long-lasting response [reviewed in Sharma and Allison 2015]. However, most tumor types have not yet responded to checkpoint-blockade immunotherapy, and for tumours that show response, there is still a subset of patients that have not received benefit [reviewed in Sharma et al. 2017]. Indeed, research has started shed light on the effect of ITH and response to checkpoint blockade immunotherapy, where it has been shown that less heterogeneous tumours tend to have a slightly better response (McGranahan et al., 2016; Miao et al., 2018).

Recent preclinical and clinical data has shown that the partial success of conventional chemotherapies can be to some extent attributed to immune-stimulatory mechanisms [reviewed in Gotwals et al. 2017]. Thus, the exploration of combination strategies where cancer cells and TME cells are targeted, with special focus on understanding the underlying variables behind response or lack or response, has become an area of intense research in the last years [reviewed in Gotwals et al. 2017; Khan and Kerbel 2018; Sharma et al. 2017]. However, a different avenue towards increasing response rates is to achieve a better understanding of cancer and the TME at diagnosis and in response to standard of care using systematic and

unbiased approaches (Binnewies et al., 2018; Miao and Van Allen, 2016; Valkenburg et al., 2018).

1.1.5 Immunogenomics of the tumour microenvironment

Different strategies have been applied to study the TME: *in vitro* studies of cell-cell communication [reviewed in Byrne et al. 2014; Wu and Swartz 2014], *in vivo* studies with model organisms trying to recreate the human disease [reviewed in Gómez-Cuadrado et al. 2017; Li et al. 2018; Lindner 2014], patient-derived xenografts (Zhao et al., 2018), and finally using tumours from clinical samples [reviewed in Binnewies et al. 2018]. All these approaches provide different levels of information and complement each other. By employing high-throughput sequencing technologies and with the help of computational analyses, more systematic studies on human data can be performed. Indeed, the generation of thousands of publicly available genomic, epigenomic, transcriptomic, and proteomic data of primary tumours in 33 tumour types through the TCGA project, and some metastatic tumours has allowed an exhaustive characterisation of tumours, including their immune landscape (Thorsson et al., 2018). Despite this massive enterprise, questions related to intra-patient TME heterogeneity, TME response to therapy, and proper benchmarks and computational methods for the characterisation of the TME in cancer remain unsolved. To overcome this limitation, further analyses on multiple tumours from the same patient, longitudinal studies, and integrative computational methods are required.

This PhD work has been focused on immunogenomic analyses of a case study of metastatic HGSOC, treatment-naive HGSOC patients with disseminated pelvic disease, the effect of chemotherapy on the TME of HGSOC, and the development of an integrative approach for the estimation of TME cells using bulk expression data along with a systematic pan-cancer benchmarking system.

1.2 High grade serous ovarian cancer

Ovarian cancer spans several different sub-types according to their histology. Importantly, each sub-type has different risk factors, cells of origin, genomic alterations, clinical features and treatments options. However, $\sim$90% of ovarian cancers fall in the category of epithelial cancers, of which high-grade serous ovarian cancer (HGSOC) is the most commonly diagnosed and lethal of the ovarian cancer subtypes. high-grade serous ovarian cancer (HGSOC)

originates from the fallopian tubes, not from the ovaries themselves [reviewed in Bowtell et al. 2015].

1.2.1 Epidemiology

Ovarian cancer is the 7th most commonly diagnosed cancer among women in the world [reviewed in Brett M. et al. 2017]. Currently, the risk for a woman to develop ovarian cancer is 1 in 75, and the probability of dying of the disease is 1 in 100 (https://seer.cancer.gov/archive/csr/1975_2013/). The disease is usually diagnosed at late stage when the 5-year relative survival is 29%. Only 15% of cases are diagnosed at stage I, when the 5-year survival rate is 92%. World wide the overall 5-year survival rate ranges between 30% and 40%, and unfortunately only 2% to 4% increase has been achieved since 1995 (Allemani et al., 2015). These alarming statistics put forward the need for more research in every angle of ovarian cancer, from early detection, to new therapeutic strategies. Nevertheless, a better understanding of the disease is crucial to increase the survival rates.

1.2.2 Molecular features

Early detection strategies in ovarian cancer have not been effective. However, genetic risk factors have been identified. Germline mutations in genes that encode proteins involved in homologous recombination DNA damage repair (e.g. BRCA1/BRCA2) increase the risk of developing ovarian cancer [reviewed in Matulonis et al. 2016]. Approximately 20% HGSOC cases harbour germline or somatic mutations in BRCA1 or BRCA2. Mutations or epigenetic silencing of other homologous recombination DNA damage repair related genes also increase risk of developing HGSOC, but are less prevalent in the population (TCGA, 2011). However, so far genomic stratification has been limited only to homologous recombination deficiency (Fong et al., 2010; Gelmon et al., 2011; Swisher et al., 2017).

In general, HGSOC is characterised by a preponderance of DNA copy-number alterations with highly complex genomic profiles (Hoadley et al., 2014; Macintyre et al., 2018), and a modest somatic missense mutation burden ($\sim$61 per exome) (Ciriello et al., 2013; Patch et al., 2015; TCGA, 2011). However, recurrent oncogenic mutations and copy number alterations are uncommon (Hoadley et al., 2014), except for TP53 mutations which are ubiquitous in HGSOC (Ahmed et al., 2010). Analyses of whole-genomes have shown that loss of RB1, NF1, and PTEN by gene breakage are common (Patch et al., 2015), as well as amplification associated fold-back inversions in tumours lacking alterations in homologous recombination genes (Wang et al., 2017). In addition, analysis of gene expression profiles has shown

predominantly differences related to the TME, although reliable only in a subset of patients (Chen et al., 2018; Verhaak et al., 2012). Importantly, it has been shown that, genetically, HGSOC is a very heterogeneous disease, even at diagnosis (Bashashati et al., 2013; De Mattos-Arruda et al., 2014; Lee et al., 2015).

1.2.3 Treatment

HGSOC is usually diagnosed at late stages, when the disease has already started to spread. In patient with symptoms, diagnosis involves physical pelvic and recto-vaginal examination, and radiographic imaging [trans-vaginal and/or abdominal ultrasonography, Computarised Tomography (CT), Magnetic Resonance Imaging (MRI), and/or Positron Emission Tomography (PET)]. The biomarker CA125 (also known as mucin-16), which is preferentially expressed by ovarian cancer cells, is also commonly used for diagnosis and to follow-up treatment response. Surgery with removal of all visible disease is recommended (Demir and Marchand, 2012), and further histological analysis is performed. Finally, ovarian cancer diagnosis is confirmed after a tissue biopsy is analysed [reviewed in Matulonis et al. 2016].

The therapeutic strategy for newly diagnosed HGSOC involves primary surgical cytoreduction (tumour debulking) followed by platinum-based chemotherapy or neoadjuvant chemotherapy Neoadjuvant Chemotherapy (NACT) - administration of chemotherapy before surgery -, then followed by interval surgery and additional chemotherapy after surgery. Currently, the most active therapeutics against newly diagnosed HGSOC are platinum (cisplatin or carboplatin) with taxane analogues (paclitaxel or docetaxel) [reviewed in Matulonis et al. 2016]. Unfortunately, resistance to chemotherapy is developed by nearly all patients diagnosed at late stages of the disease. Anti-angiogenic treatments, DNA damage repair blocking agents (poly(ADP-ribose) polymerase (PARP) inhibitors), and immunotherapy are strategies that have been used or researched for patients with recurrent disease which usually presents platinum resistance [reviewed in Cojocaru et al. 2018; Matulonis et al. 2016]. Overall, despite surgical, chemotherapy, and targeted therapeutic interventions, most patients relapse and develop resistance to therapy, which leads to premature death (Bowtell et al., 2015; Cannistra, 2004).

Analysis of data from various cancer types studied by the TCGA consortium, including ovarian cancer, has demonstrated that the number of somatic mutations and neoepitopes (**see Neoepitopes in section 1.1.2 under T cell specificity**) correlate with overall survival (Brown et al., 2014). Similarly, mutation burden and response to checkpoint-blockade immunotherapy correlate (Yarchoan et al., 2017). In HGSOC, the low somatic point mutation

load, high aneuploidy levels and high copy number alterations have been associated with lack of immunogenicity, thus HGSOC has been considered as a non-immunogenic tumour type (Bowtell et al., 2015; Zhang et al., 2018). Despite the low intrinsic immunogenicity, T cell infiltration plays a major role in predicting HGSOC survival in a primary disease setting (Ovarian Tumor Tissue Analysis (OTTA) Consortium et al., 2017; Zhang et al., 2003), and recent studies have started to shed light on the interplay between ITH and T cell interactions (Zhang et al., 2018), as well as the potential effect of chemotherapy on T cell infiltration in HGSOC (Böhm et al., 2016). Yet, check-point blockade immunotherapy has not been successful in this disease (Homicsko et al., 2016). Different mechanisms or genomic features could be at play to explain the lack of response, but besides the low intrinsic immunogenic feature of HGSOC, the high ITH in HGSOC could also be a limiting factor. This is because tumour heterogeneity increases the likelihood of selection of subclones that escape the immune system (Bhang et al., 2015; Su et al., 2012; Turke et al., 2010), in addition to the the low number of potential neoepitopes. Additionally, little is known about the effect of other treatments like chemotherapy on the cancer-immune interactions. For example, some studies suggest that chemotherapy may promote immune activation (Carson et al., 2004; Gavalas et al., 2010; Pfirschke et al., 2016), but chemotherapy may also increase genetic heterogeneity (McGranahan et al., 2016). Thus, a unified model of the effect of chemotherapy on the tumour heterogeneity and immune-tumour interactions has not yet been reached.

Importantly, due to the sample acquisition from surgical debulking as standard of care of HGSOC, this clinical setting offers the opportunity to explore and characterise systematically the TME at diagnosis and in response to treatment. Moreover, the intrinsic heterogenous nature of HGSOC provides an ideal framework for the analysis of TME heterogeneity in tumours. By integrating genomic, transcriptomic, and multiplexed imaging analyses, through state-of-the-art computational analyses, cancer-TME dynamics will start to be understood and potentially inferred.

1.3 Computational approaches for the analysis of tumour-stroma interactions

While the study of ITH has been facilitated by computational approaches to estimate the distribution and co-existence of different tumour clones from large-scale somatic mutation data, bioinformatics analysis of TME heterogeneity poses considerable technical difficulties due to the huge diversity of cell types present. This section presents a brief summary of data

types, algorithms, computational methods, and statistical approaches employed in this thesis for the study of the TME.

1.3.1 Genomics

Whole-exome sequencing

Whole Exome Sequencing (WES) sequencing targets the exons of all genes in the genome, although 100% coverage of exons is not achieved. The exome constitutes up to $\sim$1% of the genome. The process begins with the extraction of DNA from the cells, followed by its fragmentation. Then synthetic DNA linkers are added to the end of the fragments and coding sequences are selected, replicated and amplified through several rounds. The amplified fragments are then sequenced in a series of reactions in which complementary nucleotides coupled with different fluorescent dyes per base are added to clusters of identical DNA molecules. The fluorescent colour released during the elongation is imaged with a laser camera coupled with a microscope, then the fluorescent indicator is removed, and the cycle is repeated. Nucleotide sequence reads that are between 75 to 150 nucleotides in length are generated. The sequence reads are aligned to a reference genome (haploid mosaic of different DNA sequences from multiple donors, which is periodically revised), and finally each position in the exome is compared to the reference genome, and the sequences generated [reviewed in Biesecker and Green 2014]. Through this process, genetic alterations (single nucleotide variant, insertion/deletions, structural variants) in exonic regions of tumours can be sought.

Mutation calling

Mutation calling refers to the computational exercise of identifying mutations in an exome or genome and differentiating them from sequencing errors. In principle, given enough sequencing read depth (the number of times a fragment has been sequenced) and coverage (how many regions in the exome/genome are sequenced), all mutations could be *observed*. However, still *calling* the mutations, i.e. distinguishing mutations from sequencing errors and noise, has to be performed. First, low quality reads and sequencing linkers (synthetic adapters) are trimmed using read processing algorithms. Second, the cleaned reads are mapped and aligned to the reference genome. It is recommended to the perform PCR de-duplication, indel-realignment, and base quality recalibration (DePristo et al., 2011; McKenna et al., 2010). Variant calling is the last step, where real variants are discriminated from artefacts generated during library preparation, sample enrichment, sequencing, and mapping/alignment. Mutations with low variant allele frequency, due to impure sample or

small subclone population, are the most challenging to discriminate from noise [reviewed in Xu 2018].

Matched tumour-normal mutation calling is the preferred procedure for mutation calling, where the idea is to identify somatic from germline variants and artefacts using the matched normal sample. Multiple different algorithms have been developed following heuristic, Bayesian, expectation-maximisation, machine learning algorithms (random forest, Bayesian adaptive regression tree, support vector machine, and logistic regression), or a combination of approaches. Independent benchmarks have shown that MuTect (Cibulskis et al., 2013) is among the most reliable single nucleotide variant (SNV) mutation callers. MuTect was designed to improve calling low-allelic-fraction mutations (variant allele frequency (VAF) $\leq$ 0.1) by applying a Bayesian classifier. In short, MuTect models the variant allele fraction and does not assumes a heterozygous diploid event unlike other methods. MuTect identifies variants by analysing each site under two alternative models: (1) a reference model M_0, which assumes there is not variant at the site and any observed non-reference bases are due to random sequencing errors, and (2) an alternative model M_f^m, which assumes the site contains a true variant allele m at allele fraction f in addition to sequencing errors. The allele fraction f is unknown and is estimated as the fraction of tumour sample reads that support m. Then m is considered a candidate variant if the log-likelihood ratio of the data under the variant and reference models exceeds a predefined threshold that depends on the expected mutation frequency and the desired false positive rate. The threshold can be changed to control the trade-off between specificity (Number of not called variants that are artefacts) and sensitivity (Number of called variants that are mutations) (Cibulskis et al., 2013).

Copy number alteration calling

Copy Number Alterations (CNA) are commonly assessed through different technologies other than WES, like comparative genomic hybridization, single nucleotide polymorphism arrays, and low-coverage whole-genome sequencing (Scheinin et al., 2014; TCGA, 2011; Wang et al., 2004). However, performing these analyses increase cost significantly. Different methods have been developed to derived CNA from WES, and two different approaches have been followed: (i) using on-target reads or (ii) off-target reads. The advantage of using off-target reads over on-target reads is that information from non-exonic regions is captured, and the variation in the efficiency of capture baits in off-target regions presents low variation compared to on-target reads. Different statistical methods have been employed to overcome the large variation when using on-target reads (principal component analysis (PCA), hidden Markov models, singular value decomposition) (Tan et al., 2014). Nevertheless, the accuracy

of CNA calling using WES is poor compared to dedicated technologies (Tan et al., 2014). Since targeted WES targeted sequencing generally achieve between 40% to 60% efficiency (Samuels et al., 2013), off-targets reads represent a great source to calculate CNA. This is achieved by CopywriteR using the following steps: (1) exome-enriched reads are used as input (e.g. WES), (2) low quality and anomalous read pairs are removed, (3) genomic regions enriched for sequencing reads are identified and discarded in sample and reference (these correspond to exonic regions). (4) Then CopywriteR looks at the background reads and maps them into bins, (5) then compensation in removed peak regions is performed, (6) depth of coverage is corrected for GC-content and mappability, (7) then median normalisation and log_2 transformation of the corrected and compensated depth of coverage is performed. (8) Finally, substraction of log_2 transformed, corrected, and compensated depth of coverage of the normal sample is used to create relative copy number profiles (Kuilman et al., 2015). This approach showed more accuracy at calling CNA from WES than other methods based on on-target reads, and a comparable copy number detection to the obtained through CNA detection dedicated methods (Kuilman et al., 2015).

Tumour similarity

As noted before, tumours are a complex mixture of cancer and non-cancer cells. Moreover, cancer cells in the tumour are not all equal or equivalent, they are genetically heterogeneous. Leaving aside non-cancer cells, tumours are composed by a complex mixture of different subclonal cancer populations. This genetic diversity within tumours acts as a substrate for natural selection to operate and thus, tumour evolution to take place. As such, tumour evolution intervenes in immune evasion, metastasis, and resistance to therapy [reviewed in McGranahan and Swanton 2017]. Using WES of multiple regions from a tumour or from different metastasis from a patient various studies have tried to reconstruct the tumour phylogeny [reviewed in Alves et al. 2017]. However, due to the extensive ITH, unless CNA, mutation clonality, and tumour purity are properly accounted for, and clonal trees inferred, sample trees using presence/absence mutation data only provide genomic similarity between tumours or regions, but not a tumour phylogeny [reviewed in Alves et al. 2017]. Making the correct interpretation of both type of analyses could be a powerful and complementary strategy.

In this work a similarity presence/absence mutation analysis was performed using only non-synonymous mutations. A common method used for presence/absence similarity dendogram is employing unweighted pair group method with arithmetic mean (UPGMA) hierarchical clustering followed by the parsimony ratchet analysis (Nixon, 1999; Schliep, 2011). First, the

UPGMA algorithm constructs a dendogram that reflects the structure present in the pairwise similarity matrix, where the nearest two clusters are combined into a higher-level cluster. The distance between any two clusters is the mean distance between elements of each cluster. With the clusters and distance calculated, a dendogram can be generated, where the length of the branches are proportional to the similarity between clusters, and corresponds to the number of mutations between the tumours analysed. This dendogram is then passed as starting point for the parsimony ratchet heuristic method which (1) randomly selects a subset of characters, (2) each of which is given additional weight, (3) then branches are swapped using the re-weighted matrix keeping only one tree. (4) Then set all weights for the characters to the original weights, (5) followed by branch swapping. Finally return to step (2) and re-start the cycle iteratively 50 to 200 times. The last selected tree is the most likely parsimonious solution, and branch lengths are then calculated and proportional to mutation counts. As explained above, strictly speaking this approach does not provide a true phylogeny, but a similarity dendrogram based on the mutation data obtained. This is because tumour samples comprise a mixture of sub-clones and multi-regional trees built from bulk mutational profiles do not consider this heterogeneity. Thus, evolutionary inferences can be mistaken, timing of somatic mutations erroneously assigned leading to spurious parallel mutation events and/or incorrect chronological ordering of metastatic events. Importantly, this phylogeny versus dendrogram distinction in cancer has not been yet clarified at the time of the publication in chapter 2, so in that publication the analysis was mistakenly referred to as "Phylogenetic analysis", whereas "Dendrogram analysis" is the appropriate term (Jiménez-Sánchez et al., 2017).

1.3.2 ImmunoGenomics

Tumour purity

Tumour purity can be assessed using genomic and/or transcriptomic data. Using WES data genomic characterisation methods and mutation callers do not consider tumour purity and overall ploidy. Thus mutation callers provide useful information in terms of presence/absence of mutations and and estimate of VAF, but information regarding tumour cellularity requires the integration of CNA data too. ABSOLUTE is an algorithm that integrates mutation VAF data, relative CNA data, and common common tumour specific karyotypes to estimate tumour purity, absolute CNA, subclonal estimations, and ploidy of bulk tumour samples (Carter et al., 2012). In short, (1) ABSOLUTE jointly estimates tumour purity and ploidy directly from observed relative copy profiles and somatic mutation VAF if available. (2) A large and diverse sample collection of tumour specific karyotypes is used to resolve ambiguous cases.

(3) It accounts for subclonal copy-number alterations and somatic point mutations, which are common in heterogeneous tumours.

(Neo)epitope predictions

The binding affinity between HLA and epitopes depends on the position and physico-chemical properties of the anchor amino acids. Therefore, different approaches have been taken to predict if a peptide may bind or not bind to a particular HLA molecule. These approaches can be classified according to the data used to develop the prediction method: 1) *Modelling-based methods* use only sequence and structure data of peptides and HLAs (Schafroth and Floudas, 2004; Schiewe and Haworth, 2007; Tong et al., 2004), 2) *Binding pattern recognition models* trained on boolean data (binders versus non-binders) (Lauemøller et al., 2001; Mamitsuka, 1998; Nielsen et al., 2004; Reche et al., 2002), and 3) *Quantitative binding affinity models*, which are trained on HLA-peptide binding affinity datasets generated through biochemical assays (Andreatta and Nielsen, 2015; Nielsen et al., 2007; Zhang et al., 2007). Among the different methods and algorithms developed, the quantitative binding affinity model based on artificial neural network (ANN) called NetMHC has shown to be the most accurate and is a widely tool used in the community (Gulukota et al., 1997; Lin et al., 2008; Peters et al., 2006; Roomp et al., 2010; Yu et al., 2002).

ANNs are statistical supervised machine learning models that emulate the function of biological neurons to recognise linear and non-linear patterns in a set of data [reviewed in Manning et al. 2013]. The function of a neural network is to give an output after an input has been given, for example in the NetMHC algorithm, the inputs are a sequence of amino acids and the HLA subtype, and the output is a binding affinity value (Nielsen et al., 2003). This is possible because neural networks are "trained" using a set of data with known output (HLA-peptide binding affinity for NetMHC), and when the training data sets are sufficiently large, continuous data can be approximated. During the training data set processing, the neural networks adjust iteratively their error rate. Thus, with an accurate and large training set, neural networks can be used to predict outputs based on a new set of input data. However, because the learning process occurs as the networks adjust their output iteratively, the elements they recognise cannot be directly interpreted, thus neural networks operate as a black-box [reviewed in Manning et al. 2013].

For HLA-I epitope prediction, the NetMHC algorithm includes, besides the ANN algorithm, a combination of regression and hidden Markov models, sparse encoding, and Blosum encoding, which together have increased the accuracy of the predictions (Nielsen et al., 2003).

The original training set consisted of 528 HLA-peptide binding affinities from biochemical assays using 9 long peptides and 78 different HLAs allowing NetMHC to predict IC_{50} binding affinity values (Nielsen et al., 2003). However, because only 9mers were included, the prediction of HLA binding affinites of shorter or longer peptides was inaccurate, and further training using data sets from the Immune Epitope Database and Analysis Resource (IEDB) (Vita et al., 2015), the SYFPEITHI database (Rammensee et al., 1999), and new experimental data were included (Lundegaard et al., 2008a,b). To evaluate the accuracy of the predictions performed by NetMHC, 6,452 9mers covering 32 HLA alleles were used. 76% peptides were correctly predicted as binders using an $IC_{50} < 500$. The NetMHC algorithm for HLA-II epitope predictions is based on an artificial neural network-based alignment model trained on a large data set of more than 14,000 quantitative binding affinity values covering 14 HLA-DR alleles. The performance was evaluated on six independent datasets being comparable to the best publicly available HLA-II prediction methods (Nielsen and Lund, 2009). The principal caveats of the ANNs methods are: 1) the elements used for pattern recognition are usually uninterpretable, 2) reliable predictions are limited to the type of data used in the training set, which are mainly 9mers for NetMHC-I, and a subset of HLA subtypes, 3) peptide processing is not taken into account directly, 4) only primary peptide sequence information is used, and 5) machine learning algorithms always have an intrinsic error rate [reviewed in Manning et al. 2013].

Exome sequencing of tumours and identification of exonic mutations provides a source for neoepitope prediction and cancer-immune recognition analyses (Segal et al., 2008). A neoepitope prediction pipeline from sequencing data consists, briefly, of the following: 1) filtering and quality processing of sequencing reads, 2) detection and filtering of non-silent mutations, 3) mapping of mutated genes to their protein sequence, 4) *in silico* generation of mutant peptides, 5) typing of patient HLA alleles, and 6) prediction of HLA specific binding affinities (**Figure 1.4**).

Immunoediting estimation

The central paradigm of immunoediting is that it occurs during the evolution of the tumour, where negative selection of clones with mutations that bind HLA molecules could be expected. Thus, a lower number of neoepitopes than expected would suggest that an immunoediting process has occurred. A previous study by (Rooney et al., 2015) using exome sequencing and RNA sequencing of TCGA data have shown that, relative to other cancers, colorectal and kidney clear cancers have a higher neoepitope depletion in general. More recently, Mlecnik et al. applied the immunoediting model to colorectal cancer sequencing data and showed that

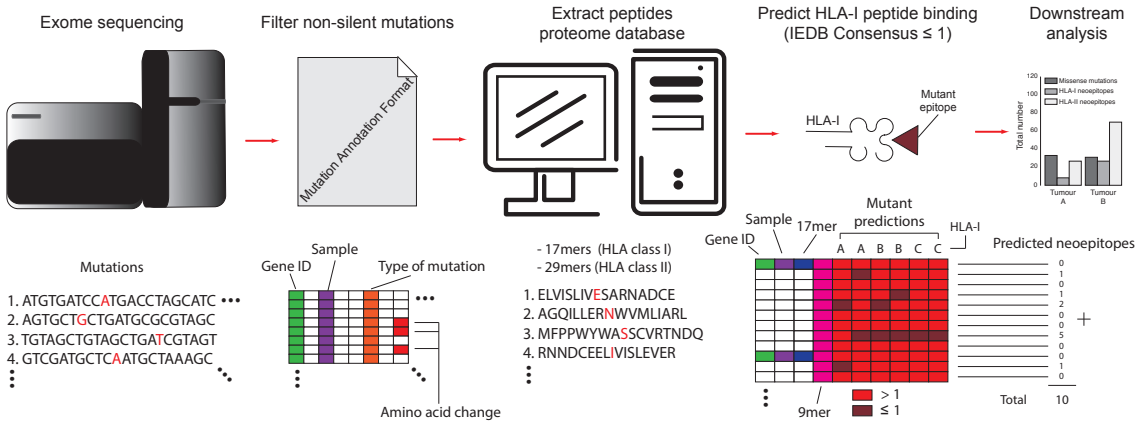


Fig. 1.4 Neoepitope prediction pipeline. Mutations detected in exome-sequencing data can be stored and organised in Mutation Annotation Format files (MAF). The missense mutations are filtered and peptide 17mers were generated using the Uniprot/Swiss-Prot human proteome as a reference. The middle amino acid (position 9^{th}) of the generated reference peptides is changed with the annotated mutated amino acid. HLA-I binding affinities are predicted for wild type (reference) and mutant peptides with NetMHC. Peptides with IC_{50} values below 500nM are considered potential epitopes (Dark blue and dark red). An epitope is counted if a peptide had an IC_{50} <500 for at least one of the six HLAs as shown in the example. Mutations are represented in red.

hypermutated samples had more tumour infiltrating T cell and neoepitope depletion. These analyses together with the observation that high mutation and predicted neoepitope loads correlate with higher patient survival rate in general (Brown et al., 2014) and better response to immunotherapy (Rizvi et al., 2014; Snyder et al., 2014) evidence that computational analyses of sequencing data and neoepitope prediction could be exploited to understand better how cancer-immune interactions occur in different tumour types and how they affect clinical outcome. However, this is still an emerging field of research and negative selection of cancer clones with specific mutations using sequencing data needs to be further verified.

Estimation and deconvolution of TME cell populations

Traditionally, cells from the TME in tumours have been quantified using immunohistochemistry (IHC), immunofluorescence (IF), and flow cytometry. However, these methods, although highly accurate, are expensive, labour intensive, and only provide cell quantifications. More recently, single cell RNA sequencing (scRNA-seq) technologies have been implemented to study the TME, offering the advantage of physically separating each individual cell and allowing the measurement of each individual cell transcriptome [reviewed in Gawad et al. 2016]. For instance, seminal studies have started to shed light on tumour and immune heterogeneity at single cell resolution (Amir et al., 2013; Azizi et al., 2018; Tirosh

et al., 2016). Also, integration of scRNA-seq with spatial localization through traditional immunohistochemistry has been done (Puram et al., 2017), and molecular mechanisms of response to immunotherapy have also been conducted using mass cytometry (Wei et al., 2017). However, the interplay and integration of ITH, TME heterogeneity, spatial localization, and their influence to response to therapy in tumour samples remains unexplored [reviewed in Yuan et al. 2017]. Moreover, the computational methods to compare dynamics of distinct cell populations at the single cell resolution still do not conform with the statistical properties of scRNA-seq data. Thus, scRNA-seq is still an emerging field and the analysis of several tumours or different tumour types is still prohibitive due to high costs. On the contrary, the great access and lower cost of bulk transcriptomics data has incited the possibility of estimating cellular populations of the TME using bioinformatics approaches. In brief, the principle behind the computational approaches to estimate TME cell populations is using a selected set of specific marker genes or expression signatures for each cell type to be estimated [reviewed in Finotello and Trajanoski 2018].

The most common approach for the analysis of marker genes is gene set enrichment analysis (GSEA), and its derivation single sample gene set enrichment analysis (ssGSEA). GSEA-based methods generate an enrichment score (ES) that is high when the genes specific for a cell type are amongst the top highly expressed in the sample of interest, and low otherwise. ssGSEA computes an ES representing the degree to which genes in a particular gene set are coordinately high- or low expression levels within a single sample compared to a random distribution of genes in the gene expression profile of the same sample (Barbie et al., 2009). With respect to the original GSEA, ssGSEA ranks the genes by their absolute expression in a sample and computes ES by integrating the differences between the empirical cumulative distribution functions of the gene ranks.

Unlike GSEA-based approaches that are semi-quantitative, deconvolution methods are quantitative. Deconvolution based methods estimate the unknown cell fractions using a signature matrix describing the cell-type-specific expression profiles. Deconvolution algorithms use a system of equations that describe the expression of each gene in a heterogeneous sample as a linear combination of the expression levels of that gene across the different cell subsets present in the sample, weighted by their relative cell fractions. However, whether the relationship between the expression levels of pure and heterogeneous samples is linear or not, is currently unknown. A crucial limitation of deconvolution methods is that cell types with higher amount of total RNA will contribute more to the cumulative expression of a heterogeneous sample and would be over-estimated as a result. Since different cell types have

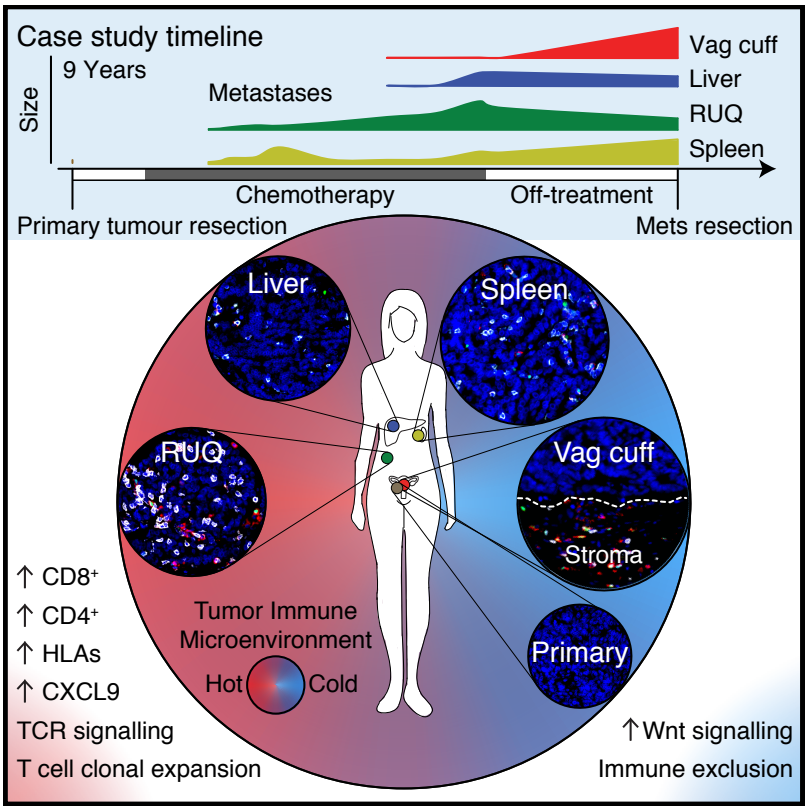
different sizes and different quantities of RNA, the final quantifications may be completely biased [reviewed in Finotello and Trajanoski 2018].

Chapter 2

Co-existence of tumour-immune microenvironments in an ovarian cancer patient

⊙ Denotes work performed by collaborators at MSKCC.
The work of this chapter has been published and has been edited for this thesis.

2.1 Graphical abstract



In brief

Distinct tumour immune microenvironments co-exist within a single individual and may help to explain the heterogeneous fates of metastatic lesions often observed post-therapy.

Highlights

- Differential progression of metastases during off-treatment period.
- Co-existence of distinct tumour-immune microenvironments within the same individual.
- Tumour regression and progression correlated with T cell infiltration and exclusion.
- Clonal neoepitopes elicited reactivity of circulating $CD8^+$ T cells.

Video abstract

To each metastasis its own immuno-environment:
https://www.youtube.com/watch?v=sIVhyPugwXA&t=214s

2.2 Summary

We present an exceptional case of a patient with high-grade serous ovarian cancer, treated with multiple chemotherapy regimens, who exhibited regression of some metastatic lesions with concomitant progression of other lesions during a treatment-free period. Using immunogenomic approaches, we found that progressing metastases were characterised by immune cell exclusion, whereas regressing and stable metastases were infiltrated by $CD8^+$ and $CD4^+$ T cells and exhibited oligoclonal expansion of specific T cell subsets. We also detected $CD8^+$ T cell reactivity against predicted neoepitopes after isolation of cells from a blood sample taken almost 3 years after the tumours were resected. These findings suggest that multiple distinct tumour immune microenvironments co-exist within a single individual and may explain in part the heterogeneous fates of metastatic lesions often observed in the clinic post-therapy.

2.3 Introduction

Several studies of smaller cohorts of patients with metastatic ovarian cancer have found that primary and metastatic lesions exhibit heterogeneity at the genomic level (Bashashati et al., 2013; De Mattos-Arruda et al., 2014; Lee et al., 2015). Supporting these findings, functional MRI-based analysis has revealed that ovarian tumours and metastatic peritoneal implants are already phenotypically heterogeneous at diagnosis (Sala et al., 2012). As tumour heterogeneity increases so does the likelihood of presence of subclones able to escape the immune system (Bhang et al., 2015; Su et al., 2012; Turke et al., 2010). Thus, immune control may be particularly challenging in ovarian cancer due to extensive heterogeneity and the low number of potential mutation-derived epitopes. The clinical challenge of tumour heterogeneity has been demonstrated recently in the context of immunotherapy: patients with less heterogeneous tumours, and hence with more clonal neoepitopes, were more likely to respond to checkpoint blockade immunotherapy than patients with heterogeneous tumours (McGranahan et al., 2016). In some settings, chemotherapy promotes immune cell homeostasis and activation (Carson et al., 2004; Gavalas et al., 2010; Pfirschke et al., 2016), tumour antigen release (Zitvogel et al., 2008), and decreased numbers of myeloid-derived suppressor cells in the tumour microenvironment (Suzuki et al., 2005). Furthermore, effector T cells have recently been implicated to play a role in abrogating fibroblast-mediated chemoresistance in a mouse model of ovarian cancer (Wang et al., 2016). Despite these findings, a unified model describing the effect of chemotherapy on the tumour heterogeneity and immune-tumour interactions has not yet been reached. A critical step toward understanding the effect of chemotherapy on advanced metastatic diseases and the immune response in humans is to analyse intra-patient matched primary and metastatic tumours (Brabletz et al., 2013).

Here we present a case study of a high-grade serous ovarian cancer patient whose different metastases exhibited concomitant regression and progression after treatment with multiple types of chemotherapy. We used whole-exome sequencing, RNA expression data, immunohistochemistry, neoepitope prediction, *in situ* T cell receptor sequencing of tumour-infiltrating immune cells, and T cell-neoepitope challenge assays with Intra-Cellular Staining (ICS) to investigate the genetic, molecular, and cellular components that potentially underlie this differential growth. In this heavily chemotherapy-treated patient, immune cell infiltration with clonal expansion of T cells, but not mutation or neoepitope number, correlated with tumour progression/regression status. Our immunogenomic analysis paints a portrait that immune infiltration and activation are different in each tumour at 2 years post-chemotherapy. Inter-site immune heterogeneity represents an important clinical challenge in the development of treatment modalities to overcome tumour heterogeneity.

2.4 Methods

Contact for reagent and resource sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by Dr Martin L. Miller (martin.miller@cruk.cam.ac.uk).

Experimental model and subject details

Human subjects research ⊙

Patient samples were collected and analysed after informed consent to the institutional tissue collection protocol, and approval by the Internal Review Board of Memorial Sloan Kettering Cancer Center (MSKCC). The biological sex of the patient is female (XX). The age of the patient at the time the primary sample was resected was 53 years old, and 60 years old at the time the metastatic samples were resected.

Method details

Whole exome sequencing ⊙

Whole exome sequencing was performed using the Illumina protocol at the Broad Institute of MIT and Harvard, Cambridge, MA, USA. Illumina sequencing of exomes was employed targeting approximately 37.7 Mb of mainly exonic territory made up of all targets from Broad Institute's Agilent exome design (Agilent SureSelect All Exon V2), all coding regions of Gencode V11 genes, and all coding regions of RefSeq gene and KnownGene tracks from the UCSC genome browser (http://genome.ucsc.edu). Data were analysed using the Broad Picard Pipeline which includes de-multiplexing and data aggregation.

The Illumina exome sequencing uses Illumina's in-solution DNA probe based hybrid selection method that uses similar principles as the Broad Institute-Agilent Technologies developed in-solution RNA probe based hybrid selection method (Fisher et al., 2011; Gnirke et al., 2009) to generate Illumina exome sequencing libraries. Pooled libraries were normalized to 2 nM and denatured using 0.2 N NaOH prior to sequencing. Flow cell cluster amplification and sequencing were performed according to the manufacturer's protocols using either the HiSeq 2000 v3 or HiSeq 2500. Each run was a 76 bp paired-end with a dual eight-base index barcode read. The sequencing depths of the samples were: normal blood sample (90% at 20 X), primary (82% at 50 X), spleen 78% at 50 X), right upper quadrant (RUQ) (60% at 50X), liver (89% at 50 X), and vaginal cuff (77% at 50 X) tumours.

Gene expression

RNA was extracted from formalin-fixed paraffin-embedded (FFPE) samples using the RecoverAll Total Nucleic Acid Isolation from Thermo Fisher Scientific (Catalogue Number: AM1975). RNA expression was assessed using the human Affymetrix Clariom D Pico assay. ⊙

Arrays were analysed using the SST-RMA algorithm in the Affymetrix Expression Console Software. Expression was determined by using the Affymetrix Transcriptome Analysis Console, and for genes displaying inconsistent expression between probes, the SRY gene signal was used as a cut-off. local regression (LOESS) normalisation across samples was implemented before differential expression analysis and ssGSEA (Tables C2S2A and C2S2C) using:

```
# R 3.4.0
library(affy) # version 1.54.0
data_norm<-normalize.loess(data,family.loess="gaussian")
```

Immunofluorescent staining ⊙

The immunofluorescent staining and cell counting were performed at Molecular Cytology Core Facility of Memorial Sloan Kettering Cancer Center using Discovery XT processor (Ventana Medical Systems) by a cytologist blinded to the sample identifiers and conditions. The tissue sections were deparaffinised with EZPrep buffer (Ventana Medical Systems), antigen retrieval was performed with CC1 buffer (Ventana Medical Systems). Sections were blocked for 30 min with Background Buster solution (Innovex) followed by avidin/biotin blocking for 8 min. Pseudocolors were applied as follows: CD4 A594, FOXP3 A488, CD8 A647; CD68 and CD3 A594 and PD-L1 A488. Cells were detected using the DAPI image, which was processed and segmented using ImageJ/FIJI (U.S. National Institutes of Health). Appropriate threshold values were set for all other markers, and the number of cells with positive signal above the threshold was counted for all single and double staining.

For multiplex staining, each marker was added consecutively in separate staining runs as follows. CD4/FoxP3/CD8: Sections were incubated with anti-CD4 (Ventana, Catalogue Number: 790-4423, 0.5 mg/ml) for 5 hr, followed by 60 min incubation with biotinylated goat anti-rabbit IgG (Vector Laboratories, Catalogue Number: PK6101) at 1:200 dilution. The detection was performed with Streptavidin-HRP D (part of DABMap kit, Ventana Medical Systems), followed by incubation with Tyramide Alexa 488 (Invitrogen, Catalogue Number:

T20922) prepared according to manufacturer instruction with predetermined dilutions. Next, slides were incubated with anti-FoxP3 (Abcam, Catalogue Number: ab20034, 5 mg/ml) for 4 hr, followed by 60 min incubation with biotinylated horse anti-mouse IgG (Vector Laboratories, Catalogue Number: MKB-22258) at 1:200 dilution. The detection was performed with Streptavidin-HRP D (part of DABMap kit, Ventana Medical Systems), followed by incubation with Tyramide Alexa Fluor 568 (Invitrogen, Catalogue Number: T20914) prepared according to manufacturer instruction with predetermined dilutions. Finally, sections were incubated with anti-CD8 (Ventana, Catalogue Number: 790-4460, 0.07 mg/ml) for 5 hr, followed by 60 min incubation with biotinylated goat anti-rabbit IgG (Vector, Catalogue Number: PK6101) at 1:200 dilution.

PDL1/CD68 or CD3: First, sections were incubated with anti-PDL1 (Cell Signaling, Catalogue Number: 13684, 5 mg/ml) for 5 hr, followed by 60 min incubation with biotinylated goat anti-rabbit IgG (Vector, Catalogue Number: PK6101) at 1:200 dilution. The detection was performed with Streptavidin-HRP D (part of DABMap kit, Ventana Medical Systems), followed by incubation with Tyramide Alexa 488 (Invitrogen, Catalogue Number: T20922) prepared according to manufacturer instruction with predetermined dilutions. Next, slides were incubated with anti-CD68 (DAKO, Catalogue Number: M0814, 0.02 mg/ml) for 5 hr, followed by 60 min incubation with biotinylated horse anti-mouse IgG (Vector Labs, Catalogue Number: MKB-22258) at 1:200 dilution, or with anti-CD3 (DAKO, Catalogue Number: A0452, 1.2 mg/ml) for 4 hr, followed by 60 min incubation with biotinylated horse anti-rabbit IgG (Vector Labs, Catalogue Number: PK6101) at 1:200 dilution. The detection was performed with Streptavidin-HRP D (part of DABMap kit, Ventana Medical Systems), followed by incubation with Tyramide Alexa Fluor 568 (Invitrogen, Catalogue Number: T20914) prepared according to manufacturer instruction with predetermined dilutions. After staining slides were counterstained with DAPI (Sigma Aldrich, Catalogue Number: D9542, 5 mg/ml) for 10 min and cover-slipped with Mowiol.

Sequenced-based HLA typing ⊙

HLA class I and class II 6-digit typing was performed at the New York Blood Center by sequence-based typing and specific sequence primers.

TCR sequencing ⊙

High-throughput sequencing of the T cell receptors present in the samples and blood of the patient was done using the immunoSEQ assay platform (Adaptive biotechnologies).

PBMC–neoepitope assay ⊙

The predicted peptides were synthesized (Genscript Corporation). peripheral blood mononuclear cell (PBMC) were cultured in complete RPMI (Core Media Preparation Facility MSKCC) with peptides at 1 mg/mL, peptide vehicle (DMSO, Sigma-Aldrich) and CEF peptide pool (2 mg/ml, C.T.L) for 21 days with peptide re-stimulation at day 7 and day 14. IL-2 (Proleukin, Chiron) and IL-15 Peprotech, Catalogue Number: 200-15) were added every 3 days at 10 IU/mL and 10 ng/mL respectively. Intracellular Cell Staining (ICS) was performed at day 14, and day 21 after 6 hr re-stimulation in the presence of monensin for 5 hr (GolgiStop, BD). Cells were then stained for 15 min with viability dye (LIVE/DEAD Fixable Aqua Dead Cell Stain Kit, ThermoFisher) at 4C followed by 30 min incubation with CD45-APC-H7 (BD PharMingen, clone 2D1), CD3-Pacific Blue (BD PharMingen, clone UCHT1), CD4-PerCP-Cy5.5 eBioscience, clone OKT4), CD8-PE (BD Biosciences, clone SK1). Cells were then fixed and permeabilised with BD Cytofix/Cytoperm (BD Biosciences) for 20 min at 4C and washed with BD Perm/Wash (BD Biosciences). The ICS was performed in BD Perm/Wash with IFN-g-FITC (eBioscience, clone GZ-4) and TNF-a-PE-Cy (eBioscience, clone MAb11) at 4C for 30 min. Samples were acquired on a BD LSRII flow cytometer (BD Biosciences) and the analysis was performed on FlowJo software (FlowJo, LLC).

Quantification and statistical analysis

Tumour volume calculation

The two axes of the CT scan measurements and the equation for the ellipsoid volume were used to estimate tumour volumes:

$$V = \frac{4}{3}\pi(a)(b)(c) \tag{2.1}$$

Where a and b are the two axes, and c is their mean.

Mutation calling

Reads with mapping quality below 30 in the BAM files were filtered out before mutation calling. SNV were called using MuTect version 1.4 (Cibulskis et al., 2013). Identified missense mutations were manually reviewed using the Integrative Genomics Viewer version 2.3.61 (Robinson et al., 2011; Thorvaldsdóttir et al., 2013).

Phylogenetic tree inference

The phylogenetic tree was generated as described in Murugaesu et al.. A binary presence/absence matrix of all non-silent mutations was used as input for the R package phangorn version 2.0.2 (Schliep, 2011). UPGMA hierarchical clustering followed by the parsimony ratchet analysis (Nixon, 1999) were implemented to build the unrooted tree, and the `acctran` function was used to determine branch lengths.

Relative copy-number alterations

To extract copy number information based on the sequenced exomes of the samples, CopywriteR version 2.2.0 (Kuilman et al., 2015) was employed in R version 3.2.3. To perform the analysis, mappability information based on the hg19 human reference genome, 20 kb bin size, and default parameters were used.

Absolute copy-number alterations and tumour purity

The absolute copy number profiles and the tumour content of the samples were inferred using the computational method ABSOLUTE version 1.0.6 (Carter et al., 2012) in R version 3.2.3. ABSOLUTE integrates segmented copy number data, pre-computed statistical models of recurrent cancer karyotypes, allelic fractions of somatic SNVs, and a probabilistic model framework to jointly estimate candidate tumour purity, ploidy values, absolute copy number data, and subclonal single nucleotide variants (Carter et al., 2012). Tumour purity and absolute copy numbers were obtained using ABSOLUTE default parameters, segmented copy number data derived from CopywriteR, and variant allele frequencies estimated by MuTect (Cibulskis et al., 2013). Best model selection was based on the guidelines provided by GenePattern and the Broad Cancer Genome Analysis group (http://www.broadinstitute.org/cancer/software/genepattern/analyzing-absolute-data). Amplifications and deep deletions were defined as copy-number alterations with at least ± 2 median absolute deviations for each sample copy-number distribution as shown in Figure S1C.

Mutation cellular prevalence

Variant allelic cellular prevalence was estimated using PyClone version 0.13.0 (Roth et al., 2014) in Python version 2.7.11. The PyClone pipeline analysis was performed jointly on all samples with their tumour purity and absolute copy number alterations estimated by ABSOLUTE. Total copy number prior probability estimate and the PyClone binomial model

were used in the analysis. The mutation variant allele frequencies, closest integer copy number alterations, and tumour purity were used as input. Mutations not present or called in the sample were set to 0. Agglomerative hierarchical cluster analysis with Euclidean distance metric and average linkage clustering was performed on the cellular prevalence values and samples. The *SREBF2*^{S120*} nonsense mutation was not included in the PyClone pipeline because its copy number data was closest to 0.

Single-sample gene set enrichment analysis

Single-sample gene set enrichment analysis (Barbie et al., 2009), a modification of standard gene set enrichment analysis (Subramanian et al., 2005), was performed on RNA measurements for each sample using the GSVA package version 1.24.1 (Hänzelmann et al., 2013) in R version 3.3.2 with parameters: `method="ssgsea"`, and `tau=0.25`. Normalised enrichment scores were generated for gene sets belonging to KEGG (Kanehisa and Goto, 2000; Kanehisa et al., 2015) and Reactome (Fabregat et al., 2016). The gene sets were obtained from MSigDB database version 5.2 (Liberzon et al., 2011). In order to identify significantly up- and down-regulated gene sets, a *p*-value was calculated for each gene set based on comparison of the enrichment score with 10,000 permutations of randomly sampled gene sets of the same size. All genes listed in the expression array were used to derive the permutated gene sets. Finally, the *p*-values were corrected using Benjamini and Hochberg method (BH). Enrichment scores were normalized across samples (Tables C2S2D and C2S2F) using:

```
# R 3.4.0
data_norm<-as.data.frame(scale(data,center=T,scale=T))
```

Immune cell gene-expression signatures

Tumour purity and total immune component in the tumour samples were analysed using the ESTIMATE algorithm method version 1.0.13 (Yoshihara et al., 2013) on the gene expression data using the option: `platform="affymetrix"` in R version 3.4.0. Then, selection of probes with the highest variance for each gene was performed to deconvolute cell type specific immune signatures. The deconvolution was achieved using CIBERSORT Jar version 1.05 (https://cibersort.stanford.edu/) with the standard LM22 signature gene file, and 1000 permutations to calculate deconvolution *p*-values (Newman et al., 2015).

Whole-exome sequencing-based HLA inference

The HLA genotyping algorithms OptiType version 1.0 (Szolek et al., 2014) and POLYSOLVER version 1.0 (Shukla et al., 2015) with default parameters were employed for HLA

class I 4-digit inference. POLYSOLVER HLA-I typing and mutation calling were performed using samtools version 0.1.19 and novocraft 3.02.05 for the alignment, and MuTect version 1.1.7 for the variant calling.

Neoepitope predictions

In silico mutant peptide generation: To predict neoepitopes, "wild-type" petide 17mers (for HLA-I) and 29mers (for HLA-II) with the affected amino acid in the middle for each missense mutation were retrieved from the GRCh37.74 human reference proteome (http://ftp.ensembl.org/pub/release-74/fasta/homo_sapiens/pep/). To generate "mutant" peptides, the affected amino acid was replaced *in silico* with the corresponding mutant amino acid.

HLA class I epitope binding predictions: Mutant peptides were used as input for the T Cell Epitope Prediction Tools included in the IEDB 3.0 (http://www.iedb.org/) (Vita et al., 2015). The HLA class I epitope binding predictions were performed using the HLA-I IEDB algorithms Consensus (Kim et al., 2012) and the artificial neural network (NetMHC) version 3.4 (Lundegaard et al., 2008b; Nielsen et al., 2003) independently, yielding same conclusions. For Consensus method – which combines NetMHC, the stabilized matrix method (Peters and Sette, 2005), and the combinatorial peptide libraries method (Sidney et al., 2008) – 9 mers with a relative percentile rank $\leq 1\%$ for each HLA-I allele were considered binders to cover most of the potential immune responses as previously suggested (Kotturi et al., 2007; Moutaftsi et al., 2006). For NetMHC, different cut-off values were evaluated independently and compared between each other. 9mers with absolute IC50 affinity values $\leq$ HLA-I specific cutoffs were considered binders (http://help.iedb.org/entries/23854373) (Paul et al., 2013). HLA-I specific cut-offs were not available for HLA-I C alleles, therefore an IC50 $\leq$ 500 nM was used instead. All mutant predicted binders were considered for the analyses, i.e., for each missense mutation, up to six binders for HLA-I (A, B, C alleles) and up to four binders for HLA-II (DQ and DR alleles). Since NetMHC gives actual nM binding affinities, and HLA-I specific cutoffs have been estimated, we used NetMHC predictions.

HLA class II epitope binding predictions: HLA class II epitope binding predictions on 15mers were performed using the HLA-II IEDB algorithms Consensus (Wang et al., 2008, 2010), NetMHCII version 2.2 (Nielsen and Lund, 2009), and Sturniolo (Sturniolo et al., 1999) since these were the only available methods for the patient HLA-II alleles. The Consensus method used the relative percentile ranks of NetMHCII and Sturniolo, and 15mers with percentile ranks $\leq 1\%$ were considered binders. 15 mers with NetMHCII IC50 $\leq$ 500 nM or

Sturniolo percentile rank $\leq 1\%$ were considered binders, which are more stringent cut-off values than the IEDB recommended 1000 nM for NetMHCII and $\leq 10\%$ percentile rank for Sturniolo. In the authors' knowledge, HLA-II specific NetMHCII cut-offs have not been reported.

Neoepitope depletion analysis

TCGA ovarian cancer null model: To analyse neoepitope depletion across the different samples, we followed the method developed by Rooney and colleagues using only expressed mutations. Commonly mutated genes were not included as indicated (Rooney et al., 2015). The method compares the samples to a data driven null model. To generate the null model and estimate neoepitope depletion, the nucleotide sequences flanking each mutation (context of the mutation) are taken into account, thus 192 possible codon mutations are considered: $(64 \text{ codons})(3 \text{ possible mutations}) = 192$ possible changes. To control for tumour type differences, we used TCGA ovarian cancer samples to generate the null model (TCGA, 2011). Context of the mutations for the TCGA ovarian cancer samples and the case study tumour samples were obtained from the assembly of the Genome Reference Consortium Human Reference 37. Only TCGA ovarian cancer samples with mutation context in all missense and silent mutations were included (n = 150). We predicted HLA-I neoepitopes of TCGA ovarian cancer samples using the same approach as for the case study samples described above.

Neoepitope depletion for each sample was calculated as follows. First, the expected number of missense mutations per silent mutation (Ns) and the expected number of predicted neoepitopes per missense mutation (Bs) were calculated using all samples (TCGA ovarian cancer samples and the patient's samples), where Ns and Bs are vectors with 192 components each:

$$Ns = \left(\frac{missense_{AAA\rightarrow ACA}}{silent_{AAA\rightarrow ACA}}, \frac{missense_{AAA\rightarrow AGA}}{silent_{AAA\rightarrow AGA}}, \ldots, \frac{missense_{TTT\rightarrow TGT}}{silent_{TTT\rightarrow TGT}}\right) \tag{2.2}$$

$$Bs = \left(\frac{neoepitope_{AAA\rightarrow ACA}}{missense_{AAA\rightarrow ACA}}, \frac{neoepitope_{AAA\rightarrow AGA}}{missense_{AAA\rightarrow AGA}}, \ldots, \frac{neoepitope_{TTT\rightarrow TGT}}{missense_{TTT\rightarrow TGT}}\right) \tag{2.3}$$

Therefore, each component of the vector Ns corresponds to the fraction of missense mutations per silent mutation, and each component of the vector Bs corresponds to the number of predicted neoepitopes per missense mutation. In both vectors, Ns and Bs, each component corresponds to the ratio of a particular codon change. The components of Ns can be computed because the counts of the mutations take into account the three possible reading frames.

Second, the count of silent mutations for each codon change (*Ss*) was calculated for each sample. Thus, *Ss* is a vector with 192 components where each component is the number of silent mutations with a particular type of codon change for a given sample:

$$Ss = (silent_{AAA \to ACA}, silent_{AAA \to AGA}, \ldots, silent_{TTT \to TGT}) \tag{2.4}$$

Third, the expected number of missense mutations (*Npred*) and the expected number of neoepitopes (*Bpred*) were calculated for each sample.

$$Npred = \sum_{m}^{Silent} Ss(m) \times Ns(m) \tag{2.5}$$

$$Bpred = \sum_{m}^{Silent} Ss(m) \times Ns(m) \times Bs(m) \tag{2.6}$$

$$m \in (AAA \to ACA, AAA \to AGA, \ldots, TTT \to TGT)$$

Where *Silent* represents the number of silent mutations.

Fourth, the expected and observed numbers of neoepitopes per missense mutation for each sample were calculated as follows:

$$Expected = \frac{Bpred}{Npred} \tag{2.7}$$

$$Observed = \frac{Bobs}{Nobs} \tag{2.8}$$

Where *Nobs* is the observed number of missense mutations and *Bobs* the predicted number of neoepitopes for each sample.

Finally, for each sample these ratios were compared and the ratio observed versus expected neoepitopes calculated:

$$R = \frac{Observed}{Expected} \tag{2.9}$$

Each sample has a ratio R, thus a distribution of $\log_2$ ratios is generated as shown in Figure S4C top panel. Empirical two-sided p-value thresholds were calculated because the ratios do not follow a normal distribution (Shapiro-Wilk test, D'Agostino-Pearson's test, and

Kolmogorov-Smirnov test) The calculations can be found in **Tables C2S4E-C2S4G**.

Permutation null model: To compare the levels of neoepitope depletion only between the patient's samples, we generated sample specific null models based on 150 random unique permutations (redundant permutations excluded) of the samples and their mutations (**Table C2S4H**). The number of permutations was selected based on the number of samples used in the TCGA ovarian cancer neoepitope depletion analysis (n = 150). Permutated and real samples were analysed together using the same approach as for the TCGA ovarian cancer neoepitope depletion analysis described above. A permutation-based null model for each sample was used to control for the number of mutations. Empirical two-sided p-value thresholds were calculated for each distribution because the ratios do not follow a normal distribution (Shapiro-Wilk test, D'Agostino-Pearson's test, and Kolmogorov-Smirnov test). The calculations can be found in **Figure 7.4** and **Table C2S4I**.

Immunogenicity predictions

Immunogenic properties of HLA class I epitopes were estimated in silico using the IEDB resource tool "MHC I Immunogenicity" (http://tools.iedb.org/immunogenicity/), which combines the chemical and physical properties of the amino acids, their position in the epitope, and the HLA-I subtype allele to estimate the immunogenicity of a given neoepitope-HLA complex (Calis et al., 2013). To compare clonal and sub-clonal predicted immunogenic properties, we used two approaches:

Absolute score comparison: Two-sided Mann-Whitney rank tests were calculated to compare absolute scores between clonal and sub-clonal predicted binders and non-binders. In **figures 7.5A-C**, n refers to the number of peptides in each category. The Mann-Whitney rank test was employed because the absolute score distributions do not follow a normal distribution (Shapiro-Wilk test, D'Agostino-Pearson's test, and Kolmogorov-Smirnov test), and because the number of peptides in each category is different a Wilcoxon signed-rank test could not be calculated.

Binomial immunogenicity comparison: Generalised Linear Model (GLM) were used to compare the probability of a peptide having immunogenic properties or not according to it clonal status and HLA binding affinity. The binomial GLM approach was considered appropriate for this setting because immunogenicity can be considered a binomial process, immunogenic or non-immunogenic. In this scenario, however, the binomial process corresponds to whether an epitope has biochemical properties associated with immunogenicity

(score ≥ 0) that outweigh properties associated with no immunogenicity (score < 0). Importantly, predicted immunogenicity scores < 0 can still elicit an immunogenic response, but overall they have less immunogenic properties than positive scores (Calis et al., 2013). To further explain variation in the intrinsic immunogenic predictions we included HLA-I binding affinity (nM) as an explanatory variable. We then calculated the probability of a peptide having a positive immunogenic score or not based on the samples' neoepitope data. No interaction between clonality and HLA-I binding affinity was found, thus the interaction was excluded from the model. The final binomial GLM formula used is:

```
# R 3.4.0
m<-glm(data=data,immunogenicity~clonality+hla_binding,family=binomial)
```

TCR sequencing analysis

Analysis of the sequences was performed on the immunoSEQ ANALYZER 3.0 (Adaptive biotechnologies). T cell rearrangements that are differentially abundant between samples were detected using the Differential Abundance tool by two-sided binomial tests with Benjamini and Hochberg multiple test correction, q-value < 0.01 was considered statistically significant.

Data and Software Availability

Requests for additional data and custom code should be directed to Dr Martin L. Miller (martin.miller@cruk.cam.ac.uk).

Whole-exome sequencing data

The accession numbers for the whole exome sequences reported in this chapter are BioSample: `SAMN06199513`, `SAMN06199514`, `SAMN06199515`, `SAMN06199516`, `SAMN06199517`, and `SAMN06199518`.

Microarray data

The microarray data discussed in this study have been deposited in NCBI's Gene Expression Omnibus (Edgar, 2002), and the accession number is GEO: `GSE92780`.

TCR sequencing data

The TCR sequencing data discussed will be provided upon request Dr Martin L. Miller.

Supplementary tables

Supplementary tables can be accessed through:
https://github.com/cansysbio/immunogenomics/tree/master/AJSPhDThesis/AppendixA/Tables/

2.5 Results

2.5.1 Case report

The patient presented here was diagnosed with stage IV HGSOC, which typically exhibits a 5 year survival of 17% (National Cancer Institute, SEER Data Base), and underwent an optimal surgical debulking followed by paclitaxel combined with first cisplatin and then carboplatin. The patient experienced recurrence after 7 months, and during a period of 3 years she was treated with multiple regimens of chemotherapy with progression of disease after each therapy (**Figures 2.1A** and **2.1B**). Her cancer was growing radio-graphically, and her CA125 was rising during treatment with topotecan when she then transitioned to best supportive care and was followed clinically with regular CA125 biomarker evaluation.

After chemotherapy treatment was stopped, she experienced an atypical course: her CA125 decreased, and after 2 years of clinical follow up, CT scans showed evidence of differential growth of metastatic lesions including a new complex cystic mass in the vaginal cuff. Because of her long treatment-free interval and abdominal discomfort, she opted to undergo another debulking procedure, which found a substantial disease burden including tumour implants on the liver capsule, the splenic hilum, right upper quadrant (RUQ), and rectovaginal space (**Figures 2.1A** and **2.1B**). Samples of the primary and four metastatic tumours were submitted for whole-exome sequencing, microarray RNA quantification, staining for protein markers by immunofluorescence, and *in situ* T cell receptor sequencing.

2.5.2 Phylogenetic analysis of somatic alterations in tumours

We performed whole-exome sequencing of normal blood and the resected samples to identify somatic mutations in the primary tumour and the metastases. Of all samples, we detected the highest mutation load in the liver and vaginal cuff metastases (**Figure 2.1C**). To infer the evolutionary relationship between the tumour samples, we used a binary presence/absence matrix of the non-silent mutations to perform a phylogenetic reconstruction based on the parsimony ratchet analysis method with branch lengths proportional to the number of non-silent mutations (Nixon, 1999; Schliep, 2011) (**Figures 2.1D** and **8.1A**). The liver and vaginal cuff tumours were genetically more heterogeneous and harboured more mutations.

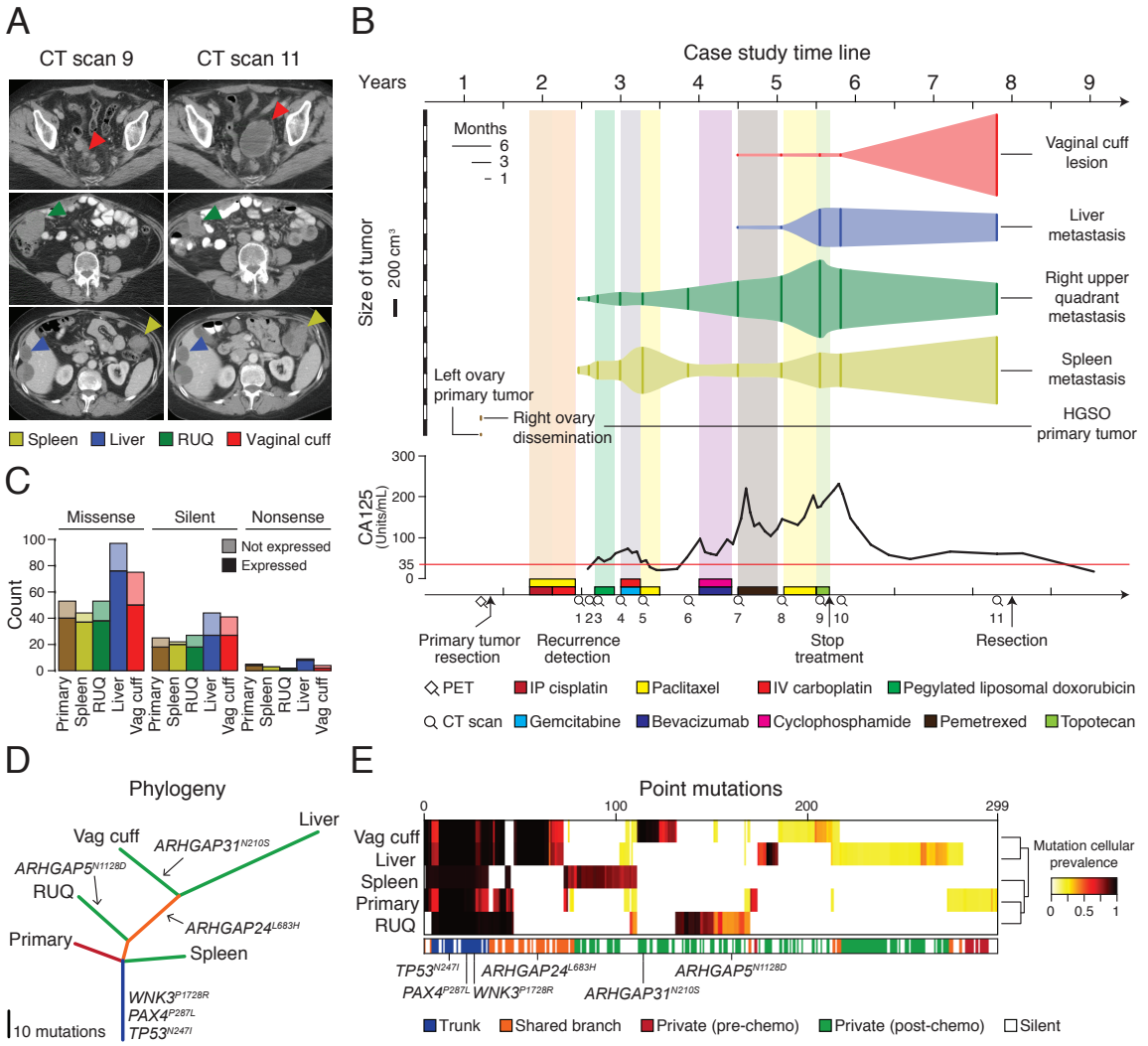


Fig. 2.1 Metastatic tumours exhibit heterogeneous growth and somatic mutation patterns after multi-line chemotherapy. (A) Representative CT scans showing concomitant progression/regression of the different resected metastatic tumours. RUQ = right upper quadrant. "Spleen" refers to the tumour deposit adjacent to the spleen. (B) CT-based volume of metastatic lesions represented with the solid vertical lines and dynamics of quantified CA125 levels with the red line indicating the CA125 upper limit of normal (35 units/ml). The x axis at the bottom shows a time-line of therapeutic interventions and clinical follow up. (C) Number of missense, silent, and nonsense mutations. (D) The phylogenetic tree represents the relationship of the samples based on binary calls of non-silent point mutations. Length of the branches is proportional to the number of mutations. Potential driver mutations are indicated. (E) Hierarchical cluster analysis (Euclidean distance metric and "average" linkage method) of the cellular prevalence of point mutations (n=299) estimated with PyClone (Roth et al., 2014).

To estimate the proportion of cancer cells identified with a given mutation (cellular prevalence), we applied PyClone (Roth et al., 2014) using CopywriteR-inferred (Kuilman et al.,

2015) DNA copy-number changes (**Figures 7.1B** and **7.1C**) and ABSOLUTE-inferred (Carter et al., 2012) tumour purity and absolute copy numbers. As expected, truncal and shared mutations were generally clonal with high cellular prevalence, whereas private mutations had medium to low cellular prevalence indicating sub-clonal status (**Figure 8.1E**). Focusing on the specific genes that were mutated across all samples, we found among the truncal mutations potential oncogenic driver alterations, including $WNK3^{P1728R}$, $PAX4^{P287L}$, and $TP53^{N247I}$ (**Figure 7.1D**). $TP53^{N247I}$ was detected with a high cellular prevalence indicating loss of heterozygosity, which was supported by our DNA copy-number analysis. Additionally, we identified other putative truncal events, including deletion of *BRCA1*, *BRCA2*, and *PTEN* and amplification of *CCNE1* (**Figure 7.1B**), which are commonly altered in serous ovarian cancer (Bowtell et al., 2015; Patch et al., 2015). Among the private mutations we detected several potential driver mutations including $RUNX3^{P246S}$ in the growing splenic lesion and $CSMD1^{G1770R}$ in the primary tumour. Several private and shared branch mutations were found in different Rho GTPase-activating genes (*ARHGAP*), which inactivate Rho and Rac signalling involved in the control of cellular motility (Bernards and Settleman, 2004; Li et al., 2014).

2.5.3 Immune-related pathways are over-expressed in regressing tumours

To evaluate whether genes involved in chemotherapy resistance were differentially altered between tumours and associated with regression and progression status, we analysed somatic alterations and gene-expression data (Affymetrix transcript array) across the samples. After analysing chemotherapy-resistance genes identified in HGSOC (Patch et al., 2015), as well as gene sets for multi-drug resistance (ABC transporters), apoptosis, and DNA-damage response, we found no clear evidence of gene-expression or somatic-alteration patterns (mutations, DNA amplification, and deep deletion) that differed between progressing (primary, vaginal cuff, and spleen) and regressing/stable tumours (RUQ and liver) (**Figures 7.2A–C**). Interestingly, there was a trend that the ABC transporter *TAP1*, which is known for its function as a transporter of cytosolic peptides to the endoplasmic reticulum for HLA class I presentation (Bahram et al., 1991; Neefjes et al., 1993; Powis et al., 1991; Suh et al., 1994), was expressed at a higher level in the regressing tumours (**Figure 2.2A**).

To further identify potential differences between samples, we analysed gene sets and pathways in an unbiased manner with ssGSEA (Barbie et al., 2009; Subramanian et al., 2005). Using permutation-based false-discovery rate, we estimated the significance of the enrich-

ment score for each pathway and performed an outlier analysis relating gene-set significance to the relative change in enrichment score between a given sample and the rest of the samples (**Figure 2.2B, Tables C2S2D-F**). The most significant and differentially enriched pathway found was the immune system pathway with a higher enrichment in the spleen and RUQ metastases and a lower enrichment in the primary and vaginal cuff tumours (**Table C2S2H**). Further indicating immune activation, the systemic lupus erythematosus pathway was highly enriched in the RUQ and liver metastases (**Table C2S2I**), whereas TCR signalling pathways were preferentially enriched in the RUQ sample alone (**Table C2S2J**). Cancer and proliferation pathways, as well as Wnt signalling, were more enriched in the primary and vaginal cuff tumours (**Table C2S2K**). No outlier gene sets were identified for the negatively enriched pathways (**Figure 7.2E** and **Tables C2S2D-F**).

To investigate the gene-expression differences between the samples on an unbiased individual gene level, we calculated the coefficient of variation of the expression levels for each gene across samples (**Table C2S2A**). We found that among the most variably expressed genes, besides lipid metabolic process related genes in the liver, the T cell chemo-attractant *CXCL9* was predominantly expressed in the RUQ and liver metastases, as well as STAT1, which has been implicated in the regulation of *CXCL9* expression (Liao et al., 1995; Satoh and Tabunoki, 2013) (**Figure 7.2D**). No relevant mutations in immune-related molecules were identified except for truncal mutations in the MHC class I polypeptide-related sequence B (MICB) (**Table C2S1A**), which is a stress-induced ligand recognized by NKG2D receptors on CD8$\alpha\beta$ and $\gamma\delta$ T cells, as well as NK cells (Bauer et al., 1999; Groh et al., 1999).

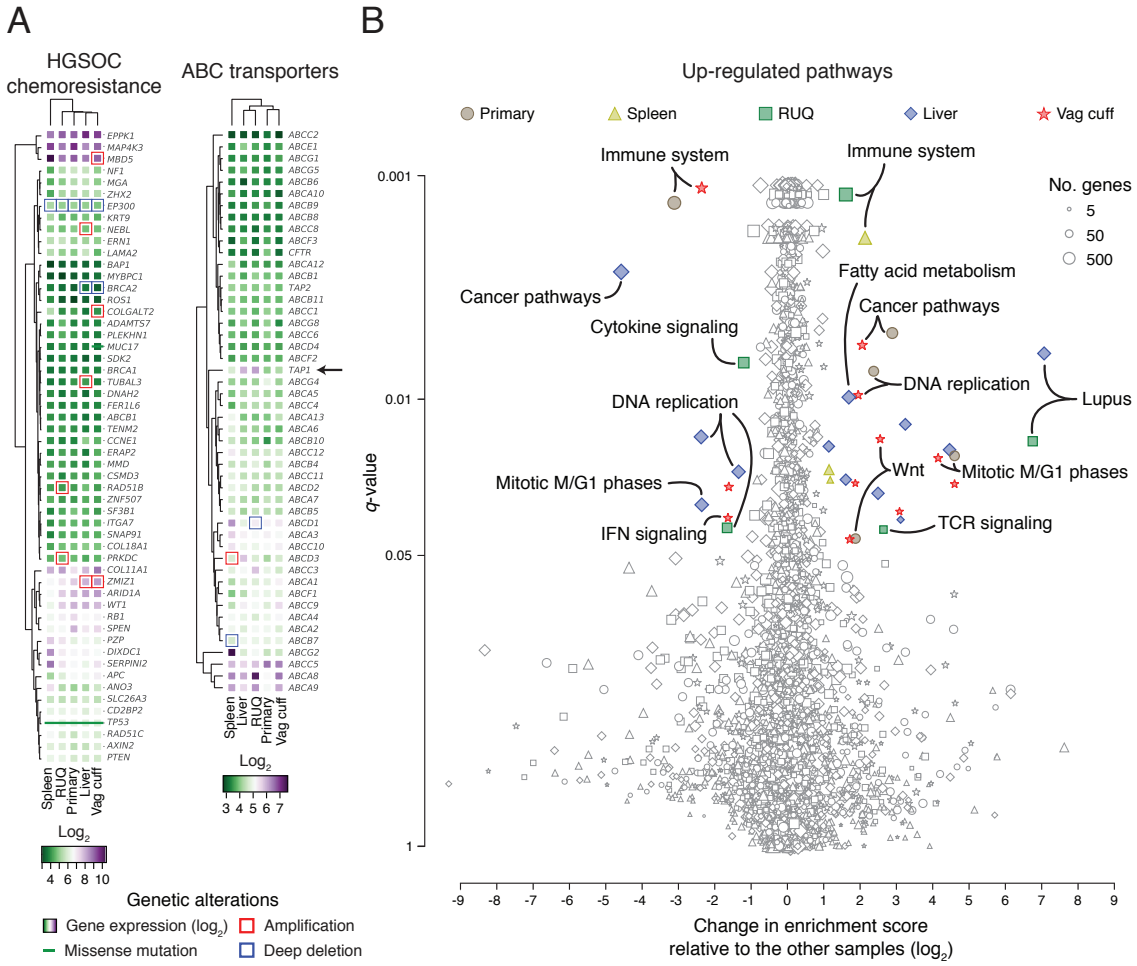


Fig. 2.2 Differential expression of immune-related pathways in heterogeneously growing tumours. (A) Expression levels and genetic alterations of genes associated with chemotherapy resistance in HGSOC (Patch et al., 2015) and multi-drug resistance. Amplification and deep deletion were defined as at least ± 2 median absolute deviations of copy-number alterations for each sample (**Figure 7.1**). (B) Single-sample gene set enrichment analysis (Barbie et al., 2009; Subramanian et al., 2005) of up-regulated pathways using the KEGG (Kanehisa and Goto, 2000; Kanehisa et al., 2015) and REACTOME (Fabregat et al., 2016) databases. Significantly enriched pathways ($q < 0.05$) with at least ± 1 log_2 change relative to the median of the other samples are coloured. FDR adjusted p-value (q-value) was calculated using the BH method.

2.5.4 Heterogeneous immune infiltration in growing and regressing lesions

To investigate the immune infiltration status of the tumours, we used ESTIMATE to analyse tumour purity and overall stromal and immune components (Yoshihara et al., 2013). The lowest tumour purities and highest immune infiltration scores were found in the RUQ, liver,

and spleen samples (**Figure 2.3A**). Furthermore, we deconvoluted the gene-expression data using CIBERSORT (Newman et al., 2015) as a first approach to dissect infiltration of specific immune cell subsets in the tumours. We found that the largest immune cell components corresponded to $CD8^+$and $CD4^+$ T cells in RUQ, liver, and spleen tumours, although the overall CIBERSORT deconvolution *p*-value was only significant for RUQ and liver tumours (**Figure 2.3B**). In contrast, the primary and vaginal cuff tumours had low immune cell ESTIMATE scores and insufficient levels of immune cell transcripts to confidently apply CIBERSORT (**Tables C2S3A-B**), together suggesting a low or absent immune component present in these tumours.

Following this analysis, samples were immuno-fluorescently co-stained for T cell markers CD4, CD8, and the T regulatory cell marker FOXP3, double stained for PD-L1 and macrophage marker CD68, as well as double stained for PD-L1 and the T cell marker CD3 (**Figures 2.3C and 7.3A**). Consistent with the transcriptomic deconvolution analyses, the primary tumour demonstrated no T cell infiltration and was negative for PD-L1 and CD68 (**Figures 2.3C, 2.3D**, and **7.3A**, and **Table C2S3C**). The vaginal cuff lesion, which was growing at the time of surgical resection, did display a T cell population; however, these cells bordered but did not infiltrate the tumour. The splenic lesion, which was also progressing at the time of resection, albeit at a much more modest rate than the vaginal cuff lesion, demonstrated a $CD8^+$ infiltrate. Finally, the RUQ and liver metastases, which were regressing and stable, respectively, at the time of surgical resection, displayed a strong $CD4^+$ and $CD8^+$ infiltrate. In summary, the transcript and IF analyses suggested that each tumour site displayed a unique tumour-immune microenvironment ranging from immune cell inclusion to exclusion.

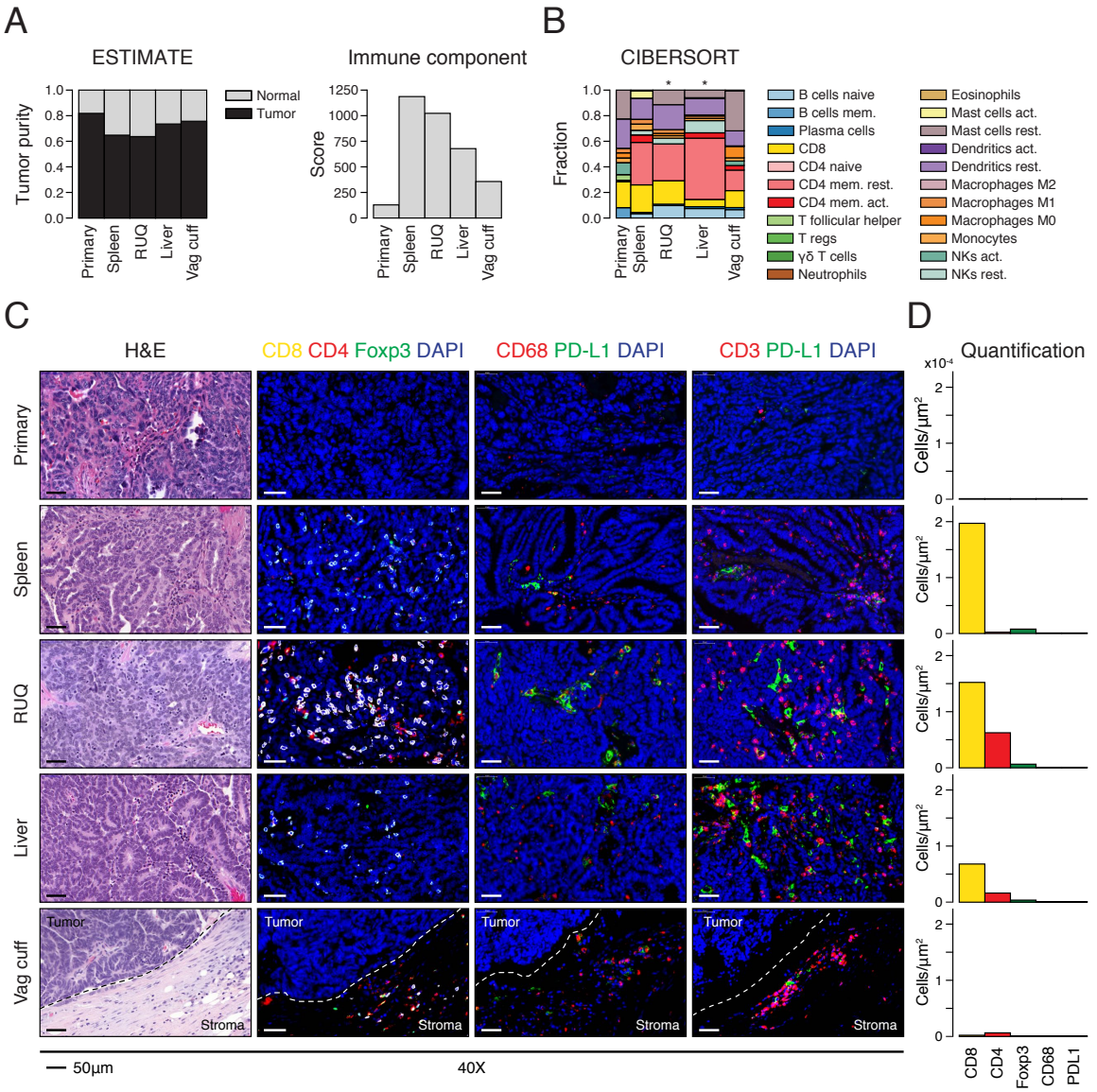


Fig. 2.3 Immune infiltration status shows heterogeneous microenvironments across tumour samples. (A) Tumour purity and immune component estimated by analysing Affymetrix-based transcriptomics (Yoshihara et al., 2013). (B) Fractions of immune cell subsets in tumour samples inferred from gene-expression data using CIBERSORT (Newman et al., 2015). Width of bars is proportional to the $-log_{10}$ p-value of the deconvolution. CIBERSORT empirical p-value, $* \; p < 0.05$. (C) Representative images of H&E staining of tumour samples and IF staining for DAPI, cytotoxic T cells ($CD8^+$), helper T cells ($CD4^+FOXP3^-$), T cells ($CD3^+$), Tregs ($CD4^+FOXP3^+$), macrophages ($CD68^+$), and immune-checkpoint PD-L1. Complete slides are shown in **Figure 7.3**. (D) Image-based cell quantification of whole slides.

2.5.5 T cell oligoclonal expansion detected in regressing metastases

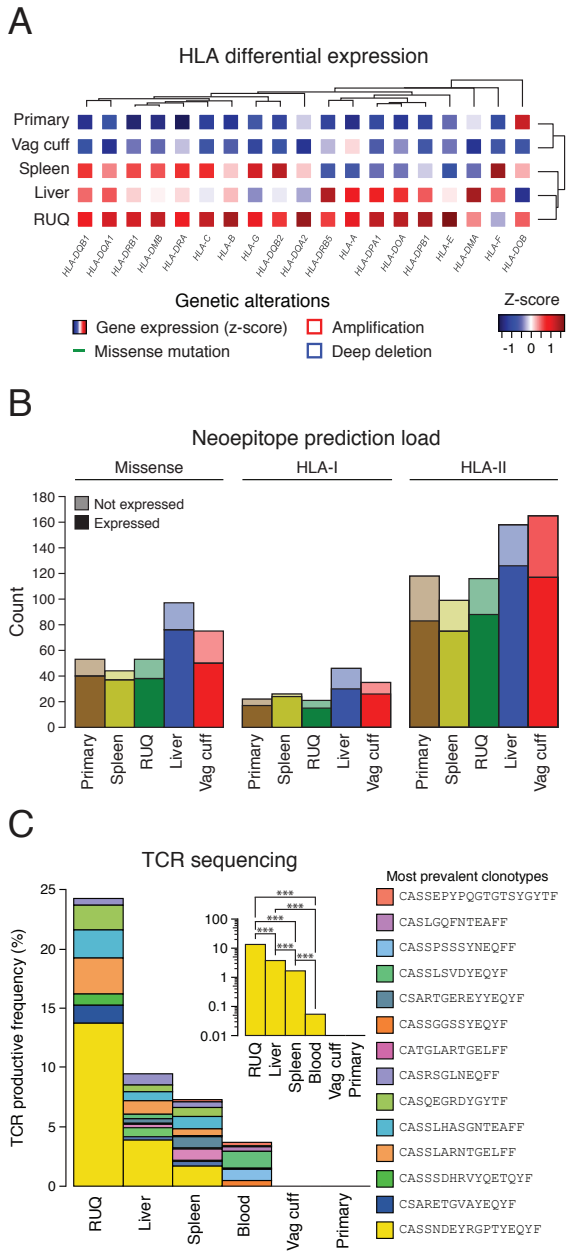


Fig. 2.4 Higher HLA expression and T cell oligoclonal expansion detected in regressing tumours. (A) HLA-I and II gene differential expression across samples. (B) Number of predicted neoepitopes per sample. (C) TCR sequencing of FFPE tumour samples and blood. The most prevalent TCR clonotypes (top 5 for each sample and blood) are shown. The blood sample was collected from the patient 550 days after secondary debulking (**Figure 7.5A**). Inset shows detection of the most frequent TCR rearrangement (`CASSNDEYRGPTYEQYF`) and its abundance comparison between samples (two-sided binomial tests with BH multiple test correction, q-value < 0.001 ***).

It is known that genetic alterations in HLA-I molecules are associated with escape of cancer cells from $CD8^+$ T cell recognition (Shukla et al., 2015). The patient's HLA alleles were determined experimentally by conventional polymerase chain reaction (PCR)-based HLA typing and computationally on exome data by OptiType (Szolek et al., 2014) and POLYSOLVER (Shukla et al., 2015) independently, yielding the same results (**Table C2S4A**). We searched copy-number alterations as well as mutations by applying POLYSOLVER, a specific computational pipeline for HLA-I typing and mutation detection in the HLA-I genes; however, no genetic alterations were detected. We then assessed gene expression and found that all HLA-I genes were expressed in the tumours (Tables C2S2A-C); however, compared to primary and vaginal cuff samples, an overall higher expression of HLA genes was observed in the RUQ and liver samples, with a lesser extent seen in the spleen sample (**Figure 2.4A**). We next estimated the neoepitope landscape of the samples by mapping missense mutations to their amino acid sequences, *in silico* generating the mutant peptide sequences, and predicting the mutant peptide-HLA binding affinities to the patient's HLAs. The predictions were performed using the NetMHC algorithm with HLA specific cut-offs for HLA-I (Lundegaard et al., 2008b; Nielsen et al., 2003; Paul et al., 2013) and consensus scores for HLA-II (Kim et al., 2012; Kreiter et al., 2015).

The tumours with the highest mutation and neoepitope loads for both HLA class I and HLA class II were the liver and vaginal cuff, which also had the highest number of missense mutations (**Figure 2.4B**). We also investigated whether there were shared neoepitopes or mutations present in the RUQ (regressing) and liver (stable) metastases alone, i.e., not present in the other tumours. No shared mutations between RUQ and liver alone were detected (**Figure 7.4A**); therefore, it did not appear that a shared neoepitope or mutation alone explained the behaviour of the non-progressing tumours sites.

As an active CD8 T cell infiltration can exert a selective pressure at the neoepitope level (DuPage et al., 2012; Matsushita et al., 2012; Teng et al., 2015; Tran et al., 2016; Verdegaal et al., 2016), we further interrogated the neoepitope landscape by analysing potential evidence of neoepitope depletion using an approach adopted from a report analysing TCGA data (Rooney et al., 2015). Relative to the other samples from the patient, the regressing RUQ tumour showed a consistent yet non-significant—tendency of neoepitope depletion (p-value < 0.1 by two-sided empirical p-value; **Figures 7.4B** and **7.4C**). This result is in line with a recent report showing neoepitope depletion in tumours with higher levels of immune signatures in colorectal cancer (Davoli et al., 2017). We then predicted the intrinsic immunogenicity of neoepitopes by analysing the biochemical properties of peptides that are predicted to be associated with T cell epitope recognition. Positive immunogenicity scores have biochemical properties associated with higher immunogenicity that outweigh properties associated with lower immunogenicity, and vice versa for negative scores (Calis et al., 2013). We observed that there was a significant effect of neoepitope clonality on the probability of a neoepitope having immunogenic properties, with clonal neoepitopes showing a lower probability of having immunogenic properties than sub-clonal predicted binders (p-value = 0.02 by chi-square test; **Figures 2.5A–D**). Using the predicted non-binders instead of binders in a control analysis (NetMHC score > HLA-I specific cut-off), the opposite trend was observed, clonal neoepitopes show a higher probability of having immunogenic properties, as there was a small but significant effect of clonal mutations being predicted as more immunogenic (p-value=0.003 by chi-square test; **Figure 2.5F**), as well as peptides with higher HLA-I affinities (chi-square test, p-value=0.0001), although the absolute differences are minor. No significant interaction between clonality and predicted HLA-I binding affinity was detected for either binders or non-binders. Although preliminary, these analyses indicate a potential negative selection process at the neoepitope level.

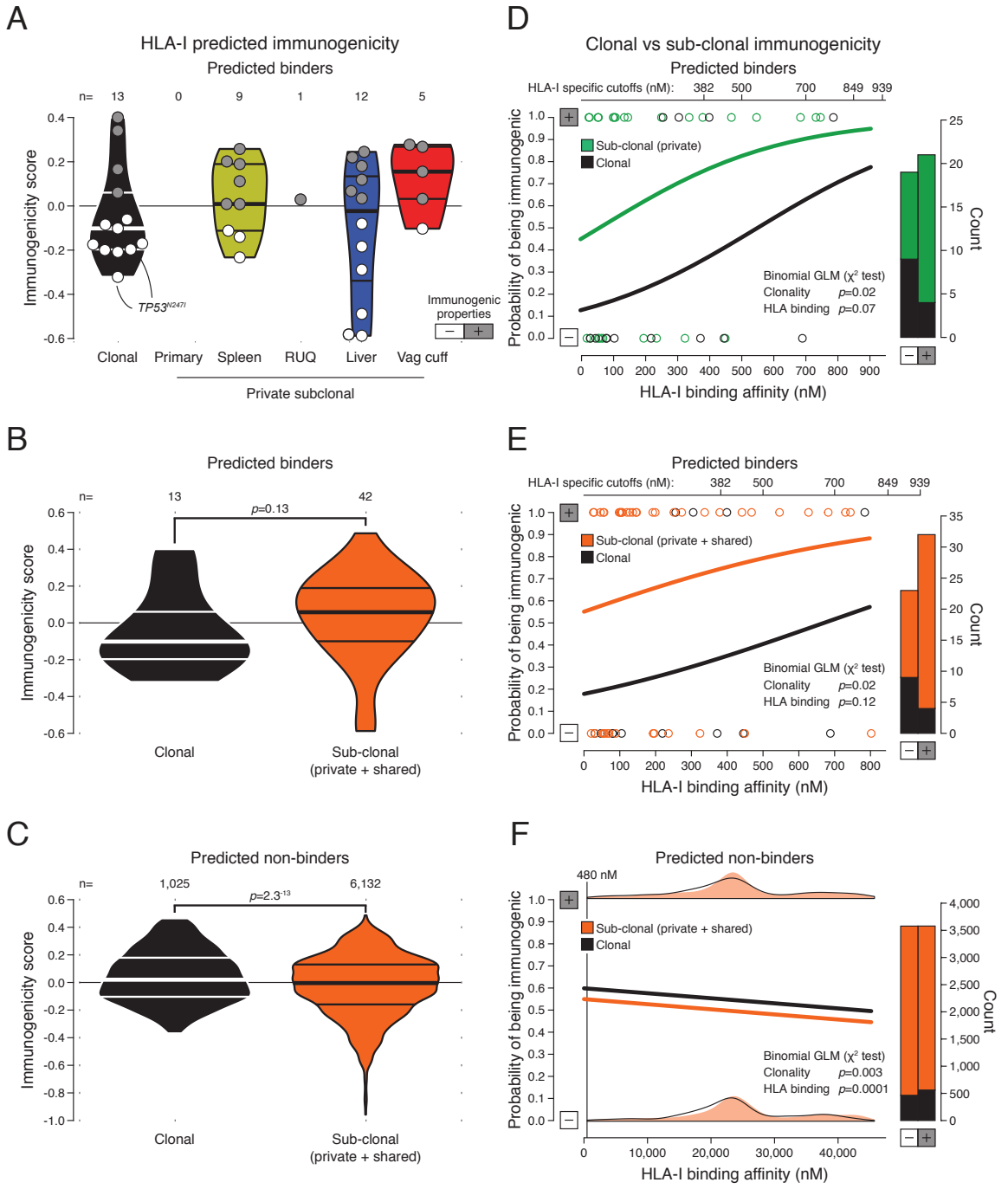


Fig. 2.5 Predicted immunogenicity of HLA class I neoepitopes. (A) Predicted immunogenic properties of trunk (clonal) and private HLA-I neoepitopes. Horizontal lines within violin plots show the median and interquartile range of the data distribution. (B and C) Comparison between clonal and sub-clonal (including shared between two or more samples but not all) predicted immunogenicity of predicted binders and non-binders (two-sided Mann-Whitney rank test). Horizontal lines within violin plots show the median and interquartile range of the data distribution. (D–F) Probability of an HLA-I neoepitope having immunogenic properties considering its clonality and HLA-I binding affinity using the neoepitope data in (A), (B), and (C), respectively. GLM = generalized linear model.

To evaluate a T cell response in the tumours, we investigated whether T cell clonal expansion could be detected in the tumour samples. To this end, we performed *in situ* TCR sequencing on each sample and on peripheral blood from the patient sampled 550 days after tumour resection (**Figures 2.4C** and **7.5A**, and **Table C2S5A**). We detected a T cell expansion in the RUQ metastasis with a dominant clone accounting for 13% of all productive T cell receptors sequenced. The expanded clone was also detected in the liver and spleen metastases and strikingly also in the blood of the patient. However, the clonal frequency in the RUQ metastasis was significantly higher than that in the other samples (q-value < 0.001 by two-sided binomial tests with BH correction). In contrast, no T cell receptors were detected in the primary and vaginal cuff tumours, further supporting their lack of T cell infiltrate.

2.5.6 Peripheral blood CD8+ T cells react against predicted neoepitopes

Since expanded T cell clones detected in the tumours were still detected in the patient's blood sampled 1 year 6 months (550 days) after resection, we decided to test whether circulating T cells could react against any of the predicted neoepitopes. We sampled blood from the patient again, this time at 2 years 8 months (978 days) after resection, and isolated PBMCs (**Figure 7.5A**). We performed an ICS assay lasting 21 days, where PBMCs were cultured with each of the mutant peptides (n = 43) predicted to have at least one HLA-I neoepitope, as a mutant peptide (17-mer) can have more than one predicted binder (9-mer) (**Figure 7.5B**). Importantly, the likelihood of observing T cell reactivity by the ICS assays is low due to the low frequency of T cell precursors in the blood and the limited representation of the total TCR repertoire in each peptide challenge experiment (5×10^5 cells per well) (**?**). Despite the high false-negative rate generally observed with the ICS assay, we found $CD8^+$ T cells reactive against several mutant peptides showing cytokine activation levels (IFN-γ and TNF-α) similar to the positive control consisting of a mixture of viral-derived epitopes (**Figures 2.6A** and **2.6B** and **Table C2S5A**). Of the top five reactive peptides detected, all had a higher mutant to wild-type predicted HLA-I binding affinity (inset, **Figure 2.6B**). With limited material available, we focused on the top hits and repeated the ICS experiment and again found reactivity with peptide 12, which was derived from a clonal mutation in $FLG2^{E1608K}$ (**Table C2S1A**), and peptide 6, which was derived from a private mutation in $LRRC8E^{C629Y}$ in the splenic tumour.

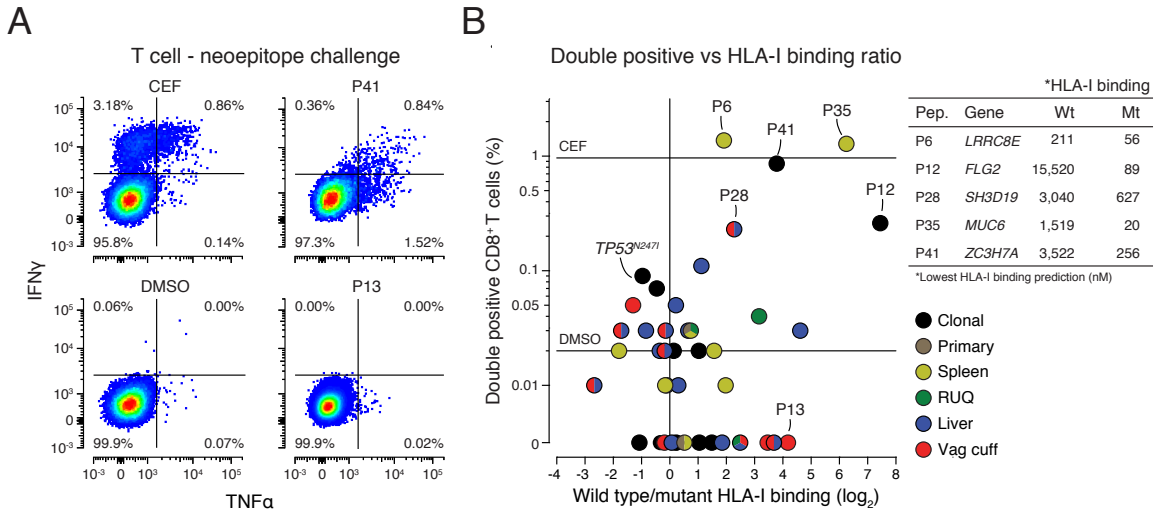


Fig. 2.6 Predicted neoepitopes with higher mutant than wild-type HLA-I binding affinity elicit a T cell response. (A) Representative scatter-plots of TNF-α and IFN-γ intracellular cytokine staining of CD8$^+$ T cells after 21 days of culture with CEF peptides or DMSO as positive and negative controls or the predicted mutant peptides (**Figure 7.5B**). CEF = Cytomegalovirus, Epstein-Barr virus, Influenza virus. (B) Percentage of CD8$^+$ T cells with double-positive intracellular staining (TNF-α and IFN-γ) after incubation with each of the 43 predicted HLA-I neoepitopes, and HLA-I predicted binding affinity wild-type to mutant ratio. Mutation in gene *FLG2E1608K* (P12) was found to be clonal after manual inspection in IGV.

2.6 Discussion

The natural history of ovarian cancer typically features remissions of decreasing length, leading to premature death (Bowtell et al., 2015). In this unusual case, the divergent fates of the tumours show an overall association with multiple molecular and cellular features at the tumour-immune interface (**Figure 2.7**). For example, the shrinking RUQ tumour was heavily infiltrated with $CD4^+$ and $CD8^+$ T cells and had evidence of active $CD8^+$ T cell surveillance with expansion of specific TCR clonotypes. The stable liver tumour also exhibited immune infiltration, but at a lower level and with fewer expanding T cell clones. The spleen tumour was growing modestly at the time of resection and presented with intermediate tumour-immune microenvironment features. Finally, the growing vaginal-cuff and the primary tumour exhibited complete immune cell exclusion. The TCR clone most prevalent in the non-progressing tumours was also detected in the blood of the patient 18 months after the metastases were resected, and clonal neoepitopes induced a $CD8^+$ T cell response from PBMCs obtained nearly 3 years after surgery. The two most extreme tumours, the RUQ and the vaginal cuff, had a consistently divergent pattern of molecular features associated with immune activation (*HLA* expression, *IFN-γ*, *CXCL9*, *TAP1*, etc.) and immune suppression (Wnt signalling), respectively. Importantly, the observed features that relate to progression/regression status are correlative and do not *per se* prove any *bona fide* mechanism nor negate the fact that chemotherapy could have influenced the divergent fates. In sum, we find evidence of distinct tumour-immune microenvironments among differentially growing metastases within the same individual.

Beyond the co-existence of tumour microenvironments across different metastases in a patient with advanced HGSOC, different questions remained unsolved, namely:

- Does intra-patient tumour-immune heterogeneity occur at baseline (i.e. before treatment intervention)?
- What are the molecular and cellular mechanisms behind the intra-patient TME heterogeneity observed?
- Which cells induce or impede immune cell infiltration in a treatment-naive setting?
- What is the effect of chemotherapy in the TME?
- What are the clinical implications of TME heterogeneity?
- Can immune-exclusion be reverted therapeutically?

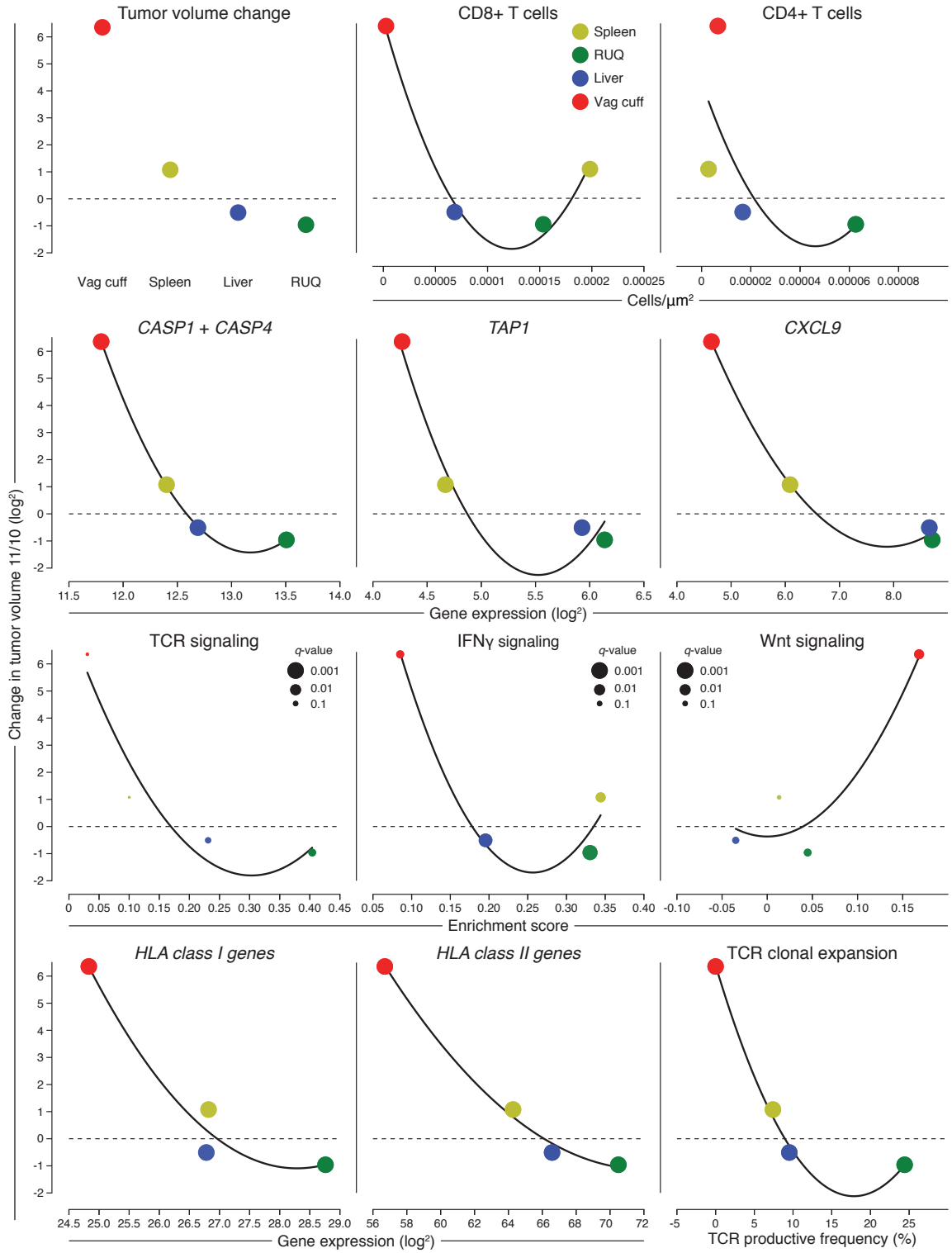


Fig. 2.7 Overall associations between tumour fates and tumour-immune microenvironmental features. Cellular and molecular associations with change in tumour growth. Change in tumour growth (y axis) was calculated by dividing the tumour volume at CT scan 11 by the tumour volume at CT scan 10 (**Figure 2.1B**). Fitted curves are 2nd order polynomial regression lines plotted for trend visualization rather than prediction purposes. *Caspase 1* and *4* are considered inflammatory caspases involved in a type of apoptosis related to immune response called pyroptosis. The enrichment score x axis and the *q*-values come from the ssGSEA analysis. HLA-I genes include *HLA-A*, *B*, *C*, *E*, and *F*. HLA-II genes include *HLA-DPA1*, *DMA*, *DRA*, *DQA1*, *DMB*, *DPB1*, *DQB2*, *DRB5*, *DRB1*, *DQB1*, and *DOA*.

2.7 Contributions

CRediT standard taxonomy: http://dictionary.casrai.org/Contributor_Roles

Conceptualisation

Ideas; formulation or evolution of overarching research goals and aims
AS, AJS, MLM, CA, JDW, ES, JK, TM, OZ.

Data curation

Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where necessary for interpreting the data itself) for initial use and later re-use
AJS.

Formal analysis

Application of statistical, mathematical, computational, or other formal techniques to analyse or synthesize study data
AJS, DM, HAV, HV.

Funding acquisition

Acquisition of the financial support for the project leading to this publication
AS, MLM.

Investigation

Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection
YL, KJP, JL, ICS, TM, SP, JR, DZ.

Methodology

Development or design of methodology; creation of models
AJS, MLM.

Project administration

Management and coordination responsibility for the research activity planning and execution
AS, MLM.

Resources

Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools
AS, MLM.

Software

Programming, software development; designing computer programs, implementation of the computer code and supporting algorithms; testing of existing code components
AJS, DM.

Supervision

Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team
AS, MLM.

Validation

Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs
AJS, SP, TM.

Visualisation

Preparation, creation and/or presentation of the published work, specifically visualization/data presentation
AJS.

Writing original manuscript draft

Preparation, creation and/or presentation of the published work, specifically writing the initial draft
AJS, AS, MLM.

Writing review & editing

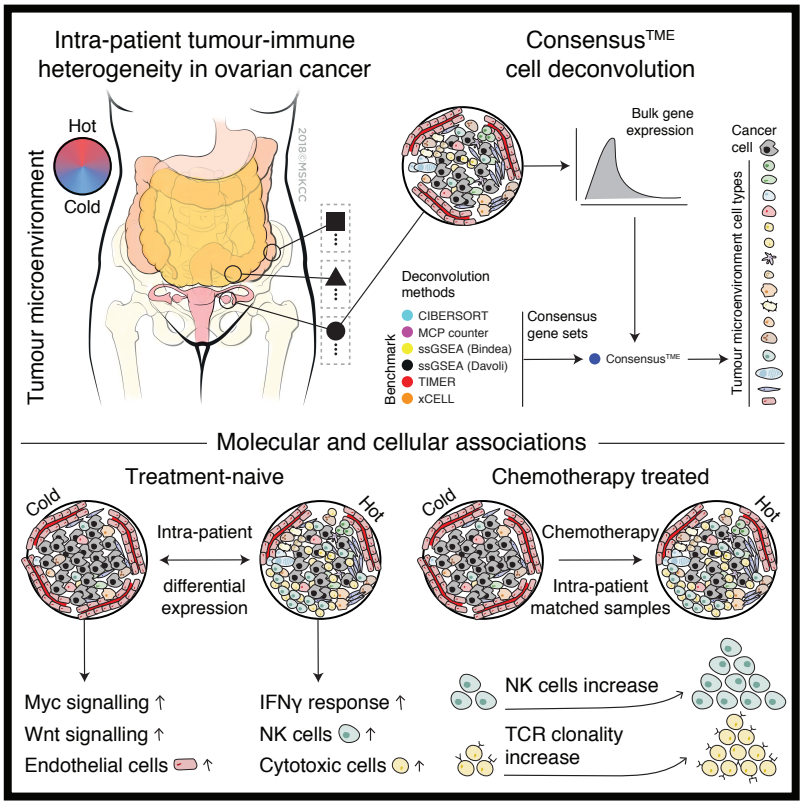
Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision
AJS, MLM, JR, MBG, DZ, AS, HV.

Chapter 3

Unravelling tumour-immune heterogeneity in advanced ovarian cancer uncovers immunogenic effect of chemotherapy

⊙ Denotes work performed by collaborators at MSKCC.

3.1 Graphical abstract



In brief

Intra-patient tumour-immune microenvironment heterogeneity is highly prevalent in treatment-naive high-grade serous ovarian cancer, and chemotherapy induces an immunogenic response in these patients.

Highlights

- Intra-patient tumour-immune heterogeneity is highly prevalent in patients with high-grade serous ovarian cancer.
- Consensus gene set signatures improve tumour microenvironment cell estimation.
- Myc and Wnt signalling pathways are enriched in immune excluded tumours.
- Neoadjuvant chemotherapy induces NK cell infiltration and T cell activation.

3.2 Summary

In metastatic cancer, the role of heterogeneity at the tumour-immune microenvironment, its molecular underpinnings and clinical relevance remain largely unexplored. To understand tumour-immune dynamics at baseline and upon chemotherapy treatment, we performed unbiased pathway and cell type-specific immunogenomics analysis of treatment-naive (38 samples from 8 patients) and paired chemotherapy treated (80 paired samples from 40 patients) high-grade serous ovarian cancer (HGSOC) samples. Whole transcriptome analysis and image-based quantification of T cells from treatment-naive tumours revealed ubiquitous variability in immune signalling and distinct immune microenvironments co-existing within the same individuals and within tumour deposits at diagnosis. To systematically explore cell type composition of the tumour microenvironment using bulk mRNA, we derived consensus immune and stromal cell gene signatures by intersecting state-of-the-art estimation/regression/enrichment methods (Consensus^{TME} v1.0), providing improved accuracy and sensitivity when compared to HGSOC immunostaining and leukocyte methylation data sets. Cell-type estimation and pathway analyses revealed that Myc and Wnt signalling associate with immune cell exclusion in untreated HGSOC. To evaluate the effect of chemotherapy on the intrinsic tumour-immune heterogeneity, we compared site-matched and site-unmatched tumours before and after neoadjuvant chemotherapy. Transcriptomic and T-cell receptor sequencing analyses showed that site-matched samples had increased cytotoxic immune activation and oligoclonal expansion of T cells after chemotherapy, which was not seen in site-unmatched samples where heterogeneity could not be accounted for. These results demonstrate that the tumour-immune interface in advanced HGSOC is intrinsically heterogeneous, and thus requires site-specific analysis to reliably unmask the impact of therapy on the tumour-immune microenvironment.

3.3 Introduction

It is unclear how the complex interplay between tumour cells and the TME and their interactions affect treatment outcome in metastatic cancer (Janssen et al., 2017; Kitamura et al., 2015; Robinson et al., 2017). Investigating this interplay in an advanced disease setting is complicated by the difficulty of obtaining multiple-site tumour samples and the finding that different tumours within the same individual can harbour distinct immune microenvironments (Jiménez-Sánchez et al., 2017; Reuben et al., 2017; Sridharan et al., 2016; Zhang et al., 2018). Moreover, interactions between different cell populations of the TME are plastic and can change dependent on extrinsic perturbations such as therapy (Wang et al., 2016).

HGSOC is ideally suited to the study of TME heterogeneity owing to its clinical presentation with multi-site abdominal disease and standardised treatment with either optimal surgical debulking at diagnosis or delayed primary surgery after NACT (Bowtell et al., 2015). Thus, in HGSOC there is a unique opportunity to study the characteristics of the TME at multiple sites, to observe variation at baseline (diagnosis), and following therapy. Furthermore, since HGSOC is typically diagnosed when dissemination has already taken place in the peritoneal cavity, this malignancy provides the basis for evaluating the ubiquitousness of intra-patient TME heterogeneity in an advanced, metastatic disease setting.

To characterise heterogeneity at the level of the TME and to begin to identify the molecular and cellular underpinnings of immune infiltration variability at diagnosis and after perturbation by chemotherapy in HGSOC, we performed a systematic analysis of >100 HGSOC samples from treatment-naive and NACT patient cohorts. Our findings confirm that HGSOC is a disease characterised by highly prevalent TME heterogeneity with distinct immune microenvironments co-existing in different tumour nests within the same individuals at diagnosis. We leverage our rich data sets to create an ensemble computational approach that integrates and improves upon existing immune and stromal cell cell-type specific estimation methods, thus enabling us to systematically characterise the TME of HGSOC before and after treatment. We identify oncogenic signalling pathways such as Myc and Wnt that associate with immune cell exclusion when comparing tumours with high cancer cell fraction (high purity) *vs* low cancer cell fraction (low purity). We find that NACT induces immune activation and specific T cell clonal expansions in local TMEs, however, intra-patient TME immune heterogeneity can mask such effects. Consequently, systemic immunomodulatory therapies may be ineffective in a subset of tumour sites, thus preventing overall patient benefit. Together, these results show that intra-patient TME heterogeneity is highly prevalent in HGSOC, which could confound clinical outcomes.

3.4 Methods

Contact for reagent and resource sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by Dr Martin L. Miller (martin.miller@cruk.cam.ac.uk).

Experimental model and subject details

Human subjects research ⊙

All patients had stage IIIC or IV HGSOC as assessed by a pathologist specialized in gynaecological malignancies. Patients signed written consent to Institutional Review Board approved bio-specimen protocols.

Treatment-naive cohort ⊙

For the treatment-naive cohort, 25 patients were consented to the study between August 2014 and March 2016. Out of these patients, 17 were excluded as they either a) withdrew from the study (n=3); b) the final pathology was not HGSOC (n=5); c) the patients had disease progression upon review of study imaging and underwent neoadjuvant chemotherapy instead of primary cytoreductive surgery (n=5); d) inadequate image-guided tissue sampling due to friable tissue (n=2); e) research imaging studies not performed due to patient cancellation (n=2). The final treatment-naive study population consisted of 8 patients with histopathologically-confirmed diagnosis of HGSOC (**Table C3S1A-B**). Each patient underwent multi-parametric MRI of the abdomen and pelvis and 18F-FDG PET/CT within 7 days immediately preceding the primary cytoreductive surgery. Volumetric regions of interest were outlined on both axial T2-weighted MR images and PET images, covering both the primary and metastatic lesions, using ImageJ/FIJI (U.S. National Institutes of Health) by a board certified radiologist with special expertise in ovarian cancer imaging. The tumour regions outlined on MRI were co-registered with those outlined on PET.

Neoadjuvant chemotherapy cohort ⊙

For the neoadjuvant chemotherapy cohort a previously established institutional database identified 152 patients with HGSOC who underwent neoadjuvant chemotherapy between 2008 and 2013. Of these patients, 149 went on to interval debulking surgery. 48 of these patients had adequate pre and post treatment formalin fixed paraffin embedded tissue samples

available. All pretreatment specimens were obtained either through core biopsy or laparoscopic biopsy, and all post treatment specimens were obtained at the time of laparotomy for interval debulking surgery. Choice of chemotherapy was at the clinician's discretion, but all patients in the cohort received a platinum and taxane based regimen (**Table C3S5A**). 40 of these paired samples yielded data for analysis, 17 of these pairs were site-matched, meaning that pre-treatment and post-treatment specimens were taken from the same anatomic site, while 23 were site-unmatched. Gene expression and T cell receptor sequencing (TCR-seq) data were generated for 28 and 37 pairs respectively (**Table C3S5B**). Samples with very low TCR sequences (n=5 samples, 10 pairs) were not included in the downstream analyses as the confidence of TCR clonality is low.

Method details

Image acquisition and analysis ⊙

The quantitative diffusion parameters D (diffusion coefficient) and f (volume fraction of blood flowing through micro-vessels) derived from the intravoxel incoherent motion MRI (Le Bihan et al., 1988) and dynamic contrast-enhanced MRI (Tofts, 1997) parameter K^{trans} (volume transfer constant between the blood plasma and the extravascular extracellular space) were generated voxel-wise using a dedicated in-house software written in Matlab (Mathworks Inc., Natick, MA, USA). The standarised uptake value (SUV) of the voxels contained within each lesion on PET were also calculated (Kinahan et al., 2009). k-means clustering algorithm (Carano et al., 2004) of the D, f, K^{trans} and SUV voxels, with the number of clusters (k) being fixed to k=3 was used to identify imaging clusters/habitats. The mean and standard deviation (mean±std. dev.) of these parameters for each cluster were calculated. To establish coherence across patients (i.e. to label each cluster with the α, β, γ greek letters, such that across patients clusters would have similar imaging features), the intra-cluster distance was calculated for each cluster. The greek letter of the clusters for each patient was assigned based on the relative value of the intra-cluster distance. Specifically, for each patient, β was assigned to the cluster which had the highest intra-cluster distance for that patient; γ was assigned to the cluster with intermediate intra-cluster distance, and α to the cluster with the lowest intra-cluster distance.

Custom made 3D tumour Moulds ⊙

For each patient, custom made 3 dimensional (3D) moulds [REF:Weigelt B, Vargas AH, Selenica P, Geyer FC, Mazaheri Y, Blecua P, Conlon N, Hoang LN, Jungbluth AA, Snyder

A, Ng CKY, Papanastasiou AD, Sosa RE, Soslow RA, Chi DS, Gardner GJ, Shen R, Reis-Filho JS, Sala E. Radiogenomics analysis of intra-tumour heterogeneity in high-grade serous ovarian cancer. BJC (under review).] were printed based on manual segmentation of the ovarian mass and metastatic implants on axial T2-weighted magnetic resonance images. The lesions were outlined on every axial slice and automatically converted into 3D models using open source software (MIPAV, National Institutes for Health, Center for Information Technology). The final 3D models of each lesion were imported into OpenSCAD (OpenSCAD, The OpenSCAD Developers), a 3D CAD modelling software, which was used to create an internal cavity that exactly shaped each lesion according to the MRI shape and contour. The custom-made 3D tumour moulds were printed using a 3D printer (MakerBot Replicator 2, MakerBot, Brooklyn, NY). The slits for slicing each lesion were placed and labelled into the molds at 10mm intervals corresponding to the slice thickness and locations of the axial T2W-weighted MR images. The mould was also labelled with left, right, anterior, posterior, superior and inferior markers to allow for proper orientation when collecting samples in the operating room.

Cluster guided specimen sampling ⊙

All 3D moulds containing the specimens were taken to the histopathology department where each lesion was sampled by a pathology fellow. Each tumour was sectioned through the mould and samples were taken according to the imaging habitats/clusters defined above. Half of the sample was sent for histopathology and the other one for immunogenomic analysis.

Immunofluorescent Staining ⊙

The immunofluorescent staining was performed at Molecular Cytology Core Facility of MSKCC using Discovery XT processor (Ventana Medical Systems). The tissue sections were deparaffinised with EZPrep buffer (Ventana Medical Systems), antigen retrieval was performed with CC1 buffer (Ventana Medical Systems). Sections were blocked for 30 minutes with Background Buster solution (Innovex), followed by avidin-biotin blocking for 8 minutes (Ventana Medical Systems).

Multiplex immunofluorescence stainings were performed as previously described (Yarilin et al., 2015). Slides were incubated with anti-FoxP3 (Abcam, Catalogue Number: ab20034, 5 ug/ml) for 4 hours, followed by 60 minutes incubation with biotinylated horse anti-mouse IgG (Vector Labs, Catalogue Number: MKB-22258) at 1:200 dilution. The detection was performed with Streptavidin-HRP D (part of DABMap kit, Ventana Medical Systems),

followed by incubation with Tyramide Alexa Fluor 488 (Invitrogen, Catalogue Number: T20922) prepared according to manufacturer instruction with predetermined dilutions. Next, sections were incubated with anti-CD4 (Ventana, Catalogue Number: 790-4423, 0.5 ug/ml) for 5 hours, followed by 60 minutes incubation with biotinylated goat anti- rabbit IgG (Vector, Catalogue Number: PK6101) at 1:200 dilution. The detection was performed with Streptavidin-HRP D (part of DABMap kit, Ventana Medical Systems), followed by incubation with Tyramide Alexa 568 (Invitrogen, Catalogue Number: T20914) prepared according to manufacturer instruction with predetermined dilutions. Finally, sections were incubated with anti-CD8 (Ventana, Catalogue Number: 790-4460, 0.07 ug/ml) for 5 hours, followed by 60 minutes incubation with biotinylated goat anti-rabbit IgG (Vector, Catalogue Number: PK6101) at 1:200 dilution. The detection was performed with Streptavidin-HRP D (part of DABMap kit, Ventana Medical Systems), followed by incubation with Tyramide Alexa 647 (Invitrogen, Catalogue Number: T20936) prepared according to manufacturer instruction with predetermined dilutions. After staining slides were counter-stained with DAPI (Sigma Aldrich, Catalogue Number: D9542, 5 ug/ml) for 10 min and cover-slipped with Mowiol.

Stained slides were digitized using Pannoramic Flash 250 (3DHistech, Hungary) using 20x/0.8NA objective. Regions of interest were drawn on the scanned images using Pannoramic Viewer (3DHistech, Hungary) and exported into tiff images. ImageJ/FIJI was used to segment DAPI-stained nuclei and count the cells with positive signal.

Nucleic Acid Isolation and Quantification ⊙

DNA and RNA were extracted from tumour areas delineated as tumour nests on H&E slides reviewed by a pathologist specialized in gynecologic malignancies using the DNeasy® and RNeasy® (Qiagen) assays, respectively. RNA expression was assessed using the human Affymetrix Clariom D Pico assay (Thermo Fisher Scientific).

T-Cell Receptor Sequencing ⊙

High-throughput in-situ sequencing of the T cell receptors present in the samples and blood of the patient was performed using the immunoSEQ assay platform (Adaptive Biotechnologies).

Quantification and statistical analysis

Gene expression analysis

RNA expression was assessed using the human Affymetrix Clariom D Pico assay. Arrays were analysed using the SST-RMA algorithm in the Affymetrix Expression Console Software. Expression was determined by using the Affymetrix Transcriptome Analysis Console. Local Regression (LOESS) normalisation across samples was implemented (Gautier et al., 2004) using:

```
# R 3.5.0
library(affy) # version 1.58.0

data_norm<-normalise.loess(data,family.loess="gaussian")
```

Single-sample gene set enrichment analysis

Single-sample gene set enrichment analysis (Barbie et al., 2009), a modification of standard GSEA (Subramanian et al., 2005), was performed on RNA measurements for each sample using the GSVA package version 1.28.0 (Hänzelmann et al., 2013) in R version 3.5.0 with parameters: `method="ssgsea"`, and `tau=0.25`. normalised enrichment scores were generated for the hallmark gene sets (Liberzon et al., 2015), immune and stromal signatures (Yoshihara et al., 2013), TME cell gene sets obtained from previous publications (Bindea et al., 2013; Davoli et al., 2017), as well as Consensus^{TME} gene sets (**Figure 8.3A**). Hallmark gene sets were obtained from MSigDB database version 6.1 (Liberzon et al., 2011).

Tumour purity and immune cell gene-expression score

Tumour purity and total immune component in the tumour samples were analysed using the ESTIMATE algorithm method version 1.0.13 (Yoshihara et al., 2013) on the gene expression data using the option: `platform="affymetrix"` for the cohort samples and `platform="illumina"` for TCGA OV samples, in R version 3.5.0.

Dimensionality reduction

The t-distributed stochastic neighbor embedding (t-SNE) and PCA dimensionality reduction methods were implemented in python version 3.6.5 using the `sklearn.manifold.TSNE` and `sklearn.decomposition.PCA` functions from the sklearn version 0.19.1 package. PCA was computed after sample wise standardization. Functions were used as follows:

```
X_tsne=tsne.TSNE(learning_rate=100,
                 n_iter=5000,
                 perplexity=5).fit_transform(loess_norm_expres.values)

X_pca=PCA(n_components=7).fit_transform(nes_standarized)
```

Analysis of T cell infiltration between cases, sites, and habitats

A linear mixed effects model analysis was performed to evaluate if there were significant differences in T cell infiltration subsets between patients, sites within patients, and habitats within tumours, and to assess whether the differences were other than random. Due to data nested dependency we employed a linear mixed effects model under the `lme4` R package (Bates et al., 2015). Cases were considered a random factor, while sites and habitats were considered fixed factors as follows:

```
# R 3.5.0
library(lme4) # version 1.1-17

glmer.fit<-glmer(cbind(cd8_counts,total_cells-cd8_counts)~
           site*habitats+(1|case),family=binomial,data=data)
```

TME cell estimation and enrichment methods

Cell estimation and enrichment methods were used to estimate levels of non-cancerous cells in the TME. The methods employed were CIBERSORT (Newman et al., 2015), MCP-counter (Becht et al., 2016), TIMER (Li et al., 2016), xCELL (Aran et al., 2017), as well as gene sets collected from two previous publications (Bindea et al., 2013; Davoli et al., 2017).

CIBERSORT:

```
# R 3.5.0
source("CIBERSORT.R") # version 1.04

cibersort<-CIBERSORT("LM22.txt","expression_data.txt",perm=1000,QN=TRUE,
           absolute=FALSE)
```

MCP-counter:

```
# R 3.5.0
library(MCPcounter) # version 1.1.0

mcp<-MCPcounter.estimate(exp_data,
     featuresType=c("affy133P2_probesets","HUGO_symbols","ENTREZ_ID")[2],
     probesets=read.table(curl("http://raw.githubusercontent.com/ebecht/
     MCPcounter/master/Signatures/probesets.txt"),sep="\t",
     stringsAsFactors=FALSE,colClasses="character"),
     genes=read.table(curl("http://raw.githubusercontent.com/ebecht/
     MCPcounter/master/Signatures/genes.txt"),sep="\t",
     stringsAsFactors=FALSE,header=TRUE,colClasses="character",
     check.names=FALSE))
```

TIMER: The TIMER web server (https://cistrome.shinyapps.io/timer/) was used for regression of TME cells (Li et al., 2017c).

xCELL: The xCELL web server version 1.1 (http://xcell.ucsf.edu/) was used to calculate enrichment of TME cells.

For the Bindea et al. and Davoli et al. gene sets, standard ssGSEA analysis was performed as previously described.

Consensus^{TME} version 1.0 gene sets

To generate the Consensus^{TME} gene sets we identified cell types that were estimated by at least 2 different methods, leading to 18 different cell types. We then intersected the gene sets that the different methods considered for the estimation of such cell types. To intersect genes used in CIBERSORT, we first filtered out genes whose weight was below 1.96 standard deviations of the mean for each of CIBERSORT cell types. In addition we combined activated states into the corresponding cell type. The only activated stated included was cytotoxic cells, which would include CD8 T cells and/or NK cells in their activated stated. The intersected genes were used to represent each cell type, and genes with a higher Pearson's correlation coefficient than -0.2 and a p-value ≤ 0.05 with tumour purity as defined by TIMER were filtered out from the gene sets (Li et al., 2016). Finally, ssGSEA was employed to calculate normalised enrichment score (NES) for each cell type as described above (**Figure 8.3A**).

TME cell estimation benchmarks

T cell subsets immunofluorescent staining benchmark: We correlated the $CD8^+$, $CD4^+$, Tregs infiltration counts with the scores generated by ESTIMATE, CIBERSORT, MCP-counter, Bindea et al., Davoli et al., TIMER, xCELL, and the consensus TME scores. For the immune score comparison, all the genes used for the estimation for each method were aggregated together into one single gene set per method except for CIBERSORT. CIBERSORT estimation $-log_{10}(p-values)$ were used as a metric for immune score comparison. CD8, CD4, and Treg counts from IF-stained data were summed and used for the comparison. Because the methods have different scoring systems and ranges we standardized (z-score) the scores to be able to compare the results across methods together. For each tumour, multiple IF-stained sections were quantified for tumour infiltrating lymphocyte (TIL)s, and we correlated all the regions quantified with the scores of each tumour, explaining the vertical patterns observed in **figure 8.3A**. Spearman's rank correlation was performed for each comparison and FDR p-value correction was applied.

TCGA OV leukocyte methylation benchmark: As an independent benchmark we used leukocyte methylation scores of TCGA ovarian cancer samples (TCGA, 2011). TCGA ovarian cancer RNA sequencing (RNA-seq) data were retrieved from the cBioPortal (Cerami et al., 2012; Gao et al., 2013) web server version 1.6.2 (http://www.cbioportal.org/).

First, estimation of cell types was performed using the different methods listed above. Spearman's rank correlations were calculated between CD8 T cell scores and the leukocyte methylation score of TCGA ovarian cancer samples, and FDR p-value correction was applied.

We further fitted multiple linear regression models to each method estimated cell types (**Figure 8.3B**). We compared the proportion of leukocyte methylation score variance that is explained by the unsupervised selected immune cells (adjusted R-squared), as well as the relative quality of the models by considering goodness of fit and model simplicity after BoxCox transformation of the leukocyte methylation scores to meet the normality of residuals assumption. As a sensitivity analysis we log transformed the leukocyte methylation scores before performing the linear regression models. In both analyses (BoxCox and log-transform), stepwise Akaike Information Criterion (AIC) variable selection was performed once normality and heteroscedasticity assumptions were checked. AIC and Bayesian Information Criterion (BIC) were employed independently to compare the fitted models for each method. Both AIC and BIC quantify information loss and penalize the number of variables. Thus, the trade-off between goodness of fit and model simplicity across methods was evaluated,

allowing us to quantify and minimize information loss. The R version and packages used for this analysis were R version 3.5.0, `gamlss_5.0-8` (Stasinopoulos et al., 2007), `leaps_3.0`, `car_3.0-0` (Fox and Weisberg, 2011), and `MASS_7.3-50` (Venables and Ripley, 2002).

Differential expression analysis

Tumour samples from the treatment-naive cohort were divided into high and low purity classes taking as a cut-off the median of tumour purity calculated for the tumour samples using ESTIMATE (Yoshihara et al., 2013). Then a differential expression analysis was performed using the R packages `limma_3.36.1` (Ritchie et al., 2015) and `Biobase_2.40.0` (Huber et al., 2015). Patient dependency was taken into account as follows:

```
# R 3.5.0
library(limma) # version limma_3.36.1
library(Biobase) # version Biobase_2.40.0

data<-read.table("expression_data.txt",sep="\t",header=T,
      row.names="Hugo_Symbol")

eset<-new("ExpressionSet", exprs=as.matrix(data))

estimate<-read.table("estimate_purity_scores",sep="\t",header=T,
          row.names="NAME")

med_purity<-median(estimate$tumourPurity)

purity<-as.data.frame(ifelse(estimate$tumourPurity > med_purity,
        "high_purity","low_purity"))

row.names(purity)<-row.names(estimate)

colnames(purity)[1]<-"purity"

patient_data<-read.table("patient_data.txt",sep="\t",header=T,
             row.names="NAME")

clinical<-merge(patient_data,purity,by="row.names")
```

```
row.names(clinical)<-row.names(purity)

clinical$Row.names<-NULL

colnames(clinical)[6]<-"purity"

# merge factors
clinical<-factor(clinical$purity)

# Make design matrix
design<-model.matrix(~0+clinical)

colnames(design)<-levels(clinical)

# estimate correlation between measurements on same subjects
corfit<-duplicateCorrelation(eset,design,block=patient_data$case)

# inter-subject correlation is input into the linear model fit
fit<-lmFit(eset,design,block=patient_data$case,
           correlation=corfit$consensus)

cm<-makeContrasts(HighPurityvsLowPurity=high_purity-low_purity,
    levels=design)

fit2<-contrasts.fit(fit,cm)

fit2<-eBayes(fit2)

results<-decideTests(fit2,adjust.method="fdr")

volcanoplot(fit2)
```

Gene ontology analysis

Gene ontology analysis of significantly up or down-regulated genes was performed using the Gene Ontology Consortium (Ashburner et al., 2000; The Gene Ontology Consortium, 2017) web server (http://www.geneontology.org/). *P*-value FDR corrections were calculated for this analysis.

ssGSEA of differential expression analysis

Further, p-values for each gene were retrieved and multiplied by -1 if the log_2 change was negative. The list of genes with their associated p-value was used to calculate hallmark and consensus TME NES through ssGSEA. Hallmark gene sets' NES were normalised by taking the exponential function. Consensus^{TME} gene sets' NES approached normality by taking the natural logarithm. The modified z-score ($\hat{z}_x$) was calculated to detect outliers in the hallmarks and Consensus^{TME} NES independently, as the $\hat{z}_x$ uses the median and the median absolute deviation ($\hat{x}(|x_i - \hat{x}|)$ to robustly measure central tendency and dispersion in small data sets (Iglewicz and Hoaglin, 1993).

$$\hat{z}_x = \frac{0.6745 \times (x_i - \hat{x})}{\hat{x}(|x_i - \hat{x}|)} \tag{3.1}$$

Paired gene set comparisons

Volcano plots: For each of the 52 hallmark and 18 Consensus^{TME} gene sets paired comparisons before and after NACT were performed. Equality of variance (Bartlett's test) and normality (Shapiro test, Kolmogorov-Smirnov test, and D-Agostino-Pearson's test) assumptions were checked to select the corresponding paired test (Paired T-test, Welch's T-test, or Wilcoxon signed-rank test). The analysis was performed using python 3.6.5 and scipy 1.1.0 (http://www.scipy.org/) (Oliphant, 2007).

Hotelling's T^2 distribution test: Multivariate Hotelling's T^2 test was performed to compare difference of CD8, NK, and cytotoxic Consensus^{TME} gene sets NES between pre- and post-NACT tumours as follows:

```
# R 3.5.0
library(Hotelling) # version Hotelling_1.0-4

data<-read.table("consensusTME_NES.txt",sep="\t",header=T,
      row.names="NAME")

matched<-data[which(data$related=="matched"),]
matched<-subset(matched,select=-c(case_number,Site,related))
unmatched<-data[which(data$related==unmatched),]
unmatched<-subset(paired,select=-c(case_number,Site,related))

fit_matched=hotelling.test(nk*cd8*cytotoxic~NACT,data=matched,perm=T)
```

```
fit_unmatched=hotelling.test(nk*cd8*cytotoxic~NACT,data=unmatched,
              perm=T)

plot(fit_matched)
plot(fit_unmatched)

fit_matched$pval
fit_unmatched$pval
```

TCR sequencing analysis

Analysis of the sequences was performed on the immunoSEQ ANALYZER 3.0 (Adaptive biotechnologies). T cell counts and TCR clonality were retrieved for statistical comparisons. T cell counts are derived from quantitative immunoSequencing of the TCR β-chain loci, in which the internal controls allow precise quantitation of sequence counts based on reads. Nucleated cell counts are determined by sequencing housekeeping genes. The fraction of T cells is determined by dividing the T cell count by the nucleated cell counts. Values for TCR productive clonality range from 0 to 1. Values near 1 represent samples with one or a few predominant rearrangements (monoclonal or oligoclonal samples) dominating the observed repertoire. TCR productive clonality values near 0 represent more polyclonal samples. TCR productive clonality is calculated by normalizing productive entropy using the total number of unique productive rearrangements and subtracting the result from 1.

LASSO regression and post-selection inference

least absolute shrinkage and selection operator (LASSO) regression analysis was performed using the `glmnet` R package (Friedman et al., 2010). Hallmark and Consensus^{TME} cell type NES of pre-NACT samples were used independently as explanatory variables, and the log_2 of the ratio post/pre NACT TCR clonality as response variable.

```
# R 3.5.0
library(glmnet) # version glmnet_2.0-16

dataH<-read.table("Hallmarks_TCRdiff.txt",sep="\t",header=T,
       row.names="ID")
dataC<-read.table("ConsensusTME_TCRdiff.txt",sep="\t",header=T,
       row.names="ID")
```

```
h<-as.matrix(subset(dataH,select=-c(clon_dif)))
c<-as.matrix(subset(dataC,select=-c(clon_dif)))
y<-as.matrix(dataH$clon_dif)

hallmark_fit<-glmnet(h,y,family="gaussian",alpha=1,standardize=T)
tmecells_fit<-glmnet(c,y,family="gaussian",alpha=1,standardize=T)

# VARIABLE SELECTION
hallmark_cvfit=cv.glmnet(h,y,family="gaussian",type.measure="mse",alpha=1,
                nfold=10)
tmecells_cvfit=cv.glmnet(c,y,family="gaussian",type.measure="mse",alpha=1,
                nfold=10)

coef(hallmark_cvfit,s="lambda.min")
coef(tmecells_cvfit,s="lambda.min")

# POST-SELECTION INFERENCE
library("selectiveInference") # version selectiveInference_1.2.4

postselinf_h=fixedLassoInf(h,y,beta_hat[-1],lambda,family="gaussian")
postselinf_c=fixedLassoInf(c,y,beta_hat[-1],lambda,family="gaussian")
```

Data and Software Availability

Requests for additional data and custom code should be directed to Dr Martin L. Miller (martin.miller@cruk.cam.ac.uk).

Supplementary tables

Supplementary tables can be accessed through:
https://github.com/cansysbio/immunogenomics/tree/master/AJSPhDThesis/AppendixB/Tables/

3.5 Results

3.5.1 Intra-patient transcriptomic heterogeneity is largely explained by immune signalling

To investigate the TME of HGSOC in a treatment-naive context, we analysed the transcriptome of 38 primary and metastatic tumour samples from 8 treatment-naive patients collected prospectively (**Figure 3.1A**, **Table C3S1A** and **Table C3S1B**). Primary tumour masses and peritoneal metastases were resected and placed on patient-specific 3D moulds created based on tumour segmentation using high resolution T2-weighted magnetic resonance magnetic resonance (MR) images. Each specimen was placed in the custom-made 3D mould in the operating theatre and was further dissected into sub-specimens according to three multiparametric imaging-based phenotypically distinct clusters, hereafter referred to as "habitats". Habitats were obtained from MR and 18F-FDG-PET imaging and were defined based on quantitative imaging features that measure water diffusion, micro-capillary perfusion, permeability and metabolic activity (**see Methods**). We first performed an unbiased clustering analysis of the whole transcriptome. We observed that overall gene expression of tumour samples was highly patient specific, irrespective of anatomical site using t-distributed stochastic neighbor embedding (t-SNE) which accounts for non-linear relationships (**Figure 3.1B**). To focus on well-defined biological processes and signalling pathways, we performed ssGSEA (Hänzelmann et al., 2013) using the hallmark gene sets (Liberzon et al., 2015), stromal and immune gene signatures, and tumour cell fraction (purity) using the ESTIMATE algorithm (Yoshihara et al., 2013). We categorised the gene sets into five classes: oncogenic, cellular stress, immune, stromal, and other.

Principal component analysis (PCA) showed that most of the gene set expression variation between samples (60% of variation) could be explained by oncogenic, immune, and stroma-associated gene sets (**Figure 3.1C, 8.1A)**. In contrast to the full transcriptome analysis, the patient specific clustering was less evident, indicating that tumours from different patients share common patterns of pathway activation and non-cancer cell infiltrates. To investigate which gene sets explained most of the observed variance, we computed the principal component feature loadings and displayed them in a variable factors map (**Figure 3.1D**). This analysis showed that PC1 (40% of variation) is largely explained by tumour purity, since immune and stromal vectors had an opposite direction to oncogenic vectors and tumour purity (immune vs oncogenic: p-value = 3×10^{-05}; stromal vs oncogenic: p-value = 1.3×10^{-03}), and PC2 (20% of variation) further showed a separation of immune, stromal,

and cellular stress vectors (immune vs stromal: p-value = 0.046; immune vs stress: p-value = 0.041; **Figure 8.1B**).

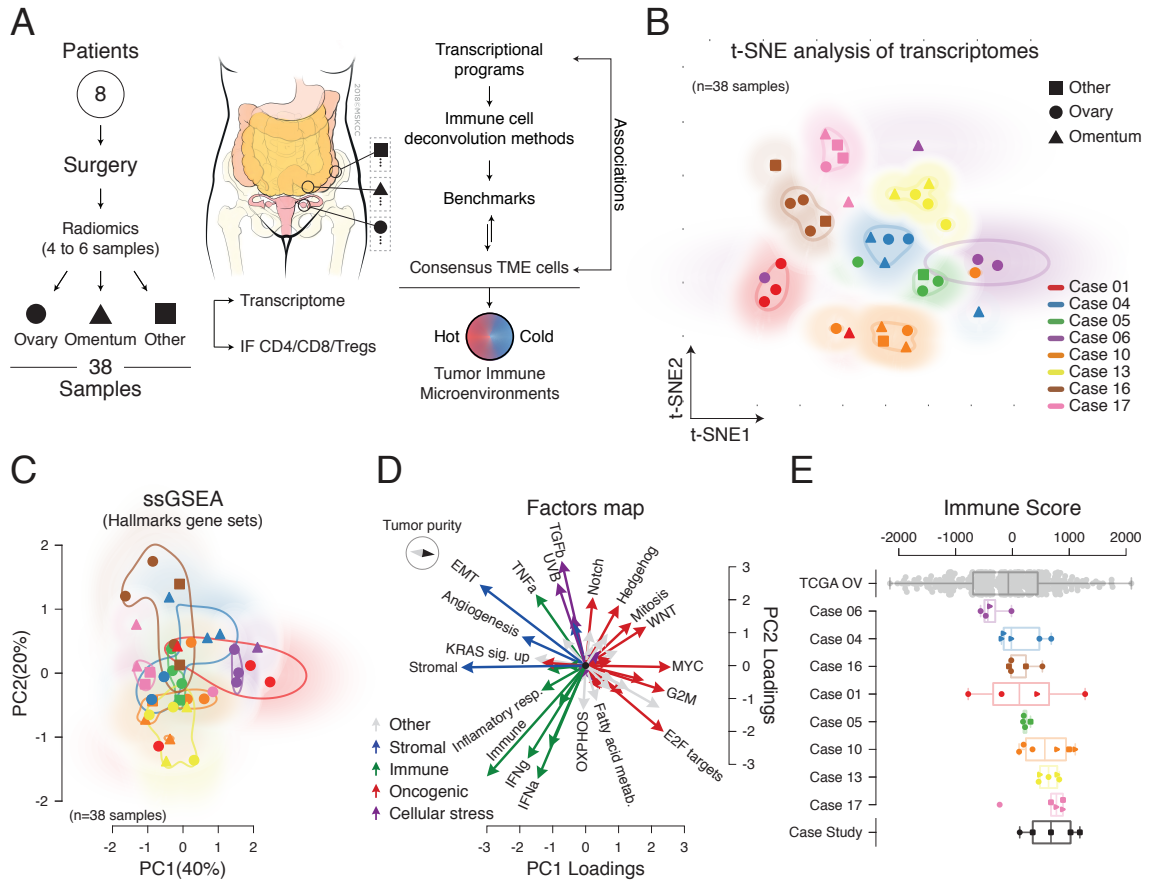


Fig. 3.1 Immune signalling contributes to the majority of the transcriptional variance observed across multiple tumour samples from treatment-naive HGSOC patients. (A) Flowchart of sample acquisition and analysis. Peritoneal metastases other than omentum were defined as "Other". (B) t-SNE analysis of overall transcription profiles of multiple HGSOC tumour samples per patient. (C) PCA of ssGSEA-based analysis of hallmark gene sets. D) Principal component feature loadings (magnitude and direction) of C are shown in the variables factor map. Vectors are coloured according to a major biological classification of hallmark gene sets. Variation across classes in PC1 (p-value=3.2^{-16}) and PC2 (p-value=0.02) after Kruskal-Wallis H-test (**Figure 8.1**). Directionality of ESTIMATE's tumour purity is represented with the map compass. E) ESTIMATE immune score across patients and samples. The Case Study samples were taken from (Jiménez-Sánchez et al., 2017). The bottom and top edges of the box plots indicate the 25th and 75th percentiles. TCGA OV (n = 301).

Since immune related pathways explained a significant amount of variation between the samples, we further investigated the extent of intra-patient immune heterogeneity by computing the ESTIMATE immune score for each sample. In addition, we included as a reference the immune scores of the samples from a HGSOC case study with >9 years of clinical history we

previously analysed (Jiménez-Sánchez et al., 2017) and the immune scores of ovarian cancer samples from The Cancer Genome Atlas (TCGA), which comprises 307 treatment-naive primary tumours (TCGA, 2011). Overall, the immune scores of our cohort fell within the range expected at the population level (**Figure 3.1E**). Some patients (01, 04, 10, and the case study) showed an intra-patient variation comparable to the inter-patient variation observed at the population level by the TCGA ovarian cancer samples, which indicates that within a single individual, complete distinct immune microenvironments can co-exist at diagnosis of HGSOC. Also, all patients in the cohort had at least one sample with similar or lower immune score than the progressing and immune excluded tumours of the case study, where distinct tumour-immune microenvironments led to different clinical outcomes (Jiménez-Sánchez et al., 2017). Importantly, consistent with our prior report, we recapitulate the observation that tumours with high immune signalling and immunosuppressive Wnt signalling tend to be mutually exclusive (**Figure 3.1D**).

3.5.2 Co-existence of distinct tumour-immune microenvironments in treatment-naive HGSOC

To further characterise the tumour microenvironment of HGSOC, we performed multicolour IF staining and quantification of $CD4^+$, $CD8^+$, and regulatory T cells ($CD4^+$ $FOXP3^+$) in at least 10 tumour regions excluding stromal areas in each sample leading to a compendium of 440 imaged and digitally quantified tumour regions (**Figures 3.2A-B**, **8.2**, and **Table C3S2A**). This multi-region and multi-site IF analysis shows that treatment-naive HGSOC patients present variation in T cell infiltration in tumour deposits, ranging from less than 1% (e.g. patient 6) to T cells accounting for more than 10% of total cells in some areas (e.g. patient 01 and 10). Furthermore, some patients' tumour deposits demonstrated marked variation in T cell infiltration within the same tumour deposit across different habitats (e.g. patient 01). We then performed a linear mixed effects model analysis (**see Methods** and **Table C3S2B**) to statistically evaluate whether there is a difference of T cell infiltration between patients, between tumours of the same patients and between habitats within tumours. We found a remarkable difference in T cell infiltration across tumours within patients (DF=2, $CD8^+$ F-value=8758; $CD4^+$ F-value=58; Tregs F-value=657) and habitats within a tumour (DF=2, $CD8^+$ F-value=1184; $CD4^+$ F-value=2870; Tregs F-value=2216). Sites and habitats also showed an important level of interaction (DF=4, $CD8^+$ F-value=5915; $CD4^+$ F-value=4466; Tregs F-value=142). This systematic T cell IF staining and computerised cell detection-counting confirm the variation in T cell infiltration across patients, across tumour samples within patients, within tumours and within habitats. Together with the transcriptome analysis,

these data show that HGSOC is intrinsically characterised by the presence of heterogeneous immune TMEs within patients and within tumours.

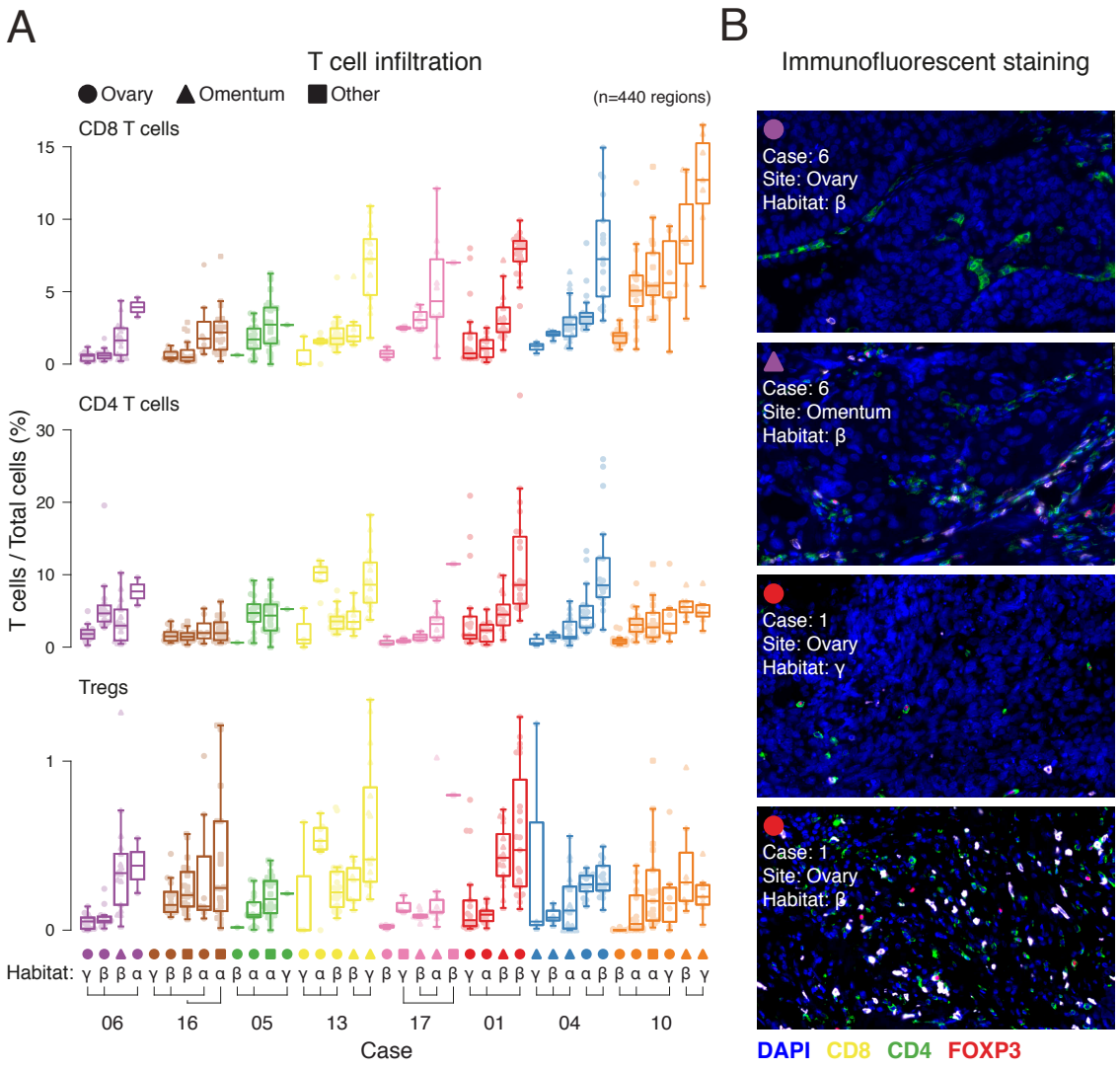


Fig. 3.2 T cell infiltrate variation across patients, within patients, and within tumours. (A) Multi-tumour sampling from 8 HGSOC patients are shown with each dot representing the percentage of T cell subsets in a quantified area within a given tumour section stained with multicolour IF for $CD8^+$, $CD4^+$ and $CD4^+FOXP3^+$. Stromal areas were excluded based on H&E stains. Patient cases are indicated by different colours. Anatomical sites of tumour deposits are indicated by different markers (circle, triangle and square). Habitats are defined by the Greek letters α, β and γ. Habitats from the same tumour are indicated by connecting lines. Box-plots are sorted according to the median of $CD8^+$ T cell infiltration across patients, sites and habitats accordingly. (B) Representative images of panel A.

3.5.3 Consensus tumour microenvironment gene sets improve cell estimation from bulk tumour mRNA

To estimate the relative abundance of different cell types in an unbiased manner using bulk RNA data, various computational approaches have been developed during the last decade (Finotello and Trajanoski, 2018). However, a comparison that objectively evaluates the performance of these approaches against one another has not been conducted and no publicly available data sets have been generated to serve as ground truths thus far (Zheng, 2017). Therefore, to test the performance of cell-type specific estimation methods, we performed a preliminary benchmark using the T cell IF quantification of the 440 regions from the 38 treatment-naive HGSOC samples as the ground truth. We used ESTIMATE for total T cell infiltration (Yoshihara et al., 2013), and compared CIBERSORT (Newman et al., 2015), TIMER (Li et al., 2016), MCP-counter (Becht et al., 2016) and xCell (Aran et al., 2017) for cell type specific estimations. We also evaluated immune gene sets that were defined based on gene expression of sorted immune populations (Bindea et al., 2013), as well as immune gene sets based on the Immunological Genome Project database (Davoli et al., 2017; Heng et al., 2008). Using ssGSEA (Hänzelmann et al., 2013), the Bindea et al. and Davoli et al. gene sets were used to calculate normalized enrichment scores (NES) of the corresponding cell types. To be able to evaluate the performance of ESTIMATE, which calculates a total immune score but does not estimate immune cell types, total immune scores from the different tools were generated (**see Methods**). Total immune scores and cell-type specific scores for $CD4^+$, $CD8^+$ and Tregs were correlated independently against the aggregated T cells and the corresponding cell type fractions (**Figure 3.3A**). Importantly, not all of the methods selected estimate $CD4^+$, $CD8^+$, and Tregs; however, for the methods that estimate all three cell types, none consistently outperformed the other methods in this independent benchmark analysis. In addition, none of these methods was able to get a significant positive correlation with Tregs, which may indicate a lower-limit threshold of sensitivity of detection, since Tregs comprised, on average, less than 1% of cells in the tumour nests (**Figure 3.2A**, **8.2**).

Since the cell-type specific estimation methods were developed independently of each other, we reasoned that generating consensus gene sets by including the genes that fall in the intersection of different cell types across the different tools could improve the cell estimation performance. Since not all methods estimate the same cell types, we focused on cell types that at least two methods estimate. We selected the union of genes, and finally removed genes whose expression levels positively correlated with tumour purity using TCGA ovarian cancer samples as a reference (**see Methods** and **Figure 8.3A**).

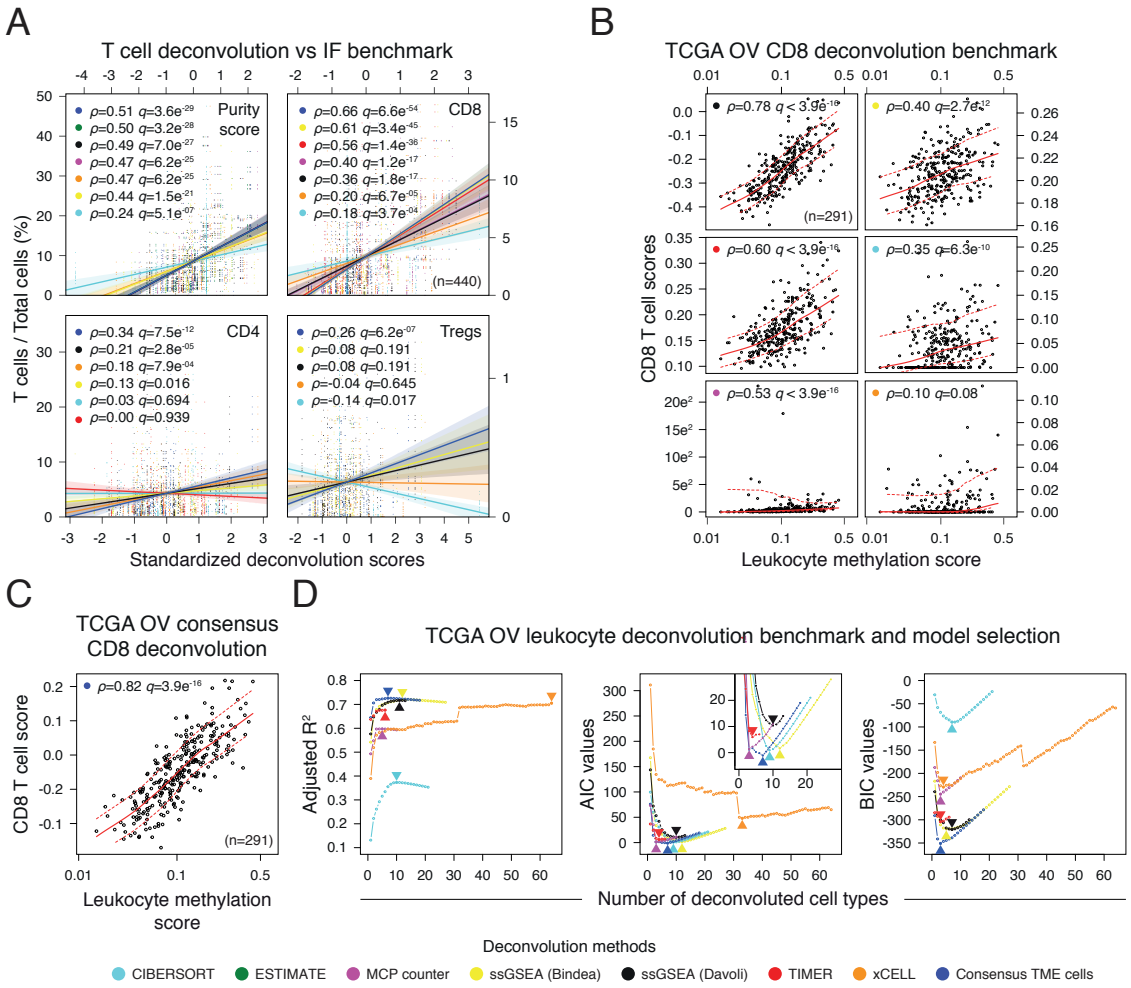


Fig. 3.3 Consensus tumour microenvironment cell enrichment method improves estimation of tumour infiltrating T cells and leukocytes. (A) Spearman's rank-order correlations between percentage of T cells (immunofluorescent staining) and the corresponding scores for each method using the treatment-naive ovarian cancer samples. Correlation coefficients (ρ) and q-values are ordered from higher and more significant to lower and less significant. Scores were standardized (z-score) to visually compare the different correlations in the same scale. IF immune score was calculated by adding up $CD8^+$, $CD4^+$, and Tregs counts as an approximation, while immune scores were calculated according to each tool (see Methods). (B, C) TCGA ovarian cancer Spearman's rank-order correlations between $CD8^+$ T cell scores and leukocyte methylation scores. D) Multiple linear regression analysis using leukocyte methylation score as response variable, and estimated cell types as explanatory variables (**see Methods**). Adjusted R2, AIC, and BIC values were calculated to compare goodness of fit and model simplicity. Arrowheads indicate best model for each method. Inset in the AIC panel shows a magnification of the best ranked models.

We called this approach Consensus^{TME} version 1.0. We correlated the ssGSEA NES of the Consensus^{TME} v.1.0 cell gene sets against the fraction of T cells quantified, and observed that the Consensus^{TME} v.1.0 gene sets consistently showed higher positive correlations

than the individual methods. In addition, the Consensus^{TME} v.1.0 Tregs NES was the only score with a significant positive correlation with the fraction of Tregs, suggesting a greater level of sensitivity obtained with Consensus^{TME} v.1.0 (**Figure 3.3A**, Spearman's ρ=0.26, q-value=6.2×10^{-07}).

To further benchmark the methods and the Consensus^{TME} v.1.0, we employed TCGA ovarian cancer leukocyte methylation scores (TCGA, 2011), which is an independent and larger patient cohort. Leukocyte methylation measurements provide orthogonal means toward estimating immune infiltration in tumours, and have been shown to significantly correlate with histological purity estimates in primary HGSOC (Carter et al., 2012). Importantly, the leukocyte methylation signature was generated by comparing methylation patterns between HGSOC tumours, normal fallopian tube samples and buffy-coat samples of female individuals, making this data set ideal for benchmarking the different cell-type specific estimation methods (Carter et al., 2012). We first performed a benchmark of all methods using the leukocyte methylation scores and the CD8$^+$ T cell scores (**Figure 3.3B**), as i) this sub-population is a major component of infiltrating leukocytes, ii) all methods tested estimate CD8$^+$ T cells, and iii) overall the CD8$^+$ IF estimations were the best correlations (**Figure 3.3A**). Among the seven methods, Consensus^{TME} v.1.0 CD8$^+$ gene signature correlated best with the CD8$^+$ T cell score (**Figure 3.3C**, Spearman's ρ=0.82 p-value=3.9×10^{-16}). However, as the leukocyte methylation score does not count CD8$^+$ T cells exclusively, we then compared the different methods in an unbiased manner by fitting a multiple linear regression model for each method using the leukocyte methylation score as a response variable and the different cellular scores as explanatory variables, followed by unsupervised nested variable selection (**see Methods** and **Figure 8.3B**). We compared the proportion of leukocyte methylation score variance that is explained by the unsupervised selected estimated cells (adjusted R^2), as well as the relative quality of the models by considering goodness of fit and model simplicity (**see Methods**). The Consensus^{TME} v.1.0 provided the highest adjusted R-squared with fewer cell types selected (Adjusted R-squared=0.73, p-value < 2.2×10^{-16}, **Figure 3.3D left panel**), as well as being selected as the simplest and most accurate model to explain leukocyte methylation (**Figure 3.3D middle and right panels**). Finally, the Consensus^{TME} v.1.0 gene sets that the systematic unbiased analysis suggested best explained leukocyte methylation, were cells expected to be present in leukocyte infiltrates, with CD8$^+$ T cells and NK cells accounting for the vast majority of the variation explained (**Table C3S3**, CD8$^+$ p-value=1.74×10^{-07}, and NK cells p-value=1.79×10^{-11}). In addition, we performed a sensitivity analysis of the leukocyte methylation benchmark, and the Consensus^{TME} v.1.0 was also the best method with the same cells explaining leukocyte methylation selected

(**Figure 8.3C** and **Table C3S3**). Together, these benchmarks show a consistent improvement of leukocyte cell estimation provided by Consensus^{TME} v.1.0 in ovarian cancer samples.

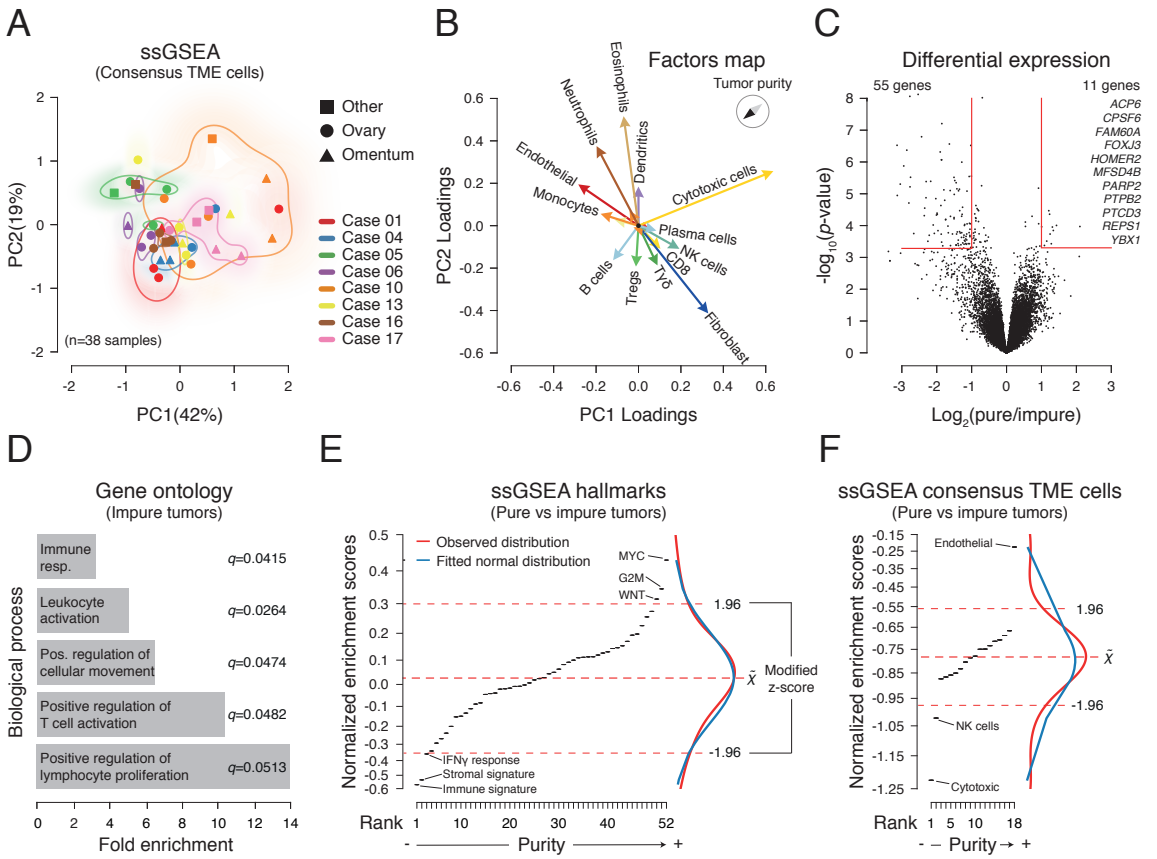


Fig. 3.4 Unbiased analysis of tumour microenvironment heterogeneity in treatment-naive HGSOC tumours. (A) PCA of ssGSEA-based analysis using Consensus^{TME} v.1.0 estimated cell gene sets. (B) Principal component feature loadings (magnitude and direction) of A. Vectors are coloured according to cell types, for example monocytes and macrophages M0, M1, M2 (orange), B cells and plasma cells (light blue), and CD8 and cytotoxic cells (yellow). (C) Differential expression analysis of high purity and low purity classified tumours using the median purity score of the cohort as a cut-off (**see Methods**). Vertical red lines indicate +/- 1 fold change of gene expression, and the horizontal line indicates the corresponding 0.05 q-value on the y-axis. D) Gene ontology analysis of significantly highly expressed genes on low purity tumours. Significantly highly expressed genes in pure tumours are not significantly over represented in any gene ontology biological process. (E, F) ssGSEA analysis of differentially expressed genes using hallmarks and Consensus^{TME} v.1.0 estimated cells, respectively (**see Methods**). Gene sets on the x-axes were ranked according to their NES (**Tables C3S4A** and **C3S4B**). Higher NES are indicative of higher purity scores. Dashed red lines indicate median and $\pm$ 1.96 median absolute deviations (modified z-score) to define outliers. Marginal density plots of observed and values fitted to a normal distribution are shown.

3.5.4 Tumour microenvironment cell decomposition and molecular comparison of high and low purity treatment-naive HGSOC

Having generated our own robust method for immune cell estimation, Consensus^{TME}, we applied it to the treatment-naive HGSOC transcript data to systematically assess if specific transcriptional programs were associated with variability in immune infiltration. We first visualized the variation across samples using the NES of estimated Consensus^{TME} gene sets of cells (**Figure 3.4A**, **8.4A**). The gene sets explaining most of the variation were cytotoxic, NK cells and fibroblast being negatively correlated with tumour purity; while endothelial, monocytes and B cells positively correlated with tumour purity (**Figure 3.4B**, **8.4B**). The NES of estimated cells of this cohort were comparable to the NES obtained in the TCGA ovarian cancer data set (**Figure 8.5**). Interestingly, the cells with highest NES in most samples were fibroblasts, highlighting the intrinsic low-immunogenic nature of HGSOC (Wang et al., 2016).

To investigate genes associated with tumours with high cellularity (pure tumours), we used the median tumour purity of the cohort to classify high and low purity tumours (**see Methods**), and performed a differential expression analysis leveraging sample-patient dependency (i.e. considering that multiple tumours come from the same individual) to increase statistical power. As expected, genes related to immune activation were significantly highly expressed in low purity tumours, but only eleven genes were significantly highly expressed in the purer tumours compared to the lowly pure ones (**Figure 3.4C**). Gene Ontology (GO) analysis showed that genes with significant higher expression in low purity tumours are significantly enriched in leukocyte proliferation and activation GO biological processes (**Figure 3.4D**), whereas no significant GO enrichment was found with the genes significantly highly expressed in pure tumours. The genes with significantly higher expression in pure tumours have been implicated in transcription [CPSF6 (Rüegsegger et al., 1998), FOXJ3 (Landgren and Carlsson, 2004)], cellular growth [HOMER2 (Tu et al., 1998; Xiao et al., 1998), REPS1 (Cantor et al., 1995; Hu and Mivechi, 2003)], glucose transport [MFSD4B (Horiba et al., 2003)], mitocondrial generation [PTCD3 (Davies et al., 2009)], aberrant proliferation [YBX1 (Frye et al., 2009; Weidensdorfer et al., 2009)], ovarian cancer initiation and progression [ACP6 (Fang et al., 2002; Hiroyama and Takenawa, 1999), PARP2 (Amé et al., 1999; Gunderson and Moore, 2015)] and TGF-beta signalling down-regulation [FAM60A (Muñoz et al., 2012; Smith et al., 2012)], which consequently can directly and indirectly promote tumourigenesis through TME immunosuppression (Colak and ten Dijke, 2017; Mariathasan et al., 2018; Tauriello et al., 2018).

To further investigate which molecular signalling pathways are more highly enriched in pure tumours, we performed ssGSEA using the adjusted *p*-values and changed the sign, positive or negative, according to the differential expression direction (see Methods). As expected, immune and stromal signatures were highly enriched in low purity tumours, in addition to interferon (IFN)-γ response. In contrast, Myc and Wnt signalling appeared to be highly enriched in pure tumours, both of which have been previously associated with immune exclusion in pre-clinical models of lung cancer (Kortlever et al., 2017) and melanoma (Spranger et al., 2015, 2016), respectively (**Figure 3.4E**). Not surprisingly, the proliferation related hallmark G2M was highly enriched in pure tumours. Of note, little or no overlapping between the G2M, Myc and Wnt hallmark gene sets was observed (**Figure 8.4C** and **Table C3S4A**; 6 out of 258 genes overlapped between G2M and Myc, and 2 out of 242 genes overlapped between G2M and Wnt signalling gene sets). Considering the TME, cytotoxic and NK estimated cells were preferentially enriched in low purity tumours, whereas endothelial cells were highly prevalent in pure tumours (**Figure 3.4F**). Since all genes in the cytotoxic gene set are included in the $CD8^+$ T cell and NK cell gene set, this suggests that a particular activation of NK cells is more prominent in low purity tumours (**Figure 8.4C**), whereas pure tumours only showed enrichment of endothelial cells (**Table C3S4B**). These observations suggest that Myc and Wnt signalling gene set enrichments in pure tumours could be considered at least partially independent of tumour proliferation, and may also contribute to immune cell exclusion as suggested by other studies (Spranger and Gajewski, 2018).

3.5.5 Chemotherapy induces immune activation in HGSOC

To investigate the effect of chemotherapy on the TME and evaluate if there is a confounding effect of the intra-patient TME heterogeneity described above, we studied the transcriptome of site-matched (n=18) and site-unmatched (n=38) primary and disseminated tumours before and after treatment with neoadjuvant platinum and taxane chemotherapy in 28 HGSOC patients (**Figure 3.5A**, **Table C3S5A** and **C3S5B**). Using t-SNE dimensionality reduction on the whole transcriptomes, we found that treated and untreated samples cluster separately (**Figure 3.5B**), in contrast to the treatment-naive samples that cluster in a patient-specific manner (**Figure 3.1B**). Using the ssGSEA NES of the hallmark gene sets of site-matched and site-unmatched samples, we observed that treated and untreated sample groups were separated by the two first principal components with 63% and 50% of variation in site-matched and site-unmatched groups, respectively (**Figures 3.5C** and **8.6A-B**). Both site-matched and site-unmatched groups showed that oncogenic and immune/stromal hallmarks contributed significantly to the variation explained by the first principal components (**Figure 8.6C-D**).

However, only site-matched PC1 reached statistical significance after paired comparison between pre- and post-NACT samples (**Figure 8.6E-F**), while also explaining more than 50% of the variation in the site-matched samples (**Figure 8.6A-B**). Interestingly, cellular stress related pathways were more enriched in site-unmatched post-NACT than site-matched samples, potentially reflecting the cellular stress generated by therapy, whereas in site-matched samples, immune related pathways dominate the variation signal (**Figures 3.5C**, **8.6C-D**). In addition, Wnt and Myc signaling showed a clear negative association to immune related gene sets in the site-matched samples, whereas no clear clustering contribution of Wnt and Myc gene sets is discernible in the site-unmatched samples (**Figure 3.5C**).

We then decomposed the TME cellular mixtures to investigate which cells were differentially present in the pre- and post-treatment tumours. Pre- and post-treatment samples also clustered separately both in site-matched and site-unmatched samples, with the first principal components accounting for 64% and 43% of variation explained, respectively (**Figures 3.5D**, **8.7A-B**). Eosinophils and cytotoxic cells showed a negative association with tumour purity in site-matched samples, in contrast to fibroblasts, B cells and macrophages (**Figures 3.5D**, **8.7C**). Interestingly, the only principal components that were significantly different between pre- and post-NACT samples were the first two principal components of the site-matched tumours (**Figure 8.7E-F**).

3.5.6 Tumour-immune microenvironment intra-patient heterogeneity masks chemotherapy induced immune activation effect

To directly evaluate differences between pre- and post-treatment samples, we performed an exploratory data analysis leveraging the possibility of performing paired comparisons using the hallmark and Consensus^{TME} v.1.0 gene set NES independently for site-matched and site-unmatched samples (**Figure 3.6A**). Site-matched samples showed a clear increase of immune pathways and Consensus^{TME} v.1.0 gene sets in post-treatment samples, while site-unmatched samples showed an increase of cellular stress pathways reflecting cellular and metabolic stress after cytotoxic drug exposure, but no difference of Consensus^{TME} v.1.0 gene sets was detected in the site-unmatched cohort. Since we observed that cytotoxic and NK cell Consensus^{TME} v.1.0 gene sets were mainly enriched in treatment-naive low purity tumours (**Figure 2.4F**), and it is known that CD8^{+} T cells play a crucial role in ovarian cancer recurrence and overall survival (Zhang et al., 2003), we performed a multivariate Student's t-statistic hypothesis test using cytotoxic, NK and CD8^{+} T cell Consensus^{TME} v.1.0 gene sets.

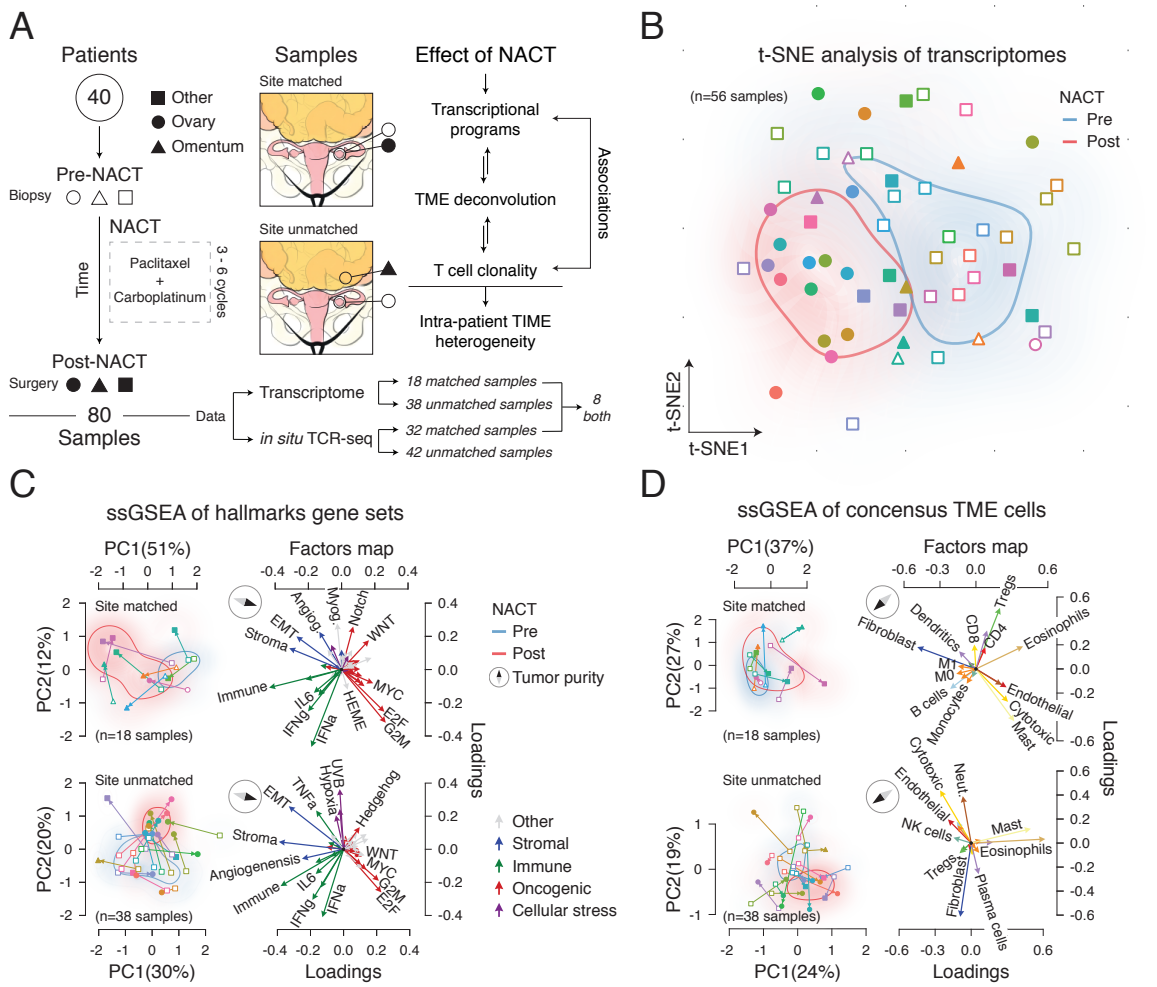


Fig. 3.5 Unbiased signalling pathway and tumour microenvironment cell decomposition analysis of chemotherapy treated HGSOC site-matched and unmatched tumour samples. (A) Flowchart of sample acquisition, clinical study design, and analysis. TME = Tumour microenvironment. (B) t-SNE analysis of overall transcription profiles of multiple HGSOC tumour samples per patient. (C, D) PCA and principal component feature projections (magnitude and direction) of ssGSEA-based analysis of hallmark and Consensus^{TME} v.1.0 gene sets respectively. Arrows in the principal component space indicate pre- to post-NACT directionality. Hallmark gene set vectors are coloured according to a major biological classification. Angiog = Angiogenesis, Myog = Myogenesis. Consensus^{TME} v.1.0 vectors are coloured according to cell types. Neut = Neutrophils.

We compared the difference between these 3 Consensus^{TME} v.1.0 gene sets in pre- and post-NACT site-matched and site-unmatched samples (**Figure 3.6B**), and observed a significant increase of these immune cell types upon chemotherapy in the site-matched samples (p-value=0.0034), but no difference in the site-unmatched samples (p-value=0.92). To further evaluate T cell infiltration and activation between pre- and post-NACT samples, we performed *in situ* TCR sequencing. Since T cell activation leads to clonal expansion of particular

T cell clonotypes, TCR clonality measures can be used as a surrogate for T cell activation upon specific (neo)antigen recognition (Jiménez-Sánchez et al., 2017; Kirsch et al., 2015; Pielou, 1966). TCR clonal expansion was significantly higher in post-NACT site-matched samples (**Figure 3.6C**, p-value=0.001), but no significant difference was observed in site-unmatched samples (p-value=0.2). T cell fraction was also significantly higher in post-NACT site-matched samples (p-value=0.03), while a slightly lower T cell fraction was observed in site-unmatched post-NACT tumours, potentially as a result of the variability of immune infiltration between omentum metastases (pre-NACT biopsies) and primary tumours (post-NACT debulking surgery). To test whether the biopsy intervention could be a confounder and lead to an immune activation *per se*, we compared post-NACT site-matched and post-NACT site-unmatched TCR clonal expansions, since the post-NACT site-unmatched tumours were not originally biopsied. No significant difference was observed in TCR clonal expansion (p-value=0.67), suggesting that T cell clonal expansion was independent of biopsy treatment and likely induced by NACT. However, a significant increase of T cell density was observed in site-matched compared to site-unmatched post-NACT tumours (p-value=3.59×10^{-05}), potentially suggesting that wound healing after the biopsy procedure could increase the influx of T cells. Together, these results provide evidence that neoadjuvant chemotherapy induces an immune activation in the local TME of HGSOC, and that intra-patient inter-site TME heterogeneity can obscure this clinically relevant observation among tumour deposits within patients.

Finally, we investigated whether hallmark pathways or Consensus^{TME} gene signatures calculated from the pre-treatment samples could explain the increase of TCR clonality upon neoadjuvant chemotherapy in site-matched samples (n=8, only 8 pre-treated samples with both gene expression and TCR-seq data were available, **see Figure 3.5A**, **Table C3S5B**). To perform the analysis in an unbiased manner, we employed LASSO regression analysis with the change of TCR clonality before and after NACT as a response variable (**see Methods**). The hallmark pathways that potentially have predictive value with positive association were the hallmarks G2M checkpoint, Myc and Wnt signalling and UV response, while apoptosis and PI3K-AKT-MTOR showed a negative association with TCR clonality increase (**Figure 3.6D**). We then performed the same analysis using the Consensus^{TME} gene signatures, where B cells, M1 macrophages and endothelial cells showed a positive association with TCR clonal expansion, while cytotoxic, CD4^{+} and plasma cells showed a negative association. In addition, the LASSO analysis selected the Consensus^{TME} v.1.0 cytotoxic signature as relevant for explaining T cell clonal expansion upon NACT, and a correlation analysis supported this association (**Figure 3.6E**).

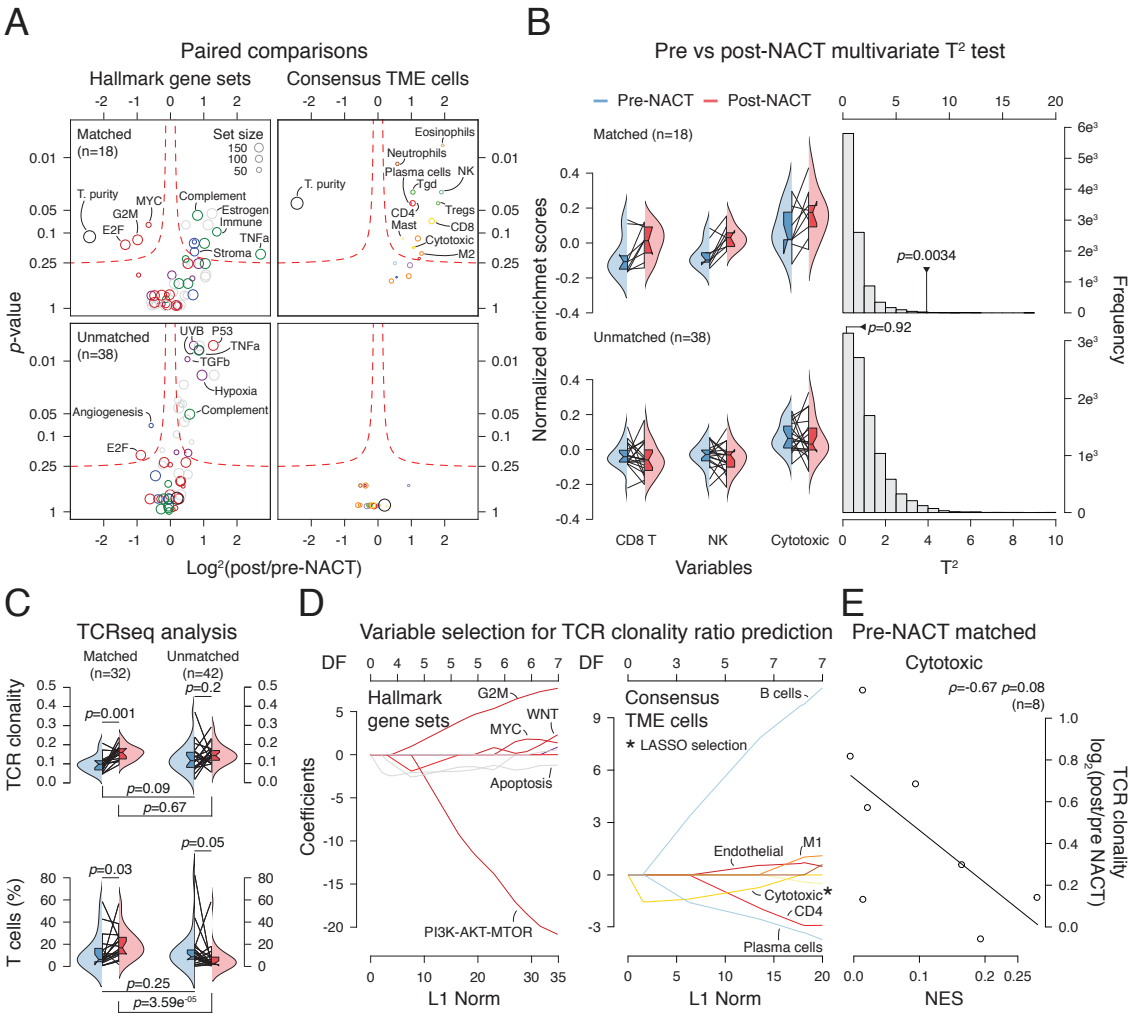


Fig. 3.6 Immune activation induced by neoadjuvant chemotherapy is evident in site-matched but not site-unmatched sample analysis. (A) Exploratory pre/post NACT paired comparisons of hallmark gene sets and Consensus^{TME} v.1.0 estimated cells (**see Methods**). (B) Multivariate T2 tests comparing pre and post-NACT CD8 T cells, NK cells, and cytotoxic NES together. (C) Comparisons of TCR productive clonality (top), and percentage of productive T cells (bottom) between pre and post-NACT site-matched and site-unmatched samples. Paired and unpaired tests were used accordingly. TCR clonality is expressed as 1-entropy with values near 1 representing samples with one or a few predominant TCR rearrangements, while values near 0 represent more polyclonal samples (D) Least Absolute Shrinkage and Selection Operator (LASSO) regression analysis using the pre-NACT matched (n=8 samples) hallmark and Consensus^{TME} v.1.0 NES as explanatory variables, and the log_2 of the TCR clonality ratio of post/pre-NACT as response variable (**see Methods**). DF = Degrees of Freedom. The variables selected by the LASSO regression are indicated with an asterisk. (E) Spearman's rank correlation of pre-NACT cytotoxic NES and log_2 of the post/pre-NACT TCR clonality ratio.

Post-selection inference, taking into account for the uncertainty of the model selection and multiplicity, corroborated that the pre-NACT NES of the Consensus^{TME} cytotoxic gene

set is a promising variable to explain T cell clonal expansion upon NACT in these eight samples (p-value=0.096, see Methods). Overall, we were able to detect T cell activation in HGSOC site-matched but not in site-unmatched samples, further suggesting that different local immune-microenvironments play a role in the response to chemotherapy treatment. Also, pre-treatment samples with low T cell infiltration have a significant increase of TCR clonality upon chemotherapy, while tumours with higher infiltration levels do not have such a high clonal expansion, an observation that could not be addressed in unmatched tumour samples, due to the intra-patient tumour-immune heterogeneity. This highlights the potential confounding effects that intra-patient tumour-immune heterogeneity imposes on comparing tumours not only from different patients but also from different sites within the same patient.

3.6 Discussion

Despite advances in surgical approaches, chemotherapy and targeted therapies, the prognosis for patients with HGSOC remains poor, with the near-inevitable development of resistance to systemic therapy. Genetic and molecular analyses of asynchronous and disseminated tumours within patients have recently started to shed light on tumour clonal dynamics and evolutionary properties of different tumour types (Johnson et al., 2014; McPherson et al., 2016; Yates et al., 2015); however, the extent of TME heterogeneity in advanced HGSOC has only begun to be revealed (Jiménez-Sánchez et al., 2017; Zhang et al., 2018). We explored the main sources of variation in the transcriptomic space among treatment-naive samples and detected that transcriptomic pathway heterogeneity is mainly explained by presence or absence of immune and stromal cells. Importantly, the degree of immune signature variation within patients was similar to the extent we observed in a case study of metastatic HGSOC, where different tumour immune microenvironments were associated with clinical outcome. In the case study (**see Chapter 2**), tumours with high immune related pathways regressed and presented evidence of T cell activation, while immune excluded tumours progressed (**Figure 2.1**) (Jiménez-Sánchez et al., 2017).

In the present study, all patients presented at least one tumour with low immune infiltration, suggesting that HGSOC is characterised by microenvironmental niches, which could underlie primary and acquired resistance to therapies (Hirata et al., 2015; Sharma et al., 2017; Wang et al., 2016). The transcriptional, imaged-based and immunofluorescence analyses show that TME heterogeneity is an intrinsic feature of HGSOC, which spans across patients, tumours within patients and within tumours. We also integrated data from different estimation and regression methods, as well as cell type-specific gene signatures to generate a consensus approach (Consensus$^{\text{TME}}$) which consistently improved the predictions on our data sets and in TCGA ovarian cancer leukocyte methylation data (**see Chapter 4**). Application of Consensus$^{\text{TME}}$ to the treatment-naive samples showed that cytotoxic and NK cells are the major populations present in low purity tumours, while endothelial cells are the main TME cell type in high purity tumours. Pathway analysis of the differentially expressed genes showed that Wnt and Myc signalling pathways were more prevalent in purer tumours, consistent with emerging data in HGSOC and other tumours and models (Damsky et al., 2011; Gounari et al., 2002; Spranger et al., 2015, 2017, 2016; Sridharan et al., 2016). Furthermore, we found that intra-patient TME heterogeneity can mask the immune activation generated by treatment with cytotoxic chemotherapy. These analyses provide firm evidence that the TME affects the extent of immune activation generated upon treatment with chemotherapy.

3.7 Contributions

CRediT standard taxonomy: http://dictionary.casrai.org/Contributor_Roles

Conceptualisation

Ideas; formulation or evolution of overarching research goals and aims
AS, AJS, MLM, ES.

Data curation

Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where necessary for interpreting the data itself) for initial use and later re-use
AJS, KL, PC.

Formal analysis

Application of statistical, mathematical, computational, or other formal techniques to analyse or synthesize study data
AJS.

Funding acquisition

Acquisition of the financial support for the project leading to this publication
AV, ES, AS, MLM.

Investigation

Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection
PC, KL, TW, YM, IN, BW, DC, ES.

Methodology

Development or design of methodology; creation of models
AJS, ES.

Project administration

Management and coordination responsibility for the research activity planning and execution
AS, MLM, ES.

Resources

Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools
AS, MLM, ES.

Software

Programming, software development; designing computer programs, implementation of the computer code and supporting algorithms; testing of existing code components
AJS, ES (the radiologic background).

Supervision

Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team
AS, MLM, ES.

Validation

Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs
AJS, DCL.

Visualisation

Preparation, creation and/or presentation of the published work, specifically visualization/data presentation
AJS.

Writing original manuscript draft

Preparation, creation and/or presentation of the published work, specifically writing the initial draft
AJS, AS, MLM.

Writing review & editing

Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision

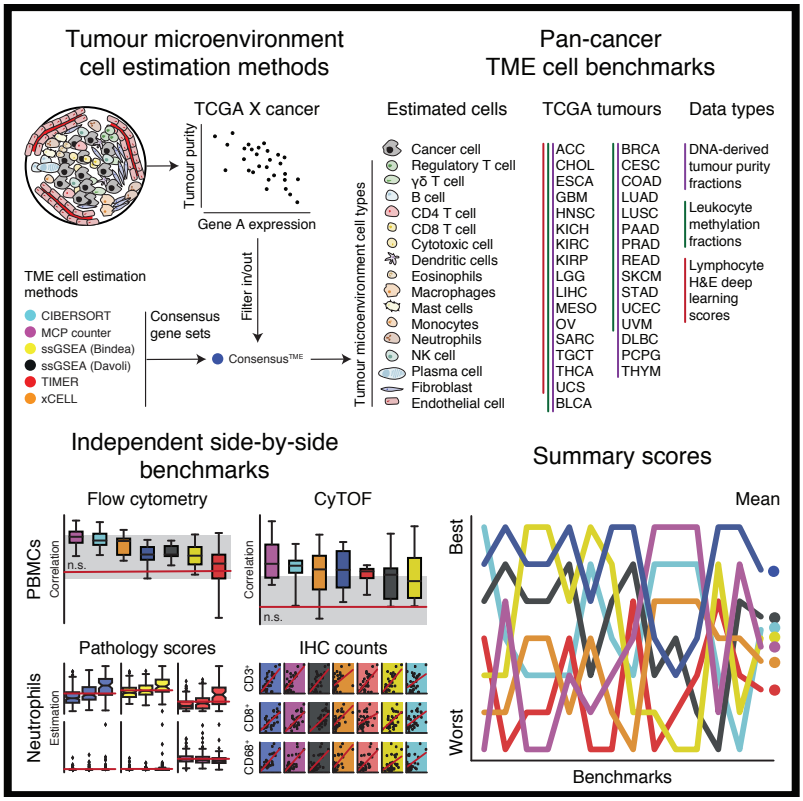
AJS, MLM, AS, KL, ES, JDB.

Chapter 4

Comprehensive benchmarking and integration of tumour microenvironment cell estimation methods

⊙ This work was performed in collaboration and as the day-to-day supervisor of Oliver Cast, a MSc rotation student in the group.

4.1 Graphical abstract



In brief

Pan-cancer and side-by-side benchmarks of six commonly used tumour microenvironment estimation methods and Consensus^{TME} integration approach.

Highlights

- Development of an enrichment-based method that integrates genes used by other methods in a cancer specific manner (Consensus^{TME}).
- Pan-cancer DNA, methylation, and H&E based benchmarks of tumour purity, leukocyte, and lymphocyte infiltration.
- Side-by-side independent methods benchmarks using PBMCs and tumour samples.
- Consensus^{TME} consistently outperforms independent methods in cancer-related benchmarks.

4.2 Summary

The tumour microenvironment comprises complex cellular compositions and interactions between cancer, immune, and stromal components which all play crucial roles in cancer. To estimate the relative abundance of different cell types in an unbiased manner using bulk tumour RNA data, various computational approaches have been developed during the last decade. However, a comparison that objectively evaluates the performance of these approaches against one another has not been conducted. Here we benchmarked six widely used tools and gene sets: Bindea et al. gene sets, Davoli et al. gene sets, CIBESORT, MCP-counter, TIMER, and xCELL. We also use Consensus^{TME}, a consensus approach that uses the union of genes that the six tools used for cell estimation, and corrects for tumour-type specificity. We benchmarked the seven tools using TCGA DNA-derived purity scores (33 tumour-types), methylation-derived leukocyte scores (30 tumour-types), and H&E deep learning derived lymphocyte counts (13 tumour-types), and individual benchmark data sets (PBMCs and 2 tumour-types). None of the seven tools outperformed the others in every single benchmark, and Consensus^{TME} was overall the top performing method in all cancer-related benchmarks. Computational methods that provide robust and accurate estimates of non-cancerous cell populations in the tumour microenvironment from tumour bulk expression data are important tools that can advance our understanding of tumour, immune, and stroma interactions, as well as potential clinical application if high accuracy estimates are achieved.

4.3 Introduction

The tumour microenvironment (TME) plays an active role in in tumour initiation, progression, metastasis, and treatment response. Thus, studying the TME is a central paradigm of cancer research. However, a great variety of stromal and immune cell types populate tumour tissues, and the role of many of these cell types is still unclear or their effect in different tumour-types is largely unknown. Traditionally, cells from the TME have been quantified using immunohistochemistry (IHC), immunofluorescence (IF), and flow cytometry, and more recently using time of flight cytometry (CyTOF). These methods, although accurate, are laborious, low throughput and require pre-selected cellular markers and therefore their application in large number of samples and measurements is challenging. Thus, their systematic application for comprehensively investigating the various different cell types in the TME in an unbiased manner is limited.

Estimation of non-cancerous cell proportions from bulk tumour samples has been performed using whole-exome sequencing, RNA-seq, or DNA methylation data. During the last decade, multiple computational approaches have been developed intending to calculate quantitatively or semi-quantitatively distinct TME cell-type population estimates (Finotello and Trajanoski, 2018). A variety of statistical frameworks and algorithmic procedures have been employed, and each method has used different benchmark data sets (Finotello and Trajanoski, 2018). In general, two different algorithmic classes into which most methods can be classified are deconvolution algorithms and gene set enrichment-based methods. Importantly, both classes rely on cell-type specific markers that are selected according to prior knowledge. The deconvolution algorithms use linear combinations of the expression values of the cell-specific genes, while gene set enrichment-based methods rank the genes of a mixture sample and compute enrichment scores as a function of the ranked selected genes. A problem, however, is that each method has claimed to outperform others, and debates about which statistical frameworks are better suited have arisen (Li et al., 2017a; Newman et al., 2017). Thus, the need for independent and more comprehensive benchmarks has been pointed out (Zheng, 2017). Additionally, with the diversity of methods developed, cell-type marker specific genes can be curated and consensus gene sets generated. Here, we developed a consensus approach (Consensus$^{\text{TME}}$) for the selection of cell-type specific gene sets after overlapping six widely used TME cell estimation methods. We performed pan-cancer benchmarks using publicly available bulk RNA sequencing data from TCGA and side-by-side comparisons using the methods' independent benchmarks. Finally, since the Consensus$^{\text{TME}}$ approach ensembles genes selected by different methods for different cell types, this makes Consensus$^{\text{TME}}$ an evolvable method by design.

4.4 Methods

Quantification and statistical analysis

Further information and requests for resources and reagents should be directed to and will be fulfilled by Alejandro Jiménez-Sánchez (alejandro.jimenezsanchez@cruk.cam.ac.uk) and Dr Martin L. Miller (martin.miller@cruk.cam.ac.uk).

Quantification and Statistical Analysis

Single-sample gene set enrichment analysis ⊙

Single-sample gene set enrichment analysis (Barbie et al., 2009), a modification of standard GSEA (Subramanian et al., 2005), was performed on RNA measurements for each sample using the GSVA package version 1.28.0 (Hänzelmann et al., 2013) in R version 3.5.0 with parameters: method = 'ssgsea', and tau = 0.25. Normalized enrichment scores were generated for the hallmark gene sets (Liberzon et al., 2015), immune and stromal signatures (Yoshihara et al., 2013), TME cell gene sets obtained from previous publications (Bindea et al., 2013; Davoli et al., 2017), as well as the consensus TME gene sets (**Figure 8.3A**). Hallmark gene sets were obtained from MSigDB database version 6.1 (Liberzon et al., 2015).

Consensus^{TME} cell gene sets ⊙

To generate the Consensus^{TME} gene sets we identified cell types that were deconvoluted by at least 2 different methods, leading to 18 different cell types. We then intersected the gene sets that the different methods considered for the deconvolution of such cell types. To intersect genes used in CIBERSORT, we first filtered out genes whose weight was below 1.96 standard deviations of the mean for each of CIBERSORT cell types. In addition we combined activated states into the corresponding cell type. The only activated stated included was cytotoxic cells, which would include CD8 and/or NK cells in their activated stated. The intersected genes were used to represent each cell type, and genes with a higher Pearson's correlation coefficient than -0.2 and a p-value ≤ 0.05 with tumour purity as defined by TIMER were filtered out from the gene sets (Li et al., 2016). Finally, ssGSEA was employed to calculate NES for each cell type as described above.

Comparison statistical metrics ⊙

The different estimation tools were compared against each other using either Kendall's rank correlation coefficient or the multiple linear regression goodness of fit metrics: adjusted

R-squared, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). AIC and BIC z-scores values were calculated to incorporate different tumour types in the comparisons since AIC and BIC values are unit-less. Differences between groups of variables were identified using one-way ANOVA with Tukey honest significant differences post-hoc tests. All statistical tests were adjusted for multiple testing using the Benjamini-Hochberg procedure to control for false discovery rate (FDR).

TCGA immune estimations ⊙

RNA-seq data was collected from cBioPortal (Cerami et al., 2012). Batch normalisation had been applied and gene expression values calculated using the "RSEM" pipeline (Li and Dewey, 2011). Four TME cell estimation methods and two published gene sets were used to produce relative abundances of immune cell types across 33 tumour types. For each method, a general immune score was also derived if it was not already provided, representing the total level of immune cell infiltration in each tumour sample.

TME cell deconvolution methods ⊙

Cell deconvolution methods were used to estimate levels of non-cancerous cells in the TME. The methods employed were CIBERSORT (Newman et al., 2015), MCP-counter (Becht et al., 2016), TIMER (Li et al., 2016), xCELL (Aran et al., 2017), as well as gene sets collected from two previous publications (Bindea et al., 2013; Davoli et al., 2017).

Bindea et al. and Davoli et al. gene sets ⊙

Gene sets provided by Bindea et al. and Davoli et al. were used with ssGSEA to provide enrichment scores for each of the immune signatures. To generate general immune scores, genes selected for immune cells where combined into one gene set for each method independently.

xCELL ⊙

The "xCell" R package (version 1.12) was used to generate immune estimates for the xCell method (Aran et al., 2017). A general immune estimation score is already provided by xCELL.

MCP-counter ⊙

Estimations for the MCP-counter method were produced using the "MCPcounter" R package (version 1.1.0) (Becht et al., 2016). Immune scores for this method were produced in a similar manner as the ssGSEA methods by creating a union of signature genes for each of the cell types. The "MCPcounter.estimate" function was used to allow for the new signature.

CIBERSORT ⊙

CIBERSORT estimations were produced using the R source code, provided on request from the web resource (Newman et al., 2015). CIBERSORT was run in "Absolute mode" (under beta development) using 100 permutations and quantile normalisation disabled as recommended for RNA-seq data. Absolute scores representing the "overall immune content" is produced natively by the algorithm in absolute mode.

TIMER ⊙

TIMER estimations were produced using R source code, available from the web resource (Li et al., 2016). Immune scores for TIMER were produced as a sum of the coefficients for each cell type.

Consensus^{TME} ⊙

Consensus^{TME} selects cell types that are estimated by at least two methods and generates a gene set for that cell type. The genes that other methods have considered as relevant for that cell type are incorporated into the gene set. Genes whose gene expression does not correlate negatively (Pearson's correlation coefficient lower than -0.2) with tumour purity, in a tumour specific manner using TCGA data, are filtered-out. Thus, each gene set represents the consensus genes for a given cell type in a particular tumour type. General immune scores for each tumour types were generated by combining the genes of the different immune cells into one gene set.

Purity score benchmark ⊙

Pan-cancer purity scores were downloaded from the NIH Genomic Data Commons (Hoadley et al., 2018). Purity scores were generated using ABSOLUTE (Carter et al., 2012) which uses copy number, variant allele frequency, and tumour specific karyotype data to calculate the cancer fraction of a tumour samples. To benchmark the immune estimation methodologies using purity of samples the immune scores were added to an independent stromal

score; calculated through the use of ESTIMATE (version 1.0.13) (Yoshihara et al., 2013). ABSOLUTE's derived tumour purity and the different methods tumour purity scores were correlated independently for each tumour type.

Leukocyte methylation benchmark ⊙

Leukocyte methylation scores were downloaded from the NIH Genomic Data Commons https://gdc.cancer.gov/about-data/publications/PanCan-CellOfOrigin (Hoadley et al., 2018). Leukocyte methylation and CD8+ T cell estimations were correlated, since CD8+ T cells are estimated by all the methods. Multiple linear regression was then employed using all leukocytes the methods estimate as explanatory variables and the leukocyte methylation scores as response variable. Leukocyte methylation data was log transformed to meet the normality and heteroscedasticity assumptions of the model. Adjusted R squared, AIC, and BIC metrics were calculated to compare the goodness of fit between the methods considering the number of variables included in the model.

Somatic single nucleotide mutation data ⊙

Somatic single nucleotide mutation data was downloaded from the Broad Institute GDAC Firehose.

H&E deep learning lymphocyte fractions benchmark ⊙

Lymphocyte fractions were generated by Saltz et al. for 13 TCGA cancer types. Multiple linear regression was applied in a similar manner as for the leukocyte methylation analysis, instead using a hyperbolic sine transformation of lymphocyte fraction as a response variable to meet normality and heteroscedasticity assumptions of the model. Models for each method were fitted using only lymphocytes as explanatory variables.

Independent methods side-by-side benchmarks ⊙

The benchmarking validation experiments for each of the methods were replicated, where possible, to match the parameters used in the original publications. Of the six methods there were four benchmarking datasets available; either online or provided by the authors. Each of the datasets contained samples with bulk gene expression values along with matched "ground truth" values. The CIBERSORT benchmarking dataset, provided by the authors on request (Newman et al., 2015), consisted of flow cytometry values of different immune cell types from PBMC samples. The xCell benchmarking datasets, SDY311 and SDY420,

were publicly available for download from ImmPort (Bhattacharya et al., 2014), and the validation data consisted of matching CyTOF quantification of immune cells from PBMC samples. The MCP-counter publication used gene expression profiles from GEO (accession number GSE39582) and IHC counts of $CD3^+$, $CD8^+$, and $CD68^+$ cells (available on request from the authors) (Becht et al., 2016). TIMER benchmark consisted of H&E stained slides from TCGA Adrenocortical Carcinoma (BLCA) study. Pathological estimations of these slides were carried out to categorise each sample into one of three categorical levels for neutrophil abundance: "Low", "Medium" or "High"; estimations are available from the TIMER online resource (Li et al., 2016). For all benchmarking experiments, except TIMER, concordance was measured using correlation between "ground truth" values and the immune estimations of each method. Due to the variation in the degree of specificity to which cell subsets were defined, summations of subsets was required to allow accurate comparisons in some cases. For the TIMER benchmark, the in-silico neutrophil estimations for each method were grouped by low, medium and high pathological estimation, then compared through ANOVA with Tukey post hoc.

Data and Software Availability

Further information and requests for resources and reagents should be directed to and will be fulfilled by Dr Martin L. Miller (martin.miller@cruk.cam.ac.uk).

4.5 Results

4.5.1 Consensus tumour-type specific superset of tumour microenvironment cell populations

Following the generation of large data sets of tumour genomic profiles, various computational tools assessing TME cell populations have been generated, each using different algorithms, gene markers, and validation benchmarks. To build on the knowledge of cell-type specific gene sets represented in the diversity of these methods, we sought an integrative strategy that incorporates cell types and genes from the collection of independent tools. Consensus^{TME} integrates cell-type specific gene markers used by independent cell estimation methods, then we employed single sample gene set enrichment analysis (ssGSEA) to compute TME cell-type and tumour specific enrichment scores from bulk expression data (**Figure 4.1A** and **Figure 8.3A**). The ssGSEA approach was selected because its treats microarray and RNA-seq values in the same way, since it is based on the ranked genes rather than the actual values. To generate Consensus^{TME}, we selected six widely used cell estimation methods that use different statistical and algorithmic frameworks, estimate different cell types, and use different genes for their estimation. The selected methods were CIBERSORT (Newman et al., 2015), TIMER (Li et al., 2016), MCP-counter (Becht et al., 2016), xCELL (Aran et al., 2017), and the gene sets generated and use in (Bindea et al., 2013; Şenbabaoğlu et al., 2016) and in (Davoli et al., 2017) here called "Bindea" and "Davoli" supersets, respectively. In brief, we first selected cells that are estimated by at least two methods. Second, we generated a gene set for each cell type by using the union of genes used by the methods to estimate that cell type, and remove genes that correlate (ρ > -0.2) with tumour purity as done before (Li et al., 2016) (**see Methods**). Therefore, Consensus^{TME} aggregates cell-type specific genes that have been independently considered relevant by different methods, and estimates their abundance in a tumour type specific manner.

4.5.2 Pan-cancer stroma, leukocyte, and lymphocyte benchmarks

To benchmark the different methods in an objective and systematic manner, we used publicly available data from between 13 and 32 tumour types comprising 9,142 tumour samples in total. First, we correlated DNA-based tumour purity scores (Carter et al., 2012; Hoadley et al., 2018) with purity scores generated by the different methods, or customly derived when not generated by default (**see Methods**). Since tumour purity does not account only for immune cell infiltration but also for other stromal cells (e.g. fibroblasts and endothelial cells) (Binnewies et al., 2018), this would affect the correlation of methods that only

estimate immune cells. Therefore, we inferred stromal non-immune related content of all samples using ESTIMATE (Yoshihara et al., 2013) and added this value to all methods' purity scores. We found that all six methods and Consensus^{TME} perform very similar to each other, with CIBERSORT, Consensus^{TME}, and Davoli as the top 3 pan-cancer negative correlations (**Figure 4.1**B). Across tumour types, the different methods performed similar and very few correlations were not statistically significant. As expected, tumour purity and leukocyte fraction show a negative correlation across tumours, however neither mutation load nor leukocyte fraction associated with the purity correlations calculated. Thus, the ability to estimate tumour purity using expression profiles is largely independent from the cancer cellularity and mutation load in the sample for any of the methods.

We further evaluated the performance of the methods by using leukocyte fractions derived from methylation data for 30 tumour types (Hoadley et al., 2018). For this analysis we fitted multiple linear regression models using only the scores of leukocyte infiltration from each method as explanatory variables and the leukocyte fraction as response variable. When comparing the adjusted coefficients of determination (R^2) of the different models, the best performing methods were Bindea, Consensus^{TME}, and Davoli (**Figure 4.1C**). Since different methods estimate different number of leukocytes, the coefficient of determination can be artificially increased by the number of variables in a model (i.e. overfitting). Thus, to more appropriately compare the models we used the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) for model selection, which penalise model complexity (i.e. number of leukocytes used in the models) less or more heavily, respectively (**Figure 4.1D**). The AIC and BIC scores showed that Consensus^{TME}, Davoli, and Bindea models perform significantly better than the other methods (models with lower AIC and BIC values are preferred). We also implemented multiple linear regression analysis using tumour-infiltrating lymphocyte counts derived from digitised H&E-stained images analysed through a deep-learning convolutional neural network approach (Saltz et al., 2018). Although low coefficient of determination values were obtained across methods, likely due to the very difficult task of computationally discriminate leukocytes on H&E images, the methods that obtained a higher coefficient of determination were Bindea, CIBERSORT, and Consensus^{TME}, while the lowest AIC and BIC values were obtained by Consensus^{TME}, Davoli, and TIMER (**Figure 4.1E**). Together, these broad pan-cancer benchmarks show a large variation in the performance of the different methods when compared to each other, and no single method consistently outperforms the others.

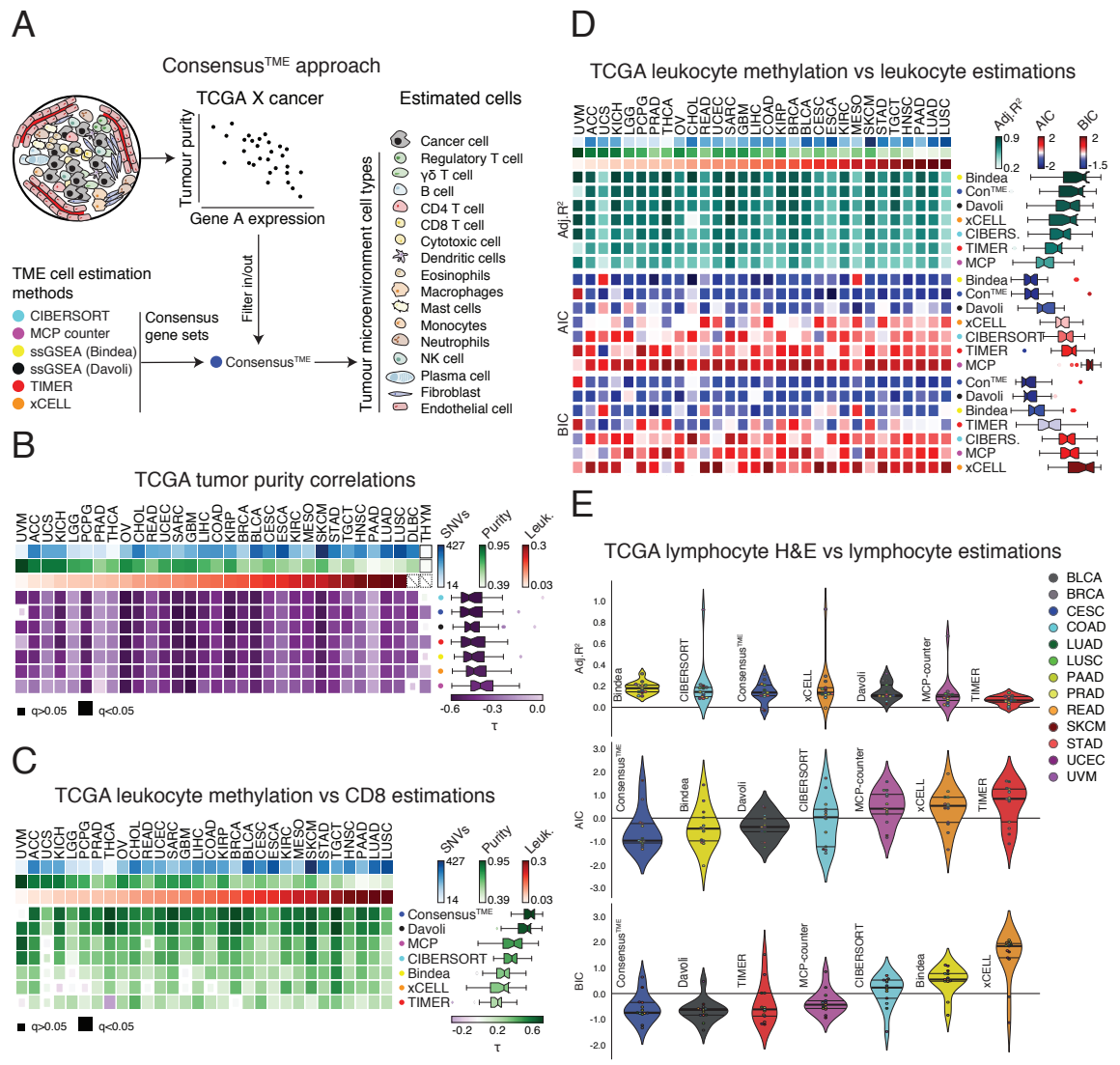


Fig. 4.1 Benchmark of methods for estimating TME cell components using purity and leukocyte data from TCGA data. (A) Bioinformatic tools benchmarked, and Consensus^{TME} development strategy (**see Methods**). (B) Kendall's correlation coefficients (τ) of DNA-derived ABSOLUTE purity scores (Carter et al., 2012; Hoadley et al., 2018) and RNA-derived estimated purity by the different methods. ESTIMATE's stromal scores were added to all the methods to account for stromal quantities, since not all methods analyse stromal cells. (C) Kendall's correlation coefficients (τ) of methylation-derived leukocytes' scores (Hoadley et al., 2018) and RNA-derived $CD8^+$ T cell estimations. (D) Multiple linear regression models of leukocyte methylation scores as response variable (Hoadley et al., 2018) and RNA-derived leukocyte estimations as explanatory variables. Column heatmaps are sorted according to leukocyte methylation scores (Left: Low, Right: High), rows are sorted according to median performance (Top: Best, Bottom: Worst). (E) Multiple linear regression models of deep learning H&E-derived lymphocyte counts as response variable (Saltz et al., 2018) and RNA-derived lymphocyte estimations as explanatory variables. Kendall's correlation significance is shown with small/large squares as indicated by the q-value multiple test corrected. Adjusted R2, AIC z-score, and BIC z-score were compared across models generated by each tool cellular estimations. Lower AIC and BIC values represent a better goodness-of-fit penalising the number of variables.

Continuation of Figure 4.1 caption ...
Median single nucleotide variants (SNVs), ABSOLUTE purity scores (purity), and leukocyte methylation scores (Leuk.) per tumour type are shown in B, C, and D. Violin plots are sorted according to median correlation coefficient (left: best, right: worst). ACC n=77, BLCA n=397, BRCA n=1052, CESC n=293, CHOL n=36, COAD n=281, DLBC n=47, ESCA n=162, GBM n=154, HNSC n=509, KICH n=66, KIRC n=498, KIRP n=285, LGG n=519, LIHC n=359, LUAD n=504, LUSC n=493, MESO n=81, OV n=293, PAAD n=159, PCPG n=165, PRAD n=473, READ n=92, SARC n=246, SKCM n=460, STAD n=403, TGCT n=155, THCA n=469, THYM n=103, UCEC n=175, UCS n=56, UVM n=80.

4.5.3 Independent methods benchmarks

All methods tested, except Bindea and Davoli gene supersets, performed their own independent benchmarks in the original publications. Thus, we collected benchmark data for CIBERSORT, xCELL, TIMER, and MCP-counter to carry out a side-by-side comparisons. We used the CIBERSORT benchmark data that consisted of peripheral blood mononuclear cell (PBMC) of 27 human subjects and quantified by flow cytometry (Newman et al., 2015). Correlations between estimated immune cell types and the flow cytometry fractions showed that the best performing methods were MCP-counter, CIBERSORT, and xCELL (**Figure 4.2A**). However, most of the correlations lacked statistical significance, and due to the different cell types estimated by the different methods it is difficult to reach a conclusion. Similarly, the xCELL benchmark data set consisted of 16 PBMC leukocyte subsets from two different studies with 61 and 104 human subjects each, where PBMCs were measured using Time of Flight Cytometry (CyTOF) (Aran et al., 2017). Again, MCP-counter, CIBERSORT, and xCELL were the methods that showed best performance in these PBMC benchmarks, however many cell types did not reach statistical significance (**Figure 4.2B**).

Finally, we used cancer-related benchmarks from TIMER (Li et al., 2016) and MCP-counter (Becht et al., 2016). For TIMER's benchmark, 404 TCGA bladder cancer samples were analysed by a pathologist who categorised them as low, medium, or high according to their neutrophil counts using H&E stained slides. Here, Consensus$^{\text{TME}}$, Bindea, and TIMER obtained the best separation between categories, but only Consensus$^{\text{TME}}$ and Bindea separated significantly the three categories after multiple test correction (**Figure 4.2C**). Interestingly, xCELL, CIBERSORT, and MCP-counter were unable to differentiate the three categories, while Davoli does not estimate neutrophils. The MCP-counter benchmark consisted of IHC digital quantification of $CD3^+$ (T cells), $CD8^+$ ($CD8^+$ T cells), and $CD68^+$ (Monocytic lineage) cell densities from 38 colorectal cancer samples. Correlations between the methods' estimations and the cellular fractions were computed (**Figure 4.2D**). Consensus$^{\text{TME}}$, MCP-counter, and Davoli methods provided the best correlations, with Consensus$^{\text{TME}}$ out-performing on the three cell types.

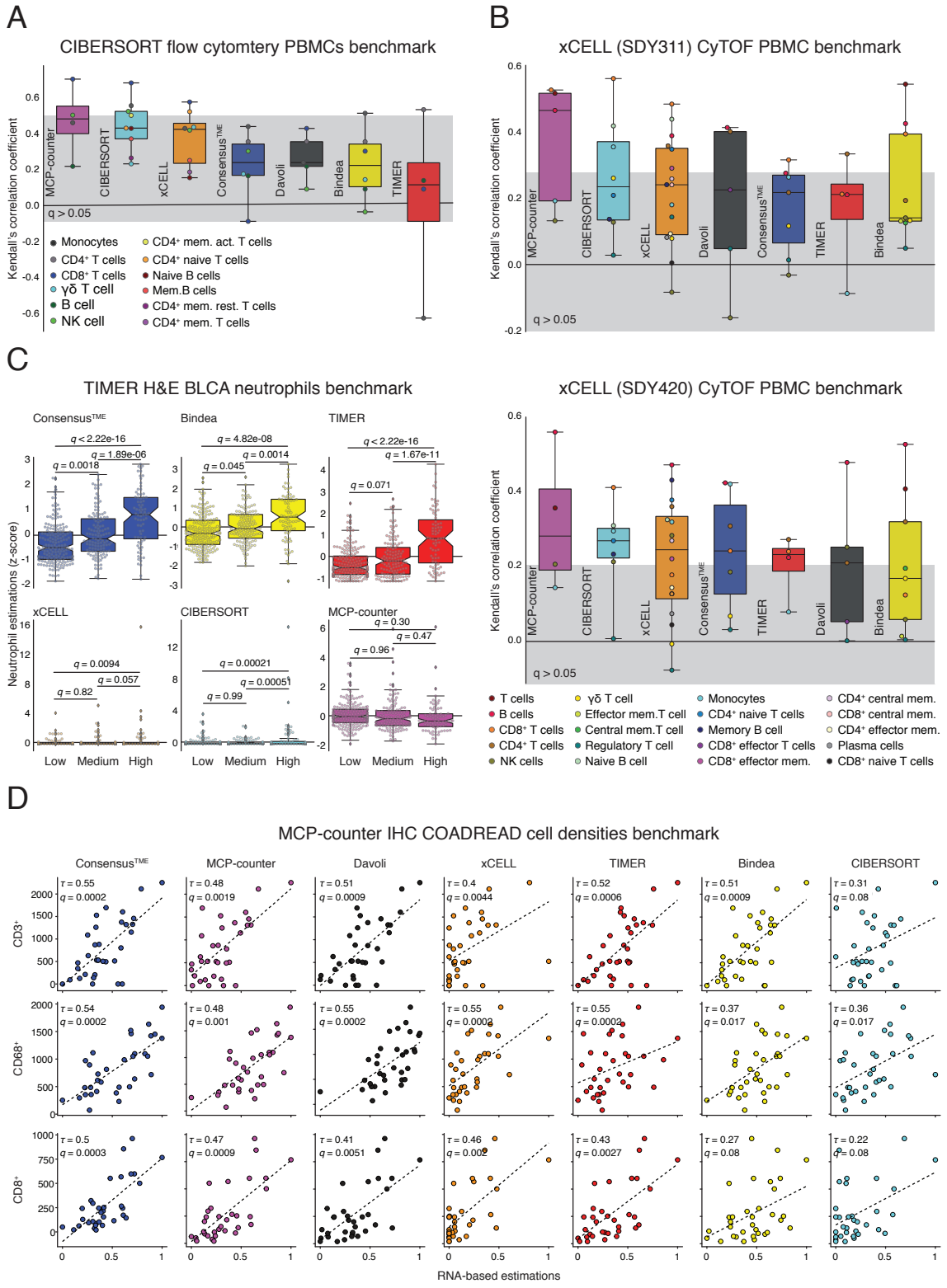


Fig. 4.2 Side-by-side benchmark using available datasets published by the individual methods. (A) Kendall's correlation coefficients (τ) of CIBERSORT PBMCs' flow cytometry (n=20 samples) and (B) xCELL PBMCs' CyTOF benchmarks. SDY311 (n=61 samples) and SDY420 (n=104 samples) are different benchmarks available from ImmPort (Bhattacharya et al., 2014). Since not all methods estimate cell activation states, cell proportions were aggregated into the cell type and use for correlation with the methods that do not estimate activation states. The grey box represents correlation coefficients that have a q-value > 0.05 (BH method). Box plots are sorted according to median correlation coefficient (Left: Best, Right: Worst).

Continuation of Figure 4.2 caption . . .
(C) Comparison between low, medium, and high categories of BLCA (n=404 samples) neutrophil haematoxylin and eosin pathology counts. One-way ANOVA with Tukey HSD post hoc tests were employed to calculate q-values. Plots are sorted according to performance (Left-Top: Best, Right-Bottom: Worst). (D) Kendall's correlation coefficients (τ) of MCP-counter colon adenocarcinoma (COAD) IHC (n=38 samples) cell densities (cellmm^2). Plots are sorted according to median correlation coefficient (Left: Best, Right: Worst).

4.6 Discussion

With the recent generation of large publicly available molecular profiling of cancer samples, a variety of computational tools for analysis of cell components of the TME have been generated. In principle, the method of choice should be based on performance, however popularity and ease of use are often reasons behind the method researchers select (Zheng, 2017). In the case of TME cell estimation from bulk expression data, this problem is magnified by the lack of objective and independent benchmark analyses, since most methods use their own benchmarks which may introduce biases and reliance on one type of data. The tools that have been developed so far for the estimation of TME cell types using tumour bulk expression data fall under two main statistical frameworks: enrichment analysis and algorithmic deconvolution (Finotello and Trajanoski, 2018). Here we performed an independent and objective benchmarking exercise comparing six of the most widely used and recent tools, and also developed Consensus^{TME}: a gene set enrichment based method that integrates cells and genes from these six different tools in order to generate a consensus gene superset that is tumour-type specific.

We performed pan-cancer benchmarks using orthogonal data types generated for TCGA samples. While DNA-derived tumour purity scores correlated negatively and similarly with RNA-derived TME estimations in all methods, leukocyte methylation scores showed some discrepancy across methods, and some lack of correlation in some cases. Also, different tumour types showed different levels of correlation with leukocyte estimations, which could be due to the leukocyte methylation signature model, which was generated by comparing pure leukocyte cells and normal tissue methylation patterns with tumour-type specific methylation patterns (Hoadley et al., 2018), and tumour infiltrating leukocytes may have different methylation patterns. Furthermore, lymphocyte deep learning H&E quantifications provided a lower association with lymphocyte RNA-derived estimations, an observation that has been reported before and considered to be in part due to the RNA-derived estimates reflect more cell counts, while spatial image-derived estimates reflect the fraction of lymphocytes per area (Saltz et al., 2018; Thorsson et al., 2018). Thus, this benchmark is inconclusive due to the uncertainty of both RNA-derived and imaged-based derived lymphocyte estimates, but it was included to achieve more comprehensive and orthogonal benchmarks. Moreover, for the application of TME cell estimation using bulk RNA tumour data, PBMC benchmarks may not be very informative as these are circulating and not necessarily infiltrating immune cells. In contrast, both BLCA and COADREAD benchmarks on neutrophils, $CD3^+$, $CD68^+$, and $CD8^+$ cells showed significant associations for some tools, particularly Consensus^{TME}. These benchmarks showed that no independent method is consistently outperforming other methods. Nevertheless, overall Consensus^{TME} ranked among the top three best performing methods in all cancer-relates benchmarks. Lastly, Consensus^{TME} is an evolvable method by conception, which means that other genes used by new methods can be added to the gene sets and tested with the already defined benchmarks, thus potentially improving its performance as different and new methods are developed.

4.7 Contributions

CRediT standard taxonomy: http://dictionary.casrai.org/Contributor_Roles

Conceptualisation

Ideas; formulation or evolution of overarching research goals and aims
AJS, MM

Data curation

Management activities to annotate (produce metadata), scrub data and maintain research data (including software code, where necessary for interpreting the data itself) for initial use and later re-use
AJS, OC

Formal analysis

Application of statistical, mathematical, computational, or other formal techniques to analyse or synthesize study data
OC, AJS

Funding acquisition

Acquisition of the financial support for the project leading to this publication
MM

Investigation

Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection
OC

Methodology

Development or design of methodology; creation of models
AJS, OC

Project administration

Management and coordination responsibility for the research activity planning and execution
AJS, MM

Resources

Provision of study materials, reagents, materials, patients, laboratory samples, animals, instrumentation, computing resources, or other analysis tools
MM

Software

Programming, software development; designing computer programs, implementation of the computer code and supporting algorithms; testing of existing code components
OC, AJS

Supervision

Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team
AJS, MM

Validation

Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs
OC, AJS

Visualisation

Preparation, creation and/or presentation of the published work, specifically visualization/data presentation
OC, AJS

Writing original manuscript draft

Preparation, creation and/or presentation of the published work, specifically writing the initial draft
AJS

Writing review & editing

Preparation, creation and/or presentation of the published work by those from the original research group, specifically critical review, commentary or revision
AJS, MM, OC

Chapter 5

Discussion

Genetic and molecular analyses of asynchronous and disseminated tumours within patients have recently started to shed light on tumour clonal dynamics and evolutionary properties of different tumour types (Johnson et al., 2014; McPherson et al., 2016; Yates et al., 2015). However, the extent of TME heterogeneity and the degree to which TME may be driven by stochastic, cellular or molecular processes is not well understood. Specifically in HGSOC, this has great relevance due to the known intra-tumour genetic heterogeneity and the lack of therapeutic success in advanced stages of the disease. Particular findings of the studies presented in this thesis may have important clinical implications if they are corroborated in larger cohorts. First, in the patient case study and then confirmed in the treatment-naive cohort, we observed coexistence of distinct TME with different immune infiltration levels across metastases within single patients. We sought factors behind the intra-patient TME heterogeneity, as these could serve as potential therapeutic targets, and also to understand fundamental properties of TME dynamics. Second, we evaluated the effect of NACT chemotherapy on the TME of HGSOC and found that intra-patient TME heterogeneity can be a confounder and has to accounted for to achieve reliable interpretations. Third, we benchmarked state-of-the-art TME cell estimation methods that use bulk tumour expression data as input, and generated a consensus method that, unlike the other methods, provides consistent high accurate estimations in pan-cancer benchmarks and individual experimental quantifications of cells in the TME. Together, these results provide a framework for the study of the TME dynamics in cancer using bulk tumour derived data and bioinformatics approaches.

5.1 HGSOC case study

In the case study, where WES data was generated, we sought differences and similarities in somatic mutations between progressing and regressing tumours. Since the patient of the case study had received multiple chemotherapeutic treatments, it could be possible that progressing tumours had acquired mutations that provided resistance to therapy. However, no specific mutations in progressing tumours that could explain the observed dichotomous progression/regression status were found. Since regressing tumours had a higher immune infiltration, we asked whether mutation or neoepitope load, or specific neoepitopes were responsible for the differences observed in immune infiltration. We found that the mutation and predicted neoepitope loads alone could not explain the different regressing/progressing behaviour of the metastatic samples. Although, a single case, it is known that HGSOC has a low mutation rate, thus the mutation load may play a less preponderant role in this disease in respect to immune evasion compared to other tumour types. In addition, since the patient of the case study did not receive immunotherapy, in principle, the selective pressure induced by the immune system onto the mutation and neoepitope landscape would be lower than in patients that have received immunotherapy treatments. Indeed, recent studies looking for mechanisms of immunotherapy resistance, have detected mutations in the antigen presentation machinery (Giannakis et al., 2016), β2-micro-globulin (Zaretsky et al., 2016), the IFN-γ pathway (Benci et al., 2016; Zaretsky et al., 2016), or HLA-I genes (Tran et al., 2016) in resistant tumours. In pan-cancer studies, using thousands of untreated primary tumours, only few mutations in the antigen presentation machinery (Rooney et al., 2015; Shukla et al., 2015), β2-micro-globulin (Challa-Malladi et al., 2011; Rooney et al., 2015), and HLA-I genes (Rooney et al., 2015; Shukla et al., 2015) were detected. Although, we cannot exclude that mutations in these pathways may occur in patients with HGSOC, in this case study we did not find mutations that could explain the dichotomous progression pattern of metastatic lesions. This result, however, opened up other avenues of possible causes behind the differential progression/regression and immune infiltrations observed. Indeed, instead of specific mutations or subclonal neoepitopes present in the regressing samples of the case study, TCR clonal expansion and T cell reactivity against clonal neoepitopes (i.e. present in all tumours) were detected. Strikingly, through an epitope T cell challenge experiment, we found that all neoepitopes that elicited a $CD8^+$ T cell response had higher mutant to wild-type HLA-I predicted binding affinity. The lack of tumour-specific somatic alterations in the regressing and stable tumours alone suggests that other factors in the tumour microenvironment may have been playing a critical role in the immune response and overall fate of the tumours.

Transcriptional signatures can be highly informative as diagnostic/prognostic resources, as well as provide insights on mechanistic underpinnings (Burel and Peters, 2018). Thus, we analysed the transcriptional profiles of the progressing and regressing tumours of the case study. We found that *STAT1* and *CXCL9* were highly expressed in the regressing metastases. *CXCL9* is well known as a potent T cell chemokine (Liao et al., 1995; Rainczuk et al., 2012), and high expression of *CXCL9* and *CXCL10* correlate with enhanced T cell infiltration of tumours and better survival of ovarian

cancer patients (Bronger et al., 2016). In contrast, when looking at gene expression signatures, the growing vaginal cuff metastasis had a higher ES in the Wnt pathway, which has been implicated as a mechanism that impairs recruitment of dendritic cells and prevents T cell infiltration in autochthonous mouse melanoma models via a CXCL9- and CXCL10-dependent mechanism (Spranger et al., 2015, 2017). Although a direct link between tumour fate and the observations found in this patient could not be proved with the available samples, this case emphasises the importance of an integrative approach to understand the molecular mechanisms governing the interaction between the tumour and its immune microenvironment (Miao and Van Allen, 2016), and points to the importance of the specific TME state, in addition to specific epitopes influencing the immune response.

As in any case study, the present work has notable limitations. It involves only one patient, and thus further studies are needed to determine whether the principles discovered here apply to other patients. Furthermore, the interplay between treatment, somatic mutations, the immune system, and heterogeneous fates of the tumours could not be untangled in this clinical case. For example, it is feasible that the multiple chemotherapy interventions for this patient contributed to shape the somatic mutations and the microenvironment of the tumours, but due to the availability of samples and descriptive nature of the study this could not be explored further. Despite such limitations, this case provides evidence for differential tumour-immune responses existing in metastases of the same individual, related not only to genetic alterations but also to the tumour-immune microenvironment, which to our knowledge has not yet been demonstrated in patients with ovarian cancer. Also, most studies on the tumour-immune microenvironment have been conducted in primary tumours (Teng et al., 2015), with the exception of a study of matched primary and metastatic tumours, which concluded that the immune con-texture globally recapitulates that of the primary (Remark et al., 2013). In contrast, the case of recurrent HGSOC presented here, clearly shows the opposite: that tumour-immune microenvironments, between primary tumour and metastases, and between metastases, can be heterogeneous within a patient. Studying the intersection between ITH and TME heterogeneity could be a path to understand the reason behind these conflicting observations.

Previous genomic and immune profiling of multiple lesions in patients have also shed light on tumour heterogeneity and its implications on tumour evolution (Gerlinger et al., 2012), disease progression (Ascierto et al., 2017), and immune control (McGranahan et al., 2016; Sridharan et al., 2016; Şenbabaoğlu et al., 2016). For example, tumours that are genetically more heterogeneous have less immune infiltrates (Şenbabaoğlu et al., 2016) and less benefit from checkpoint-blockade immunotherapies (McGranahan et al., 2016). It has been shown that T cell infiltration and gene expression of immune-related genes correlate with response to checkpoint-blockade immunotherapy in melanoma (Chen et al., 2016). Additionally, analyses of synchronous resected metastases with differential progression in patients with melanoma has shown that intra-patient metastases present not only genetic heterogeneity but also immune-infiltration heterogeneity of immune cell types and T cell clonality between samples (Reuben et al., 2017). A rapid autopsy study of a patient with

metastatic melanoma treated with anti-PD-1 therapy showed that resistant metastases over-expressed genes related to extracellular matrix and neutrophil function (Ascierto et al., 2017). Interestingly, association between Wnt signalling and lack of T cell infiltration was also observed in a patient with adenoid cystic carcinoma where serial biopsies from the same patient were analysed, and different expression profiles between primary and metastatic deposits were also detected (Sridharan et al., 2016). Finally, a plethora of molecular mechanisms and types of cells influencing the tumour-immune microenvironment have been described in different tumour types, leading to important advances in immunotherapy (Joyce and Fearon, 2015; Melero et al., 2014; Sharma et al., 2017). Unfortunately, the promise of immunotherapy has not been as successful in ovarian cancer as it has been in other tumour types (Homicsko et al., 2016) despite the fact that it has been recognised more than a decade ago that T cell infiltration is a key element for patient outcome in this disease (Zhang et al., 2003). We believe that the growing evidence of differential genomic, transcriptomic, and immune profiles between and within patients will eventually provide new key elements to target in ovarian cancer and other tumour types. However, this task will require extensive and deep systematic analyses along with longitudinal data, as the differences between metastases and coexistence of tumour-immune microenvironments within a patient are likely to be dynamic and sensitive to intrinsic (e.g., mutations and cell-cell communication) and extrinsic perturbations (e.g., prior treatment and microbiome) (Sivan et al., 2015; Vetizou et al., 2015).

5.2 Treatment-naive HGSOC

Given the observations of the case study and the mentioned limitations, we set up a follow-up study with treatment-naive HGSOC patients with metastatic disease to further evaluate intra-patient TME heterogeneity. We performed an unbiased transcriptomic analysis of primary and disseminated treatment-naive HGSOC tumours and detailed immunofluorscent quantification of T cell subpopulations. We explored the main sources of variation in the transcriptomic space among treatment naive samples and detected that transcriptomic pathway heterogeneity is mainly explained by presence of immune and stromal cells, or lack thereof. Importantly, the degree of immune signature variation within patients was similar to the one we observed in a case study of metastatic HGSOC, where different tumour immune microenvironments were associated with clinical outcome: tumours with high immune related pathways regressed and presented evidence of T cell activation, while immune excluded tumours progressed (**Figure 2.1**). In addition, all patients presented at least one tumour with low immune infiltration, suggesting that HGSOC is characterised by TME havens, which could underlie primary and acquired resistance to therapies (Hirata et al., 2015; Sharma et al., 2017). Through MR image analysis we captured immune signature differences within tumours, which was confirmed by immunofluorescence staining of T cells, and again variation of T cell infiltration within tumours was observed. Together the transcriptional, imaged-based, and immunofluorescence analyses show that TME heterogeneity is an intrinsic feature of HGSOC, which spans across patients, tumours within

patients, and within tumours. Furthermore, we found different levels of general immune infiltrates on different metastases within patients as in the case study. Having multiple tumour samples from the same patient, we were able to perform differential expression analysis between high and low purity tumours controlling for patient dependency. Pathway analysis of the differentially expressed genes confirmed the previous observation that Wnt and Myc signalling pathways were more prevalent in purer tumours. This result supports the evidence we also observed in the case study, where immune-excluded tumours showed an increase expression of Wnt signalling compared to infiltrated tumours (**Figure 2.2**). In addition, accumulated evidence in different tumour types and murine cancer models suggests that Wnt signalling can promote immune-exclusion (Spranger et al., 2015, 2017, 2016; Sridharan et al., 2016), and that the immune-exclusion was caused by a tumour cell-intrinsic oncogenic Wnt signalling activation (Damsky et al., 2011; Gounari et al., 2002; Spranger et al., 2015). Likewise, recently it was shown in a mouse model of lung cancer that Myc signalling overexpression can reprogram the immune component of the TME to a more immunosuppressed state, which was substantially dependent on NK cells (Kortlever et al., 2017). Previous work has also suggested that Myc signalling can inhibit T cell activation and infiltration (Rakhra et al., 2010), as well as directly regulating expression of immune checkpoint molecules in cancer cells (Casey et al., 2016). However, more research is required to conclusively elucidate the potential molecular and cellular mechanisms behind the association between Wnt and Myc signalling with low immune infiltration in HGSOC.

The evaluation of TME cell population in the tumours using expression data has been an area of intense research in the last decade (Finotello and Trajanoski, 2018). Previous unbiased immune enrichment and deconvolution studies have estimated relative abundances of immune cells in ovarian cancer tumours (Newman et al., 2015), and calculated survival associations (Li et al., 2016); however, no objective independent benchmarking has been performed to date, and discordant results make it difficult to judge these observations (Li et al., 2017b; Zheng, 2017; **?**). We integrated data from different enrichment/deconvolution methods and generated Consensus$^{\text{TME}}$, a consensus enrichment approach that consistently improved the predictions on our 440 quantifications IF data set of T cell subsets, and in TCGA ovarian cancer leukocyte methylation. We then evaluated which cell populations may have a higher degree of infiltrates in impure treatment-naive HGSOC tumours. The analysis using Consensus$^{\text{TME}}$ showed that cytotoxic and NK cells were the major populations present in low purity tumours, while endothelial cells were the main TME cell types in high purity tumours. A previous pan-cancer analysis showed negative correlations between somatic copy number alterations (SCNA) and estimated immune infiltrates, and that NK cells and CD8$^+$ T cell receptor pathway were the most differentially abundant immune factors between tumour samples with high versus low SCNA (Davoli et al., 2017). Despite these associations, the molecular mechanisms underlying this negative correlation between immune infiltrate and SCNA in HGSOC have not been elucidated. Interestingly, NK cell infiltration appeared to be significantly higher in tumours with low SCNAs in HGSOC (Davoli et al., 2017), and our results also showed that NK cells were more enriched than CD8$^+$ and CD4$^+$ T cell infiltration in low purity tumours, which would have low SCNA according to the previous

study (Davoli et al., 2017)), and have higher Wnt and Myc signalling. In addition, different studies have observed association between loss of *p53* function and decrease of NK cell infiltration in mouse models (Xue et al., 2007) and T cell infiltration in human breast cancers (Iannello et al., 2013). In fact, missense or nonsense mutations in *p53* are the earliest and almost ubiquitous (96%) alterations in HGSOC (Ahmed et al., 2010; TCGA, 2011). Altogether, these different lines of evidence help to clarify why HGSOC is intrinsically non-immunogenic: beyond the low somatic missense mutation load and the high SCNAs, the intrinsic oncogenic signalling of HGSOC seem to shape the TME and hinder immune infiltration of T cells and NK cells in treatment-naive tumours.

5.3 Pre/post neoadjuvant chemotherapy HGSOC

Targeting the TME has proven to be extremely successful in some cases, and thus represents a promising alternative to complement current therapeutic strategies in different tumour types including HGSOC (Hansen et al., 2016). Since the standard of care treatment in HGSOC is primary debulking surgery followed by platinum based chemotherapy or NACT followed by interval debulking surgery, there are significant potential clinical implications from understanding the effect of chemotherapy on the TME and the molecular drivers of the TME heterogeneity observed. While response rates to initial treatment are high, acquired resistance and relapse occurs in 80 to 90% of cases (Bowtell et al., 2015). A previous study investigated the effect of NACT on the activation of $CD8^+$, $CD4^+$ and Tregs in HGSOC, as well as systemic levels of cytokines (Böhm et al., 2016). This study found that patients who had good responses to NACT had a decrease of Tregs after treatment compared to poor responders. In general, there was also a trend towards higher cytolytic activity in tumours after NACT despite failing to detect significant changes on $CD8^+$ T cell counts (Böhm et al., 2016). Therefore, we applied the same unbiased approach we used with the transcriptomes of treatment-naive HGSOC samples to explore the effect of chemotherapy on the TME of matched and unmatched pre/post-NACT treated HGSOC samples. We found an increase of cytotoxic immunogenic activity after NACT in matched tumour samples but not in site-unmatched samples from the same patient. Similarly, employing TCR-seq, we found a significant increase in T cells and TCR clonality in matched samples, but no significant difference was detected in unmatched pairs. Comparing post-NACT site-matched and post-NACT site-unmatched samples indicated that the observed change in TCR clonal expansions was driven by chemotherapy and not by the biopsy itself, although we cannot formally exclude potential immunogenic effects that the biopsy procedure may have in a neoadjuvant setting. Together, these results show a clear confounding effect that spatial TME heterogeneity can cause when deriving biological interpretations if heterogeneity is not taken into account. In addition, these results indicate that besides T cells, NK cells could be a potential TME cellular target in HGSOC. Further experimental and clinical investigations would be required to evaluate this data driven hypothesis.

Having unmasked the immune activation generated by NACT, we used the matched samples to

investigate the factors in pre-treated samples that influence TCR clonal expansion induction upon NACT. Interestingly, cytotoxic cells showed a significant negative association with TCR clonal expansion, and the PI3K-AKT-MTOR pathway, which is part of the TCR signalling cascade, also showed a strong negative association. Conversely, Wnt and Myc signalling appeared as positively associated with TCR clonal expansion. Together, these results suggest that tumours with low levels of infiltration have a higher potential for T cell activation upon NACT than tumours with a previous immune presence. We hypothesise that this could be due to an already exhausted immune TME generated as a consequence of chronic immune and tumour interaction, while immune excluded tumours present a fresh environment where T cells can become active and expand. These results provide an argument for rational use of checkpoint blockade immunotherapy after NACT that warrants further clinical study. Importantly, our results also point towards NK cells being similar or more activated after NACT, further suggesting NK cells as potential therapeutic cellular target for combination therapy. Finally, this conclusion could only be drawn when intra-patient spatial TME heterogeneity was controlled for, again highlighting the necessity to take TME heterogeneity into consideration in translational studies and in clinical applications.

Taking all the clinical data together, here we have presented how TME heterogeneity is a common feature of HGSOC and how it can affect the interpretation of translational studies. We have sought to uncover molecular and cellular mechanisms behind intra-patient and intra-tumour TME heterogeneity. However, there are critical limitations to consider. Disentangling the actual mechanisms using human tumour samples represents a formidable challenge since tissue samples are limited, inter-patient variability is prominent, and mechanistic experimental validation on these patients is prohibitive. Given these constraints, these studies are descriptive in nature and relies heavily on independent observations derived by other studies that use murine tumour models to propose suitable explanations. Despite this main limitation, our unbiased analysis on human tumours not only complements *in vivo* studies, but also provides new hypotheses to further explore in a pre-clinical and clinical settings. Another major limitation is the small number of samples available compared to the large parameter space to investigate in an unbiased systematic study, which poses a challenge not only for achieving statistically meaningful results but also for the findings to be influenced by confounding variables and biological/technical noise. However, the implementation of orthogonal methods in combination with the independence of the case study, the treatment-naive, and NACT cohorts provide solid evidence of TME heterogeneity at multiple dimensions in addition to reasonable putative mechanisms behind it. Ultimately, we hope that the increasing scientific evidence would lead to better designed clinical trials, where TME heterogeneity is further evaluated, data driven combination therapies or novel therapies are tested, and tissue sections are systematically analysed and stored for future integrative unbiased analyses.

5.4 TME cell estimation from bulk tumour expression data

Beyond systematic unbiased analyses, it is important also to consider the inherent limitations and opportunities with the data obtained from these samples. It is known that tumours comprise highly complex cellular compositions and inter-cellular interactions. Thus, disentangling the underplaying mixture of cellular and molecular interactions occurring in tumours and their microenvironment is a challenging task, particularly when analysing bulk tumour data. Bulk sequencing analysis poses inherent limitations such as the need to employ deconvolution or enrichment methods to estimate cellular populations. These estimations are not 100% accurate and absolute cell counts have not been estimated yet [reviewed in (Finotello and Trajanoski, 2018)]. However, the huge amount of publicly available bulk data generated so far and the ever decreasing cost of sequencing has made the estimation of TME cell populations in bulk expression data a data mining challenge worth pursuing. Multiple computational approaches have been developed to estimate quantitatively or semi-quantitatively the different TME cell populations, using different statistical frameworks, and different using their own benchmark data sets. A problem however, is that each method claims to outperform previous ones, which has created confusion and debates in the field (Li et al., 2017a; Newman et al., 2017). The need for independent and more comprehensive benchmarks has then be rightly proposed (Zheng, 2017). After benchmarking six recently published TME cell population estimation methods using bulk expression profiles and T cell counts from IF data, as well as TCGA HGSOC cancer samples with leukocyte methylation data, we did not find a consistently outperforming method. Thus, we generated a consensus enrichment approach that integrates marker genes form the other methods and filters out genes that most likely are cancer related based on purity-gene expression correlations, we called the tool: "Consensus^{TME}". Using Consensus^{TME} we found that it consistently outperformed the other methods in our IF data set and in the TCGA HGSOC. Thus we performed an unbiased pan-cancer benchmarks, and collected independent benchmarks data sets from the different methods to comprehensively evaluate the methods performances.

Currently we are also evaluating whether we can add weights to the genes selected by Consensus^{TME}. The reason behind adding weights to genes is based on the idea that genes can be expressed by non-cancerous and cancerous cells, thus a "confidence score or weight" for each gene would potentially increase the accuracy of Consensus^{TME}, by giving greater weights to genes that are most likely expressed only by non-cancerous cells. One way to do it would be to calculate gene expression and CNA correlations, since positive correlations are most likely to represent tumour derived transcripts, while no correlation would most likely represent non-cancerous derived transcripts. We can then transform the correlations to their additive inverse and convert to a positive scale to generate weights for each gene in a tumour type specific manner using TCGA data. This also would compensate for cellular gene expression collinearity, which is an obstacle in the estimation of TME cell populations, as more genes with different weights would be added to each cell type and tumour type. Compensating biological collinearity is crucial since it is challenging to differentiate true biological from statistical

collinearity, as well as distinguish similar cell types in a complex sample (Li et al., 2017a). This is work ongoing, but preliminary results (not shown here) suggest that a the weighted version of Consensus^{TME} could improve its performance. Overall, these group of analyses show the utility and need of independent comprehensive benchmarks to assess tool performances. Additionally, by leveraging information from different tools and adding gene-specific and tumour-specific information increased the consistency in performance on the different benchmarks. Lastly, Consensus^{TME} is an evolvable method by conception, which means that other genes used by new methods can be added to the gene sets and tested with the already defined benchmarks.

Despite the progress made with the different methods created for estimation of TME cell populations, it is foreseeable that given the amount of heterogeneity in tumours, an upper level limit in accuracy will be reached at some point. Similarly, due to the difficult task of setting apart biological and statistical collinearity, and the intrinsic biological and technical noise, the level of resolution between cell types and activation states will be reached sooner or later. Moreover, these TME cell estimation methods depend completely on prior knowledge of cell types and gene markers of them, precluding the possibility of discovering new cell types or cellular states. Also, the integration of bulk RNA sequencing data and image analysis is correlative and mechanistic insights are descriptive in nature. An imminent consequence of these limitations together is that a large number of samples is often required to observe significant associations, which is not trivial for phase I and II clinical trials, since these usually recruit an order of magnitude of patients less than the required for achieving statistically significant observations using bulk data. Novel ways to study the TME with higher level of granularity and data types are required. As the fields of single-cell RNA-seq and high-dimensional image analyses continue to progress, new advances and discoveries tackling limitations of bulk genomic studies in cancer and TME dynamics are likely to be covered, discoveries confirmed, completed, and complemented (Azizi et al., 2018; van Dijk et al., 2018; Zheng et al., 2017).

Overall, the studies presented in this thesis show that the TME of HGSOC is intrinsically heterogeneous within patients and within tumours, posing an important barrier for the successful application of therapies in general, and in particular ones that target the TME, like checkpoint blockade immunotherapy. However, insights into potential mechanisms driving TME heterogeneity obtained only after controlling patient dependencies, put forward new therapeutic strategies to be explored in future studies. Furthermore, the induced immunogenicity upon NACT treatment was only unmasked after taking into account the TME heterogeneity, which otherwise acts as a confounding variable. Also, coordinating comprehensive benchmarks for TME cell population estimation methods, and leveraging upon independent methods to integrate a consensus approach is a valuable strategy to increase consistent accuracy of estimated proportions. Finally, increasing our understanding of the TME composition and dynamics has the potential to change the current paradigm of treatment and hopefully improve clinical outcomes.

Chapter 6

Conclusion

Here we showed evidence of divergent tumour genetics, TMEs, and immune activation within a single patient with advanced HGSOC. We confirmed the coexistence of TMEs within treatment-naive HGSOC patients, and reported intra-tumour levels of TME heterogeneity as well. Furthermore, the induced immunogenicity upon NACT treatment was only unmasked after taking into account the TME heterogeneity, which otherwise acts as a confounding variable. The intra-patient TME heterogeneity described here bespeaks a profound clinical challenge for therapeutic success and correct interpretations. By controlling patient dependency and accounting for intra-patient TME heterogeneity, insights into potential mechanisms driving TME heterogeneity were obtained, putting forward new therapeutic strategies to be explored in future studies. These observations, although made in small patient cohorts, may explain the frequent heterogeneous responses seen clinically and the lack of objective response to checkpoint blockade immunotherapy. In addition, we conclude that comprehensive independent benchmarks should be performed more broadly to inform the community and provide more accurate and reliable systematic analyses. Our benchmarks for TME cell population methods showed no consistent outperforming method, which prompted us to generate a consensus approach (Consensus$^{\text{TME}}$). Unlike the other methods, Consensus$^{\text{TME}}$ consistently performed among the top methods, sometimes even outperforming methods in their own data sets. Consensus$^{\text{TME}}$ was conceived to be an evolvable method which will incorporate other methods and provide the benchmarks to the community for their own tests. Given the data presented in these studies, it will be essential to understand not only how to therapeutically target ITH between and within metastases, but also how to successfully mobilise an anti-tumour immune response able to control all metastases in advanced cancers despite the broad intra-patient TME heterogeneity observed. We foresee this would require further systematic analyses of human tumour samples and robust computational methods to analyse the TME.

References

Ahmed, A. A., Etemadmoghadam, D., Temple, J., Lynch, A. G., Riad, M., Sharma, R., Stewart, C., Fereday, S., Caldas, C., DeFazio, A., Bowtell, D., and Brenton, J. D. (2010). Driver mutations in TP53 are ubiquitous in high grade serous carcinoma of the ovary. *The Journal of Pathology*, 221(1):49–56.

Alcover, A., Alarcón, B., Alarcón, A., and Bartolo, V. D. (2018). Cell Biology of T Cell Receptor Expression and Regulation. *Annual Review of Immunology*, 12:40.

Allemani, C., Weir, H. K., Carreira, H., Harewood, R., Spika, D., Wang, X.-S., Bannon, F., Ahn, J. V., Johnson, C. J., Bonaventure, A., Marcos-Gragera, R., Stiller, C., Azevedo e Silva, G., Chen, W.-Q., Ogunbiyi, O. J., Rachet, B., Soeberg, M. J., You, H., Matsuda, T., Bielska-Lasota, M., Storm, H., Tucker, T. C., and Coleman, M. P. (2015). Global surveillance of cancer survival 1995–2009: analysis of individual data for 25 676 887 patients from 279 population-based registries in 67 countries (CONCORD-2). *The Lancet*, 385(9972):977–1010.

Alves, J. M., Prieto, T., and Posada, D. (2017). Multiregional Tumor Trees Are Not Phylogenies. *Trends in Cancer*, 3(8):546–550.

Amé, J. C., Rolli, V., Schreiber, V., Niedergang, C., Apiou, F., Decker, P., Muller, S., Höger, T., Ménissier-de Murcia, J., and de Murcia, G. (1999). PARP-2, a novel mammalian DNA damage-dependent poly(ADP-ribose) polymerase. *J. Biol. Chem.*, 274(25):17860–17868.

Amir, E.-a. D., Davis, K. L., Tadmor, M. D., Simonds, E. F., Levine, J. H., Bendall, S. C., Shenfeld, D. K., Krishnaswamy, S., Nolan, G. P., and Pe'er, D. (2013). viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nature Biotechnology*, 31(6):545–552.

Andreatta, M. and Nielsen, M. (2015). Gapped sequence alignment using artificial neural networks: Application to the MHC class i system. *Bioinformatics*, 32(4):511–517.

Angell, H. and Galon, J. (2013). From the immune contexture to the Immunoscore: the role of prognostic and predictive immune markers in cancer. *Current Opinion in Immunology*, 25:261–267.

Antonio, N., Louise Bønnelykke-Behrndtz, M., Ward, L. C., Collin, J., Christensen, I. J., Steiniche, T., Schmidt, H., Feng, Y., and Martin, P. (2015). The wound inflammatory response exacerbates growth of pre-neoplastic cells and progression to cancer. *The EMBO Journal*, 34:2219–2236.

Aptsiauri, N., Carretero, R., Garcia-Lora, A., Real, L. M., Cabrera, T., and Garrido, F. (2008). Regressing and progressing metastatic lesions: Resistance to immunotherapy is predetermined by irreversible HLA class I antigen alterations. *Cancer Immunology, Immunotherapy*, 57(11):1727–1733.

Aran, D., Hu, Z., and Butte, A. J. (2017). xcell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.*, 18(1):220.

Arstila, T. P., Casrouge, a., Baron, V., Even, J., Kanellopoulos, J., and Kourilsky, P. (1999). A direct estimate of the human alphabeta T cell receptor diversity. *Science (New York, N.Y.)*, 286(5441):958–961.

Ascierto, M. L., Makohon-Moore, A., Lipson, E. J., Taube, J. M., McMiller, T. L., Berger, A. E., Fan, J., Kaunitz, G., Cottrell, T., Kohutek, Z., Favorov, A., Makarov, V., Riaz, N., Chan, T. A., Cope, L., Hruban, R. H., Pardoll, D. M., Taylor, B. S., Solit, D. B., Iacobuzio-Donahue, C. A., and Topalian, S. L. (2017). Transcriptional mechanisms of resistance to anti-PD-1 therapy. *Clinical Cancer Research*, 23(12):clincanres.0270.2016.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000). Gene ontology: tool for the unification of biology. the gene ontology consortium. *Nat. Genet.*, 25(1):25–29.

Azizi, E., Carr, A. J., Plitas, G., Cornish, A. E., Konopacki, C., Prabhakaran, S., Nainys, J., Wu, K., Kiseliovas, V., Setty, M., Choi, K., Fromme, R. M., Dao, P., McKenney, P. T., Wasti, R. C., Kadaveru, K., Mazutis, L., Rudensky, A. Y., and Pe'er, D. (2018). Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. *Cell*, 174(5):1293–1308.e36.

Bahram, S., Arnold, D., Bresnahan, M., Strominger, J. L., and Spies, T. (1991). Two putative subunits of a peptide pump encoded in the human major histocompatibility complex class II region. *Proc. Natl. Acad. Sci. USA*, 88(November):10094–10098.

Barbie, D. A., Tamayo, P., Boehm, J. S., Kim, S. Y., Moody, S. E., Dunn, I. F., Schinzel, A. C., Sandy, P., Meylan, E., Scholl, C., Fröhling, S., Chan, E. M., Sos, M. L., Michel, K., Mermel, C., Silver, S. J., Weir, B. A., Reiling, J. H., Sheng, Q., Gupta, P. B., Wadlow, R. C., Le, H., Hoersch, S., Wittner, B. S., Ramaswamy, S., Livingston, D. M., Sabatini, D. M., Meyerson, M., Thomas, R. K., Lander, E. S., Mesirov, J. P., Root, D. E., Gilliland, D. G., Jacks, T., and Hahn, W. C. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature*, 462(7269):108–112.

Bashashati, A., Ha, G., Tone, A., Ding, J., Prentice, L. M., Roth, A., Rosner, J., Shumansky, K., Kalloger, S., Senz, J., Yang, W., McConechy, M., Melnyk, N., Anglesio, M., Luk, M. T. Y., Tse, K., Zeng, T., Moore, R., Zhao, Y., Marra, M. A., Gilks, B., Yip, S., Huntsman, D. G., McAlpine, J. N., and Shah, S. P. (2013). Distinct evolutionary trajectories of primary high-grade serous ovarian cancers revealed through spatial mutational profiling. *Journal of Pathology*, 231(1):21–34.

Bates, D., Mächler, M., Bolker, B., and Walker, S. (2015). Fitting linear Mixed-Effects models using lme4. *J. Stat. Softw.*, 67(1).

Bauer, S., Groh, V., Wu, J., Steinle, a., Phillips, J. H., Lanier, L. L., and Spies, T. (1999). Activation of NK cells and T cells by NKG2D, a receptor for stress-inducible MICA. *Science (New York, N.Y.)*, 285(5428):727–729.

Bayik, D., Tross, D., Haile, L. A., Verthelyi, D., and Klinman, D. M. (2017). Regulation of the maturation of human monocytes into immunosuppressive macrophages. *Blood Advances*, 1(26):2510–2519.

Becht, E., Giraldo, N. A., Lacroix, L., Buttard, B., Elarouci, N., Petitprez, F., Selves, J., Laurent-Puig, P., Sautès-Fridman, C., Fridman, W. H., and de Reyniès, A. (2016). Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol.*, 17(1):218.

Benci, J. L., Xu, B., Qiu, Y., Wu, T. J., Dada, H., Twyman-Saint Victor, C., Cucolo, L., Lee, D. S., Pauken, K. E., Huang, A. C., Gangadhar, T. C., Amaravadi, R. K., Schuchter, L. M., Feldman, M. D., Ishwaran, H., Vonderheide, R. H., Maity, A., Wherry, E. J., and Minn, A. J. (2016). Tumor Interferon Signaling Regulates a Multigenic Resistance Program to Immune Checkpoint Blockade. *Cell*, 167(6):1540–1554.e12.

Bergers, G. and Hanahan, D. (2008). Modes of resistance to anti-angiogenic therapy.

Bergers, G., Song, S., and Francisco, S. (2005). Neuro-Oncology The role of pericytes in blood-vessel formation and maintenance 1.

Bernards, A. and Settleman, J. (2004). GAP control: Regulating the regulators of small GTPases. *Trends in Cell Biology*, 14(7):377–385.

Bhang, H.-e. C., Ruddy, D. A., Krishnamurthy Radhakrishna, V., Caushi, J. X., Zhao, R., Hims, M. M., Singh, A. P., Kao, I., Rakiec, D., Shaw, P., Balak, M., Raza, A., Ackley, E., Keen, N., Schlabach, M. R., Palmer, M., Leary, R. J., Chiang, D. Y., Sellers, W. R., Michor, F., Cooke, V. G., Korn, J. M., and Stegmeier, F. (2015). Studying clonal dynamics in response to cancer therapy using high-complexity barcoding. *Nature Medicine*, 21(5):440–8.

Bhattacharya, S., Andorf, S., Gomes, L., Dunn, P., Schaefer, H., Pontius, J., Berger, P., Desborough, V., Smith, T., Campbell, J., Thomson, E., Monteiro, R., Guimaraes, P., Walters, B., Wiser, J., and Butte, A. J. (2014). ImmPort: disseminating data to the public for the future of immunology. *Immunol. Res.*, 58(2-3):234–239.

Biesecker, L. G. and Green, R. C. (2014). Diagnostic Clinical Genome and Exome Sequencing. *New England Journal of Medicine*, 370(25):2418–2425.

Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A., Angell, H., Fredriksen, T., Lafontaine, L., Berger, A., Bruneval, P., Fridman, W., Becker, C., Pagès, F., Speicher, M., Trajanoski, Z., and Galon, J. (2013). Spatiotemporal Dynamics of Intratumoral Immune Cells Reveal the Immune Landscape in Human Cancer. *Immunity*, 39(4):782–795.

Binnewies, M., Roberts, E. W., Kersten, K., Chan, V., Fearon, D. F., Merad, M., Coussens, L. M., Gabrilovich, D. I., Ostrand-Rosenberg, S., Hedrick, C. C., Vonderheide, R. H., Pittet, M. J., Jain, R. K., Zou, W., Howcroft, T. K., Woodhouse, E. C., Weinberg, R. A., and Krummel, M. F. (2018). Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nature Medicine*, 24(5):541–550.

Böhm, S., Montfort, A., Pearce, O. M. T., Topping, J., Chakravarty, P., Everitt, G. L. A., Clear, A., McDermott, J. R., Ennis, D., Dowe, T., Fitzpatrick, A., Brockbank, E. C., Lawrence, A. C., Jeyarajah, A., Faruqi, A. Z., McNeish, I. A., Singh, N., Lockley, M., and Balkwill, F. R. (2016). Neoadjuvant chemotherapy modulates the immune microenvironment in metastases of Tubo-Ovarian High-Grade serous carcinoma. *Clin. Cancer Res.*, 22(12):3025–3036.

Bowtell, D. D., Böhm, S., Ahmed, A. A., Aspuria, P.-J., Bast, R. C., Beral, V., Berek, J. S., Birrer, M. J., Blagden, S., Bookman, M. A., Brenton, J. D., Chiappinelli, K. B., Martins, F. C., Coukos, G., Drapkin, R., Edmondson, R., Fotopoulou, C., Gabra, H., Galon, J., Gourley, C., Heong, V., Huntsman, D. G., Iwanicki, M., Karlan, B. Y., Kaye, A., Lengyel, E., Levine, D. A., Lu, K. H.,

McNeish, I. A., Menon, U., Narod, S. A., Nelson, B. H., Nephew, K. P., Pharoah, P., Powell, D. J., Ramos, P., Romero, I. L., Scott, C. L., Sood, A. K., Stronach, E. A., and Balkwill, F. R. (2015). Rethinking ovarian cancer II: reducing mortality from high-grade serous ovarian cancer. *Nature Reviews Cancer*, 15(11):668–679.

Brabletz, T., Lyden, D., Steeg, P. S., and Werb, Z. (2013). Roadblocks to translational advances on metastasis research. *Nat Med*, 19(9):1104–1109.

Brett M., R., Jennifer B., P., Thomas A., S., Brett M., R., Jennifer B., P., and Thomas A., S. (2017). Epidemiology of ovarian cancer: a review. *Cancer Biology & Medicine*, 14(1):9–32.

Bronger, H., Singer, J., Windmüller, C., Reuning, U., Zech, D., Delbridge, C., Dorn, J., Kiechle, M., Schmalfeldt, B., Schmitt, M., and Avril, S. (2016). CXCL9 and CXCL10 predict survival and are regulated by cyclooxygenase inhibition in advanced serous ovarian cancer. *British Journal of Cancer*, 115(5):553–563.

Brown, S. D., Warren, R. L., Gibb, E. a., Martin, S. D., Spinelli, J. J., Nelson, B. H., and Holt, R. a. (2014). Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. *Genome Research*, 24(5):743–750.

Burel, J. G. and Peters, B. (2018). Discovering transcriptional signatures of disease for diagnosis versus mechanism. *Nat. Rev. Immunol.*, 18(5):289–290.

Byrne, M. B., Leslie, M. T., Gaskins, H. R., and Kenis, P. J. (2014). Methods to study the tumor microenvironment under controlled oxygen conditions. *Trends in Biotechnology*, 32(11):556–563.

Calis, J. J. A., Maybeno, M., Greenbaum, J. A., Weiskopf, D., De Silva, A. D., Sette, A., Keşmir, C., and Peters, B. (2013). Properties of MHC class I presented peptides that enhance immunogenicity. *PLoS computational biology*, 9(10):e1003266.

Calle, E. E., Rodriguez, C., Walker-Thurmond, K., and Thun, M. J. (2003). Overweight, Obesity, and Mortality from Cancer in a Prospectively Studied Cohort of U.S. Adults. Technical report.

Campbell, P. J., Yachida, S., Mudie, L. J., Stephens, P. J., Pleasance, E. D., Stebbings, L. A., Morsberger, L. A., Latimer, C., McLaren, S., Lin, M.-L., McBride, D. J., Varela, I., Nik-Zainal, S. A., Leroy, C., Jia, M., Menzies, A., Butler, A. P., Teague, J. W., Griffin, C. A., Burton, J., Swerdlow, H., Quail, M. A., Stratton, M. R., Iacobuzio-Donahue, C., and Futreal, P. A. (2010). The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature*, 467(7319):1109–1113.

Cannistra, S. A. (2004). Cancer of the Ovary. *N Engl J Med*, 35124351:2519–29.

Cantor, S. B., Urano, T., and Feig, L. A. (1995). Identification and characterization of ral-binding protein 1, a potential downstream target of ral GTPases. *Mol. Cell. Biol.*, 15(8):4578–4584.

Carano, R. A. D., Ross, A. L., Ross, J., Williams, S. P., Koeppen, H., Schwall, R. H., and Van Bruggen, N. (2004). Quantification of tumor tissue populations by multispectral analysis. *Magn. Reson. Med.*, 51(3):542–551.

Carson, W. E., Shapiro, C. L., Crespin, T. R., Thornton, L. M., and Andersen, B. L. (2004). Cellular immunity in breast cancer patients completing taxane treatment. *Clinical Cancer Research*, 10(10):3401–3409.

Carter, S. L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P. W., Onofrio, R. C., Winckler, W., Weir, B. a., Beroukhim, R., Pellman, D., Levine, D. a., Lander, E. S., Meyerson, M., and Getz, G. (2012). Absolute quantification of somatic DNA alterations in human cancer. *Nature Biotechnology*, 30(5):413–421.

Casey, S. C., Tong, L., Li, Y., Do, R., Walz, S., Fitzgerald, K. N., Gouw, A. M., Baylot, V., Gütgemann, I., Eilers, M., and Felsher, D. W. (2016). MYC regulates the antitumor immune response through CD47 and PD-L1. *Science*, 352(6282):227–231.

Cerami, E., Gao, J., Dogrusoz, U., Gross, B. E., Sumer, S. O., Aksoy, B. A., Jacobsen, A., Byrne, C. J., Heuer, M. L., Larsson, E., Antipin, Y., Reva, B., Goldberg, A. P., Sander, C., and Schultz, N. (2012). The cbio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.*, 2(5):401–404.

Challa-Malladi, M., Lieu, Y. K., Califano, O., Holmes, A. B., Bhagat, G., Murty, V. V., Dominguez-Sola, D., Pasqualucci, L., and Dalla-Favera, R. (2011). Combined Genetic Inactivation of β2-Microglobulin and CD58 Reveals Frequent Escape from Immune Recognition in Diffuse Large B Cell Lymphoma. *Cancer Cell*, 20(6):728–740.

Chen, G. M., Kannan, L., Geistlinger, L., Kofia, V., Safikhani, Z., Gendoo, D. M., Parmigiani, G., Birrer, M. J., Haibe-Kains, B., and Waldron, L. (2018). Consensus on Molecular Subtypes of High-grade Serous Ovarian Carcinoma. *Clinical Cancer Research*, page clincanres.0784.2018.

Chen, P. L., Roh, W., Reuben, A., Cooper, Z. A., Spencer, C. N., Prieto, P. A., Miller, J. P., Bassett, R. L., Gopalakrishnan, V., Wani, K., De Macedo, M. P., Austin-Breneman, J. L., Jiang, H., Chang, Q., Reddy, S. M., Chen, W. S., Tetzlaff, M. T., Broaddus, R. J., Davies, M. A., Gershenwald, J. E., Haydu, L., Lazar, A. J., Patel, S. P., Hwu, P., Hwu, W. J., Diab, A., Glitza, I. C., Woodman, S. E., Vence, L. M., Wistuba, I. I., Amaria, R. N., Kwong, L. N., Prieto, V., Eric Davis, R., Ma, W., Overwijk, W. W., Sharpe, A. H., Hu, J., Andrew Futreal, P., Blando, J., Sharma, P., Allison, J. P., Chin, L., and Wargo, J. A. (2016). Analysis of immune signatures in longitudinal tumor samples yields insight into biomarkers of response and mechanisms of resistance to immune checkpoint blockade. *Cancer Discovery*, 6(8):827–837.

Cibulskis, K., Lawrence, M. S., Carter, S. L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E. S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnology*, 31(3):213–219.

Ciriello, G., Miller, M. L., Aksoy, B. A., Senbabaoglu, Y., Schultz, N., and Sander, C. (2013). Emerging landscape of oncogenic signatures across human cancers. *Nature Genetics*, 45(10):1127–1133.

Cojocaru, E., Parkinson, C., and Brenton, J. (2018). Personalising Treatment for High-Grade Serous Ovarian Carcinoma. *Clinical Oncology*, 30(8):515–524.

Colak, S. and ten Dijke, P. (2017). Targeting TGF-β signaling in cancer. *Trends Cancer Res.*, 3(1):56–71.

Cooper, M. D. (2015). The early history of B cells. *Nature Reviews Immunology*, 15(3):191–197.

Corrales, L., Matson, V., Flood, B., Spranger, S., and Gajewski, T. F. (2017). Innate immune signaling and regulation in cancer immunotherapy. *Cell Research*, 27:96–108.

Crome, S. Q., Nguyen, L. T., Lopez-Verges, S., Yang, S. Y. C., Martin, B., Yam, J. Y., Johnson, D. J., Nie, J., Pniak, M., Yen, P. H., Milea, A., Sowamber, R., Katz, S. R., Bernardini, M. Q., Clarke, B. A., Shaw, P. A., Lang, P. A., Berman, H. K., Pugh, T. J., Lanier, L. L., and Ohashi, P. S. (2017). A distinct innate lymphoid cell population regulates tumor-associated T cells. *Nature Medicine*, 23(3):368–375.

Dagogo-Jack, I. and Shaw, A. T. (2017). Tumour heterogeneity and resistance to cancer therapies. *Nature Reviews Clinical Oncology*, 15(2):81–94.

Damsky, W. E., Curley, D. P., Santhanakrishnan, M., Rosenbaum, L. E., Platt, J. T., Gould Rothberg, B. E., Taketo, M. M., Dankort, D., Rimm, D. L., McMahon, M., and Bosenberg, M. (2011). β-catenin signaling controls metastasis in braf-activated pten-deficient melanomas. *Cancer Cell*, 20(6):741–754.

Davies, S. M. K., Rackham, O., Shearwood, A.-M. J., Hamilton, K. L., Narsai, R., Whelan, J., and Filipovska, A. (2009). Pentatricopeptide repeat domain protein 3 associates with the mitochondrial small ribosomal subunit and regulates translation. *FEBS Lett.*, 583(12):1853–1858.

Davis, B. P., Rothenberg, M. E., and Rothenberg, M. E. (2014). Masters of Immunology Eosinophils and Cancer Disclosure of Potential Conflicts of Interest. *Cancer Immunol Res*, 2(1).

Davoli, T., Uno, H., Wooten, E. C., and Elledge, S. J. (2017). Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science*, 355(6322):eaaf8399.

De Mattos-Arruda, L., Bidard, F. C., Won, H. H., Cortes, J., Ng, C. K. Y., Peg, V., Nuciforo, P., Jungbluth, A. A., Weigelt, B., Berger, M. F., Seoane, J., and Reis-Filho, J. S. (2014). Establishing the origin of metastatic deposits in the setting of multiple primary malignancies: The role of massively parallel sequencing. *Molecular Oncology*, 8(1):150–158.

Demir, R. H. and Marchand, G. J. (2012). Adnexal Masses Suspected to Be Benign Treated with Laparoscopy. *JSLS : Journal of the Society of Laparoendoscopic Surgeons*, 16(1):71–84.

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., Philippakis, A. A., del Angel, G., Rivas, M. A., Hanna, M., McKenna, A., Fennell, T. J., Kernytsky, A. M., Sivachenko, A. Y., Cibulskis, K., Gabriel, S. B., Altshuler, D., and Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, 43(5):491–498.

Devarajan, E., Song, Y. H., Krishnappa, S., and Alt, E. (2012). Epithelial-mesenchymal transition in breast cancer lines is mediated through PDGF-D released by tissue-resident stem cells. *International Journal of Cancer*, 131(5):1023–1031.

Dighe, a. S., Richards, E., Old, L. J., and Schreiber, R. D. (1994). Enhanced in vivo growth and resistance to rejection of tumor cells expressing dominant negative IFN gamma receptors. *Immunity*, 1(6):447–456.

Dudley, M. E., Wunderlich, J. R., Robbins, P. F., Yang, J. C., Hwu, P., Schwartzentruber, D. J., Topalian, S. L., Sherry, R., Restifo, N. P., Hubicki, A. M., Robinson, M. R., Raffeld, M., Duray, P., Seipp, C. a., Rogers-Freezer, L., Morton, K. E., Mavroukakis, S. a., White, D. E., and Rosenberg, S. a. (2002). Cancer regression and autoimmunity in patients after clonal repopulation with antitumor lymphocytes. *Science (New York, N.Y.)*, 298(5594):850–854.

DuPage, M., Mazumdar, C., Schmidt, L. M., Cheung, A. F., and Jacks, T. (2012). Expression of tumour-specific antigens underlies cancer immunoediting.

Eberlein, C., Rooney, C., Ross, S. J., Farren, M., Weir, H. M., and Barry, S. T. (2014). E-Cadherin and EpCAM expression by NSCLC tumour cells associate with normal fibroblast activation through a pathway initiated by integrin αvβ6 and maintained through TGFβ signalling. *Oncogene*, 34:704–716.

Edgar, R. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research*, 30(1):207–210.

Eichmann, M., de Ru, A., van Veelen, P. A., Peakman, M., and Kronenberg-Versteeg, D. (2014). Identification and characterisation of peptide binding motifs of six autoimmune disease-associated human leukocyte antigen-class I molecules including HLA-B*39:06. *Tissue Antigens*, 84(4):378–388.

Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., Jassal, B., Jupe, S., Korninger, F., McKay, S., Matthews, L., May, B., Milacic, M., Rothfels, K., Shamovsky, V., Webber, M., Weiser, J., Williams, M., Wu, G., Stein, L., Hermjakob, H., and D'Eustachio, P. (2016). The reactome pathway knowledgebase. *Nucleic Acids Research*, 44(D1):D481–D487.

Fang, X., Schummer, M., Mao, M., Yu, S., Tabassam, F. H., Swaby, R., Hasegawa, Y., Tanyi, J. L., LaPushin, R., Eder, A., Jaffe, R., Erickson, J., and Mills, G. B. (2002). Lysophosphatidic acid is a bioactive mediator in ovarian cancer. *Biochim. Biophys. Acta*, 1582(1-3):257–264.

Finotello, F. and Trajanoski, Z. (2018). Quantifying tumor-infiltrating immune cells from transcriptomics data. *Cancer Immunology, Immunotherapy*, 67(7):1031–1040.

Fisher, S., Barry, A., Abreu, J., Minie, B., Nolan, J., Delorey, T. M., Young, G., Fennell, T. J., Allen, A., Ambrogio, L., Berlin, A. M., Blumenstiel, B., Cibulskis, K., Friedrich, D., Johnson, R., Juhn, F., Reilly, B., Shammas, R., Stalker, J., Sykes, S. M., Thompson, J., Walsh, J., Zimmer, A., Zwirko, Z., Gabriel, S., Nicol, R., and Nusbaum, C. (2011). A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biology*, 12(1):R1.

Fleming, C., Morrissey, S., Cai, Y., and Yan, J. (2017). $\gamma\delta$ T Cells: Unexpected Regulators of Cancer Development and Progression. *Trends in Cancer*, 3(8):561–570.

Fong, P. C., Yap, T. A., Boss, D. S., Carden, C. P., Mergui-Roelvink, M., Gourley, C., De Greve, J., Lubinski, J., Shanley, S., Messiou, C., A'Hern, R., Tutt, A., Ashworth, A., Stone, J., Carmichael, J., Schellens, J. H., de Bono, J. S., and Kaye, S. B. (2010). Poly(ADP)-Ribose Polymerase Inhibition: Frequent Durable Responses in BRCA Carrier Ovarian Cancer Correlating With Platinum-Free Interval. *Journal of Clinical Oncology*, 28(15):2512–2519.

Fox, J. and Weisberg, S. (2011). *An R Companion to Applied Regression*. SAGE Publications.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.*, 33(1):1–22.

Fritsch, E. F., Rajasagi, M., Ott, P. a., Brusic, V., Hacohen, N., and Wu, C. J. (2014). HLA-binding properties of tumor neoepitopes in humans. *Cancer immunology research*, 2(6):522–9.

Frye, B. C., Halfter, S., Djudjaj, S., Muehlenberg, P., Weber, S., Raffetseder, U., En-Nia, A., Knott, H., Baron, J. M., Dooley, S., Bernhagen, J., and Mertens, P. R. (2009). Y-box protein-1 is actively secreted through a non-classical pathway and acts as an extracellular mitogen. *EMBO Rep.*, 10(7):783–789.

Gaengel, K., Genové, G., Armulik, A., and Betsholtz, C. (2009). ATVB In Focus Developmental Biology in the Vasculature Endothelial-Mural Cell Signaling in Vascular Development and Angiogenesis.

Gao, J., Aksoy, B. A., Dogrusoz, U., Dresdner, G., Gross, B., Sumer, S. O., Sun, Y., Jacobsen, A., Sinha, R., Larsson, E., Cerami, E., Sander, C., and Schultz, N. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.*, 6(269):l1.

Gao, Y., Souza-Fonseca-Guimaraes, F., Bald, T., Ng, S. S., Young, A., Ngiow, S. F., Rautela, J., Straube, J., Waddell, N., Blake, S. J., Yan, J., Bartholin, L., Lee, J. S., Vivier, E., Takeda, K., Messaoudene, M., Zitvogel, L., Teng, M. W. L., Belz, G. T., Engwerda, C. R., Huntington, N. D., Nakamura, K., Hölzel, M., and Smyth, M. J. (2017). Tumor immunoevasion by the conversion of effector NK cells into type 1 innate lymphoid cells. *Nature Immunology*.

Gautier, L., Cope, L., Bolstad, B. M., and Irizarry, R. A. (2004). affy–analysis of affymetrix GeneChip data at the probe level. *Bioinformatics*, 20(3):307–315.

Gavalas, N. G., Karadimou, A., Dimopoulos, M. A., and Bamias, A. (2010). Immune response in ovarian cancer: How is the immune system involved in prognosis and therapy: Potential for treatment utilization. *Clinical and Developmental Immunology*, 2010.

Gawad, C., Koh, W., and Quake, S. R. (2016). Single-cell genome sequencing: current state of the science. *Nature Reviews Genetics*, 17(3):175–188.

Gelmon, K. A., Tischkowitz, M., Mackay, H., Swenerton, K., Robidoux, A., Tonkin, K., Hirte, H., Huntsman, D., Clemons, M., Gilks, B., Yerushalmi, R., Macpherson, E., Carmichael, J., and Oza, A. (2011). Olaparib in patients with recurrent high-grade serous or poorly differentiated ovarian carcinoma or triple-negative breast cancer: a phase 2, multicentre, open-label, non-randomised study. *The Lancet Oncology*, 12(9):852–861.

Gerhardt, H. and Semb, H. (2008). Pericytes: gatekeepers in tumour cell metastasis? *Journal of Molecular Medicine*, 86(2):135–144.

Gerlinger, M., Rowan, A. J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., Varela, I., Phillimore, B., Begum, S., McDonald, N. Q., Butler, A., Jones, D., Raine, K., Latimer, C., Santos, C. R., Nohadani, M., Eklund, A. C., Spencer-Dene, B., Clark, G., Pickering, L., Stamp, G., Gore, M., Szallasi, Z., Downward, J., Futreal, P. A., and Swanton, C. (2012). Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *New England Journal of Medicine*, 366(10):883–892.

Giannakis, M., Mu, X. J., Shukla, S. A., Qian, Z. R., Cohen, O., Nishihara, R., Bahl, S., Cao, Y., Amin-Mansour, A., Yamauchi, M., Sukawa, Y., Stewart, C., Rosenberg, M., Mima, K., Inamura, K., Nosho, K., Nowak, J. A., Lawrence, M. S., Giovannucci, E. L., Chan, A. T., Ng, K., Meyerhardt, J. A., Van Allen, E. M., Getz, G., Gabriel, S. B., Lander, E. S., Wu, C. J., Fuchs, C. S., Ogino, S., and Garraway, L. A. (2016). Genomic Correlates of Immune-Cell Infiltrates in Colorectal Carcinoma. *Cell Reports*, 15(4):857–865.

Gilboa, E. (1999). The makings of a tumor rejection antigen. *Immunity*, 11(3):263–270.

Gnirke, A., Melnikov, A., Maguire, J., Rogov, P., LeProust, E. M., Brockman, W., Fennell, T., Giannoukos, G., Fisher, S., Russ, C., Gabriel, S., Jaffe, D. B., Lander, E. S., and Nusbaum, C. (2009). Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature biotechnology*, 27(2):182–9.

Gómez-Cuadrado, L., Tracey, N., Ma, R., Qian, B., and Brunton, V. G. (2017). Mouse models of metastasis: progress and prospects. *Disease Models & Mechanisms*, 10(9):1061–1074.

Gotwals, P., Cameron, S., Cipolletta, D., Cremasco, V., Crystal, A., Hewes, B., Mueller, B., Quaratino, S., Sabatos-Peyton, C., Petruzzelli, L., Engelman, J. A., and Dranoff, G. (2017). Prospects for combining targeted and conventional cancer therapy with immunotherapy. *Nature Publishing Group*, 17.

Gounari, F., Signoretti, S., Bronson, R., Klein, L., Sellers, W. R., Kum, J., Siermann, A., Taketo, M. M., von Boehmer, H., and Khazaie, K. (2002). Stabilization of β-catenin induces lesions reminiscent of prostatic intraepithelial neoplasia, but terminal squamous transdifferentiation of other secretory epithelia. *Oncogene*, 21(26):4099–4107.

Groh, V., Rhinehart, R., Secrist, H., Bauer, S., Grabstein, K. H., and Spies, T. (1999). Broad tumor-associated expression and recognition by tumor-derived gamma delta T cells of MICA and MICB. *Proceedings of the National Academy of Sciences of the United States of America*, 96(12):6879–6884.

Gulukota, K., Sidney, J., Sette, a., and DeLisi, C. (1997). Two complementary methods for predicting peptides binding major histocompatibility complex molecules. *Journal of molecular biology*, 267(5):1258–1267.

Gunderson, C. C. and Moore, K. N. (2015). Olaparib: an oral PARP-1 and PARP-2 inhibitor with promising activity in ovarian cancer. *Future Oncol.*, 11(5):747–757.

Hamaï, A., Benlalam, H., Meslin, F., Hasmim, M., Carré, T., Akalay, I., Janji, B., Berchem, G., Noman, M. Z., and Chouaib, S. (2010). Immune surveillance of human cancer: If the cytotoxic T-lymphocytes play the music, does the tumoral system call the tune? *Tissue Antigens*, 75(1):1–8.

Hanahan, D. and Folkman, J. (1996). Patterns and Emerging Mechanisms Review of the Angiogenic Switch during Tumorigenesis. Technical report.

Hanahan, D. and Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. *Cell*, 144(5):646–674.

Hansen, J. M., Coleman, R. L., and Sood, A. K. (2016). Targeting the tumour microenvironment in ovarian cancer. *Eur. J. Cancer*, 56:131–143.

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-Seq data. *BMC Bioinformatics*, 14(1):7.

Heng, T. S. P., Painter, M. W., and Immunological Genome Project Consortium (2008). The immunological genome project: networks of gene expression in immune cells. *Nat. Immunol.*, 9(10):1091–1094.

Hirata, E., Girotti, M. R., Viros, A., Hooper, S., Spencer-Dene, B., Matsuda, M., Larkin, J., Marais, R., and Sahai, E. (2015). Intravital Imaging Reveals How BRAF Inhibition Generates Drug-Tolerant Microenvironments with High Integrin β1/FAK Signaling. *Cancer Cell*, 27(4):574–588.

Hiroyama, M. and Takenawa, T. (1999). Isolation of a cDNA encoding human lysophosphatidic acid phosphatase that is involved in the regulation of mitochondrial lipid biosynthesis. *J. Biol. Chem.*, 274(41):29172–29180.

Hoadley, K. A., Yau, C., Stuart, J. M., Benz, C. C., and Correspondence, P. W. L. (2018). Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell*, 173:291–304.

Hoadley, K. A., Yau, C., Wolf, D. M., Cherniack, A. D., Tamborero, D., Ng, S., Leiserson, M. D., Niu, B., McLellan, M. D., Uzunangelov, V., Zhang, J., Kandoth, C., Akbani, R., Shen, H., Omberg, L., Chu, A., Margolin, A. A., van't Veer, L. J., Lopez-Bigas, N., Laird, P. W., Raphael, B. J., Ding, L., Robertson, A. G., Byers, L. A., Mills, G. B., Weinstein, J. N., Van Waes, C., Chen, Z., Collisson, E. A., Benz, C. C., Perou, C. M., and Stuart, J. M. (2014). Multiplatform Analysis of 12 Cancer Types Reveals Molecular Classification within and across Tissues of Origin. *Cell*, 158(4):929–944.

Homicsko, K., Hospitalier, C., Vaudois, U., and De, É. P. F. (2016). Targeting Programmed Cell Death 1 in Ovarian Cancer. 33(34).

Horiba, N., Masuda, S., Takeuchi, A., Takeuchi, D., Okuda, M., and Inui, K.-I. (2003). Cloning and characterization of a novel na+-dependent glucose transporter (NaGLT1) in rat kidney. *J. Biol. Chem.*, 278(17):14669–14676.

Hu, Y. and Mivechi, N. F. (2003). HSF-1 interacts with ral-binding protein 1 in a stress-responsive, multiprotein complex with HSP90in vivo. *J. Biol. Chem.*, 278(19):17299–17306.

Huber, W., Carey, V. J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B. S., Bravo, H. C., Davis, S., Gatto, L., Girke, T., Gottardo, R., Hahne, F., Hansen, K. D., Irizarry, R. A., Lawrence, M., Love, M. I., MacDonald, J., Obenchain, V., Oleś, A. K., Pagès, H., Reyes, A., Shannon, P., Smyth, G. K., Tenenbaum, D., Waldron, L., and Morgan, M. (2015). Orchestrating high-throughput genomic analysis with bioconductor. *Nat. Methods*, 12(2):115–121.

Huppa, J. B. and Davis, M. M. (2003). T-cell-antigen recognition and the immunological synapse. *Nature reviews. Immunology*, 3(12):973–983.

Iannello, A., Thompson, T. W., Ardolino, M., Lowe, S. W., and Raulet, D. H. (2013). p53-dependent chemokine production by senescent tumor cells supports NKG2D-dependent tumor elimination by natural killer cells. *J. Exp. Med.*, 210(10):2057–2069.

Iglewicz, B. and Hoaglin, D. C. (1993). *How to Detect and Handle Outliers*. Asq Press.

Janssen, L. M. E., Ramsay, E. E., Logsdon, C. D., and Overwijk, W. W. (2017). The immune system in cancer metastasis: friend or foe? *Journal for ImmunoTherapy of Cancer*, 5(1).

Jiménez-Sánchez, A., Memon, D., Pourpe, S., Veeraraghavan, H., Li, Y., Vargas, H. A., Gill, M. B., Park, K. J., Zivanovic, O., Konner, J., Ricca, J., Zamarin, D., Walther, T., Aghajanian, C., Wolchok, J. D., Sala, E., Merghoub, T., Snyder, A., and Miller, M. L. (2017). Heterogeneous Tumor-Immune Microenvironments among Differentially Growing Metastases in an Ovarian Cancer Patient. *Cell*, 170(5):927–938.e20.

Johnson, B. E., Mazor, T., Hong, C., Barnes, M., Aihara, K., McLean, C. Y., Fouse, S. D., Yamamoto, S., Ueda, H., Tatsuno, K., Asthana, S., Jalbert, L. E., Nelson, S. J., Bollen, A. W., Gustafson, W. C., Charron, E., Weiss, W. A., Smirnov, I. V., Song, J. S., Olshen, A. B., Cha, S., Zhao, Y., Moore, R. A., Mungall, A. J., Jones, S. J. M., Hirst, M., Marra, M. A., Saito, N., Aburatani, H., Mukasa, A., Berger, M. S., Chang, S. M., Taylor, B. S., and Costello, J. F. (2014). Mutational analysis reveals the origin and therapy-driven evolution of recurrent glioma. *Science*, 343(6167):189–193.

Joyce, J. A. and Fearon, D. T. (2015). T cell exclusion, immune privilege, and the tumor microenvironment. *Science*, 348(6230):74–80.

Jung, Y., Kim, J. K., Shiozawa, Y., Wang, J., Mishra, A., Joseph, J., Berry, J. E., Mcgee, S., Lee, E., Sun, H., Wang, J., Jin, T., Zhang, H., Dai, J., Krebsbach, P. H., Keller, E. T., Pienta, K. J., and Taichman, R. S. (2013). ARTICLE Recruitment of mesenchymal stem cells into prostate tumours promotes metastasis. *Nature Communications*, 4.

Kajimura, S. (2017). Advances in the understanding of adipose tissue biology. *Nature Reviews Endocrinology*, 13(2):69–70.

Kalluri, R. and Zeisberg, M. (2006). Fibroblasts in cancer. *Nature Reviews Cancer*, 6(5):392–401.

Kanehisa, M. and Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 27(1):29–34.

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M., and Tanabe, M. (2015). KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44(D1):D457–D462.

Khan, K. A. and Kerbel, R. S. (2018). Improving immunotherapy outcomes with anti-angiogenic treatments and vice versa. *Nature Reviews Clinical Oncology*, 15(5):310–324.

Kim, Y., Ponomarenko, J., Zhu, Z., Tamang, D., Wang, P., Greenbaum, J., Lundegaard, C., Sette, A., Lund, O., Bourne, P. E., Nielsen, M., and Peters, B. (2012). Immune epitope database analysis resource. *Nucleic Acids Research*, 40(W1):525–530.

Kinahan, P. E., Doot, R. K., Wanner-Roybal, M., Bidaut, L. M., Armato, S. G., Meyer, C. R., and McLennan, G. (2009). PET/CT assessment of response to therapy: Tumor change measurement, truth data, and error. *Transl. Oncol.*, 2(4):223–230.

Kirsch, I., Vignali, M., and Robins, H. (2015). T-cell receptor profiling in cancer. *Mol. Oncol.*, 9(10):2063–2070.

Kitamura, T., Qian, B.-Z., and Pollard, J. W. (2015). Immune cell promotion of metastasis. *Nat. Rev. Immunol.*, 15(2):73–86.

Kortlever, R. M., Sodir, N. M., Wilson, C. H., Burkhart, D. L., Pellegrinet, L., Brown Swigart, L., Littlewood, T. D., and Evan, G. I. (2017). Myc cooperates with ras by programming inflammation and immune suppression. *Cell*, 171(6):1301–1315.e14.

Kotturi, M. F., Peters, B., Buendia-Laysa, F., Sidney, J., Oseroff, C., Botten, J., Grey, H., Buchmeier, M. J., and Sette, A. (2007). The CD8+ T-cell response to lymphocytic choriomeningitis virus involves the L antigen: uncovering new tricks for an old virus. *Journal of virology*, 81(10):4928–4940.

Kreiter, S., Vormehr, M., van de Roemer, N., Diken, M., Löwer, M., Diekmann, J., Boegel, S., Schrörs, B., Vascotto, F., Castle, J. C., Tadmor, A. D., Schoenberger, S. P., Huber, C., Türeci, Ö., and Sahin, U. (2015). Mutant MHC class II epitopes drive therapeutic immune responses to cancer. *Nature*, 520(7549):692–696.

Kuilman, T., Velds, A., Kemper, K., Ranzani, M., Bombardelli, L., Hoogstraat, M., Nevedomskaya, E., Xu, G., de Ruiter, J., Lolkema, M. P., Ylstra, B., Jonkers, J., Rottenberg, S., Wessels, L. F., Adams, D. J., Peeper, D. S., and Krijgsman, O. (2015). CopywriteR: DNA copy number detection from off-target sequence data. *Genome Biology*, 16(1):49.

Landgren, H. and Carlsson, P. (2004). FoxJ3, a novel mammalian forkhead gene expressed in neuroectoderm, neural crest, and myotome. *Dev. Dyn.*, 231(2):396–401.

Lauemøller, S. L., Holm, a., Hilden, J., Brunak, S., Holst Nissen, M., Stryhn, a., Østergaard Pedersen, L., and Buus, S. (2001). Quantitative predictions of peptide binding to MHC class I molecules using specificity matrices and anchor-stratified calibrations. *Tissue antigens*, 57(5):405–414.

Le Bihan, D., Breton, E., Lallemand, D., Aubin, M. L., Vignaud, J., and Laval-Jeantet, M. (1988). Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. *Radiology*, 168(2):497–505.

Lee, J.-Y., Yoon, J.-K., Kim, B., Kim, S., Kim, M. A., Lim, H., Bang, D., and Song, Y.-S. (2015). Tumor evolution and intratumor heterogeneity of an epithelial ovarian cancer investigated using next-generation sequencing. *BMC Cancer*, 15(1):85.

Li, B. and Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, 12(1):323.

Li, B., Liu, J. S., and Liu, X. S. (2017a). Revisit linear regression-based deconvolution methods for tumor gene expression data. *Genome Biology*, 18(1):127.

Li, B., Severson, E., Pignon, J.-C., Zhao, H., Li, T., Novak, J., Jiang, P., Shen, H., Aster, J. C., Rodig, S., Signoretti, S., Liu, J. S., and Liu, X. S. (2016). Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol.*, 17(1):174.

Li, H., Peyrollier, K., Kilic, G., and Brakebusch, C. (2014). Rho GTPases and cancer. *BioFactors*, 40(2):226–235.

Li, J., Byrne, K. T., Yan, F., Yamazoe, T., Chen, Z., Baslan, T., Richman, L. P., Lin, J. H., Sun, Y. H., Rech, A. J., Balli, D., Hay, C. A., Sela, Y., Merrell, A. J., Liudahl, S. M., Gordon, N., Norgard, R. J., Yuan, S., Yu, S., Chao, T., Ye, S., Eisinger-Mathason, T. K., Faryabi, R. B., Tobias, J. W., Lowe, S. W., Coussens, L. M., Wherry, E. J., Vonderheide, R. H., and Stanger, B. Z. (2018). Tumor Cell-Intrinsic Factors Underlie Heterogeneity of Immune Cell Infiltration and Response to Immunotherapy. *Immunity*, 49(1):178–193.e7.

Li, J., Wang, J., Chen, R., Bai, Y., and Lu, X. (2017b). The prognostic value of tumor-infiltrating T lymphocytes in ovarian cancer. *Oncotarget*, 8(9):15621–15631.

Li, T., Fan, J., Wang, B., Traugh, N., Chen, Q., Liu, J. S., Li, B., and Liu, X. S. (2017c). TIMER: A web server for comprehensive analysis of Tumor-Infiltrating immune cells. *Cancer Res.*, 77(21):e108–e110.

Liao, F., Rabin, R. L., Yannelli, J. R., Koniaris, L. G., Vanguri, P., and Farber, J. M. (1995). Human Mig chemokine: biochemical and functional characterization. *The Journal of experimental medicine*, 182(5):1301–14.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J., and Tamayo, P. (2015). The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Systems*, 1(6):417–425.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J. P. (2011). Molecular signatures database (MSigDB) 3.0. *Bioinformatics*, 27(12):1739–1740.

Lin, H. H., Zhang, G. L., Tongchusak, S., Reinherz, E. L., and Brusic, V. (2008). Evaluation of MHC-II peptide binding prediction servers: applications for vaccine research. *BMC bioinformatics*, 9 Suppl 12:S22.

Lindner, D. (2014). Animal Models and the Tumor Microenvironment: Studies of Tumor–Host Symbiosis. *Seminars in Oncology*, 41(2):146–155.

Lowry, L. E. and Zehring, W. A. (2017). Potentiation of Natural Killer Cells for Cancer Immunotherapy: A Review of Literature. *Frontiers in Immunology*, 8:1061.

Lundegaard, C., Lamberth, K., Harndahl, M., Buus, S., Lund, O., and Nielsen, M. (2008a). NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. *Nucleic acids research*, 36(Web Server issue):509–512.

Lundegaard, C., Lund, O., and Nielsen, M. (2008b). Accurate approximation method for prediction of class I MHC affinities for peptides of length 8, 10 and 11 using prediction tools trained on 9mers. *Bioinformatics*, 24(11):1397–1398.

Macintyre, G., Goranova, T. E., De Silva, D., Ennis, D., Piskorz, A. M., Eldridge, M., Sie, D., Lewsley, L.-A., Hanif, A., Wilson, C., Dowson, S., Glasspool, R. M., Lockley, M., Brockbank, E., Montes, A., Walther, A., Sundar, S., Edmondson, R., Hall, G. D., Clamp, A., Gourley, C., Hall, M., Fotopoulou, C., Gabra, H., Paul, J., Supernat, A., Millan, D., Hoyle, A., Bryson, G., Nourse, C., Mincarelli, L., Sanchez, L. N., Ylstra, B., Jimenez-Linan, M., Moore, L., Hofmann, O., Markowetz, F., McNeish, I. A., and Brenton, J. D. (2018). Copy number signatures and mutational processes in ovarian carcinoma. *Nature Genetics*.

Mamitsuka, H. (1998). Predicting peptides that bind to MHC molecules using supervised learning of hidden Markov models. *Proteins*, 33(4):460–474.

Manning, T., Sleator, R. D., and Walsh, P. (2013). Biologically inspired intelligent decision making: A commentary on the use of artificial neural networks in bioinformatics. *Bioengineered*, 5(2).

Mariathasan, S., Turley, S. J., Nickles, D., Castiglioni, A., Yuen, K., Wang, Y., Kadel, III, E. E., Koeppen, H., Astarita, J. L., Cubas, R., Jhunjhunwala, S., Banchereau, R., Yang, Y., Guan, Y., Chalouni, C., Ziai, J., Şenbabaoğlu, Y., Santoro, S., Sheinson, D., Hung, J., Giltnane, J. M., Pierce, A. A., Mesh, K., Lianoglou, S., Riegler, J., Carano, R. A. D., Eriksson, P., Höglund, M., Somarriba, L., Halligan, D. L., van der Heijden, M. S., Loriot, Y., Rosenberg, J. E., Fong, L., Mellman, I., Chen, D. S., Green, M., Derleth, C., Fine, G. D., Hegde, P. S., Bourgon, R., and Powles, T. (2018). TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. *Nature*, 554(7693):544–548.

Martinez, F. O. and Gordon, S. (2014). The M1 and M2 paradigm of macrophage activation: time for reassessment. *F1000prime reports*, 6:13.

Matsushita, H., Vesely, M. D., Koboldt, D. C., Rickert, C. G., Uppaluri, R., Magrini, V. J., Arthur, C. D., White, J. M., Chen, Y.-S., Shea, L. K., Hundal, J., Wendl, M. C., Demeter, R., Wylie, T., Allison, J. P., Smyth, M. J., Old, L. J., Mardis, E. R., and Schreiber, R. D. (2012). Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting. *Nature*, 482(7385):400–404.

Mattner, J. and Wirtz, S. (2017). Friend or Foe? The Ambiguous Role of Innate Lymphoid Cells in Cancer Development. *Trends in Immunology*, 38(1):29–38.

Matulonis, U. A., Sood, A. K., Fallowfield, L., Howitt, B. E., Sehouli, J., and Karlan, B. Y. (2016). Ovarian cancer. *Nature Reviews Disease Primers*, 2:16061.

McGranahan, N., Furness, A. J. S., Rosenthal, R., Ramskov, S., Lyngaa, R., Saini, S. K., Jamal-Hanjani, M., Wilson, G. A., Birkbak, N. J., Hiley, C. T., Watkins, T. B. K., Shafi, S., Murugaesu, N., Mitter, R., Akarca, A. U., Linares, J., Marafioti, T., Henry, J. Y., Van Allen, E. M., Miao, D., Schilling, B., Schadendorf, D., Garraway, L. A., Makarov, V., Rizvi, N. A., Snyder, A., Hellmann, M. D., Merghoub, T., Wolchok, J. D., Shukla, S. A., Wu, C. J., Peggs, K. S., Chan, T. A., Hadrup,

S. R., Quezada, S. A., Swanton, C., and Allen, E. M. V. (2016). Immune Checkpoint Blockade. *Science*, 351(November 2015):1463–1469.

McGranahan, N. and Swanton, C. (2017). Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell*, 168(4):613–628.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., and DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9):1297–1303.

McPherson, A., Roth, A., Laks, E., Masud, T., Bashashati, A., Zhang, A. W., Ha, G., Biele, J., Yap, D., Wan, A., Prentice, L. M., Khattra, J., Smith, M. A., Nielsen, C. B., Mullaly, S. C., Kalloger, S., Karnezis, A., Shumansky, K., Siu, C., Rosner, J., Chan, H. L., Ho, J., Melnyk, N., Senz, J., Yang, W., Moore, R., Mungall, A. J., Marra, M. A., Bouchard-Côté, A., Gilks, C. B., Huntsman, D. G., McAlpine, J. N., Aparicio, S., and Shah, S. P. (2016). Divergent modes of clonal spread and intraperitoneal mixing in high-grade serous ovarian cancer. *Nat. Genet.*, 48(7):758–767.

Melero, I., Rouzaut, A., Motz, G. T., and Coukos, G. (2014). T-cell and NK-cell infiltration into solid tumors: A key limiting factor for efficacious cancer immunotherapy. *Cancer Discovery*, 4(5):522–526.

Miao, D., Margolis, C. A., Vokes, N. I., Liu, D., Taylor-Weiner, A., Wankowicz, S. M., Adeegbe, D., Keliher, D., Schilling, B., Tracy, A., Manos, M., Chau, N. G., Hanna, G. J., Polak, P., Rodig, S. J., Signoretti, S., Sholl, L. M., Engelman, J. A., Getz, G., Jänne, P. A., Haddad, R. I., Choueiri, T. K., Barbie, D. A., Haq, R., Awad, M. M., Schadendorf, D., Hodi, F. S., Bellmunt, J., Wong, K.-K., Hammerman, P., and Van Allen, E. M. (2018). Genomic correlates of response to immune checkpoint blockade in microsatellite-stable solid tumors. *Nature Genetics*, 50(9):1271–1281.

Miao, D. and Van Allen, E. M. (2016). Genomic determinants of cancer immunotherapy. *Current Opinion in Immunology*, 41:32–38.

Mittal, D., Gubin, M. M., Schreiber, R. D., and Smyth, M. J. (2014). New insights into cancer immunoediting and its three component phases-elimination, equilibrium and escape.

Mlecnik, B., Bindea, G., Angell, H., Maby, P., Angelova, M., Tougeron, D., Church, S., Lafontaine, L., Fischer, M., Fredriksen, T., Sasso, M., Bilocq, A., Kirilovsky, A., Obenauf, A., Hamieh, M., Berger, A., Bruneval, P., Tuech, J.-J., Sabourin, J.-C., LePessot, F., Mauillon, J., Rafii, A., Laurent-Puig, P., Speicher, M., Trajanoski, Z., Michel, P., Sesboüe, R., Frebourg, T., Pagès, F., Valge-Archer, V., Latouche, J.-B., and Galon, J. (2016). Integrative Analyses of Colorectal Cancer Show Immunoscore Is a Stronger Predictor of Patient Survival Than Microsatellite Instability. *Immunity*, 44(3):698–711.

Mohan, J. F. and Unanue, E. R. (2012). Unconventional recognition of peptides by T cells and the implications for autoimmunity. *Nature Reviews Immunology*, 12(10):721–728.

Moutaftsi, M., Peters, B., Pasquetto, V., Tscharke, D. C., Sidney, J., Bui, H.-H., Grey, H., and Sette, A. (2006). A consensus epitope prediction approach identifies the breadth of murine T(CD8+)-cell responses to vaccinia virus. *Nature biotechnology*, 24(7):817–9.

Muñoz, I. M., MacArtney, T., Sanchez-Pulido, L., Ponting, C. P., Rocha, S., and Rouse, J. (2012). Family with sequence similarity 60A (FAM60A) protein is a cell cycle-fluctuating regulator of the SIN3-HDAC1 histone deacetylase complex. *J. Biol. Chem.*, 287(39):32346–32353.

Murugaesu, N., Wilson, G. A., Birkbak, N. J., Watkins, T. B. K., McGranahan, N., Kumar, S., Abbassi-Ghadi, N., Salm, M., Mitter, R., Horswell, S., Rowan, A., Phillimore, B., Biggs, J., Begum, S., Matthews, N., Hochhauser, D., Hanna, G. B., and Swanton, C. (2015). Tracking the genomic evolution of esophageal adenocarcinoma through neoadjuvant chemotherapy. *Cancer Discovery*, 5(8):821–832.

Natrajan, R., Sailem, H., Mardakheh, F. K., Arias Garcia, M., Tape, C. J., Dowsett, M., Bakal, C., and Yuan, Y. (2016). Microenvironmental heterogeneity parallels breast cancer progression: A Histology-Genomic integration analysis. *PLoS Med.*, 13(2):e1001961.

Neefjes, J., Jongsma, M. L. M., Paul, P., and Bakke, O. (2011). Towards a systems understanding of MHC class I and MHC class II antigen presentation. *Nature Reviews Immunology*, 11(12):823–836.

Neefjes, J. J., Momburg, F., and Hammerling, G. J. (1993). Selective and ATP dependent translocation of peptides by the MHC encoded transporter. *Science*, 261(May):769–771.

Newman, A. M., Gentles, A. J., Liu, C. L., Diehn, M., and Alizadeh, A. A. (2017). Data normalization considerations for digital tumor dissection. *Genome Biology*, 18(1):128.

Newman, A. M., Liu, C. L., Green, M. R., Gentles, A. J., Feng, W., Xu, Y., Hoang, C. D., Diehn, M., and Alizadeh, A. a. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nature methods*, 12(MAY 2014):1–10.

Nielsen, M. and Lund, O. (2009). NN-align. An artificial neural network-based alignment algorithm for MHC class II peptide binding prediction. *BMC bioinformatics*, 10:296.

Nielsen, M., Lundegaard, C., Blicher, T., Lamberth, K., Harndahl, M., Justesen, S., Røder, G., Peters, B., Sette, A., Lund, O., and Buus, S. (2007). NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS ONE*, 2(8).

Nielsen, M., Lundegaard, C., Worning, P., Lauemøller, S. L., Lamberth, K., Buus, S., Brunak, S., and Lund, O. (2003). Reliable prediction of T-cell epitopes using neural networks with novel sequence representations. *Protein science : a publication of the Protein Society*, 12(5):1007–1017.

Nielsen, M., Lundegaard, C., Worning, P., Sylvester Hvid, C., Lamberth, K., Buus, S., Brunak, S., and Lund, O. (2004). Improved prediction of MHC class I and class II epitopes using a novel Gibbs sampling approach. *Bioinformatics*, 20(9):1388–1397.

Nixon, K. C. (1999). The Parsimony Ratchet, a New Method for Rapid Parsimony Analysis. *Cladistics*, 15:407–414.

Ocana, A., Nieto-Jiménez, C., Pandiella, A., and Templeton, A. J. (2017). Neutrophils in cancer: prognostic role and therapeutic strategies. *Molecular Cancer*, 16(1):137.

Oliphant, T. E. (2007). Python for scientific computing. *Comput. Sci. Eng.*, 9(3):10–20.

O'sullivan, T. E., Sun, J. C., and Lanier, L. L. (2015). Review Natural Killer Cell Memory. *Immunity*, 43:634–645.

Ovarian Tumor Tissue Analysis (OTTA) Consortium, Goode, E. L., Block, M. S., Kalli, K. R., Vierkant, R. A., Chen, W., Fogarty, Z. C., Gentry-Maharaj, A., Tołoczko, A., Hein, A., Bouligny, A. L., Jensen, A., Osorio, A., Hartkopf, A., Ryan, A., Chudecka-Głaz, A., Magliocco, A. M., Hartmann, A., Jung, A. Y., Gao, B., Hernandez, B. Y., Fridley, B. L., McCauley, B. M., Kennedy, C. J., Wang, C., Karpinskyj, C., de Sousa, C. B., Tiezzi, D. G., Wachter, D. L., Herpel, E., Taran,

F. A., Modugno, F., Nelson, G., Lubiński, J., Menkiszak, J., Alsop, J., Lester, J., García-Donas, J., Nation, J., Hung, J., Palacios, J., Rothstein, J. H., Kelley, J. L., de Andrade, J. M., Robles-Díaz, L., Intermaggio, M. P., Widschwendter, M., Beckmann, M. W., Ruebner, M., Jimenez-Linan, M., Singh, N., Oszurek, O., Harnett, P. R., Rambau, P. F., Sinn, P., Wagner, P., Ghatage, P., Sharma, R., Edwards, R. P., Ness, R. B., Orsulic, S., Brucker, S. Y., Johnatty, S. E., Longacre, T. A., Ursula, E., McGuire, V., Sieh, W., Natanzon, Y., Li, Z., Whittemore, A. S., Anna, D., Staebler, A., Karlan, B. Y., Gilks, B., Bowtell, D. D., Høgdall, E., Candido dos Reis, F. J., Steed, H., Campbell, I. G., Gronwald, J., Benítez, J., Koziak, J. M., Chang-Claude, J., Moysich, K. B., Kelemen, L. E., Cook, L. S., Goodman, M. T., García, M. J., Fasching, P. A., Kommoss, S., Deen, S., Kjaer, S. K., Menon, U., Brenton, J. D., Pharoah, P. D. P., Chenevix-Trench, G., Huntsman, D. G., Winham, S. J., Köbel, M., and Ramus, S. J. (2017). Dose-Response association of CD8+ Tumor-Infiltrating lymphocytes and survival time in High-Grade serous ovarian cancer. *JAMA Oncol*, 3(12):e173290.

Parkin, J. and Cohen, B. (2001). An overview of the immune system. *Lancet*, 357(9270):1777–1789.

Patch, A.-M., Christie, E. L., Etemadmoghadam, D., Garsed, D. W., George, J., Fereday, S., Nones, K., Cowin, P., Alsop, K., Bailey, P. J., Kassahn, K. S., Newell, F., Quinn, M. C. J., Kazakoff, S., Quek, K., Wilhelm-Benartzi, C., Curry, E., Leong, H. S., Hamilton, A., Mileshkin, L., Au-Yeung, G., Kennedy, C., Hung, J., Chiew, Y.-E., Harnett, P., Friedlander, M., Quinn, M., Pyman, J., Cordner, S., O'Brien, P., Leditschke, J., Young, G., Strachan, K., Waring, P., Azar, W., Mitchell, C., Traficante, N., Hendley, J., Thorne, H., Shackleton, M., Miller, D. K., Arnau, G. M., Tothill, R. W., Holloway, T. P., Semple, T., Harliwong, I., Nourse, C., Nourbakhsh, E., Manning, S., Idrisoglu, S., Bruxner, T. J. C., Christ, A. N., Poudel, B., Holmes, O., Anderson, M., Leonard, C., Lonie, A., Hall, N., Wood, S., Taylor, D. F., Xu, Q., Fink, J. L., Waddell, N., Drapkin, R., Stronach, E., Gabra, H., Brown, R., Jewell, A., Nagaraj, S. H., Markham, E., Wilson, P. J., Ellul, J., McNally, O., Doyle, M. a., Vedururu, R., Stewart, C., Lengyel, E., Pearson, J. V., Waddell, N., DeFazio, A., Grimmond, S. M., and Bowtell, D. D. L. (2015). Whole–genome characterization of chemoresistant ovarian cancer. *Nature*, 521(7553):489–494.

Paul, S., Weiskopf, D., Angelo, M. A., Sidney, J., Peters, B., and Sette, A. (2013). HLA Class I Alleles Are Associated with Peptide-Binding Repertoires of Different Size, Affinity, and Immunogenicity. *The Journal of Immunology*, 191(12):5831–5839.

Peters, B., Bui, H. H., Frankild, S., Nielsen, M., Lundegaard, C., Kostem, E., Basch, D., Lamberth, K., Harndahl, M., Fleri, W., Wilson, S. S., Sidney, J., Lund, O., Buus, S., and Sette, A. (2006). A community resource benchmarking predictions of peptide binding to MHC-I molecules. *PLoS Computational Biology*, 2(6):0574–0584.

Peters, B. and Sette, A. (2005). Generating quantitative models describing the sequence specificity of biological processes with the stabilized matrix method. *BMC bioinformatics*, 6(1):132.

Pfirschke, C., Engblom, C., Rickelt, S., Cortez-Retamozo, V., Garris, C., Pucci, F., Yamazaki, T., Poirier-Colame, V., Newton, A., Redouane, Y., Lin, Y. J., Wojtkiewicz, G., Iwamoto, Y., Mino-Kenudson, M., Huynh, T. G., Hynes, R. O., Freeman, G. J., Kroemer, G., Zitvogel, L., Weissleder, R., and Pittet, M. J. (2016). Immunogenic Chemotherapy Sensitizes Tumors to Checkpoint Blockade Therapy. *Immunity*, 44(2):343–354.

Pielou, E. C. (1966). Species-diversity and pattern-diversity in the study of ecological succession. *J. Theor. Biol.*, 10(2):370–383.

Pietras, K. and Östman, A. (2010). Hallmarks of cancer: Interactions with the tumor stroma.

Pinilla, S., Alt, E., Abdul Khalek, F., Jotzu, C., Muehlberg, F., Beckmann, C., and Song, Y.-H. (2009). Tissue resident stem cells produce CCL5 under the influence of cancer cells and thereby promote breast cancer cell invasion. *Cancer Letters*, 284(1):80–85.

Powis, S., Townsend, A., Deverson, E., Bastin, J., Butcher, G., and Howard, J. (1991). Restoration of antigen presentation to the mutant cell line RMA-S by an MHC-linked transporter. *Nature*, 354(6354):528–31.

Puram, S. V., Tirosh, I., Parikh, A. S., Patel, A. P., Yizhak, K., Gillespie, S., Rodman, C., Luo, C. L., Mroz, E. A., Emerick, K. S., Deschler, D. G., Varvares, M. A., Mylvaganam, R., Rozenblatt-Rosen, O., Rocco, J. W., Faquin, W. C., Lin, D. T., Regev, A., and Bernstein, B. E. (2017). Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell*, 171(7):1611–1624.e24.

Rainczuk, A., Rao, J., Gathercole, J., and Stephens, A. N. (2012). The emerging role of CXC chemokines in epithelial ovarian cancer. *Reproduction*, 144(3):303–317.

Rakhra, K., Bachireddy, P., Zabuawala, T., Zeiser, R., Xu, L., Kopelman, A., Fan, A. C., Yang, Q., Braunstein, L., Crosby, E., Ryeom, S., and Felsher, D. W. (2010). CD4(+) T cells contribute to the remodeling of the microenvironment required for sustained tumor regression upon oncogene inactivation. *Cancer Cell*, 18(5):485–498.

Rammensee, H. G., Bachmann, J., Emmerich, N. P. N., Bachor, O. a., and Stevanović, S. (1999). SYFPEITHI: database for MHC ligands and peptide motifs. *Immunogenetics*, 50(3-4):213–219.

Raza, A., Franklin, M. J., and Dudek, A. Z. (2010). Pericytes and vessel maturation during tumor angiogenesis and metastasis. *American Journal of Hematology*, 85(8):593–598.

Reche, P. a., Glutting, J. P., and Reinherz, E. L. (2002). Prediction of MHC class I binding peptides using profile motifs. *Human Immunology*, 63(9):701–709.

Reits, E. a. J., Vos, J. C., Grommé, M., and Neefjes, J. (2000). The major substrates for TAP in vivo are derived from newly synthesized proteins. *Nature*, 404(6779):774–8.

Remark, R., Alifano, M., Cremer, I., Lupo, A., Dieu-Nosjean, M. C., Riquet, M., Crozet, L., Ouakrim, H., Goc, J., Cazes, A., Fléjou, J. F., Gibault, L., Verkarre, V., Régnard, J. F., Pagés, O. N., Oudard, S., Mlecnik, B., Sautés-Fridman, C., Fridman, W. H., and Damotte, D. (2013). Characteristics and clinical impacts of the immune environments in colorectal and renal cell carcinoma lung metastases: Influence of tumor origin. *Clinical Cancer Research*, 19(15):4079–4091.

Reuben, A., Spencer, C. N., Prieto, P. A., Gopalakrishnan, V., Reddy, S. M., Miller, J. P., Mao, X., De Macedo, M. P., Chen, J., Song, X., Jiang, H., Chen, P.-L., Beird, H. C., Garber, H. R., Roh, W., Wani, K., Chen, E., Haymaker, C., Forget, M.-A., Little, L. D., Gumbs, C., Thornton, R. L., Hudgens, C. W., Chen, W.-S., Austin-Breneman, J., Sloane, R. S., Nezi, L., Cogdill, A. P., Bernatchez, C., Roszik, J., Hwu, P., Woodman, S. E., Chin, L., Tawbi, H., Davies, M. A., Gershenwald, J. E., Amaria, R. N., Glitza, I. C., Diab, A., Patel, S. P., Hu, J., Lee, J. E., Grimm, E. A., Tetzlaff, M. T., Lazar, A. J., Wistuba, I. I., Clise-Dwyer, K., Carter, B. W., Zhang, J., Futreal, P. A., Sharma, P., Allison, J. P., Cooper, Z. A., and Wargo, J. A. (2017). Genomic and immune heterogeneity are associated with differential responses to therapy in melanoma. *npj Genomic Medicine*, 2(1):10.

Ribas, A. and Wolchok, J. D. (2018). Cancer immunotherapy using checkpoint blockade. *Science*, 359(6382):1350–1355.

Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., and Smyth, G. K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, 43(7):e47.

Rizvi, N. A., Hellmann, M. D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J. J., Lee, W., Yuan, J., Wong, P., Ho, T. S., Miller, M. L., Rekhtman, N., Moreira, A. L., Ibrahim, F., Bruggeman, C., Gasmi, B., Zappasodi, R., Maeda, Y., Sander, C., Garon, E. B., Merghoub, T., Wolchok, J. D., Schumacher, T. N., and Chan, T. A. (2014). Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science*, 348(March):1–10.

Robinson, D. R., Wu, Y.-M., Lonigro, R. J., Vats, P., Cobain, E., Everett, J., Cao, X., Rabban, E., Kumar-Sinha, C., Raymond, V., Schuetze, S., Alva, A., Siddiqui, J., Chugh, R., Worden, F., Zalupski, M. M., Innis, J., Mody, R. J., Tomlins, S. A., Lucas, D., Baker, L. H., Ramnath, N., Schott, A. F., Hayes, D. F., Vijai, J., Offit, K., Stoffel, E. M., Roberts, J. S., Smith, D. C., Kunju, L. P., Talpaz, M., Cieślik, M., and Chinnaiyan, A. M. (2017). Integrative clinical genomics of metastatic cancer. *Nature*, 548(7667):297–303.

Robinson, J. T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E. S., Getz, G., and Mesirov, J. P. (2011). Integrative genomics viewer. *Nature Biotechnology*, 29(1):24–26.

Roomp, K., Antes, I., and Lengauer, T. (2010). Predicting MHC class I epitopes in large datasets. *BMC bioinformatics*, 11:90.

Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell*, 160(1-2):48–61.

Rosenberg, S. a. (2014). IL-2: The First Effective Immunotherapy for Human Cancer. *Journal of immunology (Baltimore, Md. : 1950)*, 192(12):5451–8.

Rosenberg, S. A. and Restifo, N. P. (2015). Adoptive cell transfer as personalized immunotherapy for human cancer. *Science*, 348(6230):62–68.

Rosenberg, S. a., Yang, J. C., Sherry, R. M., Kammula, U. S., Hughes, M. S., Phan, G. Q., Citrin, D. E., Restifo, N. P., Robbins, P. F., Wunderlich, J. R., Morton, K. E., Laurencot, C. M., Steinberg, S. M., White, D. E., and Dudley, M. E. (2011). Durable complete responses in heavily pretreated patients with metastatic melanoma using T-cell transfer immunotherapy. *Clinical Cancer Research*, 17(13):4550–4557.

Roth, A., Khattra, J., Yap, D., Wan, A., Laks, E., Biele, J., Ha, G., Aparicio, S., Bouchard-Côté, A., and Shah, S. P. (2014). PyClone: statistical inference of clonal population structure in cancer. *Nature methods*, 11(4):396–8.

Rüegsegger, U., Blank, D., and Keller, W. (1998). Human Pre-mRNA cleavage factor im is related to spliceosomal SR proteins and can be reconstituted in vitro from recombinant subunits. *Mol. Cell*, 1(2):243–253.

Sala, E., Kataoka, M. Y., Priest, a. N., Gill, a. B., McLean, M. a., Joubert, I., Graves, M. J., Crawford, R. a. F., Jimenez-Linan, M., Earl, H. M., Hodgkin, C., Griffiths, J. R., Lomas, D. J., and Brenton, J. D. (2012). Advanced Ovarian Cancer: Multiparametric MR Imaging Demonstrates Response- and Metastasis-specific Effects. *Radiology*, 263(1):149–159.

Saltz, J., Gupta, R., Hou, L., Kurc, T., Singh, P., Nguyen, V., Samaras, D., Shroyer, K. R., Zhao, T., Batiste, R., Van Arnam, J., Cancer Genome Atlas Research Network, Shmulevich, I., Rao, A. U. K., Lazar, A. J., Sharma, A., and Thorsson, V. (2018). Spatial organization and molecular

correlation of Tumor-Infiltrating lymphocytes using deep learning on pathology images. *Cell Rep.*, 23(1):181–193.e7.

Samuels, D. C., Han, L., Li, J., Quanghu, S., Clark, T. A., Shyr, Y., and Guo, Y. (2013). Finding the lost treasures in exome sequencing data. *Trends in Genetics*, 29(10):593–599.

Satoh, J. I. and Tabunoki, H. (2013). A comprehensive profile of ChIP-Seq-based STAT1 target genes suggests the complexity of STAT1-mediated gene regulatory mechanisms. *Gene Regulation and Systems Biology*, 2013(7):41–56.

Schafroth, H. D. and Floudas, C. A. (2004). Predicting Peptide Binding to MHC Pockets via Molecular Modeling, Implicit Solvation, and Global Optimization. *Proteins: Structure, Function and Genetics*, 54(3):534–556.

Scheinin, I., Sie, D., Bengtsson, H., van de Wiel, M. A., Olshen, A. B., van Thuijl, H. F., van Essen, H. F., Eijk, P. P., Rustenburg, F., Meijer, G. A., Reijneveld, J. C., Wesseling, P., Pinkel, D., Albertson, D. G., and Ylstra, B. (2014). DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Research*, 24(12):2022–2032.

Schiewe, A. J. and Haworth, I. S. (2007). Structure-based prediction of MHC-peptide association: Algorithm comparison and application to cancer vaccine design. *Journal of Molecular Graphics and Modelling*, 26(3):667–675.

Schliep, K. P. (2011). phangorn: Phylogenetic analysis in R. *Bioinformatics*, 27(4):592–593.

Schreiber, R. D., Old, L. J., and Smyth, M. J. (2011). Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. *Science (New York, N.Y.)*, 331(6024):1565–1570.

Schubert, U., Antón, L. C., Gibbs, J., Norbury, C. C., Yewdell, J. W., and Bennink, J. R. (2000). Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. *Nature*, 404(6779):770–774.

Schumacher, T. N. and Schreiber, R. D. (2015). Neoantigens in cancer immunotherapy. *Science*, 348(6230):69–73.

Segal, N. H., Parsons, D. W., Peggs, K. S., Velculescu, V., Kinzler, K. W., Vogelstein, B., and Allison, J. P. (2008). Epitope landscape in breast and colorectal cancer. *Cancer Research*, 68(3):889–892.

Shah, S. P., Morin, R. D., Khattra, J., Prentice, L., Pugh, T., Burleigh, A., Delaney, A., Gelmon, K., Guliany, R., Senz, J., Steidl, C., Holt, R. A., Jones, S., Sun, M., Leung, G., Moore, R., Severson, T., Taylor, G. A., Teschendorff, A. E., Tse, K., Turashvili, G., Varhol, R., Warren, R. L., Watson, P., Zhao, Y., Caldas, C., Huntsman, D., Hirst, M., Marra, M. A., and Aparicio, S. (2009). Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature*, 461(7265):809–813.

Shapouri-Moghaddam, A., Mohammadian, S., Vazini, H., Taghadosi, M., Esmaeili, S.-A., Mardani, F., Seifi, B., Mohammadi, A., Afshari, J. T., and Sahebkar, A. (2018). Macrophage plasticity, polarization, and function in health and disease. *Journal of Cellular Physiology*, 233(9):6425–6440.

Sharma, P. and Allison, J. P. (2015). The future of immune checkpoint therapy. *Science*, 348(6230):56–61.

Sharma, P., Hu-Lieskovan, S., Wargo, J. A., and Ribas, A. (2017). Primary, Adaptive, and Acquired Resistance to Cancer Immunotherapy. *Cell*, 168(4):707–723.

Shiina, T., Hosomichi, K., Inoko, H., and Kulski, J. K. (2009). The HLA genomic loci map: expression, interaction, diversity and disease. *Journal of human genetics*, 54(1):15–39.

Shukla, S. a., Rooney, M. S., Rajasagi, M., Tiao, G., Dixon, P. M., Lawrence, M. S., Stevens, J., Lane, W. J., Dellagatta, J. L., Steelman, S., Sougnez, C., Cibulskis, K., Kiezun, A., Hacohen, N., Brusic, V., Wu, C. J., and Getz, G. (2015). Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nature biotechnology*, 33(11).

Sidney, J., Assarsson, E., Moore, C., Ngo, S., Pinilla, C., Sette, A., and Peters, B. (2008). Quantitative peptide binding motifs for 19 human and mouse MHC class I molecules derived using positional scanning combinatorial peptide libraries. *Immunome Research*, 4:2–15.

Sivan, A., Corrales, L., Hubert, N., Williams, J. B., Aquino-Michaels, K., Earley, Z. M., Benyamin, F. W., Man Lei, Y., Jabri, B., Alegre, M.-L., Chang, E. B., and Gajewski, T. F. (2015). Commensal Bifidobacterium promotes antitumor immunity and facilitates anti-PD-L1 efficacy. *Science*, 350(6264):1084–1089.

Smith, K. T., Sardiu, M. E., Martin-Brown, S. A., Seidel, C., Mushegian, A., Egidy, R., Florens, L., Washburn, M. P., and Workman, J. L. (2012). Human family with sequence similarity 60 member a (FAM60A) protein: a new subunit of the sin3 deacetylase complex. *Mol. Cell. Proteomics*, 11(12):1815–1828.

Smith-Garvin, J. E., Koretzky, G. A., and Jordan, M. S. (2009). T cell activation. *Annual review of immunology*, 27:591–619.

Snyder, A., Makarov, V., Merghoub, T., Yuan, J., Zaretsky, J. M., Desrichard, A., Walsh, L. a., Postow, M. a., Wong, P., Ho, T. S., Hollmann, T. J., Bruggeman, C., Kannan, K., Li, Y., Elipenahli, C., Liu, C., Harbison, C. T., Wang, L., Ribas, A., Wolchok, J. D., and Chan, T. a. (2014). Genetic Basis for Clinical Response to CTLA-4 Blockade in Melanoma. *The New England journal of medicine*, 371(23):141119140020009.

Spranger, S., Bao, R., and Gajewski, T. F. (2015). Melanoma-intrinsic β-catenin signalling prevents anti-tumour immunity. *Nature*, 523(7559):231–235.

Spranger, S., Dai, D., Horton, B., and Gajewski, T. F. (2017). Tumor-Residing Batf3 Dendritic Cells Are Required for Effector T Cell Trafficking and Adoptive T Cell Therapy. *Cancer Cell*, 31(5):711–723.e4.

Spranger, S. and Gajewski, T. F. (2018). Impact of oncogenic pathways on evasion of antitumour immune responses. *Nat. Rev. Cancer*.

Spranger, S., Luke, J. J., Bao, R., Zha, Y., Hernandez, K. M., Li, Y., Gajewski, A. P., Andrade, J., and Gajewski, T. F. (2016). Density of immunogenic antigens does not explain the presence or absence of the t-cell–inflamed tumor microenvironment in melanoma. *Proceedings of the National Academy of Sciences*, 113(48):E7759–E7768.

Sridharan, V., Gjini, E., Liao, X., Chau, N. G., Haddad, R. I., Severgnini, M., Hammerman, P., El-Naggar, A., Freeman, G. J., Hodi, F. S., Rodig, S. J., Dranoff, G., and Schoenfeld, J. D. (2016). Immune Profiling of Adenoid Cystic Carcinoma: PD-L2 Expression and Associations with Tumor-Infiltrating Lymphocytes. *Cancer Immunology Research*, 4(8):679–687.

Stasinopoulos, D. M., Mikis Stasinopoulos, D., and Rigby, R. A. (2007). Generalized additive models for location scale and shape (GAMLSS) inr. *J. Stat. Softw.*, 23(7).

Sturniolo, T., Bono, E., Ding, J., Raddrizzani, L., Tuereci, O., Sahin, U., Braxenthaler, M., Gallazzi, F., Protti, M. P., Sinigaglia, F., and Hammer, J. (1999). Generation of tissue-specific and promiscuous HLA ligand databases using DNA microarrays and virtual HLA class II matrices. *Nature biotechnology*, 17(6):555–561.

Su, K. Y., Chen, H. Y., Li, K. C., Kuo, M. L., Yang, J. C. H., Chan, W. K., Ho, B. C., Chang, G. C., Shih, J. Y., Yu, S. L., and Yang, P. C. (2012). Pretreatment Epidermal Growth Factor Receptor (EGFR) T790M mutation predicts shorter EGFR tyrosine kinase inhibitor response duration in patients with non-small-cell lung cancer. *Journal of Clinical Oncology*, 30(4):433–440.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, 102(43):15545–50.

Suh, W.-K., Cohen-Doyle, M. F., Fruh, K., Wang, K., Peterson, P. A., and Williams, D. B. (1994). Interaction of MHC class I molecules with the transporter associated with antigen processing. *Science*, 264(April):1322–1326.

Suzuki, E., Kapoor, V., Jassar, A. S., Kaiser, L. R., and Albelda, S. M. (2005). Gemcitabine selectively eliminates splenic Gr-1+/CD11b + myeloid suppressor cells in tumor-bearing animals and enhances antitumor immune activity. *Clinical Cancer Research*, 11(18):6713–6721.

Swierczak, A., Mouchemore, K. A., Hamilton, J. A., and Anderson, R. L. (2015). Neutrophils: important contributors to tumor progression and metastasis. *Cancer and Metastasis Reviews*, 34(4):735–751.

Swisher, E. M., Lin, K. K., Oza, A. M., Scott, C. L., Giordano, H., Sun, J., Konecny, G. E., Coleman, R. L., Tinker, A. V., O'Malley, D. M., Kristeleit, R. S., Ma, L., Bell-McGuinn, K. M., Brenton, J. D., Cragun, J. M., Oaknin, A., Ray-Coquard, I., Harrell, M. I., Mann, E., Kaufmann, S. H., Floquet, A., Leary, A., Harding, T. C., Goble, S., Maloney, L., Isaacson, J., Allen, A. R., Rolfe, L., Yelensky, R., Raponi, M., and McNeish, I. A. (2017). Rucaparib in relapsed, platinum-sensitive high-grade ovarian carcinoma (ARIEL2 Part 1): an international, multicentre, open-label, phase 2 trial. *The Lancet Oncology*, 18(1):75–87.

Szolek, A., Schubert, B., Mohr, C., Sturm, M., Feldhahn, M., and Kohlbacher (2014). OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics (Oxford, England)*, 30(23):3310–3316.

Tan, R., Wang, Y., Kleinstein, S. E., Liu, Y., Zhu, X., Guo, H., Jiang, Q., Allen, A. S., and Zhu, M. (2014). An Evaluation of Copy Number Variation Detection Tools from Whole-Exome Sequencing Data. *Human Mutation*, 35(7):899–907.

Tauriello, D. V. F., Palomo-Ponce, S., Stork, D., Berenguer-Llergo, A., Badia-Ramentol, J., Iglesias, M., Sevillano, M., Ibiza, S., Cañellas, A., Hernando-Momblona, X., Byrom, D., Matarin, J. A., Calon, A., Rivas, E. I., Nebreda, A. R., Riera, A., Attolini, C. S.-O., and Batlle, E. (2018). TGFβ drives immune evasion in genetically reconstituted colon cancer metastasis. *Nature*, 554(7693):538–543.

TCGA (2011). Integrated genomic analyses of ovarian carcinoma. *Nature*, 474(7353):609–615.

Teng, M. W. L., Galon, J., Fridman, W. H., and Smyth, M. J. (2015). From mice to humans: Developments in cancer immunoediting. *Journal of Clinical Investigation*, 125(9):3338–3346.

Tesone, A. J., Rutkowski, M. R., Brencicova, E., Svoronos, N., Perales-Puchalt, A., Stephen, T. L., Allegrezza, M. J., Payne, K. K., Nguyen, J. M., Wickramasinghe, J., Tchou, J., Borowsky, M. E., Rabinovich, G. A., Kossenkov, A. V., and Conejo-Garcia, J. R. (2016). Satb1 Overexpression Drives Tumor-Promoting Activities in Cancer-Associated Dendritic Cells. *Cell Reports*, 14(7):1774–1786.

The Gene Ontology Consortium (2017). Expansion of the gene ontology knowledgebase and resources. *Nucleic Acids Res.*, 45(D1):D331–D338.

Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Ou Yang, T.-H., Porta-Pardo, E., Gao, G. F., Plaisier, C. L., Eddy, J. A., Ziv, E., Culhane, A. C., Paull, E. O., Sivakumar, I. A., Gentles, A. J., Malhotra, R., Farshidfar, F., Colaprico, A., Parker, J. S., Mose, L. E., Vo, N. S., Liu, J., Liu, Y., Rader, J., Dhankani, V., Reynolds, S. M., Bowlby, R., Califano, A., Cherniack, A. D., Anastassiou, D., Bedognetti, D., Rao, A., Chen, K., Krasnitz, A., Hu, H., Malta, T. M., Noushmehr, H., Pedamallu, C. S., Bullman, S., Ojesina, A. I., Lamb, A., Zhou, W., Shen, H., Choueiri, T. K., Weinstein, J. N., Guinney, J., Saltz, J., Holt, R. A., Rabkin, C. E., Lazar, A. J., Serody, J. S., Demicco, E. G., Disis, M. L., Vincent, B. G., Shmulevich, L., Caesar-Johnson, S. J., Demchok, J. A., Felau, I., Kasapi, M., Ferguson, M. L., Hutter, C. M., Sofia, H. J., Tarnuzzer, R., Wang, Z., Yang, L., Zenklusen, J. C., Zhang, J. J., Chudamani, S., Liu, J., Lolla, L., Naresh, R., Pihl, T., Sun, Q., Wan, Y., Wu, Y., Cho, J., DeFreitas, T., Frazer, S., Gehlenborg, N., Getz, G., Heiman, D. I., Kim, J., Lawrence, M. S., Lin, P., Meier, S., Noble, M. S., Saksena, G., Voet, D., Zhang, H., Bernard, B., Chambwe, N., Dhankani, V., Knijnenburg, T., Kramer, R., Leinonen, K., Liu, Y., Miller, M., Reynolds, S., Shmulevich, I., Thorsson, V., Zhang, W., Akbani, R., Broom, B. M., Hegde, A. M., Ju, Z., Kanchi, R. S., Korkut, A., Li, J., Liang, H., Ling, S., Liu, W., Lu, Y., Mills, G. B., Ng, K.-S., Rao, A., Ryan, M., Wang, J., Weinstein, J. N., Zhang, J., Abeshouse, A., Armenia, J., Chakravarty, D., Chatila, W. K., de Bruijn, I., Gao, J., Gross, B. E., Heins, Z. J., Kundra, R., La, K., Ladanyi, M., Luna, A., Nissan, M. G., Ochoa, A., Phillips, S. M., Reznik, E., Sanchez-Vega, F., Sander, C., Schultz, N., Sheridan, R., Sumer, S. O., Sun, Y., Taylor, B. S., Wang, J., Zhang, H., Anur, P., Peto, M., Spellman, P., Benz, C., Stuart, J. M., Wong, C. K., Yau, C., Hayes, D. N., Parker, J. S., Wilkerson, M. D., Ally, A., Balasundaram, M., Bowlby, R., Brooks, D., Carlsen, R., Chuah, E., Dhalla, N., Holt, R., Jones, S. J., Kasaian, K., Lee, D., Ma, Y., Marra, M. A., Mayo, M., Moore, R. A., Mungall, A. J., Mungall, K., Robertson, A. G., Sadeghi, S., Schein, J. E., Sipahimalani, P., Tam, A., Thiessen, N., Tse, K., Wong, T., Berger, A. C., Beroukhim, R., Cherniack, A. D., Cibulskis, C., Gabriel, S. B., Gao, G. F., Ha, G., Meyerson, M., Schumacher, S. E., Shih, J., Kucherlapati, M. H., Kucherlapati, R. S., Baylin, S., Cope, L., Danilova, L., Bootwalla, M. S., Lai, P. H., Maglinte, D. T., Van Den Berg, D. J., Weisenberger, D. J., Auman, J. T., Balu, S., Bodenheimer, T., Fan, C., Hoadley, K. A., Hoyle, A. P., Jefferys, S. R., Jones, C. D., Meng, S., Mieczkowski, P. A., Mose, L. E., Perou, A. H., Perou, C. M., Roach, J., Shi, Y., Simons, J. V., Skelly, T., Soloway, M. G., Tan, D., Veluvolu, U., Fan, H., Hinoue, T., Laird, P. W., Shen, H., Zhou, W., Bellair, M., Chang, K., Covington, K., Creighton, C. J., Dinh, H., Doddapaneni, H., Donehower, L. A., Drummond, J., Gibbs, R. A., Glenn, R., Hale, W., Han, Y., Hu, J., Korchina, V., Lee, S., Lewis, L., Li, W., Liu, X., Morgan, M., Morton, D., Muzny, D., Santibanez, J., Sheth, M., Shinbrot, E., Wang, L., Wang, M., Wheeler, D. A., Xi, L., Zhao, F., Hess, J., Appelbaum, E. L., Bailey, M., Cordes, M. G., Ding, L., Fronick, C. C., Fulton, L. A., Fulton, R. S., Kandoth, C., Mardis, E. R., McLellan, M. D., Miller, C. A., Schmidt, H. K., Wilson, R. K., Crain, D., Curley, E., Gardner, J., Lau, K., Mallery, D., Morris, S., Paulauskis, J., Penny, R., Shelton, C., Shelton, T., Sherman, M., Thompson, E., Yena, P., Bowen, J., Gastier-Foster, J. M., Gerken, M., Leraas, K. M., Lichtenberg, T. M., Ramirez, N. C., Wise, L., Zmuda, E., Corcoran, N., Costello, T., Hovens, C., Carvalho, A. L., de Carvalho, A. C., Fregnani, J. H., Longatto-Filho, A., Reis, R. M., Scapulatempo-Neto, C., Silveira, H. C., Vidal, D. O., Burnette, A., Eschbacher, J., Hermes, B., Noss, A., Singh, R., Anderson, M. L., Castro, P. D., Ittmann, M., Huntsman, D., Kohl, B., Le, X., Thorp, R., Andry, C., Duffy, E. R., Lyadov, V., Paklina, O., Setdikova, G., Shabunin, A., Tavobilov, M., McPherson, C., Warnick, R., Berkowitz, R., Cramer, D., Feltmate, C., Horowitz, N., Kibel,

A., Muto, M., Raut, C. P., Malykh, A., Barnholtz-Sloan, J. S., Barrett, W., Devine, K., Fulop, J., Ostrom, Q. T., Shimmel, K., Wolinsky, Y., Sloan, A. E., De Rose, A., Giuliante, F., Goodman, M., Karlan, B. Y., Hagedorn, C. H., Eckman, J., Harr, J., Myers, J., Tucker, K., Zach, L. A., Deyarmin, B., Hu, H., Kvecher, L., Larson, C., Mural, R. J., Somiari, S., Vicha, A., Zelinka, T., Bennett, J., Iacocca, M., Rabeno, B., Swanson, P., Latour, M., Lacombe, L., Têtu, B., Bergeron, A., McGraw, M., Staugaitis, S. M., Chabot, J., Hibshoosh, H., Sepulveda, A., Su, T., Wang, T., Potapova, O., Voronina, O., Desjardins, L., Mariani, O., Roman-Roman, S., Sastre, X., Stern, M.-H., Cheng, F., Signoretti, S., Berchuck, A., Bigner, D., Lipp, E., Marks, J., McCall, S., McLendon, R., Secord, A., Sharp, A., Behera, M., Brat, D. J., Chen, A., Delman, K., Force, S., Khuri, F., Magliocca, K., Maithel, S., Olson, J. J., Owonikoko, T., Pickens, A., Ramalingam, S., Shin, D. M., Sica, G., Van Meir, E. G., Zhang, H., Eijckenboom, W., Gillis, A., Korpershoek, E., Looijenga, L., Oosterhuis, W., Stoop, H., van Kessel, K. E., Zwarthoff, E. C., Calatozzolo, C., Cuppini, L., Cuzzubbo, S., DiMeco, F., Finocchiaro, G., Mattei, L., Perin, A., Pollo, B., Chen, C., Houck, J., Lohavanichbutr, P., Hartmann, A., Stoehr, C., Stoehr, R., Taubert, H., Wach, S., Wullich, B., Kycler, W., Murawa, D., Wiznerowicz, M., Chung, K., Edenfield, W. J., Martin, J., Baudin, E., Bubley, G., Bueno, R., De Rienzo, A., Richards, W. G., Kalkanis, S., Mikkelsen, T., Noushmehr, H., Scarpace, L., Girard, N., Aymerich, M., Campo, E., Giné, E., Guillermo, A. L., Van Bang, N., Hanh, P. T., Phu, B. D., Tang, Y., Colman, H., Evason, K., Dottino, P. R., Martignetti, J. A., Gabra, H., Juhl, H., Akeredolu, T., Stepa, S., Hoon, D., Ahn, K., Kang, K. J., Beuschlein, F., Breggia, A., Birrer, M., Bell, D., Borad, M., Bryce, A. H., Castle, E., Chandan, V., Cheville, J., Copland, J. A., Farnell, M., Flotte, T., Giama, N., Ho, T., Kendrick, M., Kocher, J.-P., Kopp, K., Moser, C., Nagorney, D., O'Brien, D., O'Neill, B. P., Patel, T., Petersen, G., Que, F., Rivera, M., Roberts, L., Smallridge, R., Smyrk, T., Stanton, M., Thompson, R. H., Torbenson, M., Yang, J. D., Zhang, L., Brimo, F., Ajani, J. A., Gonzalez, A. M. A., Behrens, C., Bondaruk, J., Broaddus, R., Czerniak, B., Esmaeli, B., Fujimoto, J., Gershenwald, J., Guo, C., Lazar, A. J., Logothetis, C., Meric-Bernstam, F., Moran, C., Ramondetta, L., Rice, D., Sood, A., Tamboli, P., Thompson, T., Troncoso, P., Tsao, A., Wistuba, I., Carter, C., Haydu, L., Hersey, P., Jakrot, V., Kakavand, H., Kefford, R., Lee, K., Long, G., Mann, G., Quinn, M., Saw, R., Scolyer, R., Shannon, K., Spillane, A., Stretch, O., Synott, M., Thompson, J., Wilmott, J., Al-Ahmadie, H., Chan, T. A., Ghossein, R., Gopalan, A., Levine, D. A., Reuter, V., Singer, S., Singh, B., Tien, N. V., Broudy, T., Mirsaidi, C., Nair, P., Drwiega, P., Miller, J., Smith, J., Zaren, H., Park, J.-W., Hung, N. P., Kebebew, E., Linehan, W. M., Metwalli, A. R., Pacak, K., Pinto, P. A., Schiffman, M., Schmidt, L. S., Vocke, C. D., Wentzensen, N., Worrell, R., Yang, H., Moncrieff, M., Goparaju, C., Melamed, J., Pass, H., Botnariuc, N., Caraman, I., Cernat, M., Chemencedji, I., Clipca, A., Doruc, S., Gorincioi, G., Mura, S., Pirtac, M., Stancul, I., Tcaciuc, D., Albert, M., Alexopoulou, I., Arnaout, A., Bartlett, J., Engel, J., Gilbert, S., Parfitt, J., Sekhon, H., Thomas, G., Rassl, D. M., Rintoul, R. C., Bifulco, C., Tamakawa, R., Urba, W., Hayward, N., Timmers, H., Antenucci, A., Facciolo, F., Grazi, G., Marino, M., Merola, R., de Krijger, R., Gimenez-Roqueplo, A.-P., Piché, A., Chevalier, S., McKercher, G., Birsoy, K., Barnett, G., Brewer, C., Farver, C., Naska, T., Pennell, N. A., Raymond, D., Schilero, C., Smolenski, K., Williams, F., Morrison, C., Borgia, J. A., Liptay, M. J., Pool, M., Seder, C. W., Junker, K., Omberg, L., Dinkin, M., Manikhas, G., Alvaro, D., Bragazzi, M. C., Cardinale, V., Carpino, G., Gaudio, E., Chesla, D., Cottingham, S., Dubina, M., Moiseenko, F., Dhanasekaran, R., Becker, K.-F., Janssen, K.-P., Slotta-Huspenina, J., Abdel-Rahman, M. H., Aziz, D., Bell, S., Cebulla, C. M., Davis, A., Duell, R., Elder, J. B., Hilty, J., Kumar, B., Lang, J., Lehman, N. L., Mandt, R., Nguyen, P., Pilarski, R., Rai, K., Schoenfield, L., Senecal, K., Wakely, P., Hansen, P., Lechan, R., Powers, J., Tischler, A., Grizzle, W. E., Sexton, K. C., Kastl, A., Henderson, J., Porten, S., Waldmann, J., Fassnacht, M., Asa, S. L., Schadendorf, D., Couce, M., Graefen, M., Huland, H., Sauter, G., Schlomm, T., Simon, R., Tennstedt, P., Olabode, O., Nelson, M., Bathe, O., Carroll, P. R., Chan, J. M., Disaia, P., Glenn, P., Kelley, R. K., Landen, C. N., Phillips, J., Prados, M., Simko, J., Smith-McCune, K., VandenBerg, S., Roggin, K., Fehrenbach, A., Kendler, A., Sifri, S., Steele, R., Jimeno, A., Carey, F., Forgie, I., Mannelli, M., Carney, M., Hernandez, B., Campos, B., Herold-Mende, C.,

Jungk, C., Unterberg, A., von Deimling, A., Bossler, A., Galbraith, J., Jacobus, L., Knudson, M., Knutson, T., Ma, D., Milhem, M., Sigmund, R., Godwin, A. K., Madan, R., Rosenthal, H. G., Adebamowo, C., Adebamowo, S. N., Boussioutas, A., Beer, D., Giordano, T., Mes-Masson, A.-M., Saad, F., Bocklage, T., Landrum, L., Mannel, R., Moore, K., Moxley, K., Postier, R., Walker, J., Zuna, R., Feldman, M., Valdivieso, F., Dhir, R., Luketich, J., Pinero, E. M. M., Quintero-Aguilo, M., Carlotti, C. G., Dos Santos, J. S., Kemp, R., Sankarankuty, A., Tirapelli, D., Catto, J., Agnew, K., Swisher, E., Creaney, J., Robinson, B., Shelley, C. S., Godwin, E. M., Kendall, S., Shipman, C., Bradford, C., Carey, T., Haddad, A., Moyer, J., Peterson, L., Prince, M., Rozek, L., Wolf, G., Bowman, R., Fong, K. M., Yang, I., Korst, R., Rathmell, W. K., Fantacone-Campbell, J. L., Hooke, J. A., Kovatich, A. J., Shriver, C. D., DiPersio, J., Drake, B., Govindan, R., Heath, S., Ley, T., Van Tine, B., Westervelt, P., Rubin, M. A., Lee, J. I., Aredes, N. D., and Mariamidze, A. (2018). The Immune Landscape of Cancer. *Immunity*, 48(4):812–830.e14.

Thorvaldsdóttir, H., Robinson, J. T., and Mesirov, J. P. (2013). Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Briefings in Bioinformatics*, 14(2):178–192.

Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D., Trombetta, J. J., Rotem, A., Rodman, C., Lian, C., Murphy, G., Fallahi-Sichani, M., Dutton-Regester, K., Lin, J.-R., Cohen, O., Shah, P., Lu, D., Genshaft, A. S., Hughes, T. K., Ziegler, C. G. K., Kazer, S. W., Gaillard, A., Kolb, K. E., Villani, A.-C., Johannessen, C. M., Andreev, A. Y., Van Allen, E. M., Bertagnolli, M., Sorger, P. K., Sullivan, R. J., Flaherty, K. T., Frederick, D. T., Jané-Valbuena, J., Yoon, C. H., Rozenblatt-Rosen, O., Shalek, A. K., Regev, A., and Garraway, L. A. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science (New York, N.Y.)*, 352(6282):189–96.

Tofts, P. S. (1997). Modeling tracer kinetics in dynamic Gd-DTPA MR imaging. *J. Magn. Reson. Imaging*, 7(1):91–101.

Tong, J. C., Tan, T. W., and Ranganathan, S. (2004). Modeling the structure of bound peptide ligands to major histocompatibility complex. *Protein science*, 13(9):2523–2532.

Tran, E., Robbins, P. F., Lu, Y.-C., Prickett, T. D., Gartner, J. J., Jia, L., Pasetto, A., Zheng, Z., Ray, S., Groh, E. M., Kriley, I. R., and Rosenberg, S. A. (2016). T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer. *New England Journal of Medicine*, 375(23):2255–2262.

Tu, J. C., Xiao, B., Yuan, J. P., Lanahan, A. A., Leoffert, K., Li, M., Linden, D. J., and Worley, P. F. (1998). Homer binds a novel proline-rich motif and links group 1 metabotropic glutamate receptors with IP3 receptors. *Neuron*, 21(4):717–726.

Turke, A. B., Zejnullahu, K., Wu, Y. L., Song, Y., Dias-Santagata, D., Lifshits, E., Toschi, L., Rogers, A., Mok, T., Sequist, L., Lindeman, N. I., Murphy, C., Akhavanfard, S., Yeap, B. Y., Xiao, Y., Capelletti, M., Iafrate, A. J., Lee, C., Christensen, J. G., Engelman, J. A., and J??nne, P. A. (2010). Preexistence and Clonal Selection of MET Amplification in EGFR Mutant NSCLC. *Cancer Cell*, 17(1):77–88.

Valkenburg, K. C., de Groot, A. E., and Pienta, K. J. (2018). Targeting the tumour stroma to improve cancer therapy. *Nature Reviews Clinical Oncology*, 15(6):366–381.

van der Woude, L. L., Gorris, M. A., Halilovic, A., Figdor, C. G., and de Vries, I. J. M. (2017). Migrating into the Tumor: a Roadmap for T Cells. *Trends in Cancer*, 3(11):797–808.

van Dijk, D., Sharma, R., Nainys, J., Yim, K., Kathail, P., Carr, A. J., Burdziak, C., Moon, K. R., Chaffer, C. L., Pattabiraman, D., Bierie, B., Mazutis, L., Wolf, G., Krishnaswamy, S., and Pe'er, D. (2018). Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell*, 174(3):716–729.e27.

Varricchi, G., Galdiero, M. R., Loffredo, S., Marone, G., Iannone, R., Marone, G., and Granata, F. (2017). Are Mast Cells MASTers in Cancer? *Frontiers in Immunology*, 8:1.

Veglia, F. and Gabrilovich, D. I. (2017). Dendritic cells in cancer: the role revisited. *Current Opinion in Immunology*, 45:43–51.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer.

Verdegaal, E. M. E., de Miranda, N. F. C. C., Visser, M., Harryvan, T., van Buuren, M. M., Andersen, R. S., Hadrup, S. R., van der Minne, C. E., Schotte, R., Spits, H., Haanen, J. B. A. G., Kapiteijn, E. H. W., Schumacher, T. N., and van der Burg, S. H. (2016). Neoantigen landscape dynamics during human melanoma–T cell interactions. *Nature*, 536(7614):91–95.

Verhaak, R. G., Tamayo, P., Yang, J.-Y., Hubbard, D., Zhang, H., Creighton, C. J., Fereday, S., Lawrence, M., Carter, S. L., Mermel, C. H., Kostic, A. D., Etemadmoghadam, D., Saksena, G., Cibulskis, K., Duraisamy, S., Levanon, K., Sougnez, C., Tsherniak, A., Gomez, S., Onofrio, R., Gabriel, S., Chin, L., Zhang, N., Spellman, P. T., Zhang, Y., Akbani, R., Hoadley, K. A., Kahn, A., Köbel, M., Huntsman, D., Soslow, R. A., Defazio, A., Birrer, M. J., Gray, J. W., Weinstein, J. N., Bowtell, D. D., Drapkin, R., Mesirov, J. P., Getz, G., Levine, D. A., and Meyerson, M. (2012). Prognostically relevant gene signatures of high-grade serous ovarian carcinoma. *Journal of Clinical Investigation*.

Vetizou, M., Pitt, J. M., Daillere, R., Lepage, P., Waldschmitt, N., Flament, C., Rusakiewicz, S., Routy, B., Roberti, M. P., Duong, C. P. M., Poirier-Colame, V., Roux, A., Becharef, S., Formenti, S., Golden, E., Cording, S., Eberl, G., Schlitzer, A., Ginhoux, F., Mani, S., Yamazaki, T., Jacquelot, N., Enot, D. P., Berard, M., Nigou, J., Opolon, P., Eggermont, A., Woerther, P.-L., Chachaty, E., Chaput, N., Robert, C., Mateus, C., Kroemer, G., Raoult, D., Boneca, I. G., Carbonnel, F., Chamaillard, M., and Zitvogel, L. (2015). Anticancer immunotherapy by CTLA-4 blockade relies on the gut microbiota. *Science*, 350(6264):1079–1084.

Vigneron, N. (2015). Human Tumor Antigens and Cancer Immunotherapy. *BioMed research international*, 2015:948501.

Vita, R., Overton, J. A., Greenbaum, J. A., Ponomarenko, J., Clark, J. D., Cantrell, J. R., Wheeler, D. K., Gabbard, J. L., Hix, D., Sette, A., and Peters, B. (2015). The immune epitope database (IEDB) 3.0. *Nucleic Acids Research*, 43(D1):D405–D412.

Von Essen, M. R., Kongsbak, M., and Geisler, C. (2012). Mechanisms behind functional avidity maturation in T cells. *Clinical and Developmental Immunology*, 2012.

Vyas, J. M., Van der Veen, A. G., and Ploegh, H. L. (2008). The known unknowns of antigen processing and presentation. *Nature reviews. Immunology*, 8(8):607–18.

Wagner, M., Moro, K., and Koyasu, S. (2017). Plastic Heterogeneity of Innate Lymphoid Cells in Cancer. *Trends in Cancer*, 3(5):326–335.

Wang, M., Zhao, J., Zhang, L., Wei, F., Lian, Y., Wu, Y., Gong, Z., Zhang, S., Zhou, J., Cao, K., Li, X., Xiong, W., Li, G., Zeng, Z., and Guo, C. (2017). Role of tumor microenvironment in tumorigenesis.

Wang, P., Sidney, J., Dow, C., Mothé, B., Sette, A., and Peters, B. (2008). A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. *PLoS Computational Biology*, 4(4).

Wang, P., Sidney, J., Kim, Y., Sette, A., Lund, O., Nielsen, M., and Peters, B. (2010). Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics*, 11:568.

Wang, T.-L., Diaz, L. A., Romans, K., Bardelli, A., Saha, S., Galizia, G., Choti, M., Donehower, R., Parmigiani, G., Shih, I.-M., Iacobuzio-Donahue, C., Kinzler, K. W., Vogelstein, B., Lengauer, C., and Velculescu, V. E. (2004). Digital karyotyping identifies thymidylate synthase amplification as a mechanism of resistance to 5-fluorouracil in metastatic colorectal cancer patients. *Proceedings of the National Academy of Sciences*, 101(9):3089–3094.

Wang, W., Kryczek, I., Dostál, L., Lin, H., Tan, L., Zhao, L., Lu, F., Wei, S., Maj, T., Peng, D., He, G., Vatan, L., Szeliga, W., Kuick, R., Kotarski, J., Tarkowski, R., Dou, Y., Rattan, R., Munkarah, A., Liu, J. R., and Zou, W. (2016). Effector T cells abrogate Stroma-Mediated chemoresistance in ovarian cancer. *Cell*, 165(5):1092–1105.

Wei, S. C., Levine, J. H., Cogdill, A. P., Zhao, Y., Anang, N.-A. A., Andrews, M. C., Sharma, P., Wang, J., Wargo, J. A., Pe'er, D., and Allison, J. P. (2017). Distinct Cellular Mechanisms Underlie Anti-CTLA-4 and Anti-PD-1 Checkpoint Blockade. *Cell*, 170(6):1120–1133.e17.

Weidensdorfer, D., Stöhr, N., Baude, A., Lederer, M., Köhn, M., Schierhorn, A., Buchmeier, S., Wahle, E., and Hüttelmaier, S. (2009). Control of c-myc mRNA stability by IGF2BP1-associated cytoplasmic RNPs. *RNA*, 15(1):104–115.

Wiseman, B. S. (2002). Stromal Effects on Mammary Gland Development and Breast Cancer. *Science*, 296(5570):1046–1049.

Wu, D., Wu, P., Qiu, F., Wei, Q., and Huang, J. (2016). Human $\gamma\delta$T-cell subsets and their involvement in tumor immunity. *Nature Publishing Group*, 14:245–253.

Wu, M. and Swartz, M. A. (2014). Modeling Tumor Microenvironments In Vitro.

Xiao, B., Tu, J. C., Petralia, R. S., Yuan, J. P., Doan, A., Breder, C. D., Ruggiero, A., Lanahan, A. A., Wenthold, R. J., and Worley, P. F. (1998). Homer regulates the association of group 1 metabotropic glutamate receptors with multivalent complexes of homer-related, synaptic proteins. *Neuron*, 21(4):707–716.

Xu, C. (2018). A review of somatic single nucleotide variant calling algorithms for next-generation sequencing data. *Computational and Structural Biotechnology Journal*, 16:15–24.

Xue, W., Zender, L., Miething, C., Dickins, R. A., Hernando, E., Krizhanovsky, V., Cordon-Cardo, C., and Lowe, S. W. (2007). Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. *Nature*, 445(7128):656–660.

Yarchoan, M., Hopkins, A., and Jaffee, E. M. (2017). Tumor Mutational Burden and Response Rate to PD-1 Inhibition. *New England Journal of Medicine*, 377(25):2500–2501.

Yarilin, D., Xu, K., Turkekul, M., Fan, N., Romin, Y., Fijisawa, S., Barlas, A., and Manova-Todorova, K. (2015). Machine-based method for multiplex in situ molecular characterization of tissues by immunofluorescence detection. *Sci. Rep.*, 5:9534.

Yates, L. R., Gerstung, M., Knappskog, S., Desmedt, C., Gundem, G., Van Loo, P., Aas, T., Alexandrov, L. B., Larsimont, D., Davies, H., Li, Y., Ju, Y. S., Ramakrishna, M., Haugland, H. K., Lilleng, P. K., Nik-Zainal, S., McLaren, S., Butler, A., Martin, S., Glodzik, D., Menzies, A., Raine, K., Hinton, J., Jones, D., Mudie, L. J., Jiang, B., Vincent, D., Greene-Colozzi, A., Adnet, P.-Y., Fatima, A., Maetens, M., Ignatiadis, M., Stratton, M. R., Sotiriou, C., Richardson, A. L., Lønning, P. E., Wedge, D. C., and Campbell, P. J. (2015). Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat. Med.*, 21(7):751–759.

Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P. W., Levine, D. A., Carter, S. L., Getz, G., Stemke-Hale, K., Mills, G. B., and Verhaak, R. G. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature Communications*, 4.

Yu, K., Petrovsky, N., Schönbach, C., Koh, J. Y. L., and Brusic, V. (2002). Methods for prediction of peptide binding to MHC molecules: a comparative study. *Molecular medicine (Cambridge, Mass.)*, 8(3):137–148.

Yuan, G.-C., Cai, L., Elowitz, M., Enver, T., Fan, G., Guo, G., Irizarry, R., Kharchenko, P., Kim, J., Orkin, S., Quackenbush, J., Saadatpour, A., Schroeder, T., Shivdasani, R., and Tirosh, I. (2017). Challenges and emerging directions in single-cell analysis. *Genome Biology*, 18(1):84.

Yuen, G. J., Demissie, E., and Pillai, S. (2016). B Lymphocytes and Cancer: A Love–Hate Relationship. *Trends in Cancer*, 2(12):747–757.

Zaretsky, J. M., Garcia-Diaz, A., Shin, D. S., Escuin-Ordinas, H., Hugo, W., Hu-Lieskovan, S., Torrejon, D. Y., Abril-Rodriguez, G., Sandoval, S., Barthly, L., Saco, J., Homet Moreno, B., Mezzadra, R., Chmielowski, B., Ruchalski, K., Shintaku, I. P., Sanchez, P. J., Puig-Saus, C., Cherry, G., Seja, E., Kong, X., Pang, J., Berent-Maoz, B., Comin-Anduix, B., Graeber, T. G., Tumeh, P. C., Schumacher, T. N., Lo, R. S., and Ribas, A. (2016). Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma. *New England Journal of Medicine*, 375(9):819–829.

Zhang, A. W., McPherson, A., Milne, K., Kroeger, D. R., Hamilton, P. T., Miranda, A., Funnell, T., Little, N., de Souza, C. P. E., Laan, S., LeDoux, S., Cochrane, D. R., Lim, J. L. P., Yang, W., Roth, A., Smith, M. A., Ho, J., Tse, K., Zeng, T., Shlafman, I., Mayo, M. R., Moore, R., Failmezger, H., Heindl, A., Wang, Y. K., Bashashati, A., Grewal, D. S., Brown, S. D., Lai, D., Wan, A. N. C., Nielsen, C. B., Huebner, C., Tessier-Cloutier, B., Anglesio, M. S., Bouchard-Côté, A., Yuan, Y., Wasserman, W. W., Gilks, C. B., Karnezis, A. N., Aparicio, S., McAlpine, J. N., Huntsman, D. G., Holt, R. A., Nelson, B. H., and Shah, S. P. (2018). Interfaces of malignant and immunologic clonal dynamics in ovarian cancer. *Cell*.

Zhang, G. L., Bozic, I., Kwoh, C. K., August, J. T., and Brusic, V. (2007). Prediction of supertype-specific HLA class I binding peptides using support vector machines. *Journal of Immunological Methods*, 320(1-2):143–154.

Zhang, L., Conejo-Garcia, J. R., Katsaros, D., Gimotty, P. A., Massobrio, M., Regnani, G., Makrigiannakis, A., Gray, H., Schlienger, K., Liebman, M. N., Rubin, S. C., and Coukos, G. (2003). Intratumoral T Cells, Recurrence, and Survival in Epithelial Ovarian Cancer. *New England Journal of Medicine*, 348(3):203–213.

Zhao, Y., Shuen, T. W. H., Toh, T. B., Chan, X. Y., Liu, M., Tan, S. Y., Fan, Y., Yang, H., Lyer, S. G., Bonney, G. K., Loh, E., Chang, K. T. E., Tan, T. C., Zhai, W., Chan, J. K. Y., Chow, E. K.-H., Chee, C. E., Lee, G. H., Dan, Y. Y., Chow, P. K.-H., Toh, H. C., Lim, S. G., and Chen, Q. (2018). Development of a new patient-derived xenograft humanised mouse model to study human-specific tumour microenvironment and immunotherapy. *Gut*, pages gutjnl–2017–315201.

Zheng, C., Zheng, L., Yoo, J.-K., Guo, H., Zhang, Y., Guo, X., Kang, B., Hu, R., Huang, J. Y., Zhang, Q., Liu, Z., Dong, M., Hu, X., Ouyang, W., Peng, J., and Zhang, Z. (2017). Landscape of Infiltrating T Cells in Liver Cancer Revealed by Single-Cell Sequencing. *Cell*, 169(7):1342–1356.e16.

Zheng, S. (2017). Benchmarking: contexts and details matter. *Genome Biology*, 18(1):129.

Zimmerlin, L., Donnenberg, A. D., Rubin, J. P., Basse, P., Landreneau, R. J., and Donnenberg, V. S. (2011). Regenerative Therapy and Cancer: In Vitro and In Vivo Studies of the Interaction Between Adipose-Derived Stem Cells and Breast Cancer Cells from Clinical Isolates. *Tissue Engineering Part A*, 17(1-2):93–106.

Zitvogel, L., Apetoh, L., Ghiringhelli, F., and Kroemer, G. (2008). Immunological aspects of cancer chemotherapy. *Nature Reviews Immunology*, 8(1):59–73.

Şenbabaoğlu, Y., Gejman, R. S., Winer, A. G., Liu, M., Van Allen, E. M., de Velasco, G., Miao, D., Ostrovnaya, I., Drill, E., Luna, A., Weinhold, N., Lee, W., Manley, B. J., Khalil, D. N., Kaffenberger, S. D., Chen, Y., Danilova, L., Voss, M. H., Coleman, J. A., Russo, P., Reuter, V. E., Chan, T. A., Cheng, E. H., Scheinberg, D. A., Li, M. O., Choueiri, T. K., Hsieh, J. J., Sander, C., and Hakimi, A. A. (2016). Tumor immune microenvironment characterization in clear cell renal cell carcinoma identifies prognostic and immunotherapeutically relevant messenger RNA signatures. *Genome Biology*, 17(1):231.

Chapter 7

Appendix A

Supplementary material for Chapter 2:
Co-existence of tumour-immune microenvironments in an ovarian cancer patient

Supplementary tables

Supplementary tables can be accessed through:
https://github.com/cansysbio/immunogenomics/tree/master/AJSPhDThesis/AppendixA/Tables/

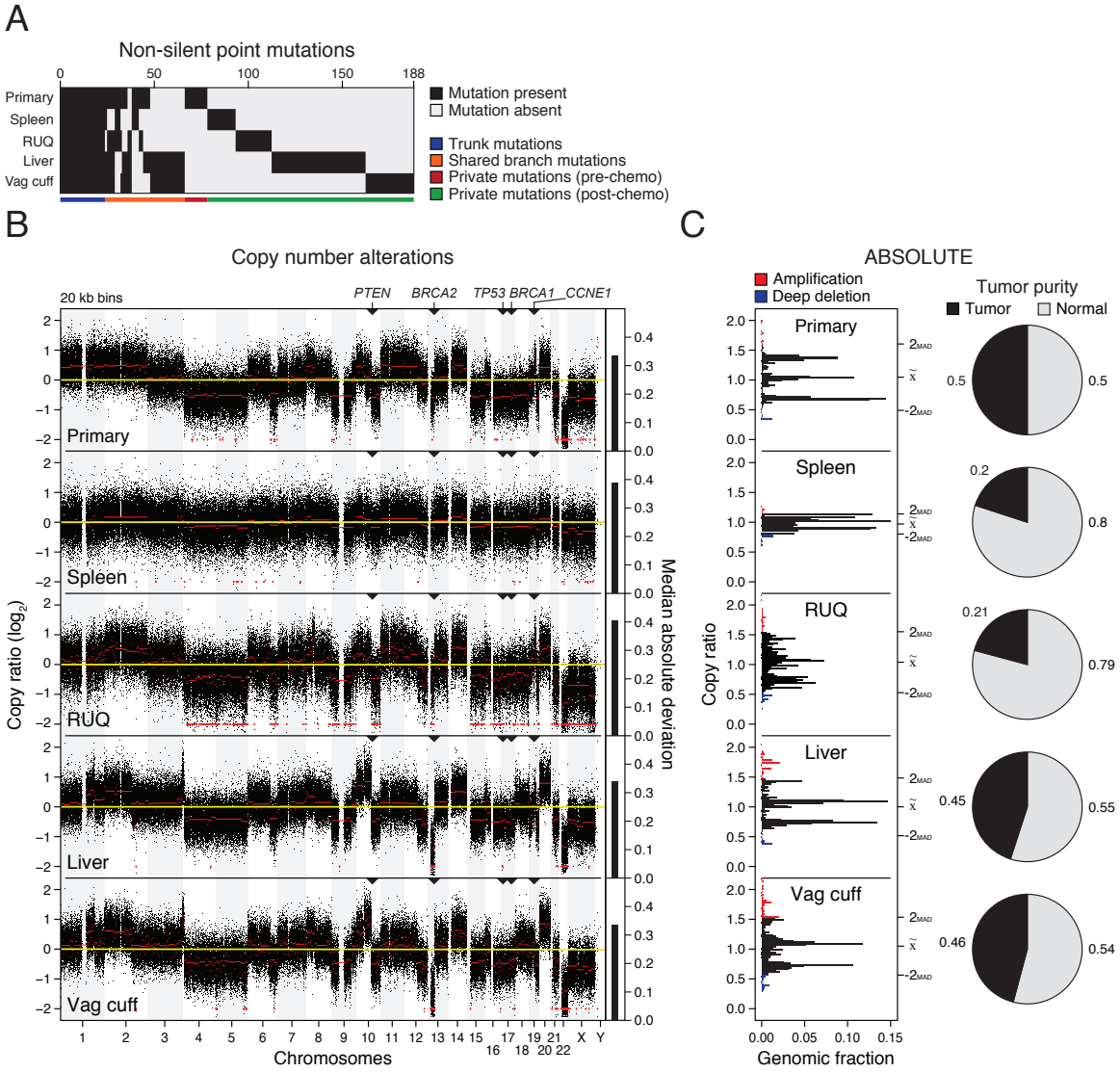


Fig. 7.1 Non-silent somatic mutations and copy-number alterations, related to Figures 2.1 and 7.2 and Table S1. (A) Binary matrix of present/absent non-silent point mutations (n=188) used for the phylogeny tree reconstruction in Figure 2.1D. (B) Relative copy-number alterations inferred from WES data of the primary and metastatic samples using CopywriteR (Kuilman et al., 2015). (C) Relative copy number profiles and tumour purity inferred after ABSOLUTE (Carter et al., 2012) analysis. Amplified and deep deleted DNA segments were defined as copy number alterations with at least ± 2 median absolute deviations for each sample. MAD = median absolute deviation.

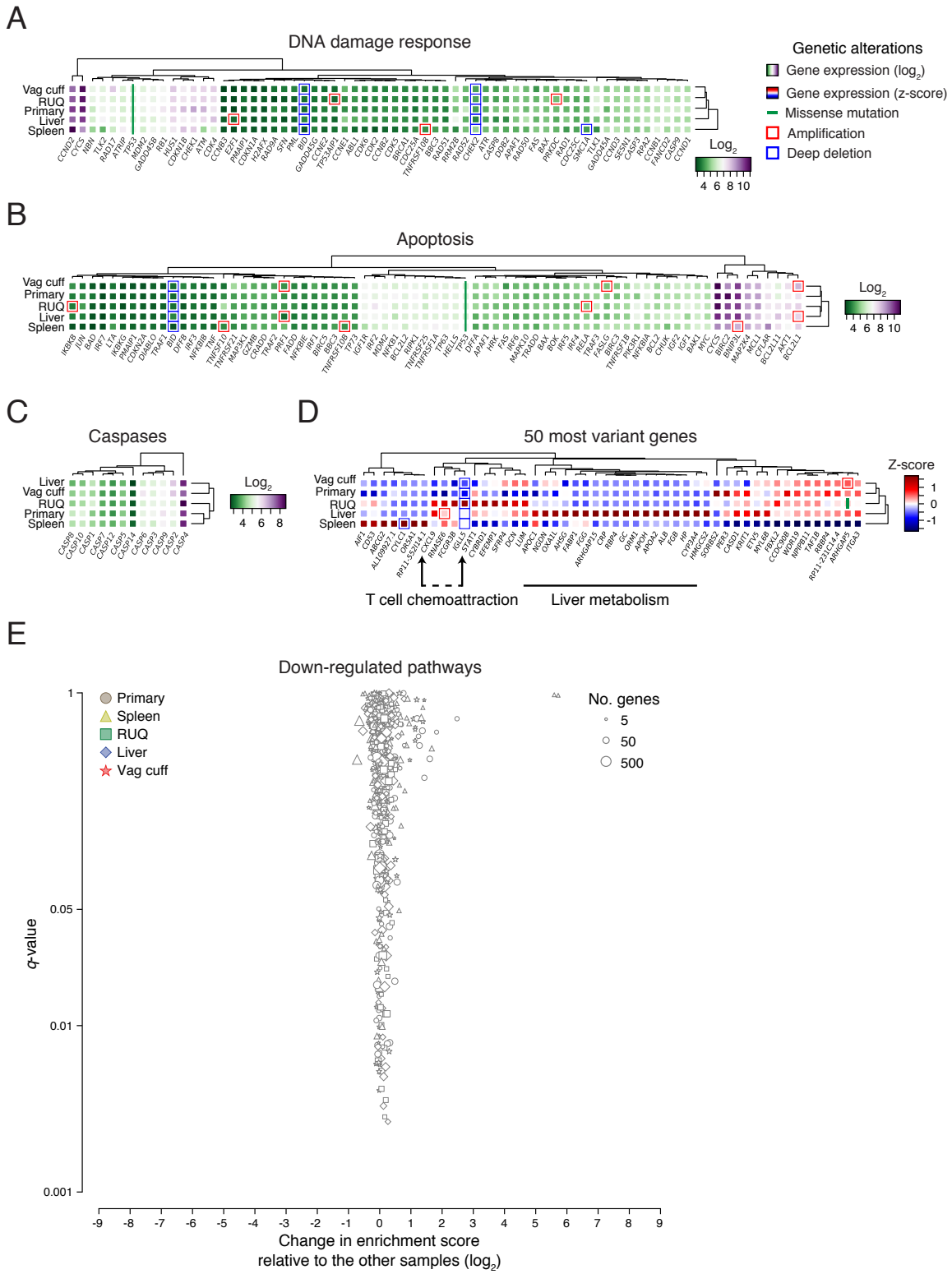


Fig. 7.2 Gene set enrichment analysis of transcript abundance and somatic alteration patterns across samples, related to Figure 2.2 and Table S2. (A–C) Gene-expression levels and genetic alterations of the DNA damage, apoptosis pathways, and caspases. (D) Expression levels of the 50 most variant genes according to their coefficient of variation. (E) Differential enrichment scores and enrichment q-values of down-regulated pathways between tumour samples. No significantly enriched pathways (q-value < 0.05) with at least ± 1 log_2 change relative to the median of the other samples were detected. FDR adjusted p-value (q-value) was calculated using the BH method.

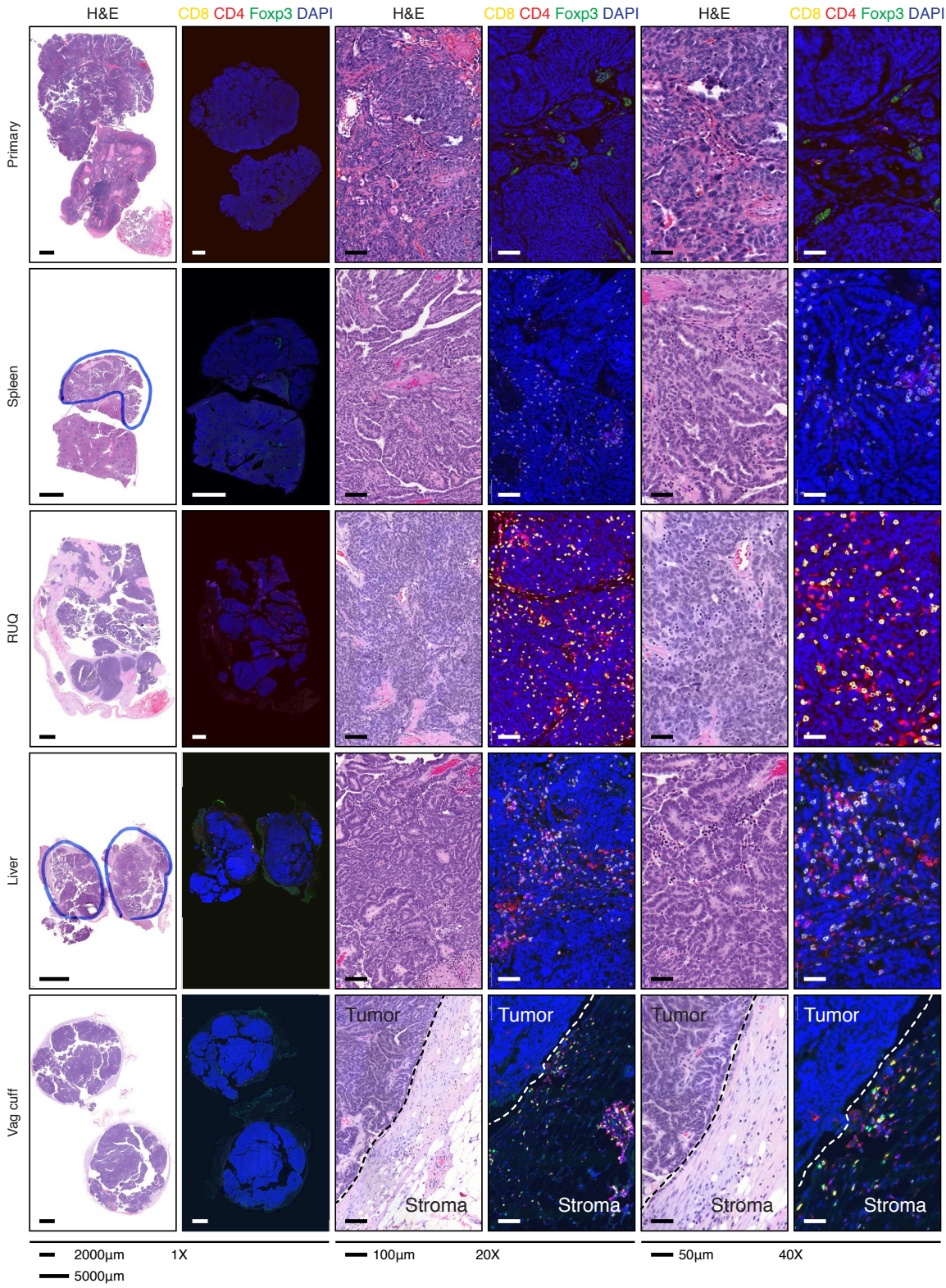


Fig. 7.3 Complete slide haematoxylin and eosin and immunofluorescent staining, related to Figure 2.3 and Table S3. Hematoxylin and eosin staining of tumour samples. Immunofluorescence staining for cytotoxic T cells ($CD8^+$), helper T cells ($CD4^+FOXP3^-$), and regulatory T cells ($CD4^+FOXP3^+$).

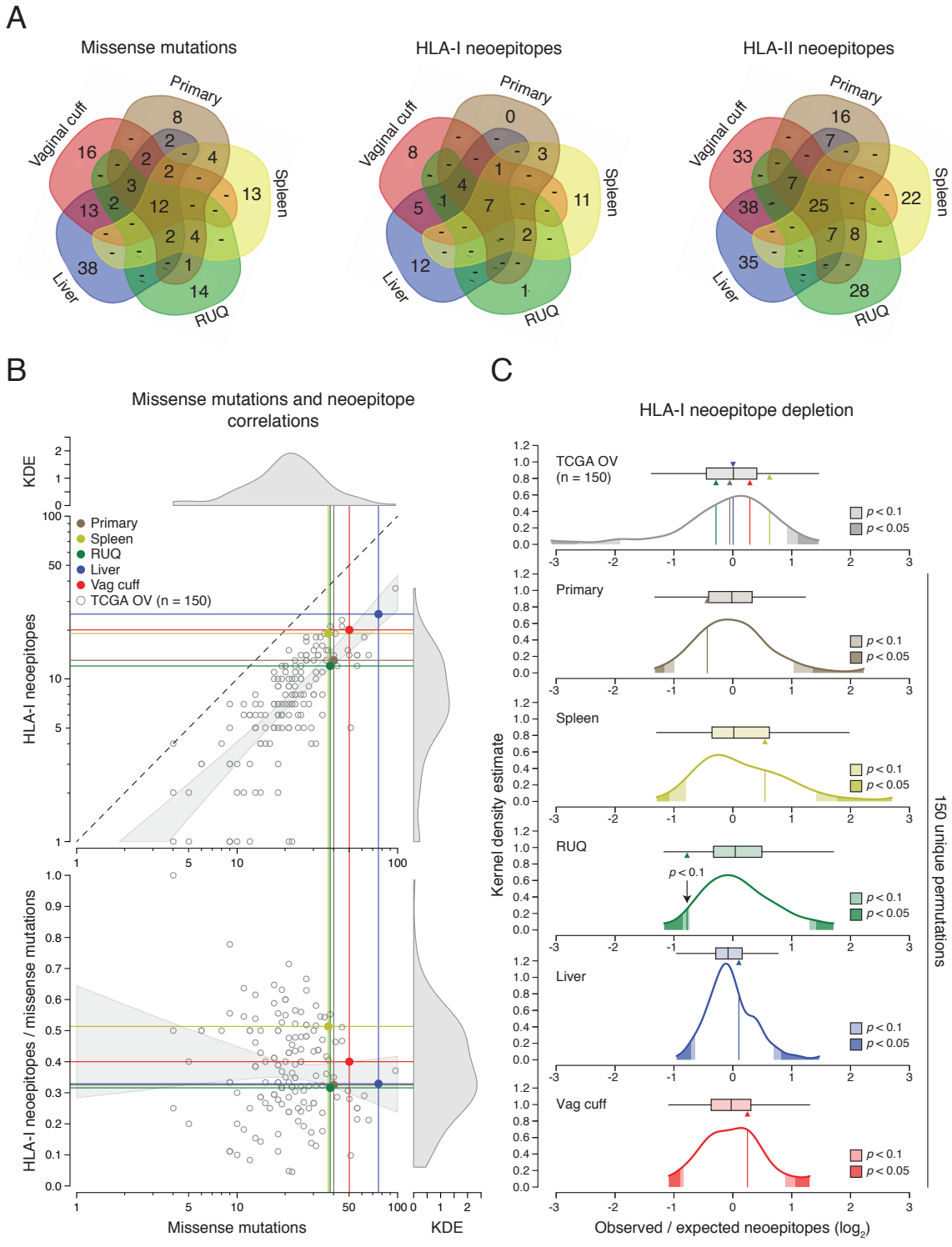


Fig. 7.4 Neoepitope distributions and HLA-I neoepitope depletion analysis, related to Figure 2.4 and Table C2S4. (A) Number of unique and overlapping expressed missense mutations, HLA-I, and II neoepitopes between samples. (B) Correlations between expressed missense mutations and predicted HLA-I neoepitopes using NetMHC. KDE = Kernel Density Estimate. (C) Top: Estimated neoepitope deviation from expected in the five tumour samples compared to TCGA ovarian cancer samples (n=150) (see Methods). Bottom: Neoepitope depletion analysis of 150 random unique permutations of the patient's tumours (primary, spleen, RUQ, liver, and vaginal cuff) and their mutations. Each sample was compared against its own 150 unique permutations to control for the number of mutations. Two-sided empirical *p*-values were calculated from each distribution.

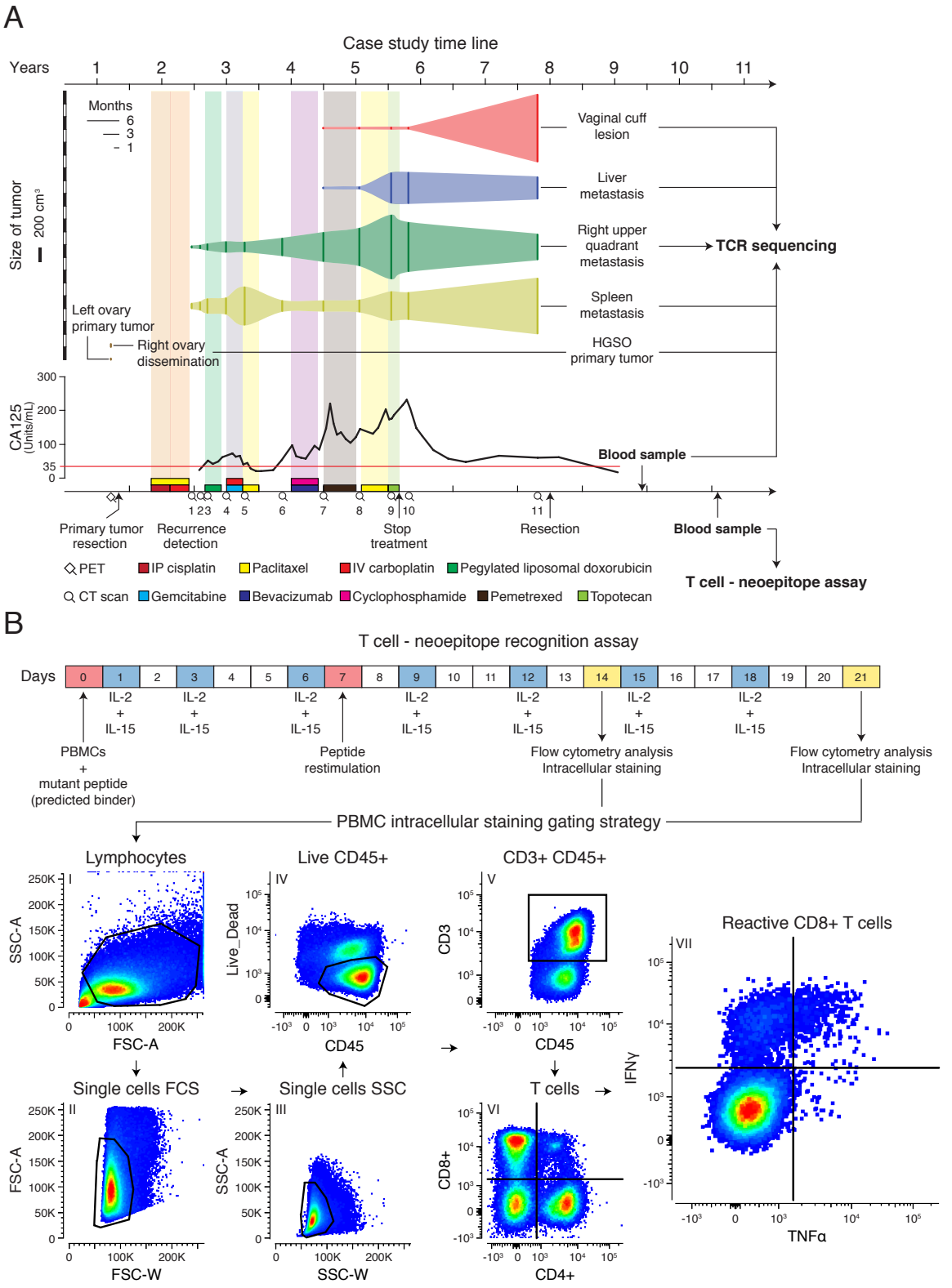


Fig. 7.5 PBMCs sample time line and T cell-neoepitope recognition assay, related to Figures 2.4 and 2.6, and Table C2S5. (A) Blood samples obtained from the patient 550 and 978 days after resection were used for TCR sequencing and T cell – neoepitope recognition assays respectively. (B) Experimental set-up and flow cytometry gating strategy for the T cell–neoepitope recognition assays (intracellular cytokine staining assay) with surface staining of $CD3^+$, $CD4^+$, $CD8^+$, $CD45^+$, and intracellular staining of IL-4^+, IFN-γ, TNF-α. PBMC = peripheral blood mononuclear cells.

Chapter 8

Appendix B

Supplementary material for Chapter 3:
Unravelling tumour-immune heterogeneity in advanced ovarian cancer uncovers immunogenic effect of chemotherapy

Supplementary tables

Supplementary tables can be accessed through:
https://github.com/cansysbio/immunogenomics/tree/master/AJSPhDThesis/AppendixB/Tables/

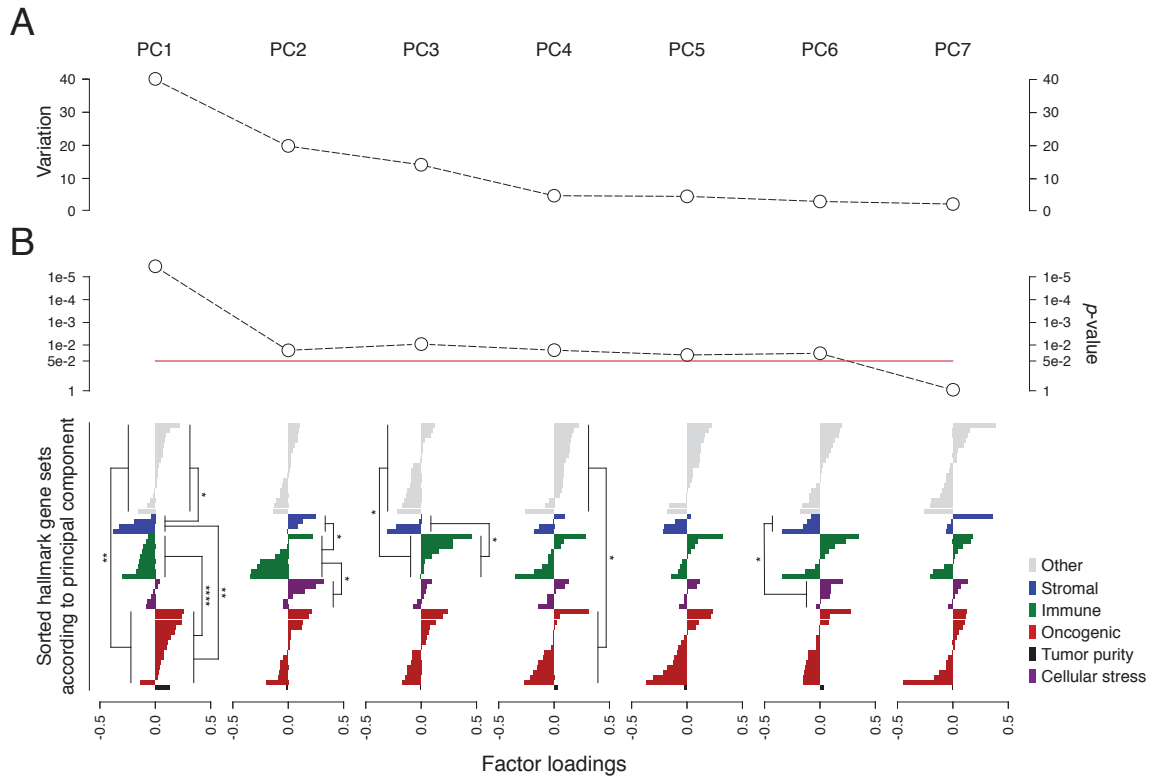


Fig. 8.1 Principal component hallmark gene sets' loadings of treatment-naive HGSOC samples, related to Figure 3.1. (A) Principal components' variance. (B) Kruskal-Wallis H-test p-values calculated for hallmark classes (top), and loadings of hallmarks for each principal component sorted in groups according to the gene sets' classes defined (bottom). Significant Dunn's Kruskal-Wallis multiple comparisons with Bonferroni correction are indicated: adjusted p-value < 0.05 *, adjusted p-value < 0.01 **, adjusted p-value < 0.001 ***, adjusted p-value < 0.0001 ****.

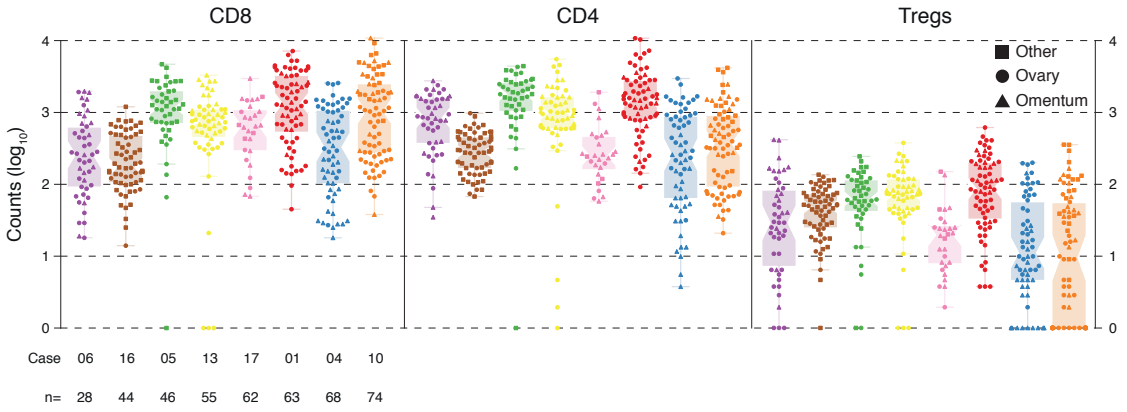


Fig. 8.2 Total counts of T cell subsets in treatment-naive HGSOC samples, related to Figure 3.2 and Table C3S2. Immunofluorescent total $CD8^+$, $CD4^+$, and Tregs counts. 440 different regions were scanned and computationally quantified.

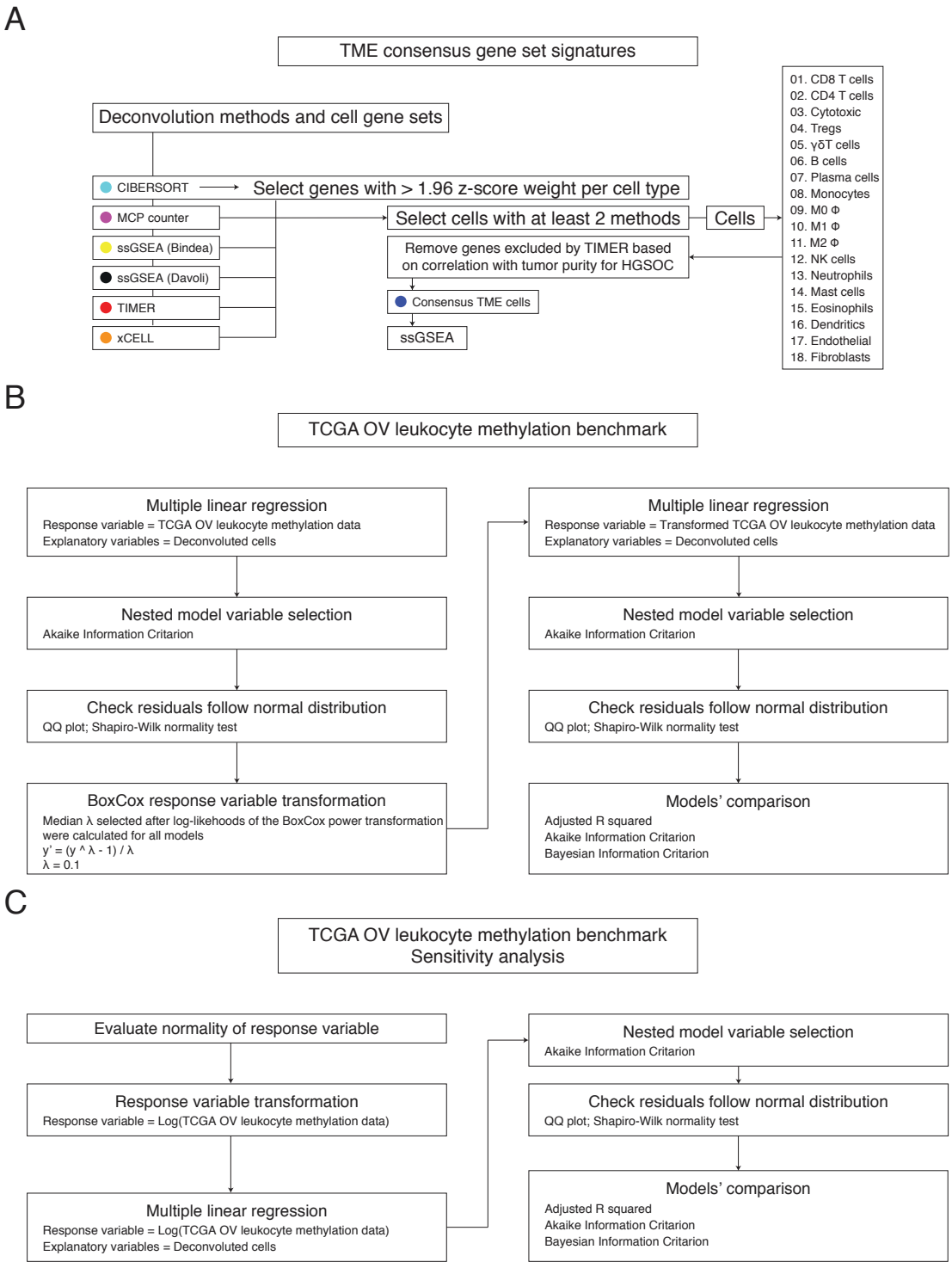


Fig. 8.3 Flowcharts for consensus TME gene set generation and TCGA OV leukocyte methylation benchmarks, related to Figure 3.3 and Table C3S3. (A) Generation of Consensus^{TME} v1.0 gene sets by taking the intersection of genes used by other methods. (B) TCGA OV leukocyte methylation benchmark using multiple linear regression through Box-Cox or (C) *log* transformation of response variable.

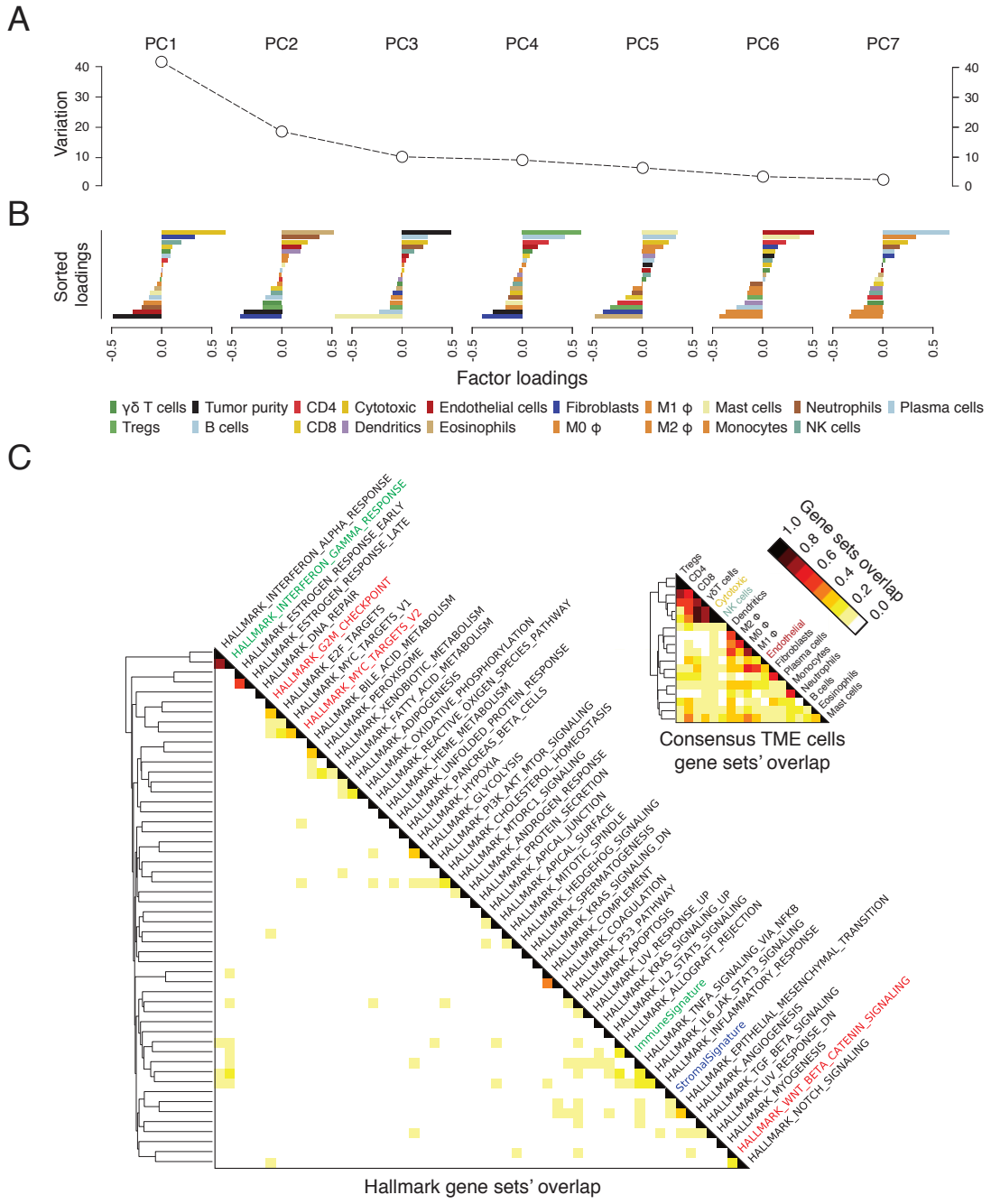


Fig. 8.4 Principal component TME cell gene sets' loadings of treatment-naive HGSOC samples and gene sets' overlaps, related to Figure 3.4. (A) Principal components' variance. (B) Loadings of hallmarks for each principal component sorted. (C) Overlap of genes between hallmark gene sets, and between consensus TME cells gene sets. Gene sets that were differentially enriched between high and low purity treatment-naive HGSOC samples (Figure 3.4E, F) are highlighted in colour.

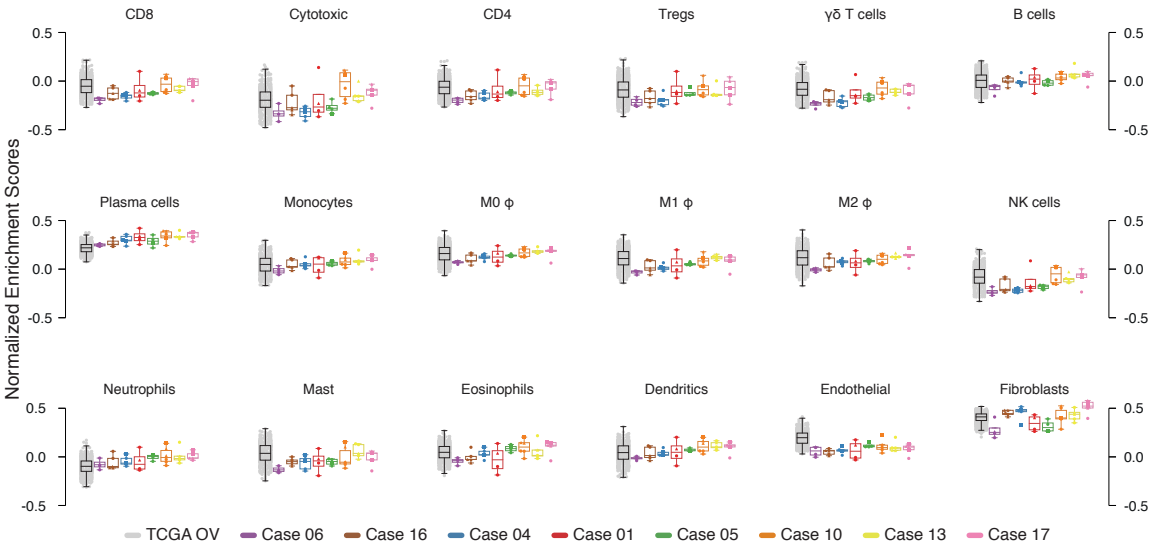


Fig. 8.5 Normalised enrichment scores of estimated TME cells in TCGA ovarian cancer and the treatment-naive HGSOC cohort, related to Figure 3.4.

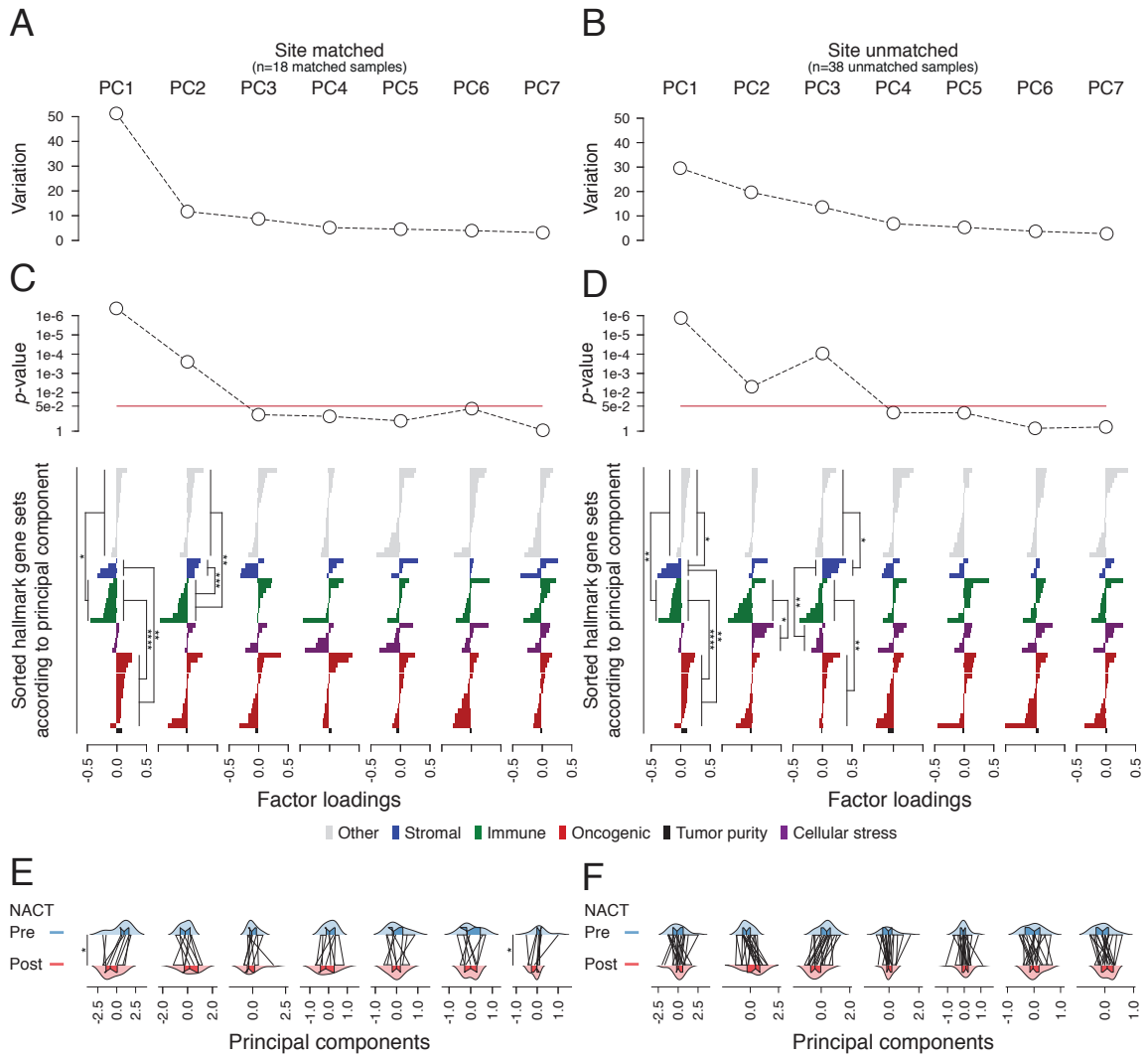


Fig. 8.6 Principal component hallmark gene sets' loadings of pre/post-NACT matched and unmatched HGSOC samples, related to Figure 3.5. (A, B) Principal components' variance. (C, D) Kruskal-Wallis H-test *p*-values calculated for hallmark classes (top), and loadings of hallmarks for each principal component sorted in groups according to the gene sets' classes defined (bottom). Significant Dunn's Kruskal-Wallis multiple comparisons with Bonferroni correction are indicated: adjusted *p*-value < 0.05 *, adjusted *p*-value < 0.01 **, adjusted *p*-value < 0.001 ***, adjusted *p*-value < 0.0001 ****. (E, F) Principal components' paired comparisons. Significant differences after Wilcoxon signed-rank tests are indicated: *p*-value < 0.05 *.

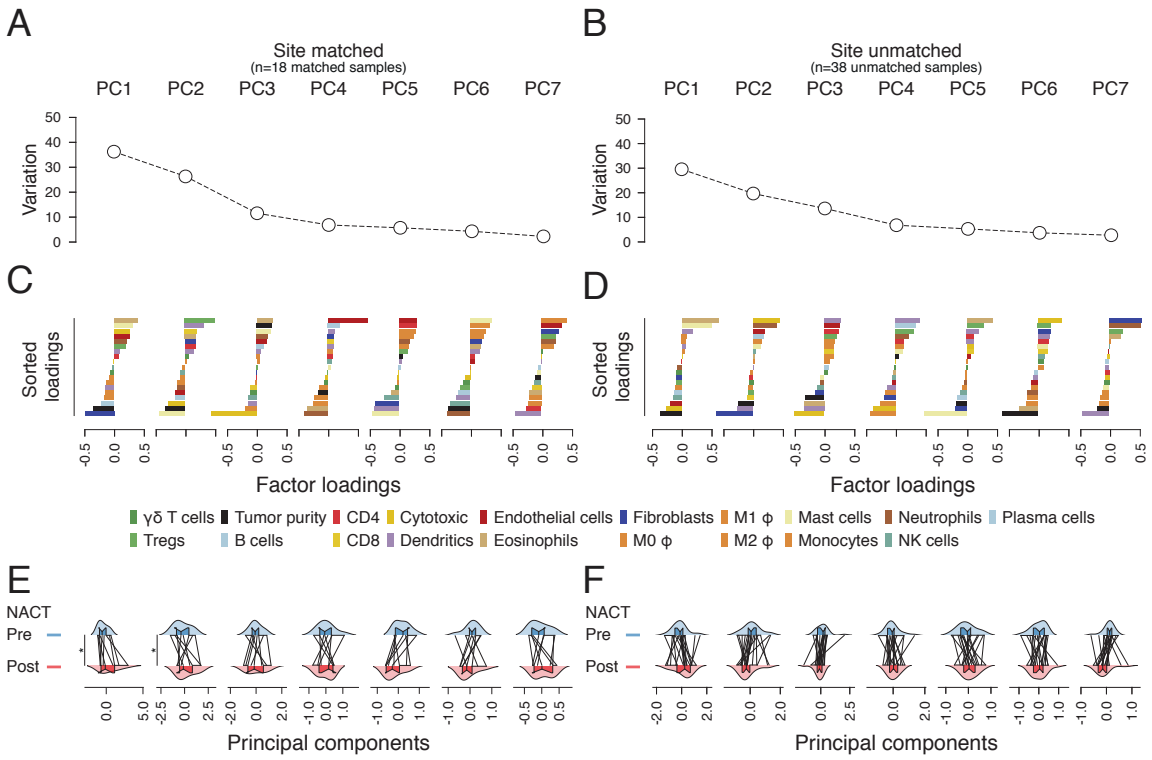


Fig. 8.7 Principal component Consensus^{TME} gene sets' loadings of pre/post-NACT matched and unmatched HGSOC samples, related to Figure 3.6. (A, B) Principal components' variance. (C, D) Loadings of estimated TME cells for each principal component sorted. (E, F) Principal components' paired comparisons. Significant differences after Wilcoxon signed-rank tests are indicated: p-value < 0.05 *.

Chapter 9

Appendix C

Supplementary figures for Chapter 4:
Comprehensive benchmarking and integration of tumour microenvironment cell estimation methods

Tetelestai