# The use of catalytically dead Cas9 to identify key transcription regulators in Triple Negative Breast Cancer

**Elisabetta Zambon**

**Corpus Christi College**

**Department of Pharmacology**

**University of Cambridge**

**This dissertation is submitted**

**for the degree of Doctor of Philosophy**

**September 2019**

*To Mum, Dad and Nicola, for their unconditional love.*

# DECLARATION

I hereby declare that my dissertation entitled 'The use of catalytically dead Cas9 to identify key transcription regulators in Triple Negative Breast Cancer' is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

I declare that this dissertation is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

In accordance with the School of the Biological Sciences guidelines, this dissertation does not exceed 60,000 words.

Elisabetta Zambon

September 2019

# ACKNOWLEDGEMENTS

Firstly, I would like to thank my supervisor, Dr Walid T. Khaled, for his guidance, incredible support, enthusiasm, patience over the past four years, and for being such a good mentor in uncountable occasions. Also, I would like to express my sincere gratitude to my supervisor at AstraZeneca, Dr Claus Bendtsen, for providing unwavering and truly invaluable scientific encouragement during my doctoral studies.

I would further like to extend my sincere gratitude to Dr Aurelie Bornot, Dr Beathe Ehrhardt, Dr Jon DeGnore, Dr Mike Firth and Dr Jonathan Cairns with whom I collaborated extensively through my degree: this project wouldn't have been possible without their help, knowledge and support. The value of the funding provided by AstraZeneca will forever be appreciated.

I would also like to thank all the members of the Khaled laboratory, past and present, for all good memories and the support throughout our time together. In particular, I would like to thank Fazal Hadi for making these four years something I will always remember with joy: I am glad we shared this experience together, I couldn't have asked for a better friend next to me.

In addition, I would like to give a special thanks to Dr Alasdair Russell, a former colleague, for believing in me since we first met five years ago, for teaching me so much when we worked together, for his precious, unconditional support and for his irreplaceable friendship.

In addition, I am extremely thankful to all the amazing people I have met during my time in Cambridge. Nelma, Silvia, Giacomo, Graham, Ammar, Semiramis, Michal, Charlotte and Flavia: everyone of you made this chapter of my life special, and no matters where life will take us, we will always find a way to be there for each other.

A special thank has to go to Sofie, Mathilde, Darcie, Tatjana and Judith, the most incredible women I have met: you girls have been family to me during these last 5 years, you have always been there, in the good and bad times, and there were moments I couldn't have made it without you. I will always smile when I will think about our crazy times together, and I will always love you deeply.

Furthermore, I would like to thank Maria Grazia, Francesca, Alberto G., Matteo, Alberto S., Renato, Nadia, Erica, Valentina and Paola, friends from a life time, for always reminding me where I come from, for always being there with open arms every time I came back.

I would also like to thank Elisabetta Petrozziello, Claudia Vivori, Elisa Magistrati and Beatrice Petrozziello: you are my dearest friends, my closest people, and there is not a single occasion I don't feel so lucky to have you in my life, even if we don't live in the same place anymore. Our friendship is one of the most valuable things I have, and spending time with you during our weekends and holidays around the world helped me to stay sane during the most difficult times of the last years.
Lastly, I would like to say a big thank you to Mum, Dad and my brother Nicola: thank you for being so supportive, so understanding, so caring and proud, even when I didn't believe in myself.

# TABLE OF CONTENT

# LIST OF TABLES

**CHAPTER 1**



**CHAPTER 2**



**CHAPTER 4**

# LIST OF FIGURES

**CHAPTER 1**



**CHAPTER 3**

# LIST OF ABBREVIATIONS AND ACRONYMS

| | |
|---|---|
| ADCC | Antibody-dependent cell-mediated cytotoxicity |
| AIs | Aromatase Inhibitors |
| AKT1 | AKT Serine/Threonine Kinase 1 |
| AL7A1 | Aldehyde Dehydrogenase 7 Family Member A1 |
| AMBIC | Ammonium hydrogen carbonate |
| AP/MS | Affinity purification/mass spectrometry |
| AR | Androgen receptor |
| ATAC-Seq | Assay for Transposase-Accessible Chromatin using sequencing |
| ATCC | American Type Culture Collection |
| ATP | Adenosine triphosphate |
| AUC | Area Under the Curve |
| AZ | AstraZeneca |
| BC | Breast cancer |
| BCA | Bicinchoninic acid assay |
| BCL11A | B-cell lymphoma/leukaemia 11A |
| BL1 | Basal-like 1 |
| BL2 | Basal-like 2 |
| BlastR | Blastocidin Resistance |
| BLBC | Basal-like breast cancer |
| BRCA1 | BRCA1 DNA Repair Associated |
| BRCA2 | BRCA2 DNA Repair Associated |
| BRD4 | Bromodomain-containing protein 4 |
| BSA | Bovine Serum Albumin |
| c-JUN | Transcription factor jun-C |
| Cas | CRISPR-associated |

| | |
|---|---|
| Cas9 | CRISPR-associated protein 9 |
| CAV1A | Caveolin-1 |
| CCND1 | Cyclin D1 |
| Cdc25 | Cell Division Cycle 25A |
| CDH1 | Cadherin 1 |
| CDK1 | Cyclin-dependent kinase 1 |
| CDK2 | Cyclin-dependent kinase 2 |
| CDK4 | Cyclin-dependent kinase 4 |
| CDK6 | Cyclin-dependent Kinase 6 |
| CDK7 | Cyclin Dependent Kinase 7 |
| CDK9 | Cyclin Dependent Kinase 9 |
| cDNA | Complementary DNA |
| CENPF | Centromere Protein F |
| CTFs | Collaborating transcription factors |
| CHD4 | Chromodomain Helicase DNA Binding Protein 4 |
| CHD8 | Chromodomain Helicase DNA Binding Protein 8 |
| ChIP | Chromatin Immunoprecipitation |
| ChIP-Seq | Chromatin immunoprecipitation Sequencing |
| CHTOP | Chromatin Target of PRMT1 Protein |
| CIGC | Cambridge Institute Genomic Core |
| CKIs | Cyclin dependent kinases inhibitors |
| CNAs | Copy Number Alterations |
| CRAPome | Contaminant Repository for Affinity Purification |
| CRISPR | Clustered Regularly Interspaced Short Palindromic Repeats |
| CRISPRa | CRISPR activation |
| CRISPRi | CRISPR interference |
| crRNA | CRISPR RNA |

| | |
|---|---|
| CRUK | Cancer Research UK |
| CRUK-CI | Cancer Research UK-Cambridge Institute |
| CSC | Cancer Stem Cells |
| CTD | Carboxy-terminal domain |
| Da | Dalton |
| dCas9 | Catalytically dead Cas9 |
| DCIS | Ductal carcinoma *in situ* |
| DMEM | Dulbecco's modified Eagle's medium |
| DNA | Deoxyribonucleic acid |
| dNTP | deoxyribose Nucleoside Triphosphate |
| Dox | Doxycycline |
| E. coli | Escherichia coli |
| E2F1 | E2F Transcription Factor 1 |
| E2F2 | E2F Transcription Factor 2 |
| E2F8 | E2F Transcription Factor 8 |
| ECD | Extracellular domain |
| ECL | Enhanced chemiluminescence |
| ECM | Extra cellular matrix |
| EDTA | Ethylenediaminetetraacetic acid |
| EGFP | Enhanced green fluorescent protein |
| EGTA | Ethylene glycol-bis (_-aminoethyl ether)-N,N,N_,N_ -tetraacetic acid |
| ELF5 | E74 Like ETS Transcription Factor 5 |
| EMT | Epithelial-Mesenchymal Transition |
| enChIP | Engineered DNA-binding molecule-mediated chromatin immunoprecipitation |
| enChIP-MS | Engineered DNA-binding molecule-mediated chromatin |

| | |
|---|---|
| | immunoprecipitation mass-spectrometry |
| enChIP-chip | Engineered DNA-binding molecule-mediated chromatin immunoprecipitation chromatin immunoprecipitation |
| ERBB2 | Erb-B2 Receptor Tyrosine Kinase 2 |
| eRNAs | Enhancer-associated RNAs |
| ESR1 | Oestrogen receptor 1 |
| EYA2 | Eyes Absent Homolog 2 |
| FACS | Fluorescence-activated cell sorting |
| FAIRE-Seq | Formaldehyde-Assisted Isolation of Regulatory Elements sequencing |
| FBS | Fetal Bovine Serum |
| FDR | False Discovery Rate |
| FGFR1 | Fibroblast Growth Factor Receptor 1 |
| FOXA1 | Forkhead Box A1 |
| FOXC1 | Forkhead Box C1 |
| FOXM1 | Forkhead Box M1 |
| G2M | G2/M Phase-specific |
| GAPDH | Glyceraldehyde 3-phosphate dehydrogenase |
| GATA3 | GATA Binding Protein 3 |
| GFP | Green fluorescent protein |
| GFTa | General transcription factors |
| GLoPro | Genomic locus proteomics |
| gRNA | guide-RNA |
| GSEA | Gene Set Enrichment Analysis |
| H2A | Histone H2A |
| H2B | Histone H2B |
| H3K36 | Histone H3 lysine 36 |

| | |
|---|---|
| HBSS | Hanks' Balanced Salt Solution |
| HCl | Hydrogen chloride |
| HDR | Homologous-directed repair |
| HER2 | Erb-B2 Receptor Tyrosine Kinase 2 |
| HOXA10 | Homeobox A10 |
| HTR | Hormone replacement therapy |
| Hz | Hertz |
| iCh-IP | Insertional ChIP |
| IBC | Invasive breast cancer |
| ICAT | Isotope-coded affinity tag |
| IGEPAL | Octylphenoxypolyethoxyethanol |
| IgG | Immunoglobulin G |
| IL-1β | Interleukin-1beta |
| IM | Immunomodulatory |
| IMI | Institute for Mathematical Innovation |
| INK4 | Inhibitors of CDK4 and CDK6 |
| IntClust | Integrative Clusters |
| IP | Immuno-precipitation |
| iTRAQ | Isobaric tags for relative and absolute quantification |
| ITRs | Inverted terminal repeat sequences |
| JUNB | Transcription factor jun-B |
| JUND | Transcription factor jun-D |
| K18 | Keratin 18 |
| K19 | Keratin 19 |
| K8 | Keratin 8 |
| kDa | kilo Dalton |
| KOH | Potassium hydroxide |

| | |
|---|---|
| KRT14 | Cytokeratin 14 |
| KRT17 | Cytokeratin 17 |
| KRT5 | Cytokeratin 5 |
| LAR | Luminal androgen receptor |
| LB | Lysogeny Broth Media |
| LC/MS | Liquid chromatography-mass spectrometry |
| LC/MS/MS | Liquid chromatography Tandem Mass-spectrometry |
| LDTFs | Lineage-determining transcription factors |
| LN | Lymph node |
| LOH | Loss of heterozygosity |
| M | Mesenchymal |
| MAP3K1 | Mitogen-Activated Protein Kinase Kinase Kinase 1 |
| MAPK | Mitogen-activated protein kinase |
| MBD3 | Methyl-CpG-binding Domain Protein 3 |
| MEK | MAPK kinase |
| MEP50 | Methylosome Protein 50 |
| METABRIC | Molecular Taxonomy of Breast Cancer International Consortium |
| MgCl2 | Magnesium chloride |
| MMP7 | Matrilysin |
| MS | Mass-spectrometry |
| MS/MS | Tandem Mass-Spectrometry |
| MSL | Mesenchymal stem-like |
| MTA1/2/3 | Metastasis-associated protein 1/2/3 |
| mTOR | Mechanistic target of rapamycin |
| MYC | Myc Proto-Oncogene |
| NaCl | Sodium chloride |

| | |
|---|---|
| NCBI | National Centre for Biotechnology Information |
| NCBI-NR | National Center for Biotechnology Information Non reduntant |
| NDRs | Nucleosome-depleted regions |
| NF-κB | Nuclear factor of κB |
| NFE2L3 | Nuclear Factor (Erythroid 2) - Like Factor 3 |
| NFIB | Nuclear Factor I B |
| NHEJ | Non-homologous end joining |
| NuRD | Nucleosome Remodeling Deacetylase |
| P-TEFb | Positive transcription elongation factor b |
| p.p.m. | Parts Per Million |
| P/S | Penicillin/Streptomycin |
| p107 | Retinoblastoma-like Protein 1 |
| p130 | Retinoblastoma-like Protein 2 |
| p65AD | p65 activation domain |
| PAM | Protospacer-adjacent motif |
| PARP | Poly-ADP ribose polymerase |
| PARPi | PARP inhibitors |
| PBS | Phosphate-buffered saline |
| PCR | Polymerase chain reaction |
| PD-L1 | Programed death-ligand 1 |
| PDIP3 | Polymerase Delta-interacting Protein 3 |
| PDXs | Patient derived xenografts |
| PI3K | Phophoinositide 3-kinase |
| PIC | Pre-initiation complex |
| PIK3CA | Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Alpha |
| PMTs | Post-translational modifications |

| | |
|---|---|
| PSM | Peptide Spectrum Match |
| PTEN | Phosphatase And Tensin Homolog |
| PVDF | Polyvinylidene fluoride or polyvinylidene difluoride |
| q-PCR | Quantitative PCR |
| Q-Q plot | Quantile-Quantile Plot |
| Rb | Retinoblastoma-associated Protein |
| RB1 | RB Transcriptional Corepressor 1 |
| RbAp46 | Retinoblastoma-associated Protein46 |
| RbAp48 | Retinoblastoma-associated Protein48 |
| RBBP4/7 | Histone-binding Protein RBBP4/7 |
| RIME | Rapid immunoprecipitation mass spectrometry of endogenous protein |
| RIPA | Radioimmunoprecipitation assay buffer |
| RNA | Ribonucleic acid |
| RNA-Seq | RNA sequencing |
| RNP | Ribonucleoprotein |
| RSC (complex) | Chromatin structure remodeling |
| RT | Reverse Transcriptase |
| RT-PCR | Reverse transcription polymerase chain reaction |
| S.D. | Standard Deviation |
| SAM | Synergistic activation mediator |
| scRNAs | scaffold RNAs |
| SDS | Sodium Dodecyl Sulphate |
| SDTFs | Signal-dependent transcription factors |
| SERM | Selective oestrogen receptor modulator |
| SET2 | Histone-lysine N-methyltransferase |
| sgRNA | single gRNA |

| | |
|---|---|
| shRNA | short hairpin RNA |
| SILAC | Stable isotope labelling by amino acids in cell culture |
| SMA | Smooth muscle actine |
| SMAD2/3 | Mothers Against Decapentaplegic Homolog 2/3 |
| SOC | Super Optimal broth with Catabolite repression media |
| SOX10 | Transcription Factor SOX-10 |
| STAT1 | Signal transducer and activator of transcription 1 |
| STATs | Signal Transducers and Activators of Transcription |
| SumAUC | Sum of the Area Under the Curve |
| SumAUCnorm | Sum of the Area Under the Curve Normalized |
| TAL | Transcription activator-like |
| TCGA | The Cancer Genome Atlas |
| TE/NaCL | Tris EDTA Sodium chloride |
| TEB | Terminal end bud |
| TERT | Telomerase reserve transcriptase |
| Tet-on | Tetracycline On |
| TF | Transcription factor |
| TFIIA | General Transcription Factor IIA |
| TFIIB | General Transcription Factor IIB |
| TFIID | Transcription factor IID |
| TFIIE | Transcription factor IIE |
| TFIIF | Transcription factor IIF |
| TFIIH | Transcription factor IIH |
| TGFβ | Tumour necrosis factor-beta |
| Tle3 | Transducin-like Enhancer Protein 3 |
| Tle4 | Transducin-like Enhancer Protein 4 |
| TNBC | Triple Negative Breast Cancer |

| | |
|---|---|
| TNF-α | Tumour necrosis factor-alpha |
| TP53 | Tumor Protein P53 |
| tracrRNA | Transactivating CRISPR RNA |
| TRE | Tet Response Element |
| TSS | Transcription Starting Site |
| tTA | Tetracycline transactivator |
| TWIST1 | Twist Family BHLH Transcription Factor 1 |
| UCSC | University of California, Santa Cruz Genome Browser |
| VP64 | Herpes simplex VP16 activation domain |
| WNT | Proto-oncogene Wnt-1 |
| XIC | Extracted ion chromatogram |
| Ybox1 | Nuclease-sensitive Element-binding Protein 1 |

# LIST OF APPENDICES

ELISABETTA ZAMBON

# THE USE OF CATALYTICALLY DEAD CAS9 TO IDENTIFY KEY TRANSCRIPTION REGULATORS IN TRIPLE NEGATIVE BREAST CANCER

## ABSTRACT

Triple-negative breast cancer (TNBC) accounts for approximately 15-20% of all breast cancer cases. It tends to be aggressive, high-grade and poorly differentiated tumour with poor clinical outcome. Lack of expression of oestrogen, progesterone receptors, and human epidermal growth factor receptor 2 make TNBC patients ineligible to hormonal therapy. For these reasons the identification of novel clinical targets still remains a priority. With the advances of multiomics, many of the genes transcriptionally upregulated in TNBC have been identified, but how they are disregulated is still unknown: the understanding of how this works at a transcriptional level could contribute to the development of a novel therapeutic approach.

We report here a new methodology to identify key transcription regulators that we applied to investigate the expression of highly expressed genes in TNBC: our approach combines RIME proteomics with CRISPR/Cas9 technology. In brief, we targeted putative regulatory regions of differentially regulated genes in TNBC compared to other subtypes of breast cancer using a catalytically dead version of the Cas9 protein (dCas9). In particular, we focused on transcription regulators like *FOXC1*, *NFIB* and *NFE2L3*. We then performed RIME proteomics to identify which proteins are in close proximity to dCas9 and thus potentially bound to these putative regulatory regions. In addition, we developed a novel, statistical approach to analyse these particular proteomic datasets based on the relative abundance of the protein of interest, and a powerful ranking method to identify biologically and therapeutic meaningful candidates.

Through this process, we identified three putative regulatory proteins, MTA2, CDK1 and CDK6, bound to all three *loci*. In here, we reported how their knockdown, performed by shRNA, directly affects the expression of the investigated genes in a panel of TNBC cell lines, and their oncogenic capacity *in vitro*. These results demonstrate the importance of these transcription regulators for TNBC biology, and they highlight the necessity of a deeper understanding of their roles in gene expression regulation.

# CHAPTER 1: INTRODUCTION

## 1.1 Introduction

Breast cancer is the most common cause of cancer death in women, with around 627,000 deaths in 2018 worldwide (Bray et al., 2018). In the UK, 1 in 8 women will develop breast cancer during their lifetime, according to the CRUK statistics of 2018. However, its mortality rates are decreasing, thanks to a significant improvement of the diagnostic techniques, surgical approaches and pharmacological treatments (World Health Organization, 2018).

### 1.1.1 Breast cancer epidemiology

The socio-economic background of the patient plays an important role in breast cancer incidence and mortality: it has been shown that women living in more socio-economically deprived environments have a lower cancer incidence, but a higher mortality (Levi et al., 2004). This is likely to be related to different access to early detection screening and treatment, as for example later diagnosis, less efficient cancer care, limited treatment options or higher risk of treatment complications.

In addition, the tumour incidence is related to different risk factor profiles including age, race, reproductive patterns, breast characteristics, hormone replacement therapy (HRT), diet, use of tobacco and alcohol (Winters et al, 2017).

Furthermore, an essential role is played by genetics: every cancer is in fact characterized by a set of genes that, if mutated, strongly associates with

tumourigenesis. These are known as driver genes (Stratton et al., 2009), and they are responsible of cancer growth advantage. These genetic changes are usually somatic, but germline mutations can predispose heritable or familial cancer. Thanks to next generation sequencing, over 30626 mutations have nowadays been reported for breast cancer (Cancer Genome Atlas N., 2012; Rajendran et Chu-Xia, 2017): the majority of them occurs at somatic level, while the germline ones are responsible for 5 to 10% of familial breast cancers types (Turnbull et Rahman, 2008). The most established model of breast cancer susceptibility is that the disease is due to several mutations in high-penetrant genes, such as *BRCA1*, *BRCA2*, *TP53*, *PTEN*, *STK11*, and *CDH1*, and a larger number of moderate penetrant genes like *CHK2*, *ATM*, *BRIP1*, and *PALB2* (Mavaddat et al., 2010; Apostolou et al., 2013). However, the contribution of low-penetrant genes for the development of sporadict breast cancer remains uncertain (Balmain et al., 2003).

Advances in high-throughput genotyping have allowed the discovery of germline mutations, primarily single-nucleotide polymorphisms (SNPs) associated with a higher risk of developing the disease. The first major genes associated with hereditary breast cancer were *BRCA1* and *BRCA2*: their mutations, inherited in an autosomal dominant way, act recessively on the cellular level as tumor suppressor genes involved in double-stranded DNA (dsDNA) break repair (Stratton et al., 2008). The lifetime risk of familial breast cancer for women carrying these mutations is 50%–85% (King et al., 2003; Shiovitz et al., 2015). Additional rare, but highly penetrant genes include: *PTEN,* where the lifetime risk is estimated around 85% (Tan et al., 2012); *TP53,* with a 25% risk by age 25 (Ray et al., 2001; Shiovitz et al., 2015); *CDH1*, with a 39% lifetime risk (Pharaoh et al., 2001; Shiovitz et al., 2015), and *STK11*, with a 32% risk by the age of 60 (Lim et al., 2004; Shiovitz et al., 2015), each conferring a distinct clinical syndrome. On the basis of mathematical modeling, it is estimated that these genes account for 25% of cases (Walsh et al., 2006). For these patients, breast awareness and breast self-exam is recommended to start at age 18. From age 25, clinical breast exam, and imaging with a combination of mammography and magnetic resonance imaging (MRI) should be carried out annually (Shiovitz et al., 2015).

The recent surge of next generation sequencing of cancer genome has led to the discovery of novel, somatic driver mutations acquired during tumorigenesis (Shah et al., 2012; Ellis et al., 2012; Stephens et al., 2012; Banerji et al., 2012): with the exception of well-known genes like *P53* and *PIK3CA* mutated in more than 30% of

breast cancer patients, these studies highlighted that the majority of the novel frequently mutated ones are present in less than 10% of the cases. In addition, they revealed the large genetic diversity among different tumours: Stephens et al. for example showed that among the 100 investigated breast cancers they identified 73 different combination of mutated genes (Stephens et al., 2012). However, these genes could be grouped in similar pathways: they in fact demonstrated that 6 cancer genes were acting in the same JUN kinase pathway (Stephens et al., 2012), while Shah et al. discovered that pathways involving P53, PIK3 and ERBB were over-represented in the mutated genes (Shah et al., 2012). These discoveries imply that even if genetically different, some tumours could have similar phenotypes because of mutations in the same pathway. In addition, in some tumours there was no obvious driver mutation, suggesting different mechanism responsible of tumour development such as for example DNA methylation (Desmedt et al., 2012).
Large multicentre projects, like The Cancer Genome Atlas (TCGA) (Cancer Genome Atlas N., 2012) and the International Cancer Genome Consortium (ICGC) (Hudson et al., 2010) have performed detailed analyses of the somatic alterations affecting tumour genomes in various cancers, including breast. Large consortia and networks, as for example COSMIC and GENIE, are collecting mutation data from different sources to implement the understanding of the mutational landscape in cancer (Forbes et al., 2016; Consortium APG, AACR Project GENIE, 2017), in order to provide evidence on potential associations of genomic information with cancer subtypes, development of metastasis and prognosis. These findings will help to increase the knowledge of the disease and the discovery of driver events.

However, the genetic diversity, the correlated different phenotypes and functional features within a patient's tumour (intra-tumour heterogeneity) and among tumours from different patients (inter-tumour heterogeneity) are characteristics that complicate diagnosis and challenge therapy for breast cancer patients.

## 1.2 Breast cancer development

The natural development, progression and characteristics of breast neoplasms, together with the normal breast architecture, are important components to understand tumour heterogeneity.

The mammary gland itself is one of the tissues that undergo a significant, dynamic remodelling during development and throughout life. Deriving from ectodermal cells, it forms a primordial ductal tree (the mammary primordium) before birth, followed by postnatal development regulated by hormones, and the ductal tree outgrowth into the mammary fat pad induced during puberty.

Further changes within the gland are related to pregnancy, with the formation of alveoli for milk secretion during lactation. At the end of this process, the ductal tree regresses to a pre-pregnancy-like state (involution) (Macias et Hinck, 2012; Sternlicht, 2006) (Fig. 1.1, from Manavathi et al., 2014).

Two major cell types form the mammary epithelium, organized in a bi-layered structure: a luminal layer, lining the duct, and a contractile myopeithelial one, with a basal location. The luminal one is formed by cells expressing keratin 8/18/19 (K8/18/19) and/or oestrogen and/or progesterone receptor (ER/PR), while the myoepithelial one by cells expressing K5/14 and/or smooth muscle actin (SMA) and/or p53 (Bissell et al., 2003; Hennighausen et Robinson, 2001).

Tumour initiation and progression involve several pathological and clinical stages, with complex series of stochastic genetic events leading towards invasive phenotype: a ductal hyperproliferation (ductal carcinoma *in situ*, or DCIS), followed by invasive carcinomas (IC), and finally metastatic disease (Allred et al., 2001; Burstein et al., 2004). Metastases usually colonize bone, lung, brain, and liver, and they are considered the main cause of breast cancer mortality.

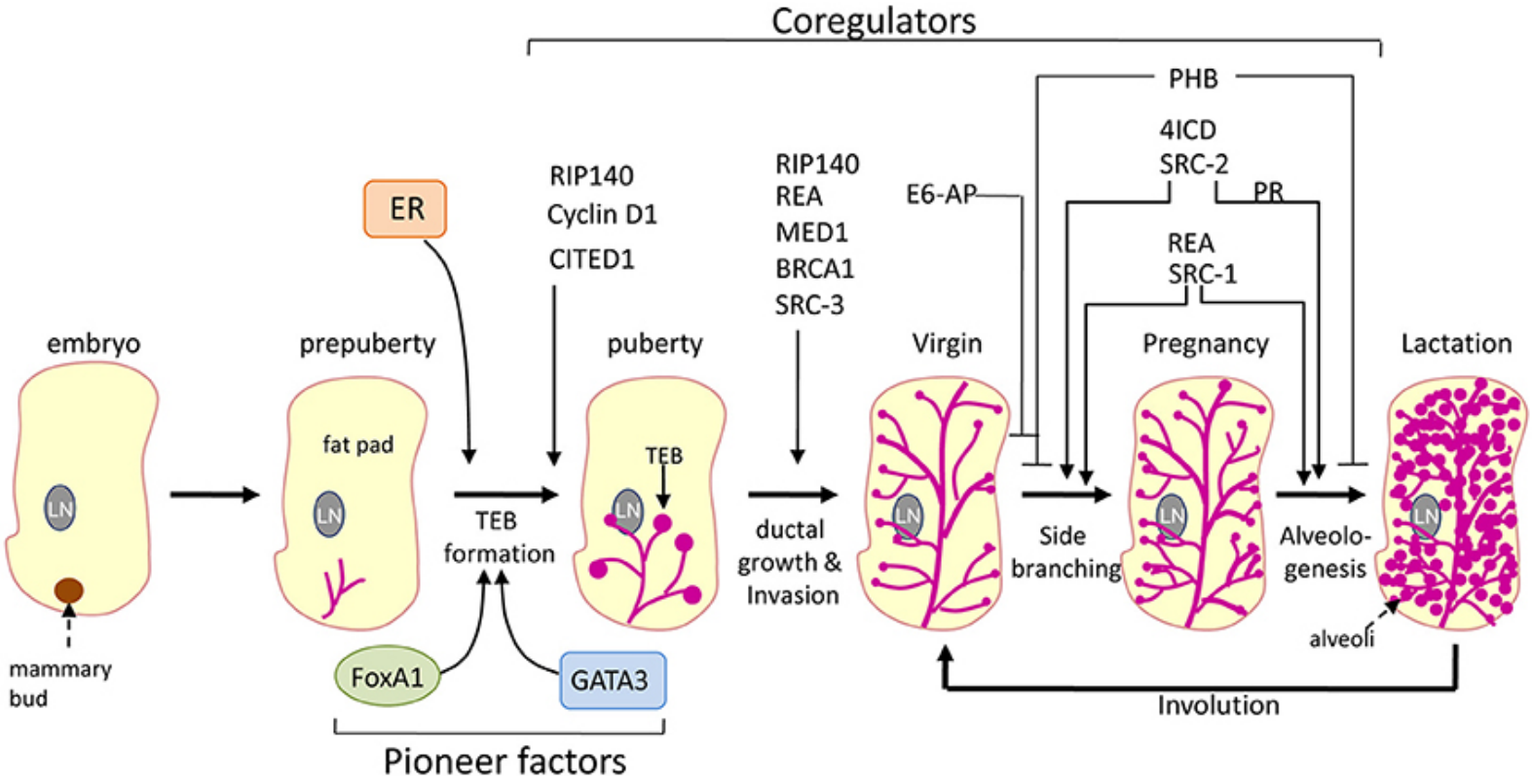



**Figure 1.1: Stages and pioneer factors of adult mammary gland development.** A schematic representation of the stages of mammary gland development from pre-puberty through to pregnancy, lactation and involution. LN, lymph node; TEB, terminal end bud (Manavathi et al., 2014).

### 1.2.1 Molecular characteristics and breast cancer classification

Nowadays, patient management is still limited to clinical characteristics and classic histological analysis. In particular, parameters such as receptor status (oestrogen-receptor, ER, progesterone-receptor, PR, and/or ERBB2 receptor, HER2) contribute to diagnostic classification (Viale et al., 2012), together with tumour size, histological grade and axillary lymph-node involvement that have been shown to correlate with clinical outcome (Dawson et al., 2013). Hormone receptor-positive breast cancers account for approximately 75-80% of all cases, while *HER2*$^{+}$ subtype for approximately 10-15% (Konecny et al., 2006). The remaining 10-15% of the cases are defined by the absence of hormone receptors and HER2 hence it is commonly known as triple negative breast cancer (TNBC).

Over the past 20 years microarray-based technologies for gene expression analysis have given a new comprehensive molecular profiling, subdividing breast cancer into at least six subtypes: normal like, luminal A (ER and PR positive), luminal B (ER, PR and HER2 positive), HER2-enriched, claudin low, and basal like by hierarchical clustering (Blows et al., 2010; Perou et al., 2000; Prat et al., 2010; Santagata et al., 2014; Sørlie et al., 2001, Sørlie et al., 2003). These expression profiles mostly reflect different clinical and pathological prognoses: using gene sets optimally selected to identify the intrinsic characteristics of breast tumours, these studies analysed large cohorts of samples and associated different mortality rates to different subtypes.

Perou and colleagues were the first ones to show that breast tumours could be classified into different subtypes according to different expression profiles (Perou et al., 2000). Using 40 breast tumours and 20 matched pairs of samples before and after doxorubicin treatment, 476 genes were selected as 'intrinsic gene set' more variable between the tumours rather than the paired samples. On the basis of expression levels and hierarchical clustering analysis, tumours were separated in two branches: one characterized by low/absent expression of ER, and one characterized by its expression. The first group was further divided into the basal-like subtype (high expression of keratins 5 and 17, laminin and fatty acid binding protein), the ERBB2+ subtype (high expression of several genes in the ERBB2 amplicon at 17q22.24 including *ERBB2* and *GRB7*), and normal breast-like group (with the highest expression of genes normally present in the adipose tissue or other nonepathilial cell types). The second group was divided into luminal subtype A

(highest expression of the *ESR1* gene, *GATA3*, *XBP1*) and luminal subtype B (low to moderate expression of luminal genes.

However, this approach requires large datasets, and cannot be applied for the classification of individual samples prospectively (Mackay et al., 2011). For this reason, 'single sample predictors' (SSPs) were developed on the basis of the correlation between the expression profile of a given sample with the centroids for each molecular subtype (usually the average expression profile of every molecular subtype) (Sørlie et al., 2003; Parker et al., 2009). In addition, on the basis of this approach, Parker et al. developed a quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR)-based or NanoString-based method (PAM50) to classify formalin-fixed paraffin-embedded (FFPE) samples into the molecular subtypes (Parker et al., 2009). These molecular signatures not only helped understanding the biological spectrum of breast cancers, but also provided diagnostic, prognostic and predictive gene signatures, useful for the identification of new therapeutic targets.

Despite the enthusiasm using this molecular taxonomy for clinical trials design and oncology practise, there are several limitations that have to be considered. Firstly, the subdivision of luminal tumours into A and B is strongly dependent on the SSP used (Weigelt et al., 2010) and highly depends on the expression of genes involved in proliferation (Parker et al., 2009), making this subgroup division artificial (Parker et al., 2009). In addition, normal breast-like cancer is now considered by some to be an invalid molecular subtype: they believe these tumours are likely to constitute an artefact of frozen tissue procurement and representation (Parker et al., 2009; Weigelt et al., 2010; Pratet al., 2010). Lastly, $HER2^+$ subtype as defined by micro-arrays does not include all cases so classified by clinically validated methods such as IHC, and vice versa (Parker et al., 2009; Weigelt et al., 2010). However, these discrepancies do not invalidate the existence of the 'intrinsic' subtypes (Perou et al., 2010): it is in fact an evolving system for classification, and its future testing will determine the prognostic and predictive power and clinical utility.

These profiles, together with integrated genomic and transcriptomic analysis of breast tumours, have revealed further subgroups with specific patters of recurrent somatic mutations, therefore specific tumour biology and distinct clinical outcomes (Stephens et al., 2012; TCGA, 2012). These type of data collected from 2000 patients have been integrated in the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC (Curtis et al., 2012)), and 10 new integrative

clusters (IntClust 1-10) were associated with distinct CNAs (Copy Number Alterations) and gene expression changes. This clearly demonstrates the heterogeneity present within tumours previously classified according to ER, PR and HER2 expression, and it further separates them into different subtype groups.

### 1.2.2 Intra-tumour heterogeneity

The complexity of the disease can challenge the accuracy of the prognosis: in fact even if biopsies are usually collected from multiple regions of the tumour, further aggressive areas may still be missed due to their scarcity and/or topological heterogeneity (Komaki et al., 2006). Furthermore, a systematic and comprehensive evaluation of the molecular features of metastases is still necessary, since they could carry different genetic and non-genetic alterations when compared to the bulk of the primary tumour (Ding et al., 2010; Shah et al., 2009).

Yates et al. demonstrated that metastases could derive from subclones in the primary cancer: thanks to a whole genome sequencing of primary tumour and metastatic biopsies, they confirmed the development of metastases from very early stage of genetic diversification of the primary tumour. This highlighted the necessity of understanding the pattern of subclonal diversification in primary tumour (Yates et al., 2015). In their study they also found this heterogeneity to be present in all major immunohistological subgroups of breast cancer, even though definitive conclusions about heterogeneity in any particular subtype couldn't be drawn (Yates et al., 2015). They sequenced mutliregion samples from 50 invasive breast cancers (27 $ER^+$, 3 $HER2^+$, and 20 triple negative) in order to determine the patterns of spatial evolution, and they identified recurrent driver mutations in oncogenes and missense substitutions in tumours suppressor genes (Lawrence et al., 2014; Stephens et al., 2012; Yates et al., 2015)

To evaluate the spatial distribution of subclones, they performed targeted gene sequencing from primary tumours and correspondent lymph node metastases when possible: the predominant pattern of heterogeneity they identified was a geographically restricted expansion in the majority of the tumours (Yates et al., 2015). In addition, they sequenced more than one focus (2-5) of 4 multifocal cancers: different foci of the same tumour were clonally related but with private mutations. This indicated that during its own growth, each focus was characterised by a complete overcome of one clone over the remaining tumour cells in that region.

Among these private mutations, *BRCA2* and *CDKN2A* inactivation, *PTEN* point mutation and *CDK6* amplification were identified (Yates et al., 2015).

In the recent years, a huge effort has been put in the parallel sequencing analysis of primary and metastatic breast cancers (Yates et al., 2015; Ng et al., 2017), revealing that a variable proportion of somatic mutations are restricted or enriched in the metastatic lesion compared to the primary tumour, even of some affecting driver genes like *PIK3CA*, *SMAD4* and *TP53* (Schrijver et al., 2018). In addition, researchers have observed marked single nucleotide and copy number differences between primary breast carcinomas and metastases (Ding et al., 2010; Shah et al., 2009). This genomic heterogeneity confirms the substantial genetic evolution acquired during disease progression, and could explain why some biomarkers specific for the primary tumour might not be informative to predict a therapeutic response.

For this reason, novel approaches have been used to investigate differences between primary site and metastases. The progressive Intensive Trial of Omics in Cancer (ITOMIC) for example was designed to enroll patients with triple negative breast cancer to a specific therapy on the basis of the molecular profile of the cancer over space (primary vs metastases) and time (progression) (Blau et al., 2016). The genome analysis from multiple biopsies have demonstrated an extensive spatial and temporal heterogeneity in single nucleotide variants, CNV, insertion or deletion polymorphisms during treatment and revealed the evolution of molecular signatures (Soon-Shiong et al., 2006).

Several phase III trials have also been conducted to evaluate the effect of extended endocrine treatment with tamoxifen (ATLAS (Davies et al., 2013)) or aromatase inhibitors (MA.17 (Goss et al., 2005), NSABP-B33 (Mamounas et al., 2008) and ABCSG (Jakesz et al., 2007)) beyond 5 years for $ER^+$ breast cancer. It is becoming clearer that longer endocrine therapy can reduce the risk of late metastases in the second decade after initial breast cancer diagnosis and treatment (Davies et al., 2013; Mamounas et al., 2008; Jakesz et al., 2007).

In addition, few retrospective studies have used multi-gene signatures to predict late recurrence risk in ER positive breast cancer: PAM50 risk-of-recurrence (ROR) score for example differentiates patients on the basis of the risk for late recurrence beyond conventional prognostic factors (Filipits et al., 2014), using PAM50 intrinsic subtypes, tumour size, proliferation, number of positive lymph nodes to categorize patients into low, intermediate or high-risk groups. EndoPredict, a qRT-PCR-based

score, combines the expression levels of proliferative and ER signalling genes for patients at risk of developing late distant metastases at 10 years of follow-up; the Breast Cancer Index (BCI), a qRT-PCR assay based on the five-gene molecular grade index (MGI) and the HOXB13/IL17B ratio) (Sgroi et al., 2013) and the 70-gene microarray prognosis signature MammaPrint® (Drukker et al., 2014) also contribute to the identification of late distant recurrences in selected subgroups. These predictors may be helpful in identifying patients for extended therapy after 5 years of initial endocrine treatment.

Some differences in the preferences of site for metastatic relapse have been identified between intrinsic subtypes of breast cancer (Smid et al., 2008; Soni et al., 2005; Kennecke et al., 2010): the skeleton is more frequent between $ER^+$, while $HER2^+$ breast cancers frequently have metastases in the brain, liver and lung (Aversa et al., 2014). On the other hand, patients with ER negative breast cancer commonly have lung metastases, but other visceral sites, like the brain, are also common among triple-negative (TNBC) or basal-like breast cancer cases [(Smid et al., 2008; Soni et al., 2015; Kennecke et al., 2010). The molecular subtype of the primary tumour could then potentially serve as a biomarker for prediction of future metastatic sites (Viale et al., 2007; Piccart-Gebhart et al., 2005). Other conventional factors associated with a higher risk of recurrence include age at the time of primary tumour diagnosis (Kollias et al., 1997), TNM staging (size, nodal status, de novo distant metastatic disease (Chiang et al., 2008)) and tumour histological grade. The risk may also vary over time: while ER negative patients usually develop metastases within 5 years, approximately 50% of recurrences in patients with $ER^+$ disease will occur after a more protracted period (beyond 5 years (Early Breast Cancer Trialists' Collaborative, 2011).

These observations suggest that metastatic process could be the convergent result of distinct genetic and epigenetic mechanisms in different patients. For this reason, further studies are necessary in order to understand it and the role of selective pressure of targeted therapies on intratumour heterogeneity.

## 1.3 Triple Negative Breast Cancer (TNBC)

One of the main challenges in treating this disease is that breast cancer is not a single entity, but a heterogeneous group of several subtypes with different biological

and clinical behaviour. Over the years, different parameters have been used to classify these tumours, but the most common ones are histopathological types in conjunction with the presence or absence of ER, PR and HER2. Hormone receptor-positive breast cancers account for approximately 75-80% of all cases, while *HER2* overexpressing subtype accounts for approximately 10-15% (Konecny et al., 2006). Triple negative breast cancer (TNBC) counts for the remaining 10-15% of the cases, and it is characterized by the absence of expression of the receptors mentioned above.

TNBC occurs more frequently in pre-menopausal women, in particular in the Africa-American female population, where it affects 39% of the patients (Carey et al., 2006), while only 16% of the Caucasian women develop it in the same age group (Trivers et al., 2009). TNBC tumours tend to be high-grade, poorly differentiated, with high mitotic and necrotic count (Ismail-Khan et Bui, 2010). They are also characterized by stromal lymphocytic infiltrate, pushing borders of invasion and increased propensity for metastases to brain and lungs (Tsuda et al., 2000), cellular pleomorphism and high nuclear-cytoplasmic ratio (Dawson et al., 2009).

Histologically, the majority of TNBCs are high-grade invasive ductal carcinomas (IDCs or invasive carcinomas of no special type), which have pushing borders, marked nuclear pleomorphism, and numerous mitoses and often have geographic zones of necrosis and brisk lymphocytic infiltrates (Foulkes et al., 2010). However, several rare histologic high-grade, special-type breast cancers are significantly enriched in TN phenotype, such as carcinomas with apocrine differentiation, carcinomas with medullary features, and metaplastic breast carcinomas (MBCs) Bertucci et al., 2006). These histologic types share with conventional TNBCs a similar genomic landscape, still maintaining some clinically relevant singularities (Geyer et al., 2017).

For example, carcinomas with apocrine differentiation are typically ER/PR negative, but with higher rate of *HER2* amplification (Vranic et al., 2010); they occur in older patients and seem to have a favorable prognosis (Mills et al., 2016). They are characterized by a higher frequency of mutations in *PIK3CA* and *PI3K* pathway genes (Lehmann et al., 2014; Weisman et al., 2016), but a lower rate of *TP53* mutations and *MYC* gains compared with other TNBCs (Weisman et al., 2016).

Carcinomas with medullary features are characterized by well-circumscribed borders, a syncytial growth pattern, and brisk lymphocytic infiltrate (Lakhani, 2012). Despite their cytologic features and high mitotic activity, they have been associated

with a favorable outcome. However, typical medullary carcinomas may not have a better prognosis than atypical ones (not fulfilling all diagnostic criteria) (Mateo et al., 2016), and precise identification of this subtype is limited by doctors' agreement (Lakhani et al., 2012).
Regarding the MBCs, these tumours are often high grade, with conspicuous nuclear pleomorphism and mitotic activity, with squamous and/or mesenchymal differentiation (Weigelt et al., 2014). They are resistant to chemotherapy and have worse outcome (Lung et al., 2010). In addition, high inter and intratumor heterogeneity have been observed at the transcriptomic and genetic levels, correlating with morphologic heterogeneity (Weigelt et al., 2015; Geyer et al., 2015).

In terms of pathological, molecular and clinical characteristics, TNBC shares similarities with the 'basal-like' subtype (BLBC). This term was chosen to define a subgroup of breast tumour cells lacking *ER*, *PR* and *HER2*, but expressing genes characteristic of normal basal/myoepithelial cells, such as *KRT5*, *KRT14* and *KRT17* (Cytokeratins 5, 14 and 17, respectively), and *EGFR* (Epidermal Growth Factor Receptor) (Foulkes et al., 2010*).* More than 90% of BLBCs are TNBCs (Cheang et al., 2015), while BLBC represents the most frequent subtype of TNBC (55–81%) (Ismail-Khan et Bui, 2010; Prat et al., 2013).

Furthermore, Lehmann et al. identified six subtypes of TNBC on the basis of gene expression analyses (Table 1.1): basal-like 1 and 2 (BL1 and BL2), immunomodulatory (IM), mesenchymal (M), mesenchymal stem-like (MSL), and luminal androgen receptor (LAR) (Lehmann et al., 2011). The BL1 subtype includes genes involved in the cell cycle and DNA damage repair, whereas the BL2 subtype is defined by higher expression of growth factor pathway genes. The IM subtype involves genes responsible for immune cell processes; the mesenchymal and MSL subtypes express genes responsible for cell motility and cellular differentiation (epithelial–mesenchymal transition), and the LAR subtype is characterized by androgen-receptor signalling. Many BL1 and BL2 tumours are associated with *BRCA* mutations and can be classified within the intrinsic basal-like subtype described by Perou et al. (Perou et al., 2000).

| TNBC subtype | Intrinsic subtype | Signaling pathways | Genetic signature | Relative overall survival | Potential therapies |
|---|---|---|---|---|---|
| BL1 | Basal-like | Cell cycle, proliferation, DNA damage pathways | *ATR, BRCA, MYC, NRAS, Ki-67* | ++ | PARP Platinum |
| BL2 | Basal-like | Growth factor Myoepithelial | *EGFR, MET, EPHA2, TP53* | + | PARP Platinum |
| M | Normal-like Claudin-low | EMT Growth factor | *Wnt, ALK, TGF-β* | + | Tyrosine kinase inhibitors PI3K/mTOR inhibitors |
| MSL | Basal-like Claudin-low | EMT Growth factor Proliferation (decreased) | *EGFR, PDGFR, ERK1/2, VEGFR2* | ++ | Tyrosine kinase inhibitors PI3K/mTOR inhibitors |
| IM | Basal-like | Immune signal | *JAK1/2, STAT1/4, IRF1/7/8, TNF* | +++ | Anti–PD-L1 inhibitors |
| LAR | Luminal HER2 | AR | *AR, FOXA1, KRT18, XBP1* | + | AR-targeted PI3K inhibitors CDK4/6 inhibitors |

**Table 1.1: Molecular Subclassification of Triple-Negative Breast Cancer.** TNBC subtype classification according to Lehmann et al., 2011 in relation to the altered gene expression profile, the intrinsic subtype, overall survival and potential matched therapies. +++: Best survival; ++: intermediate survival; +: worst survival. mTOR, mechanistic target of rapamycin; PARP, poly-ADP ribose polymerase; PD-L1, programed death-ligand 1; PI3K, phosphoinositide 3-kinase; TNBC, triple-negative breast cancer (Marotti et al., 2017; Sporikova et al., 2018).

At the genomic level, TNBC tends to be very complex, as demonstrated by the high rate of point mutations, gene amplification and deletion (Cancer Genome Atlas Network, 2012). Two large studies have focused their attention in the identification

of genetic markers that influence prognosis and prediction of the appropriate therapy (Shah et al., 2012, Koboldt et al., 2002). In the first study, exome-sequencing, RNA-sequencing, high-resolution SNP arrays and targeted deep resequencing were performed on 104 primary TNBC samples to identify patterns of somatic mutation (Shah et al., 2012): the most frequent copy number aberrations were identified for the *BRCA1/2*, *RB1* (retinoblastoma gene 1), *PTEN* (phosphatase and tensin homolog) and *EGFR* genes. *TP53* mutations were found to be the most common somatic aberration, followed by *PIK3CA*, *USH2A* (usher syndrome 2A) and *MYO3A* (myosin IIIA). However, only a minority of mutations (36%) were transcribed into mRNA (Shah et al., 2012).

In the second study, DNA copy number arrays, DNA methylation, exome sequencing, mRNA arrays, microRNA sequencing, and reverse-phase protein arrays were conduct on 463 samples from patients (Koboldt et al., 2012). In the basal-like tumors group (93 samples, 76 TNBCs), the most commonly mutated genes were *TP53*, *PIK3CA*, *MLL3* (lysine methyltransferase 2C), *AFF2* (AF4/FMR2 family member 2), *RB1* and *PTEN*. Copy number alterations were observed in several chromosomal regions or genes, as for example amplification or gain of *MYC*, *CCNE* (cyclin E1), 1q and 10p regions, loss of *PTEN*, *RB1*, *INPP4B* (inositol polyphosphate-4-phosphatase type II B) (30%), and the 8p and 5q regions (Koboldt et al., 2012). Some of these genes play a central role in the development of TNBC and are currently under investigation as promising therapeutic targets.

*BRCA1* and *BRCA2* genes are fundamental for the activation and transcriptional regulation of DNA damage (DNA double- strand break repair by homologous recombination (HRR) and the maintenance of DNA stability), control of the cell cycle, cellular proliferation and differentiation (Venkitaraman, 2002). Depending on the ethnic background and age of the investigated cohort, 10-15% of the cases carry mutations for *BRCA1* (Foulkes et al., 2003). Patients lacking *BRCA1/2* function are more sensitive to DNA-damaging agents like platinum derivatives and poly(ADP ribose) polymerase (PARP) inhibitors (Plummer, 2011): the Treating to New Targets (TNT) trial has shown an objective response rate to carboplatin compared to docetaxel in metastatic TNBC tumors with *BRCA* mutations (Kummar et al., 2012). It has been reported that among patients responding to platinum-based chemotherapy scores of allelic imbalance are higher (Watkins et al., 2015): *HORMAD1*, a cancer testis antigen involved in the promotion of non-conservative recombination in meiosis (Fukuda et al., 2010) has been identified as a novel driver of genomic instability in TNBC (Watkins et al., 2015). This protein suppresses RAD51-

dependent HR, generating micronuclei and structural chromosomal aberrations and driving in this way 53BP1-dependent non-homologous end-joining (NHEJ). In addition, the expression of *HORMAD1* correlates with a better response to HR defect-targeting agents both in TNBC cell lines and clinical trial: this might add additional information to *BRCA1/2* mutation testing for platinum treatment in TNBC patients (Watkins et al., 2015).

Furthermore, *MYC* has been shown to be frequently overexpressed in poorly differentiated tumours, driving uncontrolled proliferation or apoptosis with the cooperation of the Wnt signalling (You et al., 2002) and *BRCA1* (Wang et al. 1998) through a complex of BRCA1, Nmi, and MYC inhibiting *TERT* gene promoter activity in breast cancer (Li et al., 2002). In addition, it has been shown that these two genes cooperate to repress the transcription of *psoriasin*, a gene related to chemotherapeutic agent sensitivity (Li et al., 2002), demonstrating the fundamental role of *BRCA1* as a tumour suppressor. *MYC* overexpression and *BRCA1* loss seem highly correlated in a large portion in basal-like breast cancers (Grushko et al., 2004), suggesting this genetic combination as a possible mechanism of BLBC development.
Recent studies have also how the role of *MYC* in breast cancer tumourigenesis is dependent on the expression of PIM-1 (provirus integration site for Moloney murine leukemia virus 1) kinase 24. Horiuchi et al (Horiuchi et al., 2016) identified nine kinases required for the survival of MYC-activated non-immortalized human mammary epithelial cells: among those, PIM-1 had the greatest efficacy in maintaining survival. Subsequent analysis of distinct clinical cohorts highlighted that PIM-1 mRNA expression was significantly elevated in TNBC and its expression was associated with poor prognosis (Horiuchi et al., 2016). These results suggested how PIM-1 mediates survival, tumour growth and response to chemotherapy in cooperation with MYC in TNBC. Several research groups have generated small-molecule inhibitors targeting PIM kinases, with current preclinical and clinical trials (Blanco-Aparicio et al., 2013), demonstrating how PIM-1 may be a reliable biomarker for the diagnosis, treatment, and prognosis of TNBC, since its upregulation could be an important.molecular event during the development and progression of TNBC.

From a more clinical point of view, TNBC has an aggressive outcome, with short survival and relatively high mortality rate (Dent et al., 2007). The risk of recurrence seems to be higher during the first five years after diagnosis, with only few systemic

recurrences afterwards (Dawson et al., 2009). The high heterogeneity within patients, the lack of driver aberrations causing the pathology and the bad prognosis of the disease are the reasons why the development of new, biological and targeted treatments has been carried out. Nowadays some encouraging results have been obtained: in particular, pharmacological inhibition of transcription factors has raised increasing attention as a potential avenue for cancer treatment.

Different ways are currently available to target a transcription factor indirectly or directly. Transcription factors, as any other gene, are themselves controlled by transcription activators, repressors, epigenetic DNA or histone modifiers, so they can be inhibited or activated at the expression level (Lambert et al., 2018). In cancer treatment, a well-known example of this strategy is represented by *HOXA* cluster of transcription factors aberrantly expressed in leukaemia under the control of the MLL complex (Kawagoe et al., 1999). The aberrant MLL complex is formed by mutated or fused proteins such as HDAC, BRD4, Menin, WDR5 and PRMT1 (Steinhilber et al., 2018): many of them are targeted for cancer treatments, deregulating in this was also the control of HOXA5-10 transcription factor.

It is also possible to inhibit a transcription factor through degradation (Lambert et al., 2018): a very well established example consists on the usage of compounds like bortezomib (Velcade®) for the ubiquitin-proteasome or sumoylation processes for several tumours, including breast cancer (Liu et al., 2016; Desterro et al., 2000). However, other compounds have been developed to work at the transcription factor interaction level with other proteins. For example the partner could be another transcription factor (homo-dimers, as for STAT, BCL6; hetero-dimers, as for RUNX1/CBFβ, MYC/MAX from the basal transcription machinery; co-factor/co-activator/mediator or repressor (Nrf2/Keap1); chaperones), and the protein-protein interaction could be inhibited by small molecules like small compounds, peptidomimetics or stapled helix peptides (called PPIi, protein-protein interaction inhibitors). The tumour suppressor transcription factor p53 was the first one inhibited at the PPI level: in cancer, p53 is mutated and maintained in the cytoplasma through interaction with the Murine Double Minute 2 (mdm2) protein, also over-expressed in around 50% of all tumours (Rayburn et al., 2005). In this way, p53 is ubiquitinylated and subsequentally degraded (Zhou et al., 2001).

Targeting the transcription factor/DNA interaction is a conventional therapy still used since the first anti-cancer chemotherapies several decades ago (6-mercaptopurine was the first DNA alkylating drug for leukemia and lymphoma) and different molecules have been developed to target different binding modes of transcription

factor to the DNA helix: DNA alkylating drug (i.e platinated agents) or DNA intercalating drugs (i.e. aromatic chromophores) are just some examples of the currently available options for cancer treatment (Lambert et al., 2018).
Last possible mechanism of targeting a transcription factor is through its binding pocket, through a ligand-derived drug, as for steroid and hormonal receptors. In breast cancer, STAT5 for example is inhibited through direct interaction of an aptamer peptide (A431) mimicking its DNA binding-domain: in this way, the formation of the protein/DNA complex is inhibited and so the expression of the downstream target gene, such as cyclinD1 (Weber et al., 2013).

However, since they often operate in mutually redundant families, a broad-scale knowledge of their functions and interactors remains a necessity, and a lot of work still has to be done to increase the therapeutic strategy against this type of cancer.

## 1.4 Gene expression regulation: transcription

Nowadays, the understanding of the regulation of gene expression still remains one of the most important challenges in biology. At the transcription level, it occurs at specific transcription regulatory regions (promoters and enhancers), particular DNA sequence located upstream of gene of interest where proteins such as RNA polymerase and several transcription factors (TFs), as well as microRNAs (miRNAs) (Huang et al., 2011; Younger et Corey, 2011), are recruited to control the expression of their target genes (Latchman, 1997).

### 1.4.1 Gene transcription process

The transcription process occurs in three main steps: an initiation phase (Sainsbury et al., 2015), that involves the recognition and binding of the RNA polymerase II (Pol II) to the gene promoter sequence; an elongation phase (Jonkers et Lis, 2015), during which the RNA production occurs; and a termination one (Porrua et Libri, 2015), when the RNA and the Pol II are released from each other, and the Pol II is also released from the DNA. These steps are summarised in Fig. 1.2 (Weake et Workman, 2010). In eukaryotes, the first two phases are separated by an extra signal integration step that keeps the Pol II paused at the promoter region before the start of the active elongation (Sainsbury et al., 2015).

The initiation phase begins with the interaction between one or more transcription activator(s) and its recognition site within the promoter sequence of the target gene. These activators then sequentially recruit other co-activators and ATP-dependent nucleosome-remodelling complexes, facilitating the assembly and interaction with the DNA of the pre-initiation complex (PIC) (Roeder, 2005; Ptashne et Gann, 1997). Pol II and the general transcription factors (GTFs) TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH are members of this complex. At this point, while CDK7 (part of TFIIH complex) phosphorylates the serine 5 (S5) position of the Pol II carboxy-terminal domain (CTD), the DNA helicase XBP (another subunit of the TFIIH complex) remodels the PIC.11-15 bases of DNA at the transcription start site (TSS) are released to create a single-stranded DNA template to be introduced into the active site of Pol II (Saunders et al., 2006): this step, usually referred as promoter clearance, allows the Pol II to dissociate from some GFTs and proceed with the elongation stage.

However, after the transcription of 20-40 nucleotides into the gene, the Pol II stops at the promoter-proximal pause site: a second phosphorylation is required at the S2 of the Pol II CTD by CDK9 (a subunit of P-TEFb (Fuda et al, 2009)) in order to proceed. This modification is in fact fundamental for the formation of new binding sites for proteins involved in the mRNA processing and transcription like H3 lysine 36 (H3K36) by methylase SET2 (Egloff et Murphy, 2008).

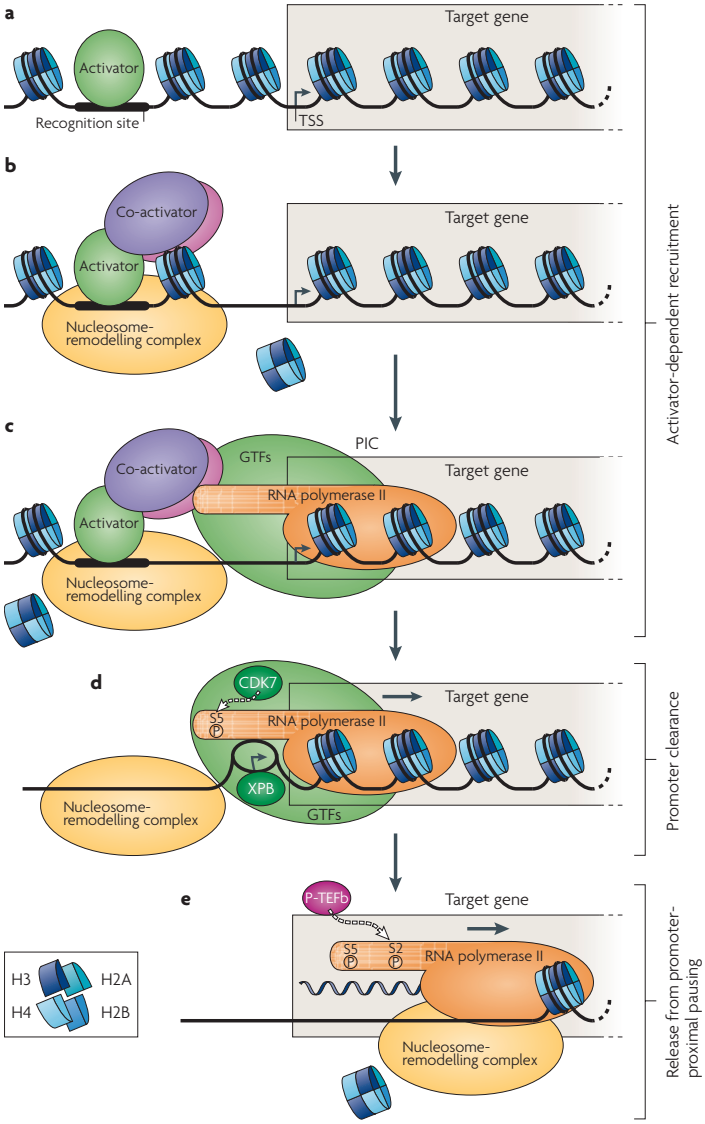



**Figure 1.2: Early steps in the transcription cycle.** a) Promoter selection is determined by the interaction of one or more transcription activator(s) with their recognition sites near target genes. b) Activation of gene expression is induced by the sequential recruitment of co-activator complexes (purple and pink) ATP-dependent nucleosome-remodeling complexes. c) Co-activators and nucleosome remodelers facilitate the recruitment of RNA polymerase II (Pol II) and the general transcription factors (GTFs, TFIIA, TFIIB, TFIID, TFIIE, TFIIF and TFIIH) to form the pre-initiation complex (PIC) on the core promoter. These first steps (a–c) constitute the activator-dependent recruitment. d) CDK7 phosphorylates the serine 5 (S5) position of the Pol II and the DNA helicase XPB remodels the PIC, and 11–15 bases of DNA at the transcription start site (TSS) are unwound to introduce a single-stranded DNA template into the active site of Pol II. This step is often referred to as promoter escape or clearance. e) Pol II transcribes 20–40 nucleotides into the gene and stops at the promoter-proximal pause site: the elongation requires a second phosphorylation at the S2 position of the Pol II by a subunit of human P-TEFb, that creates binding sites for proteins that are important for mRNA processing and transcription (Weake et Workman, 2010).

From a structural point of view, actively transcribed genes can be recognised by a specific nucleosome architecture of their promoters that allows the recruitment of the Pol II and TFs: they are characterised by AT-rich sequences that prevent the formation of a nucleosome, since they are less able to bend around a histone octane (Segal et al., 2006). In addition, chromatin remodellers like the RSC (Remodeling the Structure of Chromatin) complex maintain these regions without nucleosomes (called nucleosome-depleted regions, NDRs) by sliding them away (Hartley et Madhani, 2009).

This chromatin state is necessary, but not enough to start the gene transcription. A process of histones exchange also has to happen, performed by enzymes (adding post-translational modifications (PMTs) to histones), energy-dependent chromatin remodellers and histone chaperons. All these components are recruited through a reversible phosphorylation of an unstructured domain in the large subunit of Pol II (the carboxy-terminal domain, CTD) performed by several different kinases (Hsin et Manley, 2012).

In particular, the H2A–H2B dimer has to be replaced by the H2A.Z–H2B variant dimer into the nucleosomes flanking the NDR by the SWR complex, facilitating in this way the recruiting of chromatin remodellers and other TFs (Draker et al., 2012).

The passage of Pol II is accompanied by a rapid and continuous disruption of nucleosome structure (Jamai et al., 2007), followed by a reformation of the nucleosome array when the transcription finishes. It has been noticed that there is also a co-transcriptional H3K36 nucleosome methyl mark enrichment over coding regions that reduces the affinities of the chaperons responsible of the nucleosome dynamics over elongation (Venkatesh et al., 2012): in this way, pre-existing histones are less likely to be replaced by the newly synthesized ones.

### 1.4.2 Role of enhancers in gene transcription

Enhancers are DNA sequences distantly located from gene promoters, in a range from hundreds base pairs to megabases (Lettice et al., 2003). Initially they were identified as regions with the potential to increase basal transcription (Banerji et al., 1981), but it has been shown that they can also be transcribed by the Pol II to generate enhancer-associated RNAs (eRNAs) (Kim et al., 2010) (Fig. 1.3). These events are regulated by covalent modification of the histone tails of nucleosome,

such as methylation and acetylation. On the basis of these epigenetic modifications, it is possible to distinguish between different activations states (Ernst et Kellis, 2010): inactive enhancers, when they are characterized by compact chromatin; primed, when sequence-specific TFs bind specific sequences creating a DNase I-hypersensitive, nucleosome-free region of open chromatin; and poised enhancers, that are primed enhancers with repressive epigenetic chromatin marks.

The activation of the target gene promoter implies the recruitment of the different components of the transcriptional machinery to assembly the pre-initiation complex, start the transcription and lead to elongation. Through DNA looping, enhancers get in proximity of the promoter of the gene of interest, and they are thought to enhance the transcription process increasing the concentration of the factors that carry them out (Plank et Dean, 2014). Among these factors there are co-activator complexes, like for example the Mediator (Kagey et al., 2010) and SAGA complexes (they contain histone acetylase (HAT) Gcn5) (Grant et al, 1997), scaffold proteins like cohesion that allow a stable promoter-enhancer interaction (Kagey et al., 2010), and other factors involved in the initiation of the elongation like Brd4 (Bromodomain Containing 4) (Liu et al., 2013).

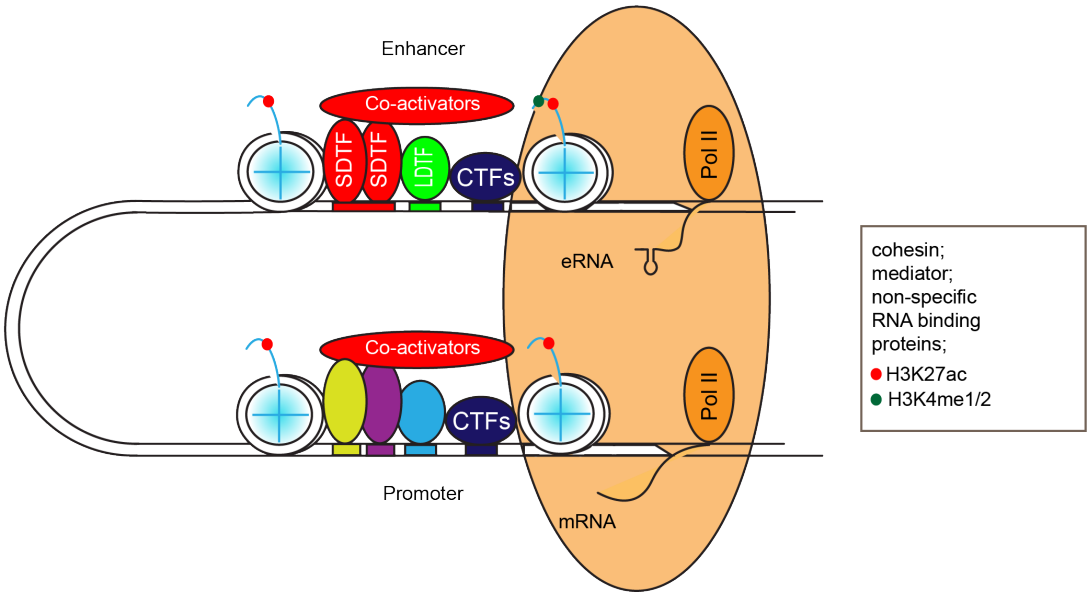



**Figure 1.3: Enhancer activation and function.** Interactions between enhancers and promoters involve structural connections (orange oval) that include cohesin and the mediator complex to promote pre-initiation complex formation and initiate transcription. Enhancer RNAs (eRNAs) could promote transcription by facilitating chromatin looping, possibly by mediating interactions with cohesion, or with protein complexes required for transcriptional elongation (for example mediator complex). LDTFs, lineage-determining transcription factors; CTFs, collaborating transcription factors; SDTFs, signal-dependent transcription factors (modified from Heinz et al., 2015).

### 1.4.3 Oncogenic transcription regulators in breast cancer

Aberrantly expressed transcription regulators that lead to tumourigenesis are defined as oncogenic elements. For breast cancer, a great part of this category is formed by transcription factors. They are usually divided in three main subgroups (Gibbs, 2000; Brivanlou et Darnell, 2002): steroid receptors, such as oestrogen receptors (Tilley et al., 2001); resident nuclear factors, such as JUNB, JUND, c-JUN (Van Dam et Castellazzi, 2001) located in the nucleus and activated by serine/threonine residue phosphorylation and co-activators, and latent cytoplasmic factors, all those factors translocating into the nucleus after activation at the cell membrane level in a receptor-ligand manner, such as STAT family (Signal Transducers and Activators of Transcription), associated with cell-cycle progression, cell survival, transformation, and angiogenesis (Calo et al., 2003).

The overexpression and/or over-activity of these oncogenic TFs have a fundamental role in cell proliferation, tumour survival and invasive behaviour. Some of these factors are crucial for breast cancer carcinogenesis, in particular for TNBC. For example NF-κB (Nuclear factor of κB) has been demonstrated to drive breast cancer development and progression (Demicco et al., 2005; Srivastava et al., 2003) and is associated with particularly aggressive ER negative and *HER2*$^{+}$ subtype known as inflammatory breast cancer (IBC) (Van Laere et al., 2006). *TP53* gene is usually mutated in 20% of the cases (Pharaoh et al., 1999), with a different prognosis and rate between different subtypes of breast cancer, but an increased rate of mutations in cancers carries the germ-line *BRCA1* and *BRCA2* mutations (Smith et al., 1999). In addition, *MYC* amplification has been observed in the more aggressive phenotype of DCIS (Aulmann et al., 2002) or in the invasive component (Aulmann et al., 2006). These are just some of the well-know altered pathways that are necessary for breast cancer growth and progression: however, many others still require further investigations.

## 1.5 Novel cancer therapies against gene transcription

Recent studies have pointed out how cancers keep an identifiable pattern of gene expression (Wang et al., 2015; Hnisz et al., 2013; Hnisz et al., 2015; Lovén et al., 2013): if a uniform gene expression is required, the tumour has to have a constant, active gene transcription. This necessity might be exploited to develop new approaches for cancer therapy: these tumours might be extremely sensitive to drugs

to inhibit transcription (Delmore et al., 2011; Chipumuro et al., 2014; Dawson et al., 2011; Chapuy et al., 2013).
Although this strategy could be difficult because of the redundancy of some pathways in non-malignant cells and tissues, recent studies have shown that the transcription of some genes is more sensitive to inhibition (Delmore et al., 2011; Kwiatkowski et al., 2014). To date, transcription factor-directed anticancer drug development has focused on membrane or cytosolic targeting of molecules acting as ligand receptors. Two successful examples can be cited as breast cancer therapies: tamoxifen, for ER-dependent breast cancers, and trastuzumab, for *HER2*$^{+}$ ones.

### 1.5.1 Breast cancer therapies: ER and HER2 examples

Oestrogen receptor is a transcription factor that regulates the expression of genes involved in timely controlled cell division during mammary gland development and during post-pubertal physiological functions, such as pregnancy (Carroll, 2016). One of the first targeted agents in the treatment of this type of tumour is the selective oestrogen receptor modulator (SERM) tamoxifen (Fisher et al., 2005): it mimics oestrogen and binds to ER, but it alters the structure and function of the transcription factor so that it is no longer able to regulate the expression of target genes (Shiau et al., 1998).

Due to gene amplification, *HER2*$^{+}$ breast cancer is characterized by the expression of HER2, a transmembrane receptor with tyrosine kinase activity that belongs to a family of four receptors (EGFR/HER1, HER2, HER3, HER4). Structural studies have shown that HER2 is always in an active confirmation that allows dimerization with the ligand-activated HER receptors (Graus-Porta et al., 1997). It is involved in regulating cell growth, survival and differentiation through activation of the PI3K/Akt and the Ras/Raf/MAPK pathways (Yarden et al., 2001).

Currently, the approved treatment for these patients is the monoclonal antibody trastuzumab, which recognizes the extracellular domain (ECD) of HER2: the binding limits the receptor's ability to activate its intrinsic tyrosine kinase, which in turn, limits the activation of many other different signaling pathways promoting cancer growth.
Although its antitumor action is not completely understood, different mechanisms have been proposed to explain the effect: trastuzumab blocks the binding of ERBB2 to the receptor, preventing in this way the activation of signalling cascade and the

regulation of the transcription of targeted genes (Lane et al., 2001). This downstream effect could also be caused by the internalization and degradation of the ERBB2 receptor after trastuzumab binding, which would then downregulate the PI3K pathway signaling and downstream mediators of cell cycle progression such as cyclin D1 (Yakes et al., 2002; Izumi et al., 2002; Valabrega et al., 2007).
Trastuzumab not only inhibits HER2 signaling pathways but also triggers immune-mediated responses against HER2-overexpressing cells through antibody-dependent cell-mediated cytotoxicity (ADCC): once trastuzumab binds the receptor on the surface of the cancer cell, activated natural killer cells bind the antibody and initiate the lysis of the cancer cell (Cooley et al., 1999). From clinical trials, trastuzumab seems to be generally well tolerated when administered after chemotherapy, although potential cardiotoxicity and resistance are major concerns.

For *HER2*$^{+}$ breast cancer resistance, it has been shown that cancer cells decrease or increase the expression of *HER2* itself (Köninki et al., 2010), *HER1* or *HER3* (Vazquez-Martin et al., 2007) to compensate, or increase the expression of some ligands like TGF-α (a ligand for EGFR/HER1) (Nahta et al., 2009). It might also arise through constitutive activation of the PI3K/Akt pathway, due to mutations in the *PIK3CA* gene and/or loss of *PTEN* (Arteaga et al., 2011). Preclinical studies showed that another treatment could inhibit cancer growth of those cells resistant to trastuzumab: lapatinib. This antibody reversibly inhibits the intracellular tyrosine kinase activity of both HER2 and HER1, suppressing downstream pathways such as MAPK/Erk1/2 and PI3K/Akt (Konecny et al., 2006). At the moment it is used in combination with anti-HER2 antibodies to enhance the apoptotic effect (O'Donovan et al., 2010).

Regarding the ER-dependent breast cancer, the tumour can become resistant downregulating ER expression. This usually happens in approximately 10-20% of the cases (Harrell et al., 2006), and ER function is substituted by additional nuclear receptors. For example, Androgen Receptor (AR) was found to be upregulated in 80-90% of *ESR1*$^{+}$ breast cancer (Peters et al., 2009), and it could initiate cell division. However, in most of the cases ER expression is retained (Harrell et al., 2006) and this transcription factor can still be functioning even in the presence of an anti-endocrine agent. Five possible mechanisms have being highlighted to explain this resistance: changes in drug metabolism and cellular secretion; upregulation of pathways that can promote ER transcriptional activity or of pathways that will make the target proteins more active; changes in the fidelity of the key proteins involved in

the ER complex; changes in the expression levels of associated proteins that are required for ER transcriptional activity (co-factors) and changes at the genetic level. In particular this last mechanism has acquired a lot of interest in the recent years.

According to the TCGA data from 962 breast cancer samples, *ESR1* mutations were present in only 0.5% of primary breast tumor cases (TCGA, 2012). The next-generation sequencing (NGS) of DNA revealed a higher prevalence (11–55%) in metastatic $ER^+$ breast cancers with prior AI exposure (Jeselsohn et al., 2015; Merenbakh-Lamin et al., 2013; Toy et al., 2013). Several works and clinical trials have showed how *ESR1* mutations are rarely detected in treatment-naïve primary tumours, while they rise up to 11-39% of the cases (according to different patient profiles) in AI-refractory tumours (Jeselsohn et al., 2015; Niu et al., 2015; Chandarlapaty et al., 2016). According to these results, it is possible to believe that mutations arise through clonal selection of low abundant resistant clones or are acquired during the disease progression under the treatment selective pressure.
The most common missense mutations are clastered in codons 537 and 538, while the most prevalent *ESR1* point mutations are Y537S and D538G (several others have been identified but at significantly lower frequencies) (Reinert et al., 2017). Mutated ER recruits its coactivators without the hormone stimulation, and its affinities for oestrogen agonist or antagonist (estradiol or tamoxifen) are decreased. In addition, the mutations alter the conformational dynamics of the ER binding loop, conferring in this way an anti-oestrogen resistance (Fanning et al., 2016*).*

Another resistance mechanism that has been reported is the acquisition of *ESR1* fusion genes. However, a detailed clinical study and prevalence is still required (Li et al., 2013). Recently Hartmaier et al. reported the identification of recurrent hyperactive *ESR1* fusion genes in breast cancers resistant to endocrine therapy (Hartmaier et al., 2018). Through mate-pair DNA sequencing and/or RNA sequencing of matching primary-metastasis-normal samples from 6 patients, they were able to identify ESR1-DAB2, ESR1-GYG1, and ESR1-SOX9 in-frame fusion transcript (found only in the lymph node metastasis, not in the primary tumour), all with ligand-independent activity and hyperactive (Hartmaier et al., 2018). These observations suggest a potential role for the distinct 3′ gene partners in determining resultant ER activity.
A similar mechanism has been reported for patients carrying *BRCA1*/*2* germline mutation resistant to PARP inhibitors/platinum salts: in these cases, different *BRCA1*/*2* intragenic deletions or reversion mutations (Edwards et al., 2008; Swisher

et al., 2008) are able to restore the reading frame, causing in this way resistance. Multiple activating *ESR1* mutations have been detected in the ctDNA samples of patients carrying activating *ESR1* fusion gene (Hartmaier et al., 2018), suggesting a novel polyclonal resistance mechanism (Christie et al., 2017).
However, these findings still have to be fully elucidated because of that limited clinical history available: it is in fact possible that some of the *ESR1* fusion genes are not transcribed and/or translated, or may have limited impact on the resistance to endocrine therapies.

To block ER function, new drugs such as Fulvestrant (Faslodex) and Aromatase Inhibitors (AIs) have been developed: the first one is a steroidal anti-oestrogen that binds ER and induces its degradation, while AIs are starving the cancer of the oestrogen ligand, acting at a metabolic level. Both of the treatments have shown effectivity in tamoxifen-resistant context (Howell et al., 2005).

Recent studies have also shown how ER requires the accumulation of many other proteins to perform different functions like, for example FOXA1 and GATA3 (Carroll et al., 2005; Eeckhoute et al., 2007): when any of these factors is specifically inhibited in breast cancer cells, ER-DNA interactions are perturbed.

FOXA1 can occupy compacted DNA without the requirement of other proteins (Cirillo et al., 1999) and in this way facilitate interactions between additional factors (such as ER) and the DNA (Carroll et al., 2005). GATA3 has been shown to be required for the morphogenesis of normal mammary glands (Asselin-Labat et al., 2006): it seems to be involved in promoting cellular differentiation and inhibiting proliferation, but becomes an essential component within the ER complex during tumour formation. In fact, when inhibited in cancer cells, ER interacts mainly with DNA binding sites demarcated by FOXA1 (Theodorou et al., 2012). This would suggest an important role of GATA3 in regulating ER-FOXA1 interactions (Carroll, 2016).

### 1.5.2 Treatment option for TNBC

Despite the success in the treatment of various subtypes of breast cancer, for TNBC there is still a lack of unique, effective therapeutic approach. Several ones are currently under investigation, and they could eventually improve the outcome for

these patients. In Fig. 1.4 some of these approaches have been summarised (Park et al., 2018).

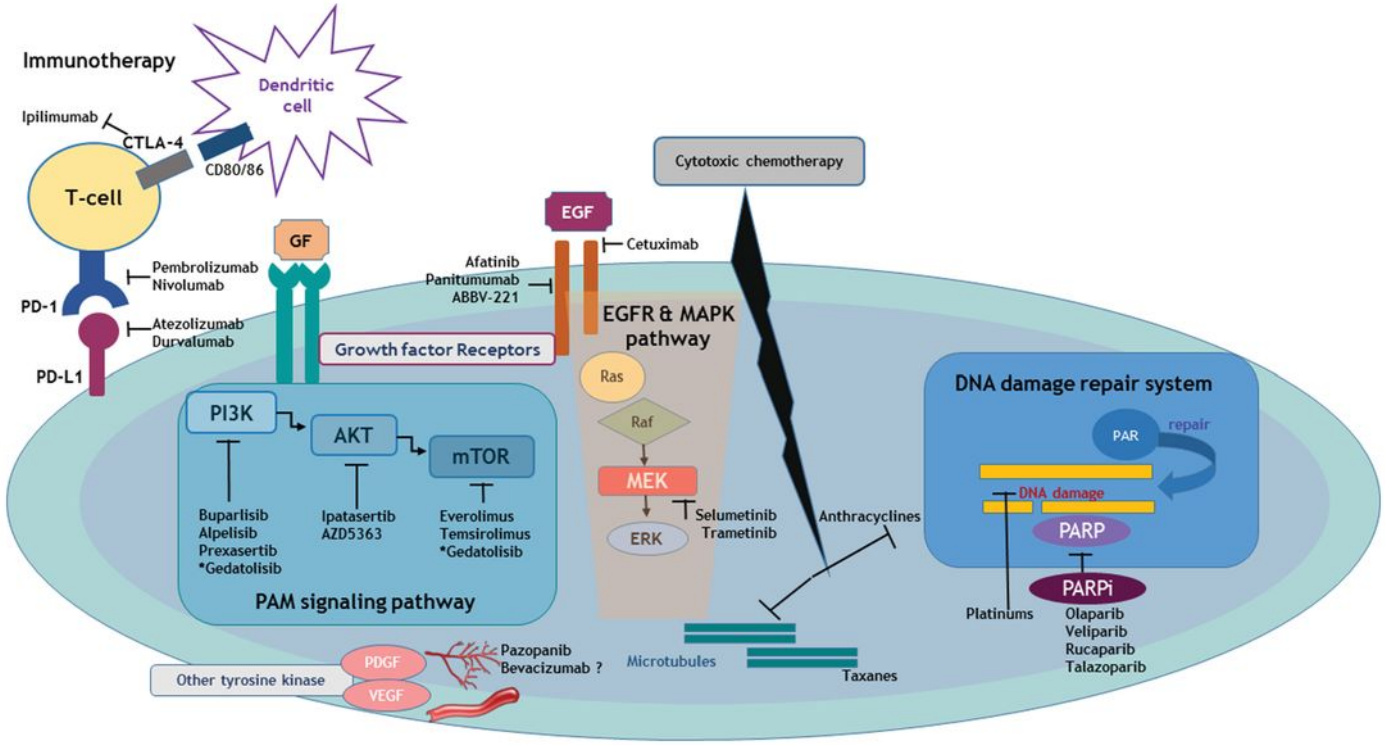



**Figure 1.4: Potential therapeutic targets in TNBC.** The blockade of the immune-checkpoint targeting PD-1, PD-L1 and CTLA-4 can boost the adaptive immune reaction. Blocking the PAM signalling pathways, together with various growth factors like EGF and MAPK signalling, affect cell cycle regulation. Platinum-based agents and PARPi affect the DNA damage repair pathway. Multikinase inhibitors involving angiogenesis or developmental process could also be potential therapeutic entities. All these investigational key targets are consistently interacting with cytotoxic effect of conventional chemotherapy. CTLA-4, cytotoxic T-lymphocyte-associated protein 4; EGF, epidermal growth factor; EGFR, EGF receptor; ERK, extracellular signal-related kinase; MAPK, mitogen-activated protein kinase; MEK, MAPK kinase; PAM, PI3K-Akt-mTOR; PARP, poly(ADP-ribose) polymerase; PARPi, PARP inhibitors; PD-1, programmed cell death protein 1; PD-L1, programmed cell death ligand 1 (Park et al., 2018)

For example, for those patients carrying the mutations affecting the PI3K–AKT signalling pathway, a promising therapeutic approach has emerged from the randomized, double-blind, phase II PAKT trial (Schmid et al., 2020): the combination of AKT inhibition and chemotherapy. In this trial, 140 women with metastatic TNBC were randomly assigned to receive paclitaxel together with a placebo or capivasertib (pan-AKT inhibitor) at the intermitted dosage of 4 days on, 3 days off. An increased of the median progression-free survival (mPFS) was observed (5.9 months compared to 4.2 months with placebo), and so the median overall survival (19.1 months versus 12.6 months). However, the effect of the treatment seems to be restricted mainly to patients carrying alteration in *PIK3CA*, *AKT1* or *PTEN*. These recent results corroborate those observed with ipatasertib, another AKT inhibitor used in the LOTUS trial (Kim et al., 2017), confirming the validity of this combination.

Another possible therapeutic approach for patients carrying *BRCA*-mutation consists of platinum agents, like carboplatin and cisplatin, leading to DNA crosslink strand breaks, which may be particularly important in these tumours lacking the homologous repair mechanism because of the mutation. A similar strategy is based on PARP (poly ADP-ribose polymerase) inhibitors, like olaparib or iniparib, that affect the activity of this polymerase, a critical enzyme for the base excision repair pathway and a key for the single-strand DNA breaks repair.

PARPi were thought to contribute to a synthetic lethality mechanism by which inhibition of two DNA repair pathways contributes to cell kill in HRR-deficient cancerous cells over normal cells (Narod, 2010). It is now known that PARPi exert their efficacy interfering with the identification of DNA damage and multiple types of repair: their effects are mainly focused during S-phase, when DNA is exposed for replication, and HRR is preferred over nonhomologous end-joining (NHEJ) for repair of DNA double-strand breaks (Schreiber et al., 2006; Murai et al., 2012).

In the last decades several clinical trials have been designed to use PARPi for breast cancer patients, with different clinical settings, such as neoadjuvant, adjuvant and metastatic (Vinayak et al., 2010). The mechanism of action (reversible or irreversible inhibition), dosing intervals (continuous or intermittent), toxicity, combination with chemotherapeutic agents are just some of the aspects of this therapy under investigations within the on going studies.

The majority of the studies have been focused on BSI-201 (known as Iniparib (Sanofi-Aventis, France) and Olaparib (AstraZeneca, UK), even though some others have been developed and their efficacy evaluated, such as ABT-888 (Veliparib (Abbott), AG014699 (Pfizer), CEP-8983 (Cephalon), and MK-4827 (Merck).

Iniparib is an intravenous (IV) irreversible PARPi (Rouleau et al., 2010), dosed intermittently, primarily used in combination with gemcitabine and carboplatin. The first results for treatment of TNBC were in 2009 (O'Shaughnessy et al., 2009) where in a randomized phase II trial treated women had improvements in the clinical benefit rate (21% vs 62%; *P*=0.0002), overall response rate (16% vs 48%; *P*=0.002), median progression-free survival (3.3 vs 6.9 months; hazard ratio [HR]=0.342; *P*<0.0001), and median overall survival. Olaparib instead, the first FDA-approved orally active PARPi is dosed continuously, mainly used in patients with *BRCA* mutation. The first phase II results came in 2009, when Tutt and colleagues showed that the response to a PARPi is dependent on *BRCA1* or *BRCA2* germline mutations rather than the tumour's phenotype (hormone receptor-positive or negative) (Tutt, et al., 2009).

Current PARPi clinical trials registered with the National Institutes of Health's United States National Library of Medicine in ClinicalTrials.gov include patients with breast cancer and are headed by monotherapy trials followed by combination trials. They are organized by type of combination (e.g. PARPi + chemotherapy) and clinical trial phase from I to III within each category, and include trial characteristics, patient population (with gBRCA1/2 bolded if a requirement for a particular trial), trial interventions and outcome measures. Combinations of iniparib with gemcitabine and carboplatin have been shown to delay TNBC progression and improve survival in clinical phase II studies (Liu, et al., 2012), but not in phase III trials (O'Shaughnessy et al., 2014). Olaparib was used in 2018 to treat germline *BRCA*-mutated, metastatic, and HER2-negative breast cancer (Le & Gelmon, 2018), but nowadays 15 clinical trials of olaparib monotherapy and combination therapy for TNBC are underway: olaparib in combination with programmed cell death-ligand 1 (PD-L1) inhibitors such as durvalumab and atezolizumab in (Roviello et al., 2016; Solinas et al., 2017); olaparib in combination with cediranib (AZD2171), inhibitor of VEGFR-2 tyrosine kinase (Wedge et al., 2005); olaparib in combination with PI3K inhibitors such as buparlisib (BKM120) and alpelisib (BYL719) (Teo et al., 2017) and olaparib in combination with oral mTORC1/2 inhibitor (vistusertib/AZD2014) or AKT inhibitor (capivasertib/AZD5363) (Ocana & Pandiella, 2017). Combination treatments have a potential effect on growth inhibition of rapidly proliferating TNBC cells affecting blood supply and blocking the molecules required for cell growth.

TNBC also exhibits higher mean quantities of tumour infiltrating lymphocytes (TILs) compared to other breast cancer subtypes both intratumourally and in adjacent

stromal tissues. Their presence has been associated with a favourable prognostic value, with complete response after neoadjuvant chemotherapy in TNBC patients and with a predictive marker for immunotherapy response (Borcherding et al., 2018). However, different TNBC subtypes show different characteristics: in particular the IM and basal-like subtypes are the ones with the higher infiltration of immune cells, antigen presenting cells and active immune pathways (Vinayak et al., 2017). In addition, the high frequency of *BRCA1* and *BRCA2* mutations is considered as another predictive marker for immunotherapy response (Borcherding et al., 2018). Among different strategies, immune-checkpoint inhibitors have shown promising results in both advanced and early-stage disease. However, response rates are very modest as single agent in advanced TNBCs and dependent on cancer type (Vikas et al., 2018). The majority of the current clinical trials have used monoclonal antibodies targeting the programmed cell death protein 1 pathway (PD-1/PD-L1) and the cytotoxic T lymphocyte-associated antigen 4 (CTLA-4), or combination strategies: these proteins are negative regulators of immune activation, and their presence in the tumour microenvironment prevents the activation of an efficient antitumor immune response (Pardoll, 2012).

PD-L1 is expressed in 20%–50% of all breast cancer subtypes and it has been associated with higher histologic grades, larger tumours, and absence of hormone receptors (Ghebeh et al., 2006). Its expression in breast cancer, and in particular in basal-like TNBC, has been associated with longer overall and disease-free survival (Sun et al., 2016). Recently, the Food and Drug Administration (FDA) approved atezolizumab (anti-PD-L1) in combination with nab-paclitaxel for PD-L1-positive advanced TNBC. Significant response has been seen in early-phase trials with anti-PD1 or anti-PD-L1, but response rates are up to 10% in unselected TNBC patients and improves only slightly to 20%–30% when patients are selected based on IHC-based PD-L1$^+$ tumours (Adams et al., 2017; Nanda et al., 2016).

CTLA-4 is a T-cell inhibitory receptor that is expressed on activated CD8$^+$ T cells and CD4$^+$ regulatory T cells expressing CD25 and Foxp3. It attenuates the T-cell immune response binding to receptors on DCs: its blockade could remove these inhibitory signals, therefore enhancing the anti-tumoural effect (Arce Vargas et al., 2018). It has been shown that TNBC is characterised by the highest incidence of TILs (20%; range, 4–37%) and the highest levels of Foxp3$^+$ Tregs cells (70%; range, 65–76%) of all breast cancer subtypes (Stanton et al., 2016): these Foxp3$^+$ Tregs may be the therapeutic targets of CTLA-4 blockade approach in TNBC treatment.

Ipilimumab has already been approved by FDA for melanoma patients and is currently being investigated in breast cancer, together with tremelimumab (Hodi et al., 2010).

Other targets that are being investigated for potential checkpoint inhibition include the BTLA, VISTA, TIM3, LAG3, and CD47 proteins but are very early in development (Havel et al., 2019). Overall, immunotherapy is considered one of the most promising therapeutic approaches with an efficient and durable effect for patients with TNBC.

Unfortunately the treatments currently available have shown limited benefits. This indicates that TNBC cannot be treated as a uniform disease: its biology likely involves multiple redundancies and pathway cross-talks, which imply that if one pathway is selectively inhibited by a therapeutic strategy, a compensatory one would be activated. Therefore, not a single targeted therapy has been approved for TNBC treatment: combining two or more agents and/or finding new molecular targets may be required for a more rational and optimal approach.

Recently Wang and colleagues have demonstrated that TNBC tumours are dependent on CDK7 (cyclin-dependent kinase 7) (Wang et al., 2015). CDK7 is not only a kinase, but also a subunit of a multi-protein basal transcription factor TFIIH: it is involved in the control of the cell cycle through phosphorylation of other CDKs (Malumbres, 2014), and regulates the initiation of the transcription by phosphorylating the RPB1 subunit of the RNA polymerase II (Malumbres, 2014). Wang and colleagues identified a selective inhibitor of CDK7 called THZ1 that also inhibits CDK7 dependent transcription of 450 genes supporting the tumourigenicity of TNBC (Fig. 1.5). These genes, encoding signalling molecules and transcription factors like TGF-β, STAT, WNT, are strongly dependent on continuous, active transcription, which is allowed by the presence of large clustered enhancer regions (super-enhancers), exceptionally regulated by transcription factors and co-factors (Hnisz et al., 2013), like for example CDK7. Targeting the transcription of this region seems to be an effective way to simultaneously suppress the expression of multiple genes fundamental for TNBC. This study thus highlights the feasibility of disrupting transcription as a therapeutic approach for TNBC treatment.

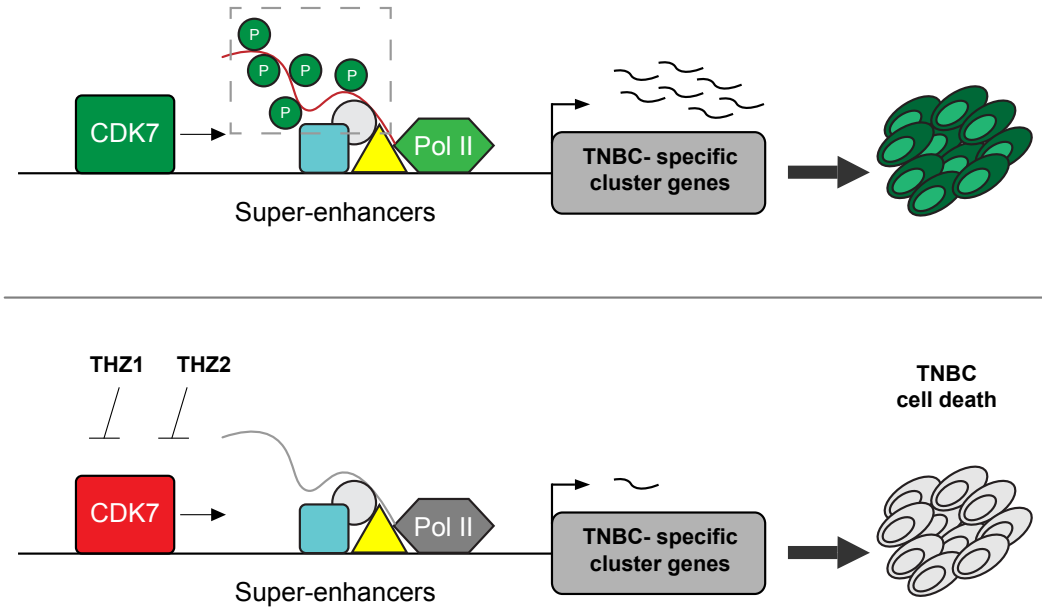



**Figure 1.5: CDK7-dependent transcription addiction in TNBC.** CDK7 regulates the continuous transcription of a cluster of TNBC genes phosphorylating the RNA polymerase II recruited at large clustered enhancer regions (super-enhancers). Selective inhibition of CDK1 through THZ1 causes a disrupted expression of this key cluster of genes, with an effect on cancer survival (modified from Wang et al., 2015).

A comprehensive identification of the transcriptional control of TNBC would aid in the development of more targeted therapies. This can be achieved by the use of novel proteomic and genomic tagging technologies, like CRISPR/Cas9, described in the following sessions.

## 1.6 Novel techniques to investigate TFs

### 1.6.1 Discovery proteomics

Mass-spectrometry (MS) based approaches have been extremely useful to identify proteins present in different cellular compartments and/or expressed at different cell cycle stages. Thanks to them, it is possible to analyse a large amount of endogenous proteins avoiding time, cost and technical limitations of other techniques in a high throughput, sensitive, dynamic and fast way.

In particular, discovery (or shotgun) proteomics is based on a bottom-up workflow (Yates et al, 2009), which consists in the identification of proteins through peptides generated by enzymatic cleavage. On the basis of their number, it is also possible to get quantitative information about the protein of interest: more specifically, relative abundances of peptides are estimated by counting the number of MS/MS spectra assigned to the same peptide/protein (Liu et al., 2004). Alternatively, the relative abundance is obtained as an integrated area of a peptide peak in extracted ion chromatogram (XIC) using dedicated software tools for data processing (Nahnsen et al., 2013). Isotopic-labelling strategies as the isotope-coded affinity tag (ICAT) (Gygi et al, 1999) and stable isotope labelling by amino acids in cell culture (SILAC) (Ong, 2002) have been introduced for quantitative comparison of biological samples. Later on, chemical labelling by isobaric tags for relative and absolute quantification (iTRAQ) (Ross et al., 2004) and dimethyl labelling protocols have been developed to improve quantification accuracy (Boersema et al., 2009). However, the variance of the relative signal intensity is sequence-dependent, making the technique not inherently quantitative.

In addition, MS can be applied for a global analysis of protein complexes: three main strategies are currently used for this purpose, summarised in Fig. 1.6.

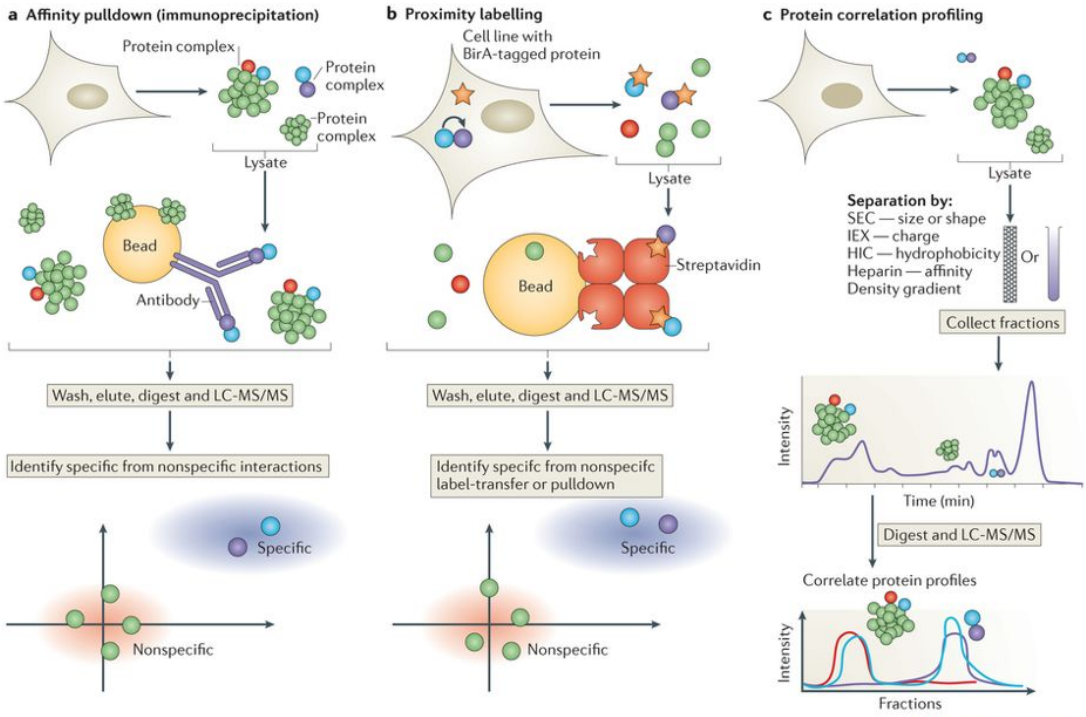



**Figure 1.6: Three main approaches for unbiased analysis of protein–protein interactions.** a) Affinity pulldown and isolation approach: it uses a specific antibody against an endogenous target protein or a tagged version of that protein to isolate it together with its interacting partners. These complexes are then eluted, digested, filtered through liquid chromatography and analysed by tandem mass spectrometry (LC-MS/MS). Statistical approaches are applied to identify specific from non-specific binding partners. b) Proximity labelling approach: cells are modified in order to ectopically express the target protein fused to a biotin ligase or a peroxidase enzyme. These enzymes can covalently transfer biotin labels to proteins that are in close proximity, so potential interactors. These biotinylated proteins can be isolated using streptavidin-conjugated beads after cell lisation. Similarly to the procedure, the isolated proteins undergo a digestion step, followed by LC-MS/MS analysis and statistical tests. c) Protein correlation profiling: coupled with techniques like chromatography and density gradient centrifugation in order to separate protein complexes according to size, density, charge or hydrophobicity, assuming that interacting proteins will co-elute. This step could involve a single type of separation or multiple ones, sequentially or in parallel. Successively, each separated fraction is digested and analysed by LC-MS/MS, generating an elution profile for each detected protein. Clustering algorithms can then identify co-eluting proteins and infer the protein complexes in the lysate. HIC, hydrophobic interaction chromatography; IEX, ion-exchange chromatography; SEC, size-exclusion chromatography (Larance et  Lamond, 2015).

The first and most widely used approach is based on the affinity pulldown strategy (immunoprecipitation) (Fig. 1.6, a): the target protein, together with its interactors, is isolated through immunoprecipitation, using a specific antibody against it or against a tag ectopically expressed (for example GFP, short peptides (FLAG or hemagglutinin) (Trinkle-Mulcahy et al., 2008). This approach has been used to examine protein complexes in a proteome-wide scale (Hubner et al., 2010; Jäger et al., 2011). It is a very sensitive method, in particular for low abundant complexes.

The second approach is an *in vivo* proximity labelling (Fig. 1.6, b): a particular cell is engineered to ectopically expressed the target protein fused to a biotin-ligase derived from bacteria (the BioID method) (Roux et al., 2012), or to a peroxidase enzyme capable of activating biotin–phenol (the APEX method) (Rhee et al., 2013). Once activated, the biotin is rapidly and covalently conjugated to nearby Lys (in the case of BioID) or to Tyr (in the case of APEX) residues. This enhances the enrichment of potential interactors thanks to a streptavidin pulldown. In this particular approach, it is possible to maximise the purity of the sample with stringent buffers and extensive washes thanks to the high-affinity interaction between biotin and streptavidin.

The third approach is based on variations of Protein Correlation Profile (Fig. 1.6, c) (Kirkwood et al., 2013): chromatography or density gradient centrifugation techniques are used to separate native protein complexes according to size, density, shape, charge and/or hydrophobicity. The cell extracts are isolated and fractionated under particular conditions to preserve protein–protein interactions. After elution, the different fractions are collected, individually processed and analysed by LC-MS/MS. Protein elution (gradient) profiles are then generated for each protein and compared with others by computational clustering to identify potential interacting proteins based on similarities. This approach can simultaneously analyse hundreds of protein complexes: it has in fact been used to study the interactome of some cell lines in combination with other protein properties like isoforms or PTMs (post transcriptional modifications) (Kirkwood et al., 2013). However, even if it has been shown that multiple chromatographic steps increase the resolution of the analysis (Havugimana et al., 2012), this method can be used only on soluble complexes.

All these approaches can confirm the presence of a protein within a complex, but not the direct or indirect interaction with any of its members. This information can be

generated through protein crosslinking step coupled with one of these methods (Leitner et al., 2014; Mohammed et al., 2013), thanks to which it has been possible to map not only direct protein-protein bindings (Fischer et al., 2013; Weisbrod et al., 2013), but also protein-RNA interactions (Kramer et al., 2014). However, this technique comes with several limitations: form a bioinformatics point of view, it is challenging to identify crosslinked peptides using MS fragmentation methods, and consequentially to estimate accurately the false-discovery rate (FDR) of each co-fragmented peptide sequence, which could come from any protein in the original, crosslinked mixture.

Nowadays, mass spectrometers offer high resolution (>400,000 mass/Δmass), high mass accuracy (<1 ppm), high sensitivity (<attomol), and high speed (12–20 Hz) (Sidoli et al., 2016). However, the results are complex to interpret, dependent on the method of acquisition and on the platform used for data analysis. Proteomics experiments are notoriously prone to produce a high proportion of false positives (Christoforou et al., 2014): in fact the signal of the protein of interest could be mixed with background noise, as other metabolites might have isobaric masses, which means the same atomic composition but different structure. In addition, reproducibility is extremely difficult to achieve: in organelle proteome studies, for example, it is almost impossible to obtain identical gradient fractions among experiments and replicates, and any experimental perturbation can affect the size or density of the organelle. The stochastic element of shotgun MS also has to be considered: the set of proteins identified in each experiment would never be identical. For these reasons, technical optimization, multiple replicates and comparative experiments are the only way to improves resolution and achieve low FDR levels (Itzhak et al., 2016; Itzhak et al., 2017).

Even if MS approaches generate a static picture of the cellular proteins map that on the other hand naturally undergoes a continuous, dynamic modification process, their informative relevance is undeniable.

### 1.6.2 Characterization of TF binding sites

Transcription factors play a fundamental role in the regulation of gene expression through direct interaction with its regulatory regions or indirect ones, together with other regulatory proteins.

Each TF can recognize several similar DNA sequences with different binding affinities (Siggers et Gordân, 2014). For this reason, TF binding specificities (the preferential binding of specific sequences) are represented as binding site motifs, a summary of the preferentially bound sequences, used to predict TF binding sites.

Currently one of the most widely used methods to study the TF-DNA binding preference is chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-Seq) (Furey, 2012). Briefly, the genomic region bound by a TF is isolated by immunoprecipitation and identified though high-throughput sequencing. ChIP-Seq signal 'peaks' are usually determined through peak calling algorithms and then analysed with software like MEME-ChIP (Kulakovskiy et al., 2010) or ChIPMunk (Machanick et Bailey, 2011) to look for enriched motifs within the pulled-down regions.

However, mapping of binding sites could be extremely difficult due to occasional low ChIP enrichment, possible clustering of binding sites in close proximity, and fragment size heterogeneity (a ChIP-Seq peak can cover even hundreds of bases, while the TF binding site are usually just few bp) (Rhee et Pugh, 2011). ChIP-exo, ChIP-nexus are improved version of the ChIP-Seq protocol to overcome these difficulties, where excess sequences are cut with exonucleases in order to have a narrower resolution of the binding sites (He et al., 2015). However these approaches are still not sufficient for a robust *de novo* motif discovery because, due to their own protocol limitation, they are not able to identify all possible binding sites, and indirect or cooperative binding events can alter the identification of a TF binding preference (Furey, 2012). Novel techniques like DNase-Seq, ATAC-Seq, and FAIRE-Seq have been developed as alternatives to investigate the transcription regulation in a TF-independent way: they identify regions with accessible chromatin using a non-specific DNA nuclease, transposase or formaldehyde crosslinking together with phenol-chloroform extraction (Boyle et al., 2011; Buenrostro et al., 2013). However, their ability in identifying TF binding sites through 'footprints' has been debated (Sung et al., 2016) and it seems to be TF-dependent: for example Sung et al. showed that DNase-Seq cannot fully capture footprints of TFs with short DNA residence time like SOX2 and glucocorticoid receptor (Sung et al., 2014).

Another method of genome tagging has been used to investigate transcription: the CRISPR/Cas9 technology.

## 1.7 CRISPR/Cas9 technology

The Clustered Regularly Interspaced Short Palindromic Repeats (CRISPRs) accompanied by CRISPR-associated (Cas) proteins is a system used by bacteria and archaea to defend themselves against foreign DNA elements (Barrangou et al., 2007; Garneau et al., 2010). There are different types of CRISPRs according to different bacteria species (Makarova et al., 2015), but the one used in this project is an engineered version of the type II CRISPR system from *Streptococcus pyogenes.* The key components of this system include the specific cleavage of double-stranded DNA mediated by Cas9 (Barrangou et al., 2007; Garneau et al., 2010), the presence of a short DNA sequence adjacent to the RNA-binding site, called protospacer-adjacent motif (PAM) as a mechanism for discriminating self from non-self (Mojica et al., 2009), and the presence of a small transactivating CRISPR RNA (tracrRNA), which directs the post-transcriptional processing and maturation of the CRISPR RNA (crRNA) through sequence complementarity (Deltcheva et al., 2011). The entire process is defined by three different stages: the spacer acquisition, CRISPR-Cas expression and DNA interference (Amitai et Sorek, 2016) (Fig. 1.7).

During the first stage, a short protospacer sequence from a previous mobile element is incorporated into the CRISPR array as a new spacer (Heler et al., 2014): which protospacer has to be used is decided by specific recognition of protospacer adjacent motifs (PAMs) present in the invading plasmid and phage genomes (Mojica et al., 2009). These motifs are a short (2–5 nucleotide) sequence essential for the cleavage of the target DNA during the interference stage (Mojica et al., 2009). During spacer acquisition, spacers are preferentially selected from protospacers that have a cognate PAM for the CRISPR–Cas system in question (Mojica et al., 2009; Heler et al., 2015).

After that, the crRNA will be generated: the CRISPR array is transcribed into a long precursor (pre-crRNA) and processed by endonucleases into mature crRNAs (single spacer surrounded by partial CRISPR repeat sequences on one/both sides (Brouns et al., 2008). At the end these mature crRNAs form complexes with Cas proteins that will be targeted for DNA degradation (van der Oost et al., 2014). The processing of the pre-crRNA transcripts involves base pairing between a small transactivating crRNA (tracrRNA) and the repeat segment of the pre-crRNA, followed by the cleavage within the repeat region by an endogenous RNase III (Deltcheva et al., 2011).

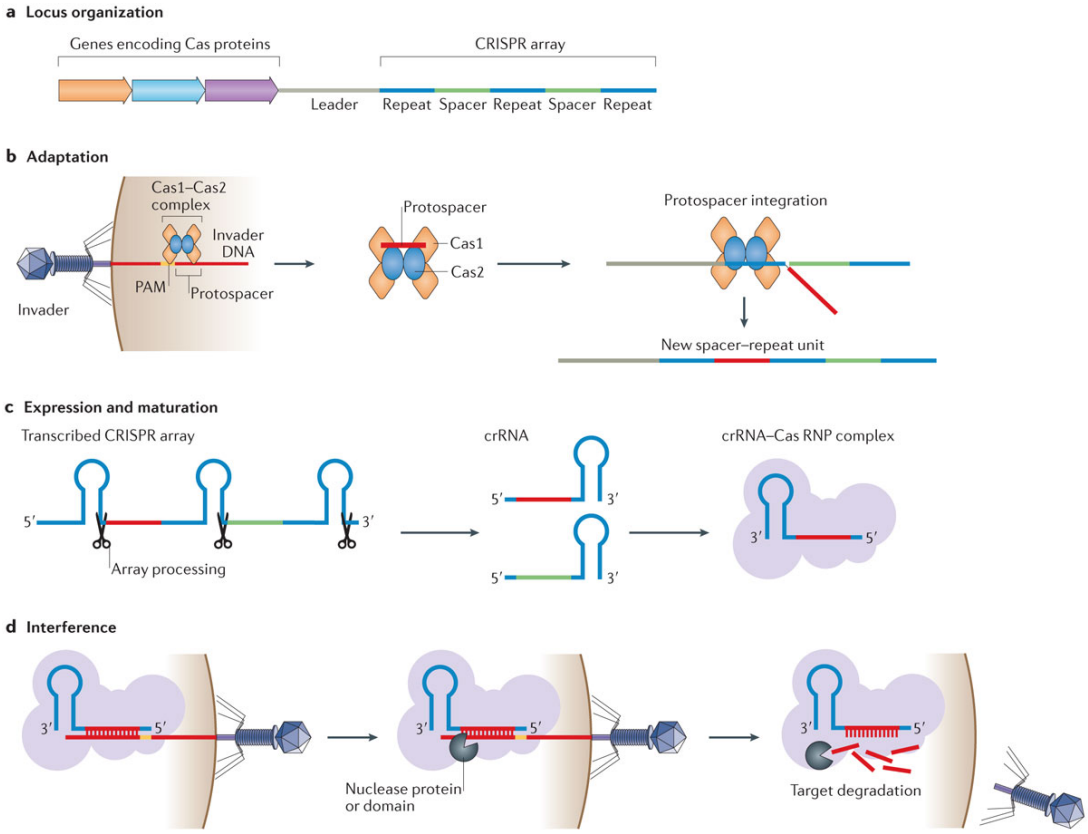



**Figure 1.7: Stages of CRISPR-Cas immunity.** a) Organization of a CRISPR–Cas locus in a bacterial or archaeal genome. The numbers, order and identities of the cas genes are different between CRISPR–Cas subtypes, and so the number of spacer–repeat units between species. b) In the adaptation stage, the Cas1–Cas2 complex (two Cas1 dimers and a single Cas2 dimer) acquires a protospacer from the invader DNA and integrates it as a new spacer into the CRISPR array, and the first repeat is duplicated. c) In the expression and maturation stage, the CRISPR array is transcribed and processed into mature CRISPR RNAs (crRNAs), containing a transcribed spacer and part of the repeat sequence. They form ribonucleoprotein (RNP) complexes with Cas proteins, different from different subtypes. d) During interference, the crRNA–Cas RNP complex identifies the target DNA through complementary base-pairing in the presence of a protospacer-adjacent motif (PAM), and it is then degraded by nuclease proteins or domains. The position of the PAM and the identity of the nuclease that degrades the target are different in different CRISPR–Cas subtypes (Amitai et Sorek, 2016).

This system utilizes the multi-functional Cas9 protein to target and degrade DNA through the guide of a dual-RNA heteroduplex made by a crRNA and a tracrRNA (Deltcheva et al., 2011). This has been used to introduce double-stranded DNA breaks in the genomes of eukaryotic cells that are site-specific and can be repaired by NHEJ (non-homologous end joining), or HDR (homology-directed repair), generating site-specific genome modifications (Mali et al., 2013).

Alternative versions have been developed to adapt to different biological systems for genome studies. For example, it can be programmed with single RNA molecule combining tracrRNA and crRNA features to cleave specific DNA sites. This small guide RNA (sgRNA) contains a designed hairpin that mimics the tracrRNA-crRNA complex (Jinek et al., 2012). The binding between the sgRNA and the target DNA causes the double-strand breaks because of the endonuclease activity of Cas9. In particular, the version we used for this project is a nuclease-deficient Cas9 (dCas9): it allows direct manipulation of the transcription process without genetically altering the DNA sequence (Qi et al., 2013). Furthermore, it allows the recruitment of diverse effector proteins for gene regulation at the transcription level (Fig. 1.8). For example it has been frequently fused to various transcription factors such as KRAB (Krüppel-associated box) (Gilbert et al., 2013) or four concatenated mSin3 interaction domains (SID4X) (Konermann et al., 2013) to enhance transcription repression (CRISPRi) (Fig. 1.8, a). It has also been used to activate the expression of target genes (CRISPRa). In this case, dCas9 is fused with VP64 (herpes simplex VP16 activation domain) or the p65 activation domain (p65AD) together with multiple gRNAs (Maeder et al., 2013), or synergistic activation mediator (SAM), like SunTag (Konermann et al., 2014) together with single gRNAs (Gilbert et al., 2013; Perez-Pinera et al., 2013) and it was enough to activate transcription (Fig. 1.8, b and c).

In addition the sgRNA can be modified into a scaffold to recruit transcriptional regulators (Zalatan et al., 2015): it could be fused to orthogonal protein-interacting RNA aptamers, which recruit specific RNA-binding proteins (RBPs) (Fig 1.8, d). These particular sgRNAs are called scaffold RNAs (scRNAs) (Zalatan et al., 2015). On the basis of the scRNA and the coupled RNA aptamer, different RBP transcriptional modulators can be recruited to different genes and have different effects.

dCas9 fusion proteins can also act as sequence-specific, synthetic epigenome modifier, varying not only the epigenetic status but also the expression of the gene

of interest. Given the great number of functional epigenetic marks, from DNA methylation to histone modifications, future studies are necessary to develop a proper dCas9-based epigenetic modifier tool (Fig. 1.8, e, f) (Hilton et al., 2015; Wang et al., 2016).

Other uses of the dCas9 protein include chromosome imaging in live cells and dissection of long-range chromatin interactions (Chen at al., 2013; Ma et al., 2015). However, the version of dCas9 we used in this project is an affinity-tagged dCas9: it is suitable for studying proteins interacting with specific portions of the genome, like for example for chromatin immunoprecipitation (ChIP) studies. In this case dCas9 is usually tagged and targeted to a specific *locus* in order to be used to pulldown the proteins associated with that region. This methodology allows the identification of protein-genome interactions at specific genomic *loci*. Recently Fujita et al. used this method to successfully characterize proteins interacting with an interferon-γ-responsive promoter (Fujita et al., 2014). Because of the genome-wide off-target dCas9-binding events, proper controls and strong results validation are necessary.

The use of the technique is flexible and programmable due to the fact that it is based on a sgRNA and it requires just a 20bp-matching region to target a specific gene.

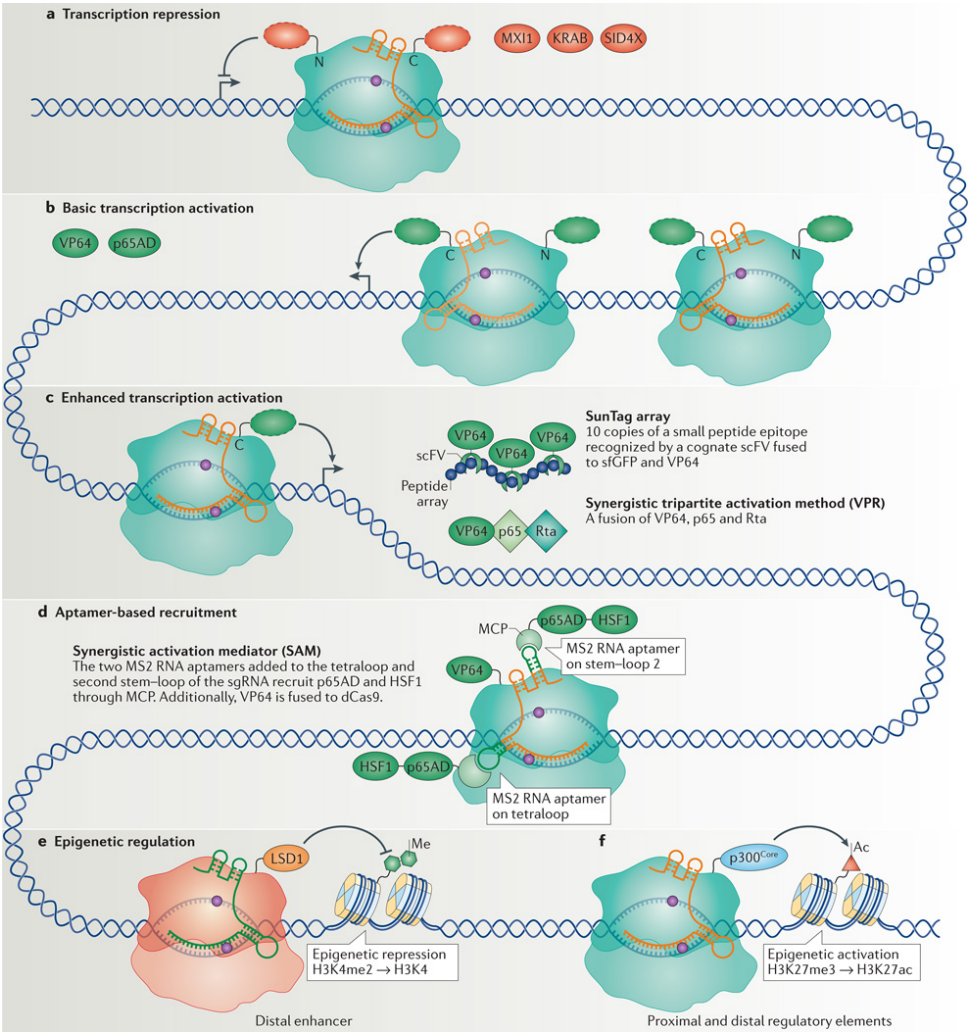



**Figure 1.8: CRISPR interference (CRISPRi) and CRISPR activation (CRISPRa) strategies.** a) Transcription repression by nuclease-deficient Cas9 (dCas9) fused with different repressor domains (red ovals), like Krüppel-associated box (KRAB) domain or four concatenated mSin3 domains (SID4X). b) Transcription activation by dCas9 fused with different activation domains (green ovals), like multiple repeats of the herpes simplex VP16 activation domain (VP64) or the NF-κB transactivating subunit activation domain (p65AD). Multiple single guide RNAs (sgRNAs; different shades of orange) are necessary to recruit multiple dCas9 fusion proteins in order to have an efficient transcription activation. c) Enhanced transcription activation: just one sgRNA to recruit one dCas9 per target gene. The SunTag array uses an array of small peptide epitopes (blue circles) fused to the C terminus of dCas9 to recruit multiple copies of single-chain variable fragment (scFV) fused to VP64. The synergistic tripartite activation method (VPR) uses a fusion of three transcription activators, VP64, p65 and the Epstein–Barr virus R transactivator (Rta), to achieve enhanced transcription activation. d) The aptamer-based recruitment system (synergistic activation mediator (SAM)) utilizes dCas9 with a sgRNA encoding MS2 RNA aptamers at the tetraloop and the second stem–loop (shown in dark green) to recruit the MS2 coat protein (MCP), fused to activators p65 and heat shock factor 1 (HSF1). Additionally, VP64 is fused to dCas9. e) Epigenetic regulation. Fusion of histone demethylase LSD1 to Cas9 removes the histone 3 Lys4 demethylation (H3K4me2) mark, for transcription repression. F) Fusion of the catalytic core of the histone acetyltransferase p300 ($p300^{Core}$) to dCas9 can acetylate H3K27 (H3K27ac), for transcription activation (Dominguez et al. 2015).

## 1.8 Aims of the present study

As described in section 1.5, targeting transcription regulation could represent a successful approach for breast cancer treatment. In particular, TNBC patients could benefit the most, since a unique, effective therapy is still missing. As shown by Wang and colleagues, the inhibition of transcription factors like CDK7 suggests a direct effect on the tumourigenity of TNBC. However, for this purpose, a deeper understanding of the transcription regulation process is still necessary.

In this thesis, the regulation of the trascription of genes highly and differentially expressed in TNBC compared to the other subtypes of breast cancer will be investigated. In particular, the main focus will be on the regulation of expression of transcription factors genes themselves.

In order to do so, a novel technique has been developed, which is the results of the combination of CRISPR/Cas9 and RIME proteomics: a catalytically inactive Cas9 (dCas9) will target potential promoter sequences of the genes of interest, and the transcription factors involved in the regulation of the gene expression will be identify through proteomic analysis. Common regulators between all or some of the studied *loci* will be further investigated in order to understand their role and importance for the biology of TNBC.

# CHAPTER 2: MATERIAL AND METHODS

## 2.1 Cell culture

The MDA-MB-231 and HS578T breast cancer cell lines, and HEK293T cell line (ATCC) were maintained in Dulbecco's modified Eagle's medium (DMEM, high glucose, Thermo Fisher Scientific, Life Technologies, Gibco), supplemented with 10% fetal bovine serum (FBS, Fetalclone III, Clontech) and 1% Penicillin/Streptomycin (P/S, Gibco) in a 37°C incubator with 5% $CO_2$. The BT549 and SUM159 breast cancer cell lines (ATCC) were maintained in Roswell Park Memorial Institute medium (RPMI 1640, Thermo Fisher Scientific, Life Technologies, Gibco), supplemented with 10% FBS and 1% P/S, in the same conditions.

All the clones derived from each cell line were grown in the same conditions as the parental one. For Doxycycline (Clontech) induction, the cells were treated 24 or 48 hours before collection for RIME, ChIP-qPCR, protein or RNA analysis. The final concentration of Doxycycline was 1µg/mL.

## 2.2 Cloning strategy

gRNAs were designed through a computational screening using the tool eCRISP (http://www.e-crisp.org/E-CRISP/) and ordered together with the complementary

from Sigma. They were designed with a specific overhang sequence for ligation into the cloning sites of the *PiggyBac* vector (PB-gRNA-Bsa1-EF1$\alpha$-RCB): the forward gRNA primers have a CTTG sequence at the 5' end and the complementary sequences an AAAC at the 3' end. They are as follows:

| Primers | Sequences |
|---|---|
| *FOXC1* forward | 5'-CTTG TGCGTAAAAAAGTCCTCGCC-3' |
| *FOXC1* complementary | 5'-AAAC GGCGAGGACTTTTTTACGCA-3' |
| *ELF5* forward | 5'-CTTG ACAGACAGGTCCGTTTGGTT-3' |
| *ELF5* complementary | 5'-AAAC AACCAAACGGACCTGTCTGT-3' |
| *SOX10* forward | 5'-CTTG CAGCTCCCAAGTCCTCTTCC-3' |
| *SOX10* complementary | 5'-AAAC GGAAGAGGACTTGGGAGCTG-3' |
| *NFIB* forward | 5'- CTTG GAAGAAGAAAAGCCAGCAAA-3' |
| *NFIB* complementary | 5'- AAAC TTTGCTGGCTTTTCTTCTTC-3' |
| *NFE2L3* forward | 5'-CTTG TGCGGCCCCTCCCACGGGCG-3' |
| *NFE2L3* complementary | 5'-AAAC CGCCCGTGGGAGGGGCCGCA-3' |

**Table 2.1: gRNA sequences designed for cloning.**

Every complementary pair of primers (1mM) was annealed for 5 minutes at 95°C in the thermocycler, and the temperature was dropped to 25°C with a decrease of 0.2°C per second. 5µg of the *PiggyBac* vector were digested at 37°C for 3 hours with the *BsaI-HF* restriction enzyme (NewEngland BioLabs) to create the ligation sites. The vector was design to have two *BsaI* restriction sites, which provide directional cloning with a single enzymatic digestion. The product was purified by agarose gel electrophoresis (1% agarose gel), and the 9000bp (base pairs) band was extracted using the QIAquick Gel Extraction Kit (Qiagen).

The linearized vector was then combined with the annealed gRNA oligos (v:v = 1:1) and ligated for 2 hours at room temperature using the T4 DNA ligase (NewEngland BioLabs) before transforming into *E. coli* bacteria.

## 2.3 Heat-shock transformation protocol of chemically competent *E. coli* cells

For transformation, 100μL of NEB 5-α Competent *E. coli* cells were thawed on ice and mixed with 5μL of the experimental DNA. Cells were incubated on ice for 30 minutes, heat-shocked in a 42°C water bath for 30 seconds and then immediately transferred on ice for 5 minutes. 950μl of SOC medium was then added to the cells and incubated at 37°C for 1 hour in a rotary shaker at 225-250rpm. For positive selection (the vector contains an Ampicillin Resistance gene), 100-200μl of transformation mixtures were plated on LB agar plates containing Ampicillin (100μg/mL) and incubated overnight at 37°C. The following day some colonies were picked and grew overnight in a rotary shaker at 225-250rpm at 37°C in a selection media (LB with Ampicillin, 10μg/mL).

## 2.4 Plasmid DNA extraction and screening for positive clones

Plasmid DNA was isolated from 10mL of overnight culture using QIAprep Spin Miniprep kit (Qiagen) according to the manufacturer's instructions. Briefly, the bacterial culture was harvested and lysed by high alkaline conditions. The plasmid DNA was then adsorbed on a QIAprep membrane, washed, eluted with nuclease free water and quantified with Nanodrop.

The oligo insertion was determined by double digestion using *NotI-HF* and *HpaI* restriction enzymes (NewEngland BioLabs). Agarose gel electrophoresis was used to identify positive clones and confirmed by DNA sequencing. 100ng/μL of every vector with the gRNAs inserted were sequenced using the following primer: 5'-AATCGCATAACTTCGTATAATGTA-3'. The sequencing was performed by the GENEWIZ sequencing service.

## 2.5 Transfection of cancer cells with lipofectamine

MDA-MB-231 cells were counted using the haemocytometer and seeded in 6-well-plates 24 hours prior to transfection in order to reach confluency on the next day. The transfection mix was prepared according to the manufacturer's protocol using a ratio of 15μL of lipofectamine LTX (with PlusTM Reagent, ThermoFisher Scientific)

for 3.5µg of total DNA to transfect (dCas9 vector, PB-TRE-multag-dvas9-mtag, and gRNA vector, PB-gRNA-Bsa1-EF1$\alpha$-RCB) and 0.5µg of Transposase expressing vector.

24 hours post transfection the cells were selected with Blasticidin (50µg/mL, ThermoFisher Scientific, Life Technologies, Gibco) and the treatment was continued for 5 consecutive days.

## 2.6 Flow cytometry analysis and sorting strategy

For flow cytometry, cells were detached as normal, spun down, washed with Hanks' Balanced Salt Solution (HBSS, calcium, magnesium) containing 1% FBS and re-suspended in 500µL of the same media, filtered and analysed at the Cell Sorter equipped with 488nm and 561nm lasers (SH800S Cell Sorter, Sony Biotechnology).

Cells were sorted into their normal growing media and plated according to the final number of cells.

## 2.7 Gene expression studies

Total RNA was extracted using the RNeasy Plus Mini Kit (QIAgen) according to the manufacturer's protocol. RNA was quantified spectrophotometrically (Nanodrop). Total cDNA was synthesized from the RNA by reverse transcription using the transcription Reverse Transcriptase (RT, Roche) following this protocol: 1.5µg of RNA was incubated at 65°C for 5 minutes with primers random hexamers (Promega). The 5X RT Buffer (Roche), RNasin Ribonucleotide Inhibitor (PROMEGA), 10mM of dNTP (Biolab) and the RT were then added to the samples and the reaction was carried out according to these cycles: 25°C for 10 minutes, 42°C for 40 minutes and 70°C for 10 minutes. The expression of selected genes was quantified by RT (real-time)-PCR using the SYBER Green Master mix (Applied Biosystem).

Gene-specific primers were chosen among the ones available on Primer Bank (https://pga.mgh.harvard.edu/primerbank/) and are as follows:

| Primers | Sequences |
|---|---|
| *GAPDH* forward | 5'- ACCCAGAAGACTGTGGATGG-3' |
| *GAPDH* complementary | 5'-TCTAGACGGCAGGTCAGGTC-3' |
| *FOXC1* forward | 5'-TGTTCGAGTCACAGAGGATCG-3' |
| *FOXC1* complementary | 5'-CAGTCGTAGACGAAAGCTCC-3 |
| *ELF5* forward | 5'-CTATGGAGGGTGAGAGCAGA-3' |
| *ELF5* complementary | 5'-GTACACTAACCTTCGGTCAACC-3' |
| *SOX10* forward | 5'-CCTCACAGATCGCCTACACC-3' |
| *SOX10* complementary | 5'-CATATAGGAGAAGGCCGAGTAGA-3' |
| *NFIB* forward | 5'-AAAAAGCATGAGAAGCGAATGTC-3' |
| *NFIB* complementary | 5'-ACTCCTGGCGAATATCTTTGC-3' |
| *NFE2L3* forward | 5'-TGGGCAAAAGCGATTAAGGG-3' |
| *NFE2L3* complementary | 5'-AGGTGAGGTCATTGCTGTCT-3' |
| *MTA2* forward | 5'-CCAAGACATCTGTGGGTCCT-3' |
| *MTA2* complementary | 5'-GTCGAAGGGAGTGAGGAGTG-3' |
| *CDK1* forward | 5'-TTTTCAGAGCTTTGGGCACT-3' |
| *CDK1* complementary | 5'-CCATTTTGCCAGAAATTCGT-3' |
| *CDK6* forward | 5'-CCAGGCAGGCTTTTCATTCA-3' |
| *CDK6* complementary | 5'-AGGTCCTGGAAGTATGGGTG-3' |

**Table 2.2: Primers designed for RT-PCR. *GAPDH* was used as a control gene to obtain normalized values.**

Assays were performed in triplicate and the results were normalized for *GAPDH* expression and then calculated as fold induction of RNA expression compared to controls.

## 2.8 Western blot analysis

Every cell line was grown in 6-well plates until confluency, treated and then washed twice with ice-cold PBS and solubilized with 50mM Hepes buffered solution, pH7.5, containing 150mM NaCl, 1.5mM MgCl2, 1mM EGTA, 10% glycerol, 1% Triton X-100, a mixture of Protease Inhibitors (Aprotinin, PMSF and Naorthovanadate, cOmpleteTM Protease Inhibitor Cocktail). Protein concentration in the supernatant was determined using the Pierce BCA Protein Assay method (Thermo Scientific). Equal amounts (30 or 50µg, depending on experiment) of the whole cell lysate were electrophoresed through a reducing SDS/7% or SDS/10% (w/v) polyacrylamide gel on the basis of the size of the investigated proteins and electroblotted onto a PVDF membrane which was probed with primary antibodies against V5 tag (rabbit polyclonal antibody to V5 tag, Abcam, #ab9116 1:5000), MTA2 (rabbit polyclonal antibody to MTA2, Abcam, #ab8106, 1:1000), CDK1 (rabbit polyclonal antibody to CDK1, Abcam, #ab131450, 1:1000), CDK6 (rabbit polyclonal antibody to CDK6, Abcam, #ab151247, 1:1000) and α-Tubulin (mouse monoclonal antibody to alpha-Tubulin, Abcam, #ab7291, 1:5000). The levels of proteins were detected after 1 hour incubation at room temperature with the horseradish peroxidase-linked secondary antibodies (anti-rabbit and anti-mouse antibodies, respectively, 1:10000), by the ECL® (enhanced chemiluminescence) System (GE Healthcare).

## 2.9 ChIP (Chromatin immuno-precipitation) and ChIP-Seq (Chromatin immuno-precipitation Sequencing)

Every clone was grown in 2 x 15-cm-dishes to 80% confluency, and fixed with 1% formaldehyde (CellStor Pot) diluted in the growing media-without serum for 10 minutes at room temperature. The crosslink was then quenched adding glycine (1M, diluted in PBS) for 5 minutes. Cells were washed once with ice-cold PBS, scraped off after the addition of PBS and Proteinase Inhibitors, and prepared for sonication.

50µL of beads per immuno-precipitation (Dynabeads™ Protein A) were used and washed once with 0.5% BSA in PBS using a magnetic rack. They were re-suspended in PBS/BSA and 2.5µg antibody per ChIP (anti-rabbit IgG antibody, Cell Signalling; rabbit polyclonal antibody to V5 tag, Abcam; rabbit polyclonal antibody to MTA2, Abcam; rabbit polyclonal antibody to CDK1, Abcam; rabbit polyclonal antibody to CDK6), and rotated overnight at 4°C. The day after, they were washed

three times with PBS/BSA to remove the unbound antibody and re-suspended in 100µL PBS/BSA just before combining them with the sonicated lysate.

Every sample was washed 2 times with different buffers (Lysis Buffer 1, Lysis Buffer 2, Table 2.3) containing Proteinase Inhibitors, and after every wash rotated at 4°C for 10 minutes after the first wash, 5 minutes after the second one, and pelleted. They were then re-suspended in Lysis Buffer 3 (Table 2.3) with Proteinase Inhibitor and sonicated for 30 seconds on, 30 seconds off for 8 cycles. To evaluate the efficacy of the sonication, 10µL of sonicated chromatin was incubated at 95°C for 5 minutes to reverse the crosslink, treated with 1µL of RNaseA (ThermoFisher Scientific) for 15 minutes at 37°C, followed by 1µL of Proteinase K (ThermoFisher Scientific) for 15 minutes at 55°C to eliminate RNA and protein contamination, and run on an agarose gel.

| Lysis Buffer 1 | Lysis Buffer 2 | Lysis Buffer 3 |
|---|---|---|
| 1M Hepes KOH,<br>pH7.5 5M NaCl,<br>0.5M EDTA<br>pH8 50% Gylcerol<br>10% IGEPAL<br>10% Triton X-100<br>ddH2O | 1M Tris HCl,<br>pH8 5M NaCl,<br>0.5M EDTA<br>pH8 0.5M EGTA<br>ddH2O | 1M Tris HCl,<br>pH8 5M NaCl,<br>0.5M EDTA,<br>pH8 0.5M EGTA<br>pH8 10% Na-deoxycholate<br>20% N-lauroylsarcosine<br>ddH2O |

**Table 2.3: Buffers used for ChIP and RIME.**

30µL of 10% Triton X-100 were added to the sonicated lysate and pelleted. The cell lysate samples were then diluted with Lysis Buffer 3 to 1mL with a final concentration of 1% Triton X-100. 25µL of the sample were taken as input and store at 4°C. The rest of the sample was added to the prepared beads and rotated overnight in the cold room.

The following day the samples were washed 5 times in RIPA buffer (150mM NaCl, 10mM Tris, pH7.2, 0.1% SDS, 1% Triton X-100, 1% NaDeoxycholate) using a magnetic rack, and 1 time in TE/NaCl (Tris 10mM, EDTA 1mM/Sodium chloride 50

mM) buffer. They were then re-suspended in 100µL of elution buffer (1% SDS, 0.1 M $NaHCO_3$), while the inputs were re-suspended in 75µL of it. To reverse the crosslink, they were incubated at 65°C overnight. The following morning the samples were centrifuged and the supernatant transferred in a new tube to eliminate the beads. They were then treated with 1µL of RNaseA for 30 minutes at 37°C, and 1µL of Proteinase K for 2 hours at 55°C. The DNA was extracted using the MinElute Reaction Cleanup Kit (QIAgen), according to the manufacturer's protocol.

The enrichment of dCas9 on targeted sequence was assessed by qPCR using the SYBER Green Master mix (Applied Biosystem). Primers were chosen among the ones available on NCBI Primer-BLAST (http://www.ncbi.nlm.nih.gov/tools/primer-blast/) to amplify a sequence of almost 200bp containing the gRNA targeting site, and they are as follows:

| Primers | Sequences |
|---|---|
| *FOXC1* forward | 5'- TCATTCGGAGGCGGTTCTCA -3' |
| *FOXC1* complementary | 5'- CAGCCGCTTAAGGAAGCATT -3 |
| *NFIB* forward | 5'- ACAAAGCAAACCAAGCAGGA -3' |
| *NFIB* complementary | 5'- GGAGGAAGAGCCTATCGCTT-3' |
| *NFE2L3* forward | 5'- ACTTCTGCTCCCAGAAAGCCT -3' |
| *NFE2L3* complementary | 5'- TCGGGAGAAGCGAAGAAGGAG-3' |

**Table 2.4: Primers designed for ChIP.**

For every ChIP-Seq experiment, the same protocol as for ChIP was followed, but with some modifications. In this case, the cell clones were maintained in DMEM supplemented with 10% fetal bovine serum tetracycline-free (Tet System Approved FBS, Clontech) and 1% P/S in a 37°C incubator with 5% $CO_2$ for 10 days before the actual experiment. Every cell line was then grown in 4 x 15-cm-dishes to 80% confluence, where two plates were used to test the sonication efficiency. 100µL of beads per immunoprecipitation (IP) were used, and re-suspended in PBS/BSA and 10µg antibody per ChIP.

ChIP DNA was sequenced on an Illumina machine by the genome core facility at CIGC (CRUK Cambridge Institute Genomic Core). ChIP-Seq data were analysed by Dr Mike Firth and Dr Jonathan Cairns at AstraZeneca (AZ), Cambridge, UK. In brief, each library was divided into two and sequenced on different lanes. Reads were subsequently run through a pipeline to remove adaptor sequences and align to the reference genome (human_g1k_v37) using mem algorithm in BWA. Next, reads from the two runs were combined into a single BAM file using samtools. Reads falling into blacklistedgenomic regions were removed using bedtools intersect before marking and removing duplicate reads using Picard tools. Next, picard tools was used to sample ~105 million reads from each BAM file. Significantly enriched genomic regions relative to input DNA were identified using MACS2 with *p*-value cutoff of 1.00e−05.

To generate the heatmaps, mapped read counts were calculated in a 10 bp window and normalised as reads per kilobase per million (RPKM mapped reads) using bamCoverage module from deeptools. This coverage file was then used to compute score matrix ± 1 kb around peak summits using computeMatrix reference-point module (from deeptools). Heatmaps of binding profiles around peak summits were generated using plotHeatmap module in deeptools. Number of overlapping peaks between samples and nearest downstream genes to peaks were determined using ODS and NDG utilities, respectively, in PeakAnnotator (version 1.4). For annotating nearest downstream genes, Homo sapiens GRCh37 (release 64) from ensembl was used.

## 2.10 RIME (Rapid Immuno-precipitation Mass spectrometry of Endogenous proteins)

Every clone was grown in 12 x 15-cm-dishes to 80% confluency, and fixed with 1% methanol-free formaldehyde (Ultra Pure, Polysciences, Inc.) diluted in the growing media without serum for 8 minutes at room temperature. To stop the crosslink, glycine (1M, diluted in PBS) was then added for 5 minutes. The cells were washed twice with ice-cold PBS, scraped off after the addition of PBS and Proteinase Inhibitors, and prepared for sonication. Every sample was washed 2 times with different buffers (Lysis Buffer 1, Lysis Buffer 2, Table 2.3) containing Proteinase Inhibitors, and rotated at 4°C for 10 minutes after the first wash, 5 minutes after the second one, and pelleted. They were then re-suspended in Lysis Buffer 3 (Table

3.4) with Proteinase Inhibitor and sonicated for thirty seconds on, thirty seconds off for 6-8 cycles. 30µL of 10% Triton X-100 were added to the sonicated lysate and pelleted. The cell lysate samples were then diluted to 1mL with a final concentration of 1% Triton X-100 and added to the prepared beads and rotated overnight in the cold room.

100µL of beads per immunoprecipitation (PureProteome Protein G Magnetic Bead System, Merck Millipore; Dynabeads Protein A, ThermoFisher Scientific) were used and washed 3 times with 0.5% BSA in PBS using a magnetic rack. They were re-suspended in PBS/BSA with 10µg antibody per IP (anti-rabbit IgG, Cell Signalling; rabbit polyclonal antibody to V5 tag, Abcam; rabbit polyclonal antibody to Bcl11a, Bethyl, A300-382A), and rotated overnight at 4°C. The day after, they were washed 3 times with PBS/BSA to remove the unbound antibody and re-suspended in 100µL PBS/BSA just before combining them with the cell lysate.

The following day the beads were washed 10 times in RIPA buffer using a magnetic rack, and 2 times in ammonium hydrogen carbonate (AMBIC, 100mM). The supernatant was then removed completely and the beads were stored in -20°C until submission for Mass Spectrometry analysis. This last part was performed by the Biological Mass Spectrometry Facility of AZ, (Waltham, Massachusetts, USA), in particular by Jon DeGnore, according to the published protocol (Mohammed H et al., 2016) with the following changes.

62ng of trypsin (Roche Applied Science, Indianapolis, IN) were used for digestion in 100mM ammonium bicarbonate (Sigma). Sample were StageTip (Thermo Scientific) desalted prior to LC/MS/MS analysis. Nanopore electrospray columns were used (ThermoFisher EasySpray #ES802). The mobile phase used for gradient elution consisted of (A) 0.1% formic acid in water and (B) 0.1% formic acid in Acetonitrile.

The 120 minute gradient consisted of 2 to 25% B for 75 minutes, then 25% to 40% B for 20 minutes, then 40% to 95% B for 10 minutes, the 10 minutes holding at 95% B, then down to 2% B for 5 minutes. All steps were at 300nL/min. Tandem mass spectra (LC/MS/MS) were acquired on a Thermo Q Exactive plus mass spectrometer (Thermo Corp., San Jose, CA). The MS/MS spectra were searched against the NCBI non-redundant protein sequence database using the PEAKS software (Bioinformatics Solutions Inc., Waterloo, ON, CA) to produce a list of proteins identified for each sample. The precursor mass tolerance was set to 15ppm

and the MS/MS fragment mass tolerance to 0.05Da. Variable search modifications were oxidation of methionine, deamidation of Asparagine or Glutamine, and acetylation of the N-term. A false positive rate of 1% was used as a cut-off for peptide identifications.

The output comprises lists of confidently identified proteins, accession numbers, protein descriptions, peptide identifications and search statistics.

## 2.11 Knockdown strategy

ShRNAs were chosen among the ones available on the MISSION™ TRC shRNA libraries and purchased from Sigma as bacterial stock together with the control vector pLKO.1-puro (SHC002V). They are as follows:

| shRNA | Sequences |
|---|---|
| MTA2 #1<br>TRCN0000013377 | 5'-<br>CCGGCCTAGATTGTAGCAGCTCCATCTCGAGAT<br>GGAGCTGCTACAATCTAGGTTTTT-3' |
| MTA2 #2<br>TRCN0000013374 | 5'-<br>CCGGCCCTCTTGAATGAGACAGATACTCGAGTA<br>TCTGTCTCATTCAAGAGGGTTTTT-3' |
| CDK1 #1<br>TRCN0000000583 | 5'-<br>CCGGGTGGAATCTTTACAGGACTATCTCGAGAT<br>AGTCCTGTAAAGATTCCACTTTTT-3' |
| CDK1 #2<br>TRCN0000196602 | 5'-<br>CCGGGTTTCCATATGTTATGTCAACCTCGAGGTT<br>GACATAACATATGGAAACTTTTTTG-3 |
| CDK6 #1<br>TRCN0000196261 | 5'-<br>CCGGGAGAAGTTTGTAACAGATATCCTCGAGGA<br>TATCTGTTACAAACTTCTCTTTTTTG-3 |
| CDK6 #2<br>TRCN0000055435 | 5'-<br>CCGGTCTGGAGTGTTGGCTGCATATCTCGAGAT<br>ATGCAGCCAACACTCCAGATTTTT-3' |

**Table 2.5: shRNAs used for knockdown.**

Bacteria were plated on LB agar plates containing Ampicillin (100µg/mL) and incubated overnight at 37°C. The following day some colonies were picked and grew overnight in a rotary shaker at 225-250rpm at 37°C in the selection media (LB with Ampicillin, 10µg/mL). Plasmid DNA was isolated from 10mL of overnight culture using QIAprep Spin Miniprep kit (Qiagen) as described in previous session. Positive clones were confirmed by double digestion using *NcoI-HF* and *EcoRI-HF* restriction enzymes (NewEngland BioLabs).

## 2.12 Lentivirus production with Addgene (3 vectors) Packaging System

8 x $10^5$ HEK293T cells (passage 0-20) were seeded on a pre gelatinized 10$cm^2$ plate (0.1 % w/v) in DMEM + 10% FBS in order to reach 70% confluency the day after. On the day of transfection, 3µg of the experimental vector were added to 3mL of OptiMEM per well, together with 3µg of packaging mix (1µg of each plasmid: pMD2.G, Addgene #12259; pMDLg/pRRE, Addgene #12251; pRSVRev, Addgene #12253) and incubated for 5 minutes at room temperature. 36µL of lipofectamine LTX were then added and after 30 minutes of incubations at room temperature the mix was added to the seeded cells, where in the meantime their media had been replaced with D10 (DMEM + 10% FBS + GlutaMax w/o P/S). After 5 hours of incubation, the media was replaced with fresh D10, and left for 48 hours to allow the virus production. The media was then collected, filtered using a 0.45µm filter cartridge and Lenti-X concentrator was added. The solution was incubated O/N at 4°C, centrifuged at 1500 x g for 45 minutes: the pellet was then resuspended in 1/20$^{th}$ of the original volume using DMEM media. The virus not used was frozen down and kept at -80°C.

Cells were then selected with 1µg/mL of puromycin for 3-5 days.

## 2.13 Colony assay

2000 cells for every cell line were counted, centrifuged at 400g for 5 minutes and resuspended in 280µL of cold matrigel matrix (Corning). 70µL of each aliquot were then seeded in a 6 well plate in order to have 3 replicates with 500 cells per well. After an incubation of 15 minutes at 37°C, 2mL of growth media was added gently to each well. Colonies were counted after 7 days.

## 2.14 Statistical analysis

Statistical analysis was performed using ANOVA followed by Holm-Šídák test to compare the means of two or more independent groups. In this way we aimed to determine whether there is statistical evidence that the associated population means are significantly different. $P<0.05$ was considered as statistically significant.

# CHAPTER 3: IDENTIFICATION OF THE GENES OF INTEREST AND SYSTEM VALIDATION

## 3.1 Introduction

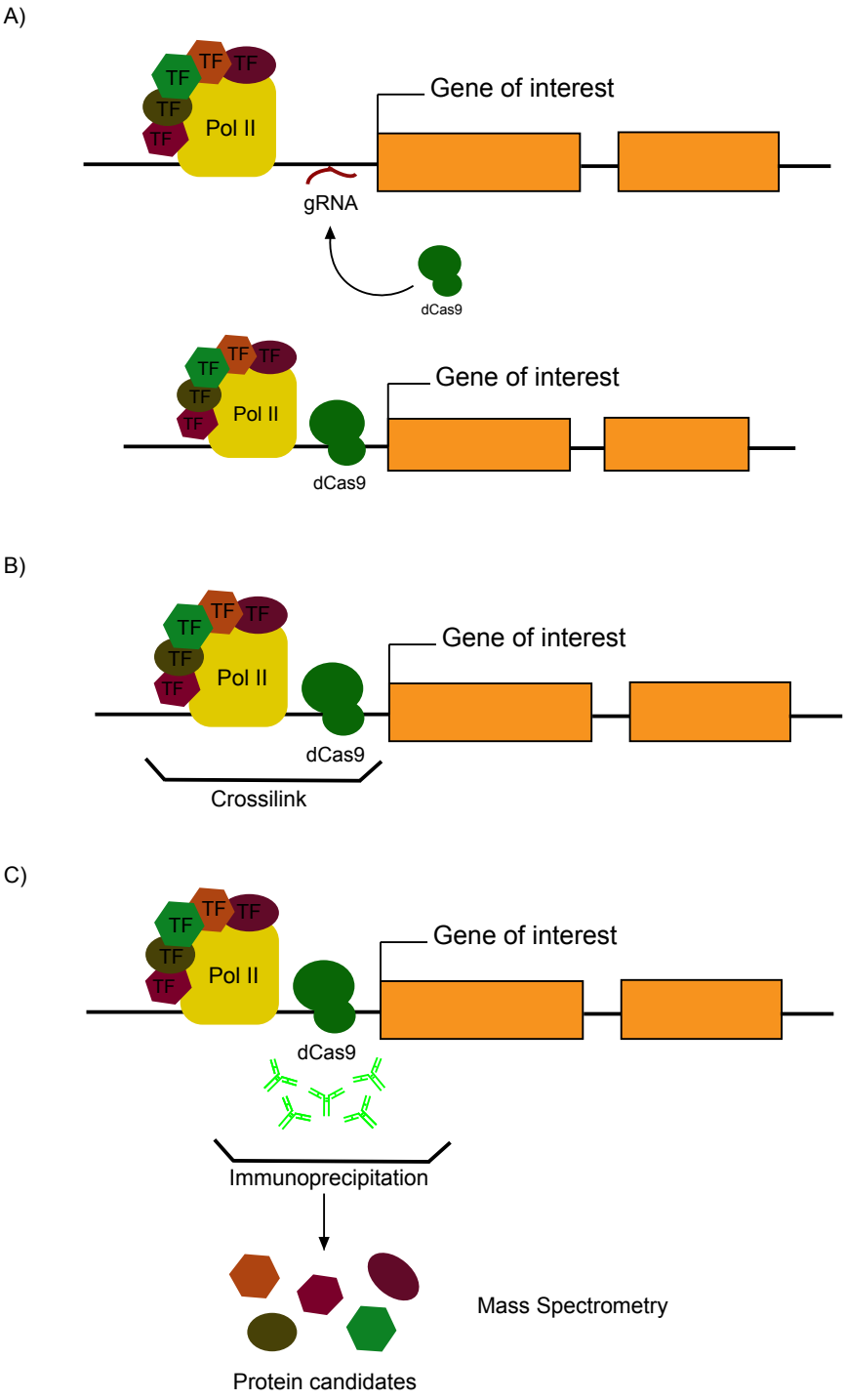
Despite the advances in diagnosis and treatment, breast cancer still remains a major health problem for women, and a high biomedical research priority. Worldwide, it is the most common cancer for women, with a very high mortality rate. One of the main challenges in treating this disease is that breast cancer is not a single entity, but a heterogeneous group of several subtypes with different biological and clinical behaviour. The most commonly used way of classifying these tumours is on the basis of the histopathological type and the expression of *ER*, *PR* and *HER2* genes: 75-80% of the cases are identifies as hormone receptor-positive breast cancers, while 10-15% of them are *HER2* overexpressing ones (Konecny et al., 2006). Triple negative breast cancer (TNBC) counts for the remaining 10-15% of the cases, and it is characterized by the absence of expression of the receptors mentioned above.

TNBCs tend to be extremely aggressive tumours, with a short survival and a relatively high mortality rate (Dent et al., 2007). In the last decades, multi-drug combination systemic therapies in the neoadjuvant and adjuvant settings have significantly improved patients' outcome, and recently significant treatment advances have been achieved with poly (ADP-ribose) polymerase (PARP) inhibitors and immunotherapy agents, although in some cases the prognosis still remains

poor. At the genomic level, this subtype has a very high genetic complexity such as high rate of point mutations, gene amplification and deletion (Cancer Genome Atlas Network, 2012). However, a common genetic alteration still has to be found, with the only exceptions of *PTEN*, *TP53* and *BRCA1* for some patients (Abramson et al., 2014; Foulkes et al., 2003).
Because of the high heterogeneity, the lack of driver aberrations causing the pathology and the relative high mortality rate, the development of new biological targeted treatments is essential.

In order to understand the complex biology behind the tumourigenity of TNBC, we combined a proteomic approach with the CRISPR/Cas9 technology (Fig. 3.1): we targeted the regulatory regions of highly expressed genes in TNBC with a catalytically inactive Cas9 protein (dCas9) to pull them down together with the associated proteins. These proteins were then identified through Mass Spectrometry (MS) and their role in the regulation of the transcription of the genes of interest was evaluated. In particular we focused on those proteins in common between all/some of the *loci* studied. For this project the proteomic approach we used was RIME (Rapid Immunoprecipitation Mass spectrometry of Endogenous proteins (Mohammed et al., 2013)).

With this investigation, we aimed to understand the transcription regulation of our genes of interest, in order to identify proteins that could be fundamental for the survival of TNBC tumours, and that could be the focus of future drug development studies.


A)
TF
TF
TF
TF
TF
Pol II
Gene of interest
gRNA
dCas9
TF
TF
TF
TF
TF
Pol II
Gene of interest
dCas9
B)
TF
TF
TF
TF
TF
Pol II
Gene of interest
dCas9
Crossilink
C)
TF
TF
TF
TF
TF
Pol II
Gene of interest
dCas9
Immunoprecipitation
Mass Spectrometry
Protein candidates

D)

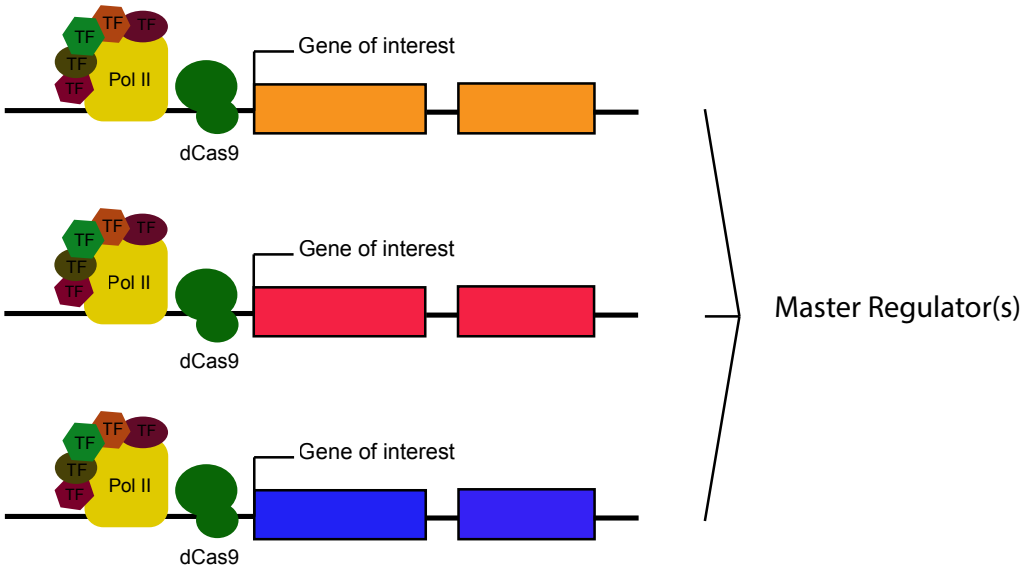



**Figure 3.1: Schematic representation of the project hypothesis.** A) The putative promoter sequence of a highly expressed gene in triple negative breast cancer (TNBC) was targeted by a gRNA to redirect the catalytic inactive Cas9 (dCas9) to a specific DNA sequence. B) The proximity of dCas9 to the transcription machinery normally recruited on the promoter was fundamental for the crosslinking step, according to RIME (Rapid Immunoprecipitation Mass spectrometry of Endogenous proteins) protocol: all the proteins physically close to each others were linked together and to dCas9. C) An antibody against dCas9 was used to immunoprecipitate it together with all the other crosslinked proteins, lately identified through Mass Spectrometry. D) The strategy was applied to several genes of interest, in order to identify one or some key transcription factor(s) that are necessary for the expression of the target gene (master regulator(s)).

## 3.2 Identification of the genes of interest

In order to identify key regulatory factors involved in the tumourigenicity of TNBC, we investigated the gene expression profile of TNBC/IntClust10 patients from the METABRIC study (Curtis et al., 2012). We set out to focus on genes upregulated in TNBC/IntClust10 but not in other subtypes of breast cancer. In addition, we narrowed our study to transcription factors as they could be new potential targets of therapeutic drugs for cancer treatment. Our analysis revealed that the top six most upregulated genes in these patients were *FOXC1*, *ELF5*, *BCL11A*, *SOX10, NFIB* and *NFE2L3* (Fig. 3.2). On the other side genes like *FOXA1*, *ESR1* and *GATA3* were not associated with TNBC (Fig. 3.2, fold negative change): their expression usually correlates with luminal subtypes of breast cancer, as shown by different expression profiling studies (Badve et al, 2007; Albergaria et al, 2009; Voduc et al, 2008)

Some of the highly regulated genes we found are well known in the literature to be specific for TNBC (*BCL11A* for example (Khaled et al., 2015)), while some others have not been associated with this disease to date (like *NFE2L3*). However, it has been shown that these genes play major roles in processes like stem cell maintenance, cell proliferation or migration (Chowdhury et al., 2017; Rhee et al., 2008), which are extremely important in a tumourigenic system.

It has to be mentioned that in this project we didn't consider *BCL11A*, although it was the third gene on our list, because it was already under examination in other projects in the laboratory. We then decided to proceed with the analysis of five genes: *FOXC1, ELF5, SOX10, NFIB* and *NFE2L3.*

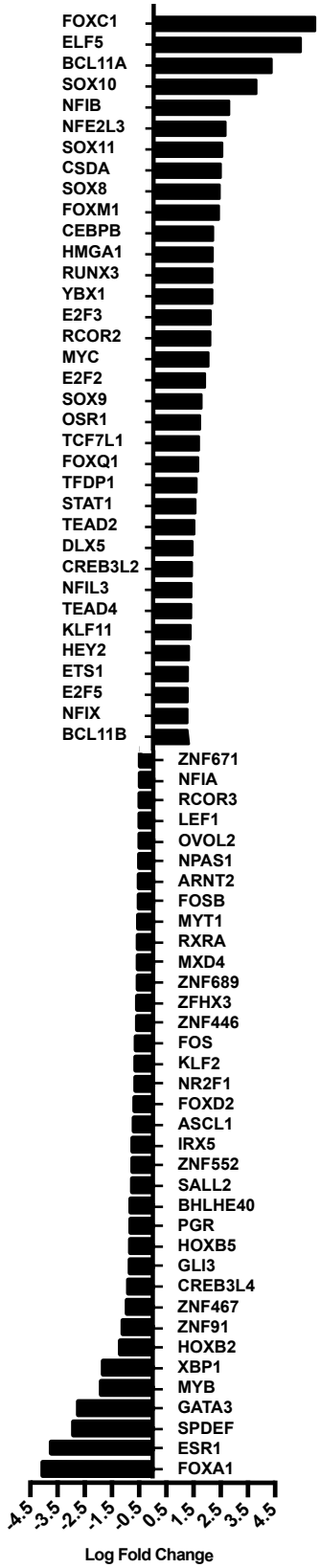
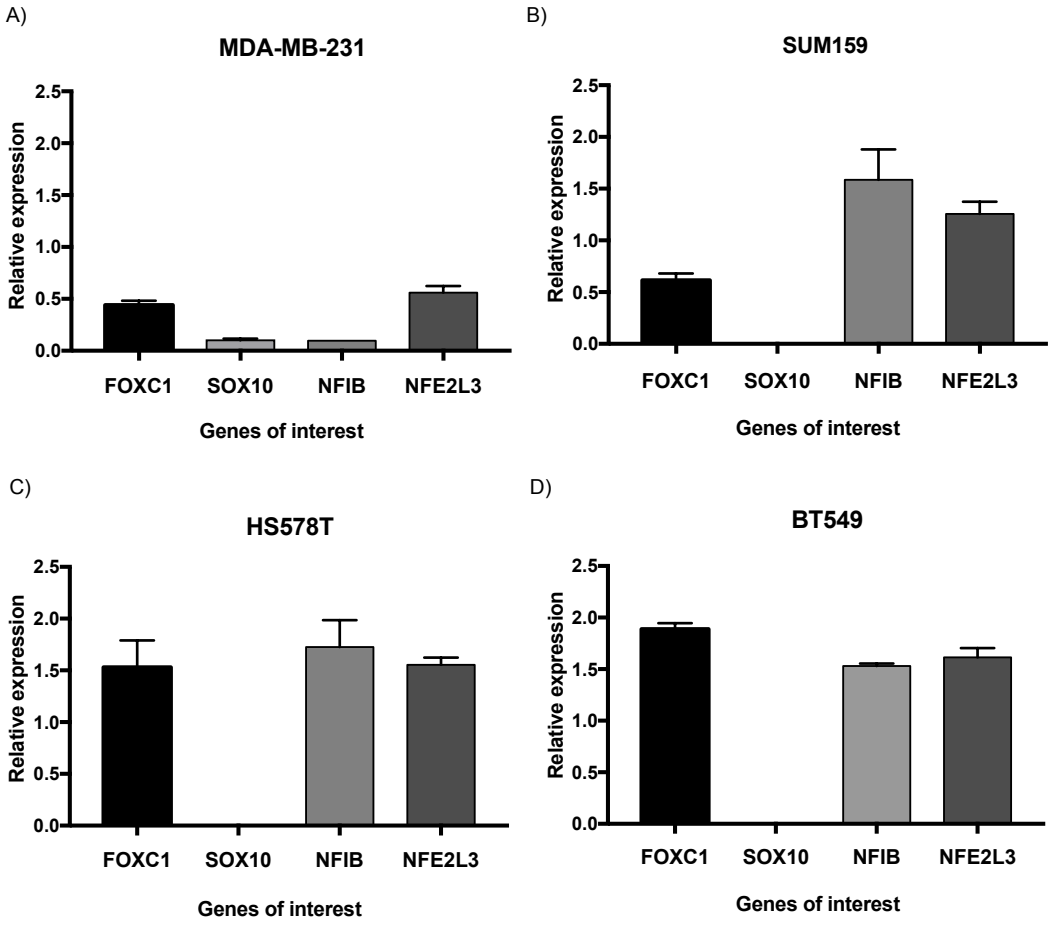



**Figure 3.2: Differentially regulated transcription factors in IntClust10 compared to the other clusters.** Values on the horizontal axis indicate the logarithmic fold change of expression between clusters. Only the top 35 highly expressed transcription factors are shown. Analysis was performed by Oscar Rueda, in Carlos Caldas' laboratory (CRUK-Cancer Institute, Cambridge).

To confirm the validity of our analysis, we investigated the basal expression of these genes in a panel of TNBC cell lines (MDA-MB-231, SUM159, HS578T and BT549) (Fig. 3.3, A, B, C and D). We were able to detect three out of five genes in all cell lines, *FOXC1*, *NFIB* and *NFE2L3*, with variable levels of expression. However, *SOX10* seemed to be expressed only in MDA-MB-231 cell line and for this reason, we decided to exclude it from further examinations, together with *ELF5*, undetected in all our cell lines, which was the second most differentially expressed transcription factor showed in TNBC. These results could be explained by the fact that the METABRIC data are from primary tumour material, while the system we used are cell lines that carry variability between themselves too. In addition, it has to be noted that the experiment was designed at that time without appropriate controls for the presence or absence of expression of the genes of interest. This limitation could have also affected the final readout (Fig. 3.3), altering the conclusions about level of expression or the association of these genes to TNBC. Among all, we decided to use MDA-MB-231 cell line as our model for our further experiments.

In addition, we compared the overall survival between TNBC patients carrying a mutated variant of the genes of interest and those without (Fig. 3.3, E) using data available from the TCGA database. Out of all of them, *FOXC1* variants seem to be the most frequent and the only ones associated with a poorer prognosis for these patients, followed by *NFIB.* Alterations in this last one in particular seem to significantly affect patient's survival in a long-term scale.


A)
MDA-MB-231
Relative expression
2.5
2.0
1.5
1.0
0.5
0.0
FOXC1
SOX10
NFIB
NFE2L3
Genes of interest
B)
SUM159
Relative expression
2.5
2.0
1.5
1.0
0.5
0.0
FOXC1
SOX10
NFIB
NFE2L3
Genes of interest
C)
HS578T
Relative expression
2.5
2.0
1.5
1.0
0.5
0.0
FOXC1
SOX10
NFIB
NFE2L3
Genes of interest
D)
BT549
Relative expression
2.5
2.0
1.5
1.0
0.5
0.0
FOXC1
SOX10
NFIB
NFE2L3
Genes of interest

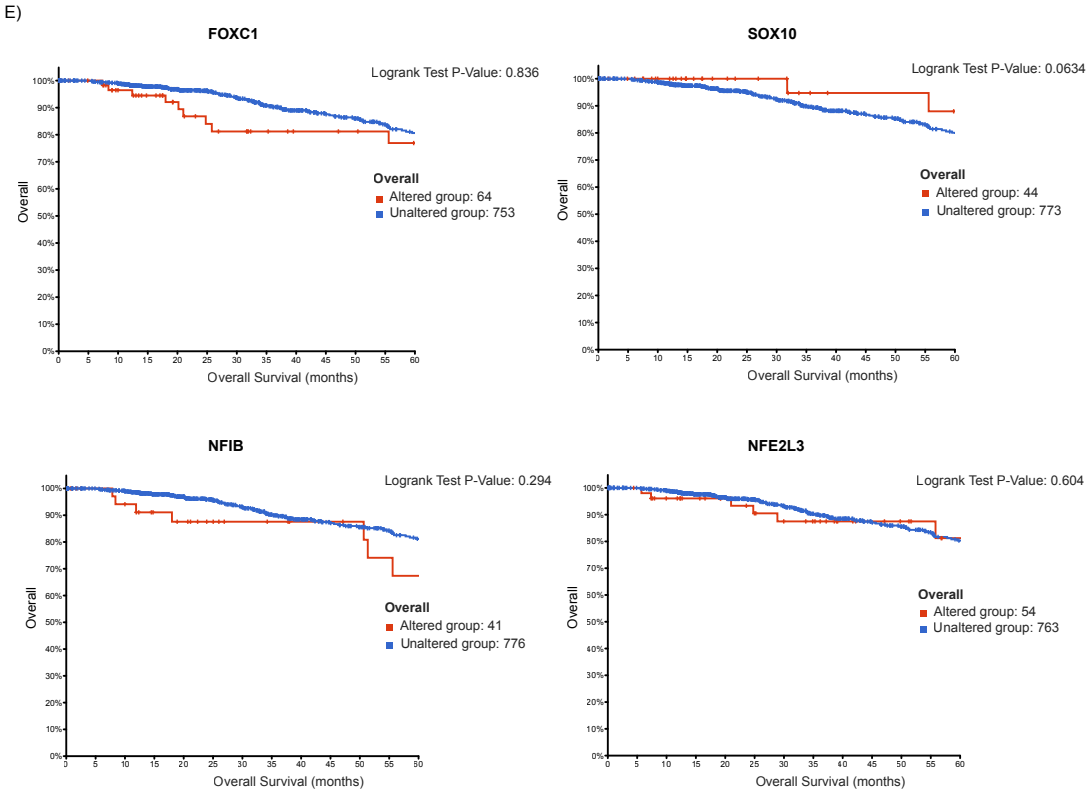



**Figure 3.3: Expression of genes of interest in TNBC cell lines panel and associated patient's survival.** mRNA expression levels of *FOXC1*, SOX10, *NFIB* and *NFE2L3* were determined by qPCR and normalized to glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*) for four different TNCB cell lines: MDA-MB-231 (A), SUM159 (B), HS578T (C) and BT549 (D). The error bars report standard deviations from triplicates. E) Comparison between the overall patient survival status between TNBC cases carrying the mutated gene of interested (altered group) and those expressing the unmutated form (unaltered group). Follow-up period: 60 months. Data modified from TCGA.

## 3.3 gRNA design strategy

To investigate the transcription regulation of the selected genes, we analysed their promoter region through the UCSC dataset (https://genome.ucsc.edu/): within it, we aimed to identify a sequence to target with our CRISPR/Cas9 approach. We based this research on some specific chromatin features and references in the literature (Gilbert et al., 2014): the presence of DNase Hypersensitivity clusters (a sequence of DNA that is sensitive to the cleavage of DNase I), histone modifications (in particular acetylation, marking a more relaxed structure of the chromatin), conservation of the DNA sequence among species and distance from the Transcription Starting Site (TSS, within 400bp upstream). All these features indicate an active chromatin state, which is most likely to be reached for transcription factor binding.

Through the online tool eCRISP (http://www.e-crisp.org/E-CRISP/) we obtained a list of possible gRNA candidates to target the desired sequence. Among them, we chose the gRNA with the highest E-Score (Efficiency-Score) and the lower number of predicted off-targets.

The position of the chosen gRNA for each gene has been reported in Fig. 3.4, 3.5 and 3.6. It is possible to notice how it varies between them: for some it is really close to the TSS (for example in *FOXC1* gene), for others more distal (as for *NFIB*). We tried to satisfy all the conditions described before in order to choose the potential best gRNA, but at the same time being flexible according to the features of every gene sequence.

A) Foxc1 gRNA position

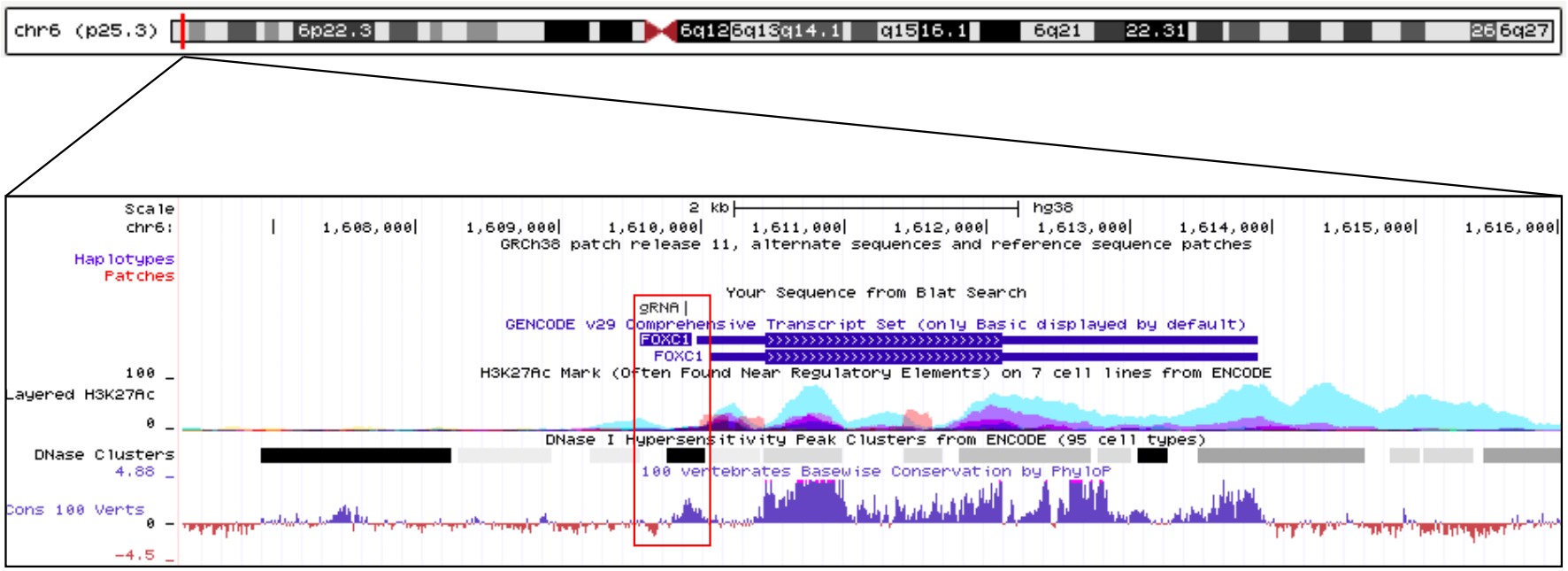



**Figure 3.4: Schematic representation of the position of gRNA within the promoter sequence of *FOXC1* gene.** The gene location on the respective chromosome is shown with a red bar. The promoter and gene sequences are magnified in order to show the position of the gRNA from the transcription starting site of the gene, highlighted with a red box. The presence of acetylation of histones (H3K27Ac) and DNA Clusters in the promoter sequence of the gene was considered to identify regions of open chromatin, so accessible to the recruitment of the transcription machinery. In addition, the level of conservation of the sequence was considered as an indication of the importance of the specific sequence among species. Every gRNA was designed within a distance from the transcription starting site of 400 bp, as previously described in Gilbert et al., 2013.

B) Nfib gRNA position

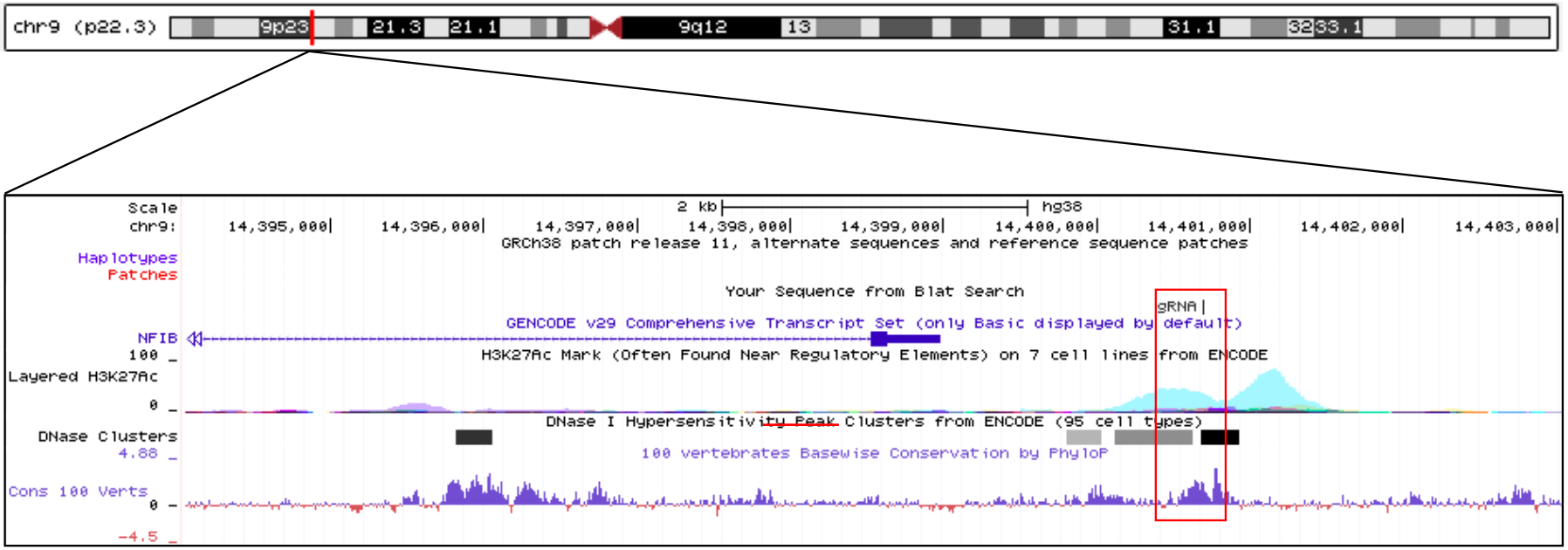



**Figure 3.5: Schematic representation of the position of gRNA within the promoter sequence of *NFIB* gene.** The gene location on the respective chromosome is shown with a red bar. The promoter and gene sequences are magnified in order to show the position of the gRNA from the transcription starting site of the gene, highlighted with a red box. The presence of acetylation of histones (H3K27Ac) and DNA Clusters in the promoter sequence of the gene was considered to identify regions of open chromatin, so accessible to the recruitment of the transcription machinery. In addition, the level of conservation of the sequence was considered as an indication of the importance of the specific sequence among species. Every gRNA was designed within a distance from the transcription starting site of 400 bp, as previously described in Gilbert et al., 2013.

C) Nfe2l3 gRNA position

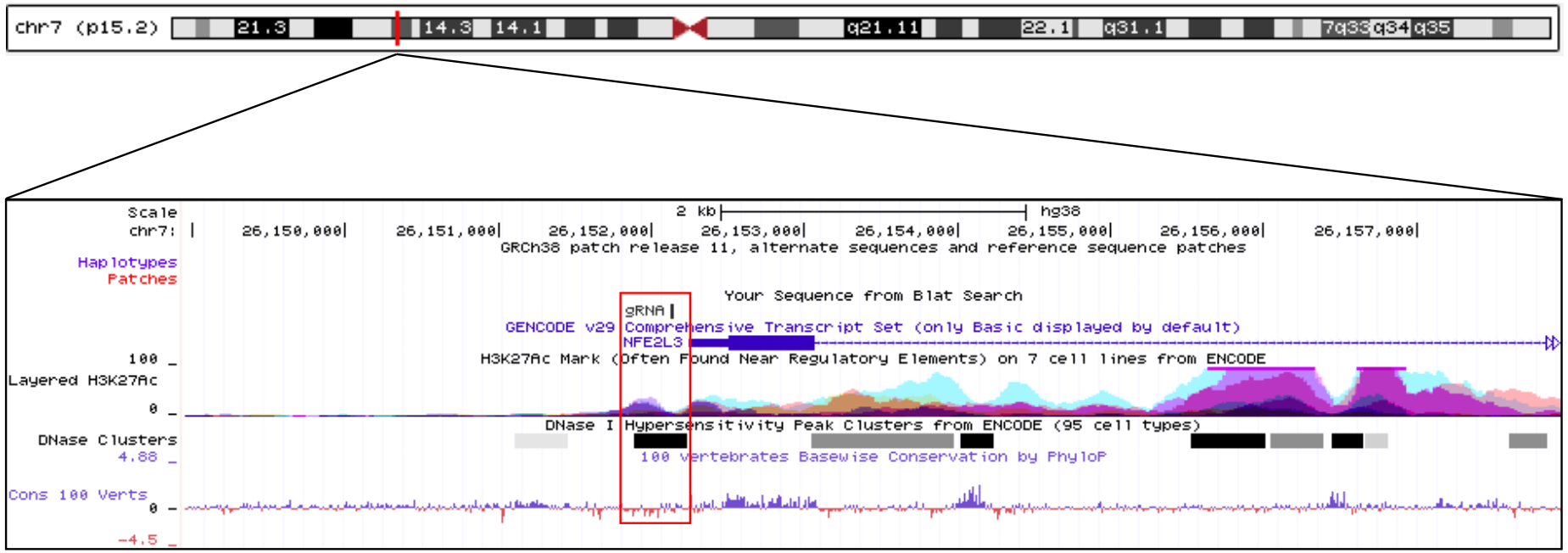



**Figure 3.6**: **Schematic representation of the position of gRNA within the promoter sequence of *NFE2L3* gene.** The gene location on the respective chromosome is shown with a red bar. The promoter and gene sequences are magnified in order to show the position of the gRNA from the transcription starting site of the gene, highlighted with a red box. The presence of acetylation of histones (H3K27Ac) and DNA Clusters in the promoter sequence of the gene was considered to identify regions of open chromatin, so accessible to the recruitment of the transcription machinery. In addition, the level of conservation of the sequence was considered as an indication of the importance of the specific sequence among species. Every gRNA was designed within a distance from the transcription starting site of 400 bp, as previously described in Gilbert et al., 2013.

To deliver dCas9 and the gRNA we used a transposon system constituted by two main vectors (Figure 3.7): the *PiggyBac* vector (PB-gRNA-Bsa1-EF1$\alpha$-RCB, Figure 3.7, A), coding for the gRNA, and the PB-TRE-multag-dcas9-mtag vector (Figure 3.7, B), coding for dCas9.

With this system we aimed for a stable integration of the vectors' DNA in the genome of the cells. Thanks to the presence of the enzyme PB transposase (transcribed by the *PBase* vector*,* courtesy of Dr Pentau Liu, Sanger), recognizing transposon-specific inverted terminal repeat sequences (ITRs) on both vectors, the DNA content is mobilized from the original site to the target genome, where they are going to integrate. This is possible through a 'cut and paste' mechanism into a genomic target region with a TTAA site.

To confirm the successful transfection and integration, both vectors were designed to deliver selection markers. The gRNA vector contains a Blasticidin-resistance gene, which allowed the cells to survive when Blasticidin was added to the culturing media, and a fluorescent marker, mCherry. The dCas9 vector contains the EGFP as a fluorescent marker, which was fused to the dCas9 (dCas9-2A-EGFP).

To allow for temporal control of dCas9 expression, we used a Tet-On inducible system (Baron et Bujard, 2000). Normally in a Tet-On system the transcription is induced by the presence of Tetracycline, which activates the recombinant Tetracycline controlled transcription factor (rTta). The rTta then binds the Tet Response Element (TRE) and initiates the gene expression. In our system the rTta was expressed under the control of the strong constitutive promoter, *EF1$\alpha$*, along with the constitutive expression of the gRNA under the control of a U6 promoter. In the other vector, the expression of dCas9 was regulated by the TRE: in this way not only the dCas9 could be expressed just when Tetracycline (or derivatives, as Doxycycline) was added to the system, but also in those cells which had been transfected with both vectors.

In addition, we decided to transfect the cells with the gRNA vector without any cloned gRNA inside, to which we will refer as 'Empty gRNA'. For some of our experiments, this was used as an internal control to evaluate the general perturbations within the cells due to the presence of this exogenous system.

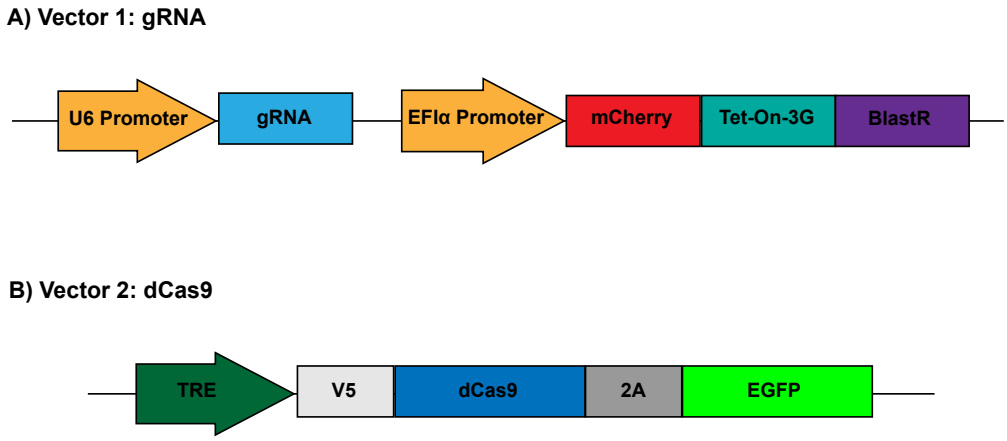



**Figure 3.7: Schematic representation of the vectors used for transfection.** A) Vector used to deliver the gRNA sequence, which is regulated by a U6 promoter. The vector also carries a mCherry gene for the fluorescence selection, a Blasticidin resistance gene (BlastR), and the tetracycline-inducible gene expression system (Tet-On-3G, third generation) for the expression of the tTA protein under the control of an EFIα promoter. B) Vector used to deliver the dCas9 sequence. It is a multi-gene expression system under the regulation of a Tetracycline Responsive Element (TRE). dCas9 is expressed together with a V5 tag and an EGFP fluorescence marker, cleavable through a 2A peptide.

## 3.4 Selection of double transfected cells

Initially the cells were selected in culture for 3 days through Blasticidin, induced for 24 hours with Doxycycline and analysed by flow cytometry for the expression of EGFP and mCherry. Initial attempts yielded low efficiency (0.02-1.9%, Fig. 3.8).

Therefore, a different strategy was developed, consisting of two cycles of transfections, each followed by population enrichment of transfected cells by FACS sorting as summarized in Fig. 3.9. Briefly, cells were transfected with both vectors and kept in culture with Blasticidin for 5 days. Just the mCherry positive cells were then sorted, in order to eliminate all those without the gRNA vector. After 7-10 days of recovery, the sorted cells were transfected for the second time with the dCas9 vector. This allowed us to increase the number of cells containing the dCas9 vector as well as the gRNA expressing vector. The cells were then expanded in culture, induced for 24 hours with Doxycycline and sorted for EGFP and mCherry signal (double positive cells) (Fig. 3.10). After the first transfection, only around 5% of the cells were mCherry positive, regardless of the gRNA analysed (Fig. 3.10, far left panel). After the second transfection we obtain higher percentage of double positive cells compared with the first attempt, but with different rates for every sample (far right panel). Although the strategy was the same, the efficiency of double transfection we obtain was different according to the gRNA analysed. The range of double positive cells varied between 5% (MDA-MB-231 + Nfe2l3 gRNA) and 10% (MDA-MB-231 + Foxc1 gRNA).

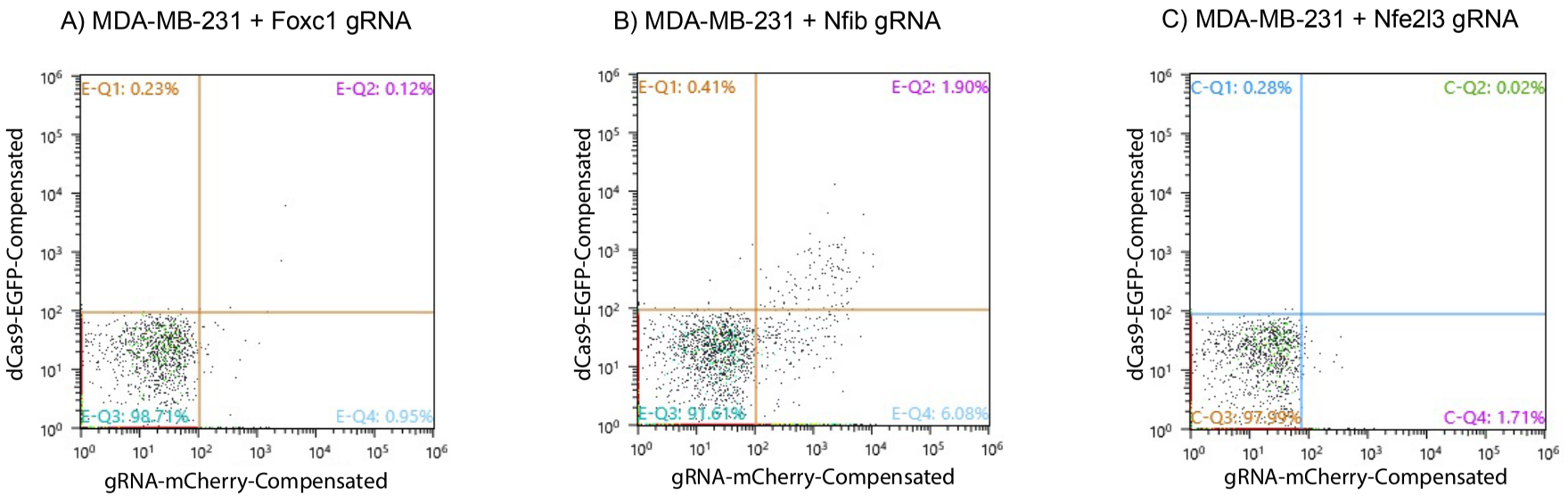



**Figure 3.8**: **First sorting attempt of the MDA-MB-231 clones.** After transfection with both mCherry-gRNA and EGFP-dCas9 vectors, cells were selected for a few days in culture with Blasticidin. Subsequently cells were induced with Doxycycline for 24 hours, and the double positive ($mCherry^{+}$ and $EGFP^{+}$) cells were sorted. MDA-MB-231 cells were transfected with different gRNA vectors (Foxc1, Nfib, Nfe2l3 gRNAs) as described previously. A) MDA-MB-231 + Foxc1 gRNA clone showing the percentage of positive cells for mCherry signal (X-axis), EGFP signal (Y-axis), and for both signals after each transfection. B) and C) represent the FACS plot analyses for MDA-MB-231 + Foxc1, MDA-MB-231 + Nfib gRNA and MDA-MB-231 + Nfe2l3 gRNA clones, respectively.

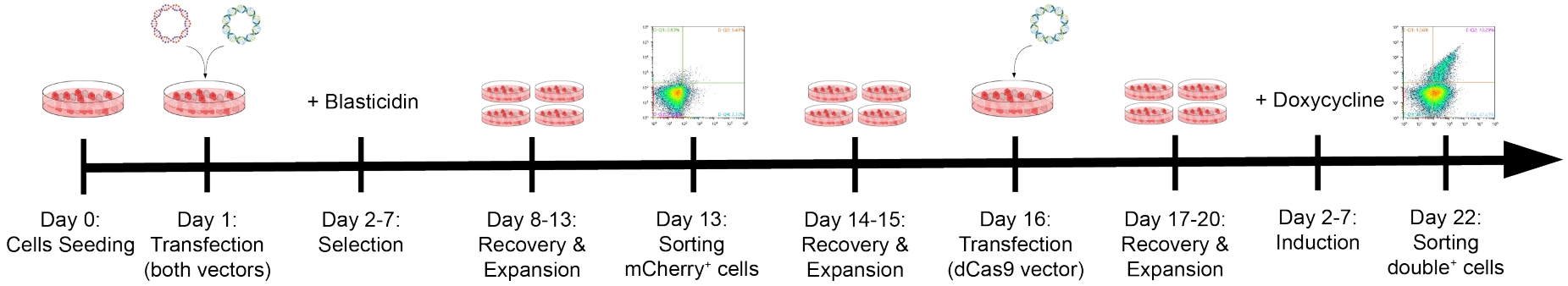
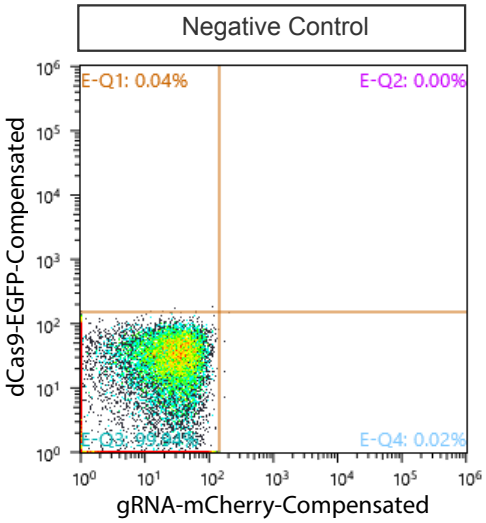
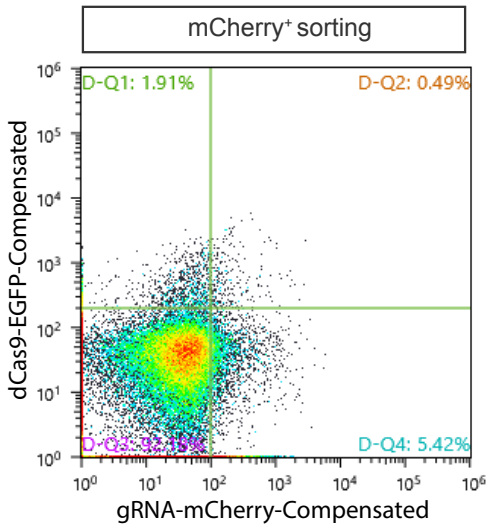


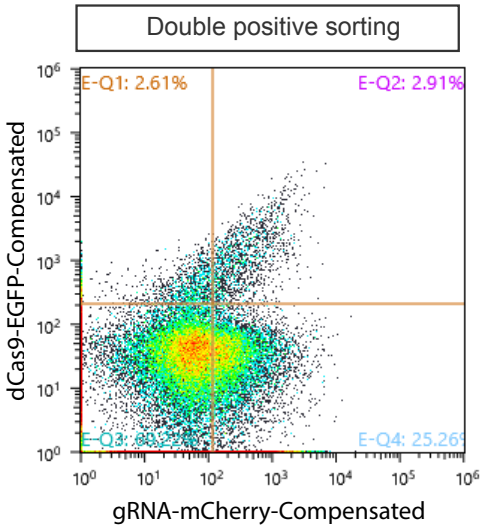

**Figure 3.9: Overall strategy of transfection and selection for clone preparation.** MDA-MB-231 cell line was seeded at day 0. At day 1, cells were transfected with both gRNA (mCherry selection marker) and dCas9 (EGFP selection marker) vectors. 24 hours later, the selection through Blastocidin started and carried on for five consecutive days. Subsequently, cells were expanded for few days until day 13, when just the mCherry$^{+}$ cells were sorted and seeded again for recovery. Three days later, they were transfected again just with the dCas9 vector. These cells were then left growing few days before inducing them for 24 hours with Doxycycline in order to express dCas9. Cells expressing both mCherry and EGFP (double positive cells) were sorted and used for further experiments.

A) MDA-MB-231

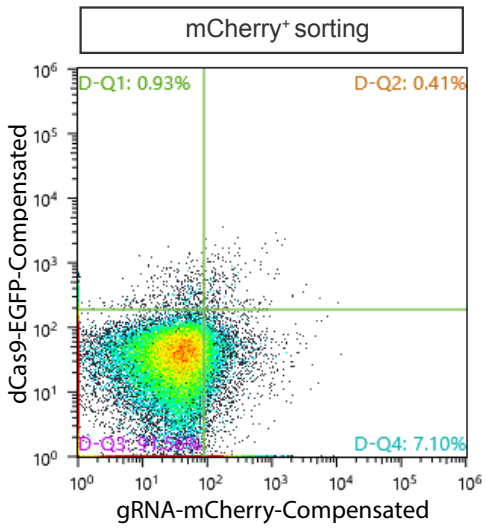


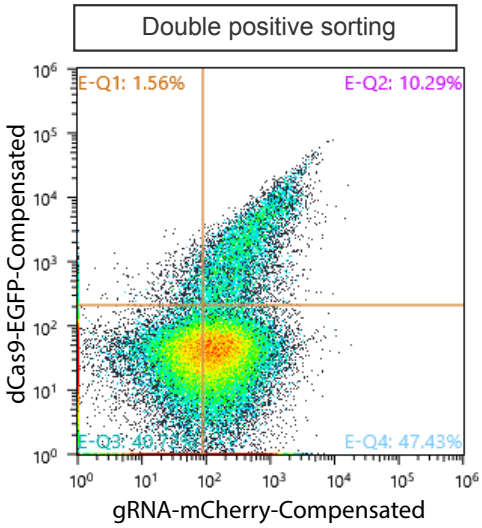

B) MDA-MB-231 + Empty gRNA



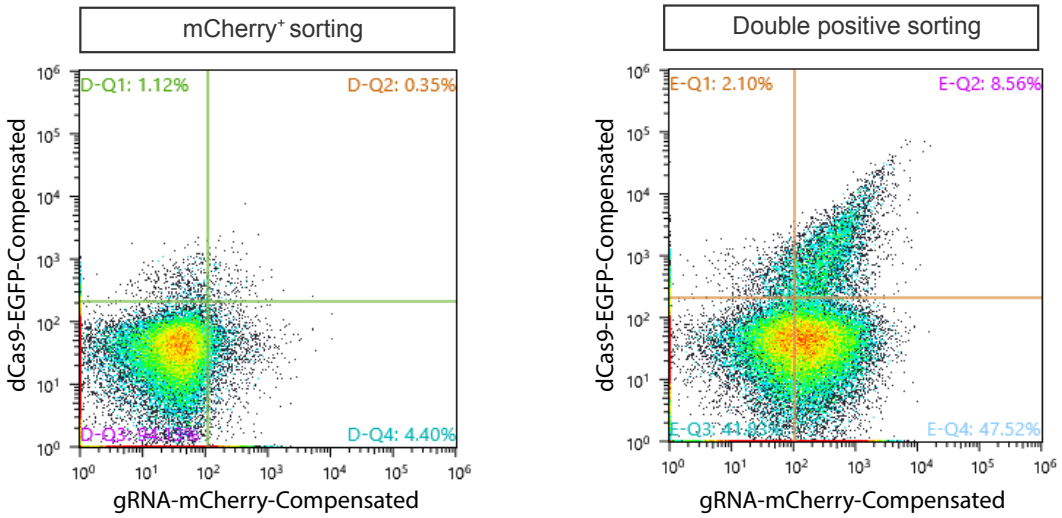




C) MDA-MB-231 + Foxc1 gRNA



D) MDA-MB-231 + Nfib gRNA




E) MDA-MB-231 + Nfe2l3 gRNA

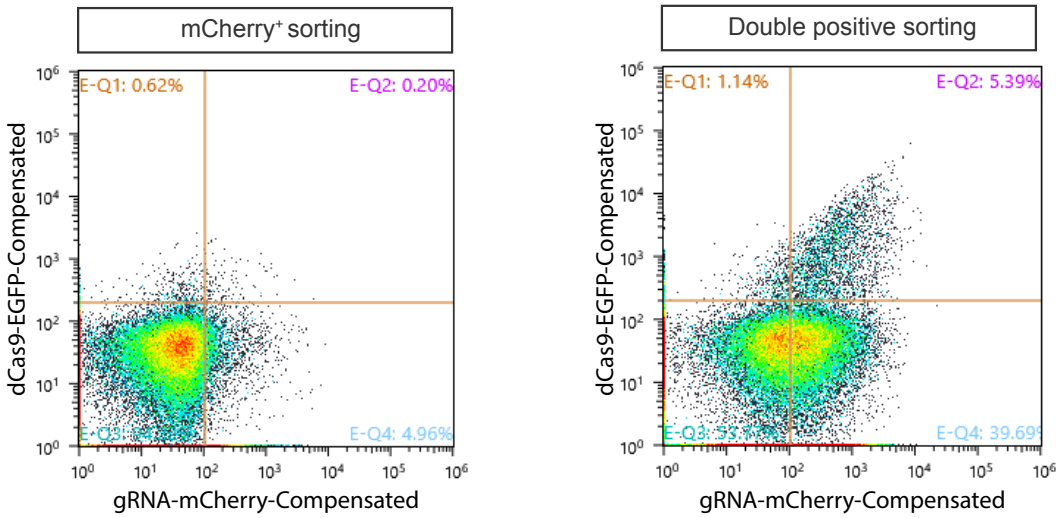



**Figure 3.10: Sorting strategies of the MDA-MB-231 clones.** After the first transfection with both mCherry-gRNA and EGFP-dCas9 vectors, just the mCherry$^{+}$ cells were sorted and expanded in culture. Subsequently, cells were transfected again with just the EGFP-dCas9 vector, induced with Doxycycline for 24 hours, and the double positive (mCherry$^{+}$ and EGFP$^{+}$) cells were sorted. MDA-MB-231 cells were transfected with different gRNA vectors (Empty, Foxc1, Nfib, Nfe2l3 gRNAs) as described previously. A) FACS plot analysis for untrasfected MDA-MB-231, used as a Negative Control. B) MDA-MB-231 + Empty gRNA clone showing the percentage of positive cells for mCherry signal (X-axis), EGFP signal (Y-axis), and for both signals after each transfection. C), D), E) and F) represent the FACS plot analyses for MDA-MB-231 + Foxc1 gRNA, MDA-MB-231 + Nfib gRNA and MDA-MB-231 + Nfe2l3 gRNA clones, respectively.

In order to determine the optimal conditions for inducing sufficient expression levels of dCas9, a time course pilot experiment was performed on MDA-MB-231 + Nfe2l3 gRNA cell line with different time points induction with Doxycycline. dCas9 was evaluated at the protein level (Fig. 3.11, A) and at the DNA binding side on the promoter sequence of the gene of interest (Fig. 3.11, B). We reported results for dCas9 expression after 6, 12, 24, 36 and 48 hours of induction.

It is possible to observe how the expression of dCas9 positively correlates with the duration of the Doxycycline induction: the longer it is, the higher the protein level. On the basis of these results, we decided to use 48 hours as the duration of induction for all further experiments.

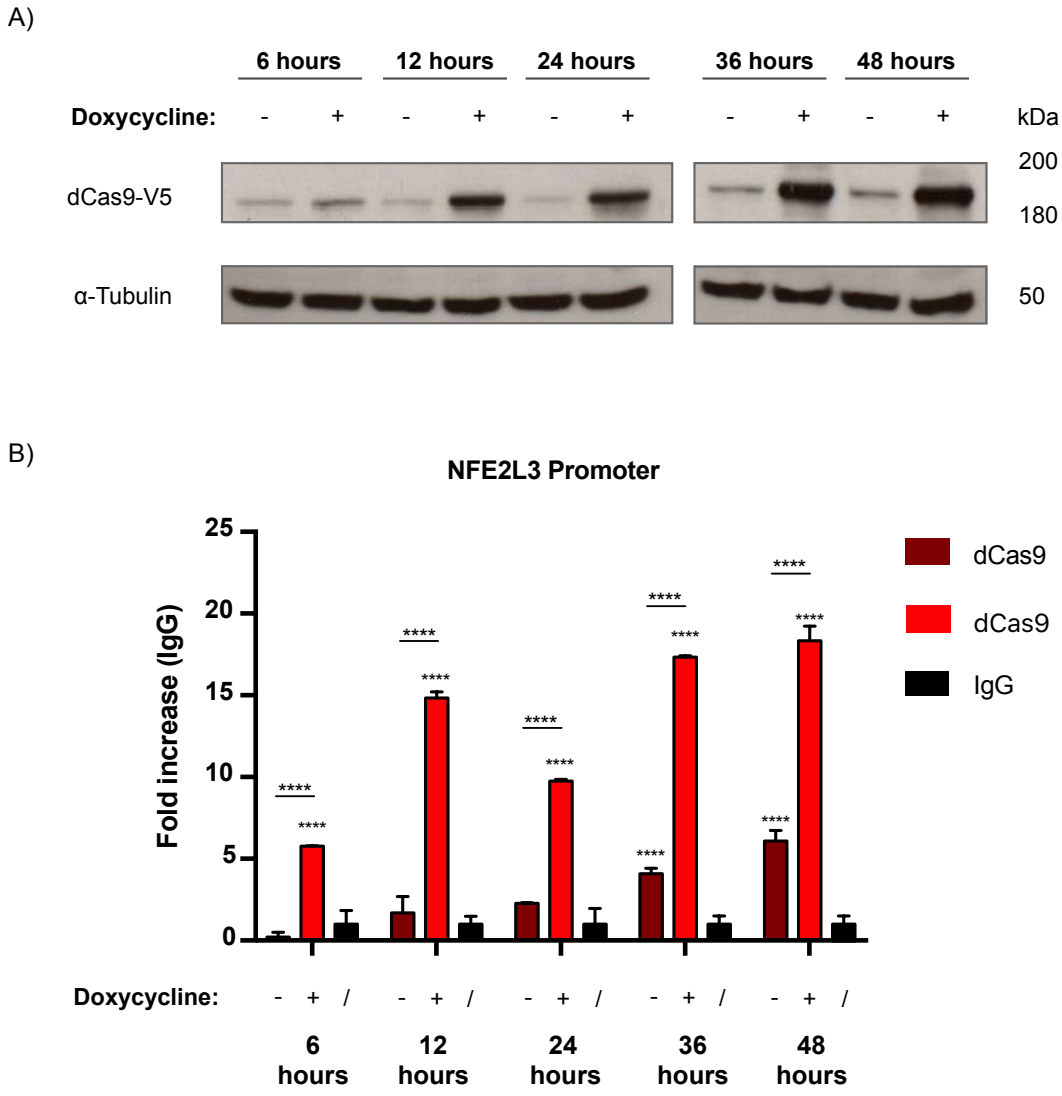



**Figure 3.11: dCas9 expression and DNA binding time course after Doxycycline induction.** MDA-MB-231 + Nfe2l3 gRNA clone was used as a representative example. Cells were induced for 6, 12, 24, 36 and 48 hours with Doxycycline (1µg/mL). A) dCas9 protein expression time course. At indicated times, cells were lysed and 50µg of protein lysates were probed by Western Blot for the expression of dCas9 and α-Tubulin (loading control). Not induced cells were collected at the same time points as a background control. B) dCas9 ChIP-qPCR time course on *NFE2L3* promoter. At indicated times, cells were crosslinked with formaldehyde and collected as described in Material and Methods to perform ChIP-qPCR. Panel shows the DNA enrichment of the dCas9 pulldowns at different time points evaluated in comparison to the respective internal IgG control, and to the not induced dCas9 pulldown. Primers for qPCR were designed in a region of 120bp flanking the gRNA targeting sequence. Two-way ANOVA test was performed between not induced and induced dCas9 and IgG, and between themselves. P value <0,05.

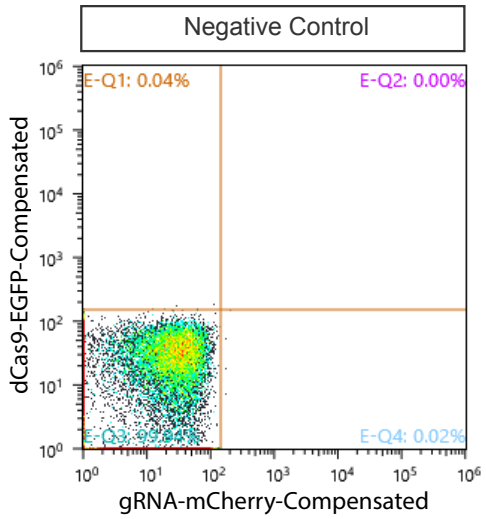
On the basis of these results, we decided to sort the double positive cells 48 hours after the addition of Doxycycline: in Fig. 3.12 the facs plots of every population of cells transfected with different gRNAs are shown, and untransfected MDA-MB-231 cells were used as a negative control.

It is interesting to notice that mCherry negative cells could be detected in sorted-mCherry$^{+}$ cells kept in culture for 7 days. This could have happened for two reasons: possible sorting error, or non-stringent sorting gate settings, or silencing of the vector which then leaded to downregulation of its expression.

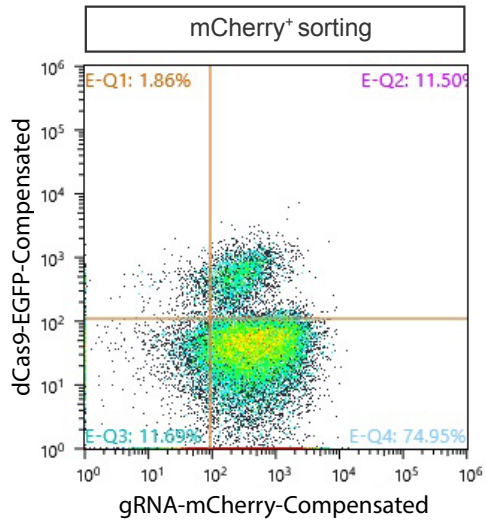
Other three strategies were tested to increase the percentage of double positive cells, all combining transfection of different DNA and sorting. Briefly, we tried to avoid a second transfection and we evaluated if just the sorting for the mCherry positive cells would have increased the percentage of double positive cells, but the results were similar to our first experiment. Because of the instability of the expression of the gRNA vector we noticed after the first sorting, we also attempted a second transfection with both vectors before sorting for the double positive. Finally we tried to deliver a single vector at the time, starting from the gRNA one, to finish with the dCas9 one just for the mCherry positive sorted cells. All these experiments worked, but the efficiency was not as good as the strategy chosen (data not shown).

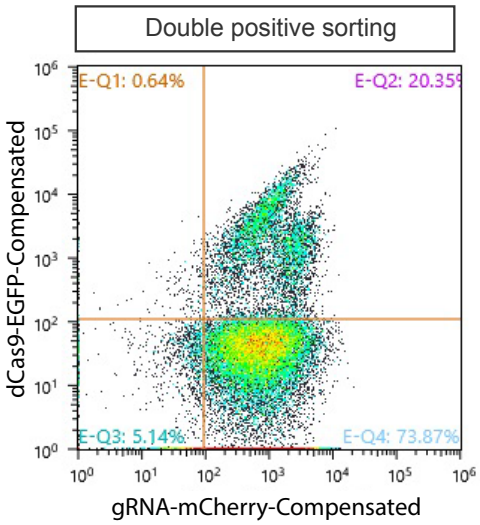
To further validate the expression of the gRNA and dCas9 vectors, we visualized the mCherry and EGFP signals: Fig. 3.13, 3.14, 3.15 and 3.16 show in particular the difference in dCas9 expression before and after induction of the double positive cells. The microscopy analysis confirmed that the vectors were stably integrated in the cells, and the inducible system worked as expected for all the transfected cell lines.

A) MDA-MB-231

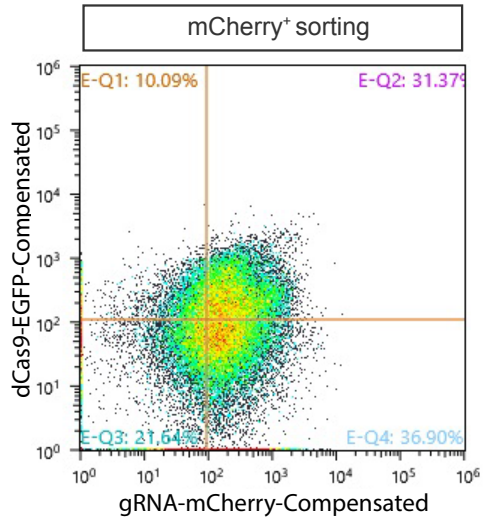


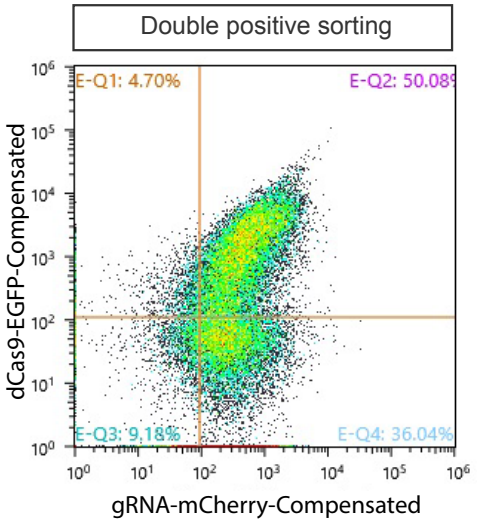

B) MDA-MB-231 + Empty gRNA



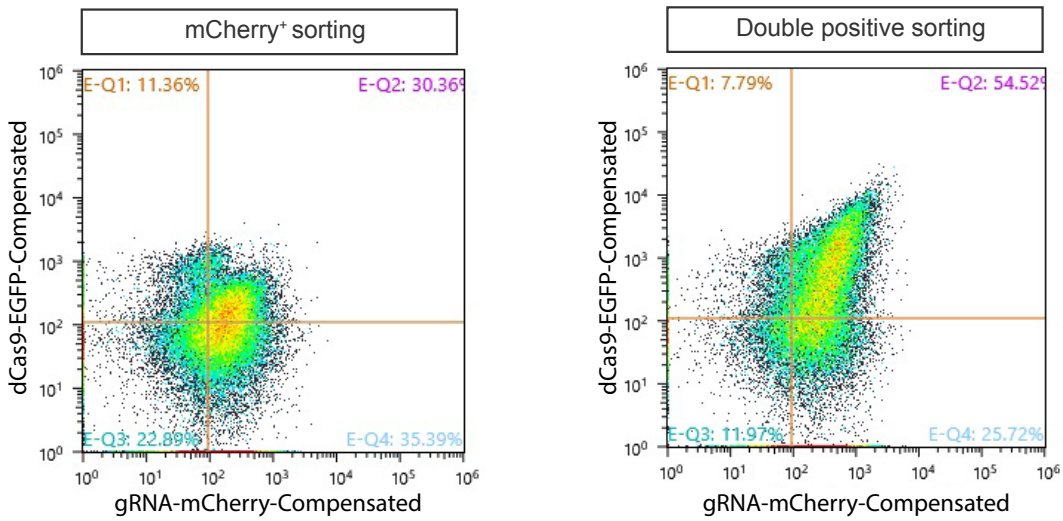




C) MDA-MB-231 + Foxc1 gRNA



D) MDA-MB-231 + Nfib gRNA




E) MDA-MB-231 + Nfe2l3 gRNA

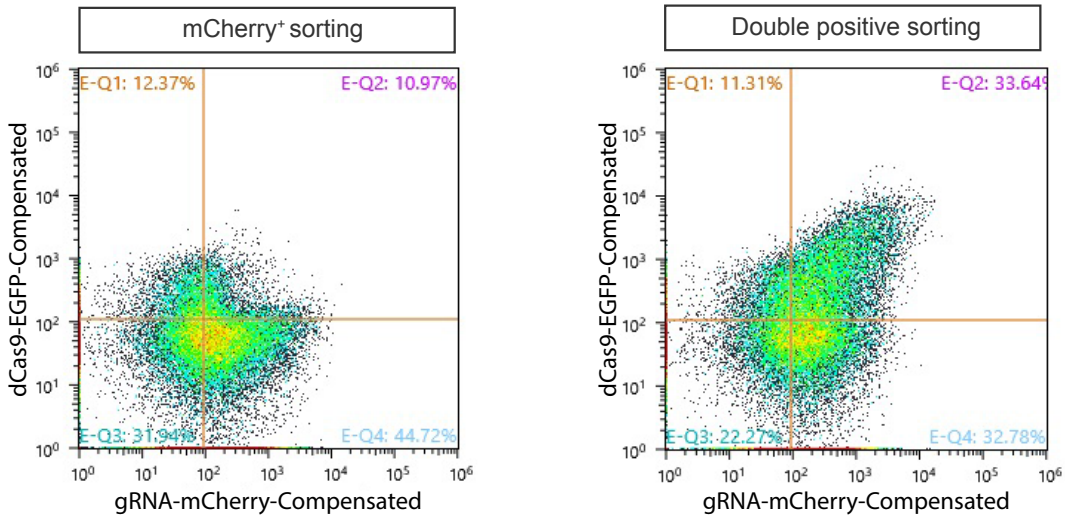



**Figure 3.12: Sorting strategy of the MDA-MB-231 clones after 48 hours induction.** After the first transfection with both mCherry-gRNA and EGFP-dCas9 vectors, just the mCherry$^+$ cells were sorted and expanded in culture. Subsequently, cells were transfected again with just the EGFP-dCas9 vector, induced with Doxycycline for 48 hours, and the double positive (mCherry$^+$ and EGFP$^+$) cells were sorted. MDA-MB-231 cells were transfected with different gRNA vectors (Empty, Foxc1, Nfib, Nfe2l3 gRNAs) as described previously. A) FACS plot analysis for untrasfected MDA-MB-231, used as a Negative Control. B) MDA-MB-231 + Empty gRNA clone showing the percentage of positive cells for mCherry signal (X-axis), EGFP signal (Y-axis), and for both signals after each transfection. C), D), E) and F) represent the FACS plot analyses for MDA-MB-231 + Foxc1 gRNA, MDA-MB-231 + Nfib gRNA and MDA-MB-231 + Nfe2l3 gRNA clones, respectively.

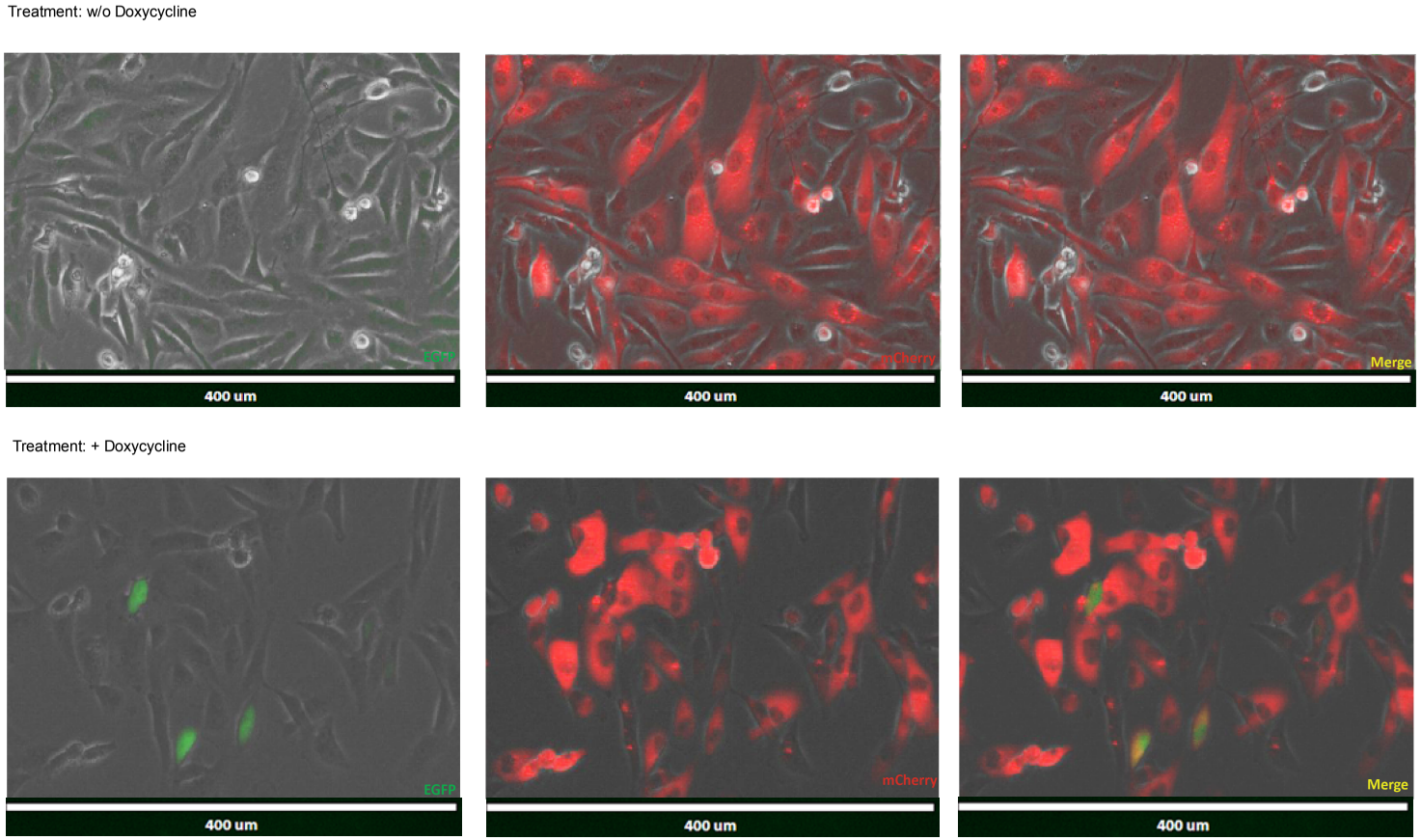



**Figure 3.13: Representative fluorescence microscopy of the expression of the vectors in MDA-MB-231 + Empty gRNA clone without and with Doxycycline treatment.** Cells were transfected with different vectors as described previously. After recovery, the double positive cells (transfected with both mCherry-gRNA and EGFP-dCas9 vectors) were seeded and treated with or w/o Doxycycline for 48 hours. They were then analysed to confirm the expression of both vectors. Far left panel: EGFP fluorescent signal. Middle panel: mCherry fluorescent signal. Far right panel: merge of both signals.

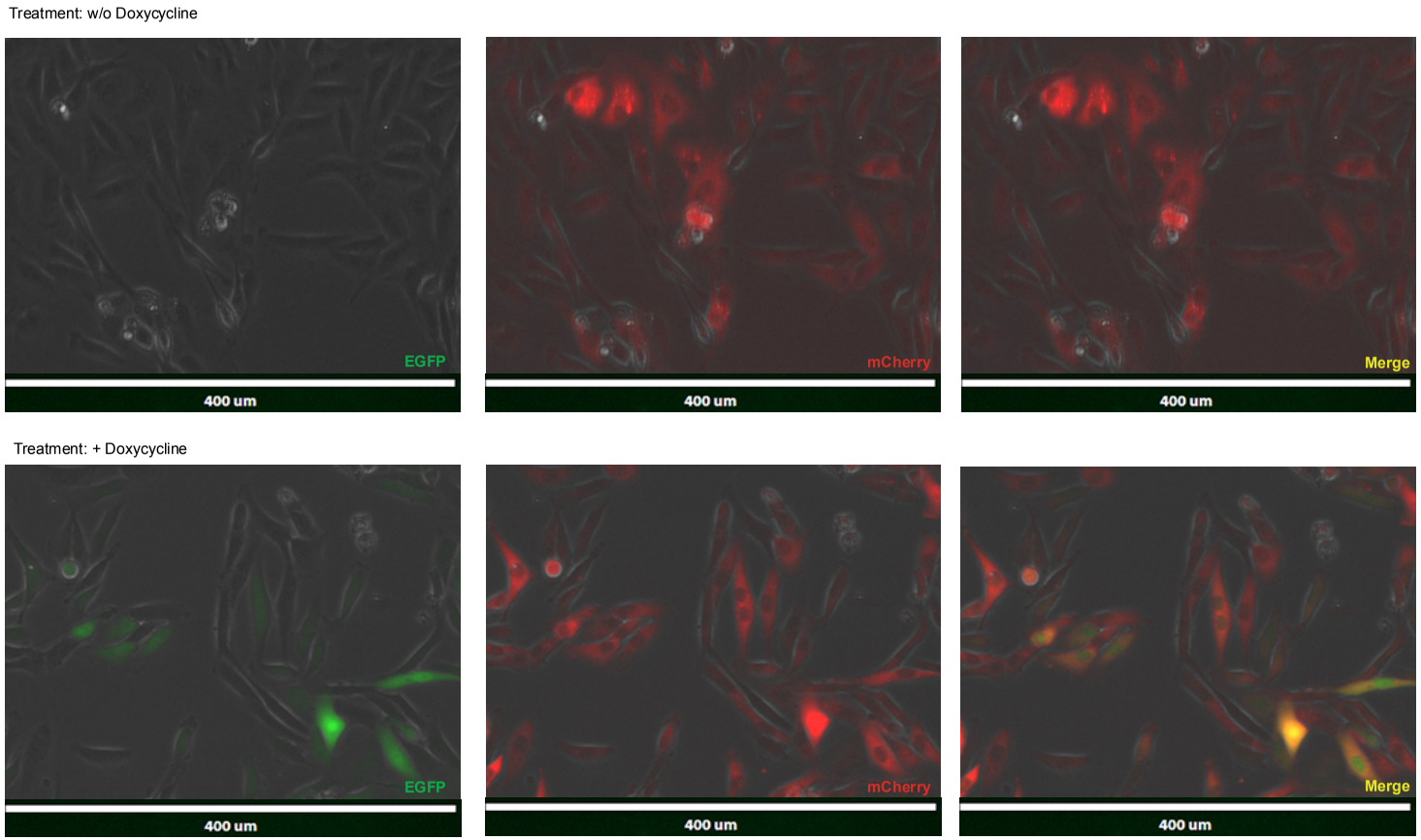



**Figure 3.14: Representative fluorescence microscopy of the expression of the vectors in MDA-MB-231 + Foxc1 gRNA clone without and with Doxycycline treatment.** Cells were transfected with different vectors as described previously. After recovery, the double positive cells (transfected with both mCherry-gRNA and EGFP-dCas9 vectors) were seeded and treated with or w/o Doxycycline for 48 hours. They were then analysed to confirm the expression of both vectors. Far left panel: EGFP fluorescent signal. Middle panel: mCherry fluorescent signal. Far right panel: merge of both signals.

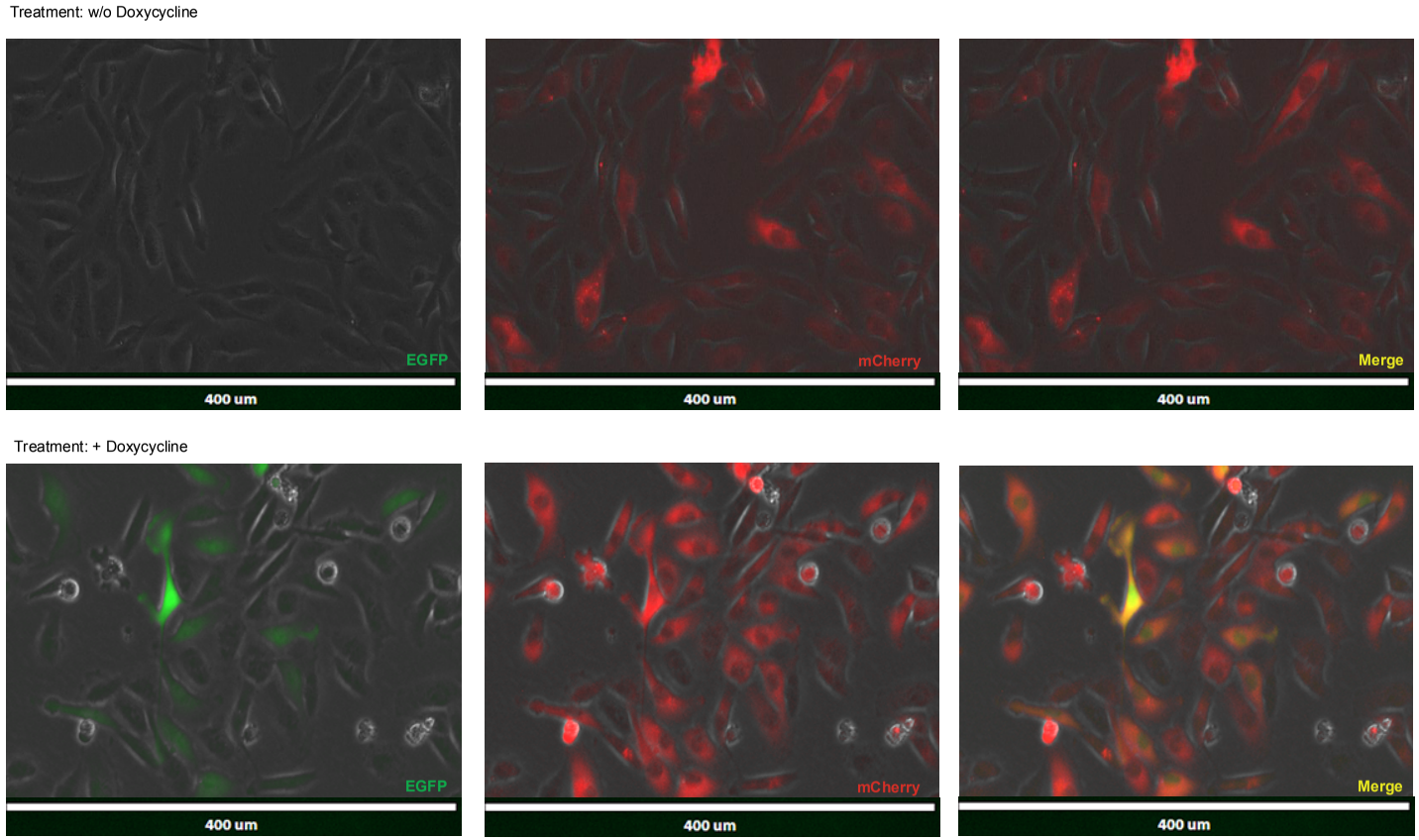



**Figure 3.15: Representative fluorescence microscopy of the expression of the vectors in MDA-MB-231 + Nfib gRNA clone without and with Doxycycline treatment.** Cells were transfected with different vectors as described previously. After recovery, the double positive cells (transfected with both mCherry-gRNA and EGFP-dCas9 vectors) were seeded and treated with or w/o Doxycycline for 48 hours. They were then analysed to confirm the expression of both vectors. Far left panel: EGFP fluorescent signal. Middle panel: mCherry fluorescent signal. Far right panel: merge of both signals.

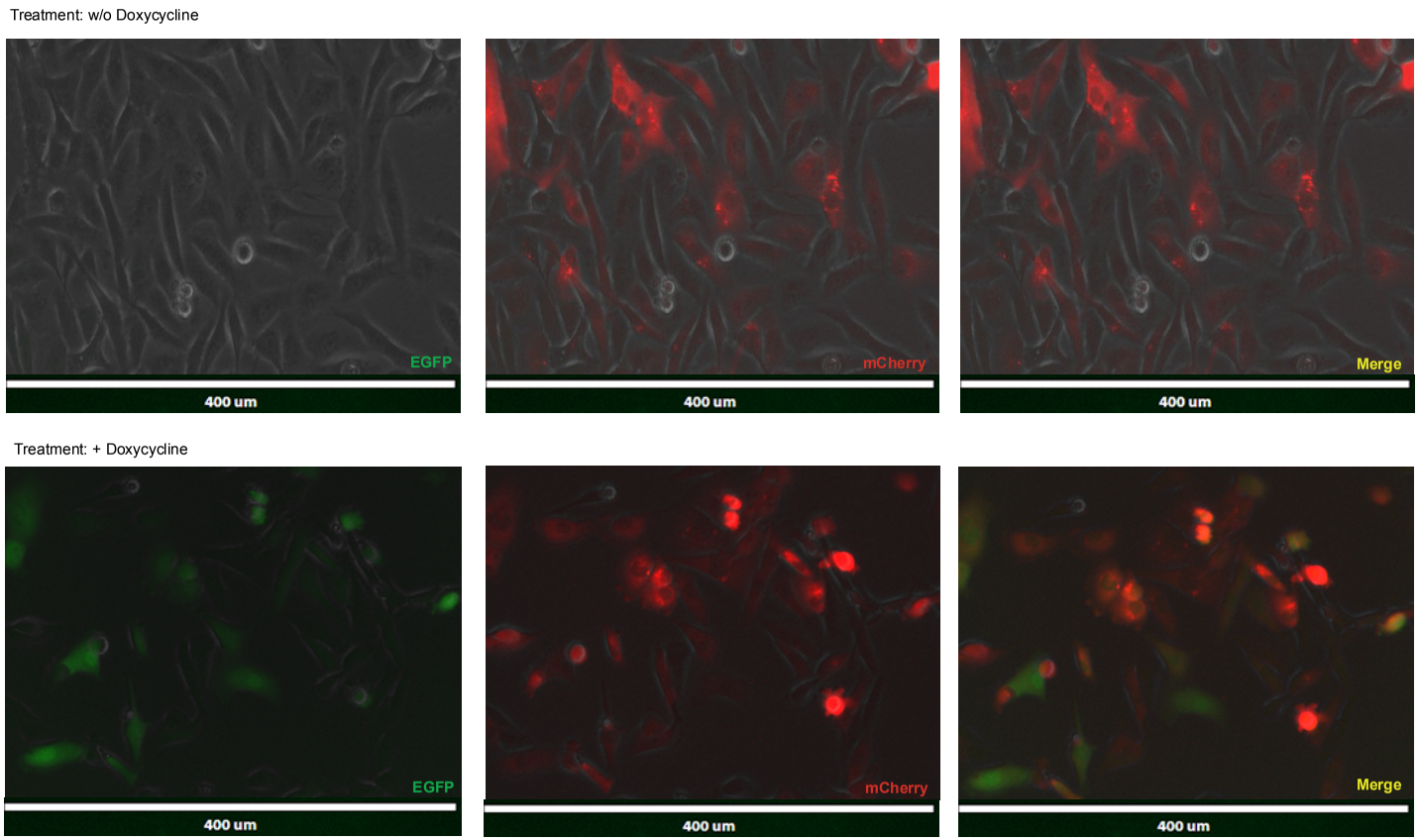



**Figure 3.16: Representative fluorescence microscopy of the expression of the vectors in MDA-MB-231 + Nfe2l3 gRNA clone without and with Doxycycline treatment.** Cells were transfected with different vectors as described previously. After recovery, the double positive cells (transfected with both mCherry-gRNA and EGFP-dCas9 vectors) were seeded and treated with or w/o Doxycycline for 48 hours. They were then analysed to confirm the expression of both vectors. Far left panel: EGFP fluorescent signal. Middle panel: mCherry fluorescent signal. Far right panel: merge of both signals.

To ensure the purity of the population and the stable integration of both vectors after a couple of weeks in culture, transfected cell lines were analysed by flow cytometry, following 48 hours induction with Doxycycline. Data are shown in Fig. 3.17.

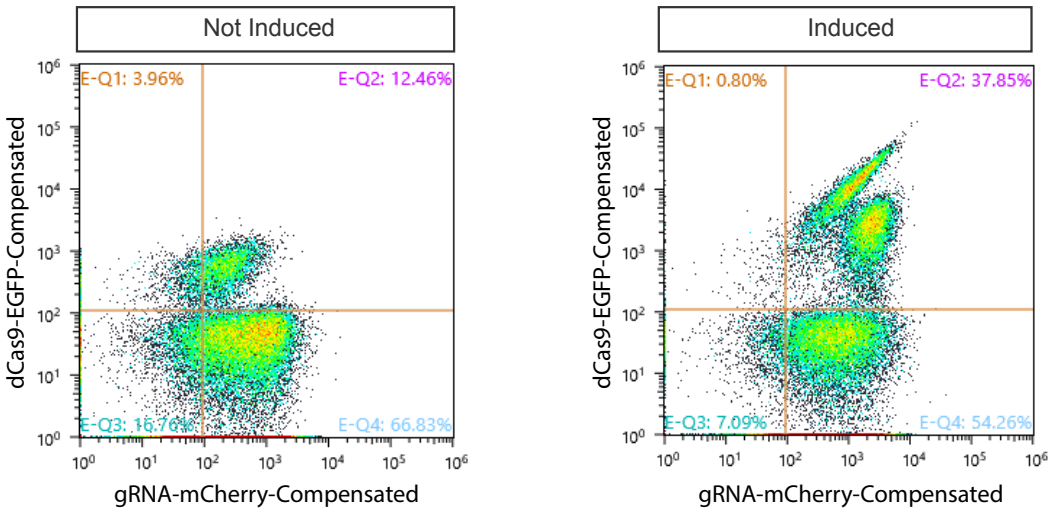
The data presented clearly show that the expression of dCas9 was significantly induced 48 hours after the addition of Doxycycline for all the transfected cell lines. Some background expression of dCas9 was detectable in cells without the Doxycycline induction: this is likely to be caused by the presence of Tetracycline derivatives in the FBS used to culture these cells. Despite this, it is evident that the percentage of double positive cells increased after induction (Fig. 3.17, E). On the basis of this information, we decided to use a tretracycline-free FBS for any further sensitive experiments.

Looking at the percentage of mCherry and EGFP positive cells after induction, it is possible to see how it varies within the samples, from a minimum of 22% (MDA-MB-231 + Foxc1 gRNA), to a maximum of 66% (MDA-MB-231 + Nfib gRNA). We believe that this difference is due to the reaction of the cells to the vector itself: for example they could shut down the expression in the area of the genome where the vector has been integrated, or the region itself where it was located might not be particularly active. In none of the sample we saw 100% of cells expressing EGFP. This was expected considering the fact that the cells were not synchronized, so they could have been in different phases of the cell cycles, therefore reflecting different levels of gene expression in different phases.

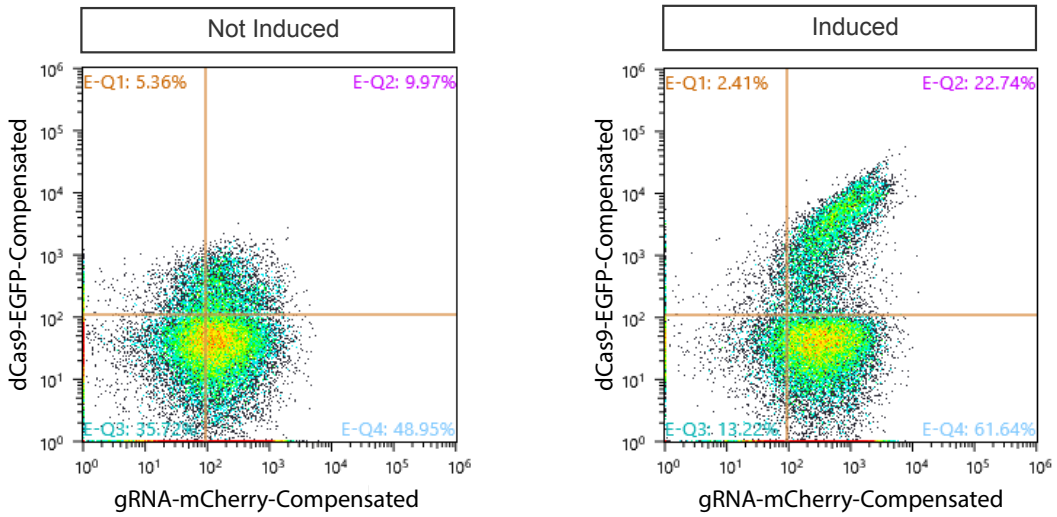
To further confirm the expression of dCas9 in our system, we also performed a Western Blot analysis on total protein lysate of cells induced with Doxycycline for 48 hours (Fig. 3.18). The level of dCas9 was compared between induced and not-induced cells, and $\alpha$-Tubulin was used as a loading control.

The data corroborated what we observed with the FACS analysis: there is a sustainable induction of dCas9 protein expression after treatment with Doxycycline. Although not quantified, it is possible to appreciate the increase of the signal in every sample.

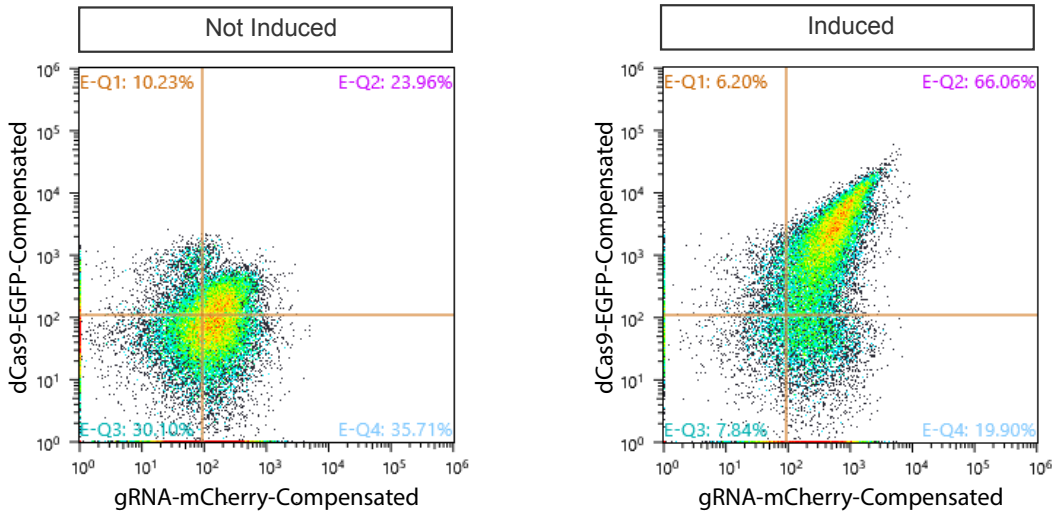
A) MDA-MB-231 + Empty gRNA



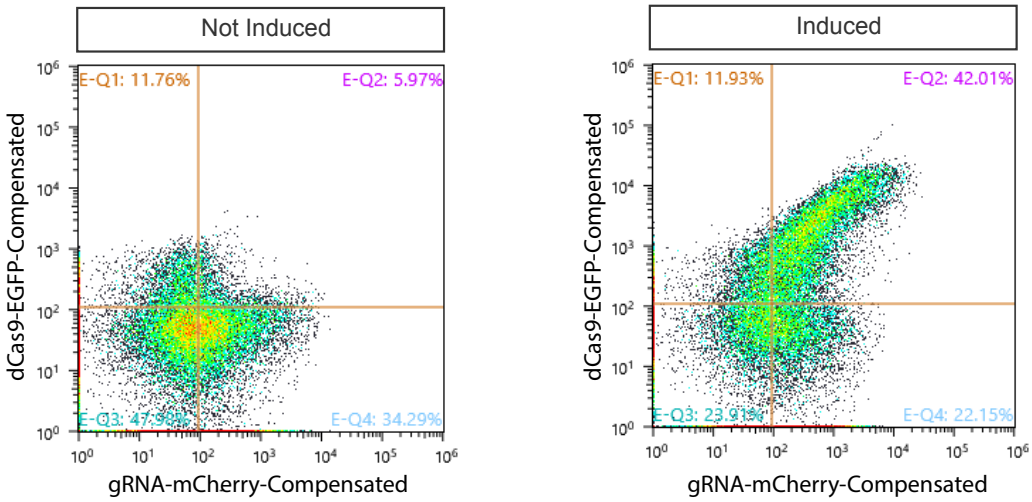

B) MDA-MB-231 + Foxc1 gRNA




C) MDA-MB-231 + Nfib gRNA



D) MDA-MB-231 + Nfe2l3 gRNA




E)

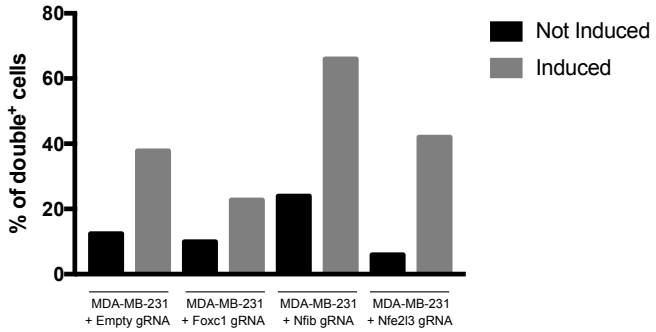



**Figure 3.17: Flow cytometry analysis of MDA-MB-231 clones before and after 48 hours induction with Doxycycline.** MDA-MB-231 cells were transfected with different gRNA vectors (Empty, Foxc1, Nfib, Nfe2l3 gRNAs) as described previously. Cells containing both gRNA (mCherry$^{+}$) and dCas9 (EGFP$^{+}$) vectors (double positive cells) were sorted and expanded in culture for 2 weeks. They were then analysed by flow cytometry in order to assess the purity of the population. A) FACS plot analysis for MDA-MB-231 + Empty gRNA clone showing the percentage of positive cells for mCherry signal (X-axis), EGFP signal (Y axis), and for both signals without (left panel) or with (right panel) Doxycycline induction. B), C) and D) represent the FACS plot analyses for MDA-MB-231 + Foxc1 gRNA, MDA-MB-231 + Nfib gRNA and MDA-MB-231 + Nfe2l3 gRNA clones respectively. E) Percentage of double$^{+}$ cells for each clone with (grey) or without (black) induction.

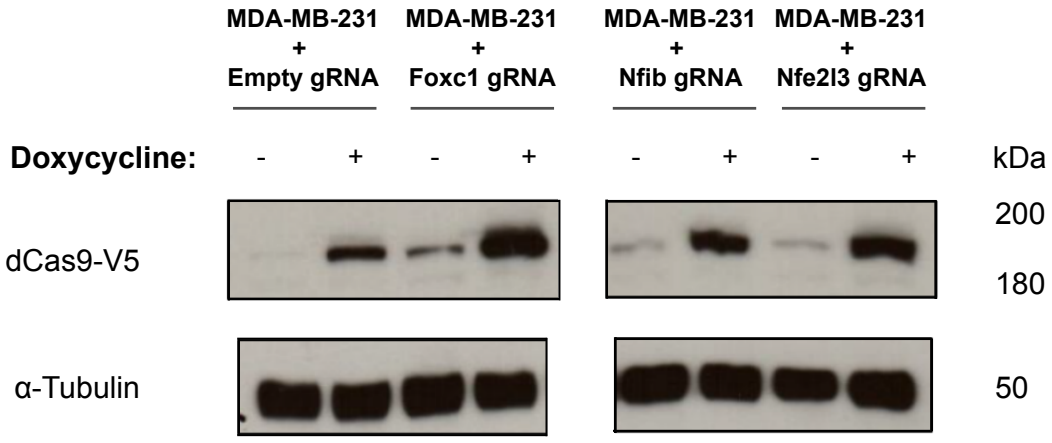



**Figure 3.18: dCas9 expression in different MDA-MB-231 clones after induction with Doxycycline.** The MDA-MB-231 + Empty gRNA, + Foxc1 gRNA, + Nfib gRNA and + Nfe2l3 gRNA clones were induced for 48 hours with Doxycycline. Cells were then lysed and 50µg of protein lysate were probed by WB for the expression of dCas9 and α-Tubulin (loading control). Not induced cells were collected at the same time points as a background control.

## 3.5 Effect of dCas9 on expression of target genes

To ensure that dCas9 was not interfering with the basal expression of the target genes, we analysed their expression in our transfected cell lines before and after Doxycycline induction for 48 hours, and we compared it with MDA-MB-231 + Empty gRNA treated in the same way (Fig. 3.19).

For all clones, it seems like Doxycycline is causing a slight change in gene expression. Looking at the MDA-MB-231 + Empty gRNA control, it is possible to see how all the genes are downregulated compared to their basal expression. On the other hand, the treatment seems to have different effects on the other clones transfected with the respective gRNA: in particular for MDA-MB-231 + Foxc1 gRNA and MDA-MB-231 + Nfib gRNA the expression of the targeted genes seems to increase. This is probably related to the fact that the dCas9 is interacting with the promoter region of those genes and interfering with their expression, possibly masking some other proteins' binding sites. MDA-MB-231 + Nfe2l3 gRNA is the only clone with a similar trend compared to the control, which could be explained by the fact that dCas9 is not masking any important region within the promoter of the gene.

Despite the effect of the presence of dCas9 on basal transcription of the targeted genes, we decided to proceed with further experiments, since the cells treated with Doxycycline were immediately collected after 48 hours and used for investigations.

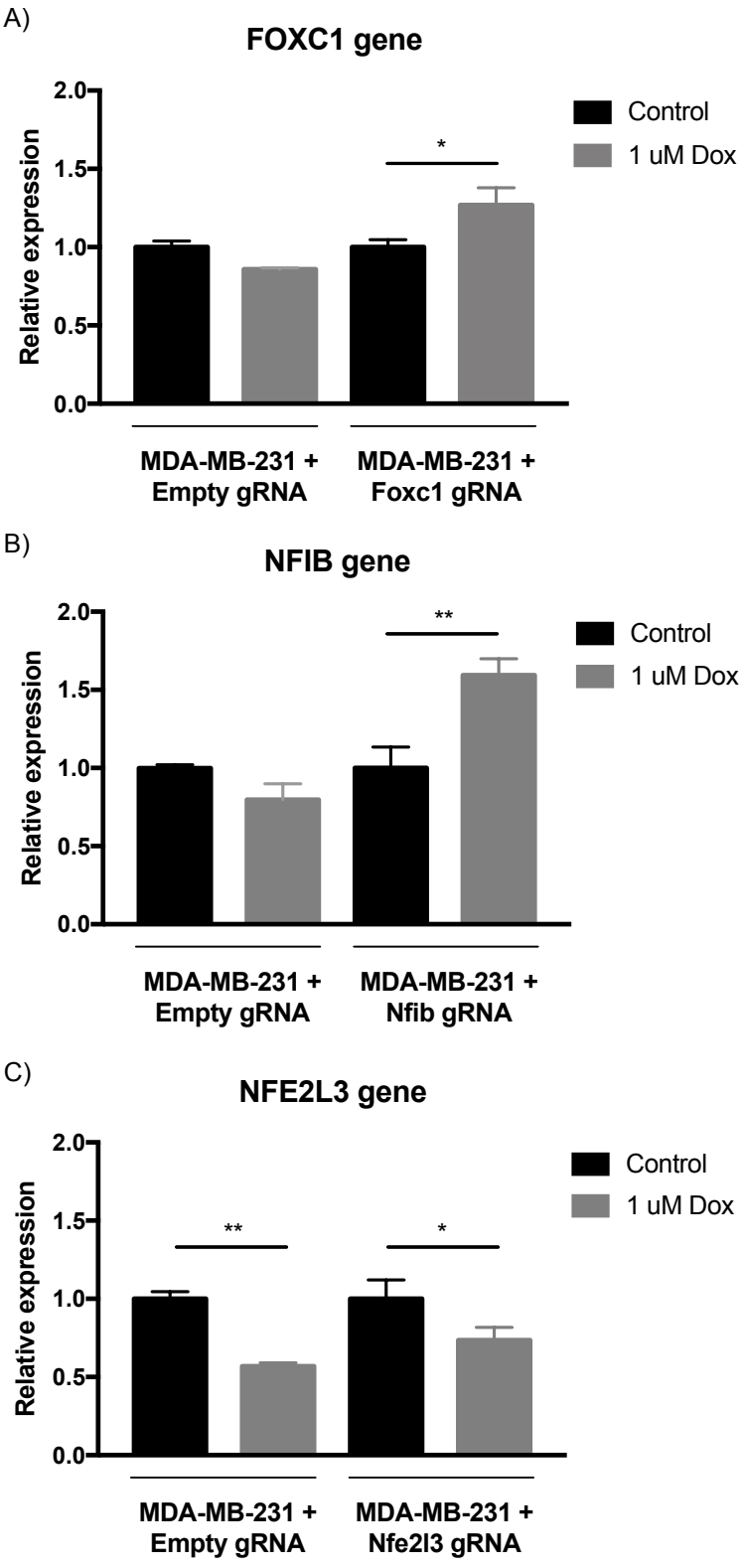



**Figure 3.19: Effect of dCas9 induction on the expression of genes of interest in MDA-MB-231 clones.** Expression of dCas9 was induced for 48 hours with Doxycycline (1µg/mL). Relative levels of *FOXC1* (A), *NFIB* (B) and *NFE2L3* (C) gene expression were quantified by qPCR and normalized to the respective *GAPDH*. The data are ratios with the respective control (untreated clone) and are reported as the mean ± S.D. of 3 technical replicates. Two-way ANOVA test was performed, P value <0,05.

To ensure the dCas9 binding to the desired sequence of DNA, ChIP (Chromatin Immuno-precipitation)–qPCR and ChIP-Seq (ChIP-Sequencing) experiments were performed. In order to do that, we induced every transfected cell line for 48 hours and compared the result with the respective not-induced control.

For the ChIP-qPCR, we amplified the DNA sequence that contains the gRNA targeting site, using primers available on NCBI Primer-BLAST (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). The results (Fig. 3.20) show the correct localization of dCas9 on the promoter sequence for all genes of interest (A) *FOXC1*, B) *NFIB*, C) *NFE2L3*), with a further confirmation for *FOXC1* by ChIP-Seq (Fig. 3.21). The enrichment of the DNA binding for both experiments was evaluated in comparison to the background binding, when dCas9 was not induced. To minimize the uncontrolled expression of dCas9, we maintained the cells in a media with a tetracycline-free FBS: in this way we wanted to reduce the perturbation of the system due to the presence of dCas9.

From the ChIP-Seq analysis it is possible to confirm that the binding of the dCas9 is not unique, which means that it also interacts with other different sequences of DNA (off-targets, Fig. 3.21, B). However, the uncontrolled DNA binding of dCas9 is extremely low, which further support the evidence of a limited interference of the used system with the overall gene transcription.

These results were extremely important because they corroborate the successful targeting of dCas9 to the region of interest. However, the off-target bindings we saw informed us that RIME on dCas9 could lead to misinterpretation: some proteins could be crosslinked to dCas9 but involved in other processes in other regions of the DNA. Taking this information into consideration, we decided to proceed with our proteomic approach, and to filter the potential candidates on the basis of their biological relevance and confidence of their identification at the mass-spectrometry level.

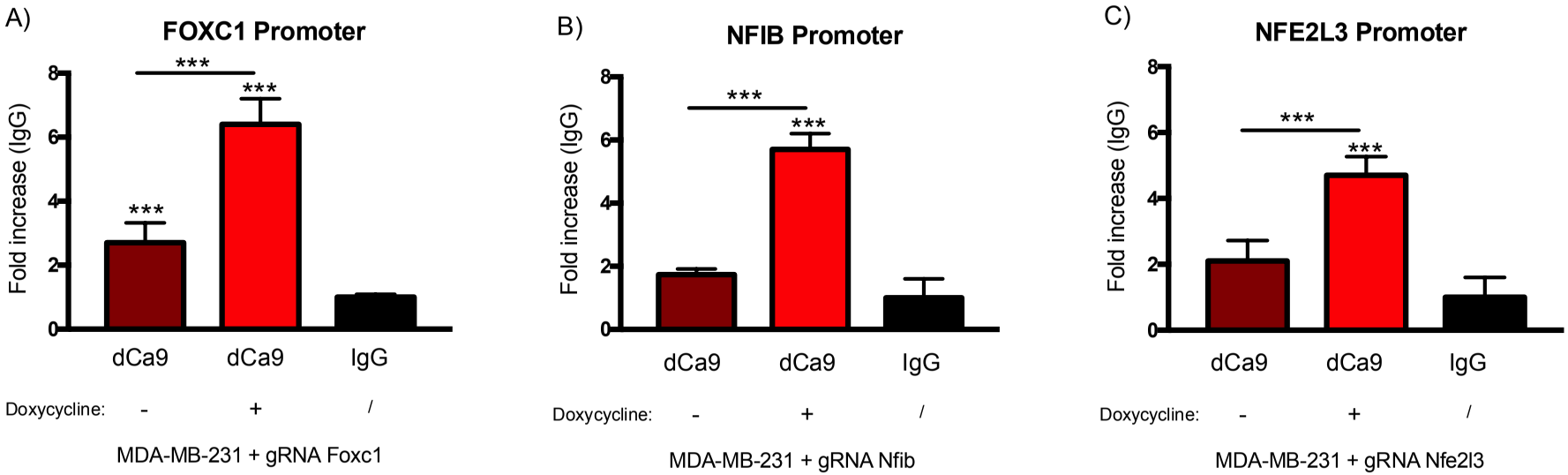



**Figure 3.20: ChIP-qPCR confirmation of dCas9 binding at the putative promoter sequence of the genes of interest.** MDA-MB-231 + Foxc1 gRNA (A), MDA-MB-231 + Nfib gRNA (B) and MDA-MB-231 + Nfe2l3 gRNA (C) clones were seeded and induced for 48 hours with Doxycycline. Cells were then crosslinked with formaldehyde and collected as described in Material and Methods to perform ChIP-qPCR. DNA enrichment of the dCas9 pulldown for every clone was evaluated in comparison to the respective internal IgG control. Primers for qPCR were designed in a region of 120bp flanking the gRNA of each respective promoter. Two-way ANOVA test was performed between not induced and induced dCas9 and IgG, and between themselves. P value <0,05.

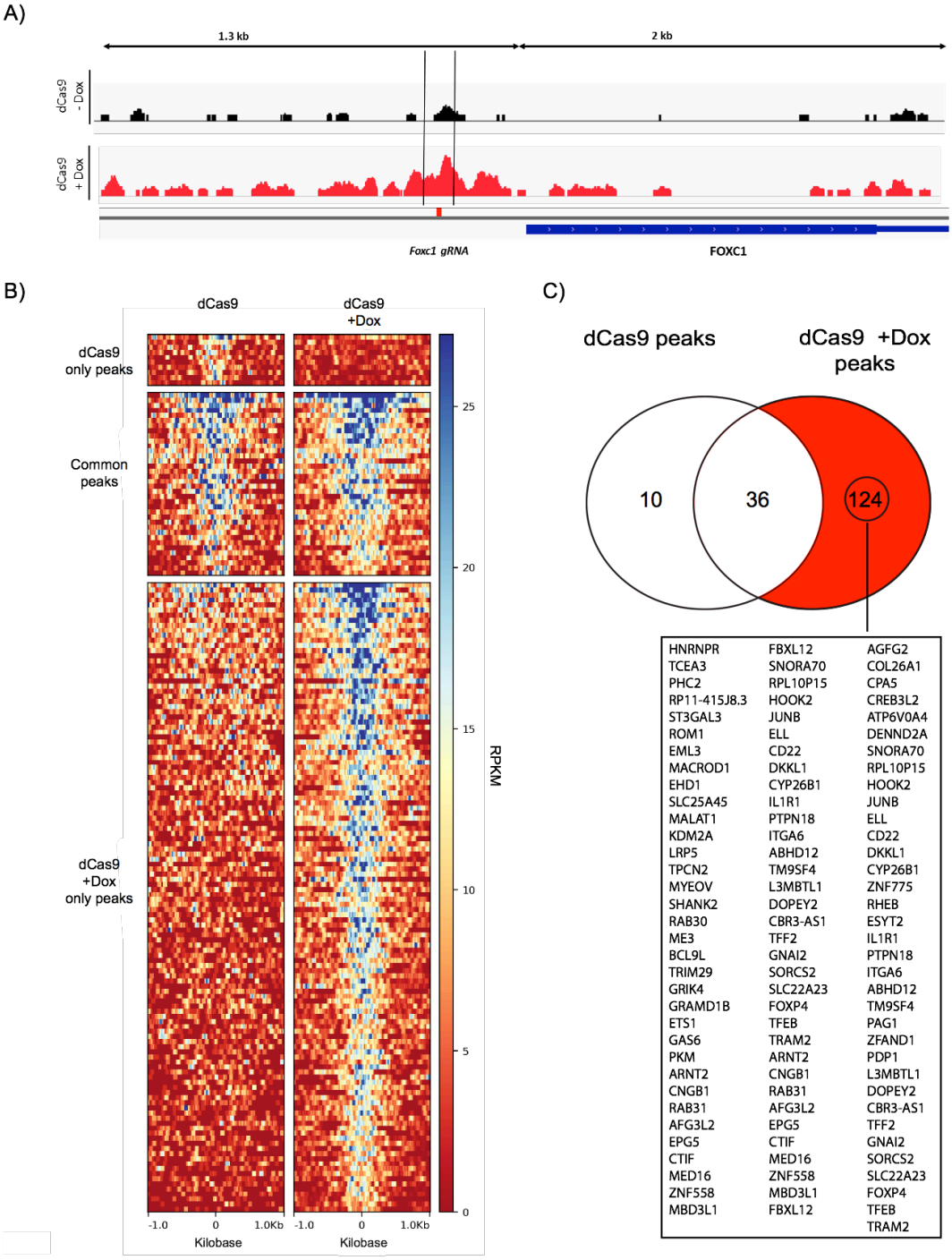



**Figure 3.21: ChIP-Seq confirmation of dCas9 binding on the putative promoter sequence of *FOXC1* gene.** MDA-MB-231 + Foxc1 gRNA clone was seeded and induced for 48 hours with Doxycycline. Cells were then crosslinked with formaldehyde and collected as described in Material and Methods to perform ChIP-Seq. A) Visualisation of differentially accessible peaks annotated to *FOXC1* in IGV after dCas9 not induced and dCas9 induced ChIP-Seq data analyses using MACS for peaks calling. B) Heat maps showing dCas9 only, dCas9 + Dox (induced) only or common peaks in dCas9 IP. C) Venn diagram indicating the unique targeted genes of dCas9 after induction. Analyses were performed by Mike Firth and Jonathan Cairns, AZ, Cambridge, UK.

## 3.6 Discussion

By investigating cancer genome data from ~2000 patients (METABRIC), we identified five highly regulated genes differentially expressed encoding for transcription factors (*FOXC1*, *ELF5*, *SOX10*, *NFIB* and *NFE2L3*), all known to have an important role in the breast, and some also in TNBC.

*FOXC1* for example promotes cancer stem cell properties by activating Smoothened (SM)-independent Hedgehog (Hh) signalling (Han et al., 2015), and can induce epithelial-mesenchymal transition (EMT) (Xia et al., 2013). It seems to play a role in TNBC's invasiveness, regulating the downstream expression of MMP7 (Han et al, 2018). Its expression positively correlates with a shorter brain and lung metastases free survival (Ray et al., 2010; Jensen et al., 2015). In addition, it seems to compete with GATA3 for binding some regions on the *ESR1* promoter, repressing the expression of this gene (Yu-Rice et al., 2016).

*ELF5*'s role in breast is well known: it regulates placentation (Donnison et al., 2005) and alveologenesis, the process during pregnancy when the mammary glands form acinar structures producing milk (Choi et al., 2009; Watson et Khaled, 2008). It also directs the differentiation of mammary progenitor cells towards the basal-like phenotype: it suppresses *ESR1* expression and a panel of other 164 genes including *FOXA1* and *GATA3* (Kalyuga et al., 2012).

*SOX10* plays important regulatory roles in promoting both stem- and EMT-like properties in mammary stem cells (Dravis et al., 2015). In particular for TNBC it induces Nestin expression (Feng et al., 2017), a stemm cell marker involved in tumour invasiveness.

On the other hand, *NFIB* activates critical MYB targets, including genes associated with apoptosis, cell cycle control, cell growth/angiogenesis and cell adhesion, forming a fusion MYB-NFIB protein (Persson et al., 2009). It also regulates the expression of genes associated with lactation such as Whey acidic protein (WAP) and α-lactalbumin (Murtagh et al., 2003).

*NFE2L3* plays a role as a transcription factor when it translocates to the nucleus in response to external stimuli (Chowdhury et al., 2017), like for example oxidant stress. In the breast, it seems to control cell proliferation activating UHMK1 (U2AF

homology motif kinase 1, a cell cycle regulator) and it has been shown to negatively correlate with metastases: its overexpression leads to inhibition of the EMT process (Sun et al., 2019).

Experimentally, we were able to target putative regulatory regions of these genes by using dCas9. This particular approach has been further developed in the last few years into a technique called enChIP (engineered DNA-binding molecule-mediated ChIP (Fujita et Fujii, 2013)), where specific genomic regions are immunoprecipitated with antibody against a tag(s) fused to an engineered DNA-binding molecule (dCas9) recognizing an endogenous DNA sequence in the genomic regions of interest. In combination with MS and NGS (next generation sequencing), it has allowed to identify novel proteins (Fujita et Fujii, 2014; Fujita et al., 2013; Hamidian et al., 2018), RNAs (Fujita et al., 2015) or genomic regions (Fujita et al., 2017) in an unbiased manner.

However, all these approaches require a deeper understanding of the off-target bindings: many of these uncontrolled binding events may be transient so insufficient for modulating transcription of nearby genes (Polstein et al., 2015), but a deeper understanding of the causes would help to develop strategies to minimize them and to improve the validation process of the results.

Furthermore, we optimized a selection strategy to obtain stable TNBC cell lines transfected with the external DNA coding for dCas9 and a gRNA, and we ensured that the system we used did not interfere with the basal expression of the targeted genes, making it an inducible one (Tet-On system). Several examples are present in the literature to support the possibility of controlling the system in a spatial and temporal manner. Polstein et al., and Nihongaki et al., for example, have shown how the expression of endogenous targeted genes can be regulated by dCas9 after illuminating cells with a blue light. They used a cryptochrome-based blue light-sensing system *CRY–CIB heterodimerizing domains* to recruit VP64 or p65AD to dCas9 to make it active in a reversible manner (Polstein et Gersbach, 2015; Nihongaki et al., 2015). But there are also examples of chemically induced dCas9, where the presence of rapamycin induces the dimerization of a split dCas9-VP64 (Zetsche et al., 2015).

We further reduced the effect of dCas9 presence on basal gene transcription inducing just those cells that were going to be collected for investigations.

## 3.7 Conclusions

Data presented in this chapter demonstrated the efficiency of targeting putative regulatory regions of genes of interest, showing a limited interference on the basal gene expression. This encouraged us to proceed with the proteomic approach in order to identify novel regulators that will be discussed in the next chapter.

# CHAPTER 4: RIME OPTIMIZATION AND STATISTICAL ANALYSIS

## 4.1 Introduction

Results presented in Chapter 3 demonstrated the ability of dCas9 to bind a specific region of the DNA within the promoter sequence of different genes of interest. They also showed how its expression was temporally regulated using a Tet-On inducible system, in order to minimize the interference of dCas9 to the basal overall transcription.

Since the aim of this study was to identify novel transcription factors involved in the regulation of the expression of *FOXC1*, *NFIB* and *NFE2L3*, we decided to use a specific proteomic approach called RIME (Rapid Immunoprecipitation Mass spectrometry of Endogenous proteins) in order to identify endogenous-interacting proteins and protein-DNA binding events (Mohammed et al., 2013; Mohammed et al. 2016; D'Santos et al., 2015).

The main features of this protocol are formaldehyde crosslinking and on-beads digestion. On one hand the usage of formaldehyde is well established not only for proteomic approaches (Sutherland, et al., 2008; Srinivasa et al., 2015), but also for chromatin immunoprecipitation (ChIP) and tissue fixation. This is because its size (~2Å) allows the permeabilization of the cell membranes without addition of extra solvents, therefore leaving the cells intact. Furthermore, it allows only proteins in

close proximity (2.3–2.7Å) to be crosslinked to each other. Thanks to the low concentration of formaldehyde and the short reaction time used in proteomic studies, unspecific crosslinks can be avoided and fixation of transient interactions allowed (Sutherland et al., 2008; Toews et al., 2008). On the other hand the on-bead digestion step allows a rapid and sensitive purification of the linked proteins.

Furthermore, RIME was suitable for our aim because the crosslink and the nuclei fraction purification make possible to identify interactions occurring in a specific cellular compartment that could also be temporary and/or weak. The technique is extremely affordable, fast and sensitive, and it allows assessing unspecific bindings simply using a parallel IgG antibody to immunoprecipitate proteins from the same lysate that will then be subtracted from the target RIME.

## 4.2 Experimental set-up

In order to perform RIME on our three genes of interest, we established a collaboration with the Biological Mass Spectrometry Facility of AZ in Waltham, USA. In particular Dr Jon DeGnore has been the reference person for the proteomic experimental part of the project and data analyses.

However, this technique had never been used there before, and the RIME samples from our laboratory had always been analysed at the Proteomic Facility at CRUK-Cancer Institute (CRUK-CI), Cambridge, UK. For these reasons, we decided to perform a pilot experiment where we repeated an investigation previously conducted in our laboratory by Dr Kyren Lazarus, where he identified BCL11A interactors in MDA-MB-231 cell line. The aim of this comparison was to evaluate the reproducibility of the data despite the different facilities. Therefore, we submitted a RIME sample to the AZ facility.

According to Mohammed et al., 2016, the reduction and alkylation processes of samples for this protocol increase the number of peptide spectrum matches (PSMs) of immunoglobulins peptides. These steps are usually performed while preparing proteins and peptides for MS analysis before the enzymatic or chemical cleavage in order to help the unfold of the protein, thereby facilitating the cut (Suttipong et a., 2017).

According to Mohammed et al., usually ~50–60 immunoglobulin PSMs are detected without any reduction and alkylation, but they increase to >500 if performed (data obtain for 10µg of bound antibody), causing a significant increased number of nonspecific proteins identified. Therefore we decided to not reduce or alkylate our RIME samples, even though this could have led to lower sequence coverage of any protein at the MS level.

The other sample was processed as described in Mohammed et al., 2016. As a mass spectrometer, we used a Thermo Q Exactive plus (Thermo Corp., San Jose, CA) rather than the LTQ Velos Orbitrap used in the paper, which allowed us to be more restrictive about the search tolerance of the ion precursors, and we analysed these first data using both Proteome Discoverer (v1.4) and PEAKS software (Bioinformatics Solutions Inc., Waterloo, ON, CA): the results are shown in Fig. 4.1. For this particular run, a cut-off of three unique PSMs was used: all the other proteins were excluded from the analysis because of the lack of confidence.

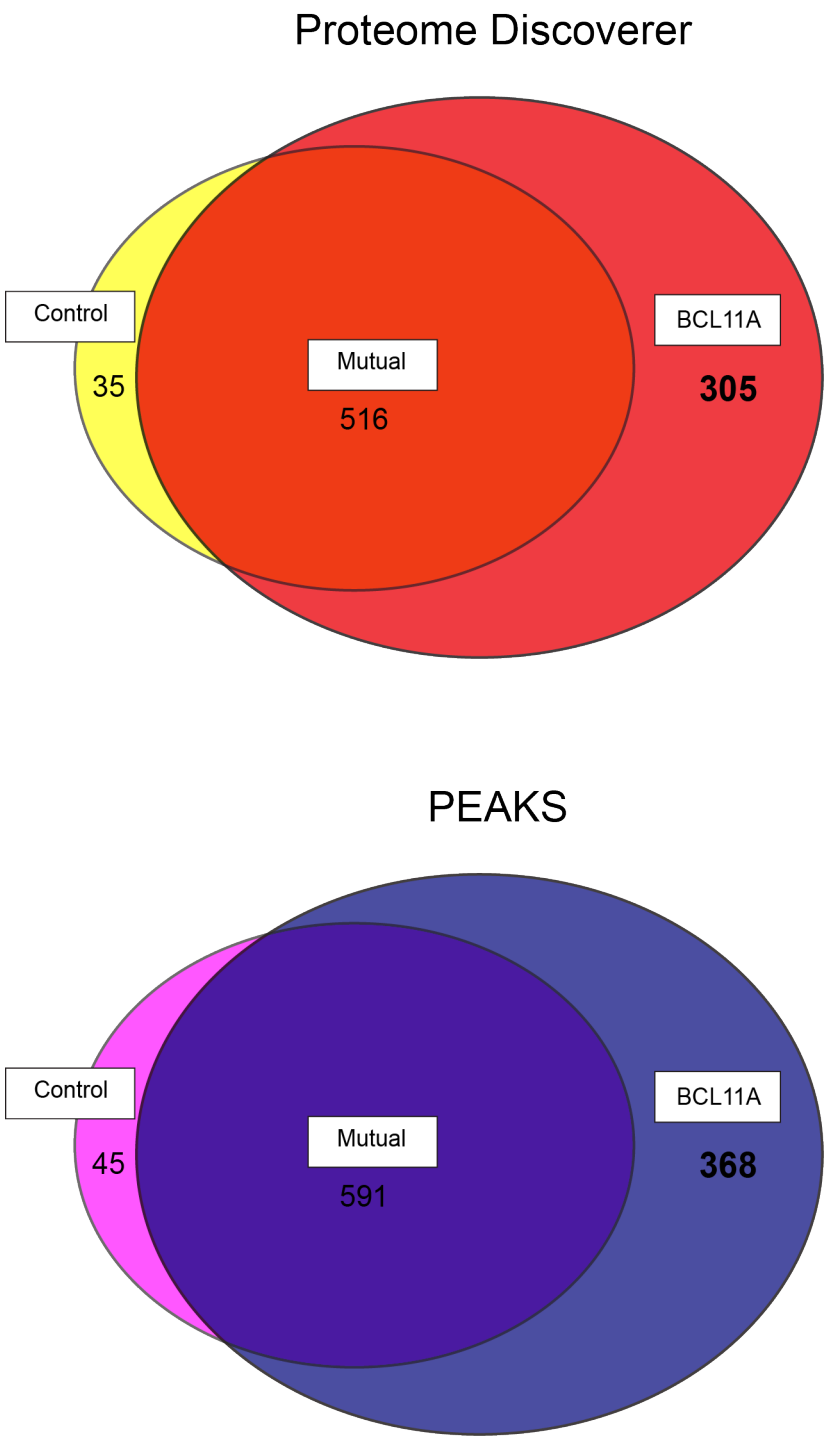



**Figure 4.1: BCL11A RIME results using Proteome Discoverer and PEAKS software for analyses.** A) Venn diagram representing the number of unique proteins identified with Proteome Discoverer potentially interacting with BCL11A (305), as well as non-specific ones (Control, 35), and proteins found in both (Mutual, 516). B) Venn diagram representing the number of unique proteins identified with PEAKS potentially interacting with BCL11A (368), non-specific ones (Control, 45), and mutual proteins (591). The IgG antibody with the same species as the BCL11A one was used as control. The analyses were performed at the Biological Mass Spectrometry Facility of AZ by Jon DeGnore.

We were extremely satisfied with the results: we were able to perform a successful RIME experiment in a different facility with different mass spectrometer. Both software seemed to give us comparable results in terms of number of proteins and overall coverage (Appendix A, Appendix B): for this reason, we decided to use PEAKS for all the other RIME experiments.

Even though a good RIME experiment is defined in Mohammed et al., 2016 by the identification of 300−900 proteins, we believed that our numbers were lower because of the higher stringency we applied to the searching parameters compared to the ones used in the published protocol. The main modifications we applied and kept for all the RIME experiments we then performed are summarized in Table 4.1.

| **Parameter** | **Value (published)** | **Value (modified)** |
|---|---|---|
| Software search | Mascot, Proteome Discoverer | PEAKS |
| Software setting: | | |
| Variable modifications | Oxidation (M)<br>Deamidation (N, Q) | Acetylation (N-term)<br>Oxidation (M)<br>Deamidation (N,Q) |
| Maximum missed cleavages | 2 | 2 |
| Precursor tolerance | 20 p.p.m | 15 p.p.m |
| MS/MS tolerance (Da) | 0.5 | 0.05 |
| Peptide FDR | 0.01 | 0.01 |

**Table 4.1: List of the modifications applied to the software settings for our RIME experiments** (adjusted from Mohammed et al., 2016)**.** Some software settings were changed in order to increase the stringency of the analysis (right column). Reference parameters published in Mohammed et al., 2016 have been reported in the left column. Parameters were changed according to the expertise of Jon DeGnore (AZ, Waltham, USA).

We then compared these results with the ones from the three replicates analysed at the Cambridge Institute Proteomic facility. Results are shown in Fig. 4.2.

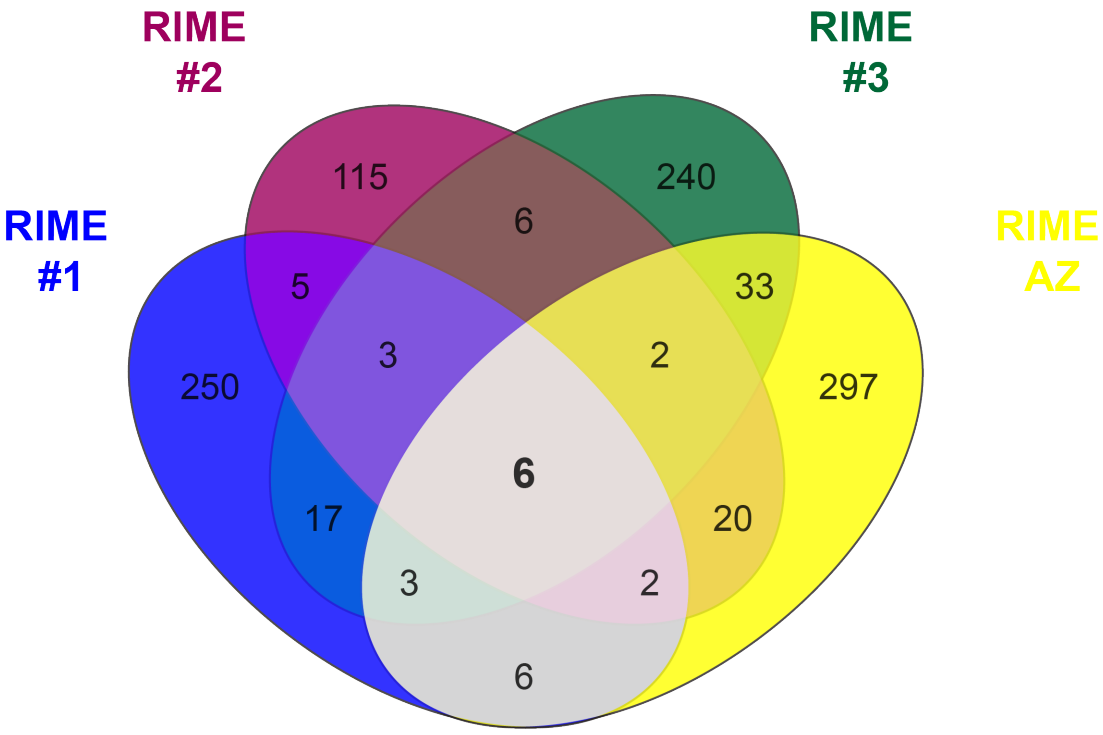



**Figure 4.2: Venn diagram representation of the common proteins identified by MS between 4 different RIME experiments on BCL11A.** Comparison between 3 RIME technical replicates (RIME #1, #2 and #3) performed by Dr Kyren Lazarus and analysed with Mascot at the Proteomic Facility at CRUK-CI (Cambridge, UK), and our RIME experiment analysed with PEAKS at the Biological Mass Spectrometry Facility of AZ (Waltham, USA) by Jon DeGnore (RIME AZ). Each RIME shows the number of unique proteins (not present in the IgG control sample) and identified by at least three PSMs (peptide spectrum matches).

Despite different facilities, different software used and different searching parameters, this pilot experiment confirmed our ability to identify 6 proteins shared with all other experiments which are: AL7A1, CHD8, CAV1A, CENPF, CHTOP and BCL11A. All these proteins were part of the 'top hits' in the previous experiments, characterized by a high number of PSMs and good coverage. More importantly, some of these interactors have been validated in the laboratory (Lazarus et al., unpublished), confirming the validity of our approach. These results encouraged us to proceed further with the collaboration.

## 4.3 RIME on dCas9 targeting *FOXC1, NFIB & NFE2L3*

In order to apply the RIME protocol to our project, we decided to run a pilot experiment to confirm the applicability of the parameters and the quality of the data themselves. Firstly, we analysed the V5-dCas9 and IgG pulldowns for the *FOXC1* gene promoter in MDA-MB-231 + Foxc1 gRNA cell line, where dCas9 was induced with Doxycycline (1µg/mL) for 48 hours, as described in Chapter 3, and the samples processed as described in Material and Methods. However, the list of proteins we identified was much shorter than the expected one from a successful RIME experiment (Table 4.2)

| UniProt Accession Number | Protein Description | | Number PSMs | Sequence coverage (%Cov:) |
|---|---|---|---|---|
| Q5VTE0 | Putative elongation factor 1-alpha-like 3 | EF1A3 | 19 | 43.073593 |
| Q13217 | DnaJ homolog subfamily C member 3 | DNJC3 | 5 | 16.071428 |
| P67809 | Nuclease-sensitive element-binding protein 1 | YBOX1 | 4 | 7.4074073 |
| P48730 | Casein kinase I isoform delta | KC1D | 4 | 12.048193 |
| Q04727 | Transducin-like enhancer protein 4 | TLE4 | 4 | 4.3984475 |
| Q04726 | Transducin-like enhancer protein 3 | TLE3 | 4 | 4.4041452 |
| Q13017 | Rho GTPase-activating protein 5 | RHG05 | 3 | 2.7962716 |
| P60983 | Glia maturation factor beta | GMFB | 3 | 29.577463 |
| O43396 | Thioredoxin-like protein 1 | TXNL1 | 3 | 12.110726 |
| Q13620 | Cullin-4B | CUL4B | 3 | 3.833516 |
| P01860 | Immunoglobulin heavy constant gamma 3 | IGHG3 | 3 | 4.244032 |
| Q9HAV0 | Guanine nucleotide-binding protein subunit beta-4 | GBB4 | 3 | 10.294118 |
| Q9BY77 | Polymerase delta-interacting protein 3 | PDIP3 | 3 | 10.451306 |

**Table 4.2: List of proteins identified through RIME on V5-dCas9 targeting the putative promoter sequence of *FOXC1* in MDA-MB-231 + Foxc1 gRNA cell line.** MDA-MB-231 + Foxc1 gRNA cells were induced for 48 hours with Doxycycline (1µg/mL), and processed according to the RIME protocol (Mohammed et al., 2016). Samples were analysed using PEAKS software at the Biological Mass Spectrometry Facility of AZ by Jon DeGnore (Waltham, USA). Proteins listed have been selected if not present in the control sample (IgG), with PSMs ≥3.

We believed that the stringent setting parameters and the high PSM cut-off we used to shortlist the candidate proteins were responsible for such a poor outcome compared to what is usually expected from a RIME experiment (300-900 proteins (Mohammed et al., 2016)). However, some of the proteins we identified have a role in the regulation of the gene transcription (for example TLE4, TLE3 (co-repressors),

or are involved in other functions like mRNA processing (YBOX1) or regulation of the translation (PDIP3).

This information was promising, supporting a good execution of the protocol itself and a potential applicability of RIME to CRISPR/Cas9. For these reasons, we decided to extend the cut-off to PSMs ≥1 (rather than ≥3, as before), which significantly increased the overall number of identified proteins, and the number of nuclear factors (Appendix C). This encouraged us to believe that our approach could be used to identify transcription factors.

For all the following RIME analyses we decided to maintain the cut-off at PSMs ≥1 in order to have a larger dataset, despite the decrease in confidence of the protein identification: with just one unique peptide calling for a protein, the certainty of a correct identification is reduced.

Furthermore, we confirmed the validity of the execution acknowledging the good coverage level we obtained for dCas9 itself (Fig. 4.3). It has to be remembered that the antibody used for this immunoprecipitation recognizes the V5 tag, not the dCas9 directly: this could explain the absence of some peptides, together with the alterations caused by the fixation process.

On the basis of these results, we decided to proceed with our first RIME experiment. Because of the intrinsic variability of the technique, we decided to analyse three technical replicates per gene and to use the internal control called MDA-MD-231 + Empty gRNA cell clone. This cell line was obtained exactly as the other cell lines, but its gRNA vector didn't contain any gRNA sequence. In this way we wanted to identify those proteins that could have reacted to the presence of dCas9, rather than being crosslinked to it because in close proximity at the gene promoter level. In Fig. 4.4 the RIME experimental workflow is reported.

**Figure 4.3: dCas9 sequence coverage obtained from RIME proteomics on *FOXC1* promoter sequence investigation.** Coverage of Q99ZW2|CAS9_STRP1 CRISPR-associated endonuclease Cas9/Csn1 OS=*Streptococcus pyogenes* serotype M1 protein evaluated using PEAKS software. Analyses were performed by Jon DeGnore (AZ, Waltham, USA). Every peptide with a match has been highlighted in grey. In blue the MS reads alignments are represented. 'O': Oxidation (M)= + 15.99.

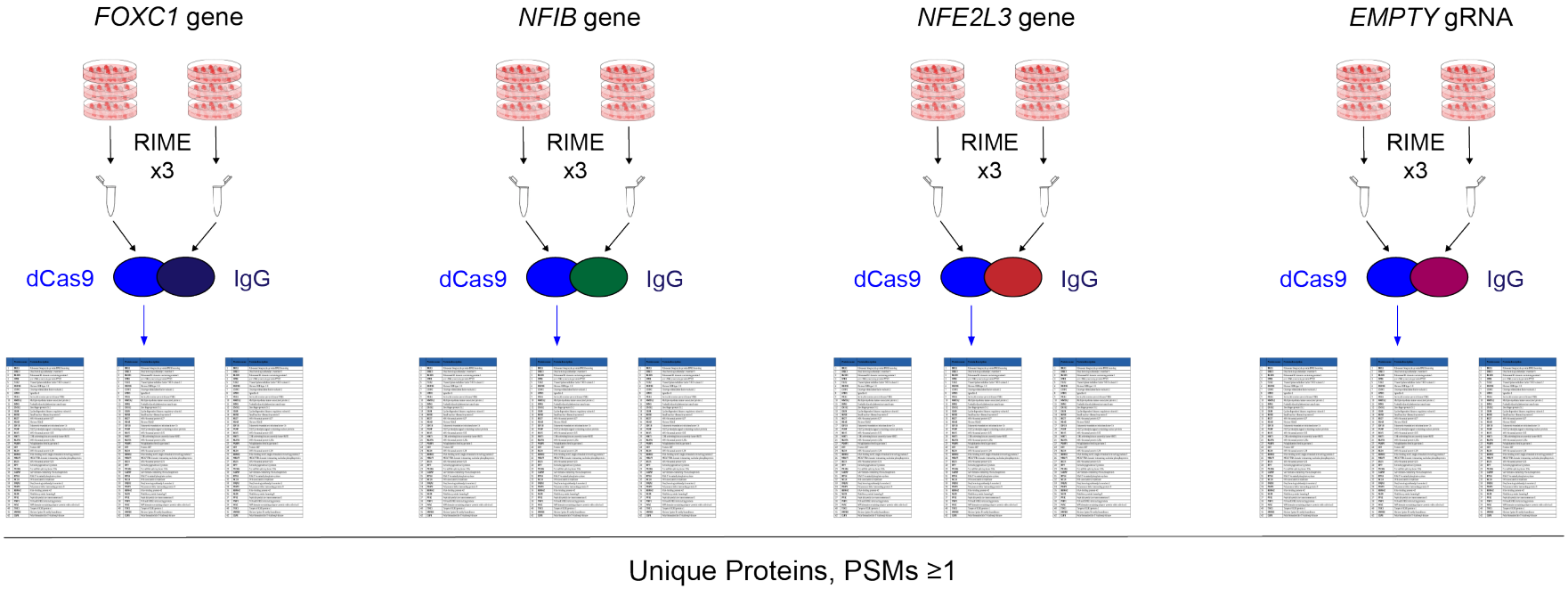



**Figure 4.4: Schematic representation of RIME experimental workflow performed on dCas9 targeting the putative promoter sequences ofgenes of interest (*FOXC1*, *NFIB*, *NFE2L3*).** The expression of dCas9 was induced for 48 hours before the cells were collected and processed as described in Material and Methods. IgG was used as a background control and just unique proteins in dCas9 with a PSM≥ 1 were considered. The MS analysis was performed using PEAKS by Jon DegGnore (AZ, Waltham, USA). The experiment was performed with N=3.

The main purpose of this experiment was to identify common transcription regulators among the three previously chosen TNBC genes. In order to do that, we looked at the unique proteins identified through MS across the replicates first, and genes later, and we excluded those ones within the Empty gRNA vector samples, considered as background. However, we realized that our analysis was too stringent.

For this reason, we tested some optimisation steps, aiming to improve the overall number of proteins, the coverage, number of PSMs per protein and the reproducibility of the data (Table 4.3).

| Parameters | Variation | Rational |
|---|---|---|
| Cell population purity | Double positive cells resorted | Higher cells number |
| Seeding confluency | Decreased | Better sonication |
| LC/MS | Longer columns | Better separation |
| Mass spectrometer | Orbitrap Fusion Lumos | Better signal |
| Beads | Types | Better signal |

**Table 4.3: Summary of the different optimizations attempted in order to improve the quality of our RIME data.** Far left column: list of parameters we modified. Middle column: type of modification applied. Far right column: rational behind this change.

Briefly, we started modifying some aspects of the sample preparation: we sorted again the double positive cells for every cell clone in order to increase the purity of the experimental population, and we aimed to reach 60-70% confluency of the cells on the day of collection to facilitate the sonication process. At the mass spectrometry level, we used longer columns to obtain a better separation of the peptides during the liquid chromatography phase, and we compared our results with a more powerful mass spectrometer (Orbitrap Fusion Lumos). However, what really had a significant impact on the quality of our data were the beads and the concentration of the antibody used.

Firstly, we decided to switch to Dynabeads magnetic beads (ThermoFisher Scientific) rather than PureProteome Magnetic Bead (Merck Millipore) following a discussion with Dr Jon DeGnore, in order to achieve a better enrichment of the immunoprecipitated protein compared to the respective IgG control.

We then assessed what the best combination of antibody concentration and type of beads (Protein A, Protein G, or Protein A + G) was for our purposes. We performed a pilot ChIP-qPCR experiment on dCas9 induced in MDA-MB-231 + Nfe2l3 gRNA cell line that showed us how we obtained the nicest enrichment when we coupled our antibodies with Protein A beads on a ratio of 1:10 (1µg of antibody every 10µL of beads) (Fig. 4.5). We used this set-up for all the other RIME experiments we performed.

All these variations lead to an increase of not only the protein numbers identified in both IgG and dCas9 pulldowns of ~40%, but also of the coverage (~50% improvement). We then submitted three RIME replicates (dCas9 and IgG immunoprecipitations) per each gene, and we run them at the mass spectrometer as previously described.

As a first approach to the data, we decided to consider just those proteins that were not present in the control sample (IgG), without any filtering on PSM. We then evaluated the reproducibility between RIME replicates per each gene, and we compared the redundant proteins with those coming from the other 2 genes. In Fig. 4.6 the results of this comparison are shown (for the full list of common proteins, see Appendix D).

Overall, this was a substantial improvement for our RIME experiment from a quantitative point of view and in terms of reproducibility.

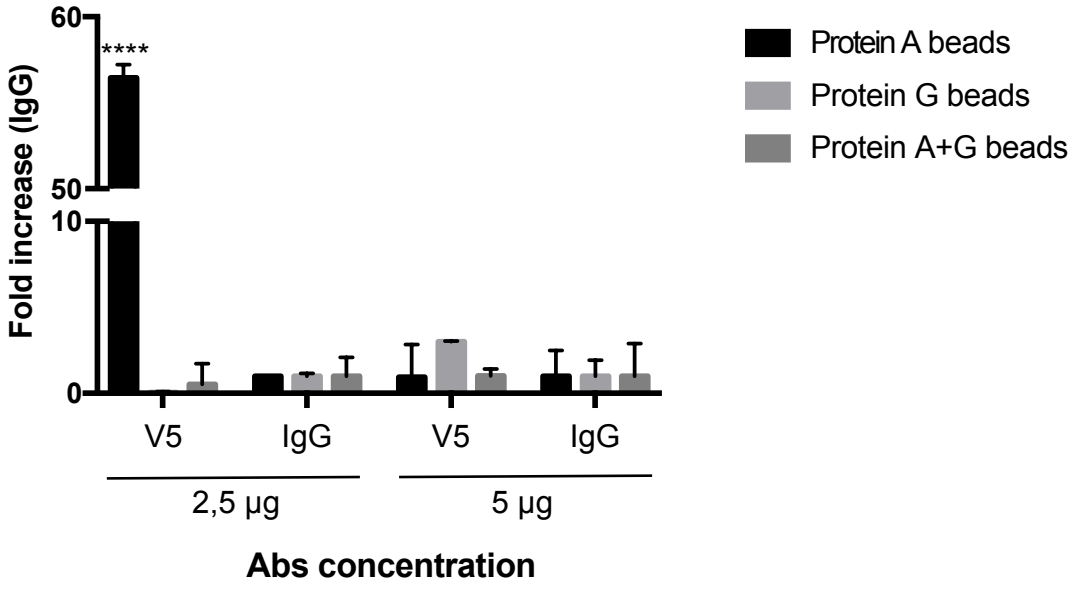



**Figure 4.5: Antibodies and beads optimisation for ChIP-qPCR on dCas9-IP, targeting the promoter sequence of the *NFE2L3* gene of interest.** ChIP-qPCR was performed to evaluate the optimal combination of type of beads and concentration of V5 antibody to pull down dCas9 targeting the promoter sequence of *NFE2L3* on MDA-MB-231 + Nfe2l3 gRNA cell line, as a reference. Three different combinations of beads were tested (Protein A, B and A+B beads) and two concentrations of antibody (2.5 and 5µg). The enrichment of the pulldown DNA was determined by qPCR and normalized to IgG. The error bars report standard deviations from duplicates. Two-way ANOVA test was performed to compare the DNA enrichment to the respective IgG. P value <0.05.

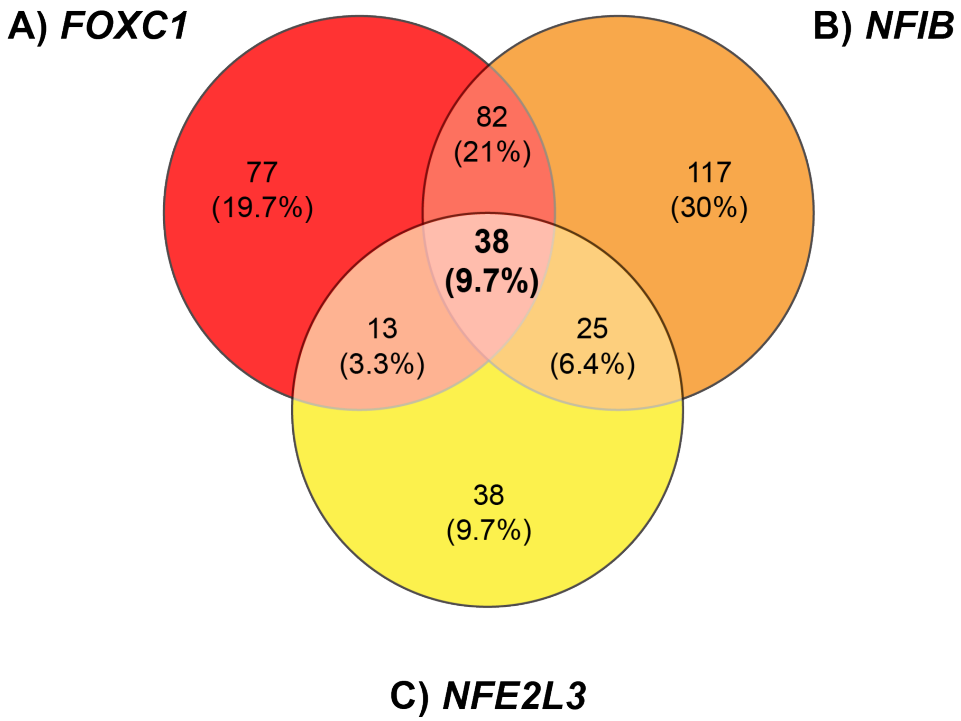



**Figure 4.6: Number of proteins in common between *FOXC1*, *NFIB*, and *NFE2L3* promoters identified through RIME.** Three RIME replicates per gene (dCas9 & IgG pulldowns) were submitted and analysed through MS by Jon DeGnore at the Biological Mass Spectrometry Facility of AZ (Waltham, USA). For every gene, a list of redundant proteins among replicates was generated and compared with the lists developed for the other genes. Overall, 38 proteins were found in common between all genes (9.7% of all the proteins identified). A) Redundant proteins among RIME replicates potentially interacting with *FOXC1* promoter: among them, 19.7% were exclusively identified within this gene; 21% were potentially in common with just *NFIB* promoter, and 3.3% with the *NFE2L3* one. B) Redundant proteins among RIME replicates potentially interacting with *NFIB* promoter: among them, 30% were exclusively identified within this gene, and 25% with the *NFE2L3* one. C) Redundant proteins among RIME replicates potentially interacting with *NFE2L3* promoter. Proteins seen just in dCas9 were considered, and no PSM filter was used. The experiment was performed with N=3.

Among the common proteins between all or some genes, we looked for those with a strong biological rational (primarily transcription factors), ideally with a strong proteomic justification (good coverage, high number of PSMs), and with a tumourigenic background. One particular protein stood out, MTA2.

This protein is a transcription regulator part of the NuRD (Nucleosome Remodelling Deacetylase) complex (Basta et Rauchman, 2015), and according to our RIME results could have been involved in the regulation of *FOXC1* and *NFIB* genes. Also, several other components of the NuRD complex were identified across different replicates of different genes (for example CHD4, MBD3, RBBP4/7), increasing its strength as a potential master regulator candidate. For these reasons, we decided to consider it as a candidate and to proceed with further validations.

## 4.4 Novel statistical approach

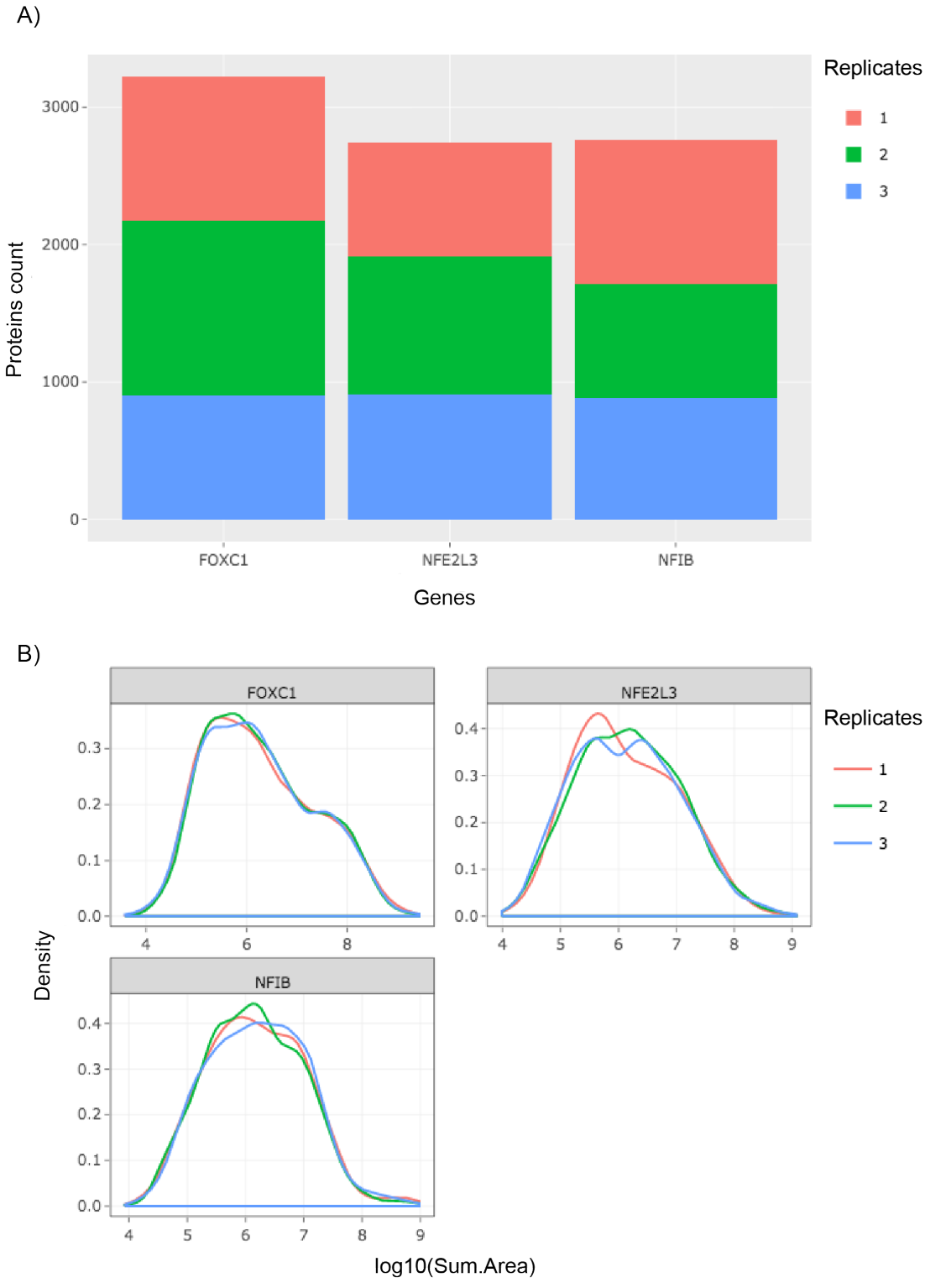
Despite the improvement we reached, the abundance in our dataset was still not comparable with the one expected from a successful RIME experiment (between 300-900 proteins, as stated in Mohammed et al., 2016). In particular we noticed that excluding *a priori* proteins from any dCas9 pulldown just because they were also present in the IgG control not only was affecting the results, but was also misleading us. We didn't in fact take into consideration that many of these proteins in common between the two RIME samples were diversified in terms of coverage and/or number of PSMs.

For these reasons, we decided to change our way to analyse the data. Here we propose a novel statistical approach developed by Dr Beate Ehrhardt (previously at AZ, Cambridge, UK, now at Institute for Mathematical Innovation (IMI), Bath, UK), in collaboration with Dr Piero Ricchiuto and Dr Aurelie Bornot (both Darwin Building, AZ, Cambridge, UK) to understand if a protein was a true dCas9 interactor.

In order to do that, we combined all RIME dCas9 samples and all the RIME IgG controls, and we evaluated if proteins were actually significantly different using the SumAUC (sum of the 'Area Under the Curve') parameter as an indicator of the relative abundance of that protein. This value is the result of the sum of the AUC of all the peptides identified for that protein, while the AUC of a peptide is a parameter

calculated by an algorithm (PEAKS in our case) on the basis of the RT (retention time) and the intensity of the peaks of the ionized peptide recorded at the MS level. The overall number of peptides used for this calculation (not just the PSMs), the number of replicates, the gene of interest and the case (dCas9 or IgG) were still taken into consideration at this point.

In Fig. 4.7 the results of this statistical exploratory analysis for our last RIME experiment using Protein A beads are shown. We firstly evaluated the distribution of the overall proteins (identified in both dCas9 and IgG pulldowns) within replicates of the same gene, and between genes: as shown in Fig 4.7, A, FOXC1 RIME sample was the one with the higher number of proteins identified, while NFE2L3 and NFIB seemed to have a lower, but comparable one. Overall, replicates among the genes show a similar level of protein identification. When looking at the density plots, it is possible to see how the distribution of the sum of the areas for all the proteins was consistent between replicates of the same gene (Fig. 4.7, B), and conditions (Fig. 4.7, C).


A)
Proteins count
3000
2000
1000
0
FOXC1
NFE2L3
NFIB
Genes
Replicates
1
2
3
B)
Density
FOXC1
NFE2L3
NFIB
0.0
0.1
0.2
0.3
0.4
4
5
6
7
8
9
log10(Sum.Area)
Replicates
1
2
3

C)

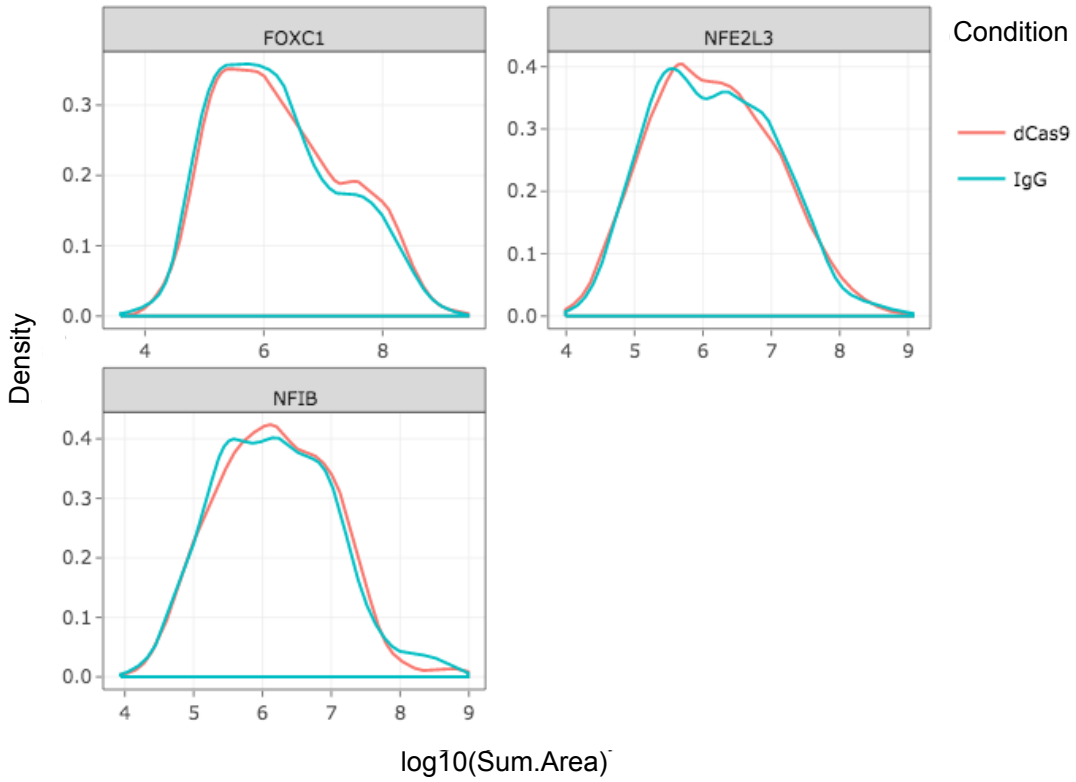


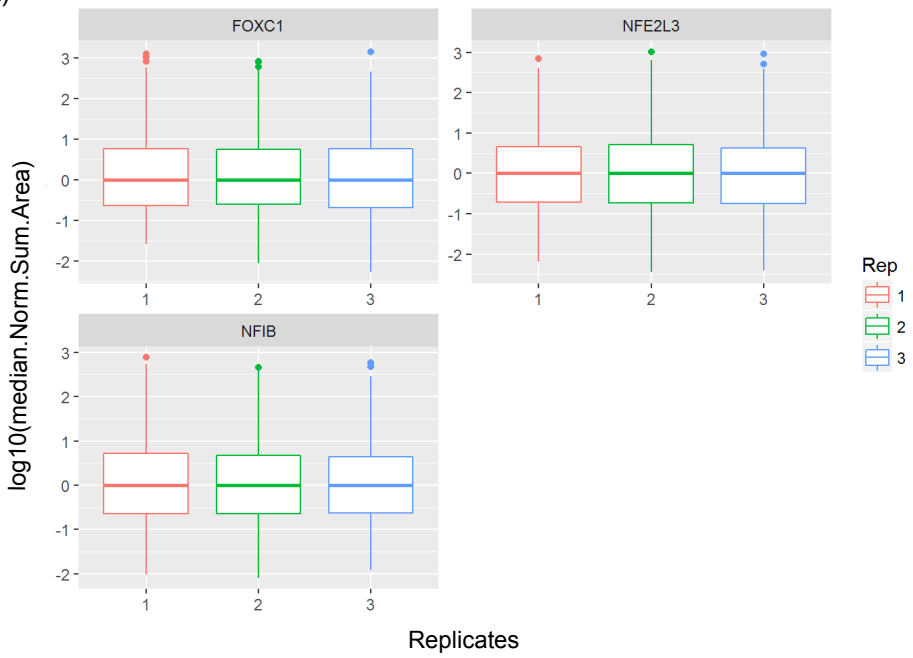

**Figure 4.7: Exploratory data analysis for dCas9 RIME experiment.** The proteomic data obtained by dCas9 and IgG immunoprecipitation on *FOXC1*, *NFIB* and *NFE2L3* promoters on the respective MDA-MB-231 cell clones were evaluated to see if proteins were significantly different from dCas9 (case) and IgG (control), on the basis of their relative abundance (Sum AUC). A) Number of proteins per gene of interest per replicate. B) Density distributions per gene of interest per replicate. C) Density distributions per gene of interest per condition. LC/MS and PEAKS analysis data were performed by Jon DeGnore (AZ, Waltham, USA). Data analysis was carried out by Beate Ehrhardt (IMI, University of Bath), Piero Ricchiuto and Aurelie Bornot (AZ, Cambridge, UK).

We normalized each sample by its own median of relative abundance, where with sample we refer to replicate, gene, condition (Fig. 4.8, A, B). However, the density distribution per gene and per condition showed a possible bimodal tendency of the distribution, definitely for *FOXC1* gene: we concluded we could not simply assume normality. Fig. 4.8, D, shows the result of this model comparison more clearly: when we evaluated all data for all genes simultaneously we observed that the number of peptides, genes, proteins, and condition (dCas9, IgG) significantly influence the relative abundance. In particular we noticed that on a Q-Q plot the points tend to curve off at the extremities: therefore the data have more extreme values than it would be expected if they truly came from a normal distribution.

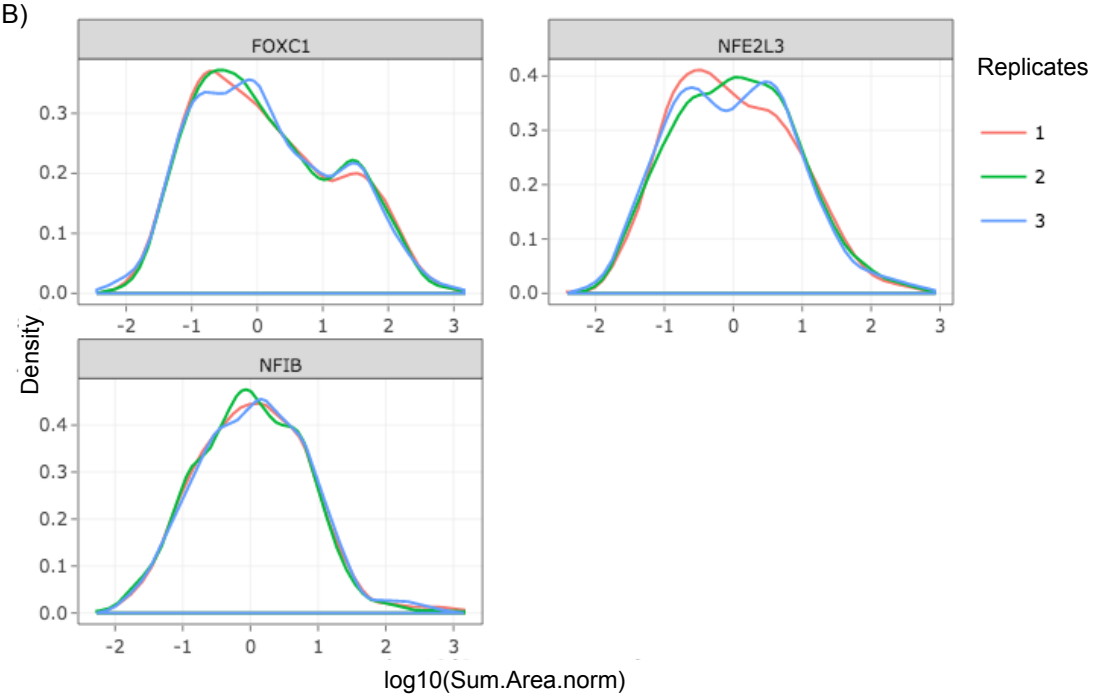
We then tested if the difference of the relative abundance of every protein was significantly different between dCas9 and IgG: to do so, we fit a linear model to the relative abundance for each protein separately. Despite the fact that the number of peptides had a significant effect on the relative abundance, we decided to exclude it from the statistical model. Since we run the test on a 'per protein' level, the information captured in the number of peptides directly related to relative abundance and any argument that the number of peptides relates to the length of proteins becomes void. We therefore did not adjust for number of peptides.

Furthermore, if the protein had been observed for more than one gene we adjusted for differences between the genes. It also has to be mentioned that we excluded those proteins that were observed but that did not get an area assigned. The statistical analysis could have been run only for those proteins with three or more replicates across genes, and observed in both dCas9 and IgG.

A)


FOXC1
NFE2L3
NFIB
log10(median.Norm.Sum.Area)
Replicates
Rep
1
2
3

B)


FOXC1
NFE2L3
NFIB
Density
log10(Sum.Area.norm)
Replicates
1
2
3

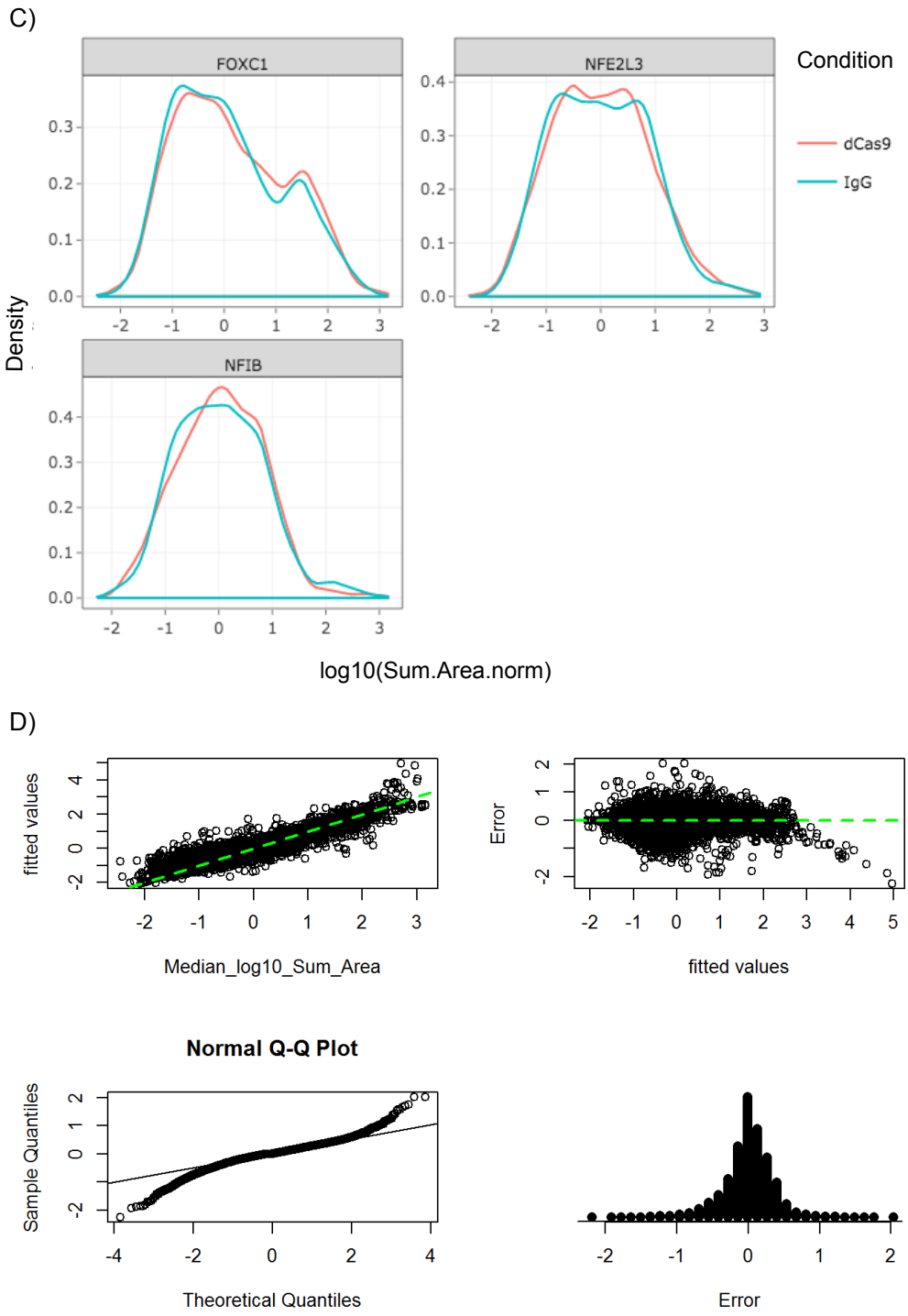



**Figure 4.8: Distribution of the relative abundance after normalization and model fitting.** Every replicate, gene, condition obtained by dCas9 and IgG immunoprecipitation on *FOXC1*, *NFIB* and *NFE2L3* promoters on the respective MDA-MB-231 cell clones were normalized by its own median of the relative abundance (Sum.Area.norm). A) Log10 of number of proteins per gene of interest per replicate after normalization. B) Density distributions per gene of interest per replicate after normalization. C) Density distributions per gene of interest per conditions after normalization. D) Overall data visualization and model comparison to normal distribution. Statistical analysis developed by Beate Ehrhardt (IMI, University of Bath), Piero Ricchiuto and Aurelie Bornot (AZ, Cambridge, UK).

The numbers of proteins excluded from the analysis for one of these reasons, and the number of proteins tested are reported in Table 4.4.

| Proteins seen: | Numbers |
|---|---|
| dCas9 only | 285 |
| IgG only | 8 |
| < 4 replicates | 816 |
| **Tested** | **475** |

**Table 4.4: Overall numbers of proteins evaluated for the statistical analysis.** Numbers of proteins excluded because only present in dCas9 (285), IgG (8), or in <4 replicates across genes (816), and number of proteins tested (475). Statistical analysis method developed by Beate Ehrhardt (IMI, University of Bath), Piero Ricchiuto and Aurelie Bornot (AZ, Cambridge, UK).

After correcting for multiple testing (FDR), the proteins reported in Appendix E showed a difference in the relative abundance between dCas9 and IgG to a significance level of 5%. We also displayed the CRAPome (Contaminant Repository for Affinity Purification) scoring and the total number of replicates. In particular, this scoring (http://www.crapome.org/, Mellacheruvu et al., 2013) was considered as an indicator of the non-specificity of interactors on the basis of published proteomic experiments: the lower the value, the more interesting the protein; the higher, the more likely the protein was just background.

Thanks to this approach we ended up having two sets of results: one list of proteins exclusively identified between dCas9 samples across the 3 genes (Appendix F), and one list of proteins (Appendix E) with a statistical significance and a specificity information. In order to choose objectively which candidates to validate, we developed a ranking method, in close collaboration with Aurelie Bornot (AZ, Cambridge, UK).

## 4.5 Identification of candidate transcriptional regulators in TNBC

In order to highlight those candidates not only biologically, but also therapeutically significant, we developed a system where proteins were ranked according to their desirability: this was calculated on the basis of a given score among several categories that we believed could have helped us in the research of a master regulator.

Desirability functions have been proposed as a way to integrate several numerous selection criteria in order to rank and select candidates (such as genes, proteins, metabolites, or lipids) from high-throughput biology experiments. Every variable is mapped to a continuous 0-1 scale, where 1 is the highest desirability, and 0 the lowest (Lazic, 2015). These functions were developed by Harrington (Harrington, 1965) and later extended by Derringer and Suich (Derringer et Suich, 1980). Nowadays they are used in cheminformatics to rank compounds (Segall, 2012; Bickerton et al., 2012), but can be applied to all those –omics technologies generating long lists of differentially expressed datasets.

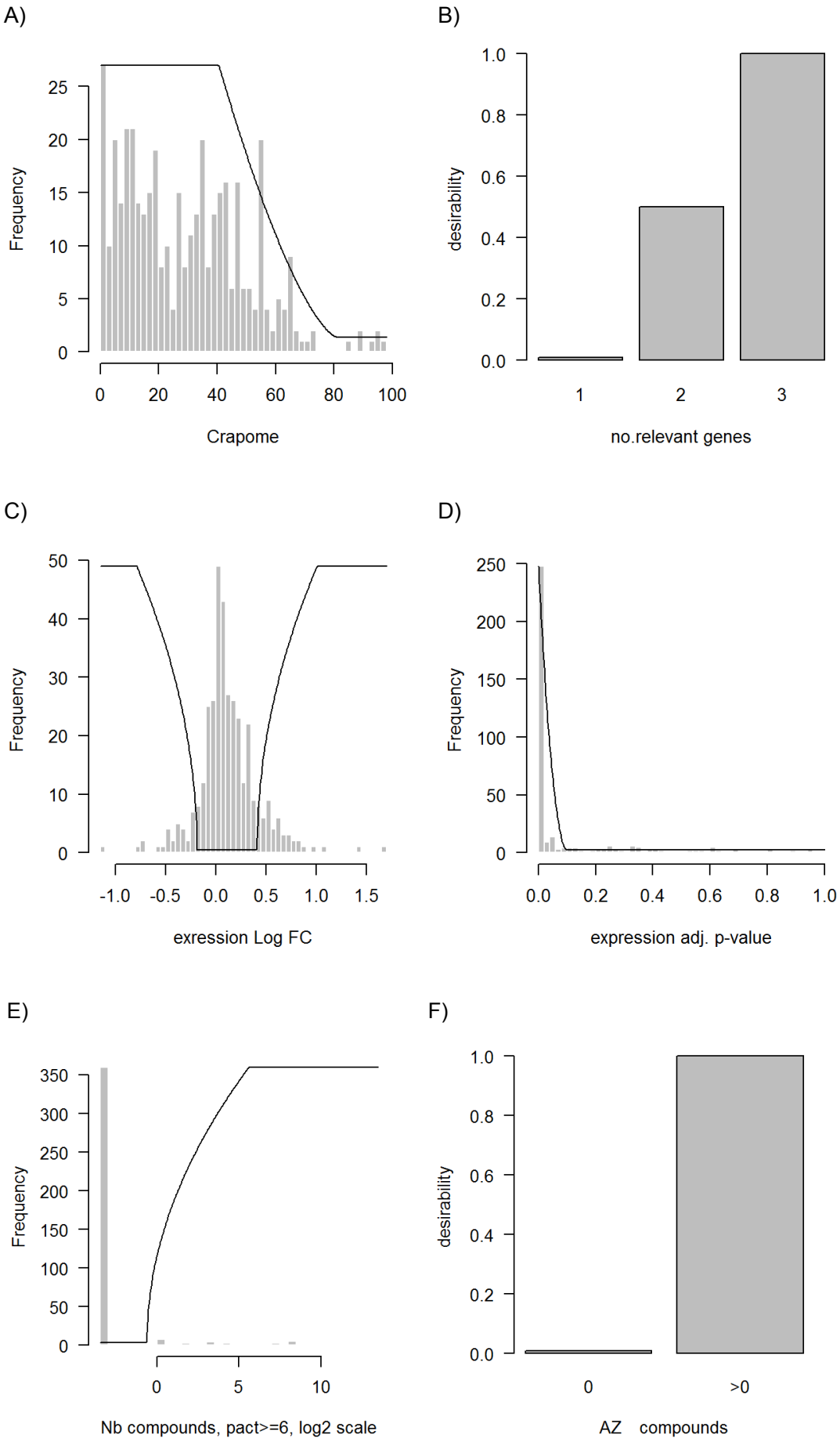
Before going into details, it has to be mentioned that we included in the analysis not only the proteins identified with the statistical method, but also those ones that were unique for the RIME dCas9 samples (Appendix F): for these last ones, a set P value of -1 was assigned.

In particular, we decided to consider the following variables as index of a protein hit desirability:

1. *CRAPome value*: as an indicator of non-specificity of the protein; the higher the CRAPome score, the lower the desirability of the protein.
2. *Number of relevant genes*: the more genes of interest had seen the questioned protein, the higher the score.
3. *Differential expression* according to METABRIC dataset.
4. *P value*: assigned after the statistical analysis previously described, or -1 if the protein was only seen in the dCas9 RIME samples.
5. *Number of compounds available*: overall number of any type of compounds available against the questioned protein, with a cut-off of activity ≥ 6.
6. *Number of AstraZeneca compounds available*: overall number of available compounds within AZ.

7. *Number of clinical compounds available*: the questioned protein was more desirable if it had compound(s) already used in the clinic.
8. *Type of relation to TNBC*: a score of 1 was assigned to those proteins with a known role in TNBC.
9. *Number of literature references to support the relation with TNBC*: among those proteins with literature references supporting their role in TNBC, we catalogued them into newer when they were lower than five overall at the time of the analysis, well known if more than that.

In Fig. 4.9 the general guidelines for this ranking are reported. Firstly, every protein was associated with a CRAPome value between 0 and 100 that predicted the probability of it to be just background: the lower the value, the more desirable the protein. We didn't take into considerations all those proteins with a CRAPome value >80 (Fig. 4.9, A). Because we were interested in finding a protein able to regulate the expression of several genes, we decided to assign a score on the basis of how many promoters had potentially shown the presence of it through RIME: 1 if it was all three, 0.5 if it was 2, and 0 if just one (Fig. 4.9, B). In addition, we correlated the level of expression of the protein with the patients' data coming from the METABRIC cluster for TNBC compared to the other subtypes of breast cancer (Fig. 4.9, C); we evaluated its differential expression between dCas9 and IgG, and we assigned a score of desirability on the basis of the P value, setting a cut off at P>0.01 (the higher the P value, the more desirable) (Fig. 4.9, D). However, one of the most important factors in our ranking system was the presence of a compound for the protein hit: we evaluated firstly the existence of any kind of compound published with an activity of at least 6 (Fig. 4.9, E), and we assigned an extra score if they were present among the AZ available compounds (Fig. 4.9, F), or already used in the clinic (Fig. 4.9, G). The final information we used to rank the hits was their relation to TNBC on the basis of the literature using '*Pathway studio'* (https://www.pathwaystudio.com/): the more published information we found about their impact in TNBC, the more confident we were about the proteins (Fig. 4.9, H). We considered the relation with TNBC well known if we could found at least five publications, new if they were lower than (Fig. 4.9, I).


A)
Frequency
Crapome
B)
desirability
no.relevant genes
C)
Frequency
exression Log FC
D)
Frequency
expression adj. p-value
E)
Frequency
Nb compounds, pact>=6, log2 scale
F)
desirability
AZ compounds

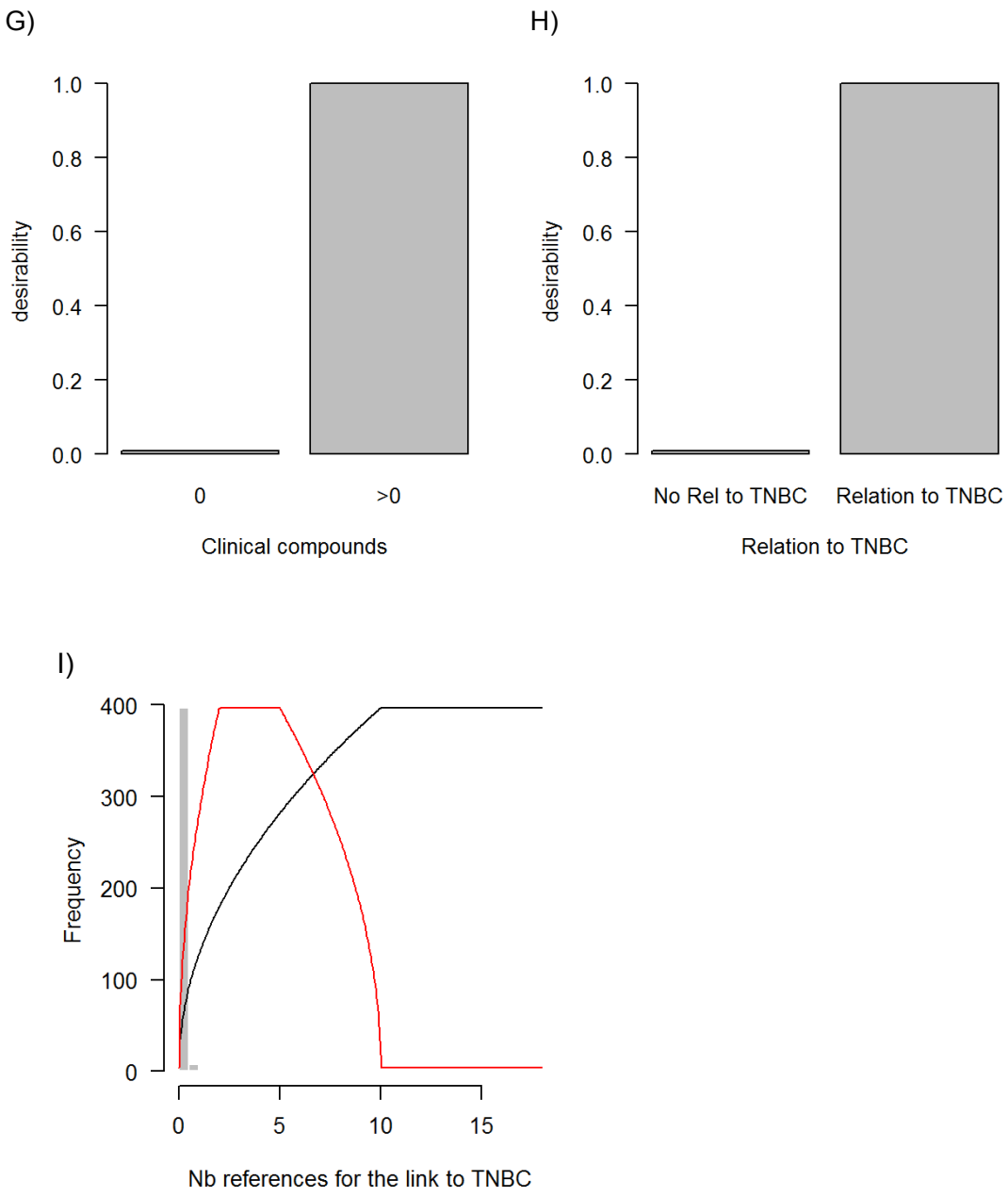



**Figure 4.9: Variables and level of desirability for RIME hits ranking.** Overview of the variables considered to rank the protein candidates identified through RIME for the three genes of interest (*FOXC1, NFIB* & *NFE2L3*). A) Distribution of proteins on the basis of their CRAPome value, with cut-off set at 80. B) Value of desirability based on the number of genes involving the protein. C) Log fold change (LFC) of expression of the protein between TNBC METABRIC IntClust compared to the others. D) Distribution of the proteins on the basis of the P value calculated with the statistical analysis. E) Distribution of the number of compounds available for the analysed protein with an activity ≥ 6. F) Value of desirability on the basis of the presence of any compound among the AZ available ones. G) Value of desirability assigned if there are compounds available in the clinic. H) Desirability based on a known relation of the protein with TNBC. I) Distribution of novel (<5) or well-known proteins on the basis of the numbers of references in the literature. The analysis was performed by Aurelie Bornot, AZ, Cambridge, UK.

With this particular dataset, we decided to take a discovery approach, and to choose candidates on the basis of their potential novelty. We wanted to validate primarily candidates whose role for the tumourigenicity of TNBC could have been new, without forgetting though the remaining scores assigned for every other variable considered (Appendix G). In Table 4.6 the top ranked candidates are shown: CDK6 and CDK1. These proteins not only were the most novel among the RIME hits identified, but were also characterized by a high level of desirability, as shown by their low CRAPome value (5.84 and 19.71, respectively), high number of RIME replicates where they were identified (7 and 6), the fact that they were only seen in the dCas9 RIME samples (set P value of -1), numbers of available compounds also within the clinic (4 each). In addition, both proteins were potentially involved in the regulation of some (*FOXC1* and *NFIB* for CDK1), or all (for CDK6) genes of interest, which made them strong master regulators candidates.

| Accession | Crapome | N° rep | Genes | P value adjust | mlogFC | N° Comps | N° Comps pact>5 | N° AZ Comps pact>5 | N° Comps Clinic | Type Relatio | N° Ref | Known | New |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Q00534 **CDK6** | 5.84 | 7 | *FOXC1 NFE2L3 NFIB* | -1 | 1.439236 | 2559 | 2095 | >0 | 4 | Regula-tion | 1 | 0.91 | 0.97 |
| P06493 **CDK1** | 19.71 | 6 | *FOXC1, NFIB* | -1 | 0.622985 | 24443 | 10953 | >0 | 4 | Regula-tion | 1 | 0.82 | 0.87 |
| | | | | | | [...] | | | | | | | |
| O94776 **MTA2** | 16.06 | 7 | *FOXC1 NFE2L3 NFIB* | -1 | 0.209349 | 0 | 0 | 0 | 0 | None | 0 | 0.04 | 0.04 |

**Table 4.6: CDK6 and CDK1 top candidates among ranked RIME hits on the basis of novelty.** CDK6 and CDK1 top ranked RIME proteins on the basis of their value of desirability according to chosen variables (CRAPome, number of replicates, genes of interest, P value, LFC of expression according to the METABRIC dataset, number of compounds available, number of compounds available with a pact>5, number of compounds available within the AZ available compounds with a pact>5, number of compounds available in the clinic, relation with TNBC, number of references published in the literature, well known candidate or newer candidate). Overall, proteins were considered primarily on the basis of their novelty. The analysis was performed by Aurelie Bornot (AZ, Cambridge, UK). As a comparison and for an overall prospective, MTA2, protein previously chosen without the statistical analysis, is here presented, even though characterized by a lower ranking.

## 4.6 Discussion

In this chapter we presented the application of RIME and CRIPR/Cas9 to identify novel transcription factors. For this purpose, we successfully established a collaboration with the Biological Mass Spectrometry Facility of AZ (Waltham, Massachusetts, USA), and in particular with an extremely talented collaborator, Dr Jon DeGnore. Despite their novelty to the technique, we managed to reproduce RIME proteomic data previously obtained and validated in our laboratory (Lazarus et al., unpublished), even if using more stringent conditions (in terms of precursor tolerance, or MS/MS tolerance for example), and different software for the analysis (PEAKS instead of Proteome Discovery).

We demonstrated how RIME proteomics can pull down dCas9 along with proteins potentially recruited within the regulatory region of interest, and how some variables like the type of beads could make such an important difference in terms of quality of the proteomic data, and quantity.

A similar technique called CAPTURE (CRISPR affinity purification *in situ* of regulatory elements) is successfully used to identify locus-specific chromatin-regulating protein complexes and long-range DNA interactions at a single-copy genomic locus (Liu et al., 2017), further confirming the applicability of our approach.

In close collaboration with Dr Beate Ehrhardt (IMI, University of Bath), Dr Piero Ricchiuto and Dr Aurelie Bornot (AZ, Cambridge, UK), we developed a novel, statistical approach to analyse RIME protein hits, based on the difference of the relative abundance (sum of the AUC, area under the curve) of a particular protein between the experimental pulldown (dCas9 in our case) and the background (IgG). This method allowed us to have an objective, scientific approach to the data despite their quality or origin. Because of the nature of the test we applied, we were able to associate a statistical significance to 475 proteins, and to identify 285 proteins exclusively pulled down with dCas9.

Furthermore, we presented here a new system we developed in close collaboration with Aurelie Bornot (AZ, Cambridge, UK) to rank potential CRISPR/Cas9-RIME candidates on the basis of their biological and therapeutic interest. This ranking strategy finds its foundations in desirability variables like CRAPome, the existence of targeting compounds (preferably in the clinic or within the AZ available compounds),

their association with the disease of interest (TNBC in our case), and the protein novelty within the field.

Among our candidates, we decided to pursue two proteins with further analyses: CDK1 and CDK6. Both these proteins belong to the cyclin dependent kinases (CDKs) family, involved in the regulation of the eukaryotic cell cycle. However, CDK1 is the only kinase essentially required for a successful completion of the cell cycle, in particular of the M-phase (Brown et al., 2015). It has been shown that CDK1 conditional knockout mice are also not viable and the derived embryonic fibroblasts show an arrest in G2 (Diril et al., 2012). Its deletion cannot be rescued by the closest relative CDK2 knock-in, suggesting it may possess unique pattern, level of expression and structural features (Satyanarayana et al., 2008). Recently Menon et al. identified though GSEA analysis of CDK1-high tumour cells from melanoma, colon and pancreatic cancer some pathway signature involving E2F, G2M, MYC and spermatogenesis, supporting a stem-like nature of these tumour cells. They also demonstrated a new role for CDK1 in regulating tumour-initiating capacity in melanoma and suggested a novel treatment strategy for cancer via interruption of CDK1 function and its protein-protein interactions (Menon et al., 2018).

On the other hand CDK6 is active in mid-G1 phase and, together with CDK4, it phosphorylates, and thus regulates the activity of tumour suppressor protein Rb and its related proteins p107 and p130. These proteins interact with the family of transcription factors known as E2 promoter binding factors (E2F1-E2F8), repressing transcription of genes that are essential for cell cycle progression (Harbour et al., 1999). It is not surprisingly then that the aberrance of the CDK4/6 cyclin D-INK4-pRb-E2F pathway is common in >80% of human cancers (Ortega et al., 2002). In addition, CDK6 phosphorylates other transcription factors such as forkhead box M1 (FOXM1), mothers against decapentaplegic homolog 2/3 (SMAD2/3), eyes absent homolog 2 (EYA2) and methylosome protein 50 (MEP50), when part of the CDK4/6 cyclin D complexes, or nuclear factors like NF-kB, linking cancers to inflammation (Buss et al., 2012; Handschick et al., 2014).

Furthermore, cdk6-null mice develop normally, suggesting a specific oncogenic role for this kinase: in fact, blockage of CDK6 by microRNAs (miRNAs) has been shown to inhibit the proliferation of several tumours like gliomas, prostate and lung and colorectal carcinoma cancer (Anderlind et al., 2010; Chen et al., 2013; Honeywell et

al., 2013; Zhu et al., 2013; Li et al., 2014). Many studies have been published suggesting the potential therapeutic benefit of CDK6 inhibitors against different types of cancer.

The third candidate we decided to validate was MTA2, less novel than the proteins just mentioned, but with a strong biological meaning. MTA2 is in fact a member of the metastasis tumour associated (MTA) family of transcriptional regulators and is a central component of the nucleosome remodelling and histone deacetylation complex (Mi-2NuRD complex) (Bowen et al., 2004). The core subunits of Mi-2/NuRD complexes, Mi-2α and Mi-2β, are SNF2-like ATPase of the chromodomain helicase DNA-binding (CHD) protein family (CHD3 and CHD4, respectively, part of a subclass of the SWI/SNF family (Eisen et al., 1995), while the other components are utilized interchangeably to produce functionally distinct complexes. They include histone deacetylases (HDAC)1/2, methyl CpG binding domain proteins (MBD)2/3, histone-binding proteins/retinoblastoma-binding proteins (RbAp46 and/or RbAp48) that might function as structural proteins providing interactive interfaces for other components of the Mi2/NuRD complex (Marhold et al., 2004) and MTA1/2/3 (Manavathi et Kumar, 2007) proteins.

MTA2 is well known to be involved in the regulation of cytoskeletal organization via modulation of the Rho signalling pathway, and to be involved in the EMT (epithelial to mesenchymal transition) process through TWIST activity regulation (Fu et al., 2010; Yang et al, 2004). It has been shown that MTA2 takes part in the regulation of the invasive behaviour for many cancers like oesophageal squamous cell carcinoma, non-small cell lung cancer and breast (Pakala et al., 2011; Weng et al., 2014; Zhang et al., 2015; Sen et al., 2014). In breast, it plays other, important roles: it supports tumour progression, as shown by the increase of its expression during the development of mammary tumours in a multi-stage model of tumour progression (from normal duct to premalignant lesions to hyperplasia to ductal carcinoma and to invasive carcinoma) (Zhang et al., 2006), and interestingly the inhibition of ERα transactivation activity, promoting the development of hormone-independent phenotypes, in collaboration with MTA1 (Mazumdar et al., 2001; Cui et al., 2006).

For these reasons we believe they could have been strong, potential master regulators candidate in TNBC.

## 4.7 Conclusion

In conclusion, we demonstrated here the applicability of this combination of CRISPR/Cas9 and RIME proteomics strategy to investigate the regulation of transcription of a gene of interest. These results supported further analyses that are presented in Chapter 5.

# CHAPTER 5: VALIDATION OF POTENTIAL TRANSCRIPTION REGULATORS

## 5.1 Introduction

In Chapter 4 we showed how we applied the RIME protocol from Mohammed and colleagues (Mohammed et al., 2013) to investigate the regulation of the transcription of three genes of interest. In particular, we demonstrated how we successfully adapted it to an exogenous protein like dCas9, increasing the signal at a mass spectrometry level.

Furthermore, we presented a novel, statistical method to analyse RIME proteomic data on the basis of the relative abundance of the identified protein. This implemented the confidence of potential candidates' selection for validation, together with a powerful ranking method based on the level of desirability assigned to every protein we developed for prioritising proteomic hits for this particular project.

We decided to follow up three RIME hits, MTA2, CDK6 and CDK1, to begin with. Our choice was based on a combination of factors: novelty of the protein candidate for TNBC tumourigenicity, and biological meaning. MTA2, for example, which is part of the Mi-NuRD complex, is known to have an important role in ERα inhibition and EMT (Mazumdar et al., 2001; Cui et al., 2006). However, Mi2/NuRd is involved in many other processes: it is a chromatin-remodelling complex combining multiple

transcriptional regulatory events, such as histone deacetylation, histone demethylation, nucleosome mobilization and recruitment of regulatory proteins. It can promote or repress transcription, on the basis of the cellular context and of the protein subunits forming it.

Regarding CDK6 and CDK1, their importance for cell cycle progression and their oncogenic roles are well established, also in breast cancer: CDKs are in fact responsible for the expression of genes by direct phosphorylation of G1/S activators (van den Heuvel et Dyson, 2008; Henley et Dick, 2012), or by their subsequent inhibition during S-phase (Bertoli et al., 2013). CDK6, together with CDK4, is responsible for entering to the cell cycle from quiescence, in collaboration with D-type cyclins and cyclin E/CDK2 (Malumbres et Barbacid, 2009). CDK2/cyclin A and CDK2/cyclin E complexes are active in S phase and beyond, while CDK1/cyclin B complexes are responsible for the final step into mitosis. It has been shown that mammalian cells require at least five CDKs to regulate interphase: CDK2, CDK3, CDK4, and CDK6, and finally CDK1 in mitosis. However, studies in mouse models have shown that mice can survive lacking the interphase CDKs (Malumbres et al., 2004; Ortega et al., 2003), since CDK1 can execute all the events necessary to drive cell division, but not the absence of this one, suggesting that for many cell types it is the only essential CDK (Santamaria et al, 2007).

In this chapter, we investigate and validate if MTA2, CDK6 and CDK1 are *bonafide* transcription regulators of *FOXC1*, *NFIB* and *NFE2L3,* and that if they play a role in the progression or maintenance of TNBC cells.

## 5.2 Localization of MTA2, CDK1 & CDK6 DNA binding

First, we investigated the expression of *MTA2*, *CDK6* and *CDK1* in a panel of TNBC cell lines, and found that they are expressed in most of them (Fig. 5.1).

Furthermore, we took into consideration the patient data set TCGA (The Cancer Genome Atlas), which has evaluated gene expression, DNA methylation and DNA copy number (CN) variation data from more than 800 patients. Pathologically, we found that the three candidates' high expression correlates with TNBC, in particular for CDK6 and CDK1, while MTA2 shows some similarities with the *HER2*$^{+}$ breast cancer (Fig. 5.2).

We then proceeded with the validation of the localization of the protein candidates within the potential promoter regions of *FOXC1*, *NFIB* and *NFE2L3*. As a first step, we performed ChIP-qPCR on every protein, and we assessed a 120bp sequence of DNA flanking the gRNA targeting site. Because we were looking for potential fundamental players in the tumourigenicity of TNBC, we took into consideration four different cell lines. In Fig. 5.3 the results of this investigation are presented.

There is general variability between the cell lines in terms of DNA enrichment of the proteins' pulldowns across the three different genes. However, *FOXC1* promoter seems to be characterized by the most significant overall binding signal of the three proteins, in particular for MTA2, followed by CDK6. Interestingly, CDK1 was not associated with *NFE2L3* gene in our proteomic dataset, but we couldn't completely exclude it according to these ChIP-qPCR data.

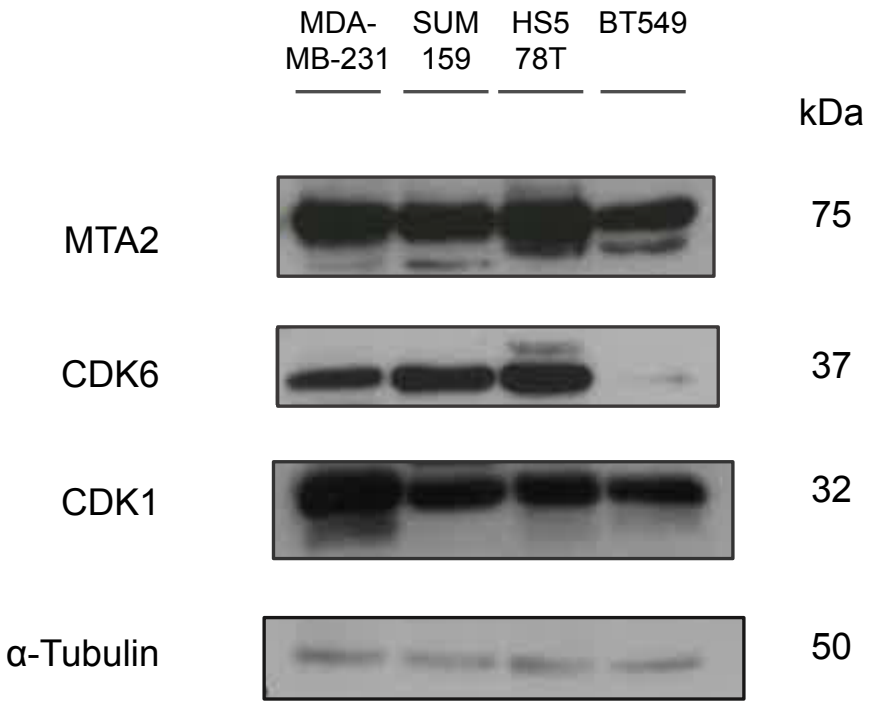



**Figure 5.1: *MTA2*, *CDK6* and *CDK1* expression in a panel of TNBC cell lines.** Cells were seeded in order to reach confluency the day after for collection and lysation. 50µg of protein lysates were probed by WB for the expression of MTA2, CDK6, CDK1 and α-Tubulin (loading control).

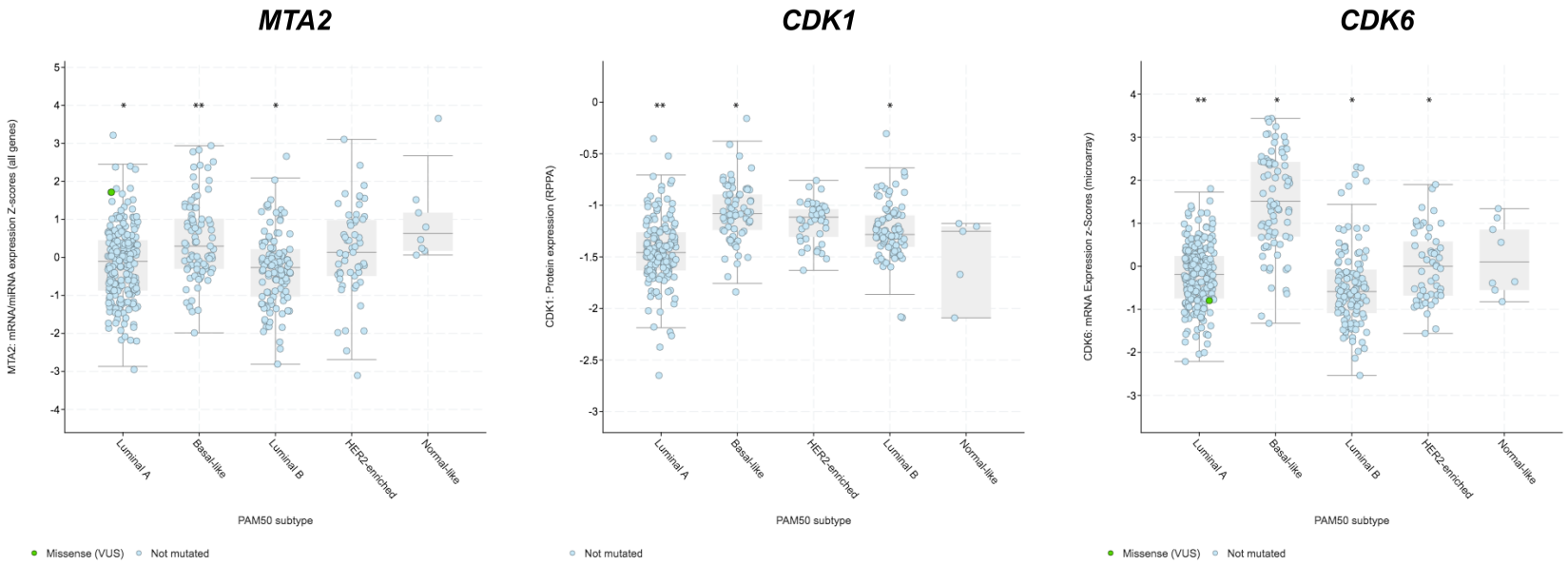



**Figure 5.2: *MTA2, CDK1 and CDK6* expression across the five molecular subtypes of breast cancer ('Normal' refers to the PAM50 subtype) in the The Cancer Genome Atlas (TCGA) dataset.** The plots show the differential expression of MTA2 (far right), CDK1 (middle) and CDK6 (far left) in comparison to other subtypes of breast cancer. The x-axis represents the PAM50 classification of the different breast cancer subtypes. The y-axis represents the mRNA level of expression according to the Z-score transformation of microarray data. For CDK1, mRNA microarray data were not available so we referred to RPPA data (reverse phase protein array). One-tailed P value was calculated and reported if <0.05.

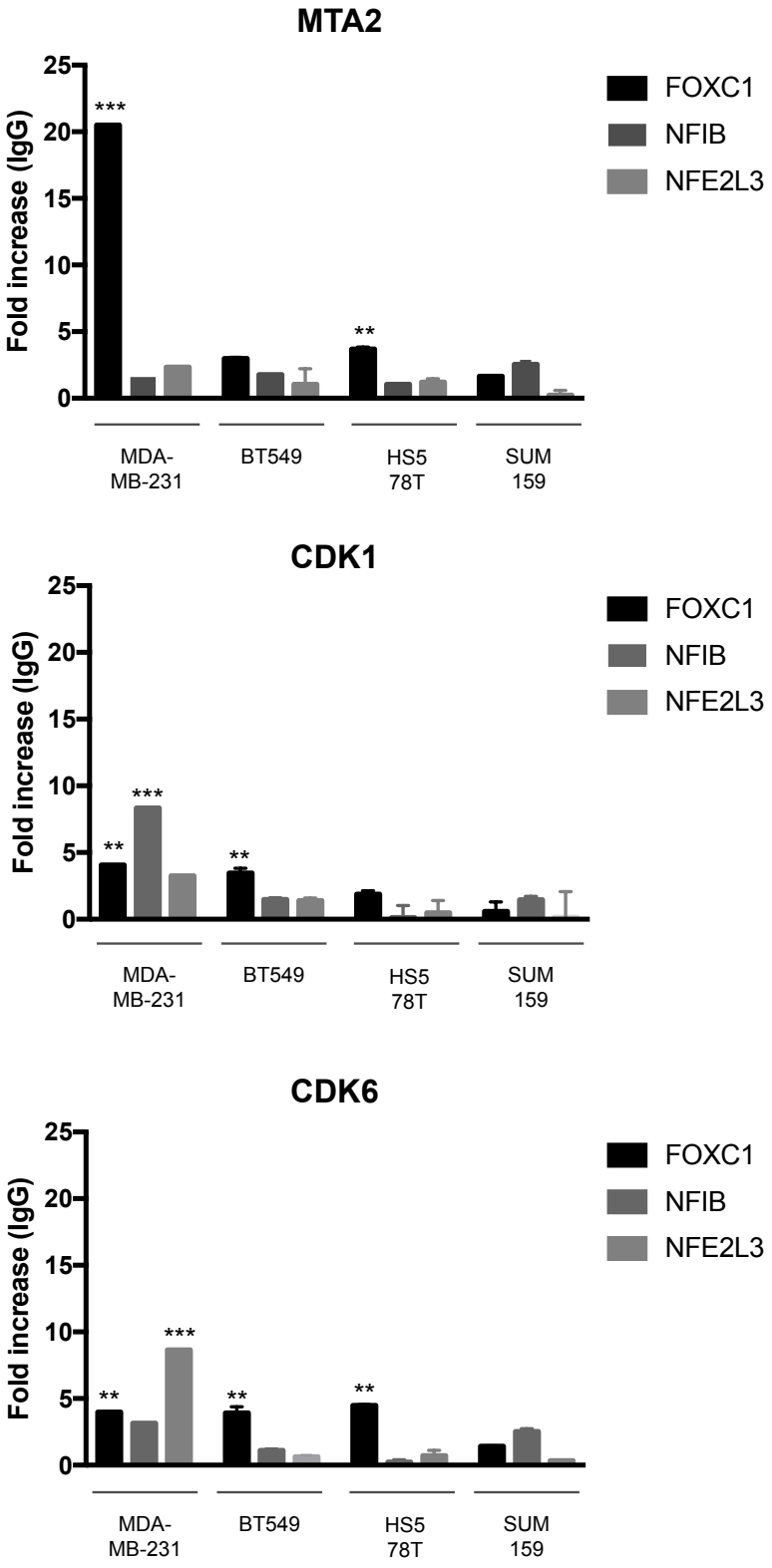



**Figure 5.3: MTA2, CDK1 & CDK6 ChIP-qPCR validation on the potential promoter sequence of *FOXC1*, *NFIB* and *NFE2L3* in a panel of TNBC cell lines.** The DNA enrichment of the MTA2, CDK1 & CDK6 for every cell line was evaluated in comparison to the respective internal IgG control. Primers for qPCR were designed in a region of 120bp flanking the gRNA of each respective promoter. The error bars report standard deviations from duplicates. Unpaired-t test was performed to compare the DNA enrichment to IgG: P value <0,05.

This experiment was extremely important for two reasons: it confirmed the recruitment of the proteins on the promoter sequences of some of the genes across multiple cell lines, and it confirmed the validity of the statistical approach combined with the ranking system. In addition, to further validate these results and to gain a global picture of where these putative transcription regulators bind on the genome, we decided to use a more powerful and accurate technique, and executed a ChIP-Seq experiment on MDA-MB-231 cell line. The ChIP-Seq analyses were performed by Dr Mike Firth and Dr Jonathan Cairns (Darwin building, AZ, Cambridge, UK).

In Fig. 5.4, the peaks identified for the three different immunoprecipitations are presented as heat maps. The results for MTA2 and CDK1 show a clear identification of wide and diverse genomic regions where the proteins bind. A deeper analysis is being conducted in order to investigate where else these proteins are interacting with the DNA, and if they could potentially be involved in the regulation of the transcription of other genes.

On the other hand, the results for CDK6 show that there were no peaks detected across the genome: this is likely to be caused by the low performance of the antibody used for sequencing. For this reason, we had to exclude CDK6 sequencing results from further analyses.

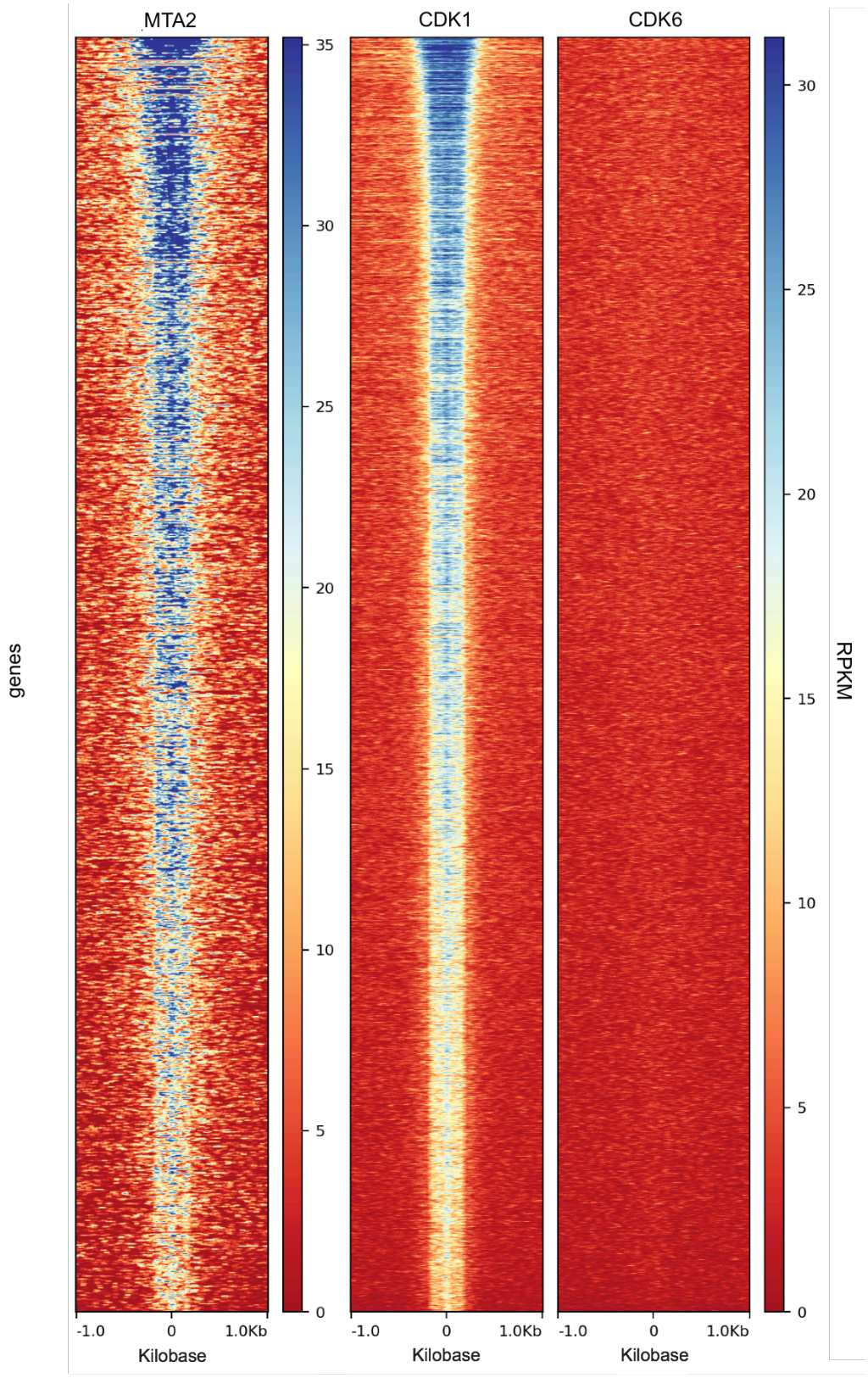



**Figure 5.4: Heat maps showing MTA2, CDK1 and CDK6 binding sites across the MDA-MB-231 cell line genome.** ChIP-Seq experiments were performed on MTA2, CDK1 and CDK6 pulldowns in order to identify their binding sites across the genome of MDA-MB-231 cell line. Total number of peaks and their intensity are shown in a horizontal window of ±1Kb. Analyses were performed by Mike Firth and Jonathan Cairns (AZ, Cambridge, UK).

In Fig. 5.5 the bindings of these proteins on the promoter sequences of *FOXC1*, *NFIB* and *NFE2L3* are shown, respectively. From a preliminary analysis it is possible to see how MTA2 and CDK1 seem to be both recruited to the putative promoter regions of *FOXC1*, *NFIB* and *NFE2L3* towards *loci* relatively close to each other. This could be an indication of a potential collaboration of these two proteins in regulating the transcription of this gene, through a direct, or indirect, interaction.

In order to understand if MTA2 and CDK1 would actually work together regulating the transcription of our genes of interest, we analysed those regions of the DNA where both proteins bind according to the ChIP-Seq dataset (Fig. 5.6, Appendix H). For this preliminary investigation, we looked at peaks overlapping within a specific and narrow range of DNA. Even though we couldn't identify any directly overlapping sequence across the regulatory regions of our targeted genes, we couldn't exclude the possibility of these two proteins collaborating to regulate the expression of the same downstream gene.

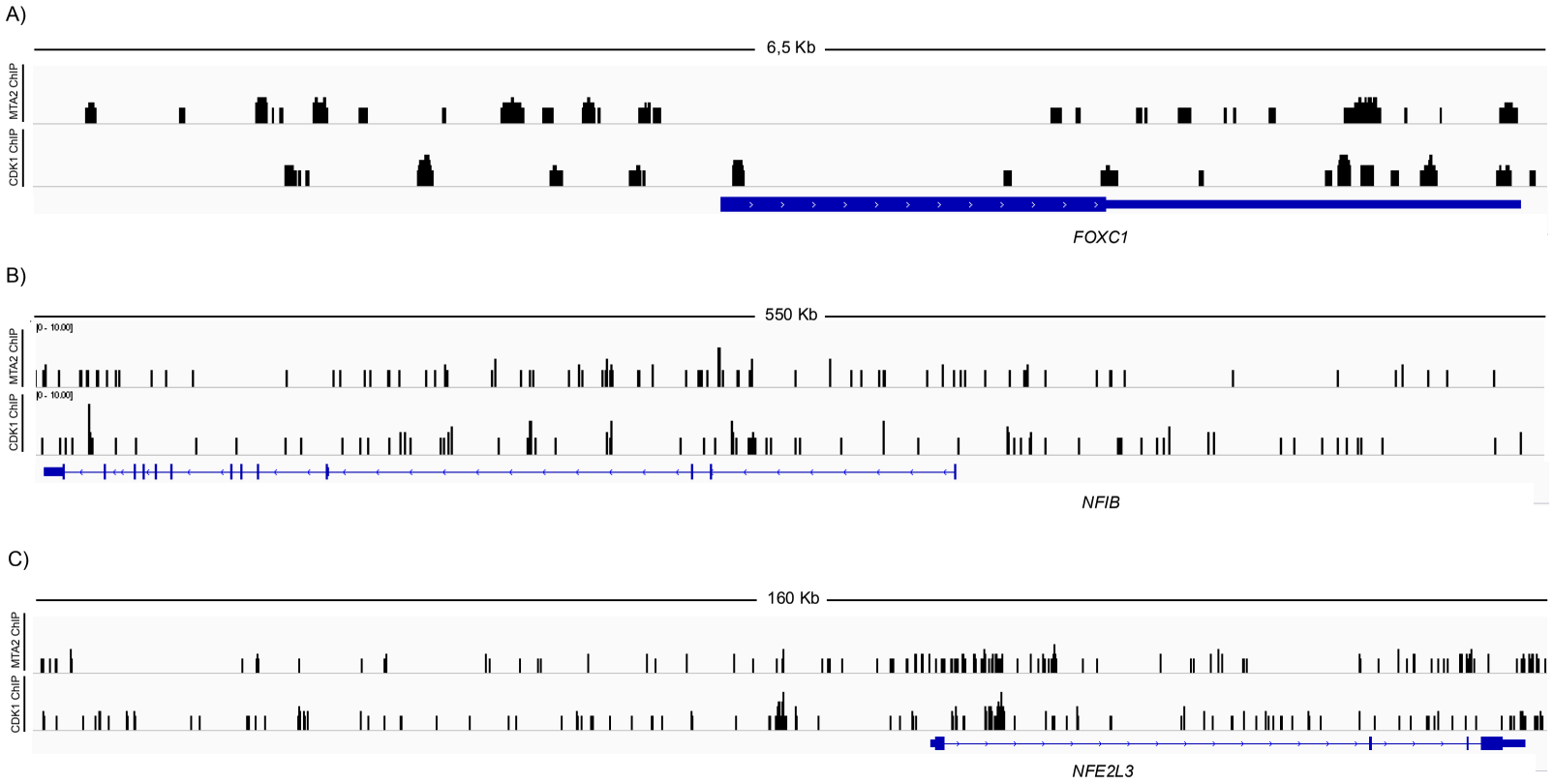



**Figure 5.5: IGV genome browser visualisation of different accessible peaks annotated for *FOXC1* (A), *NFIB* (B) and *NFE2L3* (C) after MTA2 and CDK1 ChIP-Seq data analyses.** Reads were aligned to reference genome. Peaks were called by MACS, and analysis was performed by Mike Firth and Jonathan Cairns, AZ, Cambridge, UK.

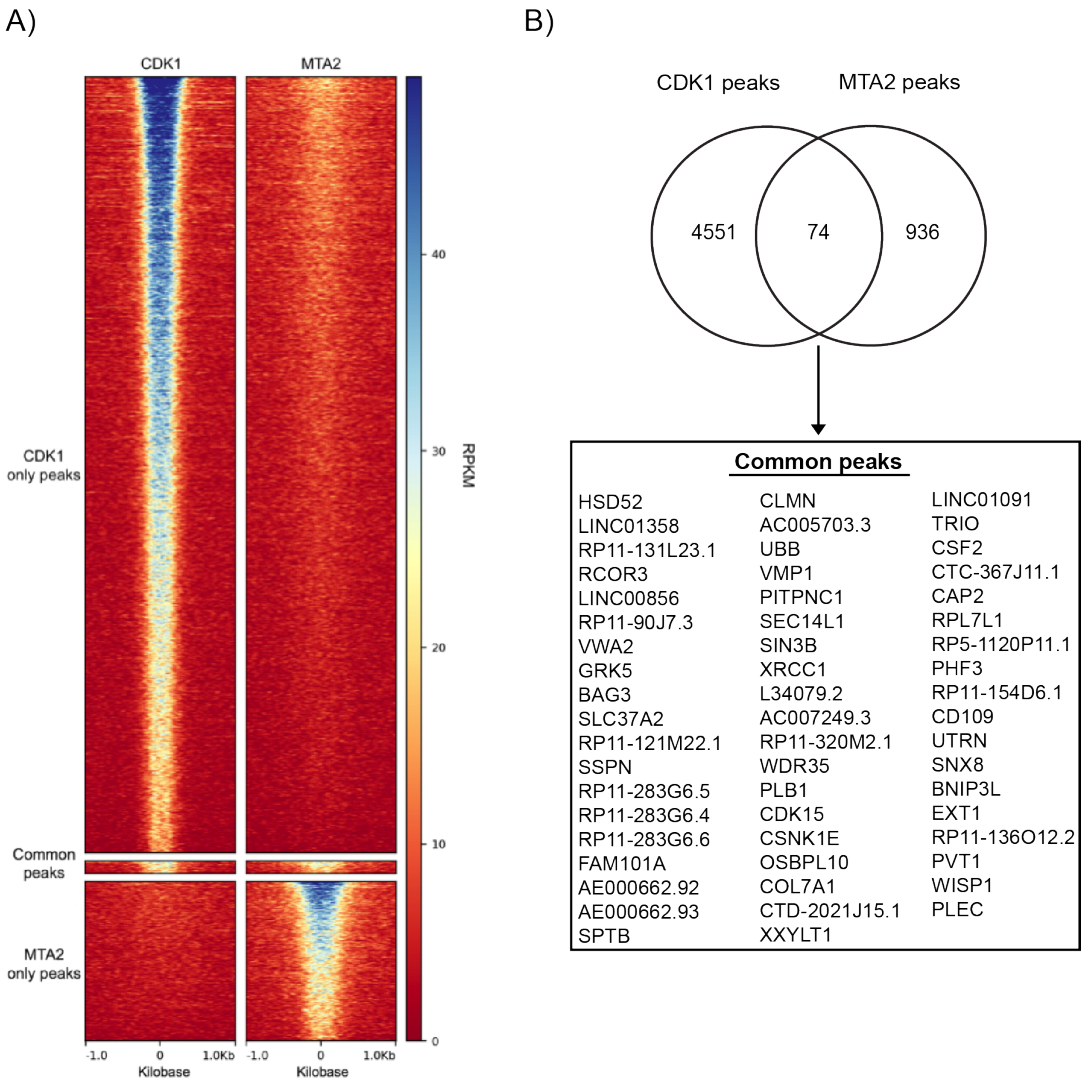



**Figure 5.6: Unique and common CDK1 and MTA2 binding sites across the MDA-MB-231 cell line genome.** MTA2 and CDK1 ChIP-Seq data were compared in order to identify those binding sites that are unique for each protein (CDK1 only peaks and MTA2 only peaks, respectively), and those that are common (Common peaks), across the genome of the MDA-MB-231 cell line. A) Heat maps showing CDK1 only, MTA2 only or common peaks in CDK1 and MTA2 IPs. Total number of peaks and their intensity are shown in a horizontal window of ±1Kb. B) Venn diagram indicating the common targeted genes of CDK1 and MTA2. Analyses were performed by Mike Firth and Jonathan Cairns, AZ, Cambridge, UK.

## 5.3 Functional validation of protein candidates

After investigating the potential localization of the protein candidates on the putative regulatory regions of *FOXC1*, *NFIB* and *NFE2L3,* we evaluated if they could be involved in the transcription regulation of these genes.

For this purpose, we performed shRNA-mediated knockdown of *MTA*, *CDK1* and *CDK6* in four different TNBC cell lines: each protein was targeted with two different shRNAs, and compared to the respective parental line transformed with a non-targeting siRNA (negative control). We confirmed and evaluated the success of the knockdown strategy at the mRNA (Fig. 5.7) and protein level (Fig. 5.8).

The data confirmed the efficacy of the knockdown for all the proteins, in particular at the mRNA level, even though some variability was reported in terms of efficiency between the cell lines, protein targeted and shRNA used. Overall, we observed a range of reduction between 60 and 90% for *MTA2*, between 40 to 90% for *CDK1*, and between 20-90% for *CDK6* at the mRNA level. Only the knockdown of *CDK6* for the HS578T cell line didn't seem to work as well as for the other cell lines at the mRNA level. However, it is possible to notice some variability among cell lines in the knockdown efficacy between two shRNAs targeting the same protein. This could be due to a combination of several reasons: the presence of a large number of off-targets for a specific shRNA, which would cause a 'dilution effect' on its activity (Arvey et al., 2010); the turnover rate of the targeted mRNA (Larsson and colleagues demonstrated that short-lived mRNAs are more insensitive to be downregulated (Larsson et al., 2010)), and the different abundance of the mRNA levels among cell lines (the more abundant the mRNA, the higher the gene-silencing effect (Hong et al., 2014). This experiment has to be repeated in order to corroborate the results.

Furthermore, we evaluated if the reduction of the three proteins had a direct effect on the expression of *FOXC1*, *NFIB* and *NFE2L3* at a gene transcription level by RT-PCR. Results are shown in Fig 5.9 (*MTA2* knockdown), Fig. 5.10 (*CDK1* knockdown) and Fig. 5.11 (*CDK6* knockdown).

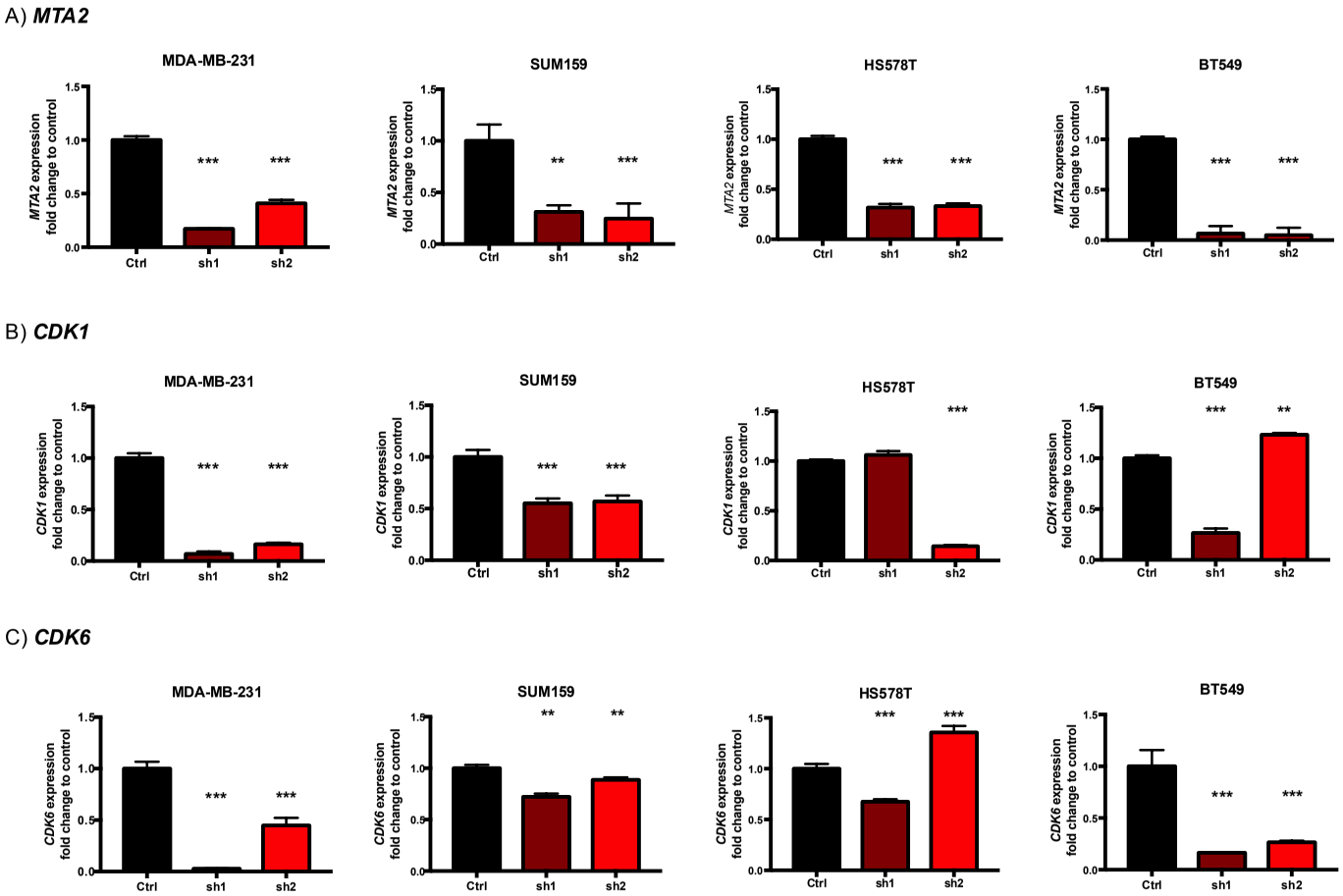
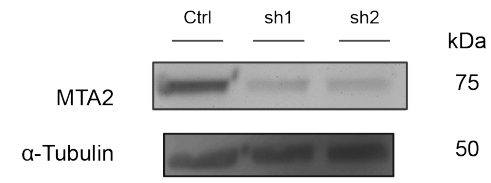


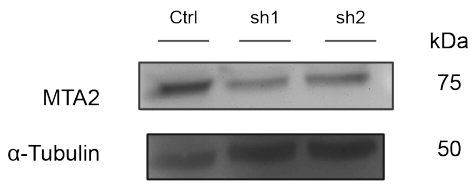

**Figure 5.7: *MTA2*, *CDK1* and *CDK6* knockdown mRNA level in a panel of TNBC cell lines.** The expression of *MTA2* (A), *CDK1* (B) and *CDK6* (C) was evaluated in the MDA-MB-231, SUM159, HS578T and BT569 transformed with the control (ctrl), sh1 and sh2 shRNAs for the three respective proteins. mRNA expression levels were determined by real-time PCR and normalized to *GAPDH*. The error bars report SD from triplicates. One-way ANOVA test was performed. P value<0.05.

## 1. MTA2 KD

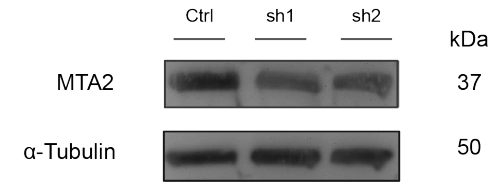
A) MDA-MB-231



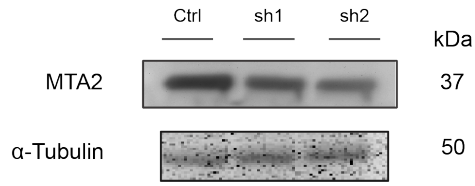

B) SUM159




C) HS578T

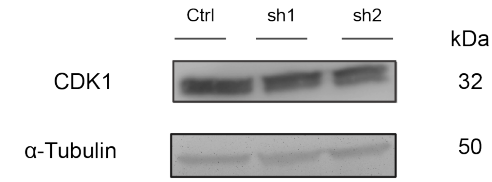



D) BT549

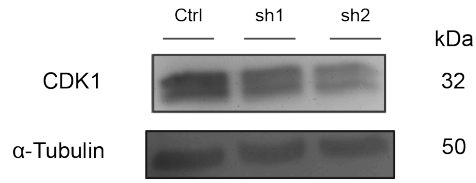



## 2. CDK1 KD

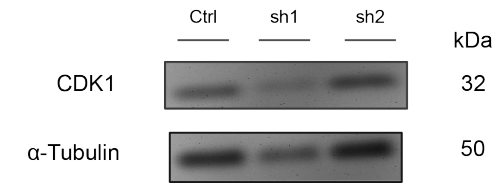
A) MDA-MB-231



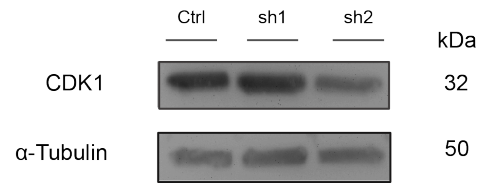

B) SUM159




C) HS578T




D) BT549



**3. CDK6 KD**

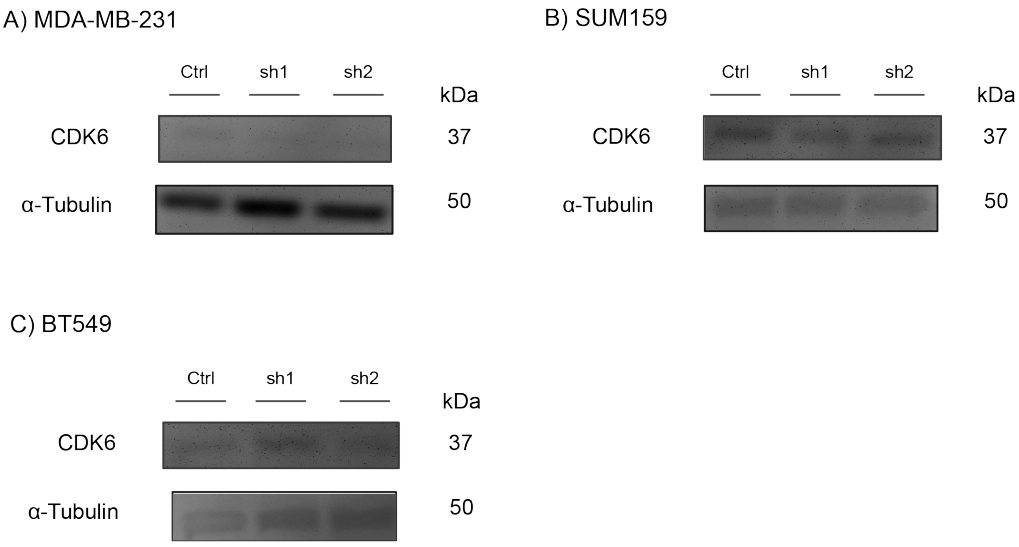



**Figure 5.8: Representative MTA2, CDK1 and CDK6 knockdown protein levels in a panel of TNBC cell lines.** MTA2, CDK1 and CDK6 protein levels were evaluated in the MDA-MB-231, SUM159, HS578T and BT569 transformed with the control (ctrl), sh1 and sh2 shRNAs for the three respective proteins. Cells were lysed and 30µg of protein lysate were probed by WB. α-Tubulin was used as a loading control.

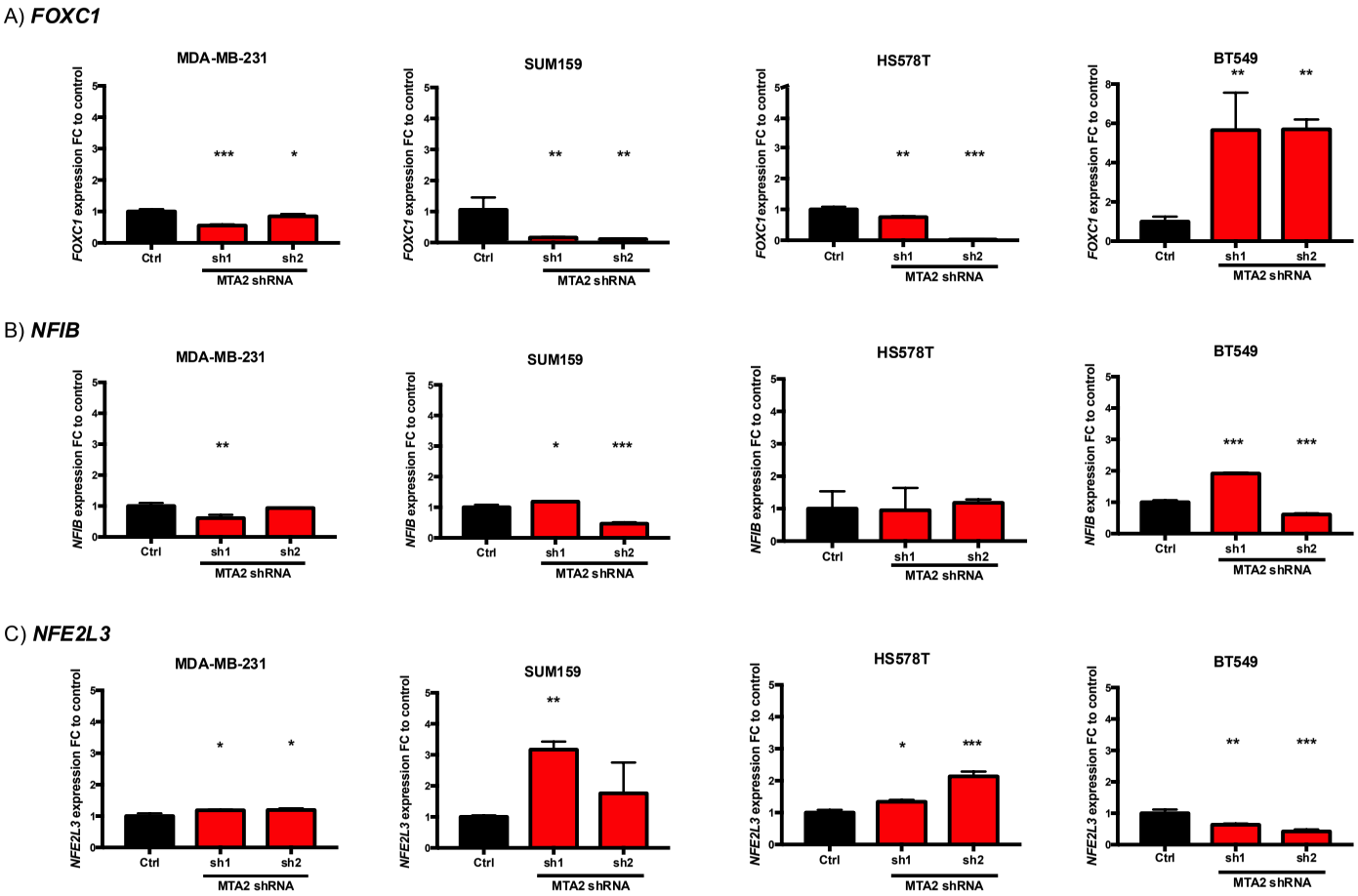



**Figure 5.9: Effect of *MTA2* knockdown on *FOXC1*, *NFIB* and *NFE2L3* gene expression in a panel of TNBC cell lines.** The expression of *FOXC1* (A), *NFIB* (B) and *NFE2L3* (C) was evaluated in the MDA-MB-231, SUM159, HS578T and BT569 transformed with the control (ctrl), sh1 and sh2 shRNAs. mRNA expression levels were determined by real-time PCR and normalized to *GAPDH*. The error bars report SD from triplicates. One-way ANOVA test was performed. P value<0.05.

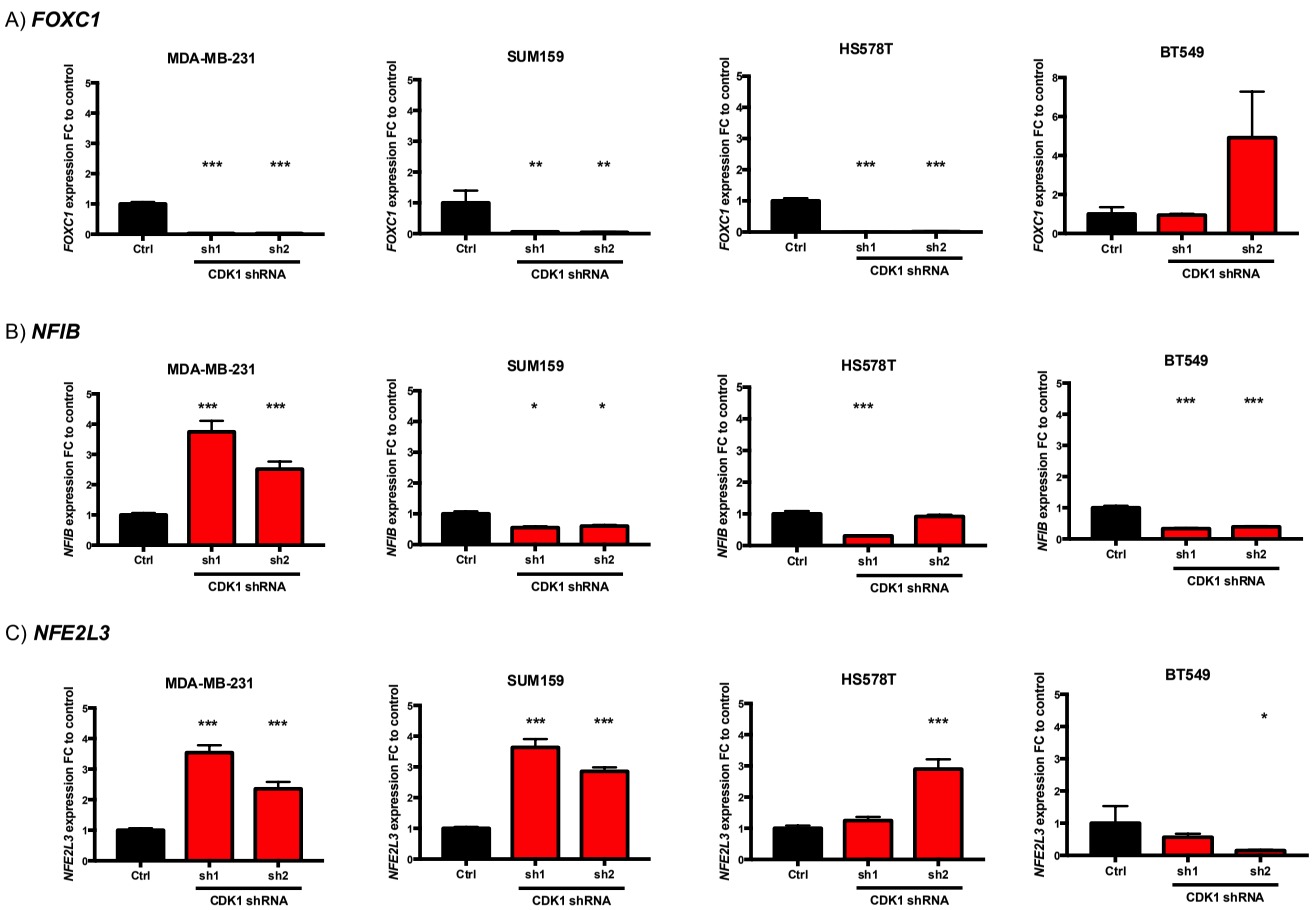



**Figure 5.10: Effect of *CDK1* knockdown on *FOXC1*, *NFIB* and *NFE2L3* gene expression in a panel of TNBC cell lines.** The expression of *FOXC1* (A), *NFIB* (B) and *NFE2L3* (C) was evaluated in the MDA-MB-231, SUM159, HS578T and BT569 transformed with the control (ctrl), sh1 and sh2 shRNAs. mRNA expression levels were determined by real-time PCR and normalized to *GAPDH*. The error bars report SD from triplicates. One-way ANOVA test was performed. P value<0.05.

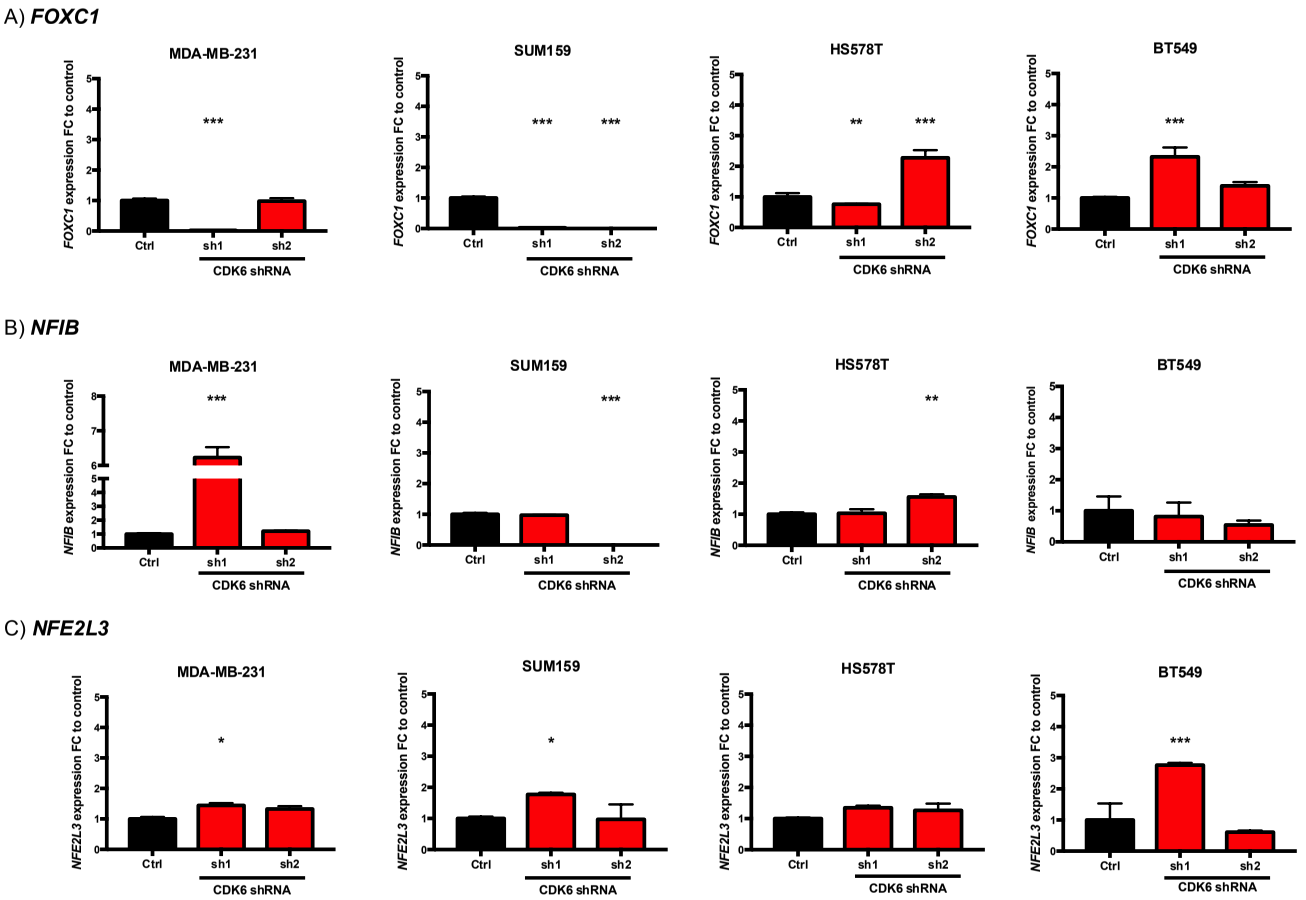



**Figure 5.11: Effect of *CDK6* knockdown on *FOXC1*, *NFIB* and *NFE2L3* gene expression in a panel of TNBC cell lines.** The expression of *FOXC1* (A), *NFIB* (B) and *NFE2L3* (C) was evaluated in the MDA-MB-231, SUM159, HS578T and BT569 transformed with the control (ctrl), sh1 and sh2 shRNAs. mRNA expression levels were determined by real-time PCR and normalized to *GAPDH*. The error bars report SD from triplicates. One-way ANOVA test was performed. P value<0.05.

The results show a potential role for MTA2, CDK1 and CDK6 in the regulation of the expression of *FOXC1*, *NFIB* and *NFE2L3* for all the cell lines analysed. In fact, when knocked down, their absence causes an altered expression of these downstream genes.

However, it is possible to see how the four cell lines are characterized by different behaviours to the knockdowns. Among them, *MTA2* knockdown had the least significant effect on transcription of our target genes (Fig. 5.9). While in contrast, *CDK1* knockdown had the most effect on our genes (Fig. 5.10). Interestingly, CDK1 seems to positively regulate the expression of *FOXC1*, and to downregulate the *NFE2L3* one, while CDK6 seems to be involved in the upregulation of the transcription of all genes in almost all the cell lines.

All these experiments were performed shortly after selection (between 5-14 days, on the basis of the cell line), and this could be a possible explanation for the observed variability: some cell lines were more sensitive than others, showing some toxicity effect to the knockdown of these proteins (e.g. arrest of cell growth and cell death), or they could have activated different alternative transcription pathways as a compensatory mechanism. In addition, the efficacy of the knockdown varied between cell lines, which could have also influenced the expression of the downstream genes. These evaluations were performed just once when this thesis was written, but they require replication in order to confirm the results obtained.

Next, we investigated if these proteins play a role in breast cancer, in particular in TNBC growth and maintenance. In order to do that, we firstly evaluated the dataset available at www.depmap.org to identify perturbation effects of *MTA2*, *CDK1* or *CDK6* knockdown (Fig. 5.12). For this purpose, we used the DEMETER2 dependency score from different cell depletion assays, where every gene gets assigned a score of 0 if not essential, and the lower the score, the higher the probability that the gene is a fundamental one in a given cell line. The score -1 is identified as the median of all pan-essential genes. As shown in Fig. 5.12, B, *CDK1* seems to be an essential gene in breast cancer, while *MTA2* and *CDK6* depletion don't seem to be as critical for this tumour.

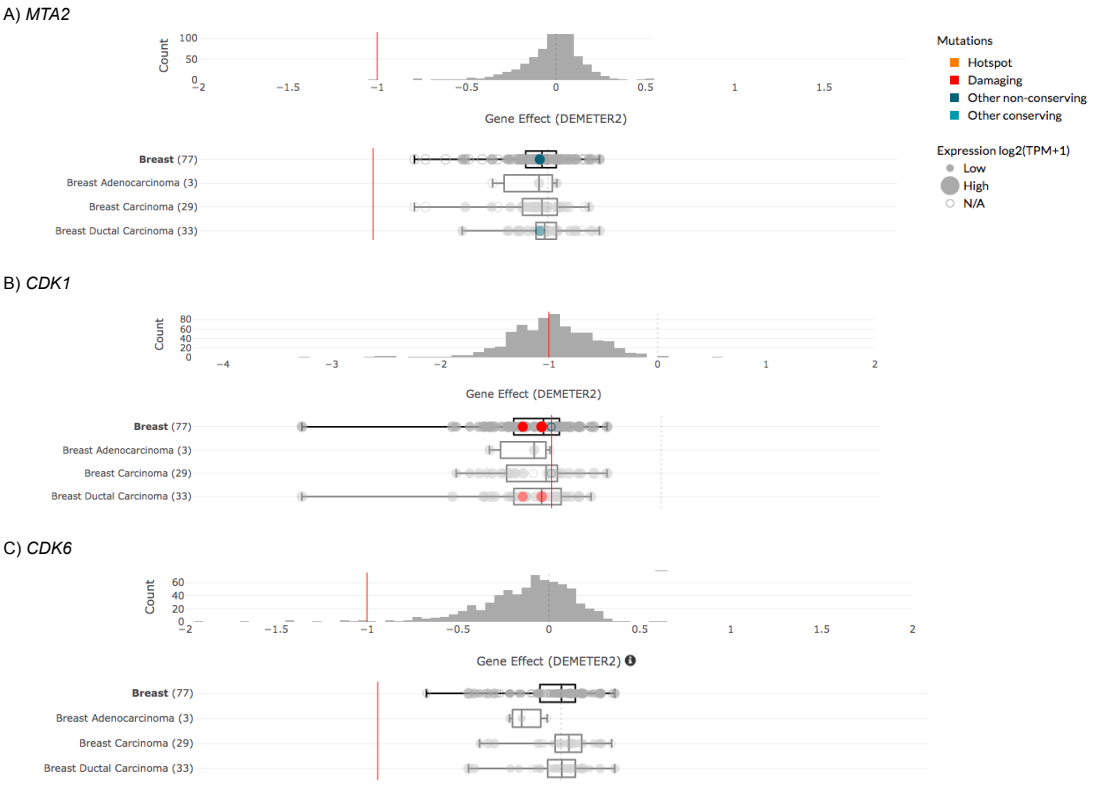



**Figure 5.12: Perturbation effects of *MTA2* (A), *CDK1* (B) and *CDK6* (C) knockdown in a combined RNAi study (Broad, Novartis, Marcotte) in breast cancer.** For each gene, the outcome from DEMETER2 dependency score from different cell depletion assays is reported. Score 0: not essential gene. Lower scores: higher probability that the gene is essential in a given cell line. Score -1 (red line): median of all pan-essential genes. Hotspot mutations: non-silent mutations hotspot in TCGA dataset; damagind mutations: mutations at start codon, or causing a frame shit, or a premature stop; other-non conserving: missense mutations; other-conserving: mutations in non coding regions. Data available at www.depmap.org.

Additionally, we performed a 3D matrigel colony formation assay experiment to evaluate if the ability of these cell lines to form colonies *in vitro* was affected by the absence of any of the protein candidates. The results, shown in Fig. 5.13, gave us information about the clonogenic ability of TNBC cells in the absence of MTA2, CDK1 and CDK6.

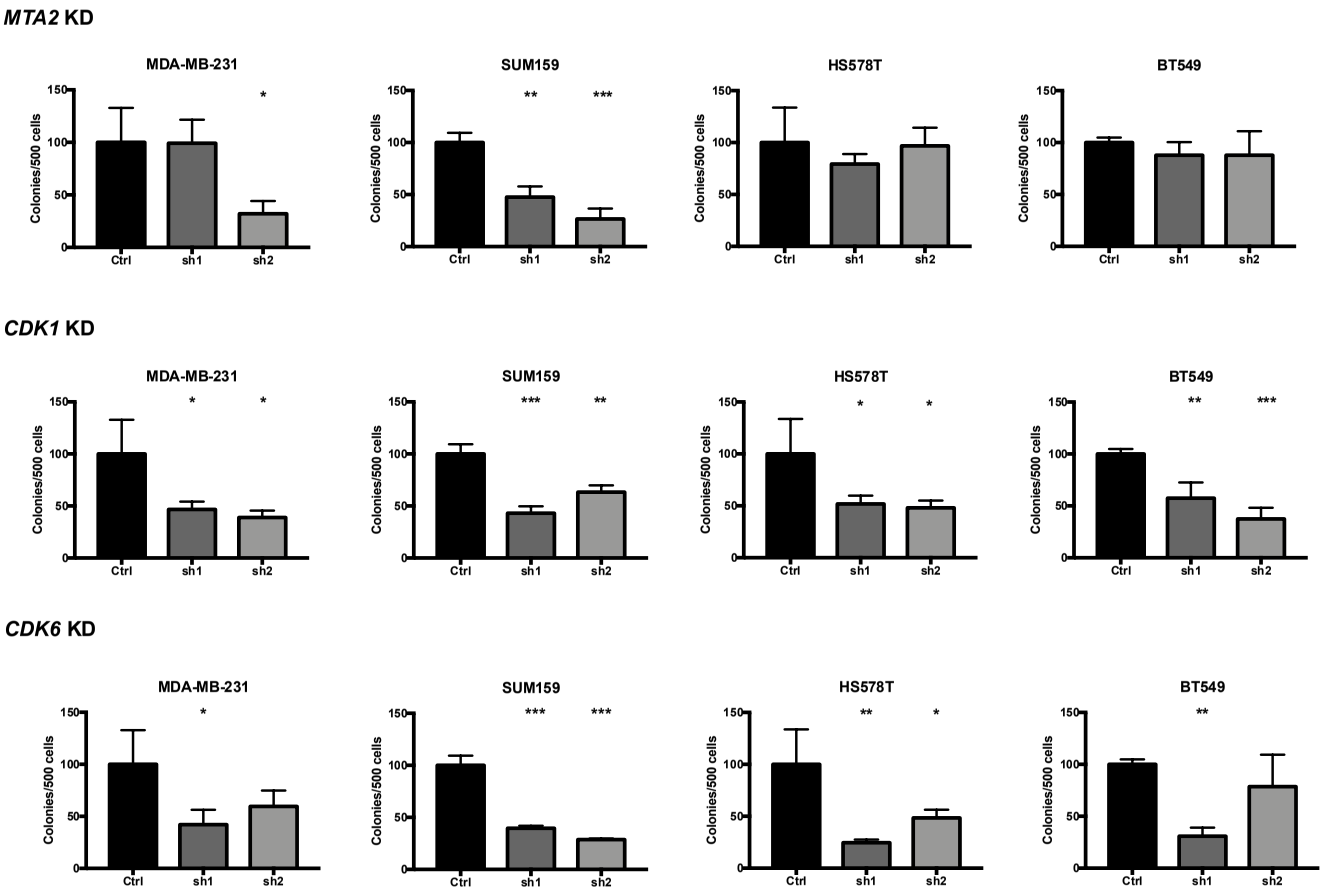



**Figure 5.13: 3D matrigel colony formation assay after *MTA2*, *CDK1* and *CDK6* knockdown in a panel of TNBC cell lines.** Graph depicting 3D matrigel assay in *MTA2*, *CDK1* and *CDK6* knockdown (KD) in MDA-MB-231, SUM159, HS578T and BT569 cell lines transformed with the respective shRNAs. Colonies were counted 7 days after seeding the cells, and normalized to the respective negative control. The error bars report SD from triplicates. One-way ANOVA test was performed. P value<0.05.

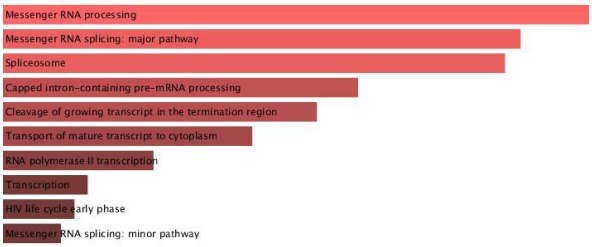
We observed that when cells are deprived of *MTA2*, their capacity of forming colonies is significantly affected for MDA-MB-231 and SUM159 (Fig. 5.13, A), while *CDK1* (Fig. 5.13, B) and *CDK6* (Fig. 5.13, C) seem to play an important role in this process for all the cell lines analysed. This suggests that these proteins could be involved and fundamental for the tumourigenicity of TNBC. It is important to note that due to time restriction, this experiment was performed once at the time this thesis was written, therefore, further validations are required in order to confirm the results obtained.

## 5.4 Future directions

It has to be kept in mind that the chosen proteins were just some of the candidates identified with this technology. Many other proteins could be important gene transcription regulators, or have a fundamental role for the biology of TNBC, but due to time restriction while writing this thesis, it was not possible to investigate them further.

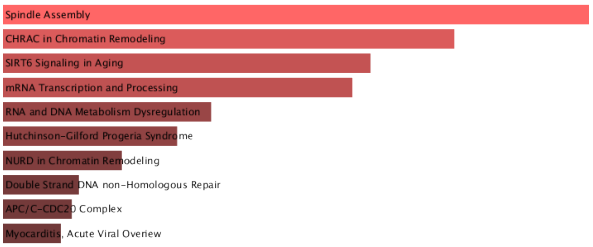
However, we performed a preliminary pathway analysis on the proteins identified through RIME proteomics associated with a statistical power (Appendix E) and those proteins identified exclusively on the dCa9 RIME samples (Appendix F). Using the Enrichr online tool (https://amp.pharm.mssm.edu/Enrichr), and in particular the comprehensive integrated and non-redundant pathway resources BioPlanet2016 and Elsevier pathway collection, we explored the potential biological mechanisms that could be regulated by these proteins. In Fig. 5.14 it is possible to see how the majority of them seemed to be involved in processes related to transcription, such as splicing, mRNA processing, mRNA transport (Fig. 5.14, A), DNA remodelling and mitotic checkpoint (Fig. 5.14, B). These results were expected, since we were aiming to target transcription-starting sites and to avoid interfering with the transcription process of the gene itself, and they further confirm the validity of our strategy. However, at this level of analysis we couldn't identify any pathway directly involved in the regulation of the gene expression.

A) BioPlanet 2016




| Index | Name | P-value | Adjusted p-value |
|---|---|---|---|
| 1 | Messenger RNA splicing: major pathway | 2.381e-12 | 1.798e-9 |
| 2 | Cleavage of growing transcript in the termination region | 6.605e-9 | 0.000001995 |
| 3 | Spliceosome | 4.368e-12 | 2.199e-9 |
| 4 | Messenger RNA processing | 1.658e-13 | 2.503e-10 |
| 5 | Integration of provirus | 0.0003661 | 0.05025 |
| 6 | Transport of mature transcript to cytoplasm | 8.208e-8 | 0.00002066 |
| 7 | HIV life cycle early phase | 0.00008338 | 0.01399 |
| 8 | Capped intron-containing pre-mRNA processing | 1.333e-9 | 5.032e-7 |
| 9 | APOBEC3G-mediated resistance to HIV-1 infection | 0.003538 | 0.3816 |
| 10 | Ran role in mitotic spindle regulation | 0.0007623 | 0.09592 |

B) Elsevier Pathway Collection




| Index | Name | P-value | Adjusted p-value |
|---|---|---|---|
| 1 | Spindle Assembly | 0.0002028 | 0.3490 |
| 2 | CHRAC in Chromatin Remodeling | 0.005240 | 1.000 |
| 3 | SIRT6 Signaling in Aging | 0.0005080 | 0.4371 |
| 4 | mRNA Transcription and Processing | 0.001741 | 0.9985 |
| 5 | RNA and DNA Metabolism Dysregulation | 0.003266 | 1.000 |
| 6 | Hutchinson-Gilford Progeria Syndrome | 0.003909 | 1.000 |
| 7 | NURD in Chromatin Remodeling | 0.005415 | 1.000 |
| 8 | Double Strand DNA non-Homologous Repair | 0.007225 | 1.000 |
| 9 | APC/C-CDC20 Complex | 0.003525 | 1.000 |
| 10 | Myocarditis, Acute Viral Overiew | 0.008248 | 1.000 |

**Figure 5.13: Preliminary pathways analysis for protein candidates identified through RIME proteomics and statistical analysis.** A) BioPlanet 2016 database investigation for enriched pathways among protein candidates. Left side: bar chart visualisation of the 10 highest enriched pathways according to the database. Rigth side: p-value of the 10 highest enriched pathways according to the database. B) Elsevier Pathway Collection database investigation for enriched pathways among protein candidates. Left side: bar chart visualisation of the 10 highest enriched pathways according to the database. Rigth side: p-value of the 10 highest enriched pathways according to the database. Data available at https://amp.pharm.mssm.edu/Enrichr.

This could be potentialy implemented by a deeper and more sophisticated investigation of different databases and online tools that couldn't be performed while this thesis was written. However, it has to be kept in mind that RIME proteomics has some intrinsic limitations, and despite our effort to implement the technique, it could be that the identification of regulatory events that are spatially and temporally limitied is affected by the low sensitivity and the low abundance of the signal.

## 5.5 Discussion

We demonstrated in this chapter the validity of the combination of CRISPR/Cas9 and RIME proteomics as a method to identify novel transcription regulators. We showed how our statistical approach, together with a powerful ranking system, could be used to highlight potentially successful hits for validation among large proteomics datasets.

In particular, we showed how the three candidates we identified in our project, MTA2, CDK1 and CDK6, correlate with TNBC patients data among the different subtypes of breast cancer, and they are recruited to the promoter sequences of the investigated genes *FOXC1*, *NFIB* and *NFE2L3*. In addition, thanks to ChIP-Seq analysis, we were able to identify many more genes regulated by these proteins. In order to complete our investigation, we need to optimize and repeat ChIP-Seq on CDK6 on MDA-MB-231: this would further enrich our dataset and possibly highlight more transcription pathways shared by these three proteins.

Our data reveal how these proteins could be actively regulating the transcription of key TNBC oncogenes. Through functional assays, we showed how their absence causes an altered expression of the downstream genes in a panel of TNBC cell lines. Despite the variability we observed, the results support a direct role of MTA2, CDK1 and CDK6 on gene transcription regulation. In particular, it is possible to speculate a potential repressing role for CDK1 and CDK6: in fact, when they are downregulated, the transcription of downstream genes is enhanced. This could be explained by the function of CDKs themselves: thanks to their phosphorylation activity, they are able to regulate the status of other proteins, for example from an inactive status to and active one (van den Heuvel et Dyson, 2008; Henley et

Dick, 2012). It could be that CDK1 and CDK6 are recruited within the promoter region of our genes of interest to regulate the activity of some target transcription factors, modulating their effect on the gene transcription. Their absence causes a lack of this regulation, therefore an enhanced expression of the downstream gene. Further investigations should be carried out in order to confirm this hypothesis and to identify the direct target of the phosphorylation activity of these CDKs.

Similarly, the absence of MTA2 caused a downregulation of the genes expression. This could be related to a change of function in the NuRD complex, of which MTA2 is a member (Bowen et al., 2004). It is in fact known that different components, as well as different tissues, can modify the role of the complex in terms of gene regulation (Manavathi et Kumar, 2007): in our case, it could be that MTA2 worked as a transcription activator for the genes of interest, and its absence promoted a decrese in gene expression. However, further experiments would need to be performed in order to confirm the recruitment of the NuRD complex on the promoter sequences, and the role of MTA2 on transcription regulation through it rather than other complexes or protein-protein interactions.

Furthermore, we demonstrated the importance of CDK1 and CDK6 for the tumourigenicity of TNBC. In fact, the ability of cell lines to form colonies *in vitro* seemed to be affected by the lack of these proteins. Over the past two decades, several works have illustrated that the dysregulation of CDKs affects tumor growth and cell proliferation (Malumbres et Barbacid, 2009). In addition, it is well known that CDKs are involved in many other processes like DNA damage repair, epigenetics, stemness, metabolism and transcriptional functions (Lim et Kaldis, 2013), indicating broader roles. In particular for breast cancer, recent studies have discovered that CDK4 and 6 also contribute to cancer stemness (Dai et al., 2016). Our results corroborate with the literature, showing how CDK1 inactivation significantly affects cell growth, probably inducing apoptotic mechanisms, as published (Goga et al., 2007; Johnson et al., 2009).

On the other hand, CDK6 has recently gained a lot of attention because of the potential efficacy of the blockade of cyclin D-CDK4/6 pathway as a therapeutic strategy for breast cancer (Arnold et Papanikolaou; 2005; Yu et al., 2001; Yu et al., 2006), which seems to induce a phenotype similar to cellular senescence (Sharpless et Sherr, 2015). However, it has mainly been associated with $ER^+$ breast cancer treatment.

TNBC has in fact been considered a poor candidate for CDK4/6 inhibitor therapy because of the frequent loss of expression of the RB protein (Herschkowitz et al., 2008) or high expression of cyclin E, two mechanisms able to confer resistance to a CDK4/CDK6 inhibition. In addition, it has been shown that many TNBC cell lines are insensitive to CDK4/6 blockade *in vitro* (Finn et al., 2009). Interestingly, we observed an effect of CDK6 inhibition in our preliminary analysis.

To support our observation, recent studies have demonstrated a potential TNBC sensitivity to CDK6: Asghar and colleagues (Asghar et al., 2017) showed that the luminal androgen receptor (LAR) subtype is affected by this treatment *in vitro* and *in vivo,* and they observed an increased CDK2 activity as a possible escaping mechanism. In addition, the simultaneous blockade of CDK4/6 and PI3K has been shown to have an effect in a variety of TNBC models (Asghar et al., 2017; Teo et al., 2017). Furthermore, it has been observed that in TNBC preclinical models the inhibition of CDK4/6 can block breast cancer metastases: Liu and colleagues in fact reported how CDK4/6-mediated activation of DUB3 (Deubiquitinating Protein 3) is essential to stabilize SNAIL1, a transcription factor involved in the EMT process (Liu et al., 2017).

Overall, these results provide rationale for further investigations on MTA2, CDK1 and CDK6 in TNBCs, in order to understand and identify downstream regulated pathways or potential resistance mechanisms.

## 5.6 Conclusion

Results presented in this chapter demonstrated the efficacy of the new methodology we developed to investigate transcription regulation. In addition, the identified proteins seem to be important for the expression of *FOXC1*, *NFIB* and *NFE2L3,* and their targeting causes alteration to cancer cell clonogenic capacity. In particular, MTA2, CDK1 and CDK6 appear to be extremely important for the biology and tumourigenicity of TNBC, making them interesting candidates for further investigations.

# CHAPTER 6: DISCUSSION

Breast cancer development and progression are characterized by several genetic and epigenetic alterations of normal and host cells interacting with the developing tumour, such as immune, vascular and stromal cells. Whether inherited or not, changes like gene loss, amplification and point mutations would normally lead to cell death, but if affecting key genes, they could promote cell survival, proliferation, invasiveness and resistance (oncogenes).

In breast cancer, loss of heterozygosity and copy number alteration seem to be involved in the transition from hyperplasia to ductal carcinoma *in situ* (DCIS) (Waldman et al., 2000). The increasing knowledge of these changes and their associated pathways has led to the development of targeted therapeutics. Several successful examples have been reported, like tamoxifen for ER-dependent breast cancers, and trastuzumab, for *HER2*$^{+}$ tumours, but TNBC still miss a unique therapeutic approach. In the last decades, patients' outcome has been significantly improved by multi-drug combination systemic therapies in the neoadjuvant and adjuvant settings, and treatment advances have been achieved with poly (ADP-ribose) polymerase (PARP) inhibitors and immunotherapy agents. However, in some cases, the prognosis still remains poor.

Recently transcription factors have gained a lot of attention as potential therapeutic targets for breast cancer because of their essential role in gene expression regulation, despite the potential, intrinsic toxicity. In particular for TNBC, Wang and colleagues have demonstrated the sensitivity of this disease to inhibition of transcription, showing how tumours are dependent on cyclin kinase 7 (CDK7), and

how its selective inhibition with THZ1 has a direct effect on cells' tumourigenity (Wang et al., 2015). This result confirms the potential efficacy of targeting transcription as a therapeutic approach also for this disease. However, in order to do it, a deeper knowledge of the TNBC biology and gene transcription regulation is fundamental.
To understand how transcription regulators drive the TNBC aggressive phenotype, we developed the current project, where we applied a combination of CRISPR/Cas9 and RIME proteomics, and we demonstrated the potential of catalytically dead version of Cas9 protein (dCas9) to explore the regulation of the transcription of a specific gene of interest when coupled with discovery proteomics.

## 6.1 Applicability of CRISPR/Cas9 to target putative regulatory regions

By investigating the METABRIC dataset (~ 2000 patients) we identified highly and differentially expressed genes in TNBC encoding for transcription factors, and we further pursued three of them (*FOXC1*, *NFIB* and *NFE2L3*) with our analyses. All these genes are known to play important roles in TNBC, with the only exception of *NFE2L3*, which is novel for this type of cancer. In particular, *FOXC1* promotes cancer stem cell properties and can induce epithelial-mesenchymal transition (EMT) (Xia et al., 2013). On the other hand, *NFIB* is involved in several different processes like cell cycle regulation, apoptosis, cell adhesion (Persson et al., 2019) and lactation (Murtagh et al., 2003).

Experimentally, we were able to target putative regulatory region of these genes using dCas9. We optimised a protocol to obtain stable TNBC cell lines transfected with exogenous DNA coding for both dCas9 and gRNA targeting the regions of interest. We demonstrated here that dCas9 binds the desired sequences causing minimal interference with the expression of the gene in question thanks to its inducible expression, regulated by a Tet-On system.

Recently, several techniques have been developed to purify specific genomic regions and to analyse molecular interactions by insertion of the recognition sequences of an exogenous DNA-binding molecule. Among these techniques, several examples are worth highlighting such as: the iChIP technology (insertional ChIP), where proteins like LexA are used for affinity purification of targeted DNA

sequences (Hoshino et Fujii, 2009), or engineered DNA-binding molecules like zinc-finger proteins (Pabo et al., 2001) and transcription activator-like (TAL) proteins (Bogdanove et Voytas, 2011) to tag a specific genomic locus. However, the CRISPR system provides the most flexible and inexpensive way to target desired genomic regions. In particular a technique called enChIP (engineered DNA-binding molecule-mediated chromatin immunoprecipitation) has been optimised to purify genomic sequences of interest (Fujita et Fujii, 2013), immunoprecipitating them with antibody against a tag(s) fused to dCas9, which is co-expressed with a guide RNA (gRNA) and recognizes endogenous DNA sequence.

All these novel technologies further demonstrate the applicability of our strategy to investigate the regulation of the transcription of a gene of interest. Furthermore, they have been recently successfully coupled with analytical methods like mass spectrometry (enChIP-MS) (Fujita et Fujii, 2014; Hamidian et al., 2018), microarray analysis (enChIP-chip), or RNA-Seq (enChIP-RNA-Seq) (Fujita et al., 2015), to perform unbiased investigation in a genome-wide scale.

Nevertheless, our strategy has shown some intrinsic limitations. Firstly, we used exclusively cell lines to investigate TNBC tumour behaviour: despite being extremely versatile, they don't fully recapitulate the complexity of the tumour and of its interaction with the microenvironment. For this reason, some of the discrepancy we saw between the METABRIC analysis and our cell line investigation could be explained. Even though better models are available to overcome this diversity, like PDXs (patient derived xenografts, (Cassidy et al., 2015)) for example, and it has been shown it is possible to genetically modify them using CRISPR/Cas9 (Hulton et al., 2019), it would be extremely difficult to reach the number of cells required for a successful RIME experiment with the conditions we used.

Secondly, the CRISPR/Cas9 system used was characterized by some undesired expression of dCas9 without induction, due to the presence of some Tetracycline derivates in the FBS: despite a particular attention before some investigations, cells were usually kept in normal media conditions. The presence of dCas9, even though minimal, could have affected the phenotype of the cells over time. For this reason, a more efficient system could be used instead to prevent any uncontrolled presence of exogenous molecules.

In relation to this, the presence of off-targets has to be considered: despite every gRNA specificity, a number of off-site bindings of the dCas9 were identified. This was expected, since the targeted promoter regions are extremely repetitive across the genome (Jordan et al., 2003; Feschotte, 2008; Huda et al., 2009), but it could be implemented using different strategies to design gRNAs, or different versions of dCas9. In addition, it could implement the signal at a proteomic level, and decrease the ratio of false positive.

## 6.2 CRISPR/Cas9 & RIME proteomics as a tool for novel transcription factor discovery

In order to investigate the transcription regulation of our genes of interest, we applied a novel proteomic approach, RIME, which was never coupled with the use of CRISPR/Cas9 before, and we demonstrated here how this combination could potentially be a successful discovery tool. In collaboration with AstraZeneca, we optimized the RIME original protocol and we developed a novel, statistical approach to analyse RIME protein hits based on their relative abundance, followed by a ranking strategy based on a desirability function in order to highlight novel potential therapeutic candidates associated with TNBC. Thanks to this approach, we were able to identify three potential regulators, MTA2, CDK1 and CDK6, never associated with *FOXC1*, *NFIB* or *NFE2L3* before.

Among the different discovery proteomic techniques, we decided to use RIME because of its extreme affordability, speed and sensitivity. However, some intrinsic limitations in the technique itself made its applicability more challenging.

In general, because of the usage of a crosslinker and the potential transitory interaction of some proteins, RIME requires optimization to find the correct experimental conditions for immunoprecipitation. For example, some antibodies might not be able to recognize their protein target because of the alteration of the epitopes after formaldehyde modification (Lindskog et al., 2005), which can obviously affect the outcome. In addition, the modifications generated by the crosslink can alter the chemical and physical properties of a peptide, making its identification at the MS level more difficult. In the particular case of formaldehyde though, the modifications it generates are known, and the MS search can be normalized taking this into consideration (Metz et al., 2003; Metz et al., 2006): it has

in fact been shown that the majority of these alterations occur at the amino-termini of lysine, tryptophan, and cysteine side chains (Toews et al., 2008).

In addition, the crosslinking can affect the on-beads digestion, performed in this case using trypsin. The enzyme's epitopes can be masked by formaldehyde modifications, causing an altered processed peptides dimension (very long, or too short), or when two proteins are linked to each other, an anomalous peptide (as a result of the fusion of parts of two interacting proteins). This would obviously interfere with the protein identification, resulting in loss of information or altered data. To overcome this problem, some proteomics protocols reverse the crosslinking with heat, but the harsh condition could also affect the quality of the proteins themselves.

Lastly, it is important to note that RIME is not a quantitative approach, but mainly exploratory: this implies that to confidently identify an interactor at the MS level, a high number of PSMs (peptide spectrum matches) and coverage of the protein are required, which are often correlated to its high abundance at the endogenous level. However, this could not be the case for proteins interacting in a specific compartment (nucleus, for example), or at a particular time point (after a certain stimulus, or at a specific stage of the cell cycle, etc.), which could make the identification more complex.

RIME has originally been developed to investigate interactors of endogenous, relatively abundant proteins, but we successfully modified it to be applied to an exogenous one. However, this decreased significantly the signal at the proteomic level, and made the identification of potential candidates more challenging. In fact, dCas9 was induced just for a short time, during which it had to translocate into the nucleus and interact with a very specific DNA sequence. Because of the importance of the genes of interest, we assumed a high level of transcription, so a high recruitment of the transcription machinery within the promoter sequence, but cells were at different stages of the cell cycle, and this could have influenced the overall gene expression level. For these reasons, despite its effectiveness, another approach could be used to implement the proteomic signals and facilitate protein discovery.

In the last six years, novel techniques have been developed to label neighbouring proteins in the cell as a powerful and complementary approach to classic affinity purification/mass spectrometry (AP/MS)-based interactome mapping. BioID and

APEX are two of the widely used approaches and are both based on generating a reactive biotin derivative that diffuses from the enzyme's active site to label proteins in the near vicinity (Rees et al., 2015; Kim et Roux, 2016). They have the ability to capture weak/transient interactions that can be lost in standard AP approaches, for both soluble and insoluble proteins, and thanks to the strength of the association of biotin with streptavidin, high-stringency protein extraction and capture methods allow minimal background contaminants. APEX in particular is characterized by a faster rate of labelling (minutes versus hours) that can facilitate the identification of dynamic changes in protein–protein associations over time.

Myers and colleagues have demonstrated how powerful this approach is if coupled with CRISPR/Cas9: they developed a novel strategy called GLoPro (genomic locus proteomics) in which they fused dCas9 to the engineered peroxidase APEX2 (Lam et al., 2014) to target a specific DNA sequence with a single gRNA (sgRNA) (Myers et al., 2018), under the control of a Tet-On system. Their strategy could be a very sensible improvement to our approach, since it lacks of the formaldehyde crosslinking and all the limitations related to it.

## 6.3 MTA2, CDK1 and CDK6 contribution to TNBC transcriptional programme

We demonstrated here that the protein candidates identified with our technology are involved in the transcription regulation of the genes we investigated. In particular, we confirmed their recruitment on the putative promoter sequences, and we reported for the first time the direct association of MTA2, CDK1 and CDK6 with *FOXC1*, *NFIB* and *NFE2L3*, showing how their disruption can affect gene expression.

Thanks to ChIP-Seq data analysis, it will be possible to identify even more genes regulated by these proteins, giving us a wider knowledge of the molecular pathways they are involved into.

Furthermore, we showed how their deregulation, in particular of CDK1 and CDK6, could affect the oncogenic ability of TNBC cell lines *in vitro*. These results confirm the importance not only of these proteins, but also of our genes of interest for the tumourigenicity of the disease, and they corroborate with the literature. MTA2, for example, has been shown to be fundamental for tumour growth and metastatic

processes not only in ER-negative breast cancer (Covington et al., 2013) through the Rho pathway, but also in many other types of tumours (gastric, non-small-cell lung cancer for example (Zhou et al., 2015; Zhang et al., 2015)).

Similarly, CDK1 and CDK6 are well known to be involved in cell cycle progression, stemness and transcriptional functions (Lim et Kaldis, 2013), and for these reasons they have gained a lot of interest in the last years as potential therapeutic targets, also for breast cancer. Several compounds are in fact available to target them, such as RO3306 for CDK1 and Palbociclib, Ribociclib and Abemaciclib for CDK6 (Pernas et al., 2018), currently in clinical use (NCT01333137 and NCT01919229 for CDK1, NCT03050398, NCT02732119 and NCT02187783 for CDK6, source: https://clinicaltrials.gov)

However, targeting transcription as a therapeutic strategy *in vivo* is complicated by the fact that cancer-related pathways are complex and sometimes interacting with each other. They are also not exclusively confined to cancer cells, but they are shared with normal tissue, causing toxicity. For these reasons we believe that a combination of therapeutic approaches could be a more efficient treatment for TNBC patients. However, in order to target the right ones, a deeper understanding of the molecular processes where key proteins are involved and how they are regulated could be extremely beneficial for a safer therapeutic approach. In support of this hypothesis and our preliminary results, it has recently been shown how the combination of CDK6 inhibitor palbociclib with chemotherapy in sequential treatments increases significantly the inhibition of cell proliferation and promotes cell death in TNBC (Cretella et al., 2019)

Our findings, even if preliminary, reveal possible, novel mechanism of tumour growth promotion by these proteins' direct regulation of *FOXC1*, *NFIB* or *NFE2L3* expression, and give rational for more investigations on the tumourigenic mechanisms regulated by MTA2, CDK1 and CDK6. In addition, it has to be kept in mind that even the targeted genes are transcription factors themselves, which implicates that their altered expression also affects the transcription regulation of many other genes.

Overall, these conclusions confirm our initial aim to investigate the regulation of gene transcription through CRISPR/Cas9 and discovery proteomics.

However, despite its demonstrated power, we believe that additional confirmation is important to increase the confidence of candidate identification when using our approach: in fact RIME itself brings an intrinsic variability, mainly related to the crosslink. In addition, our statistical approach takes into consideration the overall number of peptides identified for a protein as a reference of relative abundance, not just the PSMs: this inevitably increases the number of potential false positive. For these reasons, functional validations are necessary.

## 6.4 Future directions

CRISPR/Cas9 and RIME proteomics approach can be used to investigate the regulation of transcription of any gene of interest, as demonstrated here. However, to understand the role and the importance of any protein identified, an accurate selection of the candidate and an important validation process has to be carried out.

### 6.4.1 Implementation of CRISPR/Cas9 strategy and proteomic approach

In order to facilitate proteins identification and to reduce the background noise, an implemented but very similar strategy could be used: the GLoPro strategy (Myers et al., 2018), a system with an inducible, sgRNA guided dCas9, fused with APEX2 enzyme, able to biotinylate proximal proteins in the presence of hydrogen peroxide. The inducibility of dCas9, the short action radius of the enzyme biotinylation, and the high affinity of the streptavidin purification should significantly increase the specificity of this approach. In addition, the absence of the crosslinking step and its consequent limitations would improve the confidence of protein identification.

To further reduce the background noise, it could be possible to re-design the gRNA with more novel, developed tools, or to investigate the nature of potential off-targets using website like www.guidescan.com: in this way, the undesired binding of dCas9 across the genome could be reduced.

On a side, a deeper investigation on the quality of the gRNA used could be carried out, identifying the undesired bindings of dCas9 across the genome. This information is provided by the ChIP-Seq experiments we have performed on several

cell lines. Confrontation among them could also be very informative, in order to see if there are preferred regions or sequences for dCas9.

### 6.4.2 Identification of MTA2, CDK1 and CDK6 regulation pathways and genome-wide binding site investigation

To further understand how these proteins contribute to the TNBC transcription program, a ChIP-Seq experiment on *MTA2* and *CDK6* knockdown cells should be performed: in this way it will be possible to understand which pathways would be directly altered by their absence, and highlight potential rescue mechanisms. Due to the essentiality of *CDK1*, this experiment woudn't be informative.
In addition, these results could be coupled with a deeper analysis of the ChIP-Seq data we originated, in order to understand which regions are directly binded by these proteins. These genes could be further validated by ChIP-qPCR, and could potentially be affected by a direct inhibition of these transcription factors.

A deeper investigation should then be performed in order to assess if our protein candidates regulate different pathways in a dependent way, and what processes or tumourigenic mechanism they are involved into. This information would help to elucidate the complex biology of TNBC.

### 6.4.3 RIME proteomics to identify novel interactors

Transcription factors usually interact with other proteins to carry out their function, or they are important subunits of a bigger complex.

To understand if any of our candidates has a specific interactor, RIME proteomics should be performed. This experiment could be critical to understand if they are part of a complex, (like MTA2 is), and if they belong to the same one. Even though our preliminary data didn't show any direct interaction, it is still possible that some of these proteins are part of the same complex to regulate the expression of some genes, or that they share a similar interactor. Further validations, as for example Co-IPs, would be required in order to confirm the RIME results.

These results could help understand if some pathways are co-regulated by the same complexes or protein-protein interactions: due to the essential roles our protein candidates have, their disregulation would be difficult to consider as a

potential therapeutic strategy for TNBC. However it would be interesting to investigate the phosphorylation activity of CDK1 and CDK6 among the proteins recruited within the promoter region of the genes of interest, and to evaluate the effect of its blockade rather then directly targeting the protein itself. In addition, their interactors could potentially be more tissue- or gene-specific, providing a more efficient, potential therapeutic target. Due to the redundancy of transcription factors' activity, deeper and more sophisticated analyses would still need to be performed in order to understand the complexity of gene regulation.

### 6.4.4 *In vivo* validation of MTA2, CDK1 and CDK6's roles

Despite being a great resource, any *in vitro* system lacks in recapitulating the complexity of the tumour behaviour: for this reason, *in vivo* models like PDXs could provide novel, significant information.

Using TNBC patients' derived PDXs, it would be possible to further validate the role of these three proteins in the regulation of the transcription of the genes of interest: ChIP-qPCR could confirm the DNA binding over their promoter sequences, as well as Co-IPs could confirm potential interactions. These experiments should also be performed on those genes highlighted from the ChIP-Seq and RNA-Seq experiments, improving the significance of our results and confirming the relevance of these proteins for TNBC patients.

## 6.5 Conclusions

In conclusion, we have demonstrated here how CRISPR/Cas9 and RIME proteomics approaches can be coupled to investigate the regulation of transcription of any gene of interest. In particular, we have highlighted their applicability to identify novel therapeutic candidates. Furthermore, for the first time we have associated transcription factors like MTA2, CDK1 and CDK6 to the regulation of the transcription of three highly expressed genes in TNBC, as *FOXC1*, *NFIB* and *NFE2L3*, and we are currently investigating in which other pathways they are involved. This information could be essential to understand the efficacy of a potential drug treatment against them for TNBC patients.

# BIBLIOGRAPHY


1. Abramson, V. G., Lehmann, B. D., Ballinger, T. J., & Pietenpol, J. A. (2014). Subtyping of triple-negative breast cancer: Implications for therapy. *Cancer*, *121*(1), pp.8–16.
2. Adams, S., Schmid, P., Rugo, H.S., et al. (2017). Phase 2 study of pembrolizumab (pembro) monotherapy for previously treated metastatic triple-negative breast cancer (mTNBC): KEYNOTE-086 cohort. *A. J Clin Oncol*., 35:1008.
3. Albergaria, A., Paredes, J., Sousa, B., Milanezi, F., Carneiro, V., Bastos, J., Costa, S., Vieira, D., Lopes, N., Lam, E.W., Lunet, N., Schmitt, F. (2009). Expression of FOXA1 and GATA-3 in breast cancer: the prognostic significance in hormone receptor-negative tumours. *Breast Cancer Res* 11**,** R40.
4. Allred, D.C., Mohsin, S.K., Fuqua, S.A. (2001). Histological and biological evolution of human premalignant breast disease. *Endocr Relat Cancer.* 8(1), pp.47-61.
5. Amitai, G., & Sorek, R. (2016). CRISPR–Cas adaptation: insights into the mechanism of action. *Nature Reviews Microbiology*, *14*(2), pp.67–76.
6. Anderlind, C., Spira, A., Cardoso, W.V., Lü, J. (2010). miR-129 regulates cell proliferation by downregulating Cdk6 expression. *Cell Cycle,* 9, pp.1809-18.
7. Apostolou, P., Fostira, F. (2013). Hereditary breast cancer: the era of new susceptibility genes. *Biomed Res Int.,* 747318
8. Arce Vargas, F., Furness, A.J.S., Litchfield, K., et al. (2018). Fc effector function contributes to the activity of human anti-CTLA-4 antibodies. *Cancer Cell*, 33(4):649–663.

9. Arnold, A., & Papanikolaou, A. (2005). Cyclin D1 in Breast Cancer Pathogenesis. *Journal of Clinical Oncology*, *23*(18), pp.4215–4224.
10. Arteaga, C. L., Sliwkowski, M. X., Osborne, C. K., Perez, E. A., Puglisi, F., & Gianni, L. (2011). Treatment of HER2-positive breast cancer: current status and future perspectives. *Nature Reviews Clinical Oncology*, *9*(1), pp.16–32.
11. Arvey, A., Larsson, E., Sander, C., Leslie, C.S., Marks, D.S. (2010). Target mRNA abundance dilutes microRNA and siRNA activity. *Mol Syst Biol.;* 6():363.
12. Asghar, U. S., Barr, A. R., Cutts, R., Beaney, M., Babina, I., Sampath, D., Giltnane, J., Lacap, J.A., Crocker, L., Young, A., Pearson, A., Herrera-Abreu, M.T., Bakal, C., Turner, N. C. (2017). Single-Cell Dynamics Determines Response to CDK4/6 Inhibition in Triple-Negative Breast Cancer. *Clinical Cancer Research*, *23*(18), pp.5561–5572.
13. Asselin-Labat, M.-L., Sutherland, K. D., Barker, H., Thomas, R., Shackleton, M., Forrest, N. C., Hartley, L., Robb, L., Grosveld, F.G., van der Wees, J., Lindeman, G.J., Visvader, J. E. (2006). Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nature Cell Biology*, *9*(2), pp.201–209.
14. Aulmann, S., Adler, N., Rom, J., Helmchen, B., Schirmacher, P., Sinn, H.P. (2006). c-myc amplifications in primary breast carcinomas and their local recurrences. *J Clin Pathol,* 59, pp.424-8.
15. Aulmann, S., Bentz, M., Sinn, H.P. (2002). C-myc oncogene amplification in ductal carcinoma *in situ* of the breast. *Breast Cancer Res Treat*, 74, pp.25-31.
16. Aversa, C., Rossi, V., Geuna, E., et al. (2014). Metastatic breast cancer subtypes and central nervous system metastases. *Breast*, 23, pp. 623-628.
17. Badve, S., Turbin, D., Thorat, M.A., Morimiya, A., Nielsen, T.O., Perou, C.M., Dunn, S., Huntsman, S.D., Nakshatri, H. (2007). FOXA1 Expression in Breast Cancer—Correlation with Luminal Subtype A and Survival. Clin Cancer Res; 13(15).
18. Balmain, A., Gray, J., Ponder, B. (2003). The genetics and genomics of cancer. *Nat Genet*., :238-44.
19. Banerji, J., Rusconi, S., Schaffner, W. (1981). Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell*, 27, pp.299–308.
20. Banerji, S., Cibulskis, K., Rangel-Escareno, C., Brown, K.K., Carter, S.L., , Meyerson, M. (2012). Sequence analysis of mutations and translocations across breast cancer subtypes. *Nature*, 486(7403):405-9.

21. Baron, U., & Bujard, H. (2000). Tet repressor-based system for regulated gene expression in eukaryotic cells: Principles and advances. *Methods Enzymol*, 327, pp.401-421.
22. Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., Horvath, P. (2007). CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science*, *315*(5819), pp.1709–1712.
23. Basta, J., & Rauchman, M. (2015). The Nucleosome Remodeling and Deacetylase Complex in Development and Disease. *Translational Research*, 165(1), pp. 36–47.
24. Bertoli, C., Skotheim, J.M., de Bruin, R.A. (2013). Control of cell cycle transcription during G1 and S phases. *Nat Rev Mol Cell Biol*.; 14(8), pp.518-28.
25. Bertucci, F., Finetti, P., Cervera, N., Charafe-Jauffret, E., Mamessier, E., Adélaïde, J., Debono, S., Houvenaeghel, G., Maraninchi, D., Viens, P., Charpin, C., Jacquemier, J., Birnbaum, D. (2006). Gene Expression Profiling Shows Medullary Breast Cancer Is a Subgroup of Basal Breast Cancers. *Cancer Research*, *66*(9), pp.4636–4644.
26. Bickerton, G. R., Paolini, G. V., Besnard, J., Muresan, S., Hopkins, A. L. (2012). Quantifying the chemical beauty of drugs. *Nature Chemistry*, *4*(2), pp.90–98.
27. Bissell M.J., Rizki A., Mian I.S. (2003). Tissue architecture: the ultimate regulator of breast epithelial function. *Curr. Opin. Cell Biol,* 15, pp.753-762.
28. Blanco-Aparicio, C., Carnero, A. (2013). Pim kinases in cancer: diagnostic, prognostic and treatment opportunities. *Biochem Pharmacol.,* 85(5):629-643.
29. Blau, C.A., Ramirez, A.B., Blau, S., et al. (2016). A distributed network for intensive longitudinal monitoring in metastatic triple-negative breast cancer. J. *Natl. Compr. Canc*. Netw., 14, pp. 8-17.
30. Blows, F. M., Driver, K. E., Schmidt, M. K., Broeks, A., Leeuwen, F. E. V., Wesseling, J., … Huntsman, D. (2010). Subtyping of Breast Cancer by Immunohistochemistry to Investigate a Relationship between Subtype and Short and Long Term Survival: A Collaborative Analysis of Data for 10,159 Cases from 12 Studies. *PLoS Medicine*, *7*(5).
31. Boersema, P.J., Raijmakers, R., Lemeer, S., Mohammed, S., Heck, A.J.R. (2009). Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat. Protoc.*, 4, pp.484-494.
32. Bogdanove, A.J. & Voytas, D.F. (2011). TAL effectors: customizable proteins for DNA targeting. *Science*, 333, pp.1843-1846.
33. Borcherding, N., Kolb, R., Gullicksrud, J., Vikas, P., Zhu, Y., Zhang, W. (2018) Keeping tumors in check: a mechanistic review of clinical response and

resistance to immune checkpoint blockade in cancer. *J Mol Biol*., 430(14):2014–2029.

34. Bowen, N. J., Fujita, N., Kajita, M., & Wade, P. A. (2004). Mi-2/NuRD: multiple complexes for many purposes. *Biochimica Et Biophysica Acta (BBA) - Gene Structure and Expression*, *1677*(1-3), pp.52–57.
35. Boyle, A.P., Song, L., Lee, B.K., London, D., Keefe, D., Birney, E., Iyer, V.R., Crawford, G.E., Furey, T.S. (2011). High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. *Genome Res.*, 21, pp.456–464.
36. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., Jemal, A. (2018). Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, 68(6), pp.394–424.
37. Brivanlou, A.H. & Darnell, J.E. (2002). Signal transduction and the control of gene expression. *Science*, 295(5556).
38. Brouns, S. J. J., Jore, M. M., Lundgren, M., Westra, E. R., Slijkhuis, R. J. H., Snijders, A. P. L., Dickman, M.J., Makarova, K.S., Koonin, E.V., van der Oost, J. (2008). Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science*, *321*(5891), pp.960–964.
39. Brown, N., Korolchuk, S., Martin, M., Stanley, W., Moukhametzianov, R., Noble, M. and Endicott, J. (2015). CDK1 structures reveal conserved and unique features of the essential cell cycle CDK. *Nature Communications*, 6(1).
40. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods*., 10, pp.1213–1218.
41. Burstein, H.J., Polyak, K., Wong, J.S., Lester, S.C., Kaelin, C.M. (2004). Ductal carcinoma *in situ* of the breast. *N Engl J Med*, 350(14), pp.1430-41.
42. Buss, H., Handschick, K., Jurrmann, N., Pekkonen, P., Beuerlein, K., Müller, H., Wait, R., Saklatvala, J., Ojala, P.M., Lienhard Schmitz, M., et al. (2012). Cyclin-dependent kinase 6 phosphorylates NF-κB P65 at serine 536 and contributes to the regulation of inflammatory gene expression. *PloS One*, 7, pp.1-13.
43. Calo, V., Migliavacca, M., Bazan, V., Macaluso, M., Buscemi, M., Gebbia, N., Russo, A. (2003). Stat proteins: from normal control of cellular events to tumorigenesis. *J Cell Physiol*, 197, pp.157–168.
44. Cancer Genome Atlas N. (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, 490, pp.61–70.

45. Carey, L. A., Perou, C. M., Livasy, C. A., Dressler, L. G., Cowan, D., Conway, K., Karaca, G., Troester, M.A., Tse, C.K., Edmiston, S., Deming, S.L., Geradts, J., Cheang, M.C., Nielsen, T.O., Moorman, P.G., Earp, H.S., Millikan, R. C. (2006). Race, Breast Cancer Subtypes, and Survival in the Carolina Breast Cancer Study. *Jama*, *295*(21), pp.2492.
46. Carroll, J. S. (2016). Mechanisms of oestrogen receptor (ER) gene regulation in breast cancer. European Journal of Endocrinology, 175(1), pp. R41–R49.
47. Carroll, J. S., Liu, X. S., Brodsky, A. S., Li, W., Meyer, C. A., Szary, A. J., Eeckhoute, J., Shao, W., Hestermann, E.V., Geistlinger, T.R., Fox, E.A., Silver, P.A., Brown, M. (2005). Chromosome-Wide Mapping of Estrogen Receptor Binding Reveals Long-Range Regulation Requiring the Forkhead Protein FoxA1. *Cell*, *122*(1), pp.33–43.
48. Cassidy, J. W., Caldas, C., Bruna, A. (2015). Maintaining Tumor Heterogeneity in Patient-Derived Tumor Xenografts. *Cancer Research*, *75*(15), pp.2963–2968.
49. Chandarlapaty, S., Chen, D., He, W., Sung, P., Samoila, A., You, D., Bhatt, T., Patel, P., Voi, M., Gnant, M., Hortobagyi, G., Baselga, J., Moynahan, M.E. (2016). Prevalence of ESR1 Mutations in Cell-Free DNA and Outcomes in Metastatic Breast Cancer: A Secondary Analysis of the BOLERO-2 Clinical Trial. *JAMA Oncol.*, 2(10):1310-131521-26.
50. Chapuy, B., McKeown, M.R., Lin, C.Y., Monti, S., Roemer, M.G., Qi, J., Rahl, P.B., .., Bradner, J.E. (2013). Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell*, 24(6):777-90.
51. Cheang, M. C. U., Martin, M., Nielsen, T. O., Prat, A., Voduc, D., Rodriguez-Lescure, A., Ruiz, A., Chia, S., Shepherd, L., Ruiz-Borrego, M., Calvo, L., Alba, E., Carrasco, E., Caballero, R., Tu, D., Pritchard, K.I., Levine, M.N., Bramwell, V.H., Parker, J., Bernard, P.S., Ellis, M.J., Perou, C.M., Di Leo, A., Carey, L. A. (2015). Defining Breast Cancer Intrinsic Subtypes by Quantitative Receptor Expression. *The Oncologist*, *20*(5), pp.474–482.
52. Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G.-W., Park, J., Blackburn, E.H., Weissman, J.S., Qi, L.S., Huang, B. (2013). Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell*, *155*(7), pp.1479–1491.
53. Chen, L., Zhang, R., Li, P., Liu, Y., Qin, K., Fa, Z.Q., Liu, Y.J., Ke, Y.Q., Jiang, X.D. (2013). P53-induced microRNA-107 inhibits proliferation of glioma cells and down-regulates the expression of CDK6 and Notch-2. *Neurosci Lett* ; 534, pp.327-32.

54. Chiang, A.C., Massague, J. (2008) Molecular basis of metastasis. *N. Engl. J. Med.*, 359, pp. 2814-2820.
55. Chipumuro, E., Marco, E., Christensen, C.L., Kwiatkowski, N., Zhang, T., .., George, R.E. (2014). CDK7 inhibition suppresses super-enhancer-linked oncogenic transcription in MYCN-driven cancer. *Cell*, 159(5):1126-1139.
56. Choi, Y., Chakrabarti, R., Escamilla-Hernandez, R. Sinha, S. (2009). Elf5 conditional knockout mice reveal its role as a master regulator in mammary alveolar development: Failure of Stat5 activation and functional differentiation in the absence of Elf5. *Developmental Biology*, 329(2), pp.227-241.
57. Chowdhury, A., Katoh, H., Hatanaka, A., Iwanari, H., Nakamura, N., Hamakubo, T., Natsume, T., Waku, T. and Kobayashi, A. (2017). Multiple regulatory mechanisms of the biological function of NRF3 (NFE2L3) control cancer cell proliferation. *Scientific Reports*, 7(1).
58. Christie, E.L., Fereday, S., Doig, K., Pattnaik, S., Dawson, S.J., Bowtell, D.D.L. (2017). Reversion of BRCA1/2 Germline Mutations Detected in Circulating Tumor DNA From Patients With High-Grade Serous Ovarian Cancer.. *J Clin Oncol*., 35(12):1274-1280.
59. Christoforou, A., Arias, A. M., Lilley, K. S. (2014). Determining protein subcellular localization in mammalian cell culture with biochemical fractionation and iTRAQ 8-plex quantification. *Methods Mol. Biol.,* 1156, 157–174.
60. Cirillo, L.A., & Zaret, K.S. (1999). An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol Cell.*, 4(6), pp.961-9.
61. Consortium APG, AACR Project GENIE. (2017). Powering precision medicine through an international consortium. *Cancer Discov*., 7:818–31.
62. Cooley, S., Burns, L.J., Repka, T., Miller, J.S. (1999). Natural killer cell cytotoxicity of breast cancer targets is enhanced by two distinct mechanisms of antibody-dependent cellular cytotoxicity against LFA-3 and HER2/neu. *Exp Hematol*., 27(10):1533–41.
63. Covington, K.R., Brusco, L., Barone, I., Tsimelzon, A., Selever, J., Corona-Rodriguez, A., Brown, P., Kumar, R, Hilsenbeck, S.G., Fugua, S.A. (2013). Metastasis tumor-associated protein 2 enhances metastatic behavior and is associated with poor outcomes in estrogen receptor-negative breast cancer. *Breast Cancer Research and Treatment*, *141*(3), pp.375–384.
64. Cretella, D., Fumarola, C., Bonelli, M., Alfieri, R., Monica, S. L., Digiacomo, G., Cavazzoni, A., Galetti, M., Generali, D., Petronini, P. G. (2019). Pre-treatment

with the CDK4/6 inhibitor palbociclib improves the efficacy of paclitaxel in TNBC cells. *Scientific Reports*, 9(1).

65. Cui, Y., Niu, A., Pestell, R., Kumar, R., Curran, E. M., Liu, Y., & Fuqua, S. A. W. (2006). Metastasis-Associated Protein 2 Is a Repressor of Estrogen Receptor α Whose Overexpression Leads to Estrogen-Independent Growth of Human Breast Cancer Cells. *Molecular Endocrinology*, *20*(9), pp.2020–2035.
66. Curtis, C., Shah, S. P., Chin, S.-F., Turashvili, G., Rueda, O. M., Dunning, M. J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., Gräf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S.; METABRIC Group, Langerød, A., Green, A., Provenzano, E., Wishart, G., Pinder, S., Watson, P., Markowetz, F., Murphy, L., Ellis, I., Purushotham, A., Børresen-Dale, A.L., Brenton, J.D., Tavaré, S., Caldas, C., Aparicio, S. (2012). The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*, *486*(7403), pp.346–352.
67. D'Santos, C., Taylor, C., Carroll, J. S., Mohammed, H. (2015). RIME proteomics of oestrogen and progesterone receptors in breast cancer. *Data in Brief*, 5, pp.276–280.
68. Dai, M., Zhang, C., Ali, A., Hong, X., Tian, J., Lo, C., Aimé, N., Burgos, S.A., Ali, S., Lebrun, J.-J. (2016). CDK4 regulates cancer stemness and is a novel therapeutic target for triple-negative breast cancer. *Scientific Reports*, *6*(1).
69. Davies, C., Pan, H.C., Godwin, J. et al. (2013). Long-term effects of continuing adjuvant tamoxifen to 10 years versus stopping at 5 years after diagnosis of oestrogen receptor-positive breast cancer: ATLAS, a randomised trial. *Lancet*, 381, pp. 805-816.
70. Dawson, M.A., Prinjha, R.K., Dittmann, A., Giotopoulos, G., Bantscheff, M., .., Kouzarides, T. (2011). Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature*, 478(7370):529-33.
71. Dawson, S.-J., Rueda, O. M., Aparicio, S., & Caldas, C. (2013). A new genome-driven integrated classification of breast cancer and its implications. *The EMBO Journal*, *32*(5), pp.617–628.
72. Dawson, S., Provenzano, E., Caldas, C. (2009). Triple negative breast cancers: Clinical and prognostic implications. *European Journal of Cancer*, *45*, pp.27–40.
73. Delmore, J. E., Issa, G. C., Lemieux, M. E., Rahl, P. B., Shi, J., Jacobs, H. M., … Mitsiades, C. S. (2011). BET Bromodomain Inhibition as a Therapeutic Strategy to Target c-Myc. *Cell*, *146*(6), pp.904–917.

74. Delmore, J.E., Issa, G.C., Lemieux, M.E., Rahl, P.B., Shi, J., .., Mitsiades, C.S. (2011). BET bromodomain inhibition as a therapeutic strategy to target c-Myc.*Cell*., 146(6):904-17.
75. Deltcheva, E., Chylinski, K., Sharma, C. M., Gonzales, K., Chao, Y., Pirzada, Z. A., Eckert, M.R., Vogel, J., Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, *471*(7340), pp.602–607.
76. Demicco, E.G., Kavanagh, K.T., Romieu-Mourez, R., Wang, X., Shin, S.R., Landesman-Bollag, E., Seldin, D.C., Sonenshein, G.E. (2005). RelB/p52 NF-kappaB complexes rescue an early delay in mammary gland development in transgenic mice with targeted superrepressor IkappaB-alpha expression and promote carcinogenesis of the mammary gland. *Mol. Cell. Biol.*, *25*(22), pp.10136-10147.
77. Dent, R., Trudeau, M., Pritchard, K. I., Hanna, W. M., Kahn, H. K., Sawka, C. A., Lickley, L.A., Rawlinson, E., Sun, P., Narod, S. A. (2007). Triple-Negative Breast Cancer: Clinical Features and Patterns of Recurrence. *Clinical Cancer Research*, *13*(15), pp.4429–4434.
78. Derringer, G., & Suich, R. (1980). Simultaneous optimization of several response variables. *Journal of Quality Technology*; 12(4), pp.214–219.
79. Desmedt, C., Voet, T., Sotiriou, C., Campbell, P.J. (2012). Next-generation sequencing in breast cancer: first take home messages. *Curr Opin Oncol*., 24(6):597-604.
80. Desterro, J.M., Rodriguez, M.S., Hay, R.T. (2000). Regulation of transcription factors by protein degradation. *Cell Mol Life Sci*., 57(8-9):1207-19.
81. Ding, L., Ellis, M.J., Li, S. et al. (2010). Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature*, 464, pp. 999-1005.
82. Diril, M. K., Ratnacaram, C.K., Padmakumar, V.C., Du, T., Wasser, M., Coppola, V., Tessarollo, L., Kaldis, P. (2012). Cyclin-dependent kinase 1 (Cdk1) is essential for cell division and suppression of DNA re-replication but not for liver regeneration. *Proc. Natl Acad. Sci. USA,* 109, pp.3826–3831.
83. Dominguez, A., Lim, W., Qi, L. (2015). Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation. *Nature Reviews Molecular Cell Biology*, 17(1), pp.5-15.
84. Donnison, M., Beaton, A., Davey, H.W., Broadhurst, R., L'Huillier, P., Pfeffer, P.L. (2005). Loss of the extra embryonic ectoderm in Elf5 mutants leads to defects in embryonic patterning. *Development*, 132(10), pp.2299-2308.

85. Draker, R., Ng, M. K., Sarcinella, E., Ignatchenko, V., Kislinger, T., Cheung, P. (2012). A Combination of H2A.Z and H4 Acetylation Recruits Brd2 to Chromatin during Transcriptional Activation. *PLoS Genetics*, *8*(11), e1003047.
86. Dravis, C., Spike, B., Harrell, J., Johns, C., Trejo, C., Southard-Smith, E., Perou, C. and Wahl, G. (2015). Sox10 Regulates Stem/Progenitor and Mesenchymal Cell States in Mammary Epithelial Cells. *Cell Reports*, 12(12), pp.2035-2048.
87. Drukker, C.A., van Tinteren, H., Schmidt, M.K., et al. (2014). Long-term impact of the 70-gene signature on breast cancer outcome. Breast Cancer Res.Treat., 143, pp. 587-592.
88. Edwards, S.L., Brough, R., Lord, C.J., Natrajan, R., Vatcheva, R., Levine, D.A., Boyd, J., Reis-Filho, J.S., Ashworth, A. (2008). Resistance to therapy caused by intragenic deletion in BRCA2. *Nature*, 451(7182):1111-5.
89. Eeckhoute, J., Keeton, E. K., Lupien, M., Krum, S. A., Carroll, J. S., Brown, M. (2007). Positive Cross-Regulatory Loop Ties GATA-3 to Estrogen Receptor Expression in Breast Cancer. *Cancer Research*, *67*(13), pp.6477–6483.
90. Egloff, S., & Murphy, S. (2008). Cracking the RNA polymerase II CTD code. *Trends Genet.,* 24, pp.280–288.
91. Eisen, J.A., Sweder, K.S., Hanawalt, P.C. (1995). Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. *Nucleic Acids Res,* 23, pp.2715–2723.
92. Ellis, M.J., Ding, L., Shen, D., Luo, J., Suman, V.J, .., Mardis, E.R. (2012). Whole-genome analysis informs breast cancer response to aromatase inhibition. *Nature*, 486(7403):353-60.
93. Ernst, J., & Kellis, M. (2010). Discovery and characterization of chromatin states for systematic annotation of the human genome. *Nature biotechnology*, 28, pp.817–825.
94. Fanning, S.W., Mayne, C.G., Dharmarajan, V., Carlson, K.E., Martin, T.A., Novick, S.J., Toy, W., Green, B., Panchamukhi, S., Katzenellenbogen, B.S., Tajkhorshid, E., Griffin, P.R., Shen, Y., Chandarlapaty, S., Katzenellenbogen, J.A., Greene, G.L. (2016). Elife. Estrogen receptor alpha somatic mutations Y537S and D538G confer breast cancer endocrine resistance by stabilizing the activating function-2 binding conformation. *Elife, 5.*
95. Feng, W., Liu, S., Zhu, R., Li, B., Zhu, Z., Yang, J. and Song, C. (2017). SOX10 induced Nestin expression regulates cancer stem cell properties of TNBC cells. *Biochemical and Biophysical Research Communications*, 485(2), pp.522-528.

96. Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.*, 9, pp.397-405

97. Filipits, M., Nielsen, T.O., Rudas, M. et al. (2014). The PAM50 risk-of-recurrence score predicts risk for late distant recurrence after endocrine therapy in postmenopausal women with endocrine-responsive early breast cancer. *Clin. Cancer Res.*, 20, pp. 1298-1300.

98. Finn, R. S., Dering, J., Conklin, D., Kalous, O., Cohen, D. J., Desai, A. J., Ginther, C., Atefi, M., Chen, I., Fowst, C., Los, G., Slamon, D. J. (2009). PD 0332991, a selective cyclin D kinase 4/6 inhibitor, preferentially inhibits proliferation of luminal estrogen receptor-positive human breast cancer cell lines in vitro. *Breast Cancer Research*, *11*(5).

99. Fischer, L., Chen, Z. A. & Rappsilber, J. (2013). Quantitative cross-linking/mass spectrometry using isotope-labelled cross-linkers. *J. Proteom.,* 88, pp.120–128.

100. Fisher, B., Costantino, J. P., Wickerham, D. L., Cecchini, R. S., Cronin, W. M., Robidoux, A., … Wolmark, N. (2005). Tamoxifen for the Prevention of Breast Cancer: Current Status of the National Surgical Adjuvant Breast and Bowel Project P-1 Study. *JNCI: Journal of the National Cancer Institute*, *97*(22), pp.1652–1662.

101. Forbes, S.A., Beare, D., Bindal, N., et al. (2016). COSMIC: High-resolution cancer genetics using the catalogue of somatic mutations in cancer. *Curr Protoc Hum Genet.*, 91:10.11.1–37.

102. Foulkes, W. D. (2003). Germline BRCA1 Mutations and a Basal Epithelial Phenotype in Breast Cancer. *CancerSpectrum Knowledge Environment*, *95*(19), pp.1482–1485.

103. Foulkes, W.D., Smith, I.E., Reis-Filho, J.S. (2010). Triple-negative breast cancer. *N Engl J Med.*, 363:1938–1948.

104. Fu, J., Qin, L., He, T., Qin, J., Hong, J., Wong, J., Liao, L., Xu, J. (2010). The TWIST/Mi2/NuRD protein complex and its essential role in cancer metastasis. *Cell Research*, *21*(2), pp.275–289.

105. Fuda, N. J., Ardehali, M. B., Lis, J. T. (2009). Defining mechanisms that regulate RNA polymerase II transcription *in vivo. Nature,* 461, pp.186–192.

106. Fujita, N., Jaye, D.L., Kajita, M., Geigerman, C., Moreno, C.S., Wade, P.A. (2003). MTA3, a Mi-2/NuRD complex subunit, regulates an invasive growth pathway in breast cancer. *Cell*, 113, pp.207–219.

107. Fujita, N., Kajita, M., Taysavang, P., Wade, P.A. (2004). Hormonal regulation of metastasis-associated protein 3 transcription in breast cancer cells. *Mol Endocrinol.*, 18, pp.2937–2949.

108. Fujita, T., & Fujii, H. (2013). Efficient isolation of specific genomic regions and identification of associated proteins by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP) using CRISPR. *Biochemical and Biophysical Research Communications*, *439*(1), pp.132–136.

109. Fujita, T., & Fujii, H. (2014). Identification of Proteins Associated with an IFNγ-Responsive Promoter by a Retroviral Expression System for enChIP Using CRISPR. *PLoS ONE*, *9*(7).

110. Fujita, T., Asano, Y., Ohtsuka, J., Takada, Y., Saito, K., Ohki, R., & Fujii, H. (2013). Identification of telomere-associated molecules by engineered DNA-binding molecule-mediated chromatin immunoprecipitation (enChIP). *Scientific Reports*, 3(1).

111. Fujita, T., Yuno, M., Okuzaki, D., Ohki, R., Fujii, H. (2015). Identification of Non-Coding RNAs Associated with Telomeres Using a Combination of enChIP and RNA Sequencing. *Plos One*, 10(4).

112. Fujita, T., Yuno, M., Suzuki, Y., Sugano, S., Fujii, H. (2017). Identification of physical interactions between genomic regions by enChIP-Seq. *Genes to Cells*, 22(6), pp.506–520.

113. Fukuda, T., Daniel, K., Wojtasz, L., Toth, A., Hoog, C. (2010). A novel mammalian HORMA domain-containing protein, HORMAD1, preferentially associates with unsynapsed meiotic chromosomes. *Experimental cell research*, 316:158–71.

114. Furey T.S. (2012). ChIP-seq and beyond: new and improved methodologies to detect and characterize protein-DNA interactions. *Nat Rev Genet.,* 13, pp.840–852.

115. Garneau, J. E., Dupuis, M.-È., Villion, M., Romero, D. A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magadán, A.H., Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*, *468*(7320), pp.67–71.

116. Geyer, F.C., Pareja, F., Weigelt, B., et al. (2017). The Spectrum of Triple-Negative Breast Disease: High- and Low-Grade Lesions. *Am J Pathol*., 187(10):2139–2151.

117. Geyer, F.C., Weigelt, B., Natrajan, R., Lambros, M.B., de Biase, D., Vatcheva, R., .., Reis-Filho, J.S. (2010). Molecular analysis reveals a genetic basis for the phenotypic diversity of metaplastic breast carcinomas. *J Pathol*., 220:562–573.

118. Ghebeh, H., Mohammed, S., Al-Omair, A., et al. (2006). The B7-H1 (PD-L1) T lymphocyte-inhibitory molecule is expressed in breast cancer patients with infiltrating ductal carcinoma: correlation with important high-risk prognostic factors. *Neoplasia*, 8(3):190–198.

119. Gibbs, J.B. (2000). Mechanism-based target identification and drug discovery in cancer research. *Science*, 287(5460), pp.1969-1973.

120. Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., Lim, W.A., Weissman, J.S., Qi, L.S. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell, 154(2), pp.1455-1458.

121. Goga, A., Yang D., Tward, A.D., Morgan, D.O., Bishop, J.M. (2007). Inhibition of CDK1 as a potential therapy for tumors over-expressing MYC. *Nat Med*., 13, pp.820–827.

122. Goss, E.J., Ingle, N., Martino, S. et al. (2005). Randomized trial of letrozole following tamoxifen as extended adjuvant therapy in receptor-positive breast cancer: updated findings from NCICCTG MA.17 *J. Natl. Cancer Inst*., 97, pp. 1262-1270.

123. Grant, P.A., Duggan, L., Cote, J., Roberts, S. M., Brownell, J.E., Candau, R., Ohba, R., Owen-Hughes, T., Allis, C.D., Winston, F., Berger, S.L., Workman, J.L. (1997). Yeast Gcn5 functions in two multisubunit complexes to acetylate nucleosomal histones: characterization of an Ada complex and the SAGA (Spt/Ada) complex. *Genes & Development*, *11*(13), pp.1640–1650.

124. Graus-Porta, D. (1997). ErbB-2, the preferred heterodimerization partner of all ErbB receptors, is a mediator of lateral signaling. *The EMBO Journal*, *16*(7), pp.1647–1655.

125. Grushko, T.A., Dignam, J.J., Das, S., et al. (2004). MYC is amplified in BRCA1-associated breast cancers. *Clin Cancer Res*, 10:499-507.

126. Gygi, S.P., Rist, B., Gerber, S.A., Turecek, F., Gelb, M.H., Aebersold R. (1999). Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat. Biotechnol.*, 17, pp.994-999.

127. Hamidian, A., Vaapil, M., Stedingk, K. V., Fujita, T., Persson, C. U., Eriksson, P., Veerla, S., De Preter, K., Speleman, F., Fujii, H., Påhlman, S., Mohlin, S. (2018). Promoter-associated proteins of EPAS1 identified by enChIP-MS – A putative role of HDX as a negative regulator. *Biochemical and Biophysical Research Communications*, *499*(2), pp.291–298.

128. Han, B., Qu, Y., Jin, Y., Yu, Y., Deng, N., Wawrowsky, K., Zhang, X., Li, N., Bose, S., Wang, Q., Sakkiah, S., Abrol, R., Jensen, T., Berman, B., Tanaka, H.,

Johnson, J., Gao, B., Hao, J., Liu, Z., Buttyan, R., Ray, P., Hung, M., Giuliano, A. and Cui, X. (2015). FOXC1 Activates Smoothened-Independent Hedgehog Signaling in Basal-like Breast Cancer. Cell Reports, 13(5), pp.1046-1058.

129. Han, B., Zhou, B., Qu, Y., Gao, B., Xu, Y., Chung, S., Tanaka, H., Yang, W., Giuliano, A. and Cui, X. (2018). FOXC1-induced non-canonical WNT5A-MMP7 signaling regulates invasiveness in triple-negative breast cancer. *Oncogene*, 37(10), pp.1399-1408.

130. Handschick, K., Beuerlein, K., Jurida, L., Bartkuhn, M., Müller, H., Soelch, J., Weber, A., Dittrich-Breiholz, O., Schneider, H., Scharfe, M., et al. (2014). Cyclin-dependent kinase 6 is a chromatin-bound cofactor for NF-κB-dependent gene expression. *Mol Cell*, 53, pp.193-208.

131. Harbour, J.W., Luo, R.X., Santi, A.D., Postigo, A.A., Dean, D.C. (1999). Cdk phosphorylation triggers sequential intramolecular interactions that progressively block Rb functions as cells move through G1. *Cell*, 98, pp.859-69.

132. Harrell, J. C., Dye, W. W., Allred, D. C., Jedlicka, P., Spoelstra, N. S., Sartorius, C. A., & Horwitz, K. B. (2006). Estrogen Receptor Positive Breast Cancer Metastasis: Altered Hormonal Sensitivity and Tumor Aggressiveness in Lymphatic Vessels and Lymph Nodes. *Cancer Research*, *66*(18), pp.9308–9315.

133. Harrington J. (1965). The desirability function. *Industrial Quality Control*; 21(10), pp.494–498.

134. Hartley, P.D. & Madhani, H.D. (2009). Mechanisms that specify promoter nucleosome location and identity. *Cell*, 137, pp.445–458.

135. Hartmaier, R.J., Trabucco, S.E., Priedigkeit, N., Chung, J.H., .., Lee, A.V. (2018). Recurrent hyperactive ESR1 fusion proteins in endocrine therapy-resistant breast cancer. *Ann Oncol*., 29(4):872-880.

136. Havel, J.J., Chowell, D., Chan, T.A. (2019). The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy. *Nat Rev Cancer*. 19(3):133-150.

137. Havugimana, P. C., Hart, G. T., Nepusz, T., Yang, H., Turinsky, A. L., Li, Z., Wang, P.I., Boutz, D.R., Fong, V., Phanse, S., Babu, M., Craig, S.A., Hu, P., Wan, C., Vlasblom, J., Dar, V.U., Bezginov, A., Clark, G.W., Wu, G.C., Wodak, S.J., Tillier, E.R., Paccanaro, A., Marcotte, E.M., Emili, A. (2012). A Census of Human Soluble Protein Complexes. *Cell*, *150*(5), pp.1068–1081.

138. He, Q., Johnston, J., Zeitlinger, J. (2015). ChIP-nexus enables improved detection of in vivo transcription factor binding footprints. *Nat Biotechnol*., 33, pp.395–401.

139. Heinz, S., Romanoski, C.E., Benner, C., Glass, C.K. (2015). The selection and function of cell type-specific enhancers. *Nature Reviews Molecular Cell Biology*, *16*(3), pp.144–154.

140. Heler, R., Marraffini, L. A., Bikard, D. (2014). Adapting to new threats: the generation of memory by CRISPR-Cas immune systems. *Molecular Microbiology*, *93*(1), pp.1–9.

141. Heler, R., Samai, P., Modell, J. W., Weiner, C., Goldberg, G. W., Bikard, D., Marraffini, L. A. (2015). Cas9 specifies functional viral targets during CRISPR–Cas adaptation. *Nature*, *519*(7542), pp.199–202.

142. Henley, S.A., & Dick, F.A. (2012). The retinoblastoma family of proteins and their regulatory functions in the mammalian cell division cycle. *Cell Div*.; 7(1), pp.10.

143. Hennighausen, L., & Robinson, G. W. (2001). Signaling Pathways in Mammary Gland Development. *Developmental Cell*, *1*(4), pp.467–475.

144. Herschkowitz, J. I., He, X., Fan, C., Perou, C. M. (2008). The functional loss of the retinoblastoma tumour suppressor is a common event in basal-like and luminal B breast carcinomas. *Breast Cancer Research*, *10*(5).

145. Hilton, I. B., Dippolito, A. M., Vockley, C. M., Thakore, P. I., Crawford, G. E., Reddy, T. E., Gersbach, C. A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature Biotechnology*, *33*(5), pp.510–517.

146. Hnisz, D., Abraham, B., Lee, T., Lau, A., Saint-André, V., Sigova, A., Hoke, H. and Young, R. (2013). Super-Enhancers in the Control of Cell Identity and Disease. Cell, 155(4), pp.934-947.

147. Hnisz, D., Schuijers, J., Lin, C.Y., Weintraub, A.S., Abraham, B.J., Lee, T.I., Bradner, J.E., Young, R.A. (2015). Convergence of developmental and oncogenic signaling pathways at transcriptional super-enhancers. *Mol Cell*, 58(2):362-70.

148. Hodi, F.S., O'Day, S.J., McDermott, D.F., et al. (2010). Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med*. 363:711-723.

149. Honeywell, D.R., Cabrita, M.A., Zhao, H., Dimitroulakos, J., Addison, C.L. (2013). miR-105 inhibits prostate tumour growth by suppressing CDK6 levels. *PloS One,* 8, pp.1-12.

150. Hong, S. W., Jiang, Y., Kim, S., Li, C. J., & Lee, D. K. (2014). Target gene abundance contributes to the efficiency of siRNA-mediated gene silencing. *Nucleic acid therapeutics*, *24*(3), 192–198.

151. Horiuchi, D., Camarda, R., Zhou, A.Y., Yau, C., Momcilovic, O., Balakrishnan, S., .., Goga, A. (2016). IM1 kinase inhibition as a targeted therapy against triple-negative breast tumors with elevated MYC expression. *Nat Med*., 22(11):1321-1329.

152. Hoshino, A., & Fujii, H. (2009). Insertional chromatin immunoprecipitation: a method for isolating specific genomic regions. *J. Biosci. Bioeng.,* 108, pp.446-449.

153. Howell, A. (2005). The future of fulvestrant ('Faslodex'). *Cancer Treatment Reviews*, *31*.

154. Hsin, J.P., & Manley, J.L. (2012). The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes Dev*., 26, pp.2119–2137.

155. Huang, V., Place, R. F., Portnoy, V., Wang, J., Qi, Z., Jia, Z., Yu, A., Shuman, M., Yu, J. Li, L.-C. (2011). Upregulation of Cyclin B1 by miRNA and its implications in cancer. *Nucleic Acids Research*, *40*(4), pp.1695–1707.

156. Hubner, N. C., Bird, A. W., Cox, J., Splettstoesser, B., Bandilla, P., Poser, I., Hyman, A., Mann, M. (2010). Quantitative proteomics combined with BAC TransgeneOmics reveals in vivo protein interactions. *The Journal of Cell Biology*, *189*(4), pp.739–754.

157. Huda, A., Mariño-Ramírez, L., Landsman, D., Jordan, I. K. (2009). Repetitive DNA elements, nucleosome binding and human gene expression. *Gene*, *436*(1-2), pp.12–22.

158. Hudson, T.J., Anderson, W., Artez, A., Barker, A.D., Bell, C., Bernabé, R.R., Bhan, M.K., Calvo, F., .., Yang, H. (2010). International network of cancer genome projects. International Cancer Genome Consortium. *Nature*, 464(7291):993-8.

159. Hulton, C. H., Costa, E. A., Shah, N. S., Quintanal-Villalonga, A., Heller, G., Stanchina, E. D., Rudin, C.M., Poirier, J. T. (2019). Direct genome editing of patient-derived xenografts using CRISPR-Cas9 enables rapid in vivo functional genomics. *Biorxiv.*

160. Ismail-Khan, R., & Bui, M. M. (2010). A Review of Triple-Negative Breast Cancer. *Cancer Control*, *17*(3), pp.173–176.

161. Itzhak, D. N., Davies, C., Tyanova, S., Mishra, A., Williamson, J., Antrobus, R., Cox, J., Weekes, M.P., Borner, G. H. (2017). A Mass Spectrometry-Based Approach for Mapping Protein Subcellular Localization Reveals the Spatial Proteome of Mouse Primary Neurons. *Cell Reports*, *20*(11), pp.2706–2718.

162. Itzhak, D. N., Tyanova, S., Cox, J. & Borner, G. H. (2016). Global, quantitative and dynamic mapping of protein subcellular localization. *eLife* 5, e16950.

163. Izumi Y., Xu L., di Tomaso E., Fukumura D., et al. (2002). Tumour biology: herceptin acts as an anti-angiogenic cocktail. *Nature*, 416(6878):279–80.

164. Jäger, S., Cimermancic, P., Gulbahce, N., Johnson, J. R., Mcgovern, K. E., Clarke, S. C., … Krogan, N. J. (2011). Global landscape of HIV–human protein complexes. *Nature*, *481*(7381), pp.365–370.

165. Jakesz, R., Greil, R., Gnant, M. et al. (2007). Extended adjuvant therapy with anastrozole among postmenopausal breast cancer patients: results from the randomized Austrian Breast and Colorectal Cancer Study Group Trial 6a. *J. Natl. Cancer Inst.*, 99, pp. 1845-1850.

166. Jamai, A., Imoberdorf, R.M., Strubin, M. (2007). Continuous histone H2B and transcription-dependent histone H3 exchange in yeast cells outside of replication. *Mol. Cell* 25, pp.345–355.

167. Jensen, T., Ray, T., Wang, J., Li, X., Naritoku, W., Han, B., Bellafiore, F., Bagaria, S., Qu, A., Cui, X., Taylor, C. and Ray, P. (2015). Diagnosis of Basal-Like Breast Cancer Using a FOXC1-Based Assay. *JNCI: Journal of the National Cancer Institute*, 107(8).

168. Jeselsohn, R., Buchwalter, G., De Angelis, C., Brown, M., Schiff, R. (2015). ESR1 mutations—a mechanism for acquired endocrine resistance in breast cancer. *Nat Rev Clin Oncol.*, 12(10):573-83.

169. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science,* 337, pp.816–821.

170. Johnson, N., Cai, D., Kennedy, R.D., Pathania, S., Arora, M., Li, Y.C., D'Andrea, A.D., Parvin, J.D., Shapiro, G.I. (2009). Cdkl participates in BRCA1-dependent S phase checkpoint control in response to DNA damage. *Mol Cell.*, 35, pp.327–339.

171. Jonkers, I., & Lis, J. T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*, *16*(3), pp.167–177.

172. Jordan, I.K., Rogozin I.B., Glazko G.V., Koonin E.V. (2003). Origin of a substantial fraction of human regulatory sequences from transposable elements. *Trends Genet.*, 19, pp. 68-72.

173. Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., Berkum, N.L.V., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., Taatjes, D.J.,

Dekker, J., Young, R.A. (2010). Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, *467*(7314), pp.430–435.

174. Kalyuga, M., Gallego-Ortega, D., Lee, H., Roden, D., Cowley, M., Caldon, C., Stone, A., Allerdice, S., Valdes-Mora, F., Launchbury, R., Statham, A., Armstrong, N., Alles, M., Young, A., Egger, A., Au, W., Piggin, C., Evans, C., Ledger, A., Brummer, T., Oakes, S., Kaplan, W., Gee, J., Nicholson, R., Sutherland, R., Swarbrick, A., Naylor, M., Clark, S., Carroll, J.S. and Ormandy, C. (2012). ELF5 Suppresses Estrogen Sensitivity and Underpins the Acquisition of Antiestrogen Resistance in Luminal Breast Cancer. *PLoS Biology*, 10(12), p.e1001461.

175. Kawagoe, H., Humphries, R.K., Blair, A., Sutherland, H.J., Hogge, D.E. (1999). Expression of HOX genes, HOX cofactors, and MLL in phenotypically and functionally defined subpopulations of leukemic and normal human hematopoietic cells. *Leukemia*, 13(5):687-98.

176. Kennecke, H., Yerushalmi, R., Woods, R., et al. (2010). Metastatic behavior of breast cancer subtypes. *J. Clin. Oncol.*, 28, pp. 3271-3277.

177. Khaled, W. T., Lee, S. C., Stingl, J., Chen, X., Ali, H. R., Rueda, O. M., Hadi, F., Wang, J., Yu, Y., Chin, S.-F., Stratton, M., Futreal, A., Jenkins, N.A., Aparicio, S., Copeland, N.G., Watson, C.J., Caldas, C., Liu, P. (2015). BCL11A is a triple-negative breast cancer gene with critical functions in stem and progenitor cells. *Nature Communications*, 6(1).

178. Kim, D. I., & Roux, K. J. (2016). Filling the Void: Proximity-Based Labeling of Proteins in Living Cells. *Trends in Cell Biology*, *26*(11), pp.804–817.

179. Kim, S.B., Dent, R., Im, S.A., et al. (2017). Ipatasertib plus paclitaxel versus placebo plus paclitaxel as first-line therapy for metastatic triple-negative breast cancer (LOTUS): a multicentre, randomised, double-blind, placebo-controlled, phase 2 trial. *Lancet Oncol.*, 18(10):1360‑1372.

180. Kim, T.-K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., Markenscoff-Papadimitriou, E., Kuhl, D., Bito, H., Worley, P.F., Kreiman, G., Greenberg, M.E. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature*, *465*(7295), pp.182–187.

181. King, M.C., Marks, J.H., Mandell, J.B. (2003). Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. New York Breast Cancer Study Group. *Science*, 302(5645):643.

182. Kirkwood, K. J., Ahmad, Y., Larance, M., Lamond, A. I. (2013). Characterization of native protein complexes and protein isoform variation using size-fractionation-based quantitative proteomics. *Mol. Cell. Proteomics,* 12, pp.3851–3873.

183. Koboldt, D.C., Fulton, R.S., McLellan, M.D. et al. (2012). Comprehensive molecular por- traits of human breast tumours. *Nature*, 490:61-70.

184. Kollias, J., Elston, C.W., Ellis, I.O. et al. (1997). Early-onset breast cancer – histopathologyical prognostic considerations. *Br. J. Cancer*, 75, pp. 1318-1320.

185. Komaki, K., Sano, N., Tangoku, A. (2006). Problems in histological grading of malignancy and its clinical significance in patients with operable Breast Cancer. *Breast Cancer*, 13(3), pp.249–253.

186. Konecny, G. E., Pegram, M. D., Venkatesan, N., Finn, R., Yang, G., Rahmeh, M., Untch, M., Rusnak, D.W., Spehar, G., Mullin, R.J., Keith, B.R., Gilmer, T.M., Berger, M., Podratz, K.C., Slamon, D. J. (2006). Activity of the Dual Kinase Inhibitor Lapatinib (GW572016) against HER-2-Overexpressing and Trastuzumab-Treated Breast Cancer Cells. *Cancer Research*, *66*(3), pp.1630–1639.

187. Konermann, S., Brigham, M. D., Trevino, A. E., Hsu, P. D., Heidenreich, M., Cong, L., Platt, R.J., Scott D.A., Church G.M., Zhang, F. (2013). Optical control of mammalian endogenous transcription and epigenetic states. *Nature*, *500*(7463), pp.472–476.

188. Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., Barcena, C., Hs., P.D, Habib, N., Gootenberg, J.S., Nishimasu, H., Nureki, O., Zhang, F. (2014). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature*, *517*(7536), pp.583–588.

189. Köninki, K., Barok, M., Tanner, M., Staff, S., Pitkänen, J., Hemmilä, P., Ilvesaro, J., Isola, J. (2010). Multiple molecular mechanisms underlying trastuzumab and lapatinib resistance in JIMT-1 breast cancer cells. *Cancer Letters*, *294*(2), pp.211–219.

190. Kramer, K., Sachsenberg, T., Beckmann, B. M., Qamar, S., Boon, K.-L., Hentze, M. W., Kohlbacher, O., Urlaub, H. (2014). Photo-cross-linking and high-resolution mass spectrometry for assignment of RNA-binding sites in RNA-binding proteins. *Nature Methods*, *11*(10), pp.1064–1070.

191. Kulakovskiy, I.V., Boeva, V.A., Favorov, A.V., Makeev, V.J. (2010). Deep and wide digging for binding motifs in ChIP-Seq data. *Bioinformatics*., 26, pp.2622–2623.

192. Kummar, S., Ji, J., Morgan, R., et al. (2012). A phase I study of veliparib in combination with metronomic cyclophosphamide in adults with refractory solid tumors and lymphomas. *Clin Cancer Res*., 18:1726-34.

193. Kwiatkowski, N., Zhang, T., Rahl, P. B., Abraham, B. J., Reddy, J., Ficarro, S. B., Dastur, A., Amzallag, A., Ramaswamy, S., Tesar, B., Jenkins, C.E., Hannett, N.M., McMillin, D., Sanda, T., Sim, T., Kim, N.D., Look, T., Mitsiades, C.S., Weng, A.P., Brown, J.R., Benes, C.H., Marto, J.A., Young, R.A., Gray, N. S. (2014). Targeting transcription regulation in cancer with a covalent CDK7 inhibitor. *Nature*, *511*(7511), pp.616–620.

194. Lakhani, S.R., Ellis, I.O., Schnitt, S.J., Tan, P.H., van de Vijver, M.J. IARC; Lyon, France: 2012. WHO Classification of Breast Tumors.

195. Lam, S. S., Martell, J. D., Kamer, K. J., Deerinck, T. J., Ellisman, M. H., Mootha, V. K., & Ting, A. Y. (2014). Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nature Methods*, *12*(1), pp.51–54.

196. Lane, H.A., Motoyama, A.B., Beuvink, I., Hynes, N.E. (2001). Modulation of p27/Cdk2 complex formation through 4D5-mediated inhibition of HER2 receptor signaling. *Ann Oncol*, 12 Suppl 1:S21–2.

197. Larance, M., & Lamond, A. I. (2015). Multidimensional proteomics for cell biology. *Nature Reviews Molecular Cell Biology*, *16*(5), pp.269–280.

198. Larsson, E., Sander, C., Marks, D. (2010). mRNA turnover rate limits siRNA and microRNA efficacy. *Mol Syst Biol*.; 6():433.

199. Latchman, D.S. (1997). Transcription factors: an overview. *Int. J. Biochem. Cell Biol*., 29, pp.1305-1312.

200. Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature*; 505(7484):495-501.

201. Lazarus, K. A., Pensa, S., Bach, K., Santolla, M. F., Maggiolini, M., Cassidy, J., Batra, A.S., Bruna, A., Mohammed, H., Liu, P., Carroll, J.S., Caldas, C., Marioni, J.C., Khaled, W.T. (unpublished). Bcl11a Marks Mammary Progenitor Cells and Promotes Early Cellular Changes Associated with TNBC by Recruiting Chd8. *SSRN Electronic Journal*.

202. Lazic, S. E. (2015). Ranking, selecting, and prioritising genes with desirability functions. *PeerJ*, *3*.

203. Leach, D.R., Krummel, M.F., Allison, J.P. (1996). Enhancement of antitumor immunity by CTLA-4 blockade. *Science*, 271:1734–6.

204. Lehmann, B.D., Bauer, J.A., Chen, X., Sanders, M.E., Chakravarthy, A.B., Shyr, Y., Pietenpol, J.A. (2011). Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *J Clin Invest*., 121, pp.2750–2767.

205. Lehmann, B.D., Bauer, J.A., Schafer, J.M., Pendleton, C.S., Tang, L., Johnson, K.C., .., Pietenpol, J.A. (2014). PIK3CA mutations in androgen receptor-positive triple negative breast cancer confer sensitivity to the combination of PI3K and androgen receptor inhibitors. *Breast Cancer Res*., 16:406.

206. Leitner, A., Walzthoeni, T., Aebersold, R. (2014). Lysine-specific chemical cross-linking of protein complexes and identification of cross-linking sites using LC-MS/MS and the xQuest/xProphet software pipeline. *Nature Protoc.,* 9, pp.120–137.

207. Lettice, L.A., Heaney, S.J., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Human Molecular Genetics*, *12*(14), pp.1725–1735.

208. Levi, F., Lucchini, F., Negri, E., Vecchia, C.L. (2004). Trends in mortality from major cancers in the European Union, including acceding countries, in 2004. *Cancer*, 101, pp.2843-2850.

209. Li, C., Qi, L., Bellail, A.C., Hao, C., Liu, T. (2014). PD-0332991 induces G1 arrest of colorectal carcinoma cells through inhibition of the cyclin-dependent kinase-6 and retinoblastoma protein axis. *Oncol Lett*, 7, pp1673-8.

210. Li, H., Lee, T.H., Avraham, H. (2002). A novel tricomplex of BRCA1, Nmi, and c-Myc inhibits c-Myc-induced human telomerase reverse transcriptase gene (hTERT) promoter activity in breast cancer. *J Biol Chem*., 277:20965-73.

211. Li, S., Shen, D., Shao, J., Crowder, R., Liu, W., Prat, A., He, X., Liu, S.,.., Ellis, M.J. (2013). Endocrine-therapy-resistant ESR1 variants revealed by genomic characterization of breast-cancer-derived xenografts. *Cell Rep*., 4(6):1116-30.

212. Lim, S., & Kaldis, P. (2013). Cdks, cyclins and CKIs: roles beyond cell cycle regulation. *Development*, *140*(15), pp.3079–3093.

213. Lim, W., Olschwang, S., Keller, J.J., Westerman, A.M., Menko, F.H., .., , Houlston. R.S. (2004). Relative frequency and morphology of cancers in STK11 mutation carriers. *Gastroenterology*, 126(7):1788-9.

214. Lindskog, M., Rockberg, J., Uhlén, M., & Sterky, F. (2005). Selection of protein epitopes for antibody production. *BioTechniques*, *38*(5), pp.723–727.

215. Liu, H., Sadygov, R.G., Yates, J.R. (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem*., 76, pp.4193-4201.

216. Liu, J., Shen, J.X., Wen, X.F., Guo, Y.X., Zhang, G.J. (2016). Targeting Notch degradation system provides promise for breast cancer therapeutics. *Crit Rev Oncol Hematol*., 104():21-9.

217. Liu, T., Yu, J., Deng, M., Yin, Y., Zhang, H., Luo, K., Qin, B., Li, Y., Wu, C., Ren, T., Han, Y., Yin, P., Kim, J., Lee, S., Lin, J., Zhang, L., Zhang, J., Nowsheen, S., Wang, L., Boughey, J., Goetz, M.P., Yuan, J., Lou, Z. (2017). CDK4/6-dependent activation of DUB3 regulates cancer metastasis through SNAIL1. *Nature Communications*, *8*(1).

218. Liu, W., Ma, Q., Wong, K., Li, W., Ohgi, K., Zhang, J., Aggarwal, A., Rosenfeld, M. G. (2013). Brd4 and JMJD6-Associated Anti-Pause Enhancers in Regulation of Transcriptional Pause Release. *Cell*, *155*(7), pp.1581–1595.

219. Liu, X., Shi, Y., Maag, D.X., Palma, J.P., Patterson, M.J., Ellis, P.A., Surber, B.W., ..., Shoemaker, A.R.. (2017). Iniparib nonselectively modifies cysteine-containing proteins in tumor cells and is not a bona fide PARP inhibitor. *Clinical Cancer Research*, 18, pp. 510-523.

220. Liu, X., Zhang, Y., Chen, Y., Li, M., Zhou, F., Li, K., Gu, Z., Dickerson, K.E., Xie, S., Hon, G.C., Xuan, Z., Zhang, M.Q., Shao, Z., Xu, J. (2017). *In Situ* Capture of Chromatin Interactions by Biotinylated dCas9. *Cell*, 170(5).

221. Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell*, 153(2):320-34.

222. Lung, S.Y., Kim, H.Y., Nam, B.H., Min, S.Y., Lee, S.J., Park, C., Kwon, Y., .., Ro J. (2010). Worse prognosis of metaplastic breast cancer patients than other patients with triple-negative breast cancer. *Breast Cancer Res Treat*., 120:627–637.

223. Ma, H., Naseri, A., Reyes-Gutierrez, P., Wolfe, S. A., Zhang, S., & Pederson, T. (2015). Multicolor CRISPR labeling of chromosomal loci in human cells. *Proceedings of the National Academy of Sciences*, *112*(10), pp.3002–3007.

224. Machanick, P., & Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics*. 27, pp.1696–1697.

225. Macias, H., & Hinck, L. (2012). Mammary gland development. *Wiley Interdisciplinary Reviews: Developmental Biology*, *1*(4), pp.533–557.

226. Mackay, A., Weigelt, B., Grigoriadis, A., Kreike, B., Natrajan, R., A'Hern, R., Tan, D.S., Dowsett, M., Ashworth, A., Reis-Filho, J.S. (2011). Microarray-based class discovery for molecular classification of breast cancer: analysis of interobserver agreement. *J Natl Cancer Inst*.; 103(8):662-73.

227. Maeder, M. L., Linder, S. J., Cascio, V. M., Fu, Y., Ho, Q. H., & Joung, J. K. (2013). CRISPR RNA–guided activation of endogenous human genes. *Nature Methods*, *10*(10), pp.977–979.

228. Makarova, K. S., Wolf, Y. I., Alkhnbashi, O. S., Costa, F., Shah, S. A., Saunders, S. J., Barrangou, R., Brouns, S.J., Charpentier, E., Haft, D.H., Horvath, P., Moineau, S., Mojica, F.J., Terns, R.M., Terns, M.P., White, M.F., Yakunin, A.F., Garrett, R.A., van der Oost, J., Backofen, R., Koonin, E. V. (2015). An updated evolutionary classification of CRISPR–Cas systems. *Nature Reviews Microbiology*, *13*(11), pp.722–736.

229. Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., Dicarlo, J. E., Norville, J.L., Church, G. M. (2013). RNA-Guided Human Genome Engineering via Cas9. *Science*, *339*(6121), pp.823–826.

230. Malumbres, M. (2014). Cyclin-dependent kinases. *Genome Biology*, *15*(6), pp.122.

231. Malumbres, M., & Barbacid, M. (2009). Cell cycle, CDKs and cancer: a changing paradigm. *Nature Reviews Cancer*, *9*(3), pp.153–166.

232. Malumbres, M., Sotillo, R., Santamaria, D., Galan, J., Cerezo, A., Ortega, S., et al. (2004). Mammalian cells cycle without the D-type cyclin-dependent kinases Cdk4 and Cdk6. *Cell*, 118, pp.493–504.

233. Mamounas, E.P., Jeong, J.H., Wickerham, D.L., et al. (2008). Benefit from exemestane as extended adjuvant therapy after 5 years of adjuvant tamoxifen: intention-to-treat analysis of the National Surgical Adjuvant Breast and Bowel Project B-33 trial. *J. Clin. Oncol*., pp. 1965-1970.

234. Manavathi, B., & Kumar, R. (2007). Metastasis tumor antigens, an emerging family of multifaceted master coregulators. *J Biol Chem*., 282, pp.1529–1533.

235. Manavathi, B., Samanthapudi, V.S.K., Gajulapalli, V.N.R. (2014). Estrogen receptor coregulators and pioneer factors: the orchestrators of mammary gland cell fate and development. *Front Cell Dev Biol*.; 2: 34.

236. Marhold, J., Brehm, A., Kramer, K. (2004). The *Drosophila* methyl-DNA binding protein MBD2/3 interacts with the NuRD complex via p55 and MI-2. *BMC Mol Biol*., 5:20.

237. Marotti, J. D., Abreu, F. B. D., Wells, W. A., Tsongalis, G. J. (2017). Triple-Negative Breast Cancer. *The American Journal of Pathology*, *187*(10), pp.2133–2138.

238. Mateo, A.M., Pezzi, T.A., Sundermeyer, M., Kelley, C.A., Klimberg, V.S., Pezzi, C.M. (2016). Atypical medullary carcinoma of the breast has similar prognostic factors and survival to typical medullary breast carcinoma: 3,976 cases from the National Cancer Data Base. *J Surg Onc*ol., 114:533–536.

239. Mavaddat, N., Antoniou, A.C., Easton, D.F., Garcia-Closas, M. (2010). Genetic susceptibility to breast cancer. *Mol Oncol*, 4(3):174-91.

240. Mazumdar, A., Wang, R.A., Mishra, S.K., Adam, L., Bagheri-Yarmand, R., Mandal, M,. Vadlamudi, R.K., Kumar, R. (2001). Transcriptional repression of oestrogen receptor by metastasis-associated protein 1 corepressor. *Nat Cell Biol*., 3, pp.30–37.

241. Mellacheruvu, D., Wright Z., […] Nesvizhskii A.I. (2013). The CRAPome: a contaminant repository for affinity purification–mass spectrometry data. *Nature Methods*, 10, pp. 730–736

242. Menon, D. R., Luo, Y., Arcaroli, J. J., Liu, S., Krishnankutty, L. N., Osborne, D. G., Li, Y., Samson, J.M., Bagby, S., Tan, A., Robinson, W.A., Messersmith, W.A., Fujita, M. (2018). CDK1 Interacts with Sox2 and Promotes Tumor Initiation in Human Melanoma. *Cancer Research*, *78*(23), pp.6561–6574.

243. Merenbakh-Lamin, K., Ben-Baruch, N., Yeheskel, A., Dvir, A., Soussan-Gutman, L., Jeselsohn, R., et al. (2013). D538G mutation in estrogen receptor-α: a novel mechanism for acquired endocrine resistance in breast cancer. *Cancer Res*, 73:6856–64.

244. Metz, B., Kersten, G. F. A., Baart, G. J. E., Jong, A. D., Meiring, H., Hove, J. T., van Steenbergen, M.J., Hennink, W.E., Crommelin, D.J. Jiskoot, W. (2006). Identification of Formaldehyde-Induced Modifications in Proteins: Reactions with Insulin. *Bioconjugate Chemistry*, *17*(3), pp.815–822.

245. Metz, B., Kersten, G. F. A., Hoogerhout, P., Brugghe, H. F., Timmermans, H. A. M., Jong, A. D., Meiring, H., ten Hove, J., Hennink, W.E., Crommelin, D.J., Jiskoot, W. (2003). Identification of Formaldehyde-induced Modifications in Proteins. *Journal of Biological Chemistry*, *279*(8), pp.6235–6243.

246. Mills, A.M., Gottlieb, E.C., Wendroth, M.S., Brenin, M.C., Atkins, K.A. (2016). Pure apocrine carcinomas represent a clinicopathologically distinct androgen receptor-positive subset of triple-negative breast cancers. *Am J Surg Pathol*., 40:1109–1116.

247. Mohammed, H., D'Santos, C., Serandour, A. A., Ali, H. R., Brown, G. D., Atkins, A., Rueda, O.M., Holmes, K.A., Theodorou, V., Robinson, J.L., Zwart, W., Saadi, A., Ross-Innes, C.S., Chin, S.F., Menon, S., Stingl, J., Palmieri, C., Caldas, C., Carroll, J. S. (2013). Endogenous Purification Reveals GREB1 as a Key Estrogen Receptor Regulatory Factor. *Cell Reports*, 3(2), pp.342–349.

248. Mohammed, H., Taylor, C., Brown, G. D., Papachristou, E. K., Carroll, J. S., Dsantos, C.S. (2016). Rapid immunoprecipitation mass spectrometry of endogenous proteins (RIME) for analysis of chromatin complexes. Nature *Protocols*, 11(2), pp.316–326.

249. Mojica, F.J., Díez-Villaseñor, C., García-Martínez, J., Almendros, C. (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology*, *155*(3), pp.733–740.

250. Murai, J., Huang, S.Y., Das, B.B., Renaud, A., Zhang, Y., Doroshow, J.H., Ji, J., Takeda, S., Pommier, Y. (2012). Trapping of PARP1 and PARP2 by Clinical PARP Inhibitors. *Cancer Res*., 72(21):5588-99.

251. Murtagh, J., Martin, F. and Gronostajski, R. (2003). The Nuclear Factor I (NFI) gene family in mammary gland development and function. *J Mammary Gland Biol Neoplasia*, 2(8), pp.241-54.

252. Myers, S. A., Wright, J., Peckner, R., Kalish, B. T., Zhang, F., & Carr, S. A. (2018). Discovery of proteins associated with a predefined genomic locus via dCas9–APEX-mediated proximity labeling. *Nature Methods*, *15*(6), 437–439.

253. Nahnsen, S., Bielow, C., Reinert, K., Kohlbacher, O. (2013). Tools for label-free peptide quantification. *Mol. Cell. Proteom*., 12, pp.549-556.

254. Nahta, R., Shabaya, S., Ozbay, T., Rowe, D., Konecny L. (2009). Personalizing HER2-targeted therapy in metastatic breast cancer beyond HER2 status: what we have learned from clinical specimens. *Curr Pharmacogenomics Person Med*, 7(4), pp.263-274.

255. Nanda, R., Chow, L.Q., Dees, E.C., et al. (2016). Pembrolizumab in patients with advanced triple-negative breast cancer: Phase Ib KEYNOTE-012 study. *J Clin Oncol,* 34(21):2460–2467.

256. Narod S. (2010). BRCA mutations in the management of breast cancer: the state of the art. *Nat Rev Clin Oncol*., 7(12):702-7.

257. Ng, C.K.Y., Bidard, F.C., Piscuoglio, S., Geyer, F.C., Lim, R.S., de Bruijn, I., et al. (2017). Genetic heterogeneity in therapy-naive synchronous primary breast cancers and their metastases. *Clin Cancer Res*., 23:4402–15.

258. Nihongaki, Y., Yamamoto, S., Kawano, F., Suzuki, H., Sato, M. (2015). CRISPR-Cas9-based Photoactivatable Transcription System. *Chemistry & Biology*, 22(2), pp.169–174.

259. Niu, J., Andres, G., Kramer, K., Kundranda, M.N., Alvarez, R.H., Klimant, E., Parikh, A.R., Tan, B., Staren, E.D., Markman, M. (2015). Incidence and clinical significance of ESR1 mutations in heavily pretreated metastatic breast cancer patients. *Onco Targets Ther*., 8:3323-8.

260. O'Shaughnessy, J., Osborne, C., Pippen, J., et al. (2009). Efficacy of BSI-201, a poly (ADP-ribose) polymerase-1 (PARP1) inhibitor, in combination with gemcitabine/carboplatin (G/C) in patients with metastatic triple-negative breast cancer (TNBC): results of a randomized phase II trial [abstract 3].. Presented at 2009 *American Society of Clinical Oncology Meeting*; Orlando, FL.

261. O'Shaughnessy, J., Schwartzberg, L., Danso, M.A., Miller, K.D., Rugo, H.S., Neubauer, M., Robert, N., ..., Winer, E.P. (2014). Phase III study of iniparib plus gemcitabine and carboplatin versus gemcitabine and carboplatin in patients with metastatic triple-negative breast cancer. *Journal of Clinical Oncology*, 32, pp. 3840-3847.

262. O'Donovan, N., Byrne, A. T., O'Connor, A. E., Mcgee, S., Gallagher, W. M., Crown, J. (2010). Synergistic interaction between trastuzumab and EGFR/HER-2 tyrosine kinase inhibitors in HER-2 positive breast cancer cells. *Investigational New Drugs*, *29*(5), pp.752–759.

263. Ocana, A., Pandiella, A. (2017). Targeting oncogenic vulnerabilities in triple negative breast cancer: Biological bases and ongoing clinical studies. *Oncotarget*, 8, pp. 22218-22234.

264. Ong, S.-E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., Mann, M. (2002). Stable Isotope Labeling by Amino Acids in Cell Culture, SILAC, as a Simple and Accurate Approach to Expression Proteomics. *Molecular & Cellular Proteomics*, 1(5), pp.376–386.

265. Ortega, S., Malumbres, M., Barbacid, M. (2002). Cyclin D-dependent kinases, INK4 inhibitors and cancer. *Biochim Biophys Acta*; 1602, pp.73-87.

266. Ortega, S., Prieto, I., Odajima, J., Martin, A., Dubus, P., Sotillo, R., et al. (2003). Cyclin-dependent kinase 2 is essential for meiosis but not for mitotic cell division in mice. *Nature genetics*, 35, pp.25–31.

267. Pabo, O., Peisach, E., Grant, R.A. (2001). Design and selection of novel Cys2his2 zinc finger poroteins. *Annu. Rev. Biochem*., 70, pp. 313-340.

268. Pakala, S. B., Singh, K., Reddy, S. D. N., Ohshiro, K., Li, D.-Q., Mishra, L., & Kumar, R. (2011). TGF-β1 signaling targets metastasis-associated protein 1, a new effector in epithelial cells. *Oncogene*, *30*(19), pp.2230–2241.

269. Pardoll, D.M. (2012). The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer*, 12(4):252–264.

270. Park, J. H., Ahn, J.-H., & Kim, S.-B. (2018). How shall we treat early triple-negative breast cancer (TNBC): from the current standard to upcoming immuno-molecular strategies. *ESMO Open*, *3*(Suppl 1).

271. Park, S.Y., Kwon, H.J., Lee, H.E., Ryu, H.S., Kim, S.W., Kim, J.H., et al. (2011). Promoter CpG island hypermethylation during breast cancer progression. *Virchows Arch*, 458, pp.73–84.

272. Perez-Pinera, P., Kocak, D. D., Vockley, C. M., Adler, A. F., Kabadi, A. M., Polstein, L. R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W., Guilak, F., Crawford, G.E., Reddy, T.E., Gersbach, C. A. (2013). RNA-guided gene activation by CRISPR-Cas9–based transcription factors. *Nature Methods*, *10*(10), pp.973–976.

273. Pernas, S., Tolaney, S. M., Winer, E. P., Goel, S. (2018). CDK4/6 inhibition in breast cancer: current practice and future directions. *Therapeutic Advances in Medical Oncology*, *10*, 175883591878645.

274. Perou, C. M., Sørlie, T., Eisen, M. B., Rijn, M. V. D., Jeffrey, S. S., Rees, C. A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S.X., Lønning, P.E., Børresen-Dale, A.L., Brown, P.O., Botstein, D. (2000). Molecular portraits of human breast tumours. *Nature*, *406*(6797), pp.747–752.

275. Persson, M., Andren, Y., Mark, J., Horlings, H., Persson, F. and Stenman, G. (2009). Recurrent fusion of MYB and NFIB transcription factor genes in carcinomas of the breast and head and neck. *Proceedings of the National Academy of Sciences*, 106(44), pp.18740-18744.

276. Peters, A. A., Buchanan, G., Ricciardelli, C., Bianco-Miotto, T., Centenera, M. M., Harris, J. M., Jindal, S., Segara, D., Jia, L., Moore, N.L., Henshall, S.M., Birrell, S.N., Coetzee, G.A., Sutherland, R.L., Butler, L.M., Tilley, W. D. (2009). Androgen Receptor Inhibits Estrogen Receptor- Activity and Is Prognostic in Breast Cancer. *Cancer Research*, *69*(15), pp.6131–6140.

277. Pharaoh, P.D., Day, N.E., Caldas, C. (1999). Somatic mutations in the p53 gene and prognosis in breast cancer: a meta-analysis. *Br J Cancer*., 80, pp.1968-1973.

278. Pharaoh, P.D., Guilford, P., Caldas, C. (2001). Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. International Gastric Cancer Linkage Consortium. *Gastroenterology*, 121(6):1348-53.

279. Piccart-Gebhart, M.J., Procter, M. (2005). B. Leyland-Jones, et al. Trastuzumab after adjuvant chemotherapy in HER2-positive breast cancer. *N. Engl. J. Med*., 353, pp. 1659-1660.

280. Plank, J.L., & Dean, A. (2014). Enhancer Function: Mechanistic and Genome-Wide Insights Come Together. *Molecular cell*., 55, pp.5–14.

281. Plummer R. (2011) Poly(ADP-ribose) polymerase inhibition: a new direction for BRCA and triple-negative breast cancer? *Breast Cancer Res.* 13:218.

282. Polstein, L. R., & Gersbach, C. A. (2015). A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. *Nature Chemical Biology*, 11(3), pp.198–200.

283. Polstein, L. R., Perez-Pinera, P., Kocak, D. D., Vockley, C. M., Bledsoe, P., Song, L., Safi, A., Crawford, G.E., Reddy, T.E., Gersbach, C. A. (2015). Genome-wide specificity of DNA binding, gene regulation, and chromatin remodeling by TALE- and CRISPR/Cas9-based transcriptional activators. *Genome Research*, 25(8), pp.1158–1169.

284. Porrua, O., & Libri, D. (2015). Transcription termination and the control of the transcriptome: why, where and how to stop. *Nature Reviews Molecular Cell Biology*, *16*(3), pp.190–202.

285. Prat, A., Adamo, B., Cheang, M. C. U., Anders, C. K., Carey, L. A., & Perou, C. M. (2013). Molecular Characterization of Basal-Like and Non-Basal-Like Triple-Negative Breast Cancer. *The Oncologist*, *18*(2), pp.123–133.

286. Prat, A., Parker, J. S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J. I., He, X., Perou, C. M. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Research*, *12*(5).

287. Ptashne, M., & Gann, A. (1997). Transcriptional activation by recruitment. *Nature,* 386, pp.569–577.

288. Qi, L. S., Larson, M. H., Gilbert, L. A., Doudna, J. A., Weissman, J. S., Arkin, A. P., & Lim, W. A. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. *Cell*, *152*(5), pp.1173–1183.

289. Rajendran, B.K., & Chu-Xia, D. (2017). Characterization of Potential Driver Mutations Involved in Human Breast Cancer by Computational Approaches. *Oncotarget*, 8(30).

290. Rapakko, K., Allinen, M., Syrjakoski, K., et al. (2001). Germline TP53 alterations in Finnish breast cancer families are rare and occur at conserved mutation-prone sites. *Br J Cancer*, 84: 116–119.

291. Ray, P., Wang, J., Qu, Y., Sim, M., Shamonki, J., Bagaria, S., Ye, X., Liu, B., Elashoff, D., Hoon, D., Walter, M., Martens, J., Richardson, A., Giuliano, A. and Cui, X. (2010). FOXC1 Is a Potential Prognostic Biomarker with Functional Significance in Basal-like Breast Cancer. *Cancer Research*, 70(10), pp.3870-3876.

292. Rayburn, E., Zhang, R., He, J., Wang, H. (2005). MDM2 and human malignancies: expression, clinical pathology, prognostic markers, and implications for chemotherapy. *Curr Cancer Drug Targets*, 5(1):27-41.

293. Rees, J. S., Li, X.-W., Perrett, S., Lilley, K. S., & Jackson, A. P. (2015). Protein Neighbors and Proximity Proteomics. *Molecular & Cellular Proteomics*, *14*(11), pp.2848–2856.

294. Reinert, T., Saad, E.D., Barrios, C.H., Bines, J. (2017). Clinical Implications of ESR1 Mutations in Hormone Receptor-Positive Advanced Breast Cancer. *Front Oncol*., 7:26.

295. Rhee, D.K., Park, S.H., Jang, Y.K. (2008). Molecular signatures associated with transformation and progression to breast cancer in the isogenic MCF10 model. Genomics, 92, pp.419–428.

296. Rhee, H.-W., Zou, P., Udeshi, N. D., Martell, J. D., Mootha, V. K., Carr, S. A., Ting, A. Y. (2013). Proteomic Mapping of Mitochondria in Living Cells via Spatially Restricted Enzymatic Tagging. *Science*, *339*(6125), pp.1328–1331.

297. Rhee, H.S. & Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. *Cell*., 147, pp.1408–1419.

298. Roeder, R. G. (2005). Transcriptional regulation and the role of diverse coactivators in animal cells. *FEBS Lett.* 579, pp.909–915.

299. Ross, P.L., Huang, Y.N., Marchese, J.N., Williamson, B., Parker, K., Hatta, S., et al. (2004). Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents. *Mol. Cell. Proteom.*, 3, pp.1154-1169.

300. Rouleau, M., Patel, A., Hendzel, M.J., et al. (2010). PARP inhibition: PARP1 and beyond. *Nat Rev Cancer*, 10:293–301.

301. Roux, K. J., Kim, D. I., Raida, M. & Burke, B. (2012). A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J. Cell Biol.* 196, pp.801–810.

302. Roviello, G., Milani, M., Gobbi, A., Dester, M., Cappelletti, M.R., Allevi, G., Aguggini, S., ..., Generali, D.. (2016). A Phase II study of olaparib in breast cancer patients: Biological evaluation from a 'window of opportunity' trial. *Future Oncology*, 12, pp. 2189-2193.

303. Sainsbury, S., Bernecky, C., Cramer, P. (2015). Structural basis of transcription initiation by RNA polymerase II. *Nature Reviews Molecular Cell Biology*, *16*(3), pp.129–143.

304. Santagata, S., Thakkar, A., Ergonul, A., Wang, B., Woo, T., Hu, R., Harrell, J.C., McNamara, G., Schwede, M., Culhane, A.C., Kindelberger, D., Rodig, S., Richardson, A., Schnitt, S.J., Tamimi, R.M., Ince, T. A. (2014). Taxonomy of breast cancer based on normal cell phenotype predicts outcome. *Journal of Clinical Investigation*, *124*(2), pp.859–870.

305. Santamaria, D., Barriere, C., Cerqueira, A., Hunt, S., Tardy, C., Newton, K., et al. (2007). Cdk1 is sufficient to drive the mammalian cell cycle. *Nature*, 448, pp.811–815.

306. Satyanarayana, A., Berthet, C., Lopez-Molina, J., Coppola, V., Tessarollo, L., Kaldis, P. (2008). Genetic substitution of Cdk1 by Cdk2 leads to embryonic lethality and loss of meiotic function of Cdk2. *Development,* 135, pp.3389–3400.

307. Saunders, A., Core, L. J., Lis, J. T. (2006). Breaking barriers to transcription elongation. *Nature Rev. Mol. Cell. Biol.* 7, pp.557–567.

308. Schmid, P., Abraham, J., Chan, S., Wheatley, D., Brunt, A.M., Nemsadze, G., .., Turner, N.C. (2020). Capivasertib Plus Paclitaxel Versus Placebo Plus Paclitaxel As First-Line Therapy for Metastatic Triple-Negative Breast Cancer: The PAKT Trial. *Journal of Clinical Oncology*, 38:5, 423-433.

309. Schreiber, V., Dantzer, F., Ame, J.C., de Murcia, G. (2006). Poly(ADP-ribose): novel functions for an old molecule. *Nat Rev Mol Cell Biol*., 7(7):517-28.

310. Schrijver, WAME, Selenica, P., Lee, J.Y., et al. (2018). Mutation Profiling of Key Cancer Genes in Primary Breast Cancers and Their Distant Metastases. *Cancer Res*.; 78(12):3112–3121.

311. Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A., Field, Y., Moore, I. K., Wang, J.-P.Z., Widom, J. (2006). A genomic code for nucleosome positioning. *Nature*, *442*(7104), pp.772–778.

312. Segall, M.D. (2012). Multi-parameter optimization: identifying high quality compounds with a balance of properties. *Current Pharmaceutical Design*; 18(9), pp.1292–1310.

313. Sen, N., Gui, B., Kumar, R. (2014). Role of MTA1 in cancer progression and metastasis. *Cancer and Metastasis Reviews*, *33*(4), pp.879–889.

314. Sgroi, D.C., Sestak, I., Cuzick, J. et al. (2013). Prediction of late distant recurrence in patients with oestrogen-receptor -positive breast cancer: a prospective comparison of the breast-cancer index (BCI) assay, 21-gene recurrence score, and IHC4 in the TransATAC study population. *Lancet Oncol*., 14, pp. 1067-1070.

315. Shah, S. P., Morin, R. D., Khattra, J., Prentice, L., Pugh, T., Burleigh, A., … Aparicio, S. (2009). Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature*, *461*(7265), pp.809–813.

316. Shah, S.P., Roth, A., Goya, R., et al. (2012). The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature*, 486:395-9.

317. Sharpless, N.E., & Sherr, C.J. (2015). Forging a signature of in vivo senescence. *Nat Rev Cancer*, 15, pp.397–408.

318. Shiau, A. K., Barstad, D., Loria, P. M., Cheng, L., Kushner, P. J., Agard, D. A., Greene, G. L. (1998). The Structural Basis of Estrogen Receptor/Coactivator Recognition and the Antagonism of This Interaction by Tamoxifen. *Cell*, *95*(7), pp.927–937.

319. Shiovitz, S. & Korde L.A. (2015). Genetics of breast cancer: a topic in evolution. *Ann Oncol*., 26(7): 1291–1299.

320. Sidoli, S., Kulej, K., Garcia, B. A. (2016). Why proteomics is not the new genomics and the future of mass spectrometry in cell biology. *The Journal of Cell Biology*, *216*(1), pp.21–24.

321. Siggers, T., & Gordân, R. (2014). Protein-DNA binding: complexities and multi-protein codes. *Nucleic Acids Res*., 42, pp.2099–2111.

322. Smid, M., Wang, Y., Zhang Y., et al. (2008). Subtypes of breast cancer show preferential site of relapse. *Cancer Res*., 68, pp. 3108-3114.

323. Smith, P.D., Crossland, S., Parker, G., Osin, P., Brooks, L., Waller, J., Philp, E., Crompton, M.R., Gusterson, B.A., Allday, M.J., Crook. T. (1999). Novel p53 mutants selected in BRCA-associated tumours which dissociate transformation suppression from other wild-type p53 functions. *Oncogene*, 18, pp.2451-2459.

324. Soni, A., Ren, Z., Hameed, O. et al. (2015). Breast cancer subtypes predispose the site of distant metastases. *Am. J. Clin. Pathol*., 143, pp. 471-478;

325. Soon-Shiong, C P., Rabizadeh, S., Benz, S. et al. (2016). Integrating whole exome sequencing data with RNAseq and quantitative proteomics to better

inform clinical treatment decisions in patients with metastatic triple negative breast cancer. *Cancer Res*., 76, P6-05-08.

326. Sørlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Thorsen, T., Quist, H., Matese, J.C., Brown, P.O., Botstein, D., Lønning, P.E., Borresen-Dale, A.-L. (2001). Gene expression patterns of breast carcinomas distinguish tumour subclasses with clinical implications. *Proceedings of the National Academy of Sciences*, *98*(19), pp.10869–10874.

327. Sørlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J. S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C.M., Lønning, P.E., Brown, P.O., Børresen-Dale, A.L., Botstein, D. (2003). Repeated observation of breast tumour subtypes in independent gene expression data sets. *Proceedings of the National Academy of Sciences*, 100(14), pp.8418–8423.

328. Sporikova, Z., Koudelakova, V., Trojanec, R., Hajduch, M. (2018). Genetic Markers in Triple-Negative Breast Cancer. *Clin Breast Cancer*; 18(5):e841-e850.

329. Srinivasa, S., Ding, X., Kast, J. (2015) Formaldehyde cross-linking and structural proteomics: bridging the gap. *Methods,* 89, pp.91–98.

330. Srivastava, S., Matsuda, M., Hou, Z., Bailey, J.P., Kitazawa, R., Herbst, M.P., Horseman, N.D. (2003). Receptor activator of NF-kappaB ligand induction via Jak2 and Stat5a in mammary epithelial cells. *J. Biol. Chem.*, *278*(46), pp.46171-46178.

331. Stanton, S.E., Adams, S., Disis, M.L. (2016). Variation in the incidence and magnitude of tumor-infiltrating lymphocytes in breast cancer subtypes: a systematic review. *JAMA Oncol*., 2:1354–60.

332. Steinhilber, D., Marschalek, R. (2018). How to effectively treat acute leukemia patients bearing MLL-rearrangements? *Biochem Pharmacol*., 147():183-190.

333. Stephens, P. J., Tarpey, P. S., Davies, H., Loo, P. V., Greenman, C., Wedge, D. C., …, Stratton, M. R. (2012). The landscape of cancer genes and mutational processes in breast cancer. *Nature*, *486*(7403), pp.400–404.

334. Sternlicht, M. D. (2006). Key stages in mammary gland development: The cues that regulate ductal branching morphogenesis. *Breast Cancer Research*, 8(1).

335. Stratton, M.R., Campbell, P.J., Futreal, P.A. (2009). The cancer genome. *Nature*, 458, pp.719–24.

336. Stratton, M.R., Rahman, N. (2008). The emerging landscape of breast cancer susceptibility. *Nat Genet*, 40(1):17-22.

337. Sun, J., Zheng, Z., Chen, Q., Pan, Y., Lu, H., Zhang, H., Yu, Y., Dai, Y. (2019). NRF3 suppresses breast cancer cell metastasis and cell proliferation and is a favorable predictor of survival in breast cancer. OncoTargets and Therapy, (2), pp. 3019—3030

338. Sun, W.Y., Lee, Y.K., Koo, J.S. (2016). Expression of PD-L1 in triple-negative breast cancer based on different immunohistochemical antibodies. *J Transl Med*., 14(1):173.

339. Sung, M.H., Baek, S., Hager, G.L. (2016). Genome-wide footprinting: ready for prime time? *Nat Methods*.,13, pp.222–228.

340. Sung, M.H., Guertin, M.J., Baek, S., Hager, G.L. (2014). DNase footprint signatures are dictated by factor dynamics and DNA sequence. *Mol Cell.,* 56, pp.275–285.

341. Sutherland, B. W., Toews, J., Kast, J. (2008). Utility of formaldehyde cross-linking and mass spectrometry in the study of protein–protein interactions. *Journal of Mass Spectrometry*, 43(6), pp.699–715.

342. Suttipong, S., Xiao, H., Smeekensa, J., Wu, R. (2017). "Evaluation and Optimization of Reduction and Alkylation Methods to Maximize Peptide Identification with MS-Based Proteomics." *Molecular BioSystems*, 13(12), pp. 2574–2582.

343. Swisher, E.M., Sakai, W., Karlan, B.Y. et al. (2008). Secondary BRCA1 mutations in BRCA1-mutated ovarian carcinomas with platinum resistance. *Cancer Res*, 68(8): 2581–2586.

344. Tan, M.H., Mester, J.L., Ngeow, J., Rybicki, L.A., Orloff, M.S., Eng, C. (2012). Lifetime cancer risks in individuals with germline PTEN mutations. *Clin Cancer Res.*, 18(2):400-7.

345. TCGA. Cancer Genome Atlas Network. (2012). Comprehensive molecular portraits of human breast tumours. *Nature*, 490(7418), pp.61-70.

346. Teo, Z. L., Versaci, S., Dushyanthen, S., Caramia, F., Savas, P., Mintoff, C. P., Zethoven, M., Virassamy, B., Luen, S.J., McArthur, G.A., Phillips, W.A., Darcy, P.K., Loi, S. (2017). Combined CDK4/6 and PI3Kα Inhibition Is Synergistic and Immunogenic in Triple-Negative Breast Cancer. *Cancer Research*, *77*(22), 6340–6352.

347. Theodorou, V., Stark, R., Menon, S., Carroll, J. S. (2012). GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. *Genome Research*, *23*(1), pp.12–22.

348. Tilley, W.D., Clarke, C.L., Birrell, S.N., Bruchovsky, N. (2001). Hormones and cancer: new insights, new challenges. Trends Endocrinol. Metabol., 12(5), pp. 186-188.

349. Toews, J., Rogalski, J. C., Clark, T. J., & Kast, J. (2008). Mass spectrometric identification of formaldehyde-induced peptide modifications under in vivo protein cross-linking conditions. *Analytica Chimica Acta*, 618(2), pp.168–183.

350. Toy, W., Shen, Y., Won, H., Green, B., Sakr, R.A., Will, M., et al. (2013). ESR1 ligand-binding domain mutations in hormone-resistant breast cancer. *Nat Genet*, 45:1439–45.

351. Trinkle-Mulcahy, L., Boulon, S., Lam, Y. W., Urcia, R., Boisvert, F.-M., Vandermoere, F., Morrice, N.A., Swift, S., Rothbauer, U., Leonhardt, H., Lamond, A. (2008). Identifying specific protein interaction partners using quantitative mass spectrometry and bead proteomes. *The Journal of Cell Biology*, *183*(2), pp.223–239.

352. Trivers, K. F., Lund, M. J., Porter, P. L., Liff, J. M., Flagg, E. W., Coates, R. J., & Eley, J. W. (2009). The epidemiology of triple-negative breast cancer, including race. *Cancer Causes & Control*, *20*(7), pp.1071–1082.

353. Tsuda, H., Takarabe, T., Hasegawa, F., Fukutomi, T., Hirohashi, S. (2000). Large, central acellular zones indicating myoepithelial tumour differentiation in high-grade invasive ductal carcinomas as markers of predisposition to lung and brain metastases. *Am J Surg Pathol.,* 24(2), pp.197-202.

354. Turnbull, C., & Rahman, N. (2008). Genetic predisposition to breast cancer: past, present, and future. *Annu Rev Genomics Hum Genet.,* 9, pp.321–45.

355. Tutt, A., Robson, M., Garber, JE., et al. (2009). Phase II trial of the oral PARP inhibitor olaparib in BRCA-deficient advanced breast cancer [abstract CRA501]. Presented at 2009 *American Society of Clinical Oncology Meeting*; Orlando, FL.

356. Valabrega G., Montemurro F., Aglietta M. (2007). Trastuzumab: mechanism of action, resistance and future perspectives in HER2-overexpressing breast cancer. *Ann Oncol*., 18(6):977–84.

357. Van Dam, H., & Castellazzi, M. (2001). Distinct roles of Jun: fos and Jun: ATF dimers in oncogenesis. *Oncogene*, 20(19) pp.2453.

358. van den Heuvel, S., & Dyson, N.J. (2008). Conserved functions of the pRB and E2F families. *Nat Rev Mol Cell Biol*.; 9(9), pp.713-24.

359. Van der Oost, J., Westra, E. R., Jackson, R. N., & Wiedenheft, B. (2014). Unravelling the structural and mechanistic basis of CRISPR–Cas systems. *Nature Reviews Microbiology*, *12*(7), pp.479–492.

360. Van Laere, S.J., Van der Auwera, I., Van den Eynden, G.G., Elst, H.J., Weyler, J., Harris, A.L., van Dam, P., Van Marck, E.A., Vermeulen, P.B., Dirix, L.Y. (2006). Nuclear factor-kappaB signature of inflammatory breast cancer by cDNA microarray validated by quantitative real-time reverse transcription-PCR, immunohistochemistry, and nuclear factor-kappaB DNA-binding. *Clin. Cancer Res., 12*(11 Pt 1), pp.3249-3256.

361. Venkatesh, S., Smolle, M., Li, H., Gogol, M. M., Saint, M., Kumar, S., Natarajan, K., Workman, J. L. (2012). Set2 methylation of histone H3 lysine 36 suppresses histone exchange on transcribed genes. *Nature*, *489*(7416), pp.452–455.

362. Venkitaraman, A.R. (2002) Cancer susceptibility and the functions of BRCA1 and BRCA2. *Cell*, 108:171-82.

363. Viale, G. (2012). The current state of breast cancer classification. *Annals of Oncology*, *23*(suppl 10), pp.x207–x210.

364. Viale, G., Regan, M.M., Maiorano, E., et al. (2007). Prognostic and predictive value of centrally reviewed expression of estrogen and progesterone receptors in a randomized trial comparing letrozole and tamoxifen adjuvant therapy for postmenopausal early breast cancer: BIG 1-98. *J. Clin. Oncol*., 5, pp. 3846-3850;

365. Vikas, P., Borcherding, N., Zhang, W. (2018). The clinical promise of immunotherapy in triple-negative breast cancer. *Cancer Manag Res*, 10:6823-6833.

366. Vinayak, S., Ford, J.M. (2010). PARP Inhibitors for the Treatment and Prevention of Breast Cancer. *Current Breast Cancer Reports*, 2(4):190-197.

367. Vinayak, S., Gray, R.J., Adams, S. (2017). Association of increased tumor-infiltrating lymphocytes (TILs) with immunomodulatory (IM) triple-negative breast cancer (TNBC) subtype and response to neoadjuvant platinum-based therapy in PrECOG0105. *J Clin Oncol*., 32:1000.

368. Voduc, D., Cheang, M., Nielsen, T. (2008). GATA-3 Expression in Breast Cancer Has a Strong Association with Estrogen Receptor but Lacks Independent Prognostic Value. Cancer Epidemiol Biomarkers Prev; 17(2).

369. Vranic, S., Tawfik, O., Palazzo, J., Bilalovic, N., Eyzaguirre, E., Lee, L.M., Adegboyega, P., Hagenkord, J., Gatalica, Z. (2010). EGFR and HER-2/neu expression in invasive apocrine carcinoma of the breast. *Mod Pathol*., 23:644–653.

370. Waldman, F.M., DeVries, S., Chew, K.L., Moore, D.H., Kerlikowske, K. Ljung, B.M. (2000). Chromosomal Alterations in Ductal Carcinomas *In Situ* and Their *In Situ* Recurrences. *Journal of the National Cancer Institute*, *92*(4), pp.313–320.

371. Walsh, T., Casadei, S., Coats, K.H., Swisher, E., Stray, S.M., Higgins, J., .., King, M.C. (2006). Spectrum of mutations in BRCA1, BRCA2, CHEK2, and TP53 in families at high risk of breast cancer. JAMA, 295(12):1379-88.

372. Wang, H., Russa, M. L., Qi, L. S. (2016). CRISPR/Cas9 in Genome Editing and Beyond. *Annual Review of Biochemistry*, *85*(1), pp.227–264.

373. Wang, Q., Zhang, H., Kajino, K., Greene, M.I. (1998). BRCA1 binds c-Myc and inhibits its transcriptional and transforming activity in cells. *Oncogene*, 17:1939-48.

374. Wang, Y., Zhang, T., Kwiatkowski, N., Abraham, B., Lee, T., Xie, S., Yuzugullu, H., Von, T., Li, H., Lin, Z., Stover, D., Lim, E., Wang, Z., Iglehart, J., Young, R., Gray, N. and Zhao, J. (2015). CDK7-Dependent Transcriptional Addiction in Triple-Negative Breast Cancer. Cell, 163(1), pp.174-186.

375. Watkins, J., Weekes, D., Shah, V., et al. (2015). Genomic Complexity Profiling Reveals That HORMAD1 Overexpression Contributes to Homologous Recombination Deficiency in Triple-Negative Breast Cancers. *Cancer Discov*., 5(5):488-505.

376. Watson, C. J., & Khaled, W. T. (2008). Mammary development in the embryo and adult: a journey of morphogenesis and commitment. *Development*, *135*(6), pp.995–1003.

377. Weake, V. M., & Workman, J. L. (2010). Inducible gene expression: diverse regulatory mechanisms. *Nature Reviews Genetics*, *11*(6), pp.426–437.

378. Weber, A., Borghouts, C., Brendel, C., Moriggl, R., Delis, N., Brill, B., Vafaizadeh, V., Groner, B. (2013). The inhibition of stat5 by a Peptide aptamer ligand specific for the DNA binding domain prevents target gene transactivation and the growth of breast and prostate tumor cells. *Pharmaceuticals (Basel),* 6(8):960-8.

379. Wedge, S.R., Kendrew, J., Hennequin, L.F., Valentine, P.J., Barry, S.T., Brave, S.R., Smith, N.R.,..., Ogilvie, D.J.. (2005). AZD2171: A highly potent,

orally bioavailable, vascular endothelial growth factor receptor-2 tyrosine kinase inhibitor for the treatment of cancer. *Cancer Research*, 65, pp. 4389-4400.

380. Weigelt, B., Eberle, C., Cowell, C.F., Ng, C.K., Reis-Filho J.S. (2014). Metaplastic breast carcinoma: more than a special type. *Nat Rev Cancer,* 14:147–14.

381. Weigelt, B., Mackay, A., A'hern, R., Natrajan, R., Tan, D.S., Dowsett, M., Ashworth, A., Reis-Filho (2010). Breast cancer molecular profiling with single sample predictors: a retrospective analysis. *JS Lancet Oncol*, (4):339-49.

382. Weigelt, B., Ng, C.K., Shen, R., Popova, T., Schizas, M., Natrajan, R., Mariani, O., .., Reis-Filho, J.S. (2015). Metaplastic breast carcinomas display genomic and transcriptomic heterogeneity [corrected]. *Mod Pathol*., 28:340–351.

383. Weisbrod, C. R., Chavez, J. D., Eng, J. K., Yang, L., Zheng, C., Bruce, J. E. (2013). In VivoProtein Interaction Network Identified with a Novel Real-Time Cross-Linked Peptide Identification Strategy. *Journal of Proteome Research*, *12*(4), pp.1569–1579.

384. Weisman, P.S., Ng, C.K., Brogi, E., Eisenberg, R.E., Won, H.H., Piscuoglio, S., .., Wen H.Y. (2016). Genetic alterations of triple negative breast cancer by targeted next-generation sequencing and correlation with tumor morphology. *Mod Pathol*., 29:476–488.

385. Weng, W., Yin, J., Zhang, Y., Qiu, J., Wang, X. (2014). Metastasis-associated protein 1 promotes tumor invasion by downregulation of E-cadherin. *International Journal of Oncology*, *44*(3), pp.812–818.

386. Winters, S., Martin, C., Murphy, D., Shokar, N.K. (2017). Breast Cancer Epidemiology, Prevention, and Screening. *Progress in Molecular Biology and Translational Science Approaches to Understanding Breast Cancer*, 151, pp.1–32.

387. World Health Organization. (2018). Breast Cancer.

388. Xia, L., Huang, W., Tian, D., Zhu, H., Qi, X., Chen, Z., Zhang, Y., Hu, H., Fan, D., Nie, Y. and Wu, K. (2013). Overexpression of forkhead box C1 promotes tumor metastasis and indicates poor prognosis in hepatocellular carcinoma. *Hepatology*, 57(2), pp.610-624.

389. Yakes, F.M., Chinratanalab, W., Ritter, C.A., King, W., Seelig, S., Arteaga, C.L. (2002). Herceptin-induced inhibition of phosphatidylinositol-3 kinase and Akt Is required for antibody-mediated effects on p27, cyclin D1, and antitumor action. *Cancer Res*., 62(14):4132-41.

390. Yang, J., Mani, S. A., Donaher, J. L., Ramaswamy, S., Itzykson, R. A., Come, C., Savagner, P., Gitelman, I., Richardson, A., Weinberg, R. A. (2004). Twist, a Master Regulator of Morphogenesis, Plays an Essential Role in Tumor Metastasis. Cell, 117(7), pp.927–939.

391. Yarden, Y., & Sliwkowski, M. X. (2001). Untangling the ErbB signalling network. *Nature Reviews Molecular Cell Biology*, *2*(2), 127–137.

392. Yates, J. R., Ruse, C. I., Nakorchevsky, A. (2009). Proteomics by mass spectrometry: approaches, advances, and applications. *Annu. Rev. Biomed. Engineer.* 11, pp.49–79.

393. Yates, L.R., Gerstung, M., Knappskog, S., et al. (2015). Subclonal diversification of primary breast cancer revealed by multiregion sequencing. *Nat Med*., 21(7):751‑759.

394. You, Z., Madrid, L.V., Saims, D., Sedivy, J., Wang, C.Y. (2002). c-Myc sensitizes cells to tumor necrosis factor-mediated apoptosis by inhibiting nuclear factor kappa B transactivation. *J Biol Chem*., 277:36671-7.

395. Younger, S. T., & Corey, D. R. (2011). Transcriptional gene silencing in mammalian cells by miRNA mimics that target gene promoters. *Nucleic Acids Research*, *39*(13), pp.5682–5691.

396. Yu-Rice, Y., Jin, Y., Han, B., Qu, Y., Johnson, J., Watanabe, T., Cheng, L., Deng, N., Tanaka, H., Gao, B., Liu, Z., Sun, Z., Bose, S., Giuliano, A. and Cui, X. (2016). FOXC1 is involved in ERα silencing by counteracting GATA3 binding and is implicated in endocrine resistance. *Oncogene*, 35(41), pp.5400-5411.

397. Yu, Q., Geng, Y., Sicinski, P. (2001). Specific protection against breast cancers by cyclin D1 ablation. *Nature*, 411, pp.1017–1021.

398. Yu, Q., Sicinska, E., Geng, Y., Ahnström, M., Zagozdzon, A., Kong, Y., Gardner, H., Kiyokawa, H., Harris, L.N., Stål, O., Sicinski, P. (2006). Requirement for CDK4 kinase function in breast cancer. *Cancer Cell*, *9*(1), pp.23–32.

399. Zalatan, J. G., Lee, M. E., Almeida, R., Gilbert, L. A., Whitehead, E. H., La Russa, M., Tsai, J.C., Weissman, J.S., Dueber, J.E., Qi L.S., Lim, W. A. (2015). Engineering Complex Synthetic Transcriptional Programs with CRISPR RNA Scaffolds. *Cell*, *160*(1-2), pp.339–350.

400. Zetsche, B., Volz, S. E., Zhang, F. (2015). A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nature Biotechnology*, 33(2), pp.139–142.

401. Zhang, B., Zhang, H., Shen, G. (2015). Metastasis-associated protein 2 (MTA2) promotes the metastasis of non-small-cell lung cancer through the inhibition of the cell adhesion molecule Ep-CAM and E-cadherin. *Japanese Journal of Clinical Oncology*, *45*(8), pp.755–766.

402. Zhang, H. (2006). Metastasis Tumor Antigen Family Proteins during Breast Cancer Progression and Metastasis in a Reliable Mouse Model for Human Breast Cancer. *Clinical Cancer Research*, *12*(5), pp.1479–1486.

403. Zhang, H., Singh, R.R., Talukder, A.H., Kumar, R. (2006). Metastatic tumor antigen 3 is a direct corepressor of the Wnt4 pathway. *Genes Dev*., 20, pp.2943–2948.

404. Zhou, B.P., Liao, Y., Xia, W., Zou, Y., Spohn, B., Hung, M.C. (2001). HER-2/neu induces p53 ubiquitination via Akt-mediated MDM2 phosphorylation. *Nat Cell Biol*., 3(11):973-82.

405. Zhou, C., Ji, J., Cai, Q., Shi, M., Chen, X., Yu, Y., Zhu, Z., Zhang, J. (2015). MTA2 enhances colony formation and tumor growth of gastric cancer cells through IL-11. *BMC Cancer*, *15*(1).

406. Zhu, X., Li, Y., Shen, H., Li, H., Long, L., Hui, L., Xu, W. (2013). miR-137 inhibits the proliferation of lung cancer cells by targeting Cdc42 and Cdk6. *FEBS Lett,* 587, pp.73-81.

# APPENDICES

## APPENDIX A: EXAMPLE PROTEOME DISCOVERER DATA OF BCL11A RIME EXPERIMENT

| **UniProt Accession Number** | **Protein Description** | **Protein Name** | **# of unique pepti-des** | **Sequence coverage (%Cov:)** |
|---|---|---|---|---|
| P49454 | Centromere protein F | CENPF | 15 | 8.10% |
| P49419 | Alpha-aminoadipic semialdehyde dehydrogenase | AL7A1 | 13 | 24.68% |
| P68032 | Actin, alpha cardiac muscle 1 | ACTC | 12 | 36.34% |
| Q9HCK8 | Chromodomain-helicase-DNA-binding protein 8 | CHD8 | 12 | 9.41% |
| P55884 | Eukaryotic translation initiation factor 3 subunit B | EIF3B | 12 | 17.44% |
| Q6UB99 | Ankyrin repeat domain-containing protein 11 | ANR11 | 11 | 5.29% |
| Q8IXT5 | RNA-binding protein 12B | RB12B | 11 | 9.99% |
| O75717 | WD repeat and HMG-box DNA-binding protein 1 | WDHD1 | 11 | 15.59% |
| Q9UQE7 | Structural maintenance of chromosomes protein 3 | SMC3 | 10 | 12.49% |
| P21127 | Cyclin-dependent kinase 11B | CD11B | 9 | 9.56% |

| | | | | |
|---|---|---|---|---|
| Q9H8M2 | Bromodomain-containing protein 9 | BRD9 | 8 | 22.11% |
| Q01804 | OTU domain-containing protein 4 | OTUD4 | 8 | 8.17% |
| Q15366 | Poly(rC)-binding protein 2 | PCBP2 | 8 | 25.21% |
| P53999 | Activated RNA polymerase II transcriptional coactivator p15 | TCP4 | 8 | 51.97% |
| P26368 | Splicing factor U2AF 65 kDa subunit | U2AF2 | 8 | 36.42% |
| Q9UHB7 | AF4/FMR2 family member 4 | AFF4 | 7 | 11.44% |
| P40121 | Macrophage-capping protein | CAPG | 7 | 25.57% |
| Q96JM3 | Chromosome alignment-maintaining phosphoprotein 1 | CHAP1 | 7 | 11.08% |
| Q14839 | Chromodomain-helicase-DNA-binding protein 4 | CHD4 | 7 | 6.80% |
| O15320 | cTAGE family member 5 | CTGE5 | 7 | 11.19% |
| O15371 | Eukaryotic translation initiation factor 3 subunit D | EIF3D | 7 | 16.42% |
| Q9BY77 | Polymerase delta-interacting protein 3 | PDIP3 | 7 | 28.50% |
| Q16576 | Histone-binding protein RBBP7 | RBBP7 | 7 | 18.59% |
| Q9NTZ6 | RNA-binding protein 12 | RBM12 | 7 | 7.62% |
| Q07020 | 60S ribosomal protein L18 | RL18 | 7 | 31.91% |
| P62847 | 40S ribosomal protein S24 | RS24 | 7 | 39.10% |
| Q14683 | Structural maintenance of chromosomes protein 1A | SMC1A | 7 | 6.97% |
| Q92922 | SWI/SNF complex subunit SMARCC1 | SMRC1 | 7 | 10.95% |
| Q01130 | Serine/arginine-rich splicing factor 2 | SRSF2 | 7 | 23.08% |
| Q15029 | 116 kDa U5 small nuclear ribonucleoprotein component | U5S1 | 7 | 10.80% |
| Q6PJT7 | Zinc finger CCCH domain-containing protein 14 | ZC3HE | 7 | 13.04% |
| Q9BRD0 | BUD13 homolog | BUD13 | 6 | 14.22% |
| Q2TBE0 | CWF19-like protein 2 | C19L2 | 6 | 8.17% |
| P52907 | F-actin-capping protein subunit alpha-1 | CAZA1 | 6 | 24.48% |
| Q9NZ63 | Uncharacterized protein C9orf78 | CI078 | 6 | 29.07% |
| O75534 | Cold shock domain-containing protein E1 | CSDE1 | 6 | 10.65% |

| | | | | |
|---|---|---|---|---|
| P35659 | Protein DEK | DEK | 6 | 14.40% |
| P49411 | Elongation factor Tu, mitochondrial | EFTU | 6 | 15.27% |
| P60228 | Eukaryotic translation initiation factor 3 subunit E | EIF3E | 6 | 13.71% |
| P14625 | Endoplasmin | ENPL | 6 | 11.33% |
| P25205 | DNA replication licensing factor MCM3 | MCM3 | 6 | 10.52% |
| P55081 | Microfibrillar-associated protein 1 OS | MFAP1 | 6 | 20.73% |
| Q9Y3C1 | Nucleolar protein 16 OS | NOP16 | 6 | 43.82% |
| O60828 | Polyglutamine-binding protein 1 | PQBP1 | 6 | 39.25% |
| O94906 | Pre-mRNA-processing factor 6 | PRP6 | 6 | 8.29% |
| Q6NZI2 | Polymerase I and transcript release factor | PTRF | 6 | 12.82% |
| P62750 | 60S ribosomal protein L23a | RL23A | 6 | 37.18% |
| O15160 | DNA-directed RNA polymerases I and III subunit RPAC1 | RPAC1 | 6 | 18.50% |
| Q9Y265 | RuvB-like 1 | RUVB1 | 6 | 20.18% |
| O95391 | Pre-mRNA-splicing factor SLU7 | SLU7 | 6 | 10.92% |
| Q9H7E2 | Tudor domain-containing protein 3 | TDRD3 | 6 | 15.36% |
| Q68CZ2 | Tensin-3 | TENS3 | 6 | 8.51% |
| O14617 | AP-3 complex subunit delta-1 | AP3D1 | 5 | 5.38% |
| Q66PJ3 | ADP-ribosylation factor-like protein 6-interacting protein 4 | AR6P4 | 5 | 13.30% |
| P25705 | ATP synthase subunit alpha, mitochondrial | ATPA | 5 | 10.67% |
| Q86UU0 | B-cell CLL/lymphoma 9-like protein | BCL9L | 5 | 6.60% |
| Q9Y224 | UPF0568 protein C14orf166 | CN166 | 5 | 33.20% |
| O60716 | Catenin delta-1 | CTND1 | 5 | 5.79% |
| Q15398 | Disks large-associated protein 5 | DLGP5 | 5 | 7.45% |
| P31689 | DnaJ homolog subfamily A member 1 | DNJA1 | 5 | 19.90% |
| Q52LJ0 | Protein FAM98B | FA98B | 5 | 29.70% |
| P51116 | Fragile X mental retardation syndrome-related protein 2 | FXR2 | 5 | 10.25% |
| P49915 | GMP synthase [glutamine-hydrolyzing] | GUAA | 5 | 11.40% |

| | | | | |
|---|---|---|---|---|
| Q13418 | Integrin-linked protein kinase | ILK | 5 | 10.18% |
| P42166 | Lamina-associated polypeptide 2, isoform alpha | LAP2A | 5 | 10.81% |
| O94776 | Metastasis-associated protein MTA2 | MTA2 | 5 | 14.97% |
| Q9P2K5 | Myelin expression factor 2 | MYEF2 | 5 | 17.83% |
| Q9Y314 | Nitric oxide synthase-interacting protein | NOSIP | 5 | 22.26% |
| O00151 | PDZ and LIM domain protein 1 | PDLI1 | 5 | 24.01% |
| Q8IXK0 | Polyhomeotic-like protein 2 | PHC2 | 5 | 3.96% |
| Q8WWY3 | U4/U6 small nuclear ribonucleoprotein Prp31 | PRP31 | 5 | 15.03% |
| P62333 | 26S protease regulatory subunit 10B | PRS10 | 5 | 18.51% |
| Q9P2N5 | RNA-binding protein 27 | RBM27 | 5 | 7.45% |
| P27694 | Replication protein A 70 kDa DNA-binding subunit | RFA1 | 5 | 11.69% |
| Q15287 | RNA-binding protein with serine-rich domain 1 | RNPS1 | 5 | 20.33% |
| Q5VT52 | Regulation of nuclear pre-mRNA domain-containing protein 2 | RPRD2 | 5 | 3.97% |
| Q9NVA2 | Septin-11 | SEP11 | 5 | 14.45% |
| Q9P270 | SLAIN motif-containing protein 2 | SLAI2 | 5 | 11.36% |
| P51532 | Transcription activator BRG1 | SMCA4 | 5 | 3.28% |
| Q2TAY7 | WD40 repeat-containing protein SMU1 | SMU1 | 5 | 9.55% |
| Q8WVK2 | U4/U6.U5 small nuclear ribonucleoprotein 27 kDa protein | SNR27 | 5 | 19.35% |
| O00267 | Transcription elongation factor SPT5 | SPT5H | 5 | 7.64% |
| Q7KZ85 | Transcription elongation factor SPT6 | SPT6H | 5 | 3.19% |
| Q08170 | Serine/arginine-rich splicing factor 4 | SRSF4 | 5 | 11.74% |
| Q9P2J5 | Leucine--tRNA ligase, cytoplasmic | SYLC | 5 | 6.72% |
| P35269 | General transcription factor IIF subunit 1 | T2FA | 5 | 13.54% |
| Q9BUF5 | Tubulin beta-6 chain | TBB6 | 5 | 26.23% |
| P23193 | Transcription elongation factor A protein 1 | TCEA1 | 5 | 21.26% |

| | | | | |
|---|---|---|---|---|
| P55072 | Transitional endoplasmic reticulum ATPase | TERA | 5 | 10.92% |
| P12270 | Nucleoprotein TPR | TPR | 5 | 4.10% |
| Q13595 | Transformer-2 protein homolog alpha | TRA2A | 5 | 14.54% |
| Q6NZY4 | Zinc finger CCHC domain-containing protein 8 | ZCHC8 | 5 | 9.48% |
| Q5VUA4 | Zinc finger protein 318 | ZN318 | 5 | 2.33% |
| P61221 | ATP-binding cassette sub-family E member 1 | ABCE1 | 4 | 11.69% |
| O94929 | Actin-binding LIM protein 3 | ABLM3 | 4 | 8.05% |
| Q12904 | Aminoacyl tRNA synthase complex-interacting multifunctional protein 1 | AIMP1 | 4 | 24.68% |
| Q9NQW6 | Actin-binding protein anillin | ANLN | 4 | 5.87% |
| P12429 | Annexin A3 | ANXA3 | 4 | 15.79% |
| P46100 | Transcriptional regulator ATRX | ATRX | 4 | 2.69% |
| Q14137 | Ribosome biogenesis protein BOP1 | BOP1 | 4 | 8.71% |
| Q13895 | Bystin | BYST | 4 | 17.39% |
| Q9HC52 | Chromobox protein homolog 8 | CBX8 | 4 | 13.37% |
| Q9H6F5 | Coiled-coil domain-containing protein 86 | CCD86 | 4 | 11.39% |
| O00299 | Chloride intracellular channel protein 1 | CLIC1 | 4 | 32.78% |
| P23528 | Cofilin-1 | COF1 | 4 | 37.35% |
| Q10570 | Cleavage and polyadenylation specificity factor subunit 1 | CPSF1 | 4 | 5.13% |
| Q9BQ61 | Uncharacterized protein C19orf43 | CS043 | 4 | 47.16% |
| Q12996 | Cleavage stimulation factor subunit 3 | CSTF3 | 4 | 8.79% |
| Q9NVP1 | ATP-dependent RNA helicase DDX18 | DDX18 | 4 | 7.76% |
| Q6XZF7 | Dynamin-binding protein | DNMBP | 4 | 3.87% |
| Q14697 | Neutral alpha-glucosidase AB | GANAB | 4 | 4.77% |
| Q96RT7 | Gamma-tubulin complex component 6 | GCP6 | 4 | 4.29% |
| P51610 | Host cell factor 1 | HCFC1 | 4 | 2.70% |
| P05198 | Eukaryotic translation initiation factor 2 subunit 1 | IF2A | 4 | 9.52% |

| | | | | |
|---|---|---|---|---|
| Q15056 | Eukaryotic translation initiation factor 4H | IF4H | 4 | 31.85% |
| P13645 | Keratin, type I cytoskeletal 10 | K1C10 | 4 | 16.95% |
| P33176 | Kinesin-1 heavy chain | KINH | 4 | 5.71% |
| Q96RT1 | Protein LAP2 | LAP2 | 4 | 6.16% |
| Q9Y4Z0 | U6 snRNA-associated Sm-like protein LSm4 | LSM4 | 4 | 28.06% |
| Q9NR56 | Muscleblind-like protein 1 | MBNL1 | 4 | 17.01% |
| O43148 | mRNA cap guanine-N7 methyltransferase | MCES | 4 | 11.97% |
| Q16539 | Mitogen-activated protein kinase 14 | MK14 | 4 | 19.44% |
| O60524 | Nuclear export mediator factor NEMF | NEMF | 4 | 3.62% |
| Q9BZE4 | Nucleolar GTP-binding protein 1 | NOG1 | 4 | 8.99% |
| P55209 | Nucleosome assembly protein 1-like 1 | NP1L1 | 4 | 17.14% |
| O96028 | Histone-lysine N-methyltransferase NSD2 | NSD2 | 4 | 4.84% |
| Q08J23 | tRNA (cytosine(34)-C(5))-methyltransferase | NSUN2 | 4 | 7.17% |
| Q8TEW0 | Partitioning defective 3 homolog | PARD3 | 4 | 2.88% |
| Q6L8Q7 | 2',5'-phosphodiesterase 12 | PDE12 | 4 | 8.70% |
| Q9Y237 | Peptidyl-prolyl cis-trans isomerase NIMA-interacting 4 | PIN4 | 4 | 51.91% |
| Q12972 | Nuclear inhibitor of protein phosphatase 1 | PP1R8 | 4 | 19.66% |
| Q9Y3C6 | Peptidyl-prolyl cis-trans isomerase-like 1 | PPIL1 | 4 | 21.69% |
| Q06124 | Tyrosine-protein phosphatase non-receptor type 11 | PTN11 | 4 | 7.37% |
| Q9UHX1 | Poly(U)-binding-splicing factor PUF60 | PUF60 | 4 | 13.42% |
| Q9BTD8 | RNA-binding protein 42 | RBM42 | 4 | 14.17% |
| Q92900 | Regulator of nonsense transcripts 1 | RENT1 | 4 | 4.25% |
| P24928 | DNA-directed RNA polymerase II subunit RPB1 | RPB1 | 4 | 2.64% |
| Q9Y3B9 | RRP15-like protein | RRP15 | 4 | 15.25% |
| P60866 | 40S ribosomal protein S20 | RS20 | 4 | 29.41% |
| P62304 | Small nuclear ribonucleoprotein | RUXE | 4 | 32.61% |

| | | | | |
|---|---|---|---|---|
| | E | | | |
| Q9Y3A5 | Ribosome maturation protein SBDS | SBDS | 4 | 26.00% |
| P53992 | Protein transport protein Sec24C 3 | SC24C | 4 | 7.04% |
| Q14493 | Histone RNA hairpin-binding protein 1 | SLBP | 4 | 14.81% |
| Q53GS9 | U4/U6.U5 tri-snRNP-associated protein 2 | SNUT2 | 4 | 6.02% |
| Q8WXA9 | Splicing regulatory glutamine/lysine-rich protein 1 | SREK1 | 4 | 15.16% |
| Q08945 | FACT complex subunit SSRP1 | SSRP1 | 4 | 8.89% |
| Q9Y2Z0 | Suppressor of G2 allele of SKP1 homolog 3 | SUGT1 | 4 | 17.81% |
| O43776 | Asparagine--tRNA ligase, cytoplasmic 1 | SYNC | 4 | 11.50% |
| P23258 | Tubulin gamma-1 chain | TBG1 | 4 | 12.20% |
| Q5JTD0 | Tight junction-associated protein 1 | TJAP1 | 4 | 12.75% |
| Q15025 | TNFAIP3-interacting protein 1 | TNIP1 | 4 | 6.13% |
| Q9NXH9 | tRNA (guanine(26)-N(2))-dimethyltransferase 1 | TRM1 | 4 | 5.61% |
| Q9NW82 | WD repeat-containing protein 70 | WDR70 | 4 | 11.93% |
| Q96NC0 | Zinc finger matrin-type protein 2 | ZMAT2 | 4 | 17.59% |
| Q96ME7 | Zinc finger protein 512 | ZN512 | 4 | 6.00% |
| Q15942 | Zyxin | ZYX | 4 | 10.49% |
| P31946 | 14-3-3 protein beta/alpha | 1433B | 3 | 18.70% |
| P01892 | HLA class I histocompatibility antigen, A-2 alpha chain | 1A02 | 3 | 7.40% |
| Q9UKV8 | Protein argonaute-2 | AGO2 | 3 | 4.66% |
| O00170 | AH receptor-interacting protein | AIP | 3 | 9.39% |
| P20073 | Annexin A7 | ANXA7 | 3 | 6.35% |
| P63010 | AP-2 complex subunit beta | AP2B1 | 3 | 3.20% |
| Q9NP61 | ADP-ribosylation factor GTPase-activating protein 3 | ARFG3 | 3 | 9.50% |
| Q8NFD5 | AT-rich interactive domain-containing protein 1B | ARI1B | 3 | 2.10% |
| P06576 | ATP synthase subunit beta, mitochondrial | ATPB | 3 | 9.83% |
| Q9UIG0 | Tyrosine-protein kinase BAZ1B | BAZ1B | 3 | 1.96% |
| Q9NPI1 | Bromodomain-containing protein | BRD7 | 3 | 6.91% |

| | | | | |
|---|---|---|---|---|
| | 7 | | | |
| Q05682 | Caldesmon | CALD1 | 3 | 6.31% |
| P47756 | F-actin-capping protein subunit beta | CAPZB | 3 | 10.47% |
| Q16543 | Hsp90 co-chaperone Cdc37 | CDC37 | 3 | 13.49% |
| Q03188 | Centromere protein C | CENPC | 3 | 3.08% |
| Q9P2D1 | Chromodomain-helicase-DNA-binding protein 7 | CHD7 | 3 | 1.74% |
| Q9Y3Y2 | Chromatin target of PRMT1 protein | CHTOP | 3 | 12.10% |
| P09496 | Clathrin light chain A | CLCA | 3 | 12.10% |
| Q8N1G2 | Cap-specific mRNA (nucleoside-2'-O-)-methyltransferase 1 | CMTR1 | 3 | 5.27% |
| Q15003 | Condensin complex subunit 2 | CND2 | 3 | 2.97% |
| P35606 | Coatomer subunit beta' | COPB2 | 3 | 4.30% |
| Q9ULV4 | Coronin-1C | COR1C | 3 | 13.29% |
| O75131 | Copine-3 | CPNE3 | 3 | 6.33% |
| P21291 | Cysteine and glycine-rich protein 1 | CSRP1 | 3 | 21.76% |
| Q9P013 | Spliceosome-associated protein CWC15 homolog | CWC15 | 3 | 13.97% |
| Q9UJU6 | Drebrin-like protein | DBNL | 3 | 13.02% |
| Q16531 | DNA damage-binding protein 1 | DDB1 | 3 | 3.68% |
| Q96GQ7 | Probable ATP-dependent RNA helicase DDX27 | DDX27 | 3 | 3.64% |
| Q9UJV9 | Probable ATP-dependent RNA helicase DDX41 | DDX41 | 3 | 6.59% |
| Q9Y2R4 | Probable ATP-dependent RNA helicase DDX52 | DDX52 | 3 | 10.18% |
| Q9NY93 | Probable ATP-dependent RNA helicase DDX56 | DDX56 | 3 | 8.96% |
| Q13217 | DnaJ homolog subfamily C member 3 | DNJC3 | 3 | 6.94% |
| P49005 | DNA polymerase delta subunit 2 | DPOD2 | 3 | 7.89% |
| Q9Y295 | Developmentally-regulated GTP-binding protein 1 | DRG1 | 3 | 11.17% |
| P50570 | Dynamin-2 | DYN2 | 3 | 6.78% |
| Q9H4M9 | EH domain-containing protein 1 | EHD1 | 3 | 8.61% |
| Q99613 | Eukaryotic translation initiation factor 3 subunit C | EIF3C | 3 | 4.82% |
| Q9BSJ8 | Extended synaptotagmin-1 | ESYT1 | 3 | 5.25% |

| | | | | |
|---|---|---|---|---|
| Q9NQT5 | Exosome complex component RRP40 | EXOS3 | 3 | 26.91% |
| Q01469 | Fatty acid-binding protein, epidermal | FABP5 | 3 | 16.30% |
| Q00688 | Peptidyl-prolyl cis-trans isomerase FKBP3 | FKBP3 | 3 | 20.54% |
| Q92616 | Translational activator GCN1 | GCN1L | 3 | 1.46% |
| Q9UKJ3 | G patch domain-containing protein 8 | GPTC8 | 3 | 3.06% |
| P11166 | Solute carrier family 2, facilitated glucose transporter member 1 | GTR1 | 3 | 5.89% |
| P16104 | Histone H2AX | H2AX | 3 | 38.46% |
| P0C0S5 | Histone H2A.Z | H2AZ | 3 | 33.59% |
| Q96A08 | Histone H2B type 1-A | H2B1A | 3 | 51.18% |
| Q75N03 | E3 ubiquitin-protein ligase Hakai | HAKAI | 3 | 10.59% |
| Q92769 | Histone deacetylase 2 | HDAC2 | 3 | 7.38% |
| P51858 | Hepatoma-derived growth factor | HDGF | 3 | 16.25% |
| Q9H910 | Hematological and neurological expressed 1-like protein | HN1L | 3 | 36.32% |
| Q8WVV9 | Heterogeneous nuclear ribonucleoprotein L-like | HNRLL | 3 | 10.15% |
| P34931 | Heat shock 70 kDa protein 1-like | HS71L | 3 | 18.88% |
| O43719 | HIV Tat-specific factor 1 | HTSF1 | 3 | 4.37% |
| Q8WUF5 | RelA-associated inhibitor | IASPP | 3 | 5.31% |
| P01857 | Ig gamma-1 chain C region | IGHG1 | 3 | 7.27% |
| Q15652 | Probable JmjC domain-containing histone demethylation protein 2C | JHD2C | 3 | 1.69% |
| P30085 | UMP-CMP kinase | KCY | 3 | 16.33% |
| Q9BW19 | Kinesin-like protein KIFC1 | KIFC1 | 3 | 8.62% |
| P07195 | L-lactate dehydrogenase B chain | LDHB | 3 | 12.57% |
| Q96BZ8 | Leukocyte receptor cluster member 1 | LENG1 | 3 | 15.53% |
| Q9UPQ0 | LIM and calponin homology domains-containing protein 1 | LIMC1 | 3 | 3.79% |
| Q3MHD2 | Protein LSM12 homolog | LSM12 | 3 | 13.33% |
| Q96GA3 | Protein LTV1 homolog | LTV1 | 3 | 6.74% |
| P61326 | Protein mago nashi homolog | MGN | 3 | 30.14% |
| Q96T58 | Msx2-interacting protein | MINT | 3 | 0.79% |

| | | | | |
|---|---|---|---|---|
| Q9UBU8 | Mortality factor 4-like protein 1 | MO4L1 | 3 | 12.98% |
| Q14764 | Major vault protein | MVP | 3 | 8.40% |
| O75380 | NADH dehydrogenase [ubiquinone] iron-sulfur protein 6, mitochondrial | NDUS6 | 3 | 44.35% |
| P55769 | NHP2-like protein 1 | NH2L1 | 3 | 20.31% |
| P30419 | Glycylpeptide N-tetradecanoyltransferase 1 | NMT1 | 3 | 6.65% |
| Q9BSC4 | Nucleolar protein 10 | NOL10 | 3 | 4.65% |
| Q9Y2X3 | Nucleolar protein 58 | NOP58 | 3 | 8.88% |
| Q8TAT6 | Nuclear protein localization protein 4 homolog | NPL4 | 3 | 6.58% |
| Q8WUM0 | Nuclear pore complex protein Nup133 | NU133 | 3 | 3.03% |
| Q8N1F7 | Nuclear pore complex protein Nup93 | NUP93 | 3 | 6.11% |
| Q02218 | 2-oxoglutarate dehydrogenase, mitochondrial | ODO1 | 3 | 6.16% |
| P36957 | Dihydrolipoyllysine-residue succinyltransferase component of 2-oxoglutarate dehydrogenase complex, mitochondrial | ODO2 | 3 | 6.62% |
| P54886 | Delta-1-pyrroline-5-carboxylate synthase | P5CS | 3 | 4.91% |
| Q86YP4 | Transcriptional repressor p66-alpha | P66A | 3 | 8.37% |
| Q86U42 | Polyadenylate-binding protein 2 | PABP2 | 3 | 8.82% |
| O95340 | Bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthase 2 | PAPS2 | 3 | 7.65% |
| Q16513 | Serine/threonine-protein kinase N2 | PKN2 | 3 | 3.96% |
| O43447 | Peptidyl-prolyl cis-trans isomerase H | PPIH | 3 | 19.77% |
| Q9H2H8 | Peptidyl-prolyl cis-trans isomerase-like 3 | PPIL3 | 3 | 34.78% |
| P30041 | Peroxiredoxin-6 | PRDX6 | 3 | 23.66% |
| P25787 | Proteasome subunit alpha type-2 | PSA2 | 3 | 18.38% |
| P60900 | Proteasome subunit alpha type-6 | PSA6 | 3 | 19.92% |
| O75475 | PC4 and SFRS1-interacting | PSIP1 | 3 | 5.28% |

| | | | | |
|---|---|---|---|---|
| | protein | | | |
| Q13200 | 26S proteasome non-ATPase regulatory subunit 2 | PSMD2 | 3 | 5.40% |
| P30520 | Adenylosuccinate synthetase isozyme 2 | PURA2 | 3 | 11.18% |
| Q9H0U4 | Ras-related protein Rab-1B | RAB1B | 3 | 15.42% |
| Q3YEC7 | Rab-like protein 6 | RABL6 | 3 | 7.41% |
| Q96S59 | Ran-binding protein 9 | RANB9 | 3 | 7.82% |
| Q8NDT2 | Putative RNA-binding protein 15B | RB15B | 3 | 7.53% |
| Q5T8P6 | RNA-binding protein 26 | RBM26 | 3 | 3.38% |
| P29558 | RNA-binding motif, single-stranded-interacting protein 1 | RBMS1 | 3 | 15.27% |
| Q5TC82 | Roquin-1 | RC3H1 | 3 | 2.65% |
| P18754 | Regulator of chromosome condensation | RCC1 | 3 | 15.68% |
| Q92785 | Zinc finger protein ubi-d4 | REQU | 3 | 17.14% |
| O76021 | Ribosomal L1 domain-containing protein 1 | RL1D1 | 3 | 4.49% |
| P05388 | 60S acidic ribosomal protein P0 | RLA0 | 3 | 12.93% |
| Q9BYD1 | 39S ribosomal protein L13, mitochondrial | RM13 | 3 | 28.09% |
| Q6P1L8 | 39S ribosomal protein L14, mitochondrial | RM14 | 3 | 30.34% |
| Q9P0M9 | 39S ribosomal protein L27, mitochondrial | RM27 | 3 | 30.41% |
| Q9BRJ2 | 39S ribosomal protein L45, mitochondrial | RM45 | 3 | 11.76% |
| Q8TA86 | Retinitis pigmentosa 9 protein | RP9 | 3 | 11.76% |
| Q3B726 | DNA-directed RNA polymerase I subunit RPA43 | RPA43 | 3 | 8.58% |
| P30876 | DNA-directed RNA polymerase II subunit RPB2 | RPB2 | 3 | 4.68% |
| P62244 | 40S ribosomal protein S15a | RS15A | 3 | 27.69% |
| P82921 | 28S ribosomal protein S21, mitochondrial | RT21 | 3 | 37.93% |
| Q92541 | RNA polymerase-associated protein RTF1 homolog | RTF1 | 3 | 4.08% |
| P26447 | Protein S100-A4 | S10A4 | 3 | 30.69% |
| Q15020 | Squamous cell carcinoma antigen recognized by T-cells 3 | SART3 | 3 | 4.67% |

| | | | | |
|---|---|---|---|---|
| Q12872 | Splicing factor, suppressor of white-apricot homolog | SFSWA | 3 | 4.52% |
| O95347 | Structural maintenance of chromosomes protein 2 | SMC2 | 3 | 4.34% |
| P53814 | Smoothelin | SMTN | 3 | 5.23% |
| P09012 | U1 small nuclear ribonucleoprotein A | SNRPA | 3 | 15.25% |
| O75940 | Survival of motor neuron-related-splicing factor 30 | SPF30 | 3 | 9.24% |
| P52788 | Spermine synthase | SPSY | 3 | 13.93% |
| Q68D10 | Protein SPT2 homolog | SPT2 | 3 | 6.42% |
| P37108 | Signal recognition particle 14 kDa protein | SRP14 | 3 | 37.50% |
| Q8IX01 | SURP and G-patch domain-containing protein 2 | SUGP2 | 3 | 4.81% |
| P14868 | Aspartate--tRNA ligase, cytoplasmic | SYDC | 3 | 11.58% |
| P26639 | Threonine--tRNA ligase, cytoplasmic | SYTC | 3 | 6.50% |
| Q6P1X5 | Transcription initiation factor TFIID subunit 2 | TAF2 | 3 | 3.42% |
| P37802 | Transgelin-2 | TAGL2 | 3 | 11.06% |
| O75764 | Transcription elongation factor A protein 3 | TCEA3 | 3 | 12.64% |
| Q5QJE6 | Deoxynucleotidyltransferase terminal-interacting protein 2 | TDIF2 | 3 | 7.41% |
| Q9BWD1 | Acetyl-CoA acetyltransferase, cytosolic | THIC | 3 | 16.12% |
| Q86V81 | THO complex subunit 4 | THOC4 | 3 | 14.01% |
| Q92973 | Transportin-1 | TNPO1 | 3 | 3.56% |
| Q9HCJ0 | Trinucleotide repeat-containing gene 6C protein | TNR6C | 3 | 2.72% |
| Q5JTV8 | Torsin-1A-interacting protein 1 | TOIP1 | 3 | 9.09% |
| Q96PN7 | Transcriptional-regulating factor 1 | TREF1 | 3 | 3.75% |
| Q01081 | Splicing factor U2AF 35 kDa subunit | U2AF1 | 3 | 13.33% |
| O43818 | U3 small nucleolar RNA-interacting protein 2 | U3IP2 | 3 | 7.58% |
| P61081 | NEDD8-conjugating enzyme Ubc12 | UBC12 | 3 | 18.58% |
| P17480 | Nucleolar transcription factor 1 | UBF1 | 3 | 6.68% |

| | | | | |
|---|---|---|---|---|
| Q92890 | Ubiquitin fusion degradation protein 1 homolog | UFD1 | 3 | 12.38% |
| Q96T88 | E3 ubiquitin-protein ligase UHRF1 | UHRF1 | 3 | 5.17% |
| O60504 | Vinexin | VINEX | 3 | 7.60% |
| P55060 | Exportin-2 | XPO2 | 3 | 4.12% |
| Q7Z2W4 | Zinc finger CCCH-type antiviral protein 1 | ZCCHV | 3 | 6.21% |
| Q8N5A5 | Zinc finger CCCH-type with G patch domain-containing protein | ZGPAT | 3 | 7.91% |
| Q9UPN3 | Microtubule-actin cross-linking factor 1, isoforms 1/2/3/5 | MACF1 | 3 | 0.00% |
| P42696 | RNA-binding protein 34 | RBM34 | 2 | 9.53% |
| Q9Y285 | Phenylalanine-tRNA ligase alpha subunit | SYFA | 2 | 10.04% |

# APPENDIX B: PEAKS DATA OF BCL11A RIME EXPERIMENT

| UniProt Accession Number | Protein Name | # of unique peptides | Sequence coverage (%Cov:) |
|---|---|---|---|
| P49454 | CENPF | 17 | 8.60 |
| O75717 | WDHD1 | 16 | 20.90 |
| Q9HCK8 | CHD8 | 16 | 10.46 |
| P49419 | AL7A1 | 14 | 27.27 |
| Q9H0A0 | NAT10 | 13 | 12.98 |
| Q6PJT7 | ZC3HE | 12 | 22.55 |
| Q9BY77 | PDIP3 | 12 | 38.24 |
| P55884 | EIF3B | 11 | 14.13 |
| Q8IXT5 | RB12B | 11 | 9.99 |
| Q01804 | OTUD4 | 10 | 9.69 |
| Q2TBE0 | C19L2 | 10 | 11.86 |
| Q9BRD0 | BUD13 | 10 | 22.62 |
| O60716 | CTND1 | 9 | 14.46 |
| P53999 | TCP4 | 9 | 53.54 |
| Q03135 | CAV1 | 9 | 51.69 |
| Q9H8M2 | BRD9 | 9 | 23.12 |
| O00299 | CLIC1 | 8 | 46.06 |
| O94906 | PRP6 | 8 | 13.71 |
| O95391 | SLU7 | 8 | 18.26 |
| P21127 | CD11B | 8 | 9.56 |
| P40121 | CAPG | 8 | 32.18 |
| P42166 | LAP2A | 8 | 23.63 |
| Q01130 | SRSF2 | 8 | 38.91 |
| Q07020 | RL18 | 8 | 41.49 |
| Q15029 | U5S1 | 8 | 14.20 |
| Q7KZ85 | SPT6H | 8 | 6.26 |
| Q9UHB7 | AFF4 | 8 | 12.21 |
| Q9UQ88 | CD11A | 8 | 9.71 |
| O60828 | PQBP1 | 7 | 42.64 |
| O75534 | CSDE1 | 7 | 10.65 |
| P12270 | TPR | 7 | 4.87 |
| P23193 | TCEA1 | 7 | 24.58 |
| P49411 | EFTU | 7 | 20.35 |
| P55081 | MFAP1 | 7 | 26.42 |
| Q13595 | TRA2A | 7 | 26.60 |
| Q14839 | CHD4 | 7 | 6.22 |
| Q6NZY4 | ZCHC8 | 7 | 16.69 |
| Q8WVK2 | SNR27 | 7 | 27.10 |

| | | | |
|---|---|---|---|
| Q9H7E2 | TDRD3 | 7 | 18.13 |
| Q9NTZ6 | RBM12 | 7 | 7.62 |
| P12429 | ANXA3 | 6 | 23.22 |
| P33176 | KINH | 6 | 8.10 |
| P35269 | T2FA | 6 | 17.21 |
| P37802 | TAGL2 | 6 | 23.62 |
| P49915 | GUAA | 6 | 13.71 |
| Q12904 | AIMP1 | 6 | 31.09 |
| Q15287 | RNPS1 | 6 | 22.62 |
| Q15398 | DLGP5 | 6 | 9.69 |
| Q15434 | RBMS2 | 6 | 33.17 |
| Q17RY0 | CPEB4 | 6 | 11.80 |
| Q2TAY7 | SMU1 | 6 | 11.89 |
| Q5JTD0 | TJAP1 | 6 | 18.67 |
| Q66PJ3 | AR6P4 | 6 | 17.58 |
| Q6NZI2 | PTRF | 6 | 12.82 |
| Q8IXK0 | PHC2 | 6 | 6.99 |
| Q8NFD5 | ARI1B | 6 | 5.55 |
| Q9BQ61 | CS043 | 6 | 48.30 |
| Q9NVA2 | Sep-11 | 6 | 17.48 |
| Q9NVP1 | DDX18 | 6 | 11.49 |
| Q9NZ63 | CI078 | 6 | 29.07 |
| Q9P2N5 | RBM27 | 6 | 9.06 |
| Q9UHX1 | PUF60 | 6 | 23.08 |
| Q9Y2Z0 | SUGT1 | 6 | 21.10 |
| O00267 | SPT5H | 5 | 8.28 |
| O43148 | MCES | 5 | 15.76 |
| O43660 | PLRG1 | 5 | 10.70 |
| O43719 | HTSF1 | 5 | 6.62 |
| O75531 | BANF1 | 5 | 40.45 |
| P00367 | DHE3 | 5 | 14.70 |
| P20073 | ANXA7 | 5 | 6.76 |
| P24928 | RPB1 | 5 | 2.69 |
| P25705 | ATPA | 5 | 10.67 |
| P30520 | PURA2 | 5 | 13.16 |
| P30876 | RPB2 | 5 | 7.58 |
| P31689 | DNJA1 | 5 | 19.90 |
| P35659 | DEK | 5 | 13.60 |
| P46100 | ATRX | 5 | 3.41 |
| P49591 | SYSC | 5 | 13.42 |
| P51532 | SMCA4 | 5 | 3.28 |
| P51610 | HCFC1 | 5 | 3.54 |
| P55209 | NP1L1 | 5 | 25.58 |
| P60228 | EIF3E | 5 | 11.69 |
| P60842 | IF4A1 | 5 | 38.67 |
| P60866 | RS20 | 5 | 31.93 |

| | | | |
|---|---|---|---|
| P62333 | PRS10 | 5 | 18.51 |
| P62750 | RL23A | 5 | 37.18 |
| Q01469 | FABP5 | 5 | 24.44 |
| Q06124 | PTN11 | 5 | 10.55 |
| Q08170 | SRSF4 | 5 | 11.74 |
| Q12872 | SFSWA | 5 | 6.10 |
| Q13418 | ILK | 5 | 10.18 |
| Q15397 | K0020 | 5 | 10.49 |
| Q5VT52 | RPRD2 | 5 | 3.63 |
| Q68D10 | SPT2 | 5 | 10.07 |
| Q8N1F7 | NUP93 | 5 | 11.11 |
| Q8NDT2 | RB15B | 5 | 11.01 |
| Q8TAT6 | NPL4 | 5 | 15.63 |
| Q8WWY3 | PRP31 | 5 | 15.03 |
| Q8WXA9 | SREK1 | 5 | 15.16 |
| Q92785 | REQU | 5 | 19.69 |
| Q96ME7 | ZN512 | 5 | 9.35 |
| Q96T58 | MINT | 5 | 1.72 |
| Q96T88 | UHRF1 | 5 | 9.21 |
| Q9BTD8 | RBM42 | 5 | 20.83 |
| Q9BWD1 | THIC | 5 | 27.96 |
| Q9BZE4 | NOG1 | 5 | 10.88 |
| Q9NR56 | MBNL1 | 5 | 19.85 |
| Q9NW82 | WDR70 | 5 | 14.83 |
| Q9P2K5 | MYEF2 | 5 | 14.00 |
| Q9Y3A5 | SBDS | 5 | 28.00 |
| Q9Y3C1 | NOP16 | 5 | 41.01 |
| Q9Y3Y2 | CHTOP | 5 | 18.15 |
| O15145 | ARPC3 | 4 | 17.42 |
| O15294 | OGT1 | 4 | 5.07 |
| O43776 | SYNC | 4 | 11.50 |
| O60504 | VINEX | 4 | 9.24 |
| O60524 | NEMF | 4 | 3.62 |
| O75380 | NDUS6 | 4 | 47.58 |
| O75396 | SC22B | 4 | 25.12 |
| O75676 | KS6A4 | 4 | 9.33 |
| O75821 | EIF3G | 4 | 24.06 |
| O76021 | RL1D1 | 4 | 10.20 |
| O94929 | ABLM3 | 4 | 8.05 |
| O95232 | LC7L3 | 4 | 11.34 |
| O95573 | ACSL3 | 4 | 7.50 |
| O95793 | STAU1 | 4 | 9.01 |
| O96028 | NSD2 | 4 | 4.84 |
| P01857 | IGHG1 | 4 | 9.39 |
| P06576 | ATPB | 4 | 13.42 |
| P0C0S5 | H2AZ | 4 | 33.59 |

| | | | |
|---|---|---|---|
| P13645 | K1C10 | 4 | 12.33 |
| P14625 | ENPL | 4 | 7.85 |
| P14678 | RSMB | 4 | 18.33 |
| P14868 | SYDC | 4 | 13.77 |
| P17480 | UBF1 | 4 | 9.03 |
| P20290 | BTF3 | 4 | 37.38 |
| P23258 | TBG1 | 4 | 12.20 |
| P27824 | CALX | 4 | 16.22 |
| P28482 | MK01 | 4 | 9.44 |
| P29558 | RBMS1 | 4 | 28.08 |
| P30041 | PRDX6 | 4 | 33.04 |
| P35244 | RFA3 | 4 | 23.14 |
| P37108 | SRP14 | 4 | 47.79 |
| P42285 | SK2L2 | 4 | 6.81 |
| P52788 | SPSY | 4 | 20.22 |
| P53814 | SMTN | 4 | 6.76 |
| P53992 | SC24C | 4 | 7.04 |
| P62136 | PP1A | 4 | 17.27 |
| P62304 | RUXE | 4 | 32.61 |
| P63162 | RSMN | 4 | 18.33 |
| P68032 | ACTC | 4 | 43.77 |
| P68133 | ACTS | 4 | 43.77 |
| P78344 | IF4G2 | 4 | 9.37 |
| Q00688 | FKBP3 | 4 | 20.54 |
| Q01081 | U2AF1 | 4 | 13.75 |
| Q02218 | ODO1 | 4 | 6.94 |
| Q03188 | CENPC | 4 | 4.56 |
| Q08J23 | NSUN2 | 4 | 7.17 |
| Q12996 | CSTF3 | 4 | 8.79 |
| Q13895 | BYST | 4 | 17.39 |
| Q14137 | BOP1 | 4 | 10.99 |
| Q14493 | SLBP | 4 | 14.81 |
| Q14697 | GANAB | 4 | 4.77 |
| Q14738 | 2A5D | 4 | 7.64 |
| Q16543 | CDC37 | 4 | 21.43 |
| Q16576 | RBBP7 | 4 | 25.88 |
| Q52LJ0 | FA98B | 4 | 22.12 |
| Q53GS9 | SNUT2 | 4 | 6.02 |
| Q5JTV8 | TOIP1 | 4 | 9.09 |
| Q5QJE6 | TDIF2 | 4 | 9.13 |
| Q5T8P6 | RBM26 | 4 | 4.77 |
| Q5TC82 | RC3H1 | 4 | 4.24 |
| Q5VTR2 | BRE1A | 4 | 5.64 |
| Q6L8Q7 | PDE12 | 4 | 8.70 |
| Q6P1L8 | RM14 | 4 | 30.34 |
| Q6P1X5 | TAF2 | 4 | 3.42 |

| | | | |
|---|---|---|---|
| Q6XZF7 | DNMBP | 4 | 2.73 |
| Q70EL1 | UBP54 | 4 | 3.86 |
| Q71RC2 | LARP4 | 4 | 9.67 |
| Q71UI9 | H2AV | 4 | 33.59 |
| Q75N03 | HAKAI | 4 | 15.27 |
| Q86V81 | THOC4 | 4 | 19.07 |
| Q8IWC1 | MA7D3 | 4 | 5.71 |
| Q8IYB3 | SRRM1 | 4 | 7.85 |
| Q8N1G2 | CMTR1 | 4 | 6.47 |
| Q8NBJ5 | GT251 | 4 | 7.23 |
| Q8TEW0 | PARD3 | 4 | 2.88 |
| Q8WUM0 | NU133 | 4 | 3.81 |
| Q8WYA6 | CTBL1 | 4 | 9.06 |
| Q92616 | GCN1L | 4 | 1.91 |
| Q92890 | UFD1 | 4 | 16.61 |
| Q92973 | TNPO1 | 4 | 7.02 |
| Q96DI7 | SNR40 | 4 | 13.73 |
| Q96GQ7 | DDX27 | 4 | 3.64 |
| Q96RT7 | GCP6 | 4 | 3.52 |
| Q99426 | TBCB | 4 | 29.10 |
| Q99439 | CNN2 | 4 | 15.21 |
| Q99613 | EIF3C | 4 | 6.46 |
| Q9BTA9 | WAC | 4 | 15.30 |
| Q9BW19 | KIFC1 | 4 | 9.21 |
| Q9H2H8 | PPIL3 | 4 | 38.51 |
| Q9H4M9 | EHD1 | 4 | 14.61 |
| Q9H6F5 | CCD86 | 4 | 11.39 |
| Q9HC52 | CBX8 | 4 | 13.37 |
| Q9HCJ0 | TNR6C | 4 | 3.20 |
| Q9NRH3 | TBG2 | 4 | 12.20 |
| Q9NSY1 | BMP2K | 4 | 9.13 |
| Q9NY93 | DDX56 | 4 | 12.07 |
| Q9NYV4 | CDK12 | 4 | 3.83 |
| Q9P270 | SLAI2 | 4 | 8.78 |
| Q9P2J5 | SYLC | 4 | 5.02 |
| Q9P2R6 | RERE | 4 | 6.00 |
| Q9UBB9 | TFP11 | 4 | 5.97 |
| Q9UIG0 | BAZ1B | 4 | 3.17 |
| Q9UJV9 | DDX41 | 4 | 8.20 |
| Q9UNQ2 | DIM1 | 4 | 7.67 |
| Q9Y2R4 | DDX52 | 4 | 7.18 |
| Q9Y2X3 | NOP58 | 4 | 13.23 |
| Q9Y3B9 | RRP15 | 4 | 15.25 |
| Q9Y4Z0 | LSM4 | 4 | 28.06 |
| A0JLT2 | MED19 | 3 | 20.90 |
| A6NHR9 | SMHD1 | 3 | 1.80 |

| | | | |
|---|---|---|---|
| O00170 | AIP | 3 | 9.39 |
| O14744 | ANM5 | 3 | 7.22 |
| O43447 | PPIH | 3 | 19.77 |
| O60231 | DHX16 | 3 | 4.42 |
| O75131 | CPNE3 | 3 | 6.33 |
| O75940 | SPF30 | 3 | 9.24 |
| O94804 | STK10 | 3 | 5.48 |
| O94875 | SRBS2 | 3 | 3.55 |
| O95340 | PAPS2 | 3 | 7.65 |
| O95425 | SVIL | 3 | 2.85 |
| O95602 | RPA1 | 3 | 3.20 |
| P01892 | 1A02 | 3 | 13.97 |
| P05388 | RLA0 | 3 | 12.93 |
| P07195 | LDHB | 3 | 12.57 |
| P07237 | PDIA1 | 3 | 10.24 |
| P07339 | CATD | 3 | 11.17 |
| P08758 | ANXA5 | 3 | 12.81 |
| P09012 | SNRPA | 3 | 15.25 |
| P09496 | CLCA | 3 | 20.56 |
| P11166 | GTR1 | 3 | 5.89 |
| P11413 | G6PD | 3 | 6.60 |
| P17980 | PRS6A | 3 | 10.93 |
| P18206 | VINC | 3 | 5.03 |
| P18754 | RCC1 | 3 | 15.68 |
| P19387 | RPB3 | 3 | 19.64 |
| P23919 | KTHY | 3 | 9.43 |
| P25440 | BRD2 | 3 | 5.87 |
| P25788 | PSA3 | 3 | 20.78 |
| P26639 | SYTC | 3 | 6.50 |
| P27361 | MK03 | 3 | 15.57 |
| P30419 | NMT1 | 3 | 6.65 |
| P31930 | QCR1 | 3 | 9.58 |
| P35606 | COPB2 | 3 | 4.30 |
| P35908 | K22E | 3 | 16.28 |
| P36957 | ODO2 | 3 | 7.28 |
| P37837 | TALDO | 3 | 10.68 |
| P38606 | VATA | 3 | 9.08 |
| P43686 | PRS6B | 3 | 8.13 |
| P45974 | UBP5 | 3 | 8.74 |
| P46937 | YAP1 | 3 | 13.10 |
| P49005 | DPOD2 | 3 | 7.89 |
| P49773 | HINT1 | 3 | 34.92 |
| P50570 | DYN2 | 3 | 6.78 |
| P51116 | FXR2 | 3 | 10.70 |
| P51858 | HDGF | 3 | 16.25 |
| P52565 | GDIR1 | 3 | 33.33 |

| | | | |
|---|---|---|---|
| P54886 | P5CS | 3 | 4.91 |
| P55010 | IF5 | 3 | 8.12 |
| P56545 | CTBP2 | 3 | 13.71 |
| P61024 | CKS1 | 3 | 41.77 |
| P61163 | ACTZ | 3 | 16.22 |
| P61221 | ABCE1 | 3 | 8.35 |
| P62310 | LSM3 | 3 | 34.31 |
| P62714 | PP2AB | 3 | 13.92 |
| P63010 | AP2B1 | 3 | 3.20 |
| P67775 | PP2AA | 3 | 13.92 |
| P67809 | YBOX1 | 3 | 15.74 |
| P78406 | RAE1L | 3 | 8.15 |
| P82921 | RT21 | 3 | 37.93 |
| Q02241 | KIF23 | 3 | 7.19 |
| Q04917 | 1433F | 3 | 21.54 |
| Q05397 | FAK1 | 3 | 4.66 |
| Q05682 | CALD1 | 3 | 6.31 |
| Q12788 | TBL3 | 3 | 2.23 |
| Q12972 | PP1R8 | 3 | 14.81 |
| Q13200 | PSMD2 | 3 | 8.04 |
| Q13206 | DDX10 | 3 | 6.86 |
| Q13217 | DNJC3 | 3 | 8.13 |
| Q14444 | CAPR1 | 3 | 4.37 |
| Q14764 | MVP | 3 | 8.40 |
| Q15003 | CND2 | 3 | 2.97 |
| Q15014 | MO4L2 | 3 | 11.46 |
| Q15020 | SART3 | 3 | 4.67 |
| Q15046 | SYK | 3 | 6.70 |
| Q15427 | SF3B4 | 3 | 10.61 |
| Q15642 | CIP4 | 3 | 8.82 |
| Q15652 | JHD2C | 3 | 1.69 |
| Q16531 | DDB1 | 3 | 3.68 |
| Q16539 | MK14 | 3 | 19.44 |
| Q16637 | SMN | 3 | 13.61 |
| Q3B726 | RPA43 | 3 | 8.58 |
| Q3MHD2 | LSM12 | 3 | 13.33 |
| Q53EP0 | FND3B | 3 | 2.82 |
| Q5HYJ3 | FA76B | 3 | 13.57 |
| Q5SW79 | CE170 | 3 | 2.40 |
| Q5TGY3 | AHDC1 | 3 | 4.30 |
| Q5VWG9 | TAF3 | 3 | 4.52 |
| Q5VYS8 | TUT7 | 3 | 2.21 |
| Q6IBS0 | TWF2 | 3 | 12.61 |
| Q6RFH5 | WDR74 | 3 | 10.65 |
| Q6XE24 | RBMS3 | 3 | 15.33 |
| Q86VM9 | ZCH18 | 3 | 5.14 |

| | | | |
|---|---|---|---|
| Q8IWZ8 | SUGP1 | 3 | 4.96 |
| Q8IY67 | RAVR1 | 3 | 13.04 |
| Q8IZL8 | PELP1 | 3 | 5.40 |
| Q8IZP0 | ABI1 | 3 | 9.45 |
| Q8N5A5 | ZGPAT | 3 | 7.91 |
| Q8N6H7 | ARFG2 | 3 | 9.40 |
| Q8TA86 | RP9 | 3 | 11.76 |
| Q8TAP9 | MPLKI | 3 | 32.40 |
| Q8TAQ2 | SMRC2 | 3 | 4.94 |
| Q8TDW0 | LRC8C | 3 | 5.98 |
| Q8WUD4 | CCD12 | 3 | 21.69 |
| Q8WUF5 | IASPP | 3 | 5.31 |
| Q8WXX5 | DNJC9 | 3 | 15.77 |
| Q92522 | H1X | 3 | 19.72 |
| Q92541 | RTF1 | 3 | 4.08 |
| Q92879 | CELF1 | 3 | 12.96 |
| Q92928 | RAB1C | 3 | 26.87 |
| Q96KP4 | CNDP2 | 3 | 6.95 |
| Q96PN7 | TREF1 | 3 | 3.75 |
| Q96PU5 | NED4L | 3 | 3.90 |
| Q96Q83 | ALKB3 | 3 | 11.54 |
| Q96RT1 | LAP2 | 3 | 4.11 |
| Q96S59 | RANB9 | 3 | 7.82 |
| Q96ST2 | IWS1 | 3 | 5.25 |
| Q99816 | TS101 | 3 | 7.95 |
| Q99848 | EBP2 | 3 | 12.42 |
| Q9BPX3 | CND3 | 3 | 1.97 |
| Q9BQ67 | GRWD1 | 3 | 10.31 |
| Q9BRJ2 | RM45 | 3 | 11.76 |
| Q9BSC4 | NOL10 | 3 | 4.65 |
| Q9BWU0 | NADAP | 3 | 9.05 |
| Q9BY44 | EIF2A | 3 | 8.55 |
| Q9BYD1 | RM13 | 3 | 28.09 |
| Q9BYG3 | MK67I | 3 | 22.18 |
| Q9H0U4 | RAB1B | 3 | 26.87 |
| Q9H2U1 | DHX36 | 3 | 2.28 |
| Q9HCG8 | CWC22 | 3 | 4.63 |
| Q9NP61 | ARFG3 | 3 | 10.08 |
| Q9NPI1 | BRD7 | 3 | 6.91 |
| Q9NRX4 | PHP14 | 3 | 56.00 |
| Q9NXV6 | CARF | 3 | 6.72 |
| Q9P0L0 | VAPA | 3 | 20.48 |
| Q9P0M9 | RM27 | 3 | 30.41 |
| Q9P2D1 | CHD7 | 3 | 1.74 |
| Q9P2I0 | CPSF2 | 3 | 3.32 |
| Q9UBU8 | MO4L1 | 3 | 12.98 |

| | | | |
|---|---|---|---|
| Q9UHI6 | DDX20 | 3 | 9.83 |
| Q9UKV8 | AGO2 | 3 | 7.80 |
| Q9UMY1 | NOL7 | 3 | 15.18 |
| Q9UPW0 | FOXJ3 | 3 | 5.47 |
| Q9Y295 | DRG1 | 3 | 11.17 |
| Q9Y3B4 | SF3B6 | 3 | 28.80 |
| Q9Y3D0 | MIP18 | 3 | 31.29 |
| Q9Y3F4 | STRAP | 3 | 13.71 |

# APPENDIX C: UNIQUE NUCLEAR FACTORS PULLED-DOWN WITH DCAS9 AT THE FOXC1 GENE PROMOTER THROUGH RIME

| Protein name | Protein Description |
|---|---|
| BRX1 | Ribosome biogenesis protein BRX1 homolog |
| DNJC3 | DnaJ homolog subfamily C member 3 |
| RL1D1 | Ribosomal L1 domain-containing protein 1 |
| SPB1 | pre-rRNA processing protein FTSJ3 |
| TAF2 | Transcription initiation factor TFIID subunit 2 |
| H2B1K | Histone H2B type 1-K |
| CSTF2 | Cleavage stimulation factor subunit 2 |
| SPIN1 | Spindlin-1 |
| VRK1 | Serine/threonine-protein kinase VRK1 |
| MMTA2 | Multiple myeloma tumour-associated protein 2 |
| DIM1 | Probable dimethyl adenosine transferase |
| ZN512 | Zinc finger protein 512 |
| CKS1 | Cyclin-dependent kinases regulatory subunit 1 |
| RUXE | Small nuclear ribonucleoprotein E |
| RS27 | 40S ribosomal protein S27 |
| H2AZ | Histone H2A.Z |
| EIF1B | Eukaryotic translation initiation factor 1b |
| PCNP | PEST proteolytic signal-containing nuclear protein |
| RS15 | 40S ribosomal protein S15 |
| MAT1 | CDK-activating kinase assembly factor MAT1 |
| RL35A | 60S ribosomal protein L35a |
| PQBP1 | Polyglutamine-binding protein 1 |
| SET | Protein SET |
| RL24 | 60S ribosomal protein L24 |
| RBMS2 | RNA-binding motif, single-stranded-interacting protein 2 |
| MK67I | MKI67 FHA domain-interacting nucleolar phosphoprotein |
| RS23 | 40S ribosomal protein S23 |
| RP9 | Retinitis pigmentosa 9 protein |
| PR38A | Pre-mRNA-splicing factor 38A |
| SARNP | SAP domain-containing ribonucleoprotein |
| RTCA | RNA 3'-terminal phosphate cyclase |
| NC2A | Dr1-associated corepressor |
| DNJA2 | DnaJ homolog subfamily A member 2 |
| PDIP3 | Polymerase delta-interacting protein 3 |
| RBM42 | RNA-binding protein 42 |
| NLE1 | Notchless protein homolog 1 |
| PPIE | Peptidyl-prolyl cis-trans isomerase E |
| PSIP1 | PC4 and SFRS1-interacting protein |

| | |
|---|---|
| WAC | WW domain-containing adapter protein with coiled-coil |
| TOE1 | Target of EGR1 protein 1 |
| SMYD3 | Histone-lysine N-methyltransferase |
| CLP1 | Polyribonucleotide 5'-hydroxyl-kinase |
| ZMY11 | Zinc finger MYND domain-containing protein 11 |
| MIC60 | MICOS complex subunit MIC60 |
| PBIP1 | Pre-B-cell leukaemia transcription factor-interacting protein |
| HP1B3 | Heterochromatin protein 1-binding protein 3 |
| UBF1 | Nucleolar transcription factor 1 |
| IASPP | RelA-associated inhibitor |
| TEAD1 | Transcriptional enhancer factor TEF-1 |
| NOLC1 | Nucleolar and coiled-body phosphoprotein 1 |
| RTF1 | RNA polymerase-associated protein RTF1 |
| IKKA | Inhibitor of nuclear factor kappa-B kinase subunit alpha |
| CEBPZ | CCAAT/enhancer-binding protein zeta |
| SIK3 | Serine/threonine-protein kinase |
| RERE | Arginine-glutamic acid dipeptide repeats protein |
| SFSWA | Splicing factor, suppressor of white-apricot |
| PDS5A | Sister chromatid cohesion protein |
| SMRC2 | SWI/SNF complex subunit |
| YES | Tyrosine-protein kinase Yes |
| ZMIZ2 | Zinc finger MIZ domain-containing protein 2 |

# APPENDIX D: COMMON RIME PROTEINS AMONG FOXC1, NFBI & NFE2L3, AFTER FILTERING FOR PSM≥ 1, SUBTRACTION OF IGG, AND REDUNDANCY AMONG REPLICATES (N=3)

| UniProt Accession Number | Protein Name | Protein Description |
|---|---|---|
| Q6PJT7 | ZC3HE | Zinc finger CCCH domain-containing protein 14 |
| Q96QD9 | UIF | UAP56-interacting factor |
| O75152 | ZC11A | Zinc finger CCCH domain-containing protein 11A |
| Q53F19 | CQ085 | Uncharacterized protein C17orf85 |
| P46013 | KI67 | Antigen KI-67 |
| Q9Y5S9 | RBM8A | RNA-binding protein 8A |
| P61326 | MGN | Protein mago nashi homolog |
| Q9Y2W1 | TR150 | Thyroid hormone receptor-associated protein 3 |
| Q8N9M1 | CS047 | Uncharacterized protein C19orf47 |
| Q9BZE4 | NOG1 | Nucleolar GTP-binding protein 1 |
| Q9NYF8 | BCLF1 | Bcl-2-associated transcription factor 1 |
| Q9BZZ5 | API5 | Apoptosis inhibitor 5 |
| Q8TDD1 | DDX54 | ATP-dependent RNA helicase DDX54 |
| P27694 | RFA1 | Replication protein A 70 kDa DNA-binding subunit |
| Q9H307 | PININ | Pinin |
| Q6P6C2 | ALKB5 | RNA demethylase ALKBH5 |
| Q9ULW0 | TPX2 | Targeting protein for Xklp2 |
| Q96MU7 | YTDC1 | YTH domain-containing protein 1 |
| Q14320 | FA50A | Protein FAM50A |
| Q8WWY3 | PRP31 | U4/U6 small nuclear ribonucleoprotein Prp31 |
| Q5T8P6 | RBM26 | RNA-binding protein 26 |
| Q96EV2 | RBM33 | RNA-binding protein 33 |
| P46783 | RS10 | 40S ribosomal protein S10 |
| Q9GZR7 | DDX24 | ATP-dependent RNA helicase DDX24 |
| Q06830 | PRDX1 | Peroxiredoxin-1 |
| P14678 | RSMB | Small nuclear ribonucleoprotein-associated proteins B and B' |
| Q86U42 | PABP2 | Polyadenylate-binding protein 2 |
| Q9UKV3 | ACINU | Apoptotic chromatin condensation inducer in the nucleus |
| O43395 | PRPF3 | U4/U6 small nuclear ribonucleoprotein Prp3 |
| Q9NQ29 | LUC7L | Putative RNA-binding protein Luc7-like 1 |
| O00422 | SAP18 | Histone deacetylase complex subunit SAP18 |
| Q14011 | CIRBP | Cold-inducible RNA-binding protein |

| | | |
|---|---|---|
| Q9Y5J1 | UTP18 | U3 small nucleolar RNA-associated protein 18 homolog |
| P62318 | SMD3 | Small nuclear ribonucleoprotein Sm D3 |
| Q9Y2X3 | NOP58 | Nucleolar protein 58 |
| Q96KR1 | ZFR | Zinc finger RNA-binding protein |
| Q9NRL2 | BAZ1A | Bromodomain adjacent to zinc finger domain protein 1A |
| Q13185 | CBX3 | Chromobox protein homolog 3 |

# APPENDIX E: LIST OF PROTEINS WITH STATISTICAL SIGNIFICANCE

| Accession | Protein names | Crapome | N° reps | P value | P Value adjusted | Relevant genes |
|---|---|---|---|---|---|---|
| Q86V81 | THOC4 | 43.06569343 | 18 | 0 | 0 | FOXC1, NFE2L3, NFIB |
| Q9Y3Y2 | CHTOP | 12.40875912 | 12 | 0 | 0 | FOXC1, NFE2L3, NFIB |
| O00148 | DX39A | 33.57664234 | 12 | 0 | 0 | FOXC1, NFE2L3, NFIB |
| Q13838 | DX39B | 33.57664234 | 14 | 0 | 0 | FOXC1, NFE2L3, NFIB |
| P17844 | DDX5 | 60.0973236 | 18 | 0 | 0 | FOXC1, NFE2L3, NFIB |
| Q08211 | DHX9 | 49.8783455 | 18 | 0 | 0 | FOXC1, NFE2L3, NFIB |
| P60842 | IF4A1 | 46.22871046 | 14 | 0 | 0 | FOXC1, NFE2L3, NFIB |
| P38919 | IF4A3 | 35.76642336 | 16 | 0 | 0 | FOXC1, NFE2L3, NFIB |
| P07910 | HNRPC | 43.79562044 | 18 | 0 | 0 | FOXC1, NFE2L3, NFIB |
| P31943 | HNRH1 | 63.99026764 | 18 | 0 | 0 | FOXC1, NFE2L3, NFIB |
| Q9BY77 | PDIP3 | 19.22141119 | 11 | 0 | 0 | FOXC1, NFE2L3, NFIB |
| P08865 | RSSA | 38.92944039 | 16 | 0 | 0 | FOXC1, NFE2L3, NFIB |
| P82979 | SARNP | 18.97810219 | 15 | 0 | 0 | FOXC1, NFE2L3, NFIB |
| Q01130 | SRSF2 | 32.84671533 | 15 | 0 | 0 | FOXC1, NFE2L3, NFIB |
| P84103 | SRSF3 | 47.93187348 | 18 | 0 | 0 | FOXC1, NFE2L3, NFIB |

| | | | | | | |
|---|---|---|---|---|---|---|
| P62995 | TRA2B | 20.68126521 | 16 | 0 | 0 | FOXC1, NFE2L3, NFIB |
| Q92841 | DDX17 | 57.90754258 | 16 | 0 | 1.00E-04 | FOXC1, NFE2L3, NFIB |
| P84090 | ERH | 36.49635036 | 11 | 0 | 1.00E-04 | FOXC1, NFE2L3, NFIB |
| P09651 | ROA1 | 65.20681265 | 17 | 0 | 1.00E-04 | FOXC1, NFE2L3, NFIB |
| P61978 | HNRPK | 70.0729927 | 18 | 0 | 1.00E-04 | FOXC1, NFE2L3, NFIB |
| P23246 | SFPQ | 48.90510949 | 18 | 0 | 1.00E-04 | FOXC1, NFE2L3, NFIB |
| Q13247 | SRSF6 | 33.81995134 | 18 | 0 | 1.00E-04 | FOXC1, NFE2L3, NFIB |
| Q92499 | DDX1 | 30.4136253 | 16 | 0 | 2.00E-04 | FOXC1, NFE2L3, NFIB |
| Q5VYK3 | ECM29 | 1 | 15 | 0 | 2.00E-04 | FOXC1, NFE2L3, NFIB |
| Q6ZNL6 | FGD5 | 1 | 17 | 0 | 2.00E-04 | FOXC1, NFE2L3, NFIB |
| P51991 | ROA3 | 44.52554745 | 16 | 0 | 2.00E-04 | FOXC1, NFE2L3, NFIB |
| P14866 | HNRPL | 46.22871046 | 17 | 0 | 2.00E-04 | FOXC1, NFE2L3, NFIB |
| P19338 | NUCL | 63.50364964 | 17 | 0 | 2.00E-04 | FOXC1, NFE2L3, NFIB |
| P62847 | RS24 | 43.55231144 | 18 | 0 | 2.00E-04 | FOXC1, NFE2L3, NFIB |
| P46781 | RS9 | 41.36253041 | 17 | 0 | 2.00E-04 | FOXC1, NFE2L3, NFIB |
| Q99729 | ROAA | 50.1216545 | 13 | 0 | 3.00E-04 | FOXC1, NFE2L3, NFIB |
| Q00839 | HNRPU | 68.12652068 | 18 | 0 | 3.00E-04 | FOXC1, NFE2L3, NFIB |
| P62917 | RL8 | 38.19951338 | 18 | 0 | 3.00E-04 | FOXC1, NFE2L3, NFIB |

| | | | | | | |
|---|---|---|---|---|---|---|
| Q9Y3I0 | RTCB | 30.4136253 | 12 | 0 | 3.00E-04 | FOXC1, NFE2L3, NFIB |
| P52272 | HNRPM | 53.52798054 | 18 | 0 | 4.00E-04 | FOXC1, NFE2L3, NFIB |
| P26599 | PTBP1 | 41.84914842 | 16 | 0 | 4.00E-04 | FOXC1, NFE2L3, NFIB |
| Q6NZI2 | PTRF | 4.379562044 | 16 | 0 | 4.00E-04 | FOXC1, NFE2L3, NFIB |
| O43390 | HNRPR | 45.74209246 | 16 | 0 | 5.00E-04 | FOXC1, NFE2L3, NFIB |
| Q6ZUA9 | MROH5 | 1 | 12 | 0 | 5.00E-04 | FOXC1, NFE2L3, NFIB |
| Q15717 | ELAV1 | 24.57420925 | 15 | 1.00E-04 | 6.00E-04 | FOXC1, NFE2L3, NFIB |
| Q15424 | SAFB1 | 18.73479319 | 16 | 1.00E-04 | 6.00E-04 | FOXC1, NFE2L3, NFIB |
| P13611 | CSPG2 | 0.486618005 | 17 | 1.00E-04 | 6.00E-04 | FOXC1, NFE2L3, NFIB |
| O00571 | DDX3X | 51.58150852 | 16 | 1.00E-04 | 7.00E-04 | FOXC1, NFE2L3, NFIB |
| O43143 | DHX15 | 40.87591241 | 11 | 1.00E-04 | 7.00E-04 | FOXC1, NFE2L3, NFIB |
| P52597 | HNRPF | 59.12408759 | 13 | 1.00E-04 | 7.00E-04 | FOXC1, NFE2L3, NFIB |
| Q92945 | FUBP2 | 32.36009732 | 18 | 1.00E-04 | 7.00E-04 | FOXC1, NFE2L3, NFIB |
| P06748 | NPM | 61.31386861 | 18 | 1.00E-04 | 7.00E-04 | FOXC1, NFE2L3, NFIB |
| P61313 | RL15 | 36.25304136 | 18 | 1.00E-04 | 7.00E-04 | FOXC1, NFE2L3, NFIB |
| P39023 | RL3 | 39.6593674 | 18 | 1.00E-04 | 7.00E-04 | FOXC1, NFE2L3, NFIB |
| P04406 | G3P | 60.3406326 | 16 | 1.00E-04 | 8.00E-04 | FOXC1, NFE2L3, NFIB |
| P62829 | RL23 | 59.12408759 | 12 | 1.00E-04 | 8.00E-04 | FOXC1, NFE2L3, NFIB |

| | | | | | | |
|---|---|---|---|---|---|---|
| P61254 | RL26 | 34.30656934 | 18 | 1.00E-04 | 8.00E-04 | FOXC1, NFE2L3, NFIB |
| P62266 | RS23 | 38.92944039 | 18 | 1.00E-04 | 8.00E-04 | FOXC1, NFE2L3, NFIB |
| Q15233 | NONO | 53.52798054 | 15 | 1.00E-04 | 9.00E-04 | FOXC1, NFE2L3, NFIB |
| P62906 | RL10A | 28.46715328 | 17 | 1.00E-04 | 9.00E-04 | FOXC1, NFE2L3, NFIB |
| P26373 | RL13 | 53.04136253 | 18 | 1.00E-04 | 9.00E-04 | FOXC1, NFE2L3, NFIB |
| P08670 | VIME | 62.53041363 | 18 | 1.00E-04 | 9.00E-04 | FOXC1, NFE2L3, NFIB |
| Q13151 | ROA0 | 39.9026764 | 13 | 1.00E-04 | 0.001 | FOXC1, NFE2L3, NFIB |
| Q8WXF1 | PSPC1 | 19.9513382 | 16 | 1.00E-04 | 0.001 | FOXC1, NFE2L3, NFIB |
| Q15287 | RNPS1 | 26.52068127 | 15 | 1.00E-04 | 0.001 | FOXC1, NFE2L3, NFIB |
| P23396 | RS3 | 64.72019465 | 18 | 1.00E-04 | 0.001 | FOXC1, NFE2L3, NFIB |
| Q9BVP2 | GNL3 | 19.46472019 | 13 | 1.00E-04 | 0.0011 | FOXC1, NFE2L3, NFIB |
| P60866 | RS20 | 43.55231144 | 14 | 1.00E-04 | 0.0011 | FOXC1, NFE2L3, NFIB |
| P62701 | RS4X | 54.74452555 | 18 | 2.00E-04 | 0.0011 | FOXC1, NFE2L3, NFIB |
| Q15637 | SF01 | 19.7080292 | 11 | 1.00E-04 | 0.0011 | FOXC1, NFE2L3, NFIB |
| Q14103 | HNRPD | 52.55474453 | 16 | 2.00E-04 | 0.0012 | FOXC1, NFE2L3, NFIB |
| P38159 | RBMX | 41.11922141 | 15 | 2.00E-04 | 0.0012 | FOXC1, NFE2L3, NFIB |
| P84098 | RL19 | 51.09489051 | 18 | 2.00E-04 | 0.0012 | FOXC1, NFE2L3, NFIB |
| P35268 | RL22 | 46.71532847 | 18 | 2.00E-04 | 0.0012 | FOXC1, NFE2L3, NFIB |

| | | | | | | |
|---|---|---|---|---|---|---|
| P22626 | ROA2 | 63.74695864 | 18 | 2.00E-04 | 0.0013 | FOXC1, NFE2L3, NFIB |
| Q15365 | PCBP1 | 47.20194647 | 15 | 2.00E-04 | 0.0013 | FOXC1, NFE2L3, NFIB |
| P22087 | FBRL | 39.17274939 | 15 | 2.00E-04 | 0.0014 | FOXC1, NFE2L3, NFIB |
| P63244 | GBLP | 32.84671533 | 11 | 2.00E-04 | 0.0014 | FOXC1, NFE2L3, NFIB |
| Q07666 | KHDR1 | 33.57664234 | 17 | 2.00E-04 | 0.0014 | FOXC1, NFE2L3, NFIB |
| P62753 | RS6 | 50.60827251 | 18 | 2.00E-04 | 0.0014 | FOXC1, NFE2L3, NFIB |
| P46778 | RL21 | 32.84671533 | 12 | 2.00E-04 | 0.0015 | FOXC1, NFE2L3, NFIB |
| P62424 | RL7A | 47.93187348 | 18 | 2.00E-04 | 0.0015 | FOXC1, NFE2L3, NFIB |
| Q99848 | EBP2 | 12.89537713 | 11 | 3.00E-04 | 0.0017 | FOXC1, NFE2L3, NFIB |
| P84243 | H33 | 33.81995134 | 18 | 3.00E-04 | 0.0018 | FOXC1, NFE2L3, NFIB |
| P62826 | RAN | 37.46958637 | 13 | 3.00E-04 | 0.0018 | FOXC1, NFE2L3, NFIB |
| P63173 | RL38 | 33.81995134 | 15 | 3.00E-04 | 0.0018 | FOXC1, NFE2L3, NFIB |
| P62241 | RS8 | 55.47445255 | 17 | 3.00E-04 | 0.0018 | FOXC1, NFE2L3, NFIB |
| P23528 | COF1 | 47.68856448 | 16 | 3.00E-04 | 0.0019 | FOXC1, NFE2L3, NFIB |
| P62308 | RUXG | 11.92214112 | 12 | 3.00E-04 | 0.0019 | FOXC1, NFE2L3, NFIB |
| A8MWD9 | RUXGL | 1 | 12 | 3.00E-04 | 0.0019 | FOXC1, NFE2L3, NFIB |
| Q12906 | ILF3 | 38.44282238 | 15 | 4.00E-04 | 0.0021 | FOXC1, NFE2L3, NFIB |
| Q04118 | PRB3 | 1 | 5 | 4.00E-04 | 0.0021 | NFE2L3 |
| P49207 | RL34 | 20.4379562 | 18 | 4.00E- | 0.0021 | FOXC1, |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | 04 | | NFE2L3, NFIB |
| P18124 | RL7 | 42.57907543 | 18 | 4.00E-04 | 0.0022 | FOXC1, NFE2L3, NFIB |
| P98179 | RBM3 | 21.41119221 | 11 | 4.00E-04 | 0.0024 | FOXC1, NFE2L3, NFIB |
| Q8IVT2 | MISP | 4.866180049 | 13 | 5.00E-04 | 0.0025 | FOXC1, NFE2L3, NFIB |
| P49756 | RBM25 | 18.73479319 | 15 | 5.00E-04 | 0.0025 | FOXC1, NFE2L3, NFIB |
| Q07020 | RL18 | 46.47201946 | 18 | 5.00E-04 | 0.0025 | FOXC1, NFE2L3, NFIB |
| Q8IY81 | SPB1 | 15.32846715 | 10 | 5.00E-04 | 0.0026 | FOXC1, NFE2L3, NFIB |
| P40429 | RL13A | 35.03649635 | 14 | 5.00E-04 | 0.0027 | FOXC1, NFE2L3, NFIB |
| P62899 | RL31 | 41.36253041 | 16 | 6.00E-04 | 0.0029 | FOXC1, NFE2L3, NFIB |
| Q16629 | SRSF7 | 46.47201946 | 12 | 6.00E-04 | 0.003 | FOXC1, NFE2L3, NFIB |
| Q8TDN6 | BRX1 | 16.30170316 | 12 | 6.00E-04 | 0.0031 | FOXC1, NFE2L3, NFIB |
| P09429 | HMGB1 | 25.79075426 | 17 | 6.00E-04 | 0.0031 | FOXC1, NFE2L3, NFIB |
| P43243 | MATR3 | 41.60583942 | 17 | 6.00E-04 | 0.0031 | FOXC1, NFE2L3, NFIB |
| P61353 | RL27 | 38.68613139 | 17 | 7.00E-04 | 0.0033 | FOXC1, NFE2L3, NFIB |
| Q99829 | CPNE1 | 7.299270073 | 12 | 8.00E-04 | 0.0036 | FOXC1, NFE2L3, NFIB |
| O60869 | EDF1 | 19.46472019 | 10 | 8.00E-04 | 0.0037 | FOXC1, NFE2L3, NFIB |
| P46939 | UTRO | 6.569343066 | 13 | 9.00E-04 | 0.0039 | FOXC1, NFE2L3, NFIB |
| P63241 | IF5A1 | 30.4136253 | 16 | 9.00E-04 | 0.0042 | FOXC1, NFE2L3, NFIB |
| P36578 | RL4 | 47.93187348 | 18 | 9.00E- | 0.0042 | FOXC1, |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | 04 | | NFE2L3, NFIB |
| Q02878 | RL6 | 41.84914842 | 18 | 9.00E-04 | 0.0042 | FOXC1, NFE2L3, NFIB |
| P08621 | RU17 | 26.27737226 | 15 | 0.001 | 0.0043 | FOXC1, NFE2L3, NFIB |
| O60506 | HNRPQ | 47.93187348 | 12 | 0.001 | 0.0043 | FOXC1, NFE2L3, NFIB |
| P83731 | RL24 | 47.44525547 | 15 | 0.0011 | 0.0046 | FOXC1, NFE2L3, NFIB |
| P42766 | RL35 | 37.95620438 | 18 | 0.0011 | 0.0046 | FOXC1, NFE2L3, NFIB |
| P46777 | RL5 | 37.95620438 | 13 | 0.0011 | 0.0046 | FOXC1, NFE2L3, NFIB |
| Q02543 | RL18A | 27.49391727 | 13 | 0.0012 | 0.0048 | FOXC1, NFE2L3, NFIB |
| Q13263 | TIF1B | 43.06569343 | 10 | 0.0012 | 0.0049 | FOXC1, NFE2L3, NFIB |
| P21333 | FLNA | 54.98783455 | 15 | 0.0012 | 0.005 | FOXC1, NFE2L3, NFIB |
| P05787 | K2C8 | 57.66423358 | 18 | 0.0012 | 0.005 | FOXC1, NFE2L3, NFIB |
| Q9BQ39 | DDX50 | 28.95377129 | 11 | 0.0014 | 0.0058 | FOXC1, NFE2L3, NFIB |
| P02545 | LMNA | 22.38442822 | 18 | 0.0014 | 0.0058 | FOXC1, NFE2L3, NFIB |
| Q14980 | NUMA1 | 18.97810219 | 10 | 0.0015 | 0.0058 | FOXC1, NFE2L3, NFIB |
| P62913 | RL11 | 50.85158151 | 13 | 0.0015 | 0.0058 | FOXC1, NFE2L3, NFIB |
| P68104 | EF1A1 | 85.15815085 | 18 | 0.0016 | 0.0061 | FOXC1, NFE2L3, NFIB |
| Q5VTE0 | EF1A3 | 1 | 18 | 0.0016 | 0.0061 | FOXC1, NFE2L3, NFIB |
| P12268 | IMDH2 | 20.1946472 | 11 | 0.0016 | 0.0061 | FOXC1, NFE2L3, NFIB |
| P53999 | TCP4 | 38.92944039 | 10 | 0.0016 | 0.0063 | FOXC1, |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | NFE2L3, NFIB |
| P07477 | TRY1 | 15.81508516 | 18 | 0.0017 | 0.0064 | FOXC1, NFE2L3, NFIB |
| P46776 | RL27A | 48.17518248 | 16 | 0.0017 | 0.0064 | FOXC1, NFE2L3, NFIB |
| O75367 | H2AY | 16.54501217 | 13 | 0.0017 | 0.0065 | FOXC1, NFE2L3, NFIB |
| P62750 | RL23A | 54.50121655 | 18 | 0.0018 | 0.0065 | FOXC1, NFE2L3, NFIB |
| P61513 | RL37A | 23.84428224 | 12 | 0.0018 | 0.0065 | FOXC1, NFE2L3, NFIB |
| P15880 | RS2 | 51.09489051 | 13 | 0.0018 | 0.0066 | FOXC1, NFE2L3, NFIB |
| P11142 | HSP7C | 96.35036496 | 18 | 0.0019 | 0.0067 | FOXC1, NFE2L3, NFIB |
| P68431 | H31 | 34.79318735 | 16 | 0.0019 | 0.0068 | FOXC1, NFE2L3, NFIB |
| Q71DI3 | H32 | 34.54987835 | 16 | 0.0019 | 0.0068 | FOXC1, NFE2L3, NFIB |
| P27635 | RL10 | 35.52311436 | 15 | 0.0019 | 0.0068 | FOXC1, NFE2L3, NFIB |
| Q8NC51 | PAIRB | 42.57907543 | 17 | 0.0019 | 0.0068 | FOXC1, NFE2L3, NFIB |
| P06733 | ENOA | 54.25790754 | 17 | 0.002 | 0.0071 | FOXC1, NFE2L3, NFIB |
| Q07955 | SRSF1 | 32.60340633 | 14 | 0.002 | 0.0071 | FOXC1, NFE2L3, NFIB |
| P62937 | PPIA | 44.76885645 | 14 | 0.0021 | 0.0072 | FOXC1, NFE2L3, NFIB |
| P63220 | RS21 | 27.98053528 | 10 | 0.0021 | 0.0072 | FOXC1, NFE2L3, NFIB |
| P11940 | PABP1 | 41.84914842 | 13 | 0.0024 | 0.0081 | FOXC1, NFE2L3, NFIB |
| P61247 | RS3A | 49.39172749 | 16 | 0.0024 | 0.0082 | FOXC1, NFE2L3, NFIB |
| P60709 | ACTB | 88.32116788 | 18 | 0.0028 | 0.0094 | FOXC1, |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | NFE2L3, NFIB |
| P62979 | RS27A | 60.3406326 | 18 | 0.0028 | 0.0094 | FOXC1, NFE2L3, NFIB |
| Q9BYG3 | MK67I | 1 | 10 | 0.0029 | 0.0096 | FOXC1, NFE2L3, NFIB |
| O75607 | NPM3 | 11.92214112 | 14 | 0.003 | 0.0098 | FOXC1, NFE2L3, NFIB |
| O76021 | RL1D1 | 23.84428224 | 11 | 0.0032 | 0.0105 | FOXC1, NFE2L3, NFIB |
| P46779 | RL28 | 26.52068127 | 14 | 0.0033 | 0.0107 | FOXC1, NFE2L3, NFIB |
| Q9UKM9 | RALY | 10.70559611 | 12 | 0.0034 | 0.0108 | FOXC1, NFE2L3, NFIB |
| Q14498 | RBM39 | 39.6593674 | 10 | 0.0034 | 0.0109 | FOXC1, NFE2L3, NFIB |
| P62851 | RS25 | 47.68856448 | 17 | 0.0034 | 0.0109 | FOXC1, NFE2L3, NFIB |
| Q01081 | U2AF1 | 30.90024331 | 16 | 0.0035 | 0.0109 | FOXC1, NFE2L3, NFIB |
| P08729 | K2C7 | 37.71289538 | 18 | 0.0037 | 0.0115 | FOXC1, NFE2L3, NFIB |
| Q15366 | PCBP2 | 45.98540146 | 11 | 0.0037 | 0.0115 | FOXC1, NFE2L3, NFIB |
| P30050 | RL12 | 48.90510949 | 15 | 0.0039 | 0.0119 | FOXC1, NFE2L3, NFIB |
| P61927 | RL37 | 4.379562044 | 15 | 0.004 | 0.012 | FOXC1, NFE2L3, NFIB |
| P62891 | RL39 | 14.84184915 | 14 | 0.004 | 0.012 | FOXC1, NFE2L3, NFIB |
| Q59GN2 | R39L5 | 1 | 14 | 0.004 | 0.012 | FOXC1, NFE2L3, NFIB |
| P62244 | RS15A | 43.30900243 | 14 | 0.004 | 0.012 | FOXC1, NFE2L3, NFIB |
| P32969 | RL9 | 40.63260341 | 14 | 0.0041 | 0.0124 | FOXC1, NFE2L3, NFIB |
| P11387 | TOP1 | 26.76399027 | 10 | 0.0043 | 0.0127 | FOXC1, |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | NFE2L3, NFIB |
| Q9HB71 | CYBP | 16.30170316 | 10 | 0.0045 | 0.0133 | FOXC1, NFE2L3, NFIB |
| P62269 | RS18 | 55.71776156 | 15 | 0.0047 | 0.0139 | FOXC1, NFE2L3, NFIB |
| Q14137 | BOP1 | 12.16545012 | 10 | 0.0048 | 0.014 | FOXC1, NFE2L3, NFIB |
| P10412 | H14 | 72.74939173 | 18 | 0.0049 | 0.014 | FOXC1, NFE2L3, NFIB |
| P05783 | K1C18 | 57.90754258 | 18 | 0.0049 | 0.014 | FOXC1, NFE2L3, NFIB |
| Q6VAB6 | KSR2 | 1 | 13 | 0.0049 | 0.014 | FOXC1, NFE2L3, NFIB |
| P62280 | RS11 | 36.73965937 | 14 | 0.0049 | 0.014 | FOXC1, NFE2L3, NFIB |
| P63261 | ACTG | 88.32116788 | 12 | 0.0053 | 0.0148 | FOXC1, NFE2L3, NFIB |
| P12956 | XRCC6 | 42.82238443 | 10 | 0.0052 | 0.0148 | FOXC1, NFE2L3, NFIB |
| Q7KZF4 | SND1 | 28.71046229 | 16 | 0.0053 | 0.0149 | FOXC1, NFE2L3, NFIB |
| P16403 | H12 | 73.23600973 | 18 | 0.0055 | 0.0153 | FOXC1, NFE2L3, NFIB |
| P08708 | RS17 | 43.06569343 | 14 | 0.0061 | 0.0168 | FOXC1, NFE2L3, NFIB |
| P0CW22 | RS17L | 43.06569343 | 14 | 0.0061 | 0.0168 | FOXC1, NFE2L3, NFIB |
| Q969Q0 | RL36L | 18.24817518 | 17 | 0.0064 | 0.0176 | FOXC1, NFE2L3, NFIB |
| Q16630 | CPSF6 | 33.33333333 | 9 | 0.0067 | 0.0181 | FOXC1, NFE2L3, NFIB |
| P26641 | EF1G | 46.95863747 | 16 | 0.0067 | 0.0181 | FOXC1, NFE2L3, NFIB |
| P08779 | K1C16 | 66.66666667 | 6 | 0.0076 | 0.0205 | NFE2L3, NFIB |
| P04908 | H2A1B | 65.93673966 | 16 | 0.0084 | 0.0219 | FOXC1, NFE2L3, |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | | NFIB |
| Q93077 | H2A1C | 65.93673966 | 16 | 0.0084 | 0.0219 | FOXC1, NFE2L3, NFIB |
| Q7L7L0 | H2A3 | 65.93673966 | 16 | 0.0084 | 0.0219 | FOXC1, NFE2L3, NFIB |
| P09874 | PARP1 | 44.28223844 | 18 | 0.0083 | 0.0219 | FOXC1, NFE2L3, NFIB |
| Q7Z6E9 | RBBP6 | 9.489051095 | 11 | 0.0084 | 0.0219 | FOXC1, NFE2L3, NFIB |
| P62888 | RL30 | 27.00729927 | 12 | 0.0091 | 0.0237 | FOXC1, NFE2L3, NFIB |
| P08238 | HS90B | 67.39659367 | 18 | 0.0093 | 0.0239 | FOXC1, NFE2L3, NFIB |
| P11532 | DMD | 1.216545012 | 10 | 0.0097 | 0.025 | FOXC1, NFE2L3, NFIB |
| P62249 | RS16 | 51.33819951 | 13 | 0.0103 | 0.0262 | FOXC1, NFE2L3, NFIB |
| P35580 | MYH10 | 40.3892944 | 9 | 0.0105 | 0.0265 | FOXC1, NFE2L3, NFIB |
| Q5M775 | CYTSB | 4.622871046 | 10 | 0.0107 | 0.0272 | FOXC1, NFE2L3, NFIB |
| P62277 | RS13 | 38.44282238 | 14 | 0.0112 | 0.0281 | FOXC1, NFE2L3, NFIB |
| Q71U36 | TBA1A | 94.64720195 | 13 | 0.0112 | 0.0281 | FOXC1, NFE2L3, NFIB |
| Q14204 | DYHC1 | 31.87347932 | 12 | 0.0115 | 0.0286 | FOXC1, NFE2L3, NFIB |
| O75369 | FLNB | 41.60583942 | 14 | 0.0122 | 0.0299 | FOXC1, NFE2L3, NFIB |
| P62857 | RS28 | 44.28223844 | 14 | 0.0122 | 0.0299 | FOXC1, NFE2L3, NFIB |
| P07437 | TBB5 | 92.94403893 | 18 | 0.0122 | 0.0299 | FOXC1, NFE2L3, NFIB |
| Q9Y2T7 | YBOX2 | 42.09245742 | 4 | 0.0124 | 0.0302 | FOXC1, NFIB |
| Q14667 | K0100 | 1 | 8 | 0.0126 | 0.0305 | FOXC1, NFE2L3, NFIB |

| | | | | | | |
|---|---|---|---|---|---|---|
| Q9Y383 | LC7L2 | 1 | 12 | 0.0126 | 0.0305 | FOXC1, NFE2L3, NFIB |
| P15311 | EZRI | 27.00729927 | 13 | 0.0129 | 0.0306 | FOXC1, NFE2L3, NFIB |
| Q6FI13 | H2A2A | 65.93673966 | 16 | 0.0129 | 0.0306 | FOXC1, NFE2L3, NFIB |
| Q16777 | H2A2C | 65.93673966 | 16 | 0.0129 | 0.0306 | FOXC1, NFE2L3, NFIB |
| Q96I24 | FUBP3 | 15.32846715 | 7 | 0.0132 | 0.0312 | FOXC1, NFE2L3, NFIB |
| P07737 | PROF1 | 36.00973236 | 11 | 0.0134 | 0.0315 | FOXC1, NFE2L3, NFIB |
| P20290 | BTF3 | 8.515815085 | 9 | 0.014 | 0.0328 | FOXC1, NFE2L3, NFIB |
| Q9BUJ2 | HNRL1 | 25.79075426 | 12 | 0.0149 | 0.0345 | FOXC1, NFE2L3, NFIB |
| P47914 | RL29 | 41.11922141 | 17 | 0.0148 | 0.0345 | FOXC1, NFE2L3, NFIB |
| Q66PJ3 | AR6P4 | 26.03406326 | 13 | 0.0155 | 0.0358 | FOXC1, NFE2L3, NFIB |
| P62314 | SMD1 | 46.95863747 | 10 | 0.0161 | 0.037 | FOXC1, NFE2L3, NFIB |
| P16104 | H2AX | 56.20437956 | 16 | 0.0163 | 0.0372 | FOXC1, NFE2L3, NFIB |
| P68363 | TBA1B | 94.64720195 | 18 | 0.0164 | 0.0373 | FOXC1, NFE2L3, NFIB |
| Q14978 | NOLC1 | 29.19708029 | 10 | 0.0172 | 0.0388 | FOXC1, NFE2L3, NFIB |
| P31942 | HNRH3 | 23.35766423 | 13 | 0.0181 | 0.0407 | FOXC1, NFE2L3, NFIB |
| P62081 | RS7 | 48.66180049 | 15 | 0.0184 | 0.0412 | FOXC1, NFE2L3, NFIB |
| Q9BU76 | MMTA2 | 1 | 12 | 0.02 | 0.0446 | FOXC1, NFE2L3, NFIB |
| P18621 | RL17 | 46.22871046 | 13 | 0.0206 | 0.0458 | FOXC1, NFE2L3, NFIB |

| | | | | | | |
|---|---|---|---|---|---|---|
| P62805 | H4 | 54.74452555 | 18 | 0.0208 | 0.0459 | FOXC1, NFE2L3, NFIB |
| Q96AE4 | FUBP1 | 13.38199513 | 13 | 0.0212 | 0.0466 | FOXC1, NFE2L3, NFIB |
| Q9Y5B9 | SP16H | 18.49148418 | 11 | 0.0226 | 0.0492 | FOXC1, NFE2L3, NFIB |
| P11388 | TOP2A | 16.78832117 | 9 | 0.0226 | 0.0492 | FOXC1, NFE2L3, NFIB |

# APPENDIX F: LIST OF DCAS9 UNIQUE PROTEINS ACROSS GENES, NUMBER OF REPLICATES >4

| Accession | Protein name | Crapome | Number replicates | Relevant genes |
|---|---|---|---|---|
| P24043 | LAMA2 | 0.24 | 6 | FOXC1, NFE2L3, NFIB |
| Q9NVC6 | MED17 | 0.24 | 8 | FOXC1, NFE2L3, NFIB |
| Q8TD57 | DYH3 | 0.49 | 7 | FOXC1, NFE2L3, NFIB |
| Q14161 | GIT2 | 0.73 | 8 | FOXC1, NFE2L3, NFIB |
| Q16831 | UPP1 | 0.73 | 7 | FOXC1, NFE2L3, NFIB |
| O00442 | RTCA | 0.97 | 6 | FOXC1, NFE2L3, NFIB |
| E9PRG8 | CK098 | 1 | 8 | FOXC1, NFE2L3, NFIB |
| Q8N9M1 | CS047 | 1 | 9 | FOXC1, NFE2L3, NFIB |
| Q5T890 | ER6L2 | 1 | 8 | FOXC1, NFE2L3, NFIB |
| Q6PI47 | KCD18 | 1 | 6 | FOXC1, NFE2L3, NFIB |
| Q9BWE0 | REPI1 | 1 | 6 | FOXC1, NFIB |
| Q9UPW5 | CBPC1 | 1.46 | 8 | FOXC1, NFE2L3, NFIB |
| P21980 | TGM2 | 1.46 | 8 | FOXC1, NFE2L3, NFIB |
| Q16666 | IF16 | 2.19 | 6 | FOXC1, NFE2L3, NFIB |
| O43818 | U3IP2 | 2.43 | 6 | FOXC1, NFIB |
| Q9UMY1 | NOL7 | 2.68 | 6 | FOXC1, NFIB |
| Q6P6C2 | ALKB5 | 2.92 | 9 | FOXC1, NFE2L3, NFIB |
| O00479 | HMGN4 | 2.92 | 6 | FOXC1, NFE2L3, NFIB |
| Q5C9Z4 | NOM1 | 2.92 | 8 | FOXC1, NFE2L3, NFIB |
| Q96QD9 | UIF | 3.16 | 9 | FOXC1, NFE2L3, NFIB |
| Q9NZM1 | MYOF | 3.65 | 8 | FOXC1, NFE2L3, NFIB |
| Q15397 | K0020 | 3.89 | 6 | FOXC1, NFIB |
| Q9NR12 | PDLI7 | 3.89 | 6 | FOXC1, NFE2L3, NFIB |
| P52298 | NCBP2 | 4.62 | 7 | FOXC1, NFE2L3, NFIB |
| Q9Y3C1 | NOP16 | 4.62 | 6 | FOXC1, NFIB |
| Q9BXS6 | NUSAP | 4.62 | 6 | FOXC1, NFIB |
| O43159 | RRP8 | 4.87 | 6 | FOXC1, NFIB |
| P23193 | TCEA1 | 4.87 | 7 | FOXC1, NFE2L3, NFIB |
| Q9UNZ5 | L10K | 5.11 | 6 | FOXC1, NFE2L3, NFIB |
| Q53F19 | CQ085 | 5.35 | 9 | FOXC1, NFE2L3, NFIB |
| Q14320 | FA50A | 5.35 | 9 | FOXC1, NFE2L3, NFIB |
| Q9NWT1 | PK1IP | 5.35 | 6 | FOXC1, NFIB |
| Q9H6F5 | CCD86 | 5.6 | 7 | FOXC1, NFE2L3, NFIB |
| Q9NVU7 | SDA1 | 5.6 | 8 | FOXC1, NFE2L3, NFIB |
| P17480 | UBF1 | 5.6 | 6 | FOXC1, NFIB |
| Q96MU7 | YTDC1 | 5.6 | 9 | FOXC1, NFE2L3, NFIB |
| Q00534 | CDK6 | 5.84 | 7 | FOXC1, NFE2L3, NFIB |

| | | | | |
|---|---|---|---|---|
| Q92979 | NEP1 | 5.84 | 8 | FOXC1, NFE2L3, NFIB |
| Q6DKI1 | RL7L | 5.84 | 8 | FOXC1, NFE2L3, NFIB |
| Q9Y5J1 | UTP18 | 6.33 | 9 | FOXC1, NFE2L3, NFIB |
| Q96S55 | WRIP1 | 6.33 | 8 | FOXC1, NFE2L3, NFIB |
| P13984 | T2FB | 6.57 | 6 | FOXC1, NFIB |
| Q9Y3T9 | NOC2L | 6.57 | 6 | FOXC1, NFIB |
| Q96EV2 | RBM33 | 6.57 | 9 | FOXC1, NFE2L3, NFIB |
| Q9H7B2 | RPF2 | 6.57 | 7 | FOXC1, NFE2L3, NFIB |
| O94979 | SC31A | 6.57 | 6 | NFE2L3, NFIB |
| Q15061 | WDR43 | 6.57 | 6 | FOXC1, NFE2L3, NFIB |
| Q9BTT0 | AN32E | 6.81 | 7 | FOXC1, NFE2L3, NFIB |
| Q9NRL2 | BAZ1A | 6.81 | 9 | FOXC1, NFE2L3, NFIB |
| P06703 | S10A6 | 7.06 | 8 | FOXC1, NFE2L3, NFIB |
| Q13206 | DDX10 | 7.3 | 6 | FOXC1, NFE2L3, NFIB |
| P43490 | NAMPT | 8.27 | 6 | FOXC1, NFE2L3, NFIB |
| Q5QJE6 | TDIF2 | 8.52 | 6 | FOXC1, NFIB |
| P61326 | MGN | 9 | 9 | FOXC1, NFE2L3, NFIB |
| Q96A72 | MGN2 | 9 | 8 | FOXC1, NFE2L3, NFIB |
| O15347 | HMGB3 | 9.25 | 6 | FOXC1, NFIB |
| Q14690 | RRP5 | 9.25 | 6 | FOXC1, NFIB |
| Q15050 | RRS1 | 9.25 | 7 | FOXC1, NFE2L3, NFIB |
| Q03701 | CEBPZ | 9.49 | 6 | FOXC1, NFIB |
| P31949 | S10AB | 9.49 | 6 | FOXC1, NFE2L3, NFIB |
| Q8WWK9 | CKAP2 | 9.73 | 7 | FOXC1, NFE2L3, NFIB |
| Q9GZR7 | DDX24 | 9.73 | 9 | FOXC1, NFE2L3, NFIB |
| Q9NY93 | DDX56 | 9.73 | 6 | FOXC1, NFE2L3, NFIB |
| Q99575 | POP1 | 9.73 | 6 | FOXC1, NFIB |
| Q96T37 | RBM15 | 9.73 | 8 | FOXC1, NFE2L3, NFIB |
| Q5T8P6 | RBM26 | 9.73 | 9 | FOXC1, NFE2L3, NFIB |
| Q9Y3B9 | RRP15 | 9.73 | 6 | FOXC1, NFIB |
| P04083 | ANXA1 | 9.98 | 7 | FOXC1, NFE2L3, NFIB |
| Q96GQ7 | DDX27 | 9.98 | 8 | FOXC1, NFE2L3, NFIB |
| O15027 | SC16A | 9.98 | 6 | NFE2L3, NFIB |
| P26358 | DNMT1 | 10.22 | 6 | FOXC1, NFIB |
| Q9NQ55 | SSF1 | 10.22 | 7 | FOXC1, NFE2L3, NFIB |
| Q6PJT7 | ZC3HE | 10.22 | 9 | FOXC1, NFE2L3, NFIB |
| Q9Y2S6 | TMA7 | 10.46 | 7 | FOXC1, NFE2L3, NFIB |
| Q9NX58 | LYAR | 10.71 | 7 | FOXC1, NFE2L3, NFIB |
| Q9UKD2 | MRT4 | 10.71 | 8 | FOXC1, NFE2L3, NFIB |
| Q9H1E3 | NUCKS | 10.71 | 8 | FOXC1, NFE2L3, NFIB |
| Q9ULW0 | TPX2 | 10.71 | 9 | FOXC1, NFE2L3, NFIB |
| Q53GS9 | SNUT2 | 10.71 | 6 | FOXC1, NFIB |
| Q9H0S4 | DDX47 | 10.95 | 8 | FOXC1, NFE2L3, NFIB |
| Q13823 | NOG2 | 10.95 | 7 | FOXC1, NFE2L3, NFIB |
| P35269 | T2FA | 10.95 | 6 | FOXC1, NFIB |

| | | | | |
|---|---|---|---|---|
| Q9NZM5 | GSCR2 | 11.19 | 6 | FOXC1, NFE2L3, NFIB |
| O15226 | NKRF | 11.19 | 8 | FOXC1, NFE2L3, NFIB |
| O75152 | ZC11A | 11.19 | 9 | FOXC1, NFE2L3, NFIB |
| Q9NWH9 | SLTM | 11.44 | 8 | FOXC1, NFE2L3, NFIB |
| Q8TDD1 | DDX54 | 11.92 | 9 | FOXC1, NFE2L3, NFIB |
| Q6P1J9 | CDC73 | 12.17 | 7 | FOXC1, NFE2L3, NFIB |
| Q96EP5 | DAZP1 | 12.17 | 6 | FOXC1, NFE2L3, NFIB |
| Q86U42 | PABP2 | 12.17 | 9 | FOXC1, NFE2L3, NFIB |
| O43172 | PRP4 | 12.17 | 6 | FOXC1, NFIB |
| O75937 | DNJC8 | 12.41 | 6 | FOXC1, NFIB |
| Q9BZZ5 | API5 | 12.65 | 9 | FOXC1, NFE2L3, NFIB |
| P04637 | P53 | 12.65 | 6 | FOXC1, NFE2L3, NFIB |
| Q9NW13 | RBM28 | 13.14 | 7 | FOXC1, NFE2L3, NFIB |
| Q9Y5S9 | RBM8A | 13.14 | 9 | FOXC1, NFE2L3, NFIB |
| Q9BZE4 | NOG1 | 13.63 | 9 | FOXC1, NFE2L3, NFIB |
| Q9H307 | PININ | 13.63 | 9 | FOXC1, NFE2L3, NFIB |
| P55769 | NH2L1 | 14.36 | 6 | FOXC1, NFIB |
| O43395 | PRPF3 | 14.36 | 9 | FOXC1, NFE2L3, NFIB |
| Q7L4I2 | RSRC2 | 14.36 | 8 | FOXC1, NFE2L3, NFIB |
| O00422 | SAP18 | 14.36 | 9 | FOXC1, NFE2L3, NFIB |
| P17096 | HMGA1 | 14.84 | 6 | FOXC1, NFE2L3, NFIB |
| Q5JTH9 | RRP12 | 14.84 | 7 | FOXC1, NFE2L3, NFIB |
| Q9P2N5 | RBM27 | 15.09 | 8 | FOXC1, NFE2L3, NFIB |
| Q14684 | RRP1B | 15.09 | 6 | FOXC1, NFIB |
| O14776 | TCRG1 | 15.57 | 6 | FOXC1, NFE2L3, NFIB |
| Q9NVP1 | DDX18 | 16.06 | 8 | FOXC1, NFE2L3, NFIB |
| O94776 | MTA2 | 16.06 | 7 | FOXC1, NFE2L3, NFIB |
| Q96KR1 | ZFR | 16.55 | 9 | FOXC1, NFE2L3, NFIB |
| Q13573 | SNW1 | 16.79 | 7 | FOXC1, NFE2L3, NFIB |
| Q01780 | EXOSX | 17.03 | 6 | FOXC1, NFIB |
| Q9H0A0 | NAT10 | 17.03 | 7 | FOXC1, NFE2L3, NFIB |
| O75494 | SRS10 | 17.27 | 9 | FOXC1, NFE2L3, NFIB |
| Q9UKV3 | ACINU | 17.76 | 9 | FOXC1, NFE2L3, NFIB |
| Q9UQE7 | SMC3 | 17.76 | 6 | FOXC1, NFE2L3, NFIB |
| O95218 | ZRAB2 | 17.76 | 6 | FOXC1, NFIB |
| O60832 | DKC1 | 18 | 8 | FOXC1, NFE2L3, NFIB |
| Q13242 | SRSF9 | 18.73 | 8 | FOXC1, NFE2L3, NFIB |
| P16949 | STMN1 | 18.98 | 6 | FOXC1, NFIB |
| Q14683 | SMC1A | 19.22 | 7 | FOXC1, NFE2L3, NFIB |
| P46013 | KI67 | 19.46 | 9 | FOXC1, NFE2L3, NFIB |
| P37108 | SRP14 | 19.46 | 6 | FOXC1, NFE2L3, NFIB |
| P06493 | CDK1 | 19.71 | 6 | FOXC1, NFIB |
| P61956 | SUMO2 | 19.71 | 8 | FOXC1, NFE2L3, NFIB |
| Q9BXP5 | SRRT | 19.95 | 7 | FOXC1, NFE2L3, NFIB |
| P27694 | RFA1 | 20.44 | 9 | FOXC1, NFE2L3, NFIB |

| | | | | |
|---|---|---|---|---|
| Q13185 | CBX3 | 20.68 | 9 | FOXC1, NFE2L3, NFIB |
| P56537 | IF6 | 20.68 | 8 | FOXC1, NFE2L3, NFIB |
| Q9Y2X3 | NOP58 | 21.9 | 9 | FOXC1, NFE2L3, NFIB |
| P41091 | IF2G | 22.14 | 7 | FOXC1, NFE2L3, NFIB |
| P46087 | NOP2 | 22.14 | 8 | FOXC1, NFE2L3, NFIB |
| O00567 | NOP56 | 23.36 | 6 | FOXC1, NFIB |
| Q05519 | SRS11 | 23.6 | 6 | NFE2L3, NFIB |
| Q9H0D6 | XRN2 | 23.6 | 8 | FOXC1, NFE2L3, NFIB |
| P23526 | SAHH | 23.84 | 6 | FOXC1, NFE2L3, NFIB |
| P39748 | FEN1 | 23.84 | 8 | FOXC1, NFE2L3, NFIB |
| Q13428 | TCOF | 25.06 | 6 | FOXC1, NFIB |
| Q14011 | CIRBP | 26.03 | 9 | FOXC1, NFE2L3, NFIB |
| P18077 | RL35A | 26.28 | 6 | FOXC1, NFE2L3, NFIB |
| Q9P258 | RCC2 | 27.01 | 6 | FOXC1, NFIB |
| O43684 | BUB3 | 27.25 | 6 | FOXC1, NFE2L3, NFIB |
| P17980 | PRS6A | 27.49 | 7 | FOXC1, NFE2L3, NFIB |
| Q92769 | HDAC2 | 27.98 | 6 | FOXC1, NFE2L3, NFIB |
| Q8N163 | CCAR2 | 28.71 | 6 | FOXC1, NFE2L3, NFIB |
| Q9NQ29 | LUC7L | 28.71 | 9 | FOXC1, NFE2L3, NFIB |
| O43809 | CPSF5 | 28.95 | 6 | FOXC1, NFE2L3, NFIB |
| O95232 | LC7L3 | 29.68 | 8 | FOXC1, NFE2L3, NFIB |
| E9PAV3 | NACAM | 30.41 | 7 | FOXC1, NFE2L3, NFIB |
| Q13765 | NACA | 30.41 | 6 | FOXC1, NFE2L3 |
| P15924 | DESP | 30.66 | 7 | FOXC1, NFE2L3, NFIB |
| P62841 | RS15 | 30.66 | 7 | FOXC1, NFE2L3, NFIB |
| P42167 | LAP2B | 30.9 | 6 | FOXC1, NFIB |
| Q9Y230 | RUVB2 | 34.31 | 6 | FOXC1, NFE2L3, NFIB |
| Q8WWY3 | PRP31 | 35.28 | 9 | FOXC1, NFE2L3, NFIB |
| Q9UHX1 | PUF60 | 35.28 | 7 | FOXC1, NFE2L3, NFIB |
| P17661 | DESM | 38.44 | 6 | FOXC1, NFE2L3, NFIB |
| Q96E39 | RMXL1 | 38.69 | 7 | FOXC1, NFE2L3, NFIB |
| P14678 | RSMB | 41.36 | 9 | FOXC1, NFE2L3, NFIB |
| Q9NYF8 | BCLF1 | 41.61 | 9 | FOXC1, NFE2L3, NFIB |
| O75533 | SF3B1 | 41.61 | 8 | FOXC1, NFE2L3, NFIB |
| P62318 | SMD3 | 42.82 | 9 | FOXC1, NFE2L3, NFIB |
| P46783 | RS10 | 43.07 | 9 | FOXC1, NFE2L3, NFIB |
| Q14240 | IF4A2 | 43.31 | 6 | FOXC1, NFE2L3, NFIB |
| Q9Y2W1 | TR150 | 43.31 | 9 | FOXC1, NFE2L3, NFIB |
| Q06830 | PRDX1 | 61.56 | 9 | FOXC1, NFE2L3, NFIB |

# APPENDIX G: RIME HITS RANKING: FIRST 100 TOP CANDIDATES RANKED ON A NOVELTY BASIS

| Accession | Crapome | N° rep | genes | p. adjust | mlogFC | N° Cpds | N° Cpd pact>5 | N° AZ Cpd pact>5 | N° Cpd Clinic | type | N°ref | known | newer |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Q00534\|CDK6_HUMAN | 5.84 | 7 | FOXC1, NFE2L3, NFIB | -1 | 1.439236227 | 2559 | 2095 | 5 | 4 | Regulation | 1 | 0.91 | 0.97 |
| P06493\|CDK1_HUMAN | 19.71 | 6 | FOXC1, NFIB | -1 | 0.622985294 | 24443 | 10953 | 11 | 4 | Regulation | 1 | 0.82 | 0.87 |
| P11388\|TOP2A_HUMAN | 16.79 | 9 | FOXC1, NFE2L3, NFIB | 0.0492 | 0.854372122 | 2077 | 117 | 0 | 121 | Regulation | 1 | 0.61 | 0.65 |
| P09874\|PARP1_HUMAN | 44.28 | 18 | FOXC1, NFE2L3, NFIB | 0.0219 |  | 8819 | 6670 | 6 | 8 | Regulation | 10 | 0.96 | 0.61 |
| P04637\|P53_HUMAN | 12.65 | 6 | FOXC1, NFE2L3, NFIB | -1 |  | 37387 | 3580 | 0 | 8 | ClinicalTrial,Regulation | 18 | 0.63 | 0.39 |
| Q92769\|HDAC2_HUMAN | 27.98 | 6 | FOXC1, NFE2L3, NFIB | -1 | 0.728621906 | 7117 | 4224 | 0 | 8 | NoRelto TNBC | 0 | 0.30 | 0.30 |
| P43490\|NAMPT_HUMAN | 8.27 | 6 | FOXC1, NFE2L3, NFIB | -1 | 0.499884987 | 9340 | 8848 | 0 | 3 | NoRelto TNBC | 0 | 0.29 | 0.29 |
| P26358\|DNMT1_HUMAN | 10.22 | 6 | FOXC1, NFIB | -1 | 0.591521458 | 2445 | 190 | 0 | 3 | NoRelto TNBC | 0 | 0.28 | 0.28 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P04406\|G3P_HUMAN | 60.34 | 16 | FOXC1, NFE2L3, NFIB | 0.0008 | 0.758192562 | 682 | 18 | 0 | 1 | NoRelto TNBC | 0 | 0.25 | 0.25 |
| P23526\|SAHH_HUMAN | 23.84 | 6 | FOXC1, NFE2L3, NFIB | -1 | 0.217309025 | 699 | 307 | 0 | 2 | NoRelto TNBC | 0 | 0.21 | 0.21 |
| P62937\|PPIA_HUMAN | 44.77 | 14 | FOXC1, NFE2L3, NFIB | 0.0072 | 0.042381598 | 765 | 378 | 0 | 8 | NoRelto TNBC | 0 | 0.20 | 0.20 |
| P21333\|FLNA_HUMAN | 54.99 | 15 | FOXC1, NFE2L3, NFIB | 0.005 | 0.174363056 | 291 | 291 | 0 | 0 | Regulation | 3 | 0.18 | 0.19 |
| P07437\|TBB5_HUMAN | 92.94 | 18 | FOXC1, NFE2L3, NFIB | 0.0299 | 0.542417078 | 781 | 173 | 0 | 30 | NoRelto TNBC | 0 | 0.18 | 0.18 |
| P19338\|NUCL_HUMAN | 63.5 | 17 | FOXC1, NFE2L3, NFIB | 0.0002 | 0.31215718 | 0 | 0 | 0 | 1 | CellExpression | 1 | 0.16 | 0.17 |
| P11387\|TOP1_HUMAN | 26.76 | 10 | FOXC1, NFE2L3, NFIB | 0.0127 | 0.012205904 | 1623 | 285 | 0 | 50 | NoRelto TNBC | 0 | 0.14 | 0.14 |
| Q16831\|UPP1_HUMAN | 0.73 | 7 | FOXC1, NFE2L3, NFIB | -1 | 0.005089477 | 301 | 79 | 0 | 2 | NoRelto TNBC | 0 | 0.14 | 0.14 |
| P12268\|IMDH2_HUMAN | 20.19 | 11 | FOXC1, NFE2L3, NFIB | 0.0061 | -0.411410207 | 2058 | 1386 | 0 | 0 | NoRelto TNBC | 0 | 0.14 | 0.14 |
| P39748\|FEN1_HUMAN | 23.84 | 8 | FOXC1, NFE2L3, NFIB | -1 | 0.617906432 | 18871 | 851 | 0 | 0 | NoRelto TNBC | 0 | 0.14 | 0.14 |
| Q9NZM1\|MYOF_HU | 3.65 | 8 | FOXC1, NFE2L3, | -1 | -0.446525 | 0 | 0 | 0 | 0 | Regulation | 1 | 0.12 | 0.13 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MAN | | | NFIB | | 596 | | | | | | | | |
| Q9H1E3\|NUCKS_HUMAN | 10.71 | 8 | FOXC1, NFE2L3, NFIB | -1 | 0.10407036 | 2244 | 284 | 3 | 0 | NoRelto TNBC | 0 | 0.13 | 0.13 |
| P08621\|RU17_HUMAN | 26.28 | 15 | FOXC1, NFE2L3, NFIB | 0.0043 | -0.223690007 | 0 | 0 | 0 | 1 | NoRelto TNBC | 0 | 0.13 | 0.13 |
| Q9BXS6\|NUSAP_HUMAN | 4.62 | 6 | FOXC1, NFIB | -1 | 0.627675524 | 0 | 0 | 0 | 0 | Regulation | 3 | 0.12 | 0.13 |
| P06748\|NPM_HUMAN | 61.31 | 18 | FOXC1, NFE2L3, NFIB | 0.0007 | | 13557 | 5708 | 21 | 0 | NoRelto TNBC | 0 | 0.12 | 0.12 |
| P46013\|KI67_HUMAN | 19.46 | 9 | FOXC1, NFE2L3, NFIB | -1 | 0.426718768 | 0 | 0 | 0 | 0 | Regulation | 6 | 0.12 | 0.12 |
| P06733\|ENOA_HUMAN | 54.26 | 17 | FOXC1, NFE2L3, NFIB | 0.0071 | 0.794191533 | 2 | 1 | 0 | 0 | NoRelto TNBC | 0 | 0.10 | 0.10 |
| P08670\|VIME_HUMAN | 62.53 | 18 | FOXC1, NFE2L3, NFIB | 0.0009 | 0.437449241 | 1 | 0 | 0 | 1 | NoRelto TNBC | 0 | 0.10 | 0.10 |
| P02545\|LMNA_HUMAN | 22.38 | 18 | FOXC1, NFE2L3, NFIB | 0.0058 | 0.31086473 | 36141 | 6147 | 0 | 0 | NoRelto TNBC | 0 | 0.1 | 0.1 |
| P17980\|PRS6A_HUMAN | 27.49 | 7 | FOXC1, NFE2L3, NFIB | -1 | 0.167019816 | 125 | 83 | 0 | 0 | NoRelto TNBC | 0 | 0.1 | 0.1 |
| P27694\|RFA1_HUMAN | 20.44 | 9 | FOXC1, NFE2L3, NFIB | -1 | 0.341587719 | 245 | 56 | 0 | 0 | NoRelto TNBC | 0 | 0.1 | 0.1 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P21980\|TGM2_HUMAN | 1.46 | 8 | FOXC1, NFE2L3, NFIB | -1 | 0.185262094 | 7192 | 416 | 0 | 0 | NoRelto TNBC | 0 | 0.09 | 0.09 |
| P11940\|PABP1_HUMAN | 41.85 | 13 | FOXC1, NFE2L3, NFIB | 0.0081 | 0.411681485 | 843 | 314 | 0 | 0 | NoRelto TNBC | 0 | 0.09 | 0.09 |
| P38919\|IF4A3_HUMAN | 35.77 | 16 | FOXC1, NFE2L3, NFIB | 0 | 0.138252245 | 56 | 27 | 0 | 0 | NoRelto TNBC | 0 | 0.09 | 0.09 |
| Q14683\|SMC1A_HUMAN | 19.22 | 7 | FOXC1, NFE2L3, NFIB | -1 | -0.141746662 | 0 | 0 | 0 | 0 | Regulation | 1 | 0.09 | 0.09 |
| P41091\|IF2G_HUMAN | 22.14 | 7 | FOXC1, NFE2L3, NFIB | -1 | 0.133219699 | 8 | 8 | 0 | 0 | NoRelto TNBC | 0 | 0.09 | 0.09 |
| P16403\|H12_HUMAN | 73.24 | 18 | FOXC1, NFE2L3, NFIB | 0.0153 | -0.476225517 | 11 | 8 | 0 | 0 | NoRelto TNBC | 0 | 0.09 | 0.09 |
| Q9NQ55\|SSF1_HUMAN | 10.22 | 7 | FOXC1, NFE2L3, NFIB | -1 | | 40 | 6 | 0 | 0 | NoRelto TNBC | 0 | 0.09 | 0.09 |
| Q15717\|ELAV1_HUMAN | 24.57 | 15 | FOXC1, NFE2L3, NFIB | 0.0006 | 0.177883213 | 15 | 3 | 0 | 0 | NoRelto TNBC | 0 | 0.09 | 0.09 |
| Q07666\|KHDR1_HUMAN | 33.58 | 17 | FOXC1, NFE2L3, NFIB | 0.0014 | 0.067469558 | 36 | 36 | 0 | 0 | NoRelto TNBC | 0 | 0.08 | 0.08 |
| Q71U36\|TBA1A_HUMAN | 94.65 | 13 | FOXC1, NFE2L3, NFIB | 0.0281 | 0.338243876 | 728 | 152 | 2 | 0 | NoRelto TNBC | 0 | 0.08 | 0.08 |
| O43172\|PRP4_HUM | 12.17 | 6 | FOXC1, NFIB | -1 | 0.150382745 | 5 | 3 | 0 | 0 | NoRelto TNBC | 0 | 0.08 | 0.08 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| AN | | | | | | | | | | | | | |
| P60842\|IF4A1_HUMAN | 46.23 | 14 | FOXC1, NFE2L3, NFIB | 0 | | 8 | 3 | 0 | 0 | NoRelto TNBC | 0 | 0.08 | 0.08 |
| P68104\|EF1A1_HUMAN | 85.16 | 18 | FOXC1, NFE2L3, NFIB | 0.0061 | 0.047344463 | 11 | 10 | 0 | 1 | NoRelto TNBC | 0 | 0.08 | 0.08 |
| Q15365\|PCBP1_HUMAN | 47.2 | 15 | FOXC1, NFE2L3, NFIB | 0.0013 | 0.143778023 | 2 | 2 | 0 | 0 | NoRelto TNBC | 0 | 0.08 | 0.08 |
| P08865\|RSSA_HUMAN | 38.93 | 16 | FOXC1, NFE2L3, NFIB | 0 | 0.065765317 | 1 | 1 | 0 | 0 | NoRelto TNBC | 0 | 0.08 | 0.08 |
| P23528\|COF1_HUMAN | 47.69 | 16 | FOXC1, NFE2L3, NFIB | 0.0019 | 0.208933807 | 1 | 1 | 0 | 0 | NoRelto TNBC | 0 | 0.07 | 0.07 |
| P08238\|HS90B_HUMAN | 67.4 | 18 | FOXC1, NFE2L3, NFIB | 0.0239 | 0.181627528 | 1813 | 959 | 0 | 0 | NoRelto TNBC | 0 | 0.07 | 0.07 |
| O00422\|SAP18_HUMAN | 14.36 | 9 | FOXC1, NFE2L3, NFIB | -1 | -0.774650629 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| Q14011\|CIRBP_HUMAN | 26.03 | 9 | FOXC1, NFE2L3, NFIB | -1 | -1.110894726 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| P11532\|DMD_HUMAN | 1.22 | 10 | FOXC1, NFE2L3, NFIB | 0.025 | 0.056003077 | 0 | 0 | 0 | 3 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| Q9ULW0\|TPX2_HUMAN | 10.71 | 9 | FOXC1, NFE2L3, NFIB | -1 | 1.099977504 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P07477\|TRY1_HUMAN | 15.82 | 18 | FOXC1, NFE2L3, NFIB | 0.0064 | -0.000892578 | 8477 | 3478 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| P46939\|UTRO_HUMAN | 6.57 | 13 | FOXC1, NFE2L3, NFIB | 0.0039 | -0.043355989 | 0 | 0 | 0 | 2 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| Q9BZE4\|NOG1_HUMAN | 13.63 | 9 | FOXC1, NFE2L3, NFIB | -1 | 0.965905846 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| P22626\|ROA2_HUMAN | 63.75 | 18 | FOXC1, NFE2L3, NFIB | 0.0013 | -0.096630363 | 1 | 1 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| O75369\|FLNB_HUMAN | 41.61 | 14 | FOXC1, NFE2L3, NFIB | 0.0299 | -0.713406086 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| P04083\|ANXA1_HUMAN | 9.98 | 7 | FOXC1, NFE2L3, NFIB | -1 | 0.818941475 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| P08729\|K2C7_HUMAN | 37.71 | 18 | FOXC1, NFE2L3, NFIB | 0.0115 | 0.811377915 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| P20290\|BTF3_HUMAN | 8.52 | 9 | FOXC1, NFE2L3, NFIB | 0.0328 | -0.527273631 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| Q6VAB6\|KSR2_HUMAN | 1 | 13 | FOXC1, NFE2L3, NFIB | 0.014 | -0.035123262 | 20 | 17 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| P84243\|H33_HUMAN | 33.82 | 18 | FOXC1, NFE2L3, NFIB | 0.0018 | -0.499038876 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| O60832\|DKC1_HU | 18 | 8 | FOXC1, NFE2L3, | -1 | 0.729023914 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MAN | | | NFIB | | | | | | | | | | |
| P09651\|ROA1_HUMAN | 65.21 | 17 | FOXC1, NFE2L3, NFIB | 0.0001 | 0.065488567 | 1 | 1 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| P17096\|HMGA1_HUMAN | 14.84 | 6 | FOXC1, NFE2L3, NFIB | -1 | 0.021796769 | 0 | 0 | 0 | 0 | Regulation | 1 | 0.06 | 0.06 |
| Q7L4I2\|RSRC2_HUMAN | 14.36 | 8 | FOXC1, NFE2L3, NFIB | -1 | -0.084986913 | 0 | 0 | 0 | 0 | Regulation | 1 | 0.06 | 0.06 |
| O75607\|NPM3_HUMAN | 11.92 | 14 | FOXC1, NFE2L3, NFIB | 0.0098 | 0.703137464 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| Q14980\|NUMA1_HUMAN | 18.98 | 10 | FOXC1, NFE2L3, NFIB | 0.0058 | -0.459706335 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| P06703\|S10A6_HUMAN | 7.06 | 8 | FOXC1, NFE2L3, NFIB | -1 | 0.664536461 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| Q9H7B2\|RPF2_HUMAN | 6.57 | 7 | FOXC1, NFE2L3, NFIB | -1 | 0.652865116 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| P49207\|RL34_HUMAN | 20.44 | 18 | FOXC1, NFE2L3, NFIB | 0.0021 | 0.651378428 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| P40429\|RL13A_HUMAN | 35.04 | 14 | FOXC1, NFE2L3, NFIB | 0.0027 | -0.396639156 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| O00148\|DX39A_HUMAN | 33.58 | 12 | FOXC1, NFE2L3, NFIB | 0 | 0.638998333 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P24043\|LAMA2_HUMAN | 0.24 | 6 | FOXC1, NFE2L3, NFIB | -1 | -0.389470058 | 1 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| Q9NWH9\|SLTM_HUMAN | 11.44 | 8 | FOXC1, NFE2L3, NFIB | -1 | -0.385685743 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| Q9NX58\|LYAR_HUMAN | 10.71 | 7 | FOXC1, NFE2L3, NFIB | -1 | 0.619657015 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| P53999\|TCP4_HUMAN | 38.93 | 10 | FOXC1, NFE2L3, NFIB | 0.0063 | -0.369622243 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| P22087\|FBRL_HUMAN | 39.17 | 15 | FOXC1, NFE2L3, NFIB | 0.0014 | 0.606127189 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| P37108\|SRP14_HUMAN | 19.46 | 6 | FOXC1, NFE2L3, NFIB | -1 | -0.360199008 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| Q16666\|IF16_HUMAN | 2.19 | 6 | FOXC1, NFE2L3, NFIB | -1 | 0.595680282 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| P15311\|EZRI_HUMAN | 27.01 | 13 | FOXC1, NFE2L3, NFIB | 0.0306 | -0.346938262 | 16 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| P13611\|CSPG2_HUMAN | 0.49 | 17 | FOXC1, NFE2L3, NFIB | 0.0006 | -0.342411417 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| P31949\|S10AB_HUMAN | 9.49 | 6 | FOXC1, NFE2L3, NFIB | -1 | 0.576932267 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| Q969Q0\|RL36L_HU | 18.25 | 17 | FOXC1, NFE2L3, | 0.0176 | -0.303341 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| MAN | | | NFIB | | 162 | | | | | | | | |
| Q15061\|WDR43_HUMAN | 6.57 | 6 | FOXC1, NFE2L3, NFIB | -1 | 0.54617118 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| P46087\|NOP2_HUMAN | 22.14 | 8 | FOXC1, NFE2L3, NFIB | -1 | 0.529283327 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| Q9H6F5\|CCD86_HUMAN | 5.6 | 7 | FOXC1, NFE2L3, NFIB | -1 | 0.52790997 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| Q9NW13\|RBM28_HUMAN | 13.14 | 7 | FOXC1, NFE2L3, NFIB | -1 | 0.524771825 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| O15027\|SC16A_HUMAN | 9.98 | 6 | NFE2L3, NFIB | -1 | -0.568319769 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| P61326\|MGN_HUMAN | 9 | 9 | FOXC1, NFE2L3, NFIB | -1 | 0.509331947 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| Q8TDN6\|BRX1_HUMAN | 16.3 | 12 | FOXC1, NFE2L3, NFIB | 0.0031 | 0.501433548 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| O60869\|EDF1_HUMAN | 19.46 | 10 | FOXC1, NFE2L3, NFIB | 0.0037 | -0.2561233 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| P62308\|RUXG_HUMAN | 11.92 | 12 | FOXC1, NFE2L3, NFIB | 0.0019 | 0.498979714 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| Q9Y2X3\|NOP58_HUMAN | 21.9 | 9 | FOXC1, NFE2L3, NFIB | -1 | 0.495899605 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Q92979\|NEP1_HUMAN | 5.84 | 8 | FOXC1, NFE2L3, NFIB | -1 | 0.470572942 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| O00571\|DDX3X_HUMAN | 51.58 | 16 | FOXC1, NFE2L3, NFIB | 0.0007 | 0.02289704 | 80 | 22 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| Q6ZNL6\|FGD5_HUMAN | 1 | 17 | FOXC1, NFE2L3, NFIB | 0.0002 | -0.214139015 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| O95232\|LC7L3_HUMAN | 29.68 | 8 | FOXC1, NFE2L3, NFIB | -1 | -0.212021667 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| P68363\|TBA1B_HUMAN | 94.65 | 18 | FOXC1, NFE2L3, NFIB | 0.0373 | 0.31456492 | 716 | 142 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| P11142\|HSP7C_HUMAN | 96.35 | 18 | FOXC1, NFE2L3, NFIB | 0.0067 | 0.285458532 | 161 | 34 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |
| Q9NWT1\|PK1IP_HUMAN | 5.35 | 6 | FOXC1, NFIB | -1 | 0.529972805 | 0 | 0 | 0 | 0 | NoRelto TNBC | 0 | 0.06 | 0.06 |

# APPENDIX H: MTA2 AND CDK1 COMMON PEAKS IDENTIFIED THROUGH CHIP-SEQ ANALYSIS

| Seq Names | Start | End | Score | Seq Names | Start | End | Where | Symbol |
|---|---|---|---|---|---|---|---|---|
| chr1 | 55429434 | 55429658 | 0.71886 | chr1 | 55428001 | 55431001 | enhancer | NA |
| chr1 | 59151082 | 59151315 | 0.61045 | chr1 | 59133331 | 59151119 | intron | HSD52 |
| chr1 | 59151082 | 59151315 | 0.61045 | chr1 | 59133331 | 59198182 | intron | HSD52 |
| chr1 | 59151082 | 59151315 | 0.61045 | chr1 | 59151120 | 59151241 | exon | HSD52 |
| chr1 | 59151082 | 59151315 | 0.61045 | chr1 | 59133331 | 59203094 | intron | HSD52 |
| chr1 | 59151082 | 59151315 | 0.61045 | chr1 | 59133331 | 59208010 | intron | HSD52 |
| chr1 | 59151082 | 59151315 | 0.61045 | chr1 | 59132582 | 59184430 | intron | LINC01358 |
| chr1 | 59151082 | 59151315 | 0.61045 | chr1 | 59150202 | 59152199 | promoter_flanking_region | NA |
| chr1 | 85298209 | 85298495 | 1.6266 | chr1 | 85277739 | 85376765 | intron | RP11-131L23.1 |
| chr1 | 85298209 | 85298495 | 1.6266 | chr1 | 85297002 | 85298999 | promoter_flanking_region | NA |
| chr1 | 211259020 | 211259300 | 0.5961 | chr1 | 211258491 | 211260107 | intron | RCOR3 |
| chr1 | 211259020 | 211259300 | 0.5961 | chr1 | 211258526 | 211259578 | intron | RCOR3 |
| chr1 | 211259020 | 211259300 | 0.5961 | chr1 | 211259279 | 211259560 | five_prime_utr | RCOR3 |
| chr1 | 211259020 | 211259300 | 0.5961 | chr1 | 211259279 | 211259726 | exon | RCOR3 |
| chr1 | 211259020 | 211259300 | 0.5961 | chr1 | 211258000 | 211261401 | promoter | NA |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr10 | 15200669 | 15200938 | 0.60408 | chr10 | 15200002 | 15202999 | promoter_flanking_region | NA |
| chr10 | 31940989 | 31941232 | 0.48901 | chr10 | 31939802 | 31941999 | promoter_flanking_region | NA |
| chr10 | 78520955 | 78521190 | 1.22506 | chr10 | 78249071 | 78525329 | intron | LINC00856 |
| chr10 | 78520955 | 78521190 | 1.22506 | chr10 | 78179247 | 78525329 | intron | LINC00856 |
| chr10 | 78520955 | 78521190 | 1.22506 | chr10 | 78406700 | 78525329 | intron | RP11-90J7.3 |
| chr10 | 78520955 | 78521190 | 1.22506 | chr10 | 78238305 | 78528750 | intron | RP11-90J7.3 |
| chr10 | 78520955 | 78521190 | 1.22506 | chr10 | 78466179 | 78525329 | intron | RP11-90J7.3 |
| chr10 | 78520955 | 78521190 | 1.22506 | chr10 | 78521001 | 78521400 | ctcf_binding_site | NA |
| chr10 | 78520955 | 78521190 | 1.22506 | chr10 | 78520955 | 78521469 | open_chromatin | NA |
| chr10 | 110414567 | 110415026 | 0.10265 | chr10 | 110413402 | 110415999 | promoter_flanking_region | NA |
| chr10 | 114271671 | 114271909 | 0.96656 | chr10 | 114261296 | 114272739 | intron | VWA2 |
| chr10 | 114271671 | 114271909 | 0.96656 | chr10 | 114271201 | 114272401 | enhancer | NA |
| chr10 | 119253831 | 119254068 | 0.30664 | chr10 | 119207970 | 119326515 | intron | GRK5 |
| chr10 | 119253831 | 119254068 | 0.30664 | chr10 | 119250002 | 119255599 | promoter_flanking_region | NA |
| chr10 | 119667659 | 119667874 | 0.04108 | chr10 | 119651856 | 119669850 | intron | BAG3 |
| chr10 | 119667659 | 119667874 | 0.04108 | chr10 | 119657569 | 119669850 | intron | BAG3 |
| chr10 | 119667659 | 119667874 | 0.04108 | chr10 | 119666802 | 119668999 | promoter_flanking_region | NA |
| chr11 | 34371503 | 34371726 | 0.05178 | chr11 | 34371165 | 34372399 | promoter_flanking_region | NA |
| chr11 | 69122770 | 69123097 | 0.11619 | chr11 | 69121672 | 69124199 | promoter_flanking_region | NA |
| chr11 | 125079116 | 125079383 | 0.38631 | chr11 | 125079112 | 125079247 | exon,cds | SLC37A2 |
| chr11 | 125079116 | 125079383 | 0.38631 | chr11 | 125079248 | 125079683 | intron | SLC37A2 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr11 | 130383658 | 130383868 | 0.39025 | chr11 | 130369043 | 130393494 | intron | RP11-121M22.1 |
| chr11 | 130383658 | 130383868 | 0.39025 | chr11 | 130383400 | 130384600 | enhancer | NA |
| chr12 | 26274766 | 26275073 | 1.30365 | chr12 | 26252921 | 26287482 | intron | SSPN |
| chr12 | 26274766 | 26275073 | 1.30365 | chr12 | 26252921 | 26298264 | intron | SSPN |
| chr12 | 26274766 | 26275073 | 1.30365 | chr12 | 26231986 | 26319601 | intron | RP11-283G6.5 |
| chr12 | 26274766 | 26275073 | 1.30365 | chr12 | 26214778 | 26326383 | intron | RP11-283G6.4 |
| chr12 | 26274766 | 26275073 | 1.30365 | chr12 | 26273901 | 26274900 | five_prime_flank | RP11-283G6.6 |
| chr12 | 26274766 | 26275073 | 1.30365 | chr12 | 26273802 | 26276199 | promoter_flanking_region | NA |
| chr12 | 85465093 | 85465334 | 0.84325 | chr12 | 85465073 | 85465589 | open_chromatin | NA |
| chr12 | 124233414 | 124233654 | 0.14343 | chr12 | 124149792 | 124311817 | intron | FAM101A |
| chr12 | 124233414 | 124233654 | 0.14343 | chr12 | 124149792 | 124235544 | intron | FAM101A |
| chr12 | 124233414 | 124233654 | 0.14343 | chr12 | 124232402 | 124234611 | promoter_flanking_region | NA |
| chr14 | 22556427 | 22556668 | 0.98877 | chr14 | 22555640 | 22556639 | five_prime_flank | AE000662.92 |
| chr14 | 22556427 | 22556668 | 0.98877 | chr14 | 22556640 | 22556842 | exon | AE000662.92 |
| chr14 | 22556427 | 22556668 | 0.98877 | chr14 | 22556311 | 22556522 | exon | AE000662.93 |
| chr14 | 22556427 | 22556668 | 0.98877 | chr14 | 22556523 | 22556791 | intron | AE000662.93 |
| chr14 | 64806588 | 64806828 | 0.43819 | chr14 | 64805091 | 64822946 | intron | SPTB |
| chr14 | 64806588 | 64806828 | 0.43819 | chr14 | 64806202 | 64807029 | promoter_flanking_region | NA |
| chr14 | 95265501 | 95265711 | 0.36738 | chr14 | 95230134 | 95307513 | intron | CLMN |
| chr14 | 95265501 | 95265711 | 0.36738 | chr14 | 95230134 | 95319710 | intron | CLMN |
| chr14 | 95265501 | 95265711 | 0.36738 | chr14 | 95260603 | 95296133 | intron | CLMN |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr14 | 95265501 | 95265711 | 0.36738 | chr14 | 95264402 | 95266399 | promoter_flanking_region | NA |
| chr16 | 48964320 | 48964590 | 0.66186 | chr16 | 48963262 | 48964760 | promoter_flanking_region | NA |
| chr16 | 86951884 | 86952179 | 0.94826 | chr16 | 86951601 | 86953199 | promoter_flanking_region | NA |
| chr17 | 15273147 | 15273354 | 1.12975 | chr17 | 15272291 | 15273290 | five_prime_flank | AC005703.3 |
| chr17 | 15273147 | 15273354 | 1.12975 | chr17 | 15272719 | 15274614 | promoter_flanking_region | NA |
| chr17 | 16381345 | 16381554 | 0.23606 | chr17 | 16381341 | 16381515 | exon,five_prime_utr | UBB |
| chr17 | 16381345 | 16381554 | 0.23606 | chr17 | 16381465 | 16381679 | exon,five_prime_utr | UBB |
| chr17 | 16381345 | 16381554 | 0.23606 | chr17 | 16381516 | 16381901 | intron | UBB |
| chr17 | 16381345 | 16381554 | 0.23606 | chr17 | 16381030 | 16381901 | intron | UBB |
| chr17 | 16381345 | 16381554 | 0.23606 | chr17 | 16381037 | 16381901 | intron | UBB |
| chr17 | 16381345 | 16381554 | 0.23606 | chr17 | 16381290 | 16381462 | exon,five_prime_utr | UBB |
| chr17 | 16381345 | 16381554 | 0.23606 | chr17 | 16381185 | 16381901 | intron | UBB |
| chr17 | 16381345 | 16381554 | 0.23606 | chr17 | 16381185 | 16382358 | intron | UBB |
| chr17 | 16381345 | 16381554 | 0.23606 | chr17 | 16381463 | 16381901 | intron | UBB |
| chr17 | 16381345 | 16381554 | 0.23606 | chr17 | 16381152 | 16382151 | three_prime_flank | RP11-138I1.4 |
| chr17 | 16381345 | 16381554 | 0.23606 | chr17 | 16380200 | 16382001 | promoter | NA |
| chr17 | 55433073 | 55433322 | 0.2755 | chr17 | 55432602 | 55434599 | promoter_flanking_region | NA |
| chr17 | 59756148 | 59756477 | 0.26479 | chr17 | 59735474 | 59764970 | intron | VMP1 |
| chr17 | 59756148 | 59756477 | 0.26479 | chr17 | 59735474 | 59808795 | intron | VMP1 |
| chr17 | 59756148 | 59756477 | 0.26479 | chr17 | 59738948 | 59765051 | intron | VMP1 |
| chr17 | 59756148 | 59756477 | 0.26479 | chr17 | 59738948 | 59764970 | intron | VMP1 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr17 | 59756148 | 59756477 | 0.26479 | chr17 | 59752002 | 59758199 | promoter_flanking_region | NA |
| chr17 | 59786676 | 59786964 | 0.44025 | chr17 | 59735474 | 59808795 | intron | VMP1 |
| chr17 | 59786676 | 59786964 | 0.44025 | chr17 | 59773886 | 59808795 | intron | VMP1 |
| chr17 | 59786676 | 59786964 | 0.44025 | chr17 | 59782002 | 59793599 | promoter_flanking_region | NA |
| chr17 | 67441296 | 67441495 | 0.80391 | chr17 | 67379393 | 67532801 | intron | PITPNC1 |
| chr17 | 67441296 | 67441495 | 0.80391 | chr17 | 67378203 | 67532801 | intron | PITPNC1 |
| chr17 | 67441296 | 67441495 | 0.80391 | chr17 | 67440002 | 67444799 | promoter_flanking_region | NA |
| chr17 | 77125426 | 77125725 | 1.02836 | chr17 | 77089272 | 77142645 | intron | SEC14L1 |
| chr17 | 77125426 | 77125725 | 1.02836 | chr17 | 77093348 | 77142645 | intron | SEC14L1 |
| chr17 | 77125426 | 77125725 | 1.02836 | chr17 | 77117402 | 77130365 | promoter_flanking_region | NA |
| chr19 | 16877019 | 16877260 | 0.92906 | chr19 | 16876939 | 16877639 | exon | SIN3B |
| chr19 | 16877019 | 16877260 | 0.92906 | chr19 | 16877162 | 16877335 | exon | SIN3B |
| chr19 | 16877019 | 16877260 | 0.92906 | chr19 | 16876579 | 16877544 | intron | SIN3B |
| chr19 | 16877019 | 16877260 | 0.92906 | chr19 | 16876953 | 16877375 | open_chromatin | NA |
| chr19 | 43562801 | 43563012 | 0.06599 | chr19 | 43561021 | 43574909 | intron | XRCC1 |
| chr19 | 43562801 | 43563012 | 0.06599 | chr19 | 43561021 | 43575407 | intron | XRCC1 |
| chr19 | 43562801 | 43563012 | 0.06599 | chr19 | 43561021 | 43580364 | intron | XRCC1 |
| chr19 | 43562801 | 43563012 | 0.06599 | chr19 | 43554805 | 43574909 | intron | XRCC1 |
| chr19 | 43562801 | 43563012 | 0.06599 | chr19 | 43561021 | 43592867 | intron | L34079.2 |
| chr19 | 43562801 | 43563012 | 0.06599 | chr19 | 43561602 | 43564599 | promoter_flanking_region | NA |
| chr2 | 10452293 | 10452650 | 0.91955 | chr2 | 10452319 | 10452469 | exon | AC007249.3 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr2 | 10452293 | 10452650 | 0.91955 | chr2 | 10451319 | 10452318 | five_prime_flank | AC007249.3 |
| chr2 | 10452293 | 10452650 | 0.91955 | chr2 | 10452470 | 10453868 | intron | AC007249.3 |
| chr2 | 10452293 | 10452650 | 0.91955 | chr2 | 10451328 | 10452327 | three_prime_flank | RP11-320M2.1 |
| chr2 | 10452293 | 10452650 | 0.91955 | chr2 | 10452001 | 10452800 | enhancer | NA |
| chr2 | 19911810 | 19912027 | 0.51074 | chr2 | 19910260 | 19913708 | exon | WDR35 |
| chr2 | 19911810 | 19912027 | 0.51074 | chr2 | 19910260 | 19913557 | three_prime_utr | WDR35 |
| chr2 | 19911810 | 19912027 | 0.51074 | chr2 | 19911698 | 19912308 | tf_binding_site | NA |
| chr2 | 28584009 | 28584215 | 0.37252 | chr2 | 28582506 | 28585760 | intron | PLB1 |
| chr2 | 28584009 | 28584215 | 0.37252 | chr2 | 28582802 | 28585599 | promoter_flanking_region | NA |
| chr2 | 37776692 | 37776940 | 0.24589 | chr2 | 37775401 | 37777001 | enhancer | NA |
| chr2 | 38396148 | 38396347 | 0.79317 | chr2 | 38395869 | 38396531 | open_chromatin | NA |
| chr2 | 201798315 | 201798548 | 0.9522 | chr2 | 201790661 | 201800783 | intron | CDK15 |
| chr2 | 201798315 | 201798548 | 0.9522 | chr2 | 201790661 | 201806428 | intron | CDK15 |
| chr2 | 201798315 | 201798548 | 0.9522 | chr2 | 201798090 | 201799244 | promoter_flanking_region | NA |
| chr22 | 38313384 | 38313698 | 1.43467 | chr22 | 38303249 | 38314081 | intron | CSNK1E |
| chr3 | 31969003 | 31969283 | 0.45962 | chr3 | 31879831 | 32046490 | intron | OSBPL10 |
| chr3 | 31969003 | 31969283 | 0.45962 | chr3 | 31879831 | 31980898 | intron | OSBPL10 |
| chr3 | 31969003 | 31969283 | 0.45962 | chr3 | 31879831 | 31969402 | intron | OSBPL10 |
| chr3 | 31969003 | 31969283 | 0.45962 | chr3 | 31876513 | 31980898 | intron | OSBPL10 |
| chr3 | 31969003 | 31969283 | 0.45962 | chr3 | 31967202 | 31970399 | promoter_flanking_region | NA |
| chr3 | 36701759 | 36702010 | 0.19312 | chr3 | 36701735 | 36702414 | open_chromatin | NA |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr3 | 48595608 | 48595807 | 0.41569 | chr3 | 48595268 | 48596267 | five_prime_flank | COL7A1 |
| chr3 | 98922857 | 98923207 | 0.65577 | chr3 | 98904995 | 99018055 | intron | CTD-2021J15.1 |
| chr3 | 195105423 | 195105624 | 0.21739 | chr3 | 195070112 | 195122025 | intron | XXYLT1 |
| chr3 | 195105423 | 195105624 | 0.21739 | chr3 | 195070112 | 195153886 | intron | XXYLT1 |
| chr3 | 195105423 | 195105624 | 0.21739 | chr3 | 195070112 | 195156448 | intron | XXYLT1 |
| chr3 | 195105423 | 195105624 | 0.21739 | chr3 | 195105346 | 195105771 | tf_binding_site | NA |
| chr4 | 5038240 | 5038466 | 0.86428 | chr4 | 5036202 | 5038799 | promoter_flanking_region | NA |
| chr4 | 123832418 | 123832634 | 1.06512 | chr4 | 123829188 | 123863574 | intron | LINC01091 |
| chr4 | 123832418 | 123832634 | 1.06512 | chr4 | 123829188 | 123925183 | intron | LINC01091 |
| chr4 | 123832418 | 123832634 | 1.06512 | chr4 | 123830802 | 123834399 | promoter_flanking_region | NA |
| chr5 | 14268231 | 14268525 | 0.86046 | chr5 | 14143883 | 14270824 | intron | TRIO |
| chr5 | 14268231 | 14268525 | 0.86046 | chr5 | 14183960 | 14270824 | intron | TRIO |
| chr5 | 14268231 | 14268525 | 0.86046 | chr5 | 14263202 | 14270804 | promoter_flanking_region | NA |
| chr5 | 42985834 | 42986057 | 0.14794 | chr5 | 42984402 | 42986399 | promoter_flanking_region | NA |
| chr5 | 132073620 | 132073829 | 0.12371 | chr5 | 132072790 | 132073789 | five_prime_flank | CSF2 |
| chr5 | 132073620 | 132073829 | 0.12371 | chr5 | 132073790 | 132073823 | five_prime_utr | CSF2 |
| chr5 | 132073620 | 132073829 | 0.12371 | chr5 | 132073790 | 132073982 | exon | CSF2 |
| chr5 | 132073620 | 132073829 | 0.12371 | chr5 | 132073824 | 132073982 | cds | CSF2 |
| chr5 | 150477404 | 150477603 | 1.49474 | chr5 | 150475760 | 150485228 | intron | CTC-367J11.1 |
| chr5 | 150477404 | 150477603 | 1.49474 | chr5 | 150476801 | 150478199 | promoter_flanking_region | NA |
| chr5 | 173455362 | 173455603 | 0.68005 | chr5 | 173451202 | 173457799 | promoter_flanking_region | NA |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr5 | 173457001 | 173457215 | 0.40407 | chr5 | 173451202 | 173457799 | promoter_flanking_region | NA |
| chr6 | 17417814 | 17418019 | 1.77466 | chr6 | 17393747 | 17421554 | intron | CAP2 |
| chr6 | 17417814 | 17418019 | 1.77466 | chr6 | 17417202 | 17419599 | promoter_flanking_region | NA |
| chr6 | 33840183 | 33840460 | 1.10424 | chr6 | 33839402 | 33841076 | promoter_flanking_region | NA |
| chr6 | 42879722 | 42879977 | 0.38534 | chr6 | 42879618 | 42879951 | exon | RPL7L1 |
| chr6 | 42879722 | 42879977 | 0.38534 | chr6 | 42879635 | 42879937 | five_prime_utr | RPL7L1 |
| chr6 | 42879722 | 42879977 | 0.38534 | chr6 | 42879682 | 42879937 | five_prime_utr | RPL7L1 |
| chr6 | 42879722 | 42879977 | 0.38534 | chr6 | 42879682 | 42879951 | exon | RPL7L1 |
| chr6 | 42879722 | 42879977 | 0.38534 | chr6 | 42879635 | 42879951 | exon | RPL7L1 |
| chr6 | 42879722 | 42879977 | 0.38534 | chr6 | 42879639 | 42879951 | exon | RPL7L1 |
| chr6 | 42879722 | 42879977 | 0.38534 | chr6 | 42879929 | 42879951 | exon | RPL7L1 |
| chr6 | 42879722 | 42879977 | 0.38534 | chr6 | 42879933 | 42879937 | five_prime_utr | RPL7L1 |
| chr6 | 42879722 | 42879977 | 0.38534 | chr6 | 42879920 | 42879951 | exon | RPL7L1 |
| chr6 | 42879722 | 42879977 | 0.38534 | chr6 | 42879924 | 42879951 | exon | RPL7L1 |
| chr6 | 42879722 | 42879977 | 0.38534 | chr6 | 42879933 | 42879951 | exon | RPL7L1 |
| chr6 | 42879722 | 42879977 | 0.38534 | chr6 | 42879938 | 42879951 | cds | RPL7L1 |
| chr6 | 42879722 | 42879977 | 0.38534 | chr6 | 42879952 | 42880557 | intron | RPL7L1 |
| chr6 | 42879722 | 42879977 | 0.38534 | chr6 | 42879952 | 42880860 | intron | RPL7L1 |
| chr6 | 42879722 | 42879977 | 0.38534 | chr6 | 42879952 | 42883318 | intron | RPL7L1 |
| chr6 | 42879722 | 42879977 | 0.38534 | chr6 | 42879952 | 42883450 | intron | RPL7L1 |
| chr6 | 42879722 | 42879977 | 0.38534 | chr6 | 42879200 | 42881001 | promoter | NA |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr6 | 44043158 | 44043368 | 0.49055 | chr6 | 44001117 | 44074496 | intron | RP5-1120P11.1 |
| chr6 | 44043158 | 44043368 | 0.49055 | chr6 | 44041202 | 44044807 | promoter_flanking_region | NA |
| chr6 | 47135752 | 47135954 | 0.73707 | chr6 | 47134002 | 47136458 | promoter_flanking_region | NA |
| chr6 | 63635791 | 63636096 | 0.32917 | chr6 | 63635823 | 63636150 | exon,five_prime_utr | PHF3 |
| chr6 | 63635791 | 63636096 | 0.32917 | chr6 | 63634820 | 63635819 | five_prime_flank | PHF3 |
| chr6 | 63635791 | 63636096 | 0.32917 | chr6 | 63635820 | 63636150 | exon,five_prime_utr | PHF3 |
| chr6 | 63635791 | 63636096 | 0.32917 | chr6 | 63635825 | 63636150 | exon,five_prime_utr | PHF3 |
| chr6 | 63635791 | 63636096 | 0.32917 | chr6 | 63635836 | 63636150 | exon,five_prime_utr | PHF3 |
| chr6 | 63635791 | 63636096 | 0.32917 | chr6 | 63635400 | 63637801 | promoter | NA |
| chr6 | 71248478 | 71248706 | 1.66823 | chr6 | 71233658 | 71284859 | intron | RP11-154D6.1 |
| chr6 | 71248478 | 71248706 | 1.66823 | chr6 | 71247802 | 71251799 | promoter_flanking_region | NA |
| chr6 | 73803547 | 73803780 | 1.0143 | chr6 | 73803302 | 73803965 | intron | CD109 |
| chr6 | 73803547 | 73803780 | 1.0143 | chr6 | 73803302 | 73806843 | intron | CD109 |
| chr6 | 73803547 | 73803780 | 1.0143 | chr6 | 73803200 | 73804201 | enhancer | NA |
| chr6 | 144415880 | 144416268 | 0.11308 | chr6 | 144403185 | 144421877 | intron | UTRN |
| chr6 | 144415880 | 144416268 | 0.11308 | chr6 | 144415601 | 144416599 | promoter_flanking_region | NA |
| chr7 | 2255729 | 2255970 | 0.47301 | chr7 | 2255170 | 2256873 | intron | SNX8 |
| chr7 | 2255729 | 2255970 | 0.47301 | chr7 | 2255819 | 2256190 | open_chromatin | NA |
| chr8 | 8288569 | 8288827 | 0.46645 | chr8 | 8286202 | 8289918 | promoter_flanking_region | NA |
| chr8 | 26451397 | 26451614 | 0.19394 | chr8 | 26408377 | 26505306 | intron | BNIP3L |
| chr8 | 26451397 | 26451614 | 0.19394 | chr8 | 26447801 | 26452278 | promoter_flanking_region | NA |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr8 | 41221552 | 41221766 | 0.16648 | chr8 | 41221391 | 41222128 | open_chromatin | NA |
| chr8 | 82257863 | 82258095 | 0.4962 | chr8 | 82257704 | 82258628 | open_chromatin | NA |
| chr8 | 117811330 | 117811531 | 0.08973 | chr8 | 117807378 | 117812871 | intron | EXT1 |
| chr8 | 117811330 | 117811531 | 0.08973 | chr8 | 117810955 | 117811589 | open_chromatin | NA |
| chr8 | 125513058 | 125513395 | 0.69125 | chr8 | 125473315 | 125540909 | intron | RP11-136O12.2 |
| chr8 | 125513058 | 125513395 | 0.69125 | chr8 | 125511802 | 125514199 | promoter_flanking_region | NA |
| chr8 | 127898887 | 127899098 | 0.3132 | chr8 | 127796009 | 127939507 | intron | PVT1 |
| chr8 | 127898887 | 127899098 | 0.3132 | chr8 | 127890999 | 127932464 | intron | PVT1 |
| chr8 | 127898887 | 127899098 | 0.3132 | chr8 | 127890999 | 127989161 | intron | PVT1 |
| chr8 | 127898887 | 127899098 | 0.3132 | chr8 | 127890999 | 127939507 | intron | PVT1 |
| chr8 | 127898887 | 127899098 | 0.3132 | chr8 | 127898202 | 127901599 | promoter_flanking_region | NA |
| chr8 | 133217883 | 133218125 | 0.10502 | chr8 | 133213144 | 133220580 | intron | WISP1 |
| chr8 | 133217883 | 133218125 | 0.10502 | chr8 | 133213144 | 133225389 | intron | WISP1 |
| chr8 | 133217883 | 133218125 | 0.10502 | chr8 | 133213144 | 133227410 | intron | WISP1 |
| chr8 | 133217883 | 133218125 | 0.10502 | chr8 | 133191214 | 133225389 | intron | WISP1 |
| chr8 | 133217883 | 133218125 | 0.10502 | chr8 | 133191214 | 133227410 | intron | WISP1 |
| chr8 | 133217883 | 133218125 | 0.10502 | chr8 | 133216202 | 133218599 | promoter_flanking_region | NA |
| chr8 | 143940293 | 143940582 | 1.96392 | chr8 | 143938693 | 143946384 | intron | PLEC |
| chr8 | 143940293 | 143940582 | 1.96392 | chr8 | 143938693 | 143958594 | intron | PLEC |
| chr8 | 143940293 | 143940582 | 1.96392 | chr8 | 143938693 | 143973402 | intron | PLEC |
| chr8 | 143940293 | 143940582 | 1.96392 | chr8 | 143938693 | 143975176 | intron | PLEC |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| chr8 | 143940293 | 143940582 | 1.96392 | chr8 | 143938693 | 143950183 | intron | PLEC |
| chr8 | 143940293 | 143940582 | 1.96392 | chr8 | 143938693 | 143942391 | intron | PLEC |
| chr8 | 143940293 | 143940582 | 1.96392 | chr8 | 143938693 | 143943778 | intron | PLEC |
| chr8 | 143940293 | 143940582 | 1.96392 | chr8 | 143938693 | 143944647 | intron | PLEC |
| chr8 | 143940293 | 143940582 | 1.96392 | chr8 | 143938693 | 143946349 | intron | PLEC |
| chr8 | 143940293 | 143940582 | 1.96392 | chr8 | 143938693 | 143953725 | intron | PLEC |
| chr1 | 156505014 | 156505238 | 0.30513 | NA | NA | NA | NA | NA |
| chr5 | 143246065 | 143246322 | 0.4611 | NA | NA | NA | NA | NA |
| chr6 | 1481425 | 1481651 | 0.09676 | NA | NA | NA | NA | NA |