

Investigating molecular vulnerabilities in *TP53*-mutant transformed myeloproliferative neoplasms



Aimee Paterson
St Edmund Hall
University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Trinity 2024

Acknowledgements

Firstly, I would like to thank my supervisor Professor Adam Mead for his support, patience and guidance throughout this project. Thank you for fostering an enthusiastic laboratory environment and giving me this opportunity to advance my scientific knowledge. Thank you also to Professor Anindita Roy, for her scientific contributions throughout my DPhil.

This project wouldn't have been possible without Alba Rodriguez-Meira, whose impressive work this project follows. Thank you for your guidance and kindness during my first year in Oxford.

I am grateful to Shady Awad for his collaboration during the later stages of this project. Your input and team spirit were greatly valued, I wish you every success. I also thank Abdullah Khan for his support and generation of organoids for this project.

A number of core facility members at the WIMM deserve recognition for their input to this project. Thank you, Ryan Beveridge, for teaching me everything needed for virus production, Christian Babbs and Zoe Holloway for their genome engineering endeavours and finally, Kevin Clark and Sally-Anne Clarke, for their support with flow cytometry experiments.

I would like to thank every member of the HSCB laboratory for their friendship and unwavering support. You made every day in the lab better and I couldn't have spent 4 years with greater colleagues. I wish you all the brightest futures. A special thank you to Lauren, whose PhD timeline and thesis submission coincided with my own, your solidarity and friendship helped immensely.

To Mack, thank you for your *truly* unshakable patience and for making Oxford feel like home. To Gran and Grandad, thank you for your encouragement. I can always count on you to make my day brighter. To Grandma and Grandpa, I wish I could share this achievement with you, I know you will be proud.

To my parents, thank you for supporting my every step. You never doubted me - although never did we think I would make it here! I couldn't have done this without you. This thesis is dedicated to you both.

Declarations

I declare this thesis is entirely my own work, unless stated otherwise. Experiments and analysis were designed and performed by myself, and where other researchers contributed, their contributions are clearly acknowledged in the relevant sections:

1. Generation and bioinformatic analysis of the TARGET-seq dataset described in Chapter 3 was performed by Dr. Alba Rodriguez-Meira
2. *Ex vivo* murine HSPC CRISPR screening data was shared by Professor Adam Wilkinson for generation of Figure 4.2A in Chapter 4
3. All human bone marrow organoids described in Chapter 5 (Sections 5.3.3 and 5.3.4) were generated by Dr Abdullah Khan
4. Pre-processing of RNA-seq libraries for differential gene expression analysis in Chapter 6 (Section 6.3.1) was performed by Antonio Rodriguez-Romera
5. Dr Sean Wen analysed bulk RNA-seq data for splicing events using the MARVEL pipeline in Chapter 6 (Section 6.3.1) and generated Figures 6.9, 6.10 and 6.12
6. Dr Shady Awad carried out BCL-XL experiments described in Chapter 6 (Section 6.3.2) and generated Figure 6.17A
7. R-loop interrogation in Chapter 6 (Section 6.3.3) was complete in collaboration with Dr Chiara Beghè in the laboratory of Professor Natalie Gromak at the Sir William Dunn School of Pathology, University of Oxford
8. Preparation and imaging of γ H2AX-stained cells in Chapter 6 (Section 6.3.3) was complete with the assistance of Antonio Rodriguez-Romera

Abstract

Myeloproliferative neoplasms (MPN) are chronic haematological disorders characterised by activating mutations within the JAK-STAT signalling pathway. MPN are a heterogeneous spectrum of stem cell-derived neoplasm defined by excess clonal proliferation of mature myeloid entities. Clinical complications include thrombohemorrhagic events, extramedullary haematopoiesis and leukaemic transformation. Patients who undergo leukaemic transformation have poor survival outcomes, with no effective therapeutic options. Patients with *TP53*-mutations, termed *TP53*-mutant secondary acute myeloid leukaemic (*TP53*-sAML) have an increasingly poor prognosis. Thus, *TP53*-sAML remains an unmet clinical need and novel therapeutic strategies are necessary to improve patient outcomes. In this thesis, I therefore investigated molecular vulnerabilities in transformed *TP53*-mutant MPN with the potential for therapeutic targeting.

This thesis investigates the molecular vulnerabilities of *TP53*-sAML via analysis of a single-cell multi-omic TARGET-seq dataset, including the intra-patient comparison between *TP53*-wildtype (WT) haematopoietic stem and progenitor cells (HSPC) and *TP53*-mutant leukaemic stem cells (LSC). In addition to this intra-patient comparison, analysis included comparisons between *TP53*-mutant LSC and healthy donor controls and *TP53*-WT MPN patients. In all comparisons, I identified a cohort of RNA-binding proteins (RBP) to be significantly upregulated in *TP53*-mutant LSC. Furthermore, all upregulated RBP were identified to be essential for cancer cell line survival via *in silico* analysis and validated by candidate RBP-targeted gene perturbation screens.

From this RBP cohort, I identified PCBP2 as a candidate RBP with the potential for therapeutic targeting. Evaluation of each candidate genes therapeutic index was based on extensive literature review and access to an *ex vivo* murine HSPC gene perturbation screen. Importantly, I showed *PCBP2* depletion in *TP53*-WT healthy donor HSPC had limited effect on HSPC proliferation, differentiation or clonogenic-capacity *in vitro*.

Conventional *in vitro* liquid culture techniques were found to be insufficient to sustain survival and proliferation of primary *TP53*-sAML cells for meaningful periods of time.

Attempting to perturb gene expression added further complexity to primary cell cultures. Therefore, to investigate the requirement of PCBP2 in primary samples, I adopted a human bone marrow organoid engraftment model. *Ex vivo* culturing of primary *TP53*-sAML cells greatly improved cell survival and proliferation, generating a tractable system for validating *PCBP2* gene perturbation *ex vivo*.

Mechanistically, PCBP2 is a multifunctional RBP involved in numerous aspects of post-transcriptional gene regulation. To investigate the mechanisms by which PCBP2 is required to support *TP53*-sAML, RNA-seq data from a cancer cell line model was analysed following *PCBP2* depletion for differentially expressed genes and alternative splicing events. Splice analysis showed the extent of PCBP2 regulation with a number of isoforms switching events relevant to myeloid malignancies. Further mechanistic investigation, including identification of PCBP2-binding transcripts and protein-binding partners will help to validate molecular requirements of PCBP2 in *TP53*-sAML.

In summary, I identified a cohort of essential RBP significantly upregulated and enriched in *TP53*-sAML through analysis of a single-cell multi-omic TARGET-seq dataset. PCBP2 was selected as a candidate RBP of particular interest, due to evidence suggesting PCBP2 is non-essential in *TP53*-WT HSPC. To validate the selective requirement of PCBP2 in *TP53*-mutant LSC, I adopted a human bone marrow organoid model which I showed to outperform liquid culture conditions in supporting the survival and proliferation of *TP53*-sAML samples. This system has not previously been shown to support the proliferation and survival of primary AML, highlighting the importance of this work.

Contents

<i>List of Figures</i>	<i>xiii</i>
<i>List of Tables</i>	<i>xvi</i>
<i>Abbreviations</i>	<i>xviii</i>
Chapter 1 Introduction	1
1.1 The haematopoietic system	1
1.2 Myeloproliferative neoplasms	3
1.2.1 Classification of myeloproliferative neoplasms.....	3
1.2.2 Philadelphia-chromosome negative myeloproliferative neoplasms	3
1.2.3 Cell-of-origin: myeloproliferative neoplasms as a stem cell driven disease	4
1.2.4 Myeloproliferative neoplasms: A tractable model for clonal evolution in cancer	5
1.2.5 Disease initiating mutations in myeloproliferative neoplasms	5
1.2.6 Additional mutations contributing to myeloproliferative disease	8
1.3 Blast-phase myeloproliferative neoplasms	9
1.3.1 Incidence of transformation and associated risk factors	9
1.3.2 Genetic heterogeneity in BP-MPN	11
1.3.3 Single-cell technologies to unravel intratumoural heterogeneity	12
1.3.4 Therapeutic treatment of BP-MPN.....	15
1.4 TP53-mutant secondary acute myeloid leukaemia (TP53-sAML)	17
1.4.1 p53, guardian of the genome	17
1.4.2 p53 in normal and malignant haematopoiesis	21
1.4.3 TP53-mutant leukaemic evolution from chronic-phase MPN	22
1.5 Summary and aims	26
Chapter 2 Methods	27
2.1 Cell line assays	27
2.1.1 Cell line maintenance.....	27

2.1.2 Lentiviral production	28
2.1.3 Lentiviral transduction.....	31
2.1.4 Quantitative real time polymerase chain reaction	32
2.1.5 Western blotting	34
2.1.6 RNA-binding Protein Targeted CRISPR screening.....	36
2.1.7 CRISPR-mediated gene knockout in cell lines.....	42
2.1.8 Annexin V/DAPI apoptosis assay.....	43
2.1.9 Competitive cell growth assay	44
2.1.10 Bulk cell editing assay.....	45
2.1.11 Evaluation of dsDNA: RNA hybrids via S9.6 slot blot	46
2.1.12 γ H2AX immunofluorescence for double-stranded breaks	49
2.1.13 RNA-seq.....	50
2.2 Primary patient assays	51
2.2.1 Processing of human samples	51
2.2.2 Fluorescent activated cell analysis and sorting.....	52
2.2.3 Lentiviral transduction.....	53
2.2.4 Quantitative real time polymerase chain reaction	54
2.2.5 Multilineage differentiation assay.....	55
2.2.6 Colony forming unit (CFU) assay	57
2.2.7 CRISPR-mediated gene knockout in primary cells.....	58
2.2.8 Single-cell colony genotyping assay	59
2.2.9 Primary patient cell engraftment in human bone marrow organoids	60
2.3 Bioinformatic analysis.....	64
2.3.1 TARGET-seq analysis	64
2.3.2 DepMap analysis	64
2.3.3 Gene ontology.....	65
2.3.4 Analysis of external AML datasets.....	65
2.3.5 Identification of essential genes (CRISPR screening)	66
2.3.6 Bulk RNA-Seq analysis.....	67
2.3.7 Bulk-level splicing analysis.....	68
2.3.8 Statistical analysis	70
Chapter 3 RNA-binding proteins are overexpressed in TP53-sAML.....	71

3.1 Introduction	71
3.1.1 An introduction to RNA-binding proteins.....	71
3.1.2 RBP in normal and malignant haematopoiesis.....	74
3.1.3 RBP in <i>TP53</i> -mutant myeloid malignancies.....	78
3.1.4 Hypothesis.....	78
3.1.5 Summary and aims	78
3.2 Results.....	79
3.2.1 Patient cohort.....	79
3.2.2 Identifying overexpressed genes in <i>TP53</i> -mutant LSC.....	80
3.2.3 Essential genes	81
3.2.4 Validating RNA-binding protein dependencies in <i>TP53</i> -sAML	83
3.3 Discussion	86
<i>Chapter 4 PCBP2: a novel molecular vulnerability in TP53-sAML</i>	91
4.1 Introduction	91
4.1.1 An introduction to PCBP2	91
4.1.2 The role of PCBP2 in cancer.....	94
4.1.3 PCBP2 in normal and malignant haematopoiesis	95
4.1.4 Hypothesis.....	98
4.1.5 Summary and aims	98
4.2 Results.....	99
4.2.1 Investigating the requirement of candidate RBP in <i>TP53</i> -WT HSPC.....	99
4.2.2 Evaluating <i>PCBP2</i> expression and effect in publicly available AML cohorts..	101
4.2.3 Investigating the expression and dependency of <i>PCBP2</i> across cancer cell line models	105
4.2.4 PCBP2 presents a functional requirement in <i>JAK2/TP53</i> -mutant cell line models	108
4.3 Discussion	112
<i>Chapter 5 Establishing functional assays to assess the role of PCBP2 in TP53-sAML</i>	120
5.1 Introduction	120
5.1.1 Unknown functional role of PCBP2 in primary cells.....	120

5.1.2 Available models to study the primary cell requirement of PCBP2.....	120
5.1.3 Hypothesis.....	125
5.1.4 Summary and aims	125
5.2 Results.....	126
5.2.1 PCBP2 is indispensable for <i>TP53</i> -WT HSPC proliferation and differentiation	126
5.2.2 Conventional liquid culture conditions are insufficient to assess <i>PCBP2</i> depletion in primary <i>TP53</i> -sAML	131
5.2.3 Engraftment of primary <i>TP53</i> -sAML samples into a human bone marrow organoid model enhances cell survival and proliferation.....	135
5.2.4 A human bone marrow organoid model facilitates functional assessment of <i>PCBP2</i> knockdown in primary cells	141
5.3 Discussion	147
<i>Chapter 6 Investigating PCBP2 mechanisms of action in TP53-sAML</i>	<i>154</i>
6.1 Introduction	154
6.1.1 PCBP2 and regulation of gene transcription	154
6.1.2 PCBP2 and post transcriptional gene regulation.....	154
6.1.3 PCBP2 and protein-protein interactions	158
6.1.4 Hypothesis.....	160
6.1.5 Summary and aims	160
6.2 Results.....	162
6.2.1 Investigating the impact of PCBP2 on transcript stability and splicing in <i>TP53</i> -sAML.....	162
6.2.2 Investigating the requirement of PCBP2 in <i>BCL2L1</i> regulation.....	177
6.2.3 Investigation the potential role of PCBP2 in R-loop biology.....	178
6.3 Ongoing mechanistic work.....	182
6.4 Discussion	184
<i>Chapter 7 Conclusions and future work.....</i>	<i>187</i>
7.1 Conclusions	187

7.1.1 A cohort of RBP, essential for cancer cell survival, are enriched in <i>TP53</i> -sAML	187
7.1.2 PCBP2 is a molecular vulnerability with potential for therapeutic targeting in <i>TP53</i> -sAML.....	187
7.1.3 Human bone marrow organoids provide a tractable system for investigating gene perturbations in <i>TP53</i> -sAML patient samples	188
7.1.4 PCBP2 is a multifunctional RBP with various roles in post-transcriptional gene regulation	188
7.2 Wider relevance of this work	189
7.3 Limitations of this work.....	189
7.4 Future work.....	190
7.4.1 What is the role of other candidate RBP in <i>TP53</i> -sAML?	190
7.4.2 Is PCBP2 required for <i>TP53</i> -mutant leukaemic transformation?	190
7.4.3 Is PCBP2 essential for <i>TP53</i> -sAML survival?	191
7.4.4 What are the molecular mechanisms through which PCBP2 supports <i>TP53</i> -sAML survival and proliferation?	191
7.5 Final summary.....	192
<i>Appendix</i>	193
<i>References</i>	203

List of Figures

- 1.1. An updated haematopoietic hierarchy
- 1.2. Resolving intra-patient tumour heterogeneity using TARGET-seq
- 1.3. Genetic and non-genetic determinants of *TP53*-mutant leukaemic transformation in MPN
- 2.1. CRISPR-Cas9 structure and pathways for genome engineering
- 2.2. Principles of apoptosis Annexin V/DAPI assays
- 2.3. Flow cytometry gating strategy for Annexin V/DAPI apoptosis assay
- 2.4. Flow cytometry gating strategy for competitive cell growth assays
- 2.5. Flow cytometry gating strategy for multilineage differentiation assays
- 2.6. Representative CFU colonies
- 2.7. Flow cytometry gating strategy for assessment of transduced primary cell populations following organoid engraftment
- 3.1. Processes involved in mRNA metabolism
- 3.2. Patient cohort for TARGET-seq analysis
- 3.3. Venn diagram of commonly upregulated genes in *TP53*-mutant LSC
- 3.4. Investigating the *JAK2/TP53*-mutant cancer cell requirement of overexpressed genes
- 3.5. Schematic of CRISPR-Cas9 screening in *JAK2/TP53*-mutant AML model cell lines
- 3.6. RBP-targeted CRISPR-Cas9 screening
- 4.1. PCBP2 expression in normal tissues and mRNA/protein architecture
- 4.2. Requirement and expression of PCBP2 in WT-HSPC
- 4.3. *PCBP2* expression in the BeatAML patient cohort stratified by *TP53*-mutation status
- 4.4. Survival probability by *PCBP2* expression in a *TP53*-stratified secondary AML cohort from the BeatAML dataset
- 4.5. *PCBP2* expression and impact on survival in the TCGA-LAML cohort stratified by *TP53*-mutation status
- 4.6. PCBP2 expression across cancer types
- 4.7. Interrogating expression and requirement of *PCBP2* in cancer cell lines using the DepMap database

- 4.8. *PCBP2* depletion in cancer cell line models by lentiviral shRNA
- 4.9. Effect of *PCBP2* depletion on model cell line survival
- 4.10. Competitive disadvantage of *PCBP2* depleted cell lines
- 4.11. *PCBP2*-knockout cells are lost from culture overtime
- 5.1. The bone marrow microenvironment
- 5.2. Workflow for primary HSPC transduction
- 5.3. *PCBP2* depletion in *TP53*-WT HSPC via lentiviral shRNA
- 5.4. Effect of *PCBP2* depletion on cord blood HSPC
- 5.5. Effect of *PCBP2*-KO on HD HSPC single-cell colony formation
- 5.6. Transducing HD CD34+ HSPC with all-in-one Cas9-*PCBP2*-sgRNA lentiviral vector
- 5.7. *PCBP2* depletion in *TP53*-sAML patient samples via lentiviral shRNA
- 5.8. *TP53*-sAML cells fail to expand in multilineage differentiation cultures or generate colonies in methylcellulose-based CFU assays
- 5.9. Workflow for primary cell engraftment into human bone marrow derived organoids
- 5.10. Human bone marrow organoids engrafted with CellTrace-stained *TP53*-sAML samples or HD controls compared to liquid culture conditions
- 5.11. The effect of human bone marrow organoid engraftment on primary cell proliferation
- 5.12. The effect of human bone marrow organoid engraftment on primary cell viability
- 5.13. Workflow for human organoid engraftment with *PCBP2*-depleted primary cells
- 5.14. Lentiviral transduction of *TP53*-sAML samples with *PCBP2*-targeting shRNA
- 5.15. Transduced LSC from *TP53*-sAML patients are engraftable and tractable in human bone marrow organoids
- 5.16. Effect of *PCBP2* depletion on viability and proliferation of LSC from *TP53*-sAML samples following 7-days organoid engraftment
- 6.1. Major steps in the U2 spliceosome pathway
- 6.2. Mechanistic roles of *PCBP2*
- 6.3. Comparing *PCBP2* knockdown in HEL cells at 72-hours and 96-hours post-transduction
- 6.4. Principle component analysis (PCA) and *PCBP2* counts
- 6.5. Identifying differentially expressed genes using DESeq2
- 6.6. Gene set enrichment analysis (GSEA) from RNA-seq data

- 6.7. Investigating the requirement of PCBP2 in regulation of interferon stimulated genes
- 6.8. Investigating the essentiality of differentially expressed genes in *JAK2/TP52*-mutant cell line models
- 6.9. MARVEL exon-level splice event types
- 6.10. Identifying alternative splice events (ASE) in *PCBP2*-depleted cells
- 6.11. Visualisation of *PCBP2* transcripts alternative last exon (ALE) events
- 6.12. Investigating gene expressing-splicing dynamics
- 6.13. Visualisation of *RUNX1* skipped exon and *STAT3* alternative 3' splice site events
- 6.14. Investigating the essentiality of alternatively spliced genes in *JAK2/TP53*-mutant cancer cell line models
- 6.15. Dependency of *JAK2/TP53*-mutant erythroleukemia HEL cells on anti-apoptotic proteins following *PCBP2*-KD
- 6.16. Investigating R-loop formation following PCBP2 depletion via slot blot using dsRNA: DNA specific S9.6 antibody
- 6.17. Detecting double-strand breaks via γ H2AX staining in *PCBP2*-KD HEL cells.
- A.3. Cas9-sgRNA-GFP all-in-one vector
- A.4. shRNA-RFP lentiviral vector

List of Tables

- 1.1. Patient, disease and treatment-derived risk factors of leukaemic transformation in MPN
- 2.1. *JAK2* and *TP53* mutation status in HEL and SET-2 cell lines
- 2.2. shRNA lentiviral vectors
- 2.3. Master mix for cell line cDNA synthesis
- 2.4. Thermocycler conditions for cell line cDNA synthesis
- 2.5. Master mix for qPCR
- 2.6. Thermocycler conditions for qPCR
- 2.7. Primary and secondary antibodies
- 2.8. Target genes and sgRNA sequences for RBP-targeted CRISPR screen
- 2.9. PCR1 primer sequences for RBP-targeted CRISPR screening libraries
- 2.10. PCR1 reaction mix for CRISPR screening library preparation
- 2.11. PCR1 thermocycler conditions
- 2.12. PCR2 primer sequences for CRISPR screening libraries
- 2.13. PCR2 reaction mix for CRISPR screening library preparation
- 2.14. PCR2 thermocycler conditions
- 2.15. *PCBP2* and *AAVS1* sgRNA sequences for RNP nucleofection
- 2.16. PCR primers for amplification of sgRNA edited loci
- 2.17. Reaction mix for PCR amplification of edited loci
- 2.18. Thermocycler conditions for *PCBP2* and *AAVS1* amplification
- 2.19. Cell lysis buffer for S9.6 slot blot
- 2.20. Nuclear lysis buffer for S9.6 slot blot
- 2.21. Master mix for primary gene-specific PCR amplification and cDNA synthesis
- 2.22. 20X TaqMan probe mix and probe ID
- 2.23. Thermocycler conditions for gene-specific PCR-amplification and cDNA synthesis
- 2.24. Liquid culture multilineage differentiation media (1X)
- 2.25. Antibodies for determination of multilineage differentiation
- 2.26. Media conditions for single-cell colony assays
- 2.27. Buffer for single-cell colony lysis
- 2.28. Master mix for single cell colony genotyping

- 2.29. Thermocycler conditions for single-cell colony genotyping
- 2.30. AML-specific media for organoid engraftment
- 4.1. Summary of candidate RNA-binding proteins
- A.1. Primary patient cohort
- A.2. List of all 96 commonly upregulated genes in *TP53*-mutant LSC
- A.5. DESeq2 results for 76 commonly differentially expressed genes
- A.6. Top 50 alternatively splice events following *PCBP2*-KD in HEL cells
- A.7. Complete list of differential gene expression-splice relationships

Abbreviations

ALE	Alternative last exon
AFE	Alternative first exon
AML	Acute myeloid leukaemia
ASE	Alternative splice event
A3SS	Alternative 3' splice site
A5SS	Alternative 5' splice site
BP	Blast-phase
CALR	Calreticulin
CFU	Colony forming unit
CML	Chronic myeloid leukaemia
CNA	Copy number alterations
CNN-LOH	Copy number neutral loss-of-heterozygosity
CP	Chronic-phase
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DAPI	4', 6-diamidino-2-phenylidole
DDR	DNA-damage response
DEG	Differentially expressed gene
DSB	Double-strand break
ET	Essential thrombocytopenia
GFP	Green fluorescent protein
GOF	Gain-of-function
GSEA	Gene set enrichment analysis
HD	Healthy donor
HDR	Homology-directed repair
HMA	Hypomethylating agents
hnRNP	Heterogeneous nuclear ribonucleoproteins
HSC	Haematopoietic stem cell
HSPC	Haematopoietic stem and progenitor cell
ISG	Interferon-stimulated genes
JAK2	Janus Kinase 2
KD	Knockdown
KO	Knockout
MOI	Multiplicity of infection

MPL	Myeloproliferative leukaemia protein (thrombopoietin receptor)
MPN	Myeloproliferative neoplasms
MXE	Mutually exclusive exon
NHEJ	Non-homologous end joining
NMD	Nonsense-mediated decay
OS	Overall survival
PBS	Phosphate buffer saline
PCBP2	Poly(rC)-binding protein 2
PCA	Principal component analysis
PDX	Patient-derived xenograft
PMF	Primary myelofibrosis
PSI	Percentage spliced-in
PTGR	Post-transcriptional gene regulation
PTM	Post-transcriptional modification
PPT	Polypyrimidine tract
PV	Polycythaemia vera
RBD	RNA-binding domain
RBP	RNA binding protein
RFP	Red fluorescent protein
RI	Retained intron
RNP	Ribonucleoprotein
rpm	Rotations per minute
RT	Room temperature
sAML	Secondary acute myeloid leukaemia
SE	Skipped exon
shRNA	short hairpin RNA
sgRNA	single guide ribonucleic acid
snRNP	Small nuclear ribonuclear proteins
STAT	Signal transducer and activator of transcription
TCGA	The Cancer Genome Atlas
TU	Transducing units
TP53	Tumour protein p53
TRIBE	Targets of RNA-binding proteins identified by editing
WT	Wild-type
UTR	Untranslated region
7AAD	7-Aminoactinomycin

Chapter 1 Introduction

1.1 The haematopoietic system

Haematopoiesis is a tightly regulated and hierarchical system essential for the production and maintenance of all mature blood cell lineages. The haematopoietic system is established and maintained by haematopoietic stem cells (HSC) at the apex of the haematopoietic hierarchy. HSC reside in the bone marrow and are a long-lived cell population with the capacity to continually replenish the haematopoietic system of all shorter-lived blood cell lineages, through a fine balance of self-renewal and differentiation¹.

The classical model of the haematopoiesis proposes a small number of multipotent HSC differentiate into mature blood cell populations via a step-wise series of discrete intermediate progenitor populations. In this discrete differentiation landscape, each intermediate progenitor possesses increasingly restricted differentiation capacity and loss of self-renewal with obvious boundaries between cell populations¹. However, this classical hierarchical model has been superseded in recent years through the application of single-cell approaches, and it is now appreciated haematopoietic differentiation follows a hierarchical continuum with gradual transitions between cellular nodes².

Interrogating the haematopoietic hierarchy at the single-cell resolution has also shown substantial heterogeneity exists in the HSPC compartment, previously masked in population-based studies. Key alterations to the classical understanding of haematopoietic differentiation include: (1) megakaryocytes emerging directly from HSC and bypassing intermediate progenitors, (2) the first bifurcation point during haematopoietic differentiation separates megakaryocyte, erythroid, eosinophil, basophil, mast (EBM) progenitors from other myeloid and lymphoid progenitors and (3) megakaryocyte and erythroid cells sharing close differentiation trajectories through late stages of haematopoietic development² (**Figure 1.1**).

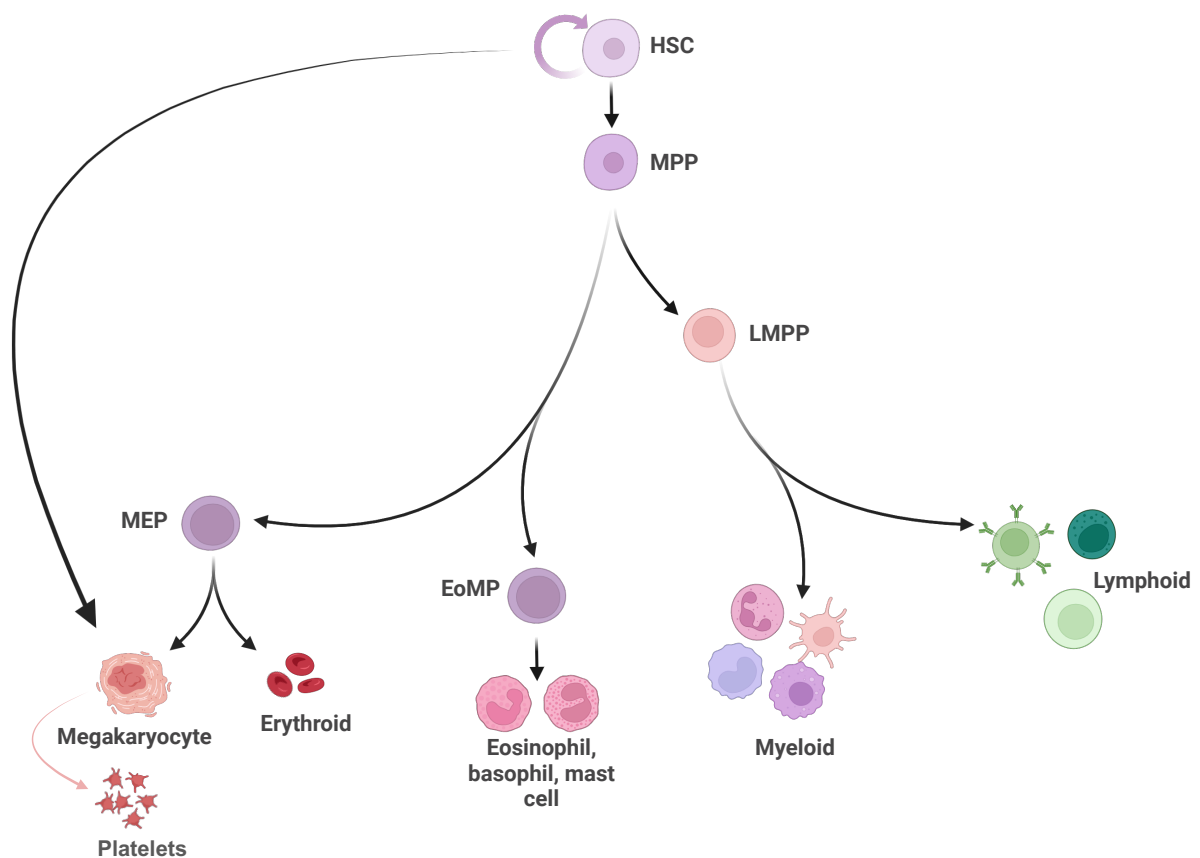


Figure 1.1. An updated haematopoietic hierarchy. HSC differentiate by gradually transitioning between cellular nodes, losing multipotency and becoming increasingly lineage primed. Lineage-primed cells may exist in HSC/MPP populations and megakaryocytes can emerge directly from HSC. LMPP, lymphoid-primed multipotent progenitor; MEP, megakaryocyte-erythroid progenitor; EoMP, eosinophil-basophil-mast cell progenitor. Adapted from Psaila and Mead, 2019³⁷⁴ and created in BioRender.com.

1.2 Myeloproliferative neoplasms (MPN)

1.2.1 Classification of myeloproliferative neoplasms

The World Health Organisation (WHO) classifies myeloproliferative neoplasms (MPN) into two main categories based on the presence or absence of the *BCR-ABL1* translocation³. The *BCR-ABL1* translocation is prerequisite for the development and diagnosis of chronic myeloid leukaemia (CML)⁴. This diagnostic requirement means CML can be regarded as a distinct clinical disease entity while the *BCR-ABL1* negative MPN can be categorised based on a spectrum of clinical symptoms, molecular characteristics and genetic signatures. This thesis focuses on *BCL-ABL1* negative MPN and these are described in detail below.

1.2.2 Philadelphia-chromosome negative myeloproliferative neoplasms

Philadelphia-chromosome negative MPN are a heterogenous spectrum of stem-cell derived neoplasms defined by excessive clonal proliferation of mature myeloid entities comprising polycythaemia vera (PV), essential thrombocytosis (ET) and primary myelofibrosis (MF)³.

Clinical presentation of MPN varies depending on the lineage effected by aberrant proliferation. Overproduction of red blood cells, erythrocytosis, is characteristic of PV while thrombocytosis, the excess production of platelets, is the main clinical manifestation of ET. MF is a disease defined by fibrotic scarring of the bone marrow and 80-90% of MF patients present with splenomegaly due to extramedullary haematopoiesis, which compensates for the defective bone marrow microenvironment⁵. Patients diagnosed with MPN will typically have systemic symptoms such as weight loss, night sweats and fatigue.

MPN is a disease that predominantly presents in patient over 60-years old⁶. Collectively, classical MPN have a combined annual incidence rate of 2.58 per 100,000 with ET being most frequent, followed by PV and then PMF (annual incidence rates 1.03, 0.84 and 0.47

per 100,00, respectively)⁷. Patient survival following MPN diagnosis is reduced relative to the general population due to health complications, primarily thromboembolic events, fibrous scarring of the bone marrow and leukaemic transformation. Median survival is disease-dependent and overall lower than the general population⁸. Patients above 65-years old have life expectancies of 20-years, 14-years and 6-years for ET, PV and PMF respectively⁸. Fibrotic transformation from ET or PV is relatively rare, and more common in PV patients (16%), than in ET (13%)⁶, however once transformation occurs patients are very difficult to treat. Leukaemic transformation to secondary acute myeloid leukaemia (sAML) predominantly occurs in patients with MF (10-year risk 10-20%) followed by PV and then ET (10-year risk 2-4% and 1%, respectively)⁹. Hence, patients with MF are more likely to undergo leukaemic transformation and are presented with an increasingly poor prognosis. Overall, MPN are a spectrum of chronic malignancies in the general population but present individuals with complications which reduce overall survival and significantly impact quality of life through daily symptoms such as fatigue, pruritis and life changing thrombohemorrhagic events.

1.2.3 Cell-of-origin: myeloproliferative neoplasms as a stem cell driven disease

Acquisition of an MPN-initiating mutation within a single HSC provides a selective advantage of the mutant “MPN stem cell” over normal HSC. This selective advantage drives clonal expansion and excessive myeloid differentiation which leads to the MPN phenotype¹⁰. Unlike in acute leukaemia, where myeloid progenitor cells can require the self-renewal capacity required to initiate and propagate disease, MPN can only be established and sustained by somatic mutations acquired in immunophenotypically defined long-term HSC¹⁰.

HSC are highly heterogeneous in their self-renewal capacity and possess intrinsic lineage bias's, including myeloid-restricted and platelet-bias programmes^{11,12}. This intrinsic lineage bias may, in part, explain why two MPN patients, harbouring identical driver mutations, display distinct clinical phenotypes¹⁰. Hence, the cell-of-origin in which the MPN-initiating mutation occurs is an important consideration when evaluating MPN disease heterogeneity.

1.2.4 Myeloproliferative neoplasms: a tractable model for clonal evolution in cancer

As a clonal disorder originating for the HSC compartment, MPN present as excellent tractable disease models for the study of genetic evolution and cancer stem cell biology, particularly since the advent of single-cell technologies¹³. This is due to the long disease latency in MPN patients and the ease of accessibility to serial peripheral blood and bone marrow samples allowing clonal evolution to be studied overtime in primary patient samples. Furthermore, the disease-driving mutations in MPN are well-defined and MPN are highly clonal meaning genetic heterogeneity, mutational order and disease progression can all be studied throughout the disease's genetic evolution.

1.2.5 Disease initiating mutations in myeloproliferative neoplasms

The majority of MPN patients present with mutually exclusive activating mutations within the JAK-STAT (Janus kinase/signal transducer and activator of transcription) signalling pathway. JAK-STAT constitutes a central signalling pathway for human development and homeostasis, regulating diverse processes including haematopoiesis, immunity, inflammatory responses and apoptosis¹⁴.

Activation of the JAK-STAT pathway occurs via a myriad of cytokines and growth factors to their cognate transmembrane receptors. Ligand binding leads to conformational changes in receptors which activate intracellular receptor-bound JAK proteins to phosphorylate one another. Trans-phosphorylated JAK proteins then phosphorylate cytoplasmic STAT proteins, leading to their homo- or hetero-dimerization and nuclear translocation. In the nucleus, activated STAT molecules bind target DNA sequences to regulate transcription and modulate expression of downstream target genes. Activated signalling molecules include chemokines, interferon-stimulated genes (ISG), caspases and cyclin-dependent kinases¹⁴.

There are 4 JAK proteins (JAK1-4) with JAK2 having the greatest implications in haematopoiesis. In mouse models, germline deletion of *JAK2* leads to embryonic lethality, and adult mice with conditional *JAK2* deletion suffer the same phenotype¹⁵. In adult mice,

lethality results from rapid HSC depletion and subsequent bone marrow failure, highlighting the critical role of JAK2 in HSC maintenance and function¹⁵. More specifically, JAK2 is required for HSPC response to stem cell factor (SCF) via the receptor tyrosine kinase, Kit, facilitating HSPC proliferation¹⁶. JAK2 also regulates erythroid progenitor proliferation, differentiation and maturation following binding of erythropoietin (EPO) to the erythropoietin receptor (EPOR)¹⁷. STAT1, STAT3 and STAT5A/B are downstream effectors of JAK2, where STAT3 and STAT5A/B are important for HSC function including proliferation, survival and self-renewal, and STAT1 is specifically associated with megakaryopoiesis¹⁸. Hence, while loss of *JAK2* leads to devastating bone marrow failure, uncontrolled activation of JAK-STAT signalling drives myelopoiesis and subsequent myeloproliferative disease.

Mutually exclusive activating mutations in JAK-STAT signalling genes, *JAK2*, *MPL* and *CALR*, can be found in all MPN subtypes with mutation frequency dependent on MPN classification. A small cohort of MPN cases have been identified as “triple-negative” for JAK-STAT driver mutations¹⁹. However, some triple-negative cases may be misclassified due to low-level mutation frequencies below the detection threshold for clinical diagnostic laboratories. Hence, unclassified triple-negative MPN cases remain the focus of sequencing efforts.

Biochemically, all three proteins have very different roles in canonical JAK-STAT signalling. JAK2, described above, is a non-receptor protein tyrosine kinase sensitive to activation by growth factors or cytokine binding to cognate receptors. MPL, otherwise known as thrombopoietin receptor (TPOR), is a type 1 homodimeric cytokine receptor activated by thrombopoietin (TPO). And CALR is a housekeeping gene important for protein folding and calcium ion homeostasis, whose role in the JAK-STAT signalling pathway only occurs following acquisition of an activating mutation which drives its ability to activate MPL. Importantly, all three genes, *JAK2*, *MPL* and *CALR*, when mutated, converge within the JAK-STAT signalling pathway to potentiate myeloproliferative disease.

JAK2 is located on chromosome 9p24 and the V617F mutation predominates the *JAK2* mutation profile for MPN. Biochemically, the JAK2 V617 residue is located in the

pseudokinase JH2 domain, a domain which interacts with the JH1 kinase domain to maintain JAK2 in an inactive state. The JAK2 V617F substitution leads to subtle changes in JH2 domain structure, preventing interaction between JH1 and JH2 domains and retention of the inactive state, leading to constitutively active JAK2 signalling²⁰. JAK2 V617F closely associates with the cytoplasmic domain of key haematopoietic signalling receptors for ligands including, TPO, EPO, GM-CSF, and hence the frequency of *JAK2 V617F* mutations across all MPN subtypes is substantial. In PV, JAK2 V617F is a hallmark disease feature, with 95-97% of patients presenting with the mutation. In MF and ET, JAK2 V617F drives disease in 50-60% of cases⁵.

CALR resides on chromosome 19p13.2 and encodes calreticulin, a housekeeping gene important for protein folding and calcium ion homeostasis within the endoplasmic reticulum. In 2013, somatic *CALR* mutations were identified to be responsible for MPN initiation in a large proportion of patients lacking *JAK2* or *MPL* mutations using whole genome sequencing^{21,22}. Specifically, *CALR* mutations occur in ~30-40% of ET and MF patients^{21,22}. The requirement of mutant CALR for MPN pathogenesis was surprising given CALR is not a cytokine receptor, nor does it have reported roles in canonical JAK-STAT signalling. Type I (52bp deletion) and type II (5bp insertion) *CALR* mutations in terminal exon 9 both lead to a 1bp frameshift affecting the C-terminus. The novel *CALR*-mutant C-terminal tail lacks an endoplasmic reticulum-retention peptide and in contrast to the negatively-charged C-terminus of non-mutant CALR, the mutant C-terminal peptide is populated by positively charged amino acids^{23,24}. The positively charged C-terminus is responsible for the association of mutant CALR with MPL, leading to constitutive activation of MPL and downstream JAK-STAT signalling^{23,24}. The lack of PV cases with *CALR* mutations can be explained by the exclusive association of mutant CALR with the megakaryocyte lineage restricted receptor, MPL.

The MPL cytokine receptor, *MPL*, is located on chromosome 1p3 and is a key cytokine receptor required for HSC function and regulation of megakaryocyte and platelet production^{25,26}. Upon TPO binding, MPL receptors dimerise conferring an active conformation, leading to transphosphorylation of associated JAK2 proteins and subsequent downstream signalling. *MPL* mutations are less frequent than *JAK2* or *CALR* mutations in MPN, with mutations driving disease in 6-10% of PMF and 3-5% of ET

cases⁵. The lack of activating *MPL* mutations in PV, characterised by overproduction of red blood cells, can be explained by both the lack of *MPL* expression in erythroid lineages and the essential role of *MPL* in megakaryopoiesis with mutations in *MPL* causing thrombocytosis. Novel *MPL* mutations continue to be reported via whole genome sequencing in patients who had previously been reported triple-negative for mutations in all three MPN driver genes^{19,27}.

Thus, mutually exclusive mutations within the JAK-STAT signalling pathway are responsible for initiation of myeloproliferation disease. Independent of JAK-STAT driver mutation or clinical presentation, transcriptomic analysis has shown an activated JAK-STAT signalling signature in all MPN patients studied, highlighting the convergent role of driver genes to constitutively activate the JAK-STAT signalling pathway in MPN²⁸.

1.2.6 Additional mutations contributing to myeloproliferative disease

Over 50% of MPN patients present with additional somatic driver mutations in non-JAK-STAT signalling pathways²⁹. Such mutations occur in genes with known contributions to myeloid malignancies including epigenetic regulators (*DNMT3A*, *TET2*, *ASXL1*), spliceosome components (*SRSF2*, *SF3B1*, *U2AF1*) and other prominent cancer genes (*KRAS*, *NRAS*, *TP53*)³⁰. Additional mutations are most frequently identified in *TET2* (~12%), *ASXL1* (~7%) and *DNMT3A* (~6%)³⁰. Mutation events in highly prevalent cancer-driving genes such as *TP53* and *RAS*-genes, mutated in 50% and 19% of all cancers, respectively, are rarer occurrences in MPN^{31,32}.

The order of mutation acquisition in non-JAK-STAT signalling genes may influence the heterogenous phenotypes observed in MPN. For example, order of mutations in MPN cases with co-occurring *JAK2* and *TET2* mutations can influence clinical presentation and HSPC biology²⁹. Using colony assays, researchers found “*JAK2*-first” clones were more likely to occur in patients with PV and colonies showed increased erythroid/megakaryocyte differentiation. In contrast, patients with “*TET2*-first” clones were older at clinical presentation and colonies found to expand rapidly but with reduced differentiation potential²⁹. This model likely explains the significantly earlier clinical

disease presentation in *JAK2*-first compared to *TET2*-first MPN cases, where *JAK2*-first patients' rapid expansion of progenitor populations leads to overt abnormal blood counts. Similarly, the phenotype of *JAK2*-mutant MPN can be influenced by the pattern of *DNMT3A*-mutation acquisition where "*DNMT3A*-first" patients frequently present with a PV phenotype compared to the ET phenotype acquired by "*JAK2*-first" patients³³.

Copy number alterations (CNA) also contribute to myeloproliferative disease. High resolution single-nucleotide polymorphism (SNP) arrays have shown chromosomal aberrations to be present in over 60% of MPN cases^{34,35}. Chromosome 9p acquired uniparental disomy (aUPD), a form of copy-number neutral loss-of-heterozygosity (CNN-LOH) involving duplication of the mutant *JAK2V617F* allele, is the most prevalent chromosomal aberration in MPN and occurs most frequently in PV cases (30-50%), where *JAK2V617F* gene dosage contributes to the PV phenotype^{36,37}. Additional specific lesions, such as loss of chromosome 7 or chromosome 17p, are known to be associated with disease progression in MPN^{34,35}.

Inherited and germline factors can also contribute to MPN development. A population-based study of eleven-thousand MPN patients showed a 5 to 7-fold elevated risk of MPN amongst patient's first-degree relatives, highlighting the existence of inherited and highly penetrant susceptibility loci predisposing to MPN³⁸. The *JAK2* 46/1 haplotype has been well described as major risk factor for development of MPN, and in most cases, *JAK2 V617F* arises on the 46/1 haplotype³⁹.

1.3 Blast-phase myeloproliferative neoplasms (BP-MPN)

1.3.1 Incidence of transformation and associated risk factors

Leukaemic transformation of MPN is characterised by peripheral blood cytopenias and accumulation of undifferentiated myeloblasts (10-19% in accelerated-phase and >20% in blast-phase) within the bone marrow or peripheral blood. The risk of leukaemic transformation in MPN varies, with a higher incidence in MF (10-20% after 10-years) than in ET and PV (~1-5%)⁹. Patients who transform have a highly aggressive and treatment-refractory leukaemia with a median overall survival of 3-6 months^{40,41}.

Patient, disease and treatment-related characteristics can contribute to increased risk of MPN transformation and their contribution varies depending of the MPN presented. Risk factors for transformation, comprehensively reviewed by Bewersdorf *et al*⁴², are summarised in **Table 1.1**.

Patient factor
Advanced patient age (>60 vs >70 years)
Disease factors
Extended disease duration
Risk higher for patients with PMF compared to ET and PV
Platelet count <150×10 ⁹ /L, leukocytosis, transfusion dependence, peripheral blasts ≥3%
Splenomegaly of ≥20cm from left costal margin or requirement for splenectomy
Bone marrow reticulin fibrosis
Abnormal karyotype
CALR mutations present lower risk
High-risk mutations (eg, TP53, ASXL1, SRSF2, IDH2, EZH2)
Higher DIPSS, MYSEC-PM or MIPSS70 score
Treatment factors
Prior treatment with alkylating agents and pipobroman
Lower risk in patients treated with interferon

Table 1.1. Patient, disease and treatment-derived risk factors of leukaemic transformation in MPN. Modified from Bewersdorf *et al* 2022⁴². DIPSS: Dynamic International Prognostic Scoring System; MYSEC-PM: myelofibrosis secondary to PV and ET Collaboration Prognostic Model; MIPSS: Mutation-Enhanced International Prognostic Score System

MPN prognosis, including leukaemic transformation, can be informed by risk stratification scoring systems such as the Molecular International Prognostic Scoring System (MIPSS). For prognostic scoring of overt MF, MIPSS70+ includes clinical, cytogenetic and molecular information⁴³. Significant risk factors defined by MIPSS70+ include leukocytosis (25×10⁹/L), thrombocytopenia (platelets < 100×10⁹/L), bone marrow fibrosis grade ≥2, peripheral blood blasts ≥2%, high-risk mutations and unfavourable karyotype⁴³. Unfavourable karyotypes identified as high-risk prognostic factors in leukaemic transformation include trisomy 8, monosomy 7 or 5, 7q-, 5q-, inv(17q), inv(3), 12p- or 11q23, as complex karyotypes (≥3 events), alone or in pairs⁴⁴. Somatic mutations also contribute to MF transformation, with mutations in *ASXL1*, *SRSF2* or *IDH1/2*, significantly limiting leukaemia-free survival⁴⁵. Risk factors for leukaemic transformation in PV and ET vary subtly from MF due to the differing biological and clinical disease parameters. The PV scoring system, MIPSS-PV, attributes risk to age >67, thrombosis history, leucocyte counts ≥15×10⁹/L and the presence of *SRSF2* mutations^{46,47}. Abnormal karyotype has also been reported to adversely limit leukaemic-

free survival in PV^{47,48}. For ET, MIPSS-ET risk factors include age >60, male sex, leucocyte count $\geq 11 \times 10^9/L$ and high-risk mutations specifically within *SRSF2*, *SF3B1*, *U2AF1* or *TP53*^{46,49}. Independent studies further support the adverse effect high-molecular risk mutations present for transformation-free survival in ET, specifically splice factor and *TP53* mutations⁵⁰.

Historical treatments for MPN have been associated with leukaemic transformation including radioactive phosphorus (P^{32}) and alkylating agents, chlorambucil and pipobroman^{47,51,52}. Hydroxyurea is a current standard-care, non-alkylating cytoreductive reagent for the management of MPN symptoms. Hydroxyurea interferes with DNA synthesis, bringing into question its therapeutic utility in MPN, given its mutagenic and leukaemogenic potential. However, multiple studies have refuted the leukaemogenic potential of hydroxyurea treatment in MPN, showing no significant difference in transformation risk between HU-treated and non-HU treated MPN cohorts^{53–56}.

1.3.2 Genetic heterogeneity in BP-MPN

Progression from CP-MPN to BP-MPN is associated with increased clonal complexity, with patients harbouring more somatic mutations compared to CP-MPN^{57,58}. Furthermore, compared to *de novo* AML, striking molecular and cytogenetic differences are observed in BP-MPN. *NPM1* and *FLT3* mutations are frequently implicated in *de novo* AML, yet rarely observed in BP-MPN^{57,58}. In contrast, the most commonly occurring mutations observed in BP-MPN include *ASXL1*, *SRSF2*, *RUNX1*, *TET2* and *TP53*⁵⁹.

Cytogenetic abnormalities are also relatively infrequent in *de novo* AML, with less than 50% of patients presenting with copy number alterations (CNA) or structural variants⁵⁷. In comparison, these events are a frequent occurrence in BP-MPN and contribute to the increasing clonal complexity and genetic heterogeneity observed upon progression from CP-MPN^{59–62}. High-resolution single nucleotide polymorphism arrays have identified frequent events associated with BP-MPN including 5q, 17p and 21q deletions, monosomy 7, trisomy 8 and gains of chr1q^{60–63}. Commonly altered regions can be mapped to genes with known leukaemogenic functions for example, *RUNX1* within chr21q-deletions and the TP53 inhibitor, *MDM4*, in chr1q-amplified regions⁶¹. Prominent CNN-LOH events and

deletions of 17p are observed in BP-MPN, the loci in which *TP53* resides^{60,62,63}. Furthermore, CNN-LOH or 17p deletion are associated with complex karyotype, progression to BP-MPN and reduced overall survival following transformation^{60,62,63}. In a multivariate analysis, including clinical and genetic parameters, the presence of ≥ 4 somatic mutations, ≥ 3 cytogenetic abnormalities or mutant *TP53* correlated with reduced overall survival in BP-MPN, highlighting the importance of additional genetic determinants, including specific mutations, in the clinical course of BP-MPN⁵⁹.

1.3.3 Single-cell technologies to unravel intratumoural heterogeneity

To address the vast intratumoural genetic and non-genetic heterogeneity in BP-MPN requires the application of advanced single-cell techniques. While historic single-cell techniques, including flow cytometry analysis of surface markers/intracellular proteins and karyotyping, are still routinely used in assessment and diagnosis of haematological malignancies, immense progress has been made in the development of advanced single-cell technologies.

Single-cell techniques range from single modality approaches – focusing on DNA, transcriptome and epigenome level analysis – to “multi-omic” approaches capable of capturing multiple layers of molecular information. Multi-omic techniques are of particular importance in cancer biology, where multiple layers of information are necessary to comprehensively evaluate intratumoural heterogeneity, define clonal hierarchies, identify driver mutations and study cancer evolution. For example, techniques such as Genotyping of Transcriptomes (GoT) and TARGET-seq capture transcriptional and genetic information at the single-cell level, allowing mutant-specific transcriptional signatures to be defined^{64,65}. Both GoT and TARGET-seq have generated novel insights in MPN biology, where mutant cells lack or overexpress surface markers distinct from the non-malignant clone^{64,65}. GoT, a droplet-based methodology, uses gene-specific primers to amplify target transcripts and shared cell-specific barcodes to couple mutation and transcriptomic information from cDNA. Using this high-throughput methodology, researchers profiled >38,000 CD34+ HSPC from *CALR*-mutant ET patients and found *CALR* mutations, originating from the HSC pool, provide a cell-intrinsic fitness

advantage to myeloid progenitors, enriching myeloid progenitor populations with *CALR*-mutant clones⁶⁵. It should be noted that, for GoT to call genetic variants from cDNA transcripts, high expression of the mutant transcript is required for sufficient mutation capture and GoT cannot identify variants which lead to transcript loss (nonsense or frameshift mutations). Difficulties can also be experienced in reliably identifying WT and heterozygous cells, due to high allelic dropout frequencies and/or the allele-specific transcript expression.

TARGET-seq, a method developed in the Mead group, circumvents limitations of GoT, achieving high-sensitivity mutation detection through low allelic dropout rates and targeted mutation analysis from both cDNA and gDNA⁶⁴. As such, TARGET-seq allows mutation detection and simultaneous unbiased whole-transcriptome analysis from individual cells, with integration of cell surface proteomics data via index sorting of single-cells by fluorescently activated cell sorting (FACS; **Figure 1.2**)⁶⁴.

TARGET-seq is a plate-based methodology, with single-cells sorted into 96 or 384-well plates, reducing the throughput compared to droplet-based methodologies such as GoT^{64,65}. However, through allelic-resolution mutation detection, TARGET-seq has the capacity to distinguish malignant subclones from admixed WT cells, with reliable determination of true WT and heterozygous-mutant populations. Applying TARGET-seq to MPN biology facilitates intra-patient transcriptomic comparisons between WT and mutant HSPC, enabling identification of mutation-specific transcriptomic signatures. The specific overexpression of genes in mutant cells populations allows potential new therapeutic targets to be identified. Overexpressed genes may be essential for cancer cell survival, be required for resistance to current therapeutic strategies or present a synthetic lethal requirement in combination with other target genes. Thus, single-cell multi-omic techniques allow researchers to identify novel molecular vulnerabilities that can be exploited to selective target mutant cell populations.

Other advances in single-cell multi-omic techniques include combining transcriptomics and cell surface protein expression (CITE-Seq; Cellular Indexing of Transcriptomes and Epitopes by Sequencing)⁶⁶, transcriptomics and chromatin accessibility (10X Multiome), genotyping and protein (Mission Bio Tapestry) and genotyping and chromatin accessibility (GoT-ChA; genotyping of targeted loci with single-cell chromatin

accessibility or GTAC; genotyping with the assay for transposase-accessible chromatin)^{67,68}. Applied to MPN these techniques have the potential to decipher clonal evolution, disease progression and therapeutic vulnerabilities. For example, GoT-ChA, was applied to *JAK2*^{V617F}-mutant CD34+ cells derived from MPN patients to investigate how somatic mutations disrupt chromatin accessibility profiles governing cellular fate and commitment⁶⁸. Izzo *et al*⁶⁸ showed the presence of a *JAK2*^{V617F}-mutation in HSPC directed cell-intrinsic and cell-type specific changes in the epigenetic landscape of malignant cells, increasing chromatin accessibility in genes responsible for a pro-inflammatory state and myeloid-priming.

Single-cell multi-omic technologies have not yet been integrated into clinical pathways and routine clinical care, however, their potential to inform diagnosis, treatment response and resistance, disease progression and relapse risk is clear. For example, the ability of single-cell techniques to identify molecular vulnerabilities specific to only mutant clones with the potential for therapeutic targeting has been shown. Using a TARGET-seq approach, G6B is a cell surface protein which has been identified as aberrantly expressed on mutant HSPC in MF which can mediate ablation of the mutant MF clone via G6B-targeting immunotherapy⁶⁹.

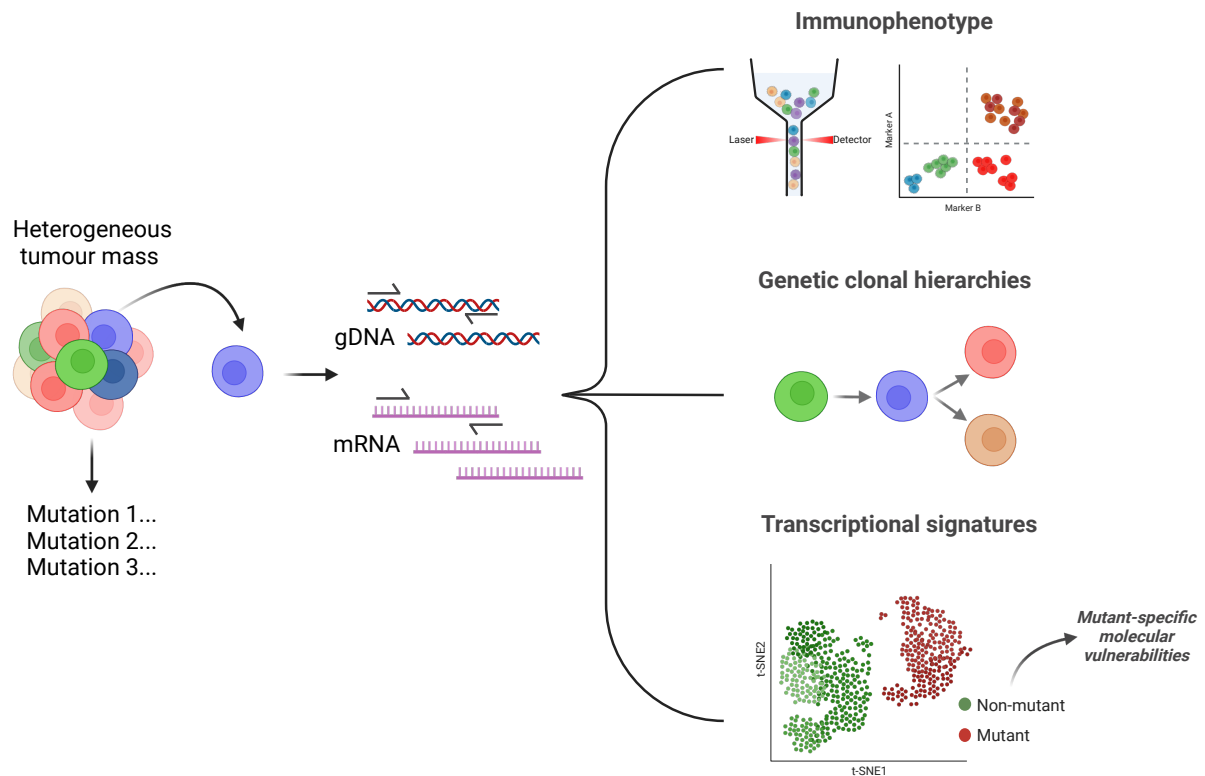


Figure 1.2. Resolving intra-patient tumour heterogeneity using TARGET-seq. TARGET-seq facilitates allelic resolution mutation detection and parallel whole transcriptome analysis from individual immunophenotyped cells. Identification of mutant-specific transcriptional signatures potentiates development of selective therapeutic strategies targeting mutant-specific molecular vulnerabilities. Illustration adapted from Rodriguez-Meira *et al*⁶⁴.

1.3.4 Therapeutic treatment of BP-MPN

BP-MPN is a significant unmet clinical need. Survival estimates following disease progression to BP-MPN are dismal, with a median overall survival (OS) time of <6 months and 1-year and 3-year survival rates of <20% and <5%⁴¹. Allogenic haematopoietic stem cell transplant (allo-HSCT) remains the only treatment to provide any long-term benefit, although survival rates remain low (3-year 18-32%, 5-year 10%)^{41,70}. Additionally, access to allo-HSCT is limited by confounding factors including a patient's age, competing comorbidities and access to a suitable donor. Induction chemotherapy treatments for AML, such as cytarabine and daunorubicin, also have limited impact on BP-MPN overall survival (median survival 4.9-months⁴⁰), and responses are not durable without

subsequent transplant^{40,71}. The observed lack of response in BP-MPN to traditional AML therapies is likely due to the pathogenetic disparity between the *de novo* and transformed disease.

Patients unsuitable for transplantation and/or who cannot tolerate induction chemotherapy regimens are offered less intensive therapies, including hypomethylating agents (HMA; azacytidine and decitabine). Azacytidine has been shown to improve OS, however, with limited response durations (median OS 11-months)⁷². Response and OS outcomes following decitabine treatment are comparable to those observed via intensive chemotherapy and HMA therapy is therefore increasingly recognised as an appropriate alternative due to improved tolerability^{73,74}. A combination of HMA and venetoclax, an inhibitor of the BCL-2 anti-apoptotic protein, has been shown to have deep and durable remissions in older patients previously untreated for AML^{75,76}. The response to HMA + venetoclax in BP-MPN however, does not reflect the treatments superior outcomes observed in AML, with a median survival time of 8-months and no significant difference in OS when compared to HMA treatment alone⁷⁷. Finally, despite its ability to reduce spleen size, alleviate disease symptoms and improve OS in MF, the oral JAK1/2 inhibitor, Ruxolitinib, has shown minimal disease-modifying activity in BP-MPN^{78,79}. Thus, it is clear BP-MPN remains a significantly unmet clinical need and it is therefore necessary to identify improved therapeutic strategies.

A number of emerging therapies for the treatment of BP-MPN are currently under assessment. For example, Ruxolitinib and decitabine combination therapy has shown synergistic effects *in vivo*⁸⁰ leading to phase I and II clinical trials in BP-MPN^{81,82}. Results from the phase II clinical trial including 25 participants showed a 44% overall response rate and median survival of 9.5 months through Ruxolitinib + decitabine combination therapy, an improvement on historical induction therapy data⁸¹. A number of novel therapeutic strategies have also entered clinical trial. These include a multicentre phase 1b/2 study (NCT04113616) evaluating the efficacy of Navtemadlin in BP-MPN⁸³. Navtemadlin selectively inhibits MDM2, a negative regulator of p53, and has the capacity to restore p53 function in *TP53*-WT cohorts. Preclinical support of this trial comes from prolonged survival following Navtemadlin treatment in BP-MPN *TP53*-WT patient-derived xenograft (PDX) models⁸⁴. Other novel targeted therapies in preclinical

development include LSD1 inhibitor, CPI-482. LSD1 (lysine-specific demethylase 1) catalyses the removal of histone H3 lysine 4 methylation at gene enhancers and transcriptional start sites, leading to transcriptional 'reprogramming' and alternate cellular fates⁸⁵. In AML cell line screens, LSD-1 was shown to be highly responsive in *JAK2*-mutant cell lines and efficacy verified in *JAK^{V617F}/TP53*-null mouse models⁸⁵.

Given the dismal prognosis outcomes in BP-MPN, innovative treatments strategies are crucial to improve disease outcomes. New combination strategies and ongoing clinical trials using novel therapeutic agents show some promise and will continue to improve. Furthermore, application of single-cell multi-omic techniques has the capacity to unravel genetic and molecular heterogeneity driving disease transformation and maintenance, with the potential to uncover novel vulnerabilities for therapeutic targeting.

1.4 *TP53*-mutant secondary acute myeloid leukaemia (*TP53*-sAML)

1.4.1 p53, guardian of the genome

Forty-five years since its discovery, p53 has risen to prominence and with >120,000 p53-relevant publications on PubMed, it has become one of the most extensively studied genes/proteins in history. Originally discovered in 1979 by multiple independent researchers⁸⁶, p53, encoded by *TP53* and located on chromosome 17p, was identified as an oncogene responsible for promoting cellular transformation. Evidence suggesting p53 was in fact, a tumour suppressor, began to emerge in the 1980's, where studies showed *TP53* was inactivated by retroviral insertion or extensively rearranged upon transformation⁸⁶. It was subsequently discovered many cDNA clones were derived from cancer cells lines overexpressing mutant *TP53* or murine *Trp53* alleles⁸⁶. Evidence resolving *TP53* as a bona fide tumour suppressor came in the 1990's when germline *TP53* mutations were found to be responsible for early-cancer onset in Li-Fraumeni syndrome⁸⁷, and highly-penetrant cancers were observed in the first *Trp53*-knockout mouse model⁸⁸. While *TP53* has been confidently ascertained to be a tumour suppressor and mutations known to compromise WT p53 activity, this traditional view of p53 function fails to account for the many *TP53* variants reported to have GOF properties⁸⁹.

Thus, *TP53* is an atypical cancer gene and can be characterised as a tumour suppressor which when mutated can possess additional oncogenic properties.

Overview of p53 structure, regulation and function

p53 is a sequence-specific DNA-binding transcription factor containing 5 structural domains; an N-terminal transactivation domain, proline-rich domain, central DNA-binding domain (DBD), tetramerization domain and C-terminal regulatory domain⁹⁰. In normal unstressed cells, low level p53 expression is maintained by the ubiquitin ligase, MDM2. Regarded as an important gatekeeper of p53 expression, MDM2 negatively regulates p53 expression via either sterically blocking the p53 N-terminal transactivation domain or initiating proteasomal degradation via ubiquitination of the C-terminal domain⁹¹. Interestingly, *MDM2* is also a transcriptional target of p53, creating a negative feedback loop and limiting excessive p53 expression in non-stressed cells. p53 is stabilised in response to cellular stress, including DNA damage, replication stress and oncogenic signalling⁹¹. For example, in response to DNA-damage, DNA-damage response (DDR) kinase proteins, ATM and ATR, phosphorylate p53 and block MDM2-mediated degradation⁹². Activation of p53 also commonly occurs indirectly via degradation of MDM2.

Upon activation by cellular stress, p53 has two prominent canonical functions; to induce cell cycle arrest and repair DNA-damage or initiate apoptosis⁹³. Cell cycle arrest by p53 is primarily mediated by transcriptional activation of *CDKN1A*, the cyclin-dependent kinase inhibitor, p21⁹³. Activated p21 binds cyclin E/Cdk2 and cyclin D/Cdk4 preventing their ability to hyper-phosphorylate the retinoblastoma protein (Rb) and release the bound E2F1 transcription factor, critical for transcription of target genes responsible for DNA replication and G₁ cell-cycle progression⁹³. Downstream of p53-mediated cell cycle arrest, p53 orchestrates a variety of DDR-mechanisms to facilitate DNA-repair⁹⁴. As a regulator of apoptosis, p53 directly activates a number of genes within the intrinsic BCL-2 regulated apoptosis pathway including pro-apoptotic BH-3-only proteins (Puma, Noxa), cell death effectors (Bax and Bak) and caspase-9 scaffold protein (APAF-1)⁹⁵. Thus, these protective mechanisms prevent mutation acquisition and genetic evolution of damaged cells to drive cancer initiation and progression.

Above is a rather simplistic overview of p53 functions when in fact p53 directly regulates the transcription of >500 genes to control diverse cellular processes⁹¹. Extensive research documenting the “non-canonical” roles of p53 depicts a bewildering array of biological functions including roles in metabolism, repression of pluripotency and iron-dependent ferroptosis⁹⁰. This highlights the remarkably flexible and highly context dependent nature of the p53 response, determined by specific sets of regulators and situational control over distinct sets of target genes⁹⁰. The p53 response can also be cell-type specific⁹⁶, further emphasising the necessity of single-cell genomic techniques to unravel the varying p53-responses in different cell populations. Therefore, despite an increasingly comprehensive view of the p53 network, a complete understanding of p53’s context-dependent roles and regulation requires further investigation⁹⁰.

p53 in cancer biology

TP53 is the most frequently mutated gene in human cancer with over half of all cancers possessing a mutant *TP53* allele³¹. Missense mutations are most prominent in the *TP53* locus, and a large frequency can be found within p53’s DBD. Notable missense mutations within DBD hotspots account for 30% of all *TP53* missense mutations and include R175, G245, R248, R249, R273, and R282⁹⁷. Missense mutations can be classified by whether they interfere with DNA binding activities (e.g., R248W and R273H) or conformational stability (e.g., R175H, G245S, and R249S).

The functional effect of a *TP53* mutation can be either (1) loss-of-function (LOF) where WT tumour suppressor functions are simply lost, (2) dominant-negative effect (DNE), where mutant p53 interferes and hinders the function of remaining WT p53 function or (3) gain-of-function (GOF), where mutant p53 acquires additional oncogenic functions beyond p53 inactivation alone⁹⁷. Mutation-saturation screens have unequivocally shown the DNE of certain *TP53* missense mutations to enhance cellular fitness⁹⁸. In contrast, uncovering which p53 mutants harbour true GOF activities, and their mechanisms, is more complex due to the vast and highly context dependent effector network of mutant p53. GOF mutants are therefore the subject of intense research⁹⁹.

Regardless of the *TP53*-missense mutation acquired, a *TP53*-heterozygous mutant state is primarily transient, and ultimately leads to inactivation of the residual WT allele either via biallelic inactivation or loss-of-heterozygosity through a cytogenetic deletion event. Complete inactivation of *TP53* leads to loss of p53-regulated tumour suppressive functions, including an accumulation of mutations and replication errors in the absence of activated DDR-pathways and a lack of protection against missense mutant *TP53* GOF activities in pre-malignant cells, ultimately leading to oncogenic transformation.

TP53 mutations are an independent prognostic marker of poor OS and treatment response in an array of cancers¹⁰⁰. Given the centrality and impact of *TP53* mutations/inactivation in human cancer, restoration of WT p53 is an attractive therapeutic strategy and can effectively lead to tumour regression *in vivo*¹⁰¹. For decades p53 has been determined “undruggable”, due to its lack of classical targeting features and the complexity surrounding re-establishing WT p53 function for a broad-spectrum of mutants¹⁰².

Confirmational p53 mutants are attractive candidates for the development of small molecule inhibitors to restore WT p53 structure and function. Many of these small molecule inhibitors target a relatively broad spectrum of p53 mutants, however, approaches targeting specific p53 mutants are also being investigated. For example, ZMC1 is a small molecule inhibitor targeting the p53^{R175H} confirmational mutant. Restoring the WT p53 structure, ZMC1, was found to promote p53-dependent apoptosis and tumour regression *in vivo* in a mutant-allele specific manner¹⁰³. Confidence in the application of mutant-specific inhibitors can be drawn from FDA approval of a KRAS^{G12C} inhibitor, because like p53, KRAS is considered another challenging therapeutic target. Alternative therapeutic approaches are gaining traction, including immunotherapies targeting p53-mutant neoantigens and p53-based genetic therapies¹⁰². However, while alternative approaches circumvent many challenges met by small molecules, they present their own challenges. For example, immunotherapy is hindered by the paucity of neoantigen presentation and gene-therapies present considerable challenges in vector delivery and restricted application to *TP53*-null cancers¹⁰². While it is clear a one-drug-fits-all approach is highly unlikely to be successful in *TP53*-mutant cancers, given the high

frequency of certain *TP53* mutations, effective treatment's targeting a handful of specific mutations could benefit a significant number of patients.

1.4.2 p53 in normal and malignant haematopoiesis

p53 in normal haematopoiesis

Trp53-null mice develop with normal haematological parameters¹⁰⁴. Despite this, various roles of p53 in stem cell biology have been described including involvement in HSC quiescence and self-renewal. Under steady-state haematopoiesis, most HSC adopt a quiescent state to preserve their long-term function and self-renewal capacity. Highly primitive and quiescent long-term HSC (LT-HSC) populations have the greatest expression of *Trp53* mRNA compared to progenitor populations in murine models¹⁰⁵. In *Trp53*-null mice, LT-HSC populations undergo expansion within the bone marrow¹⁰⁵⁻¹⁰⁷. Concordant with this expansion, *Trp53*-null mice show a significant reduction in quiescent HSC populations and an increase in the population of HSC entering the cell cycle¹⁰⁵. Mechanistically, p53 was identified to regulate HSC quiescence through p53-target genes *Gfi-1* and *Necdin*¹⁰⁵. Importantly, *Trp53*-null HSC showed reduced engraftment ability compared to WT counterparts in competitive repopulation and single-cell serial transplantation assays, suggesting the absence of p53 compromises HSC functional fitness¹⁰⁶. Thus, WT p53 has been identified as an important mediator of HSC quiescence and self-renewal.

Similar to the reduced intrinsic reconstitution ability of *Trp53*-null HSC, constitutive activation of p53 is also detrimental to long-term haematopoietic fitness^{108,109}. The highly proliferative nature of the haematopoietic system makes it increasingly vulnerable to DNA-damage and accumulation of mutations with age. Apoptosis is the primary response of haematopoietic cells to DNA-damage¹⁰⁵ and depletion of HSC from the bone marrow of mouse models with constitutively activated p53 was shown to be mediated by p53-dependent activation of pro-apoptotic protein, Puma¹⁰⁸. An independent study showed activating mutations in p53 also led to a premature ageing phenotype due to reduced HSC proliferation capacity and impaired self-renewal¹⁰⁹. Cells with constitutive p53 activity are also increasingly sensitive to exogenous stress, such as low-level ionising radiation, leading to excessive cell elimination from the haematopoietic system¹¹⁰. Therefore,

insufficient p53 activation favours cell survival and proliferation but impedes self-renewal capacity and leaves cells vulnerable to genomic instability and an increased likelihood of oncogenic transformation. However, increased p53 dosage can compromise haematopoietic homeostasis leading to excessive elimination of cells and eventual bone marrow failure, concluding careful *TP53* gene dosage is crucial for steady-state haematopoiesis.

p53 in malignant haematopoiesis

p53 is mutated in over 50% of all cancers, however, the incidence of *TP53* mutations and deletions in haematological cancers is considerably lower than that observed in solid tumours. For examples, *TP53* is mutated in 3-8% of *de novo* AML¹¹¹, <3% of acute lymphoid leukaemia (ALL) cases¹¹² and 10-12% of multiple myeloma cases¹¹³. Nonetheless, in these malignancies, *TP53* mutations associate with inferior clinical outcomes¹¹¹⁻¹¹³.

Despite the lower frequency of *TP53* mutations in haematological malignancies, *TP53*-mutant myeloid malignancies are recognised as a distinct clinical entity by the International Consensus Classification (ICC)¹¹⁴. *TP53*-mutant myeloid malignancies can be characterised by a high frequency of CNA and complex chromosomal aberrations^{115,116}. It has also been shown that common mutations in other subtypes of AML, for example *DNMT3A*, *NPM1* and *FLT3*, are less frequent in *TP53*-mutant myeloid malignancies¹¹⁶. In contrast to *de novo* AML, *TP53*-mutations are reported at a higher frequency in both therapy-related AML and BP-MPN (30-40% and 25-35%, respectively)^{80,117}. Overall, independent of whether the *TP53*-mutation occurs in *de novo* or secondary myeloid malignancies, *TP53*-mutant myeloid malignancies have inferior survival outcomes compared to *TP53*-WT disease^{118,119}.

1.4.3 *TP53*-mutant leukaemic evolution from chronic-phase MPN

Approximately 10-20% of MPN patients undergo leukaemic transformation to BP-MPN, otherwise known as secondary acute myeloid leukaemia (sAML)⁹. *TP53* mutations are detected in 25-35% of MPN patients who experience leukaemic transformation^{74,80,120,121}.

Overall survival of patients with *TP53*-sAML is exceeding poor (OS 4.4-months)⁷⁴, largely due to the lack of effective therapeutic strategies. While clinically patients who transform are described to have BP-MPN, in a research setting, sAML and BP-MPN are used more interchangeably. Therefore, when describing *TP53*-mutant leukaemic evolution from CP-MPN, this thesis will refer henceforth to this disease entity as *TP53*-sAML.

At the time of leukaemic transformation, *TP53*-mutant clones predominantly have a variant allele frequency (VAF) nearing 50%^{80,121}, indicating loss of residual WT *TP53* allele. Mouse models have imitated this *TP53*-sAML phenotype, where a combination of *JAK^{V617F}* and *TP53* loss leads to highly penetrant and clonally-dominant myeloid leukaemia *in vivo*⁸⁰. In contrast, the *TP53* VAF in CP-MPN clones is frequently low (<15%) and in this lower range may be undetectable depending on sequencing sensitivity^{80,121,122}. Thus, low burden *TP53*-mutations in CP-MPN may be greater than published prior to advances in high-throughput sequencing technology. Using an ultra-deep next-generation sequencing (NGS) approach, the occurrence of low burden *TP53*-mutations has been analysed. Results showed a considerable 15% of CP-MPN patients harboured a low VAF *TP53*-mutation (VAF 0.2-11.6%)¹²². Strikingly, only 6% (3/50) of CP-MPN patients harbouring a *TP53*-mutation underwent leukemic transformation¹²², suggesting *TP53*-mutation alone is insufficient for disease progression.

Single-cell technologies have advanced our understanding of clonal evolution and disease progression^{63,123,124}. Using the TARGET-seq methodology, Rodriguez-Meira *et al*^{63,64} deciphered the genetic and non-genetic determinants of *TP53*-sAML disease progression following acquisition of the initial *TP53*-mutation. Results showed *TP53* missense mutations were invariably associated with loss of the residual WT allele and complex CNA. All *TP53*-sAML patients presented with CNA and the majority of patients (87%) showed complex karyotype, defined by ≥ 3 CNA⁶³. Thus, dominant leukaemic clones present a strong selective pressure for complete *TP53* inactivation, either via biallelic inactivation or LOH, and complex cytogenetic evolution.

Non-genetic drivers of *TP53*-mutant leukaemic transformation were also investigated via transcriptional analysis of serial patient samples⁶³. *TP53*-mutant HSPC identified in CP-MPN are the genetic ancestors of *TP53*-multihit LSC identified upon transformation. To

investigate molecular drivers of disease progression, Rodriguez-Meira *et al*⁶³ analysed the transcriptomic profiles of *TP53*-heterozygous HSPC from CP-MPN patients who do and do not transform. Aberrant inflammatory signalling was identified in *TP53*-heterozygous HSPC of patients who retrospectively transform, highlighting a previously unappreciated role of chronic inflammatory signaling in *TP53*-mutant clonal evolution towards *TP53*-sAML. Competitive transplant experiments in murine models validated the role of chronic inflammation in *TP53*-mutant leukaemia progression, where engrafted *Trp53*^{R172H/+} cells were selectively enriched, at the detriment of the competing *Trp53*^{+/+} population, in the presence of chronic inflammatory stimuli⁶³. Furthermore, *Trp53*^{R172H/+} cells gained karyotypic abnormalities compared to *Trp53*^{+/+} wildtype competitors, reflecting the complex cytogenetic evolution observed in primary samples upon transformation.

Overall, transformation of *TP53*-mutant MPN is associated with a grave prognosis and a lack of effective therapeutic strategies. Unravelling intratumoural heterogeneity using single-cell multi-omic techniques has identified genetic and molecular events driving *TP53*-mutant disease transformation (**Figure 1.3**). In the presence of WT *TP53*, HSC undergo inflammation-induced^{125–127} and DNA-damage-associated¹²⁸ attrition, whereas *TP53*-mutant HSC are rescued from DNA-damage and instead accumulate CNA and complex karyotypes that underlie disease progression. Thus, these findings have the potential to improve disease prognosis and inform the development of effective therapeutic strategies to prevent leukaemic transformation of MPN.

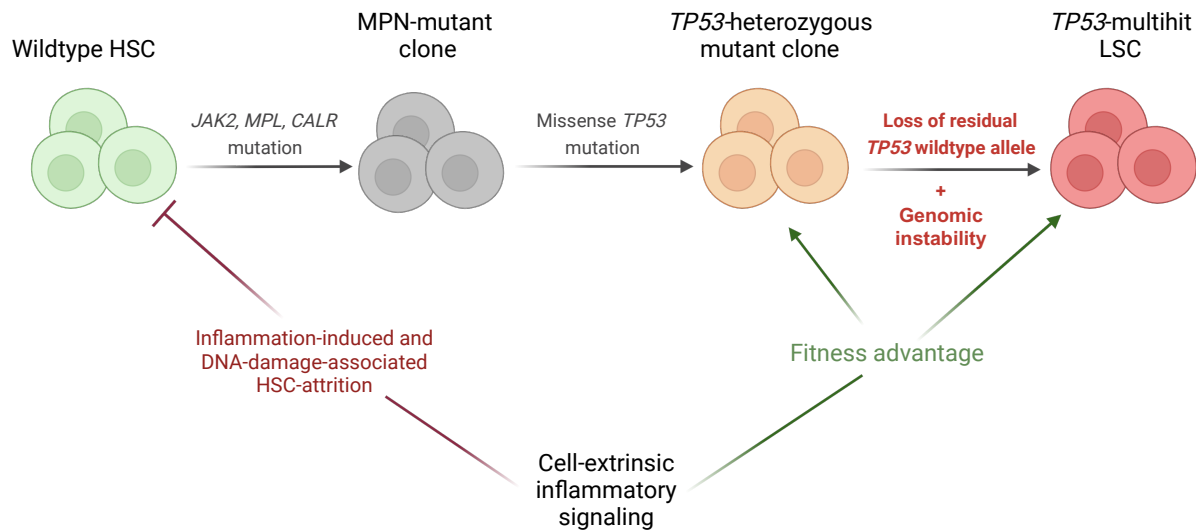


Figure 1.3. Genetic and non-genetic determinants of *TP53*-mutant leukaemic transformation in MPN. Wildtype HSC gain activating mutations within JAK-STAT signalling pathway driving myeloproliferative disease. Genetically, leukaemic transformation of the MPN clone requires inactivation of wildtype *TP53* and associated cytogenetic evolution. Cell-extrinsic inflammation selects against *TP53*-wildtype HSC, leading to their attrition, while driving genetic instability in *TP53*-mutant clones and disease progression. Illustration adapted from Rodriguez-Meira *et al*⁶³ and created on BioRender.com.

1.5 Summary and aims

While the genetic hierarchy and molecular determinants of *TP53*-mutant MPN transformation have been explored, the requirements for *TP53*-sAML maintenance remain unknown. The overarching aim of this thesis was therefore to identify and validate LSC-specific molecular vulnerabilities for therapeutic targeting in *TP53*-sAML, and ultimately to revert a patient's disease status to CP-MPN, which is manageable via therapies such as the JAK2 inhibitor, Ruxolitinib. The single-cell multi-omic TARGET-seq dataset generated by Rodriguez-Meira *et al*⁶³ containing *TP53*-sAML, MPN and healthy donor (HD) controls presents a unique opportunity to investigate the yet unexplored molecular requirements of *TP53*-sAML. Finally, *TP53*-sAML remains a disease with dire prognosis and unmet clinical need. Novel therapeutic candidates are therefore crucial to improve treatment and disease outcomes.

Thus, the overall aims of this thesis are specified below:

1. Identify putative molecular vulnerabilities in *TP53*-mutant LSC in *TP53*-sAML
2. To validate the requirement of molecular vulnerabilities in cell line models and primary patient samples
3. Investigate the mechanisms by which specific vulnerabilities are necessary for *TP53*-sAML maintenance

It is important to note this is a general introduction and each subsequent chapter is paired with a relevant section-specific introduction.

Chapter 2 Methods

2.1 Cell line assays

2.1.1 Cell line maintenance

Suspension cell lines, HEL and SET-2, were cultured in RPMI-1640 including L-Glutamine (R8758, Sigma-Aldrich). Media was further supplemented with 1% penicillin/streptomycin (P/S) and 10% or 20% foetal bovine serum (FBS; F9665, Sigma-Aldrich) for HEL or SET-2 cells, respectively. Adherent cell line, HEK293T, was cultured with high glucose Dulbecco's Modified Eagle Medium (DMEM; 11960044, Gibco) supplemented with 10% FBS, 1% L-Glutamine and 1% P/S.

Cell lines were passaged weekly, seeded at $\sim 0.4 \times 10^6$ /mL and incubated at 37°C with 5% CO₂. Cell lines were passaged for a maximum of two months, after which new stocks were thawed. All cell lines were tested for mycoplasma contamination at regular 3-month intervals. Cell line authentication was conducted via targeted sequencing of known mutations by Sanger Sequencing (**Table 2.1**).

Cell culture was conducted under laminar flow with all equipment and reagents entering the hood being cleaned thoroughly with 70% industrial methylated spirits (IMS) to maintain sterile conditions.

HEL cell line		
Gene	Mutation	Allelic State
<i>JAK2</i>	V617F (g.1849 G>T)	Homozygous
<i>TP53</i>	M133K (c.398 T>A)	Homozygous

SET-2 cell line		
Gene	Mutation	Allelic State
<i>JAK2</i>	V617F (g.1849 G>T)	Heterozygous
<i>TP53</i>	R248W (c.742 C>T)	Heterozygous
<i>TP53</i>	c.599+2T>A	Heterozygous

Table 2.1. *JAK2* and *TP53* mutation status in HEL and SET-2 cell lines

2.1.2 Lentiviral production

Transfection of packaging cells

To achieve stable expression of exogenous sequences, for example single-guide RNA (sgRNA) or short hairpin RNA (shRNA), a lentiviral approach can be used. To increase lentiviral safety and reduce biosafety risks, lentivirus components for virus production are split across multiple plasmids. Here, a second-generation lentiviral system was used, requiring independent transfer, envelope and packaging plasmids. The replication incompetent transfer plasmid includes the sequence of interest flanked by long terminal repeat sequences to aid genomic integration. The psPAX2 packaging plasmid, encodes three viral genes, *gag*, *pol* and *tat* which are responsible for encoding viral structures, reverse transcriptase and integrase components plus transcriptional trans-activators for upregulation of gene expression, respectively. The envelope plasmid, pMDG.2, encodes the envelope protein, VSV-G, which recognises a low-density lipoprotein receptor (LDL-R) on the target cells surface and facilitates receptor-mediated endocytosis of viral particles. Ubiquitous expression of the LDL-R allows lentiviral transduction of a wide variety of cell types. Additionally, the ability of lentivirus to transduce quiescent types, unlike retrovirus, makes this approach attractive in the study and manipulation of HSPC populations.

Lentivirus was generated using a polyethylenimine (PEI) transfection protocol and concentrated by ultracentrifugation to improve viral titres for cell transduction. PEI is a synthetic, positively-charged polymer which aids transfection via condensing DNA into positively-charged particles capable of binding anionic cell surfaces to subsequently be internalised by endocytosis.

The day prior to transfection, HEK293T cells at 80-90% confluence were washed with PBS and incubated for 5-minutes with trypsin (Sigma-Aldrich, T4049) to promote disassociation from the culture dish. Following disassociation, trypsin was neutralised with complete DMEM, supplemented with additional 1X non-essential amino acids (NEAA; Gibco, 11140050) and cells in suspension counted by haemocytometer. Each 15cm plate was seeded with 1.2×10^7 HEK293T cells in 25mL complete DMEM media including 1X NEAA. The following day, cells were transfected via PEI-mediated

transfection. To prepare PEI-DNA complexes for transfection (per 15cm plate), 1mL OptiMEM (Gibco, 31985062) was incubated with 33uL PEIpro (Polyplus, 101000017) for 5-minutes at room temperature (RT). In parallel, for transfer plasmids <15Kb, 13.3ug transfer plasmid, 10ug psPAX2 packaging and 6.7ug pMD2.G envelope plasmid were mixed. After incubation, Opti-MEM containing PEIpro was added to the plasmid mixture and incubated at RT for 15-minutes to allow PEI-DNA complexes to form. PEI-DNA complexes were then added dropwise across each 15cm dish and returned to the incubator. 5-6 hours post transfection, viral media was removed and replaced with 10mL pre-warmed complete DMEM including 1X NEAA with great care taken not to detach cells. Plates were subsequently returned to the incubator at 37°C with 5% CO₂.

48-hours after transfection, viral media was harvested and replaced with a further 10mL complete DMEM including 1X NEAA before being returned to the incubator for a further 24-hours. Viral harvests were kept safely contained at 4°C. The following day (72-hours post-transfection), viral media was again collected and combined with previous harvest for concentration by ultracentrifugation.

Lentivirus concentration via ultracentrifugation

Upon harvest, viral media was prepared for ultracentrifugation. Viral media was centrifuged at 2000 $\times g$ for 5-minutes at 4°C and supernatant filtered through 0.45uM Millex-HC PVDF filter (Millipore, SLHV033RS) to remove any cell debris. Viral supernatant was layered onto a sucrose gradient for ultracentrifugation. Briefly, 5mL 20% sucrose (dissolved in sterile PBS and filtered through 0.22uM filter) was added to a 38.5mL Open-Top Thinwall Ultra-Clear Tube (Beckman Coulter, 344058) before slowly layering 33mL viral supernatant. Hence, following this protocol, each ultracentrifuge tube requires viral harvest from two 15cm plates and knowing this, virus production could be scaled accordingly. Ultracentrifuge tubes were added to rotor buckets and loaded into an Optima™ XPN-80 (Beckman Coulter, A95765) with SW 32 Ti Swinging-Bucket Rotor (Beckman Coulter, 369694) and centrifuged at 24,000 rpm for 2-hours at 4°C. Prior to loading, ultracentrifuge tubes were weighed against each other and volumes adjusted as necessary to ensure the centrifuge was correctly balanced.

Following centrifugation, supernatant was poured into 2% Virkon under laminar flow, while viral pellets remain at the bottom of ultracentrifuge tubes. Viral pellets were resuspended in PBS with volumes ranging between 100uL-400uL depending on expected viral titre. Ultracentrifuge tubes were then covered in parafilm for sterility and incubated on ice for 2-hours with agitation every 20-minutes by vortexing. After incubation, virus was aliquoted and snapped frozen on dry ice before long-term storage at -80°C.

Determining lentiviral titre

To determine lentiviral titre, a functional titre approach was used. This method was favoured over a physical titre, due to the increased accuracy of a functional titre to predict viral particles capable of cellular transduction. All viral constructs contained a fluorescent reporter allowing titre to be determined by flow cytometry.

Depending on estimated viral titre, HEK293T cells were transduced with the appropriate dilution series of lentiviral particles. For highly concentrated virus in the range of 1×10^8 to 1×10^9 transducing units/mL (TU/mL), dilutions 1:100, 1:200, 1:400, 1:800, 1:16000 provided robust titre data. The day prior to transduction, HEK293T cells were seeded into a 12-well plate at 0.9×10^6 cells per well in 1mL complete DMEM media. Sufficient wells were seeded for each dilution per lentiviral construct to be run in duplicate. After seeding, cells were returned to the 37°C incubator with 5% CO₂ until the following day.

Twenty-four hours later, lentivirus was thawed and the appropriate dilution series prepared in complete DMEM including 8ug/mL polybrene (Sigma Aldrich, 107689). Polybrene is a cationic polymer which neutralizes the electrostatic repulsion between lentiviral particles and the negatively-charged cell surface leading to improved transduction efficiency. Media was removed from HEK293T cells seeded the previous day in 12-well plates, and 500uL of each viral dilution gently added to each appropriate well. With each batch of titrations, one well was transduced with a “standard” control where the viral titre was known. This standard helped to resolve any discrepancies between expected and actual viral titres. An additional two wells were required to determine the average cell count, which was used to calculate the viral titre. Following transduction, plates were returned to 37°C incubator with 5% CO₂ for 96-hours.

Ninety-six hours post-transduction, media was removed, cells trypsinized and resuspended in FACS buffer (PBS + 10% FBS) including 1:10,000 DAPI (BD Biosciences, 564907) for acquisition on the BD LSRFortessa Analyzer (BD Biosciences). Live (DAPI-negative), transduced (fluorescence-positive) populations were gated, with gates set using untransduced and single-stained controls. Finally, viral titre in transducing units (TU/mL) was calculated using **Equation 2.1**.

$$\text{Transducing Units } \left(\frac{\text{TU}}{\text{mL}} \right) = \frac{\left(\frac{\% \text{ Fluorescence}}{100} \right) \times \text{Number of Transduced Cells} \times \text{Dilution Factor}}{\text{Transduction Volume (mL)}}$$

Equation 2.1. Calculating the number of functional transducing lentiviral units

2.1.3 Lentiviral transduction

All lentiviral vectors were prepared as per Section 2.1.2. Lentiviral transfer vectors encoding shRNA sequences were purchased from Horizon Discovery and shRNA sequences are described in **Table 2.2**. The plasmid map can be found in **Appendix 4**.

shRNA ID	Target Region	shRNA Sequence (5'-3')	Clone ID (Horizon Discovery)
PCBP2 KD #1	ORF	TGCATAAGTAGCCGGATGG	V3SVHS09_4782013
PCBP2 KD #2	3'UTR	AACCTGGAATCGCTGACTG	V3SVHS09_10167844
Scrambled-Control	-	-	VSC11716

Table 2.2. shRNA lentiviral vectors

Suspension cell lines, HEL and SET-2, were transduced with lentiviral vectors. Multiplicity of infection (MOI), referring to the number of lentiviral particles expected to integrate per cell, varied depending on experimental requirements and is stated in appropriate method sections. The number of cells per transduction was dependent on experimental requirements and ranged between 0.5×10^6 and 2.0×10^6 cells. Cells were suspended in suitable complete media, as per Section 2.1.1, including 8ug/mL polybrene to improve transduction efficiency. Lentivirus was recovered from -80°C and thawed. The volume of virus (TU_{TOTAL}) was calculated based on the viral titre (TU/mL) as per **Equation 2.2**.

$$\text{TU}_{\text{TOTAL}} = \frac{\text{MOI} \times \text{Cell Number}}{\text{Viral Titre (TU/mL)}}$$

Equation 2.2. Calculating virus volumes for lentiviral transduction

Virus was added to cell suspension and plates centrifuged at 1000 x *g* for 30-minutes at 35°C. Cells were then incubated at 37°C with 5% CO₂ and 3-6 hours after transduction, 50% additional complete media was added to cultures. The following day, viral supernatant was removed and replaced with fresh media to avoid undue cell toxicity.

2.1.4 Quantitative real time polymerase chain reaction

Total RNA extraction

For RNA extraction from cell lines, a RNeasy Mini Kit (Qiagen, 74104) was used as per the manufacturer's instructions. RNA was quantified using a Qubit RNA Quantification, Broad Range (BR) kit (Invitrogen, Q10210) and Qubit Fluorometer 2.0 before storing at -80°C.

Reverse Transcription Polymerase Chain Reaction (RT-PCR)

To prepare complementary DNA (cDNA) for quantitative PCR from cell lines, a High-Capacity cDNA Reverse Transcription Kit (ThermoFisher, 4368813) was used as per the manufacturer's instructions. Per sample, RNA input was normalised, most frequently between 500-1000ng, in RNase/DNase-free H₂O to a final volume of 10uL. A master mix including Reverse Transcriptase, dNTP mix and random primers was prepared as per **Table 2.3** and 10uL added per sample to a final volume of 20uL. Thermocycler conditions for cDNA synthesis are detailed in **Table 2.4**. After synthesis, cDNA was diluted 1:10 with DNase/RNase-free H₂O and stored at -20°C. To prevent degradation, RNA was kept on ice throughout reaction preparation.

Reagent	Volume per Sample (uL)	Catalog Number
10X RT Buffer	2	High Capacity cDNA Reverse Transcription Kit (Invitrogen, 4368813)
25X dNTP Mix	0.8	
10 RT Random Primers	2	
MultiScribe Reverse Transcriptase	1	
RNase/DNase-Free Water	4.2	
Total	10	

Table 2.3. Master mix for cell line cDNA synthesis

Temperature (°C)	Time
25	10 mins
37	120 mins
85	5 mins
4	∞

Table 2.4. Thermocycler conditions for cell line cDNA synthesis

Quantitative polymerase chain reaction (qPCR)

To quantitatively analyse mRNA levels on a per sample basis, amplification of cDNA was complete using TaqMan probes. A master mix was prepared per TaqMan probe, including TaqMan Universal PCR Master Mix (ThermoFisher, 4304437) (**Table 2.5**). A comprehensive list of probes used throughout this project are detailed in **Table 2.22**.

Reagent	Volume per Reaction (uL)	Catalog Number
TaqMan Universal PCR Master Mix	7.5	4304437 (ThermoFisher)
TaqMan Probe	0.75	(Invitrogen, detailed in Table 2.22)
RNase/DNase-Free Water	4.75	10977035 (Invitrogen)
Total	13	

Table 2.5. Master mix for qPCR

Per reaction, 13uL master mix was loaded into a MicroAmp Fast Optical 96-well Reaction Plate with Barcode (Life Technologies, 4346906) followed by 2uL cDNA. To normalise gene expression between samples, each sample was assayed for a reference gene, either *B2M* or *GAPDH*, to exclude variation arising from sample preparation and loading. Each sample was run in duplicate to provide technical replicates. Negative control wells, replacing cDNA with H₂O, were also included to identify any potential contamination. Plates were sealed and briefly centrifuged prior to loading into a StepOne Real-Time PCR System (Applied Biosystems, 4376357). Thermocycle conditions were as described in **Table 2.6** and relative gene expression calculated using the $2^{-\Delta\Delta C_t}$ method¹²⁹.

Temperature (°C)	Time	Cycles
50	2 mins	1
95	10 mins	1
95	15 sec	40
60	1 min	

Table 2.6. Thermocycle conditions for qPCR

2.1.5 Western blotting

Protein preparation and quantification

Per sample, 1-million cells were collected and centrifuged at $300 \times g$ for 5-minutes at 4°C. Supernatant was removed and cells washed twice with ice-cold PBS to remove any minor contaminants. Lysis buffer was prepared fresh, including 24.5uL RIPA buffer (Sigma-Aldrich, R0278) and 0.5uL Protease and Phosphate Inhibitor Cocktail (Merck, PPC1010) per sample and incubated for 45-minutes at 4°C. Following lysis, samples were centrifuged at $10,000 \times g$ for 15-minutes at 4°C and supernatant collected. Extracted protein was quantified using a Qubit Protein Assay Kit (Invitrogen, Q33211) and Qubit Fluorometer 2.0 as per the manufacturer's instructions

SDS-PAGE electrophoresis

To denature proteins for sodium dodecyl-sulphate polyacrylamide (SDS-PAGE) electrophoresis, 30-50ug protein was prepared per sample. Protein was mixed with 10X Bolt Sample Reducing Agent (Invitrogen, B0009) and 4X Bolt LDS Sample Buffer (Invitrogen, B0007) and incubated at 95°C for 10-minutes to denature proteins. Samples were loaded into a pre-cast Bolt 4 to 12%, Bis-Tris Mini-Protein gel, 30uL for 15-well gels (Invitrogen, NW04125BOX) and 35uL for 17-well gels (Invitrogen, NW04127BOX) and subject to SDS-PAGE. Anionic SDS associates with the denature protein based on molecular weight and protein separation therefore occurs due to varying protein migration towards the positively charged anode during electrophoresis. Proteins were run alongside 3uL Colour Pre-Stained Protein Standard (NEB, P7719) in 1X Bolt MES SDS Running Buffer (Invitrogen, B0002) for 35-minutes at 200V using a Mini Gel Tank (Invitrogen, A25977).

Membrane transfer

Following protein separation, proteins were transferred from the polyacrylamide gel onto a nitrocellulose membrane (Invitrogen, LC2000) using a wet electroblotting method. For membrane transfer, blotting pads, Whatmann filter paper and nitrocellulose membrane were submerged into transfer buffer (1X Tris-Glycine Transfer Buffer (CST, 12539), 70% H₂O and 20% methanol). Following submersion, assembly of the transfer sandwich into

the blot module was as follows: blotting pad, x1 Whatmann filter paper, polyacrylamide gel, nitrocellulose membrane, x1 Whatmann filter paper, blotting pad. Order of assembly ensured the nitrocellulose membrane was placed between the positive cathode and polyacrylamide gel, allowing proteins to pass from the gel and onto the membrane by electrophoretic transfer. Proteins were transferred at 10V for 60-minutes in 1X transfer buffer using a Mini Gel Tank as above.

Immunoblotting

Transfer sandwiches were disassembled and membranes blocked with 5% milk in 0.1% TBS-T for 60-minutes on a shaker at RT. TBS-T formulation was 1X TRIS-buffered saline (TBS; ThermoFisher, J62780-K3) in 90% H₂O with 0.1% Tween20. Following blocking, membranes were incubated with primary antibodies against PCBP2, or BCL-X_L overnight at 4°C with gentle agitation. Membranes were then washed 3-times with 0.1% TBS-T prior to incubation with horseradish peroxidase (HRP)-conjugated secondary antibody for 1-hour at RT with agitation. Following incubation, membranes were again washed with 0.1% TBS-T three times. To detect HRP-conjugated antibodies, membranes were covered with SuperSignal West Pico PLUS Chemiluminescent Substrate (ThermoScientific, 34577) for 5-minutes and developed using an ChemiDoc MP Imaging System (Bio-Rad). After visualisation, immuno-detection was repeated for loading control, tubulin. HRP-conjugated alpha-tubulin antibody was incubated with membranes for 1-hour at RT with gentle agitation. Membranes were washed five times with 0.1% TBS-T before development as above. Protein expression was quantified using ImageJ software (version 1.53k) and normalised to tubulin expression. Primary and secondary antibodies are described in **Table 2.7**.

Antibody	Host Species	Dilution	Catalog Number
PCBP2	Rabbit	1:500	Abcam, ab184962
BCL-X _L	Rabbit	1:1000	Cell Signaling Technology, 2764T
Alpha-Tubulin, HRP-conjugated	Rabbit	1:2000	Cell Signaling Technology, 9099
Alpha-Tubulin, HRP-conjugated	Mouse	1:2000	Cell Signaling Technology, 12351S
Anti-Rabbit IgG, HRP-conjugated, secondary	Rabbit	1:10,000	Cell Signaling Technology, 7074S
Anti-Mouse IgG, HRP-conjugated, secondary	Mouse	1:10,000	Cell Signaling Technology, 7076S

Table 2.7. Primary and secondary antibodies

2.1.6 RNA-binding protein targeted CRISPR screening

Genome engineering using Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9

Genome editing of precise loci via CRISPR-Cas9 has become an incredibly powerful technique in molecular biology. Adapted from the microbial adaptive immune system, the CRISPR-Cas9 system utilises RNA-guided nucleases to cleave targeted genomic loci. The Type II CRISPR system, derived from *Streptococcus pyogenes* and adapted for use in eukaryotic cells, comprises two main components; Cas9-nuclease and a crRNA: tracrRNA duplex¹³⁰. crRNA encodes a 20-nucleotide sequence complementary to target DNA and the double-stranded tracrRNA is necessary for tethering the duplex to Cas9 for double-stranded cleavage of target loci¹³⁰. Synthetically engineered sgRNA comprise key features of the crRNA: tracrRNA duplex creating a simple two-component system for genome editing comprising Cas9 and sgRNA.

The 20-nucleotide crRNA sequence binds to complementary DNA immediately preceding a 5'NGG PAM site and Cas9-nuclease subsequently cleaves genomic DNA ~3bp upstream of the PAM creating a double strand break (DSB). Following DNA damage, DNA is repaired by one of two major DNA repair pathways; non-homologous end joining (NHEJ) or homology-directed repair (HDR). In the absence of a repair template for high-fidelity HDR, DSB are repaired via the error-prone NHEJ pathway. NHEJ leads to insertion/deletion mutations, creating frameshift mutations and premature stop codons constituting a gene knockout (**Figure 2.1**).

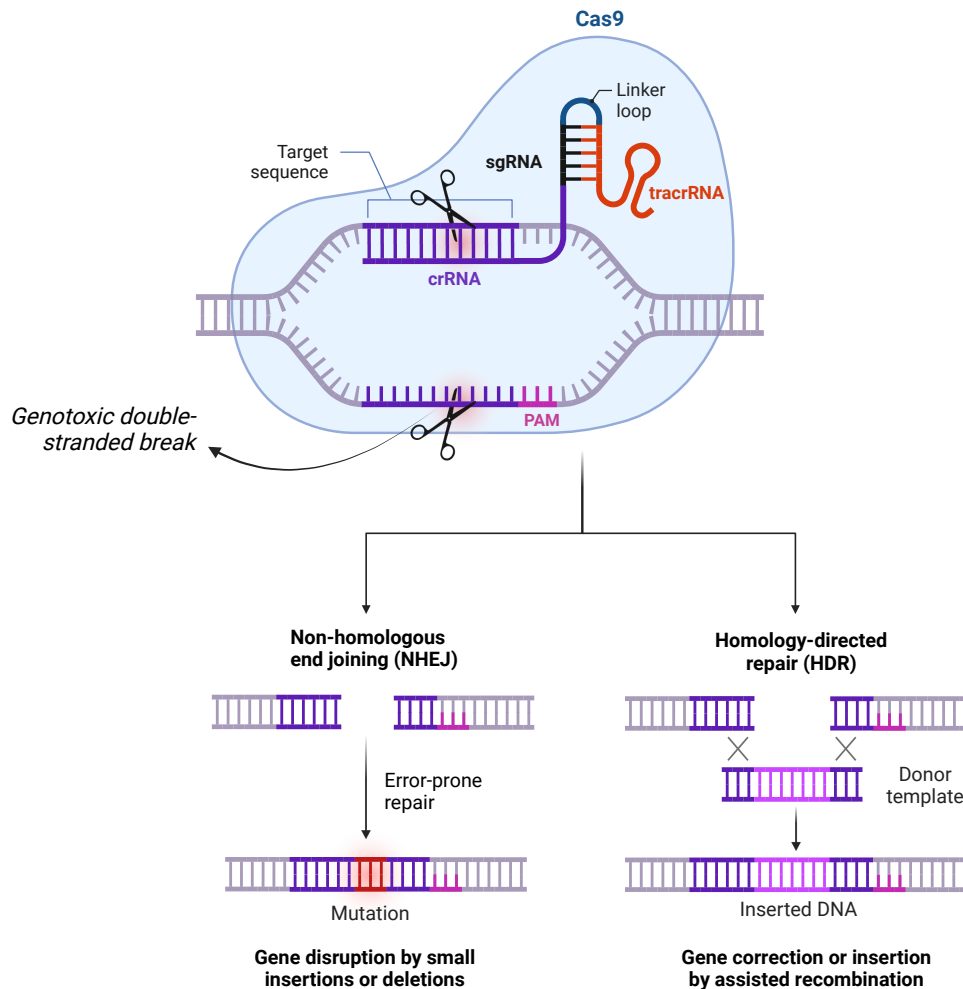


Figure 2.1. CRISPR-Cas9 structure and pathways for genome engineering. Single-guide RNA (sgRNA) comprises a 3' double-stranded tracrRNA (red) for tethering to Cas9 and a 20-nucleotide crRNA sequence (purple) complementary to target loci adjacent to a PAM sequence (pink). sgRNA binding to complementary DNA directs DNA cleavage by Cas9. In the absence of a donor template for homology-directed repair (HDR), double-stranded breaks are repaired via non-homologous end joining (NHEJ), an error prone pathway leading to insertion/deletion mutations at target loci. Illustration created using BioRender.com.

CRISPR screen design

Guides were selected based on MinLibCas9, an optimised, minimal genome-wide CRISPR-Cas9 library¹³¹. Per RBP, three sgRNA were selected. A smaller number of sgRNA targeting positive control genes were selected than usually required due to the predicted essentiality of target genes and inclusion of common positive controls within the target gene cohort (i.e., *RPS14*, *RPS27*). Negative controls included guides targeting genomic regions where editing is known to have no effect such as the *AAVS1* locus or defined

intronic and intergenic regions. Non-targeting sgRNA targeted genes absent from the human genome, for example the *LacZ* loci. A list of sgRNA sequences and the corresponding target gene can be found in **Table 2.8**. Overall, the screen included a total of 61 sgRNA targeting 22 genes, 12 of which were RBP-of-interest.

For transduction of cell lines, an all-in-one plasmid was selected (Addgene plasmid #57827) including sgRNA, Cas9 and fluorescent marker (GFP), initially described by Heckl *et al*¹³² (**Appendix 3**). To satisfy RNA Polymerase III requirements for transcription from the plasmid backbones U6 promoter, guide sequences were adapted to start with one, but ideally two guanine (G) nucleotides¹³³. Guides adapted to meet RNA Polymerase III requirements are specified within **Table 2.8**. Cloning of the sgRNA plasmid pool was complete by Dr Philip Hublitz, previous Head of the Genome Engineering Services at the Weatherall Institute of Molecular Medicine, University of Oxford.

Cell line screening

Prior to screening, lentivirus was titrated in model HEL and SET-2 cell lines to determine the multiplicity of infection (MOI) necessary for 20-30% of viable cells to be transduced. A transduction efficiency within this range ensures only one sgRNA is incorporated per cell. Cells were transduced as per Section 2.1.3.

After titration, cells were transduced at 20,000X coverage (cells/sgRNA). This was excessive, however, due to the small number of guides within the library, the number of cells transduced was scaled up for experimental ease. As a rough estimate of cell starting number for transduction, the number of sgRNA (n=61) was multiplied by the desired coverage (20,000X). An extra 70-80% cells were then included to account for the desired 20-30% transduction efficiency plus losses during transduction, sorting and replating.

Target Gene	Sequence Modification (5')	sgRNA Sequence (5'-3')	Sequence Modification (3')	sgRNA Subset
SRSF11		GGACGATACTCACTCCGGC		Target RBP
SRSF11	G	GAAC TAGTACCTGGGTAAG		Target RBP
SRSF11		GGTCAAGAAGCAGATCGAGA		Target RBP
SFPQ	GG	CATCCTCCGTGATATCAGC		Target RBP
SFPQ	G	GCCGCCTCGATGCGCGGCT		Target RBP
SFPQ		GGAAGTCGTGGAGGCCGCG		Target RBP
PRPF40A	G	GAGTTATGTCTCAATGC		Target RBP
PRPF40A		GGCATGCATTATGCCCAAT		Target RBP
PRPF40A		GGGACGGGAGCTGACCGTGG		Target RBP
HNRNPM	G	GACCAATGCACGTCAAGA		Target RBP
HNRNPM		GGCGACGGAGATCAAAATGG		Target RBP
HNRNPM		GGCATGGAGCGCATGGGCGC		Target RBP
RPS28	GG	CTTTCCTCCAGTCAACA		Target RBP
RPS28	G	GTAGGTGCGCGTGGAAATTCA		Target RBP
RPS28		GGGGACCCGTAGGTGCGCG		Target RBP
RPS17	G	GCTCCGCAACAAGATAGCA		Target RBP
RPS17	GG	TATCCAGAAAGTTTACCTC		Target RBP
RPS17		GGATGAATTTCAAAACGCT		Target RBP
ILF3	G	GTCCGACTGGATAGACGAGC		Target RBP
ILF3		GGAACCACTTGGCGTGTCCGG		Target RBP
ILF3		GGCGTGTGGAGGGACGCCA		Target RBP
PCBP2	GG	TTGTCAATGATCATAGCAA		Target RBP
PCBP2	GG	TTTACTCCAAAGAGTACA		Target RBP
PCBP2		GGCTTGGGCGGTACGGGA		Target RBP
YBX1		GGCTTTTGTAGGTGCGG		Target RBP
YBX1	G	GTCTTGCAAGGAATGACACCA		Target RBP
YBX1		GGACCATACTCGGGAATCG		Target RBP
YARS1	G	GAGTGGCTCTTATACCC		Target RBP
YARS1	GG	ACCACGGGCAACACATG		Target RBP
YARS1		GGAGTTCTAGAAGTTCCCAT		Target RBP
SRSF10	GG	CTTACTCTTTCGATCCCC		Target RBP
SRSF10	GG	CTGACGAACAGAGCGTGT		Target RBP
SRSF10		GGCGAGTGTAGAAATCAAG		Target RBP
RBBP6	GG	TAAGTCGCAGAGGAGATG		Target RBP
RBBP6	GG	CTGCTTCTTTAAGTCGCAG		Target RBP
RBBP6		GGAGATGTGGAGCCCATCAA		Target RBP
AAVS1		GGGGCCACTAGGGACAGGAT		Negative Control
AAVS1		GGGAGGGAGAGCTTGGCAGG		Negative Control
Intergenic	G	GACGAGGATGTATTGGGG		Negative Control
Intergenic	G	GCACATGTGCAAAACGTAGG		Negative Control
Intergenic		GGGGTAAGCTTACAAATCTC	C	Negative Control
Intergenic		TTCTATCATGTAGACTCTTC	C	Negative Control
Intronic	GG	AAGACTGTTAAGGATAGCT		Negative Control
Intronic		GGACAAAATCCGGAAGGCTA		Negative Control
Intronic		GGCAATGGTTGCTACTTATA		Negative Control
Intronic	G	GTGTTGACCGTAGGATGGGT		Negative Control
LacZ	G	GCGCGTCGTGATTAGCGCCG		Non-Targeting Control
LacZ		GGGCGGCTTCGTCTGGGAC		Non-Targeting Control
LacZ		GGTGAAATTATCGATGAGCG		Non-Targeting Control
Luciferase		GGTAACCCAGTAGATCCAG		Non-Targeting Control
Luciferase	G	GACCAACGCCCTTGATTGACA		Non-Targeting Control
Luciferase		GGCATGCGAGAATCTGACGC		Non-Targeting Control
Luciferase		GGCTATGAAGAGATACGCC		Non-Targeting Control
Yeast		GGAGTAAAACCTCTACAAATG		Non-Targeting Control
Aquaporin 1	GG	TGATTATGATCTCTAGTCA		Non-Targeting Control
POLR2A		GGTCGTGTCGGAGCTAA		Essential Control
POLR2A	GG	TCCAGGATACATCCAGCGG		Essential Control
TUBB	GG	ATTGTAGTACACAGAGATG		Essential Control
TUBB		GGTCAGTCTGGGCGAGGTA	CC	Essential Control
RPL14	G	GTGTTCTAGGCTTTGGTCTGA		Essential Control
RPL14	GG	CATCTCAAGTTTCCGCAC		Essential Control

Table 2.8. Target genes and sgRNA sequences for RBP-targeted CRISPR screen

Forty-eight hours after transduction, transduced cells were sorted via fluorescently activated cell sorting (FACS) to maintain 20,000X coverage in culture ($\sim 1.22 \times 10^6$ cells). Cells were collected for DNA extraction on days 0, 7, 14 and 21 at 150 to 500-fold library representation^{131,134}. Coverage for library representation was kept consistent for technical replicates across timepoints but varied between independent experiments.

Genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen, 69504), as per the manufacturer's instructions, and eluted in RNase/DNase-free H₂O. A nested PCR approach was used for library preparation where PCR1 enriched for sgRNA regions integrated into the genome and PCR2 further amplified target sequences and ligated P5/P7 adaptors for Illumina sequencing. PCR1 forward primers included a 3' stagger sequence to increase diversity during sequencing since PCR1 primers annealed the ubiquitous U6 promoter sequence upstream of the diverse sgRNA sequences. PCR1 primers also included sequences complementary to PCR2 primers for subsequent annealing (**Table 2.9**). PCR1 reactions were prepared as described in **Table 2.10** and thermocycler conditions in **Table 2.11**. PCR1 products were cleaned up using AmPure XP beads (Beckman Coulter, A63882) at a 1:1 ratio, as per the manufacturer's instructions. DNA was then quantified using a dsDNA Quantitation, Broad Range kit (ThermoFisher, Q32850) and Qubit Fluorometer 2.0, as per the manufacturer's instructions.

Primer	Sequence (5'-3')
PCR1 F1	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG ttgtggaaggacgaaacaccg
PCR1 F1 (+A)	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG Attgtggaaggacgaaacaccg
PCR1 F1 (+GA)	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG GAttgtggaaggacgaaacaccg
PCR1 F1 (+CGA)	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG CGAttgtggaaggacgaaacaccg
PCR1 F1 (+TCGA)	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG TCGAttgtggaaggacgaaacaccg
PCR1 F1 (+ATCGA)	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG ATCGAttgtggaaggacgaaacaccg
PCR1 R1	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG gccatatctgggaccatctgt

Table 2.9. PCR1 primers sequences for RBP-targeted CRISPR screening libraries.
Upper case nucleotides are complementary to PCR2 primers, orange nucleotides represent stagger sequence and bold lower-case nucleotides anneal ubiquitous U6 promoter region integrated into gDNA

Reagent	Volume per Reaction (uL)	Catalog Number
5X KAPA2G Robust HotStart ReadyMix	12.5	KK5702 (Sigma, Roche)
10uM Forward Primer Mix	1.25	Custom designed primers described in Table 2.9 (Invitrogen)
10uM Reverse Primer	1.25	
Template gDNA*	As required (up to 10uL)	-
PCR-grade H ₂ O	Up to 25uL	10977035 (Invitrogen)

Table 2.10. PCR1 reaction mix for CRISPR screening library preparation.
*Template gDNA for PCR1 representing 150-500 cells/sgRNA equates to 55-183ng gDNA (given 167 cells = 1ng DNA)

PCR1			
Step	Temperature (°C)	Time	Cycles
Initial Denaturation	98	3 min	
Denaturation	95	20 sec	30 cycles
Annealing	63	15 sec	
Extension	72	1 min	
Final Extension	72	10 min	
Cooling	4	∞	

Table 2.11. PCR1 thermocycler conditions

During PCR2, Nextera XT adapters (Illumina, 15055294) were then ligated to PCR1 products for next generation sequencing. Nextera XT adaptors included 5' P5 (5'-AATGATACGGCGACCACCGAGATCTACAC-3') and P7 (5'-CAAGCAGAAGACGGCATACGAGAT-3') sequences for binding to the flow chamber followed by variable 8bp i5 and i7 indexes, respectively, to uniquely barcode each sample. Indexes used for CRISPR screening are described in **Table 2.12**.

P7 Primer	Primer Sequence (5'-3')
P7-index1	CAAGCAGAAGACGGCATACGAGATTAAGGCGAGTCTCGTGGGCTCGG
P7-index2	CAAGCAGAAGACGGCATACGAGATCGTACTAGGTCTCGTGGGCTCGG
P7-index3	CAAGCAGAAGACGGCATACGAGATAGGCAGAA GTCTCGTGGGCTCGG
P7-index4	CAAGCAGAAGACGGCATACGAGATTCCTGAGCGTCTCGTGGGCTCGG
P7-index5	CAAGCAGAAGACGGCATACGAGATGGACTCCTGTCTCGTGGGCTCGG
P7-index6	CAAGCAGAAGACGGCATACGAGATTAGGCATGGTCTCGTGGGCTCGG
P7-index7	CAAGCAGAAGACGGCATACGAGATCTCTACGTCTCGTGGGCTCGG
P7-index8	CAAGCAGAAGACGGCATACGAGATCGAGGCTGGTCTCGTGGGCTCGG
P7-index9	CAAGCAGAAGACGGCATACGAGATAAGAGGCA GTCTCGTGGGCTCGG
P7-index10	CAAGCAGAAGACGGCATACGAGATGTAGAGGAGTCTCGTGGGCTCGG
P7-index10	CAAGCAGAAGACGGCATACGAGATGCTCATGAGTCTCGTGGGCTCGG

P5 Primer	Primer Sequence (5'-3')
P5-index1	AATGATACGGCGACCACCGAGATCTACACCTCTATTCGTCGGCAGCGTC
P5-index2	AATGATACGGCGACCACCGAGATCTACACTATCCTCTTCGTCGGCAGCGTC
P5-index3	AATGATACGGCGACCACCGAGATCTACACAGAGTAGATCGTCGGCAGCGTC
P5-index4	AATGATACGGCGACCACCGAGATCTACACACTGCATATCGTCGGCAGCGTC
P5-index5	AATGATACGGCGACCACCGAGATCTACACAAGGAGTATCGTCGGCAGCGTC
P5-index6	AATGATACGGCGACCACCGAGATCTACACCTAAGCCTTCGTCGGCAGCGTC

Table 2.12. PCR2 primer sequences for CRISPR screening libraries. Blue nucleotides represent Nextera XT adapter sequences for binding to the flow chamber. Black nucleotides represent 8bp variable indexes and green sequences are complementary to PCR1 primers for ligation.

To ligate Nextera XT primers (Illumina, 15055294) to PCR1 products for sequencing, 50ng of PCR1 product was added to the reaction mix in **Table 2.13** using the thermocycler conditions described in **Table 2.14**. PCR2 products were again cleaned-up using 1:1 ratio of DNA product: AmPure XP beads prior to quantification. Following sample

quantification, sample molarity was determined and samples pooled to generate a 4nm pool for sequencing. As a final quality check, library size was confirmed using a TapeStation D1000 with Agilent D1000 ScreenTape (5067-5582) and reagents (5067-5583). Three-hundred cycle paired-end sequencing was performed using a MiSeq Micro Kit V2 (Illumina, #MS-102-2002), including PhiX in sample sequencing accounting for 10% of total reads.

Reagent	Volume per Reaction (uL)	Catalog Number
DNA (50ng)	As required	-
Nextera i7 Index	5ul	15055294 (Illumina)
Nextera i5 Index	5ul	
5X KAPA2G Robust HotStart ReadyMix	25ul	KK5702 (Sigma, Roche)
PCR-grade H ₂ O	Up to 25uL	10977035 (Invitrogen)

Table 2.13. PCR2 reaction mix for CRISPR screening library preparation

PCR2			
Step	Temperature (°C)	Time	Cycles
Initial Denaturation	95	3 min	6 cycles
Denaturation	95	30 sec	
Annealing	55	30 sec	
Extension	72	30 sec	
Final Extension	72	5 min	
Cooling	4	∞	

Table 2.14. PCR2 thermocycler conditions

2.1.7 CRISPR-mediated gene knockout in cell lines

To achieve gene knockout (KO) in cell lines, chemically synthesised sgRNA and recombinant Cas9 protein were precomplexed to form a ribonucleoprotein (RNP) and delivered into cells via nucleofection. For editing HEL and SET-2 cell lines, the Cell Line Nucleofector Kit V (VA-1003, Lonza) was used. Briefly, 1-million cells per sample were collected and washed twice with PBS. During washing, 3ug Alt-R S.p Cas9 Nuclease V3 (1081059, Integrated Technologies) was mixed with 1.6ug 2'O-methyl 3'phosphorothioate modified sgRNA (Synthego) and incubated at 21°C for 10-minutes to form RNP. Single-guide RNA were modified to include 2'O-methyl and 3' phosphorothioate internucleotide linkages to improve RNA stability. Cell pellets were resuspended in 100uL Nucleofector solution (82uL Nucleofector solution plus 18uL

Supplement), mixed with RNP complexes and transferred to a 100uL Nucleocuvette. Cells were electroporated using the Nucleofector II/2b Device (Lonza) using program X-05. After nucleofection, 100uL pre-warmed cell line-specific media, lacking antibiotics, was added to the Nucleocuvette and cells left to recover for 10-minutes. Following recovery, cells were collected and cultured at 1×10^6 /mL and incubated at 37°C with 5% CO₂.

For all cell line experiments, cells were edited using one of two sgRNA targeting *PCBP2* in parallel with *AAVS1* control sgRNA. *AAVS1* is a safe harbour site, found within intron 1 of *PPP1R12C*, where editing is not expected to affect gene function. sgRNA sequences can be found in **Table 2.15**.

sgRNA	sgRNA Sequence (5'-3')
PCBP2 sgRNA-1	GCTTTGAAGATGGCATTAGT
PCBP2 sgRNA-2	TTTGTCAATGATCATAGCAA
AAVS1 sgCTL	ggggccactagggacaggat

Table 2.15. *PCBP2* and *AAVS1* sgRNA sequences for RNP nucleofection

2.1.8 Annexin V/DAPI apoptosis assay

The effect of *PCBP2*-KD on cell viability was investigated using an Annexin V Apoptosis Detection Kit (eBioscience, 88-8007-74). Using this assay, early and late apoptosis can be defined. Annexin V binds phosphatidylserine, a phospholipid which presents on the cell membrane only during apoptosis, allowing cells in early apoptosis to be identified. During later stages of apoptosis, cell membrane integrity degrades and a DNA stain, such as DAPI, can enter the cell and bind dsDNA. Upon cell death, complete loss of membrane integrity during necrosis prevents phosphatidylserine presentation and hence cells can only be positive for the DNA stain (**Figure 2.2**).

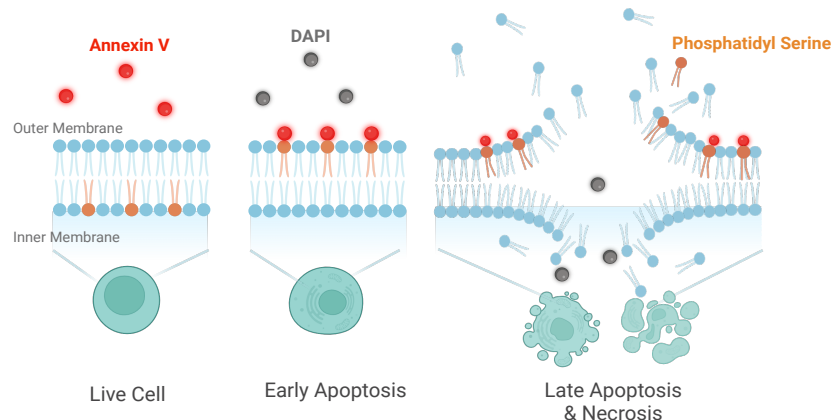


Figure 2.2. Principles of apoptosis Annexin V/DAPI assay.

Up to 1-million cells were transduced with lentiviral shRNA vectors targeting *PCBP2* or a scrambled control at an MOI 5 for Annexin V/DAPI apoptosis assays. The effect of *PCBP2*-KD on HEL and SET-2 cell line apoptosis was investigated 5-days post-transduction. Briefly, cells were collected and resuspended in 1mL Annexin V Binding Buffer (eBiosciences, BMS500BB) and stained with 1:100 Annexin V-APC (Invitrogen, 17-8007-74) for 15-minutes at RT in darkness. To stop the reaction, 4mL Annexin V Binding Buffer was added per sample and cells centrifuged at 500 x *g* for 5-minutes. Supernatant was removed and samples resuspended in 200mL Annexin V Binding Buffer including 1:10,000 DAPI. Samples were acquired via flow cytometry with gating as illustrated in **Figure 2.3**.

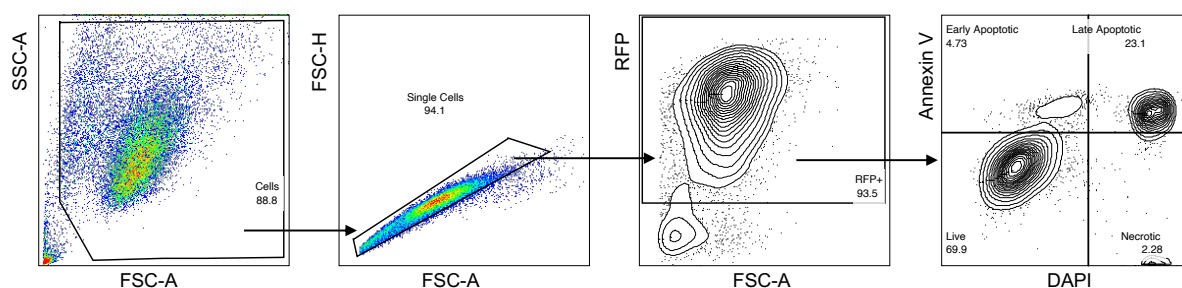


Figure 2.3. Flow cytometry gating strategy for Annexin V/DAPI apoptosis assay.

2.1.9 Competitive cell growth assay

Competitive cell growth (CCG) assays were used to determine the effect of *PCBP2*-KD on long-term cell proliferation compared to non-transduced cell populations in liquid co-

culture¹³⁵. Transduced populations can be identified via the expression of the fluorescent marker, RFP, and therefore if PCBP2 is required for cell proliferation we expect *PCBP2*-KD (RFP-positive cells) to be lost from culture overtime. Cultures were maintained over 35-40 days and loss of RFP-positive cells over time was assayed by flow cytometry.

Cells were transduced at MOI 0.5 for competitive cell growth assays. A lower MOI than used for apoptosis assays ensured transduction efficiency was <90%, allowing a sufficient non-transduced population to be present and competitive in culture against transduced populations. Three-days following transduction, cells were seeded at 0.4×10^6 /mL in 3mL media and passaged every 3-4 days. On passage days, 500uL cell suspension was collected for assaying RFP expression via flow cytometry using a BD LSRFortessa (BD Biosciences) analyser as per the gating strategy in **Figure 2.4**.

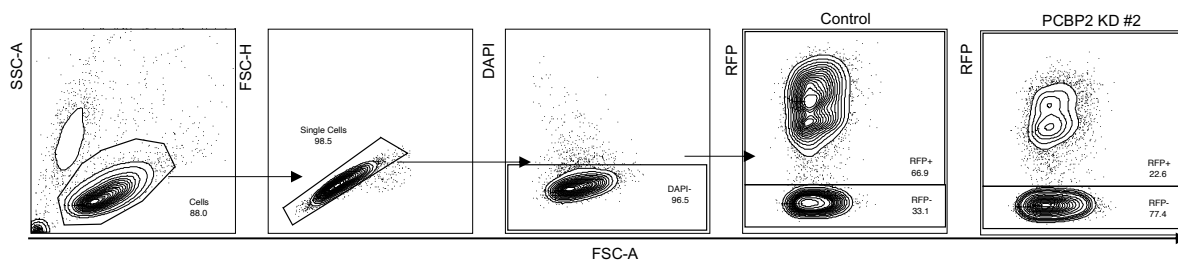


Figure 2.4. Flow cytometry gating strategy for competitive cell growth assays.

2.1.10 Bulk cell editing assay

HEL and SET-2 cells were edited using one of two sgRNA targeting *PCBP2* or with *AAVS1* control sgRNA as described in Section 2.1.7. Following editing, cells were cultured at 0.4×10^6 cells/mL and passaged every 3-4 days. DNA was extracted on day 7 and day 21 using the DNeasy Blood and Tissue Kit (Qiagen, 69506) as per the manufacturer's instructions. Following DNA extraction, 50ng DNA per sample was amplified, using the appropriate *PCBP2* or *AAVS1* primer pair (**Table 2.16**). DNA was added to the PCR mix described in **Table 2.17**, and amplified using thermocycle conditions described in **Table 2.18**. Following amplification, PCR products were cleaned up using a 1:1 ratio of AmPure XP beads, quantified and sent for Sanger sequencing at the WIMM Sequencing Facility. Chromatogram traces were analysed using the Synthego ICE tool (Synthego Performance Analysis, ICE Analysis, 2019. v3.0. Synthego), which compares edited sequences to wild-type control traces to infer percentage editing.

Primer	Primer Sequence (5'-3')
PCBP2 F1	tgctgtaggctatctttgagga
PCBP2 R1	aagtccatggcaaaactcaaactg
AAVS1 F1	CTTGCTTTCTTTGCCTGGAC
AAVS1 R1	GATCAGTGAAACGCACCAGA

Table 2.16. PCR primers for amplification of sgRNA edited loci

Reagent	Volume per Reaction	Catalog Number
5X KAPA2G Robust HotStart ReadyMix	12.5 uL	KK5702 (Sigma-Aldrich)
10uM Forward Primer	1.25 uL	Custom oligonucleotides (Table 2.16, Invitrogen)
10uM Reverse Primer	1.25 uL	
Template gDNA (50ng)	-	-
PCR-grade H ₂ O	up to 25 uL	10977035 (ThermoFisher)

Table 2.17. Reaction mix for PCR amplification of edited loci

Step	Temperature (°C)	Time	Cycles
Initial Denaturation	98	3 min	35
Denaturation	95	20 sec	
Annealing	63 (PCBP2) // 56 (AAVS1)	15 sec	
Extension	72	1 min	
Final Extension	72	1 min	
Cooling	4	∞	

Table 2.18. Thermocycler conditions for PCBP2 and AAVS1 amplification

2.1.11 Evaluation of dsDNA: RNA hybrids via S9.6 slot blot

Slot blots were performed in collaboration with Dr Chiara Beghè in Professor Natalie Gromak's laboratory at the Sir William Dunn School of Pathology, University of Oxford

Relative changes in dsDNA: RNA hybrid abundance in control vs. PCBP2-KD HEL cells was investigated. Cells were transduced with lentiviral PCBP2-shRNA or scrambled control as described in Section 2.1.3. Two-million HEL cells were transduced per construct at MOI 5.

Nucleic Acid Extraction

Three-days post-transduction, samples were collected and washed twice with 1mL PBS, centrifuging at 500 x *g* for 5-minutes at 4°C. Pellets were resuspended in 800uL cell lysis buffer (**Table 2.19**) in a clean 1.5mL Eppendorf tube and incubated on ice for 10-minutes.

Reagent	Final Concentration	Catalog Number
KCl (Potassium Chloride)	85 mM	AM9640G (Invitrogen)
Pipes pH 8.0	5 mM	15444009 (Thermo Scientific Chemicals)
NP-40	0.50%	28324 (Thermo Scientific)
RNase/DNase-free H ₂ O	up to final volume	10977035 (Thermo Scientific)

Table 2.19. Cell lysis buffer for S9.6 slot blot

Nuclei were then pelleted and resuspended in 800uL nuclear lysis buffer (**Table 2.20**) using a pipette to produce a very dense DNA solution and incubated on ice for 10-minutes. To digest protein, 5uL proteinase K (Roche, 031158001) was mixed with the nuclear lysis for 3-hours at 55°C. After 1-hour, an additional 3uL proteinase K was added, and a further 2uL added after 2-hours. After 3-hours if the DNA solution remained dense, incubation at 55°C was extended and solution liquified by manual pipetting.

Reagent	Final Concentration	Catalog Number
Tris-HCl pH 8.0	50 mM	15893661 (Thermo Scientific)
EDTA pH 8.0	5 mM	324506 (Millipore)
SDS (Sodium dodecyl sulfate)	1%	28364 (Thermo Scientific)
RNase/DNase-free H ₂ O	up to final volume	10977035 (Thermo Scientific)

Table 2.20. Nuclear lysis buffer for S9.6 slot blot

To precipitate dsDNA, 5M potassium acetate (Sigma, P1190) was added to each sample and mixed by inversion. The resulting milky white solution was then centrifuged at 10,000 $\times g$ for 5-minutes at 4°C where a white precipitate forms. Supernatant was collected, transfer to a clean 1.5mL Eppendorf tube and mixed with 800uL isopropanol by inversion. At this point, the DNA becomes visible in suspension. DNA was pelleted by centrifugation at max speed for 20-minutes at 4°C before being gently washed with 1mL 75% ethanol, mixing by inversion. DNA was centrifuged at 10,000 $\times g$ for 10-minutes at 4°C and supernatant removed without disrupting the pellet. The DNA pellet was left to airdry before being dissolved in 100uL RNase/DNase-free H₂O and transferred to a clean 1.5mL Eppendorf tube. Extracted DNA was quantified via nanodrop prior to S9.6 immunoprecipitation.

S9.6 immunoprecipitation

The S9.6 antibody is used for the high-specificity detection of DNA: RNA hybrids. To confirm S9.6 detection is RNA-DNA hybrid specific, control samples are prepared including RNase H, an enzyme with the capacity to resolve hybrids. Hence, control samples including RNase H should show no, or limited, hybrid detection. Reaction mixes, containing 1ug DNA, 1X RNase H buffer and 1.7U RNase H per 1ug DNA (NEB, M0297) up to 100uL in RNase/DNase-free H₂O were prepared. Reaction mixture for RNase H-negative samples replaced RNase H with additional H₂O up to 100uL. RNase H treated samples were incubated in the thermoshaker at 500 rpm for 2.5-hours at 37°C. Untreated samples, not subjected to RNase H treatment or downstream processing, were prepared by adding RNase/DNase-free H₂O to 1ug DNA to a total of 100uL.

Prior to loading samples, 4 Whatmann filter papers and a Hybond N+ membrane (Amersham, RPN203B; 7.5cm, 11.5cm) were soaked in 1X TE (50mM Tris-HCl pH 8.0 (ThermoFisher, 15893661), 5mM EDTA pH 8.0 (Millipore, 324506) in RNase/DNase-free H₂O) and layered into the BioDot SF Cell (BR, 160). The membrane was layered on top of the filters and the vacuum for the BioDot SF Cell turned ON. The principle of the slot blot is that the vacuum pulls DNA through to the membrane, on which DNA can then be probed. Once the vacuum had started, 200uL 1X TE was washed through each cell of the slot blot using a multichannel pipette.

To load slot blot, 400uL 1X TE was added to each sample. In addition to untreated, RNase H+ and RNase H- samples, loading controls of known DNA concentration (1.5ug, 1ug and 0.5ug) were prepared. Loading controls were important to determine whether running of the slot blot apparatus and transfer of DNA to the membrane has been successful. Per sample, 200uL DNA suspension was added per allotted cell of the slot blot machine. Once samples had passed through, the membrane was washed three times with 200uL 1X TE. After washing, the nylon membrane was removed and air dried at RT for 15-minutes before being wrapped in transparent film and exposed to 0.12 J/m² UV for 2.5 minutes to cross-link DNA to the membrane and improve signal detection.

Immunodetection

Membranes were blocked with 2% milk in PBS for 1-hour at RT followed by incubation overnight with 0.02ug/uL S9.6 antibody (provided by Professor Gromak's Lab) in 2%

milk in PBS at 4°C with agitation. The following day, membranes were washed 3-times for 7-minutes with 2% milk in PBS. Membranes were then incubated with 1:3000 secondary anti-mouse in 2% milk in PBS for 1-hour at RT before repeating washing with PBS. Membranes were developed using Pierce ECL Western Blotting Substrate (ThermoFisher, 32106) as per the manufacturer's instructions. Membranes were then transferred to a western blotting cassette and developed in a dark room using CL-Exposure Film (ThermoFisher, 34089). Quantification of S9.6 intensity was determined using ImageJ 1.53k.

2.1.12 γ H2AX immunofluorescence for double-stranded breaks

Immunofluorescent images and data processing was kindly complete by Antonio Rodriguez-Romera.

γ H2AX immunofluorescence was used to investigate the prevalence of DSB in control versus *PCBP2*-shRNA KD #2 HEL cells. Given high processing power required for imaging, only *PCBP2*-shRNA KD #2 was compared to controls. HEL cells were transduced at MOI 5.

Under laminar flow, 24-well plates were coated with 1:10 Poly-L-Lysine (Sigma, P8920) diluted in tissue culture grade H₂O and incubated at RT for 15-minutes. After incubation, Poly-L-Lysine solution was removed, wells washed twice with tissue culture grade H₂O and cover slips left to dry. Once dry, 100,000 HEL cells in complete media were added per well and returned to 37°C incubator with 5% CO₂ overnight, allowing cells to settle and spread.

The following day, cells were gently washed with 500uL PBS prior to fixation in 4% paraformaldehyde (PFA) for 10-minutes. After fixation, wells were washed three-times with PBS and cells permeabilised with 0.1% Triton X-100 (Sigma-Aldrich, X100) in PBS for 5-minutes at RT. Cells were washed again and blocked for 1-2 hours with block buffer (1% BSA in PBS). After blocking, primary γ H2AX antibody (Merck, 05-636) was added 1:400 in 200uL block buffer per sample. An unstained sample contained no γ H2AX antibody. Samples were left to incubate overnight at 4°C on a rocker. Positive control cells were irradiated at 2Gy in irradiator and left to rest in the incubator for 1-hour prior to fixation.

The following day, cells were washed once with PBS and incubated for 5-minutes with 1:1000 Hoechst 33343 in 200uL PBS. Wells were washed once with tissue-culture grade H₂O and coverslips carefully mounted onto slides using ProLong Gold Antifade Mountant (Invitrogen, P36930). Slides were left overnight at 4C to allow media to harden.

Images were acquired with a Leica THUNDER (20× objective). Number of γ H2AX foci per cell were counted using a custom script and the imaging analysis software Fiji.

2.1.13 RNA-seq

Library preparation for RNA-sequencing was performed by Azenta Life Sciences

Samples for RNA-seq transduced with shRNA constructs at MOI 5 (**Table 2.2**) were collected ninety-six hours post-transduction. RFP-positive populations were sorted via FACS and a minimum of 1-million cells collected for RNA extraction using a RNeasy Mini Kit (Qiagen, 74104). RNA was quantified using a Qubit RNA Quantification, Broad Range kit (Invitrogen, Q10210) and Qubit Fluorometer 2.0. RNA integrity was checked using an Agilent D1000 ScreenTape (5067-5582) and reagents (5067-5583) on a TapeStation 4200. 2ug total RNA was prepared and shipped to Azenta for library preparation and sequencing.

Briefly, RNA sequencing libraries were prepared using the NEBNext Ultra II RNA Library Prep Kit for Illumina (New England Biolabs, E7770S) as per the manufacturer's instructions. mRNA was first enriched with oligo(dT) beads (NEB, E7490) and enriched mRNA fragmented for 15-minutes at 94°C. Complementary DNA was synthesised, end repaired and 3' ends adenylated. Universal adaptors were ligated to cDNA fragments followed by indexes during PCR amplification of libraries. The quality of cDNA traces and indexed libraries was assessed using a NGS Fragment Kit (Agilent, DNF-473-0500) on an Agilent 5300 Fragment Analyzer and quantified using a Qubit 4.0 Fluorometer. Libraries were pooled and paired-end reads of 150bp sequenced by a NovaSeq 6000 (Illumina) with a median depth of 60-million reads per sample.

2.2 Primary patient assays

2.2.1 Processing of human samples

Samples were banked under the ethical approval of the INForMeD Study (REC:199833, University of Oxford) and the Weatherall Institute of Molecular Medicine's Human Tissue Association licensing number (12433). All patients and healthy donors provided written consent in accordance with the Declaration of Helsinki for sample collection and use in research. To be included in this DPhil project, patients required a BP-MPN diagnosis and the identification of a *TP53*-mutation by clinical diagnostics.

Mononuclear cells (MNC) and HSPC were banked following the INForMeD standard operating procedure as follows. Peripheral blood and bone marrow were diluted 1:1 with INForMeD buffer (PBS including 0.5%FBS and 2mM EDTA (46-034-Cl, Corning)) and passed through a 70um nylon strainer to remove solid particles. To perform MNC isolation, 20mL diluted sample was layered onto 10mL Ficoll-Plaque (17-5442-03, GE Healthcare) in a 50mL falcon tube and centrifuged at 400 $\times g$ for 30-minutes with no break to isolate cells by gradient centrifugation. During centrifugation, granulocytes and erythrocytes sediment in the bottom layer and MNC sit between the Ficoll-Plaque and plasma layer. MNC are collected from the interphase, diluted in up to 50mL INForMeD buffer and counted using trypan blue. After counting, cells were resuspended at a concentration of 5-30 million cells per mL in FBS including 10% Dimethyl Sulfoxide (DMSO; 67-68-5, Sigma-Aldrich), and aliquoted into barcoded cryovials for cryopreservation.

CD34-enriched HSPC populations were collected from healthy donor apheresis samples where patients had received G-CSF treatment. After counting MNC, up to 1×10^8 cells were collected for CD34-enrichment using the Human CD34 MicroBead Kit UltraPure (130-100-453, Miltenyi Biotec). Cells were pelleted at 300 $\times g$ for 10-minutes at 4°C and resuspended in a mixture of 100uL FcR Blocking reagent and 100uL CD 34 MicroBeads UltraPure to be incubated at 4°C for 30-minutes. After incubation, cells were washed with 10mL INForMeD Buffer, pelleted and supernatant aspirated. Cells were resuspended in 500uL INForMeD Buffer and CD34-expressing cells separated from the remaining MNC

fraction using an LS column and MACS Separator. After enrichment, the CD34+ fraction was resuspended in freezing media at 0.5-1 million cells per 1mL and cryopreserved.

Cryopreserved samples were thawed in a 37°C water bath. After thawing, cells were transferred to a 15mL falcon and 1mL sterile FBS added dropwise. Each sample was diluted with a further 8mL thaw buffer constituting IMDM (Gibco, 21056023) supplemented with 20% FBS and 1mg/mL DNase I, Bovine Pancreas (Millipore, 260913). Samples were centrifuged at 300 $\times g$ for 10-minutes, washed twice with 10mL FACS buffer (IMDM + 10% FBS) and counted via haemocytometer. Cells for transduction or CRISPR-Cas9 editing the following day were resuspended in cytokine-rich recovery media at approximately 0.5×10^6 /mL and placed in the 37°C incubator with 5% CO₂ overnight to recover from the thawing process. Recovery media consisted of StemSpan SFEM (09650, StemCell Technologies) supplemented with 1% P/S, 100ng/mL SCF (300-07, Peprotech), 100ng/mL FLT3-L (300-19, Peprotech) and 50ng/mL TPO (300-18, Peprotech). For primary cell editing (Section 2.2.8), StemSpan was not supplemented with P/S due to increased membrane permeability which enhances antibiotic uptake and cell cytotoxicity.

2.2.2 Fluorescent activated cell analysis and sorting

Bulk and single-cell fluorescently activated cell sorting (FACS) was performed using a BD Fusion I or Fusion II (Becton Dickinson) machine. For analysis only, samples were run on a BD LSRFortessa (BD Biosciences). All analysis and sorting experiments included appropriate single stain and fluorescence minus one (FMO) controls. For primary experiments, cells were resuspended in FACS buffer constituting IMDM plus 10% FBS. For cell line experiments, IMDM was replaced by PBS. To exclude dead cells, DAPI (BD Biosciences, 564907) or 7AAD (1:100 from working stock; Cayman Chemical, 11397) were added to FACS buffer prior to sample acquisition. FACS instruments were operated by a member of the Flow Cytometry Facility at the Weatherall Institute of Molecular Medicine, University of Oxford. Where applicable, antibodies and gating strategies are described in the relevant methods sections. All flow cytometry data was analysed using FlowJo software (version 10.9.0, BD Biosciences).

2.2.3 Lentiviral transduction

Lentiviral vectors for primary cell transduction were identical to those used in cell lines and sequences were previously described in **Table 2.2**.

Primary samples were thawed and placed in cytokine-rich recovery media overnight as described in Section 2.2.1. The following day samples were harvested and viable CD34-positive populations selected via FACS or column-enrichment. Antibodies for FACS are detailed in **Table 2.25**. Non-tissue culture treated plates (Falcon) were coated with 25ug/mL RetroNectin (Takara, T100B), covered with parafilm and left at 4°C overnight. RetroNectin replaces polybrene for primary cell transductions, facilitating increased transduction efficiency and reduced cell toxicity. RetroNectin, a recombinant human fibronectin fragment, constitutes three domains, a Heparin-binding domain (H-domain) responsible for binding virus, and CS-1 and C-domains which bind cell surface integrin receptors VLA-4 and VLA-5, respectively. Therefore, RetroNectin enhances primary cell transduction by promoting physical colocalization of lentivirus with target cell membranes. Following overnight incubation, RetroNectin-coated wells were blocked with 2% BSA (bovine serum albumin) in sterile PBS for 30-minutes and subsequently washed once with PBS.

To further aid transduction efficiency of HSPC populations, a “double transduction” approach was taken. Virus was first bound to RetroNectin-coated plates via centrifugation before cell suspension containing additional virus was added prior to repeated centrifugation. For over 50K or 100K CD34-positive HSPC, transductions took place in 48- or 96-well non-tissue culture treated plates (Falcon, 351178;351172), respectively. Briefly, 1×10^8 TU/mL virus was added per well and plates centrifuged at $1000 \times g$ for 1-hour at 32°C, allowing viral particles to bind RetroNectin. After centrifugation, viral supernatant was discarded in 2% Virkon and plates gently washed with sterile PBS. Lentivirus at MOI 100 was then added to cell suspensions and transferred to wells containing RetroNectin-bound-virus. Centrifugation was repeated and plates subsequently placed in 32°C incubator with 5% CO₂ for 24 to 72-hours depending on the experiment performed.

Twenty-four hours after transduction, cells were collected, centrifuged at 500 $\times g$ for 5-minutes at RT and viral supernatant safely discarded. Transduced cells were resuspended in pre-warmed cytokine-rich recovery media described in Section 2.2.1 and cells returned to the incubator for a further 48-hours.

Ninety-six hours post-transduction, CD34-positive RFP-positive cell populations were prepared for FACS. Downstream functional experiments included confirmation of PCBP2-KD via RT-qPCR (Section 2.2.4), short-term liquid culture differentiation assay (Section 2.2.5), colony forming unit assay (CFU; Section 2.2.6) and engraftment into iPSC-derived human bone marrow organoids (Section 2.2.9)

2.2.4 Quantitative real time polymerase chain reaction

RNA extraction and cDNA synthesis

For primary samples, fewer cells were available for RNA extraction and subsequent cDNA synthesis compared to cell lines. Thus, a SuperScript III One-Step RT-PCR System (Invitrogen, 12574026) was used to perform cDNA synthesis and gene-specific PCR amplification in a single reaction with between 0.01pg-1ug total RNA. Briefly, 100 cells (viable, CD34/RFP double-positive) per sample were sorted into 0.5mL PCR strip tubes containing 10uL reaction mix (**Table 2.21**). A 20X TaqMan probe mix (**Table 2.22**) was diluted in TE buffer (Invitrogen, AM9858) to 0.2X for addition to the final reaction mix. At all times, pre- and post-sorting, strip tubes were kept on ice to avoid *Taq* Polymerase degradation.

Reagent	Volume per Sample (uL)	Catalog Number
SUPERase-In RNase Inhibitor	0.1	Invitrogen, AM2694
TE Buffer	1.2	Invitrogen, AM9858
0.2X Probe Mix	2.5	Described in Table 2.22
2X Reaction Mix	5	Invitrogen, 12574026
SuperScript III RT/Platinum <i>Taq</i> Mix	1.2	
Total	10	

Table 2.21. Master mix for primary gene-specific PCR amplification and cDNA synthesis

Probe	Volume for 20X Master Mix (uL)	Assay ID
PCBP2	1	Hs01590472_mH
ACTB	1	Hs99999903_m1
IFNB1	1	Hs01077958_s1
IFIT1	1	Hs03027069_s1
CXCL10	1	Hs00171042_m1
IL-6	1	Hs00174131_m1
GAPDH	1	Hs03929097_g1
RPS18	1	Hs02387368_g1
RPL13A	1	Hs04194366_g1
B2M	1	Hs00187842_m1
Total	10	

Table 2.22. 20X TaqMan probe mix and probe ID. List includes all TaqMan probes used during this thesis project.

After sorting, strip tubes were placed in a ProFlex PCR System (ThermoFisher) thermocycler for pre-amplification and cDNA synthesis (**Table 2.23**). After amplification, cDNA was diluted 1:10 with DNase/RNase-free H₂O and stored at -20°C. For quantitative evaluation of transcript expression via qPCR, see Section 2.1.4 where qPCR steps have been previously described.

Temperature (°C)	Time	Cycles	Description
50	15 mins	1	Reverse Transcription
95	2 mins	1	Inactivation of Reverse Transcriptase/Activation of <i>Taq</i> Polymerase
95	15 sec	20	Specific target amplification
60	4 mins		

Table 2.23. Thermocycler conditions for gene-specific PCR amplification and cDNA synthesis

2.2.5 Multilineage differentiation assay

CD34-enriched cord blood (CB) samples were transduced with either PCBP2-targeting or scrambled control lentiviral shRNA vectors, as described in Section 2.2.3. Ninety-six hours post-transduction, cells were sorted via FACS using DAPI (1:10,000) for viability, RFP to select for transduced populations and CD34-APCef780 (eBioscience, 47-0349-42) for isolation of HSPC populations. Five-hundred viable, transduced CD34-positive cells were sorted into each well of a 96-well round bottom plates containing 50uL non-supplemented StemSpan (09650, StemCell Technologies). Sorted cells were then placed

in a 37°C incubator with 5% CO₂ to recover. After 1-2 hours, 50uL 2X multi-lineage differentiation media was added per well (**Figure 2.24**). Outer empty wells were filled with sterile PBS to prevent losses via evaporation. Four-days post-sort, an additional 50uL 1X multi-lineage differentiation media was added per well.

Component	Concentration (1X)	Catalog number
StemSpan SFEM		09650 (StemCell Technologies)
P/S	1%	-
BIT9500	20%	9500 (StemCell Technologies)
SCF	100 ng/mL	300-07 (Peprotech)
TPO	100 ng/mL	300-18-10 (Peprotech)
IL6	50 ng/mL	200-06 (Peprotech)
FLT3-L	20 ng/mL	300-19 (Peprotech)
G-CSF	20 ng/mL	300-23 (Peprotech)
GM-CSF	20 ng/mL	300-03 (Peprotech)
IL3	10 ng/mL	200-03 (Peprotech)
EPO	1 IU/mL	Janssen, erythropoietin alpha, clinical grade

Table 2.24. Liquid culture multilineage differentiation media (1X)

Multi-lineage differentiation was analysed via flow cytometry on day-12 of culture. Briefly, 200uL FACS buffer was added per well and plates centrifuged at 500 x *g* for 5-minutes. Supernatant was discarded and 1.56uL multilineage differentiation antibody mix added to the remaining ~15uL media per well for staining (**Table 2.25**).

Antibody	Clone	Dilution	Catalog Number
CD34-APC-ef780	4H11	1:100	47-0349-42 (eBioscience)
CD41a-PE-Cy7	HIP8	1:75	25-0419-42 (eBioscience)
CD42b-BV650	HIP1	1:100	303925 (eBioscience)
CD71-AF700	M-A712	1:30	563769 (BD Biosciences)
CD16-BV786	3G8	1:42	302046 (eBioscience)
CD11b-APC	ICRF44	1:300	17-0118-42 (eBioscience)

Table 2.25 Antibodies for determination of multilineage differentiation. List includes all antibodies used in this project for flow cytometry analysis and sorting

Samples were incubated at 4°C for 20-minutes protected from light. Following incubation, 200uL buffer was added per well and plates centrifuged at 500 x *g* for 5-minutes. Samples were resuspended in 200uL buffer including 1:10,000 DAPI and acquired using a BD High Throughput Sampler on a BD LSRFortessa (BD Biosciences). The strategy for gating of differentiated populations is detailed in **Figure 2.5**.

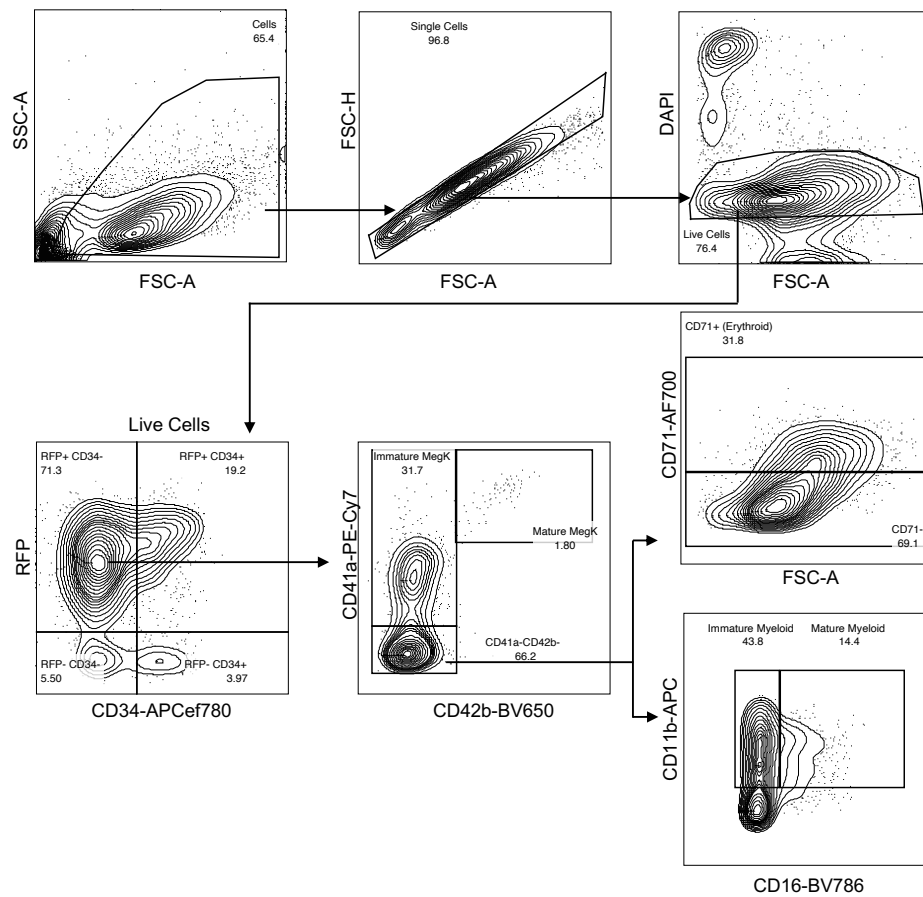


Figure 2.5. Flow cytometry gating strategy for multilineage differentiation assays

2.2.6 Colony forming unit (CFU) assay

Following transduction of cord blood samples as described above in Section 2.2.3, 500 viable CD34/RFP-positive cells were seeded into methylcellulose-based media. Briefly, CD34-positive transduced populations were sorted by FACS into 1.5mL Eppendorf's containing 100uL FACS buffer. Per replicate, 500 cells in minimal FACS buffer were added to 1.5mL Methocult H4435 Media (Stem Cell Technologies, 04435) in a 15mL falcon. Methocult media was vortexed vigorously to homogenise cell suspension and left until bubbles dissipated. Upon settling, 1.5mL cell suspension was transferred to a TC-treated 6-well plate using a blunt needle and 2mL syringe. Empty wells were filled with sterile PBS and CFU plates placed in 37°C incubator with 5% CO₂ undisturbed for 14-days. On day-14, colonies were morphologically scored under the microscope. Representative colonies can be found in **Figure 2.6**.

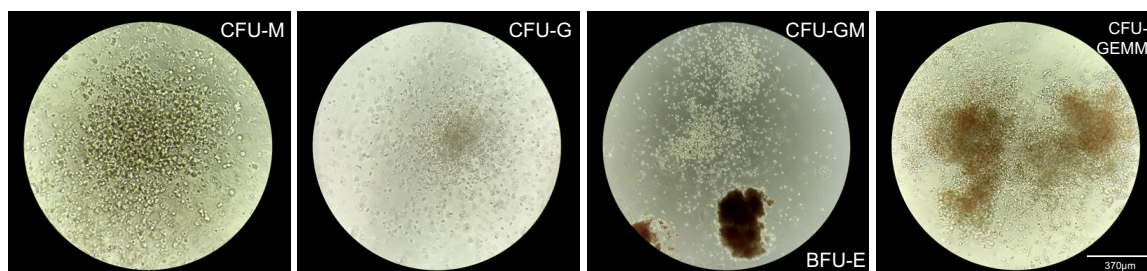


Figure 2.6. Representative CFU colonies. Colonies derived from multilineage progenitor cells during CFU assay. Colony classifications: CFU-M, macrophage; CFU-G, granulocyte; CFU-GM, granulocyte/macrophage; BFU-E, burst-forming unit erythroid; CFU-GEMM, granulocyte/erythroid/macrophage/megakaryocyte

2.2.7 CRISPR-mediated gene knockout in primary cells

Similar to cell line editing described in Section 2.1.7, an RNP-nucleofection approach was used to KO target genes in CD34-enriched HSPC populations. Guides for primary cell editing are identical to those used for cell lines described in **Table 2.15**. Primary samples were thawed the day prior to editing as described in Section 2.2.1. To increase the likelihood of *PCBP2* being rendered inactive, two *PCBP2*-targeting sgRNA were combined during nucleofection to generate concurrent DSB and maximise KO efficiency in primary samples.

Primary cells were edited using the Amaza P3 Primary Cell 4D-Nucleofector X Kit (V4XP-3032, Lonza) with up to 100,000 cells edited per sgRNA. Briefly, RNP complexes were prepared by mixing 1.6µg sgRNA and 3µg Cas9 followed by a 10-minute incubation at 21°C. *PCBP2* sgRNA were complexed individually and then mixed. Cell pellets were each resuspended in 20µL complete Nucleofector Solution (16.4µL Nucleofector solution + 3.6µL Supplement) before mixing with RNP complexes. 20µL cell suspensions containing RNP complexes were then transferred to a Nucleocuvette vessel and electroporated in a 4D-Nucleofector X Unit (AAF-1003X, Lonza) using program, DZ100. Immediately after electroporation, cells were removed from the Nucleocuvette and transferred to 500µL pre-warmed antibiotic-free recovery media (StemSpan SFEM, 100ng/mL SCF, 100ng/mL FLT-3L and 50ng/mL TPO) in a 24-well plate and incubated at 37°C with 5% CO₂ to recover from editing. At minimum, 24-hours post-nucleofection DNA was extracted using a DNeasy Blood and Tissue Kit (Qiagen, 69504) as per the manufacturer's instructions and PCR amplification complete as describe below in Section 2.2.8.

2.2.8 Single-cell colony genotyping assay

Primary cells were edited as per Section 2.2.7. Twenty-four hours after editing, single viable CD34-positive cells were sorted into 60-well Terasaki plates (Thermo Scientific, 163118) containing 20uL media (**Table 2.26**). Following sorting, plates were placed in a TC-box to reduce evaporation and placed in a 37°C incubator with 5% CO₂.

Component	Concentration (1x)	Catalog Number
StemSpan SFEM	-	09650 (StemCell Technologies)
Penicillin/Streptomycin	1%	-
BIT9500	20%	9500 (StemCell Technologies)
SCF	10 ng/mL	300-07 (Peprotech)
FLT3-L	10 ng/mL	300-19 (Peprotech)
TPO	10 ng/mL	300-18-10 (Peprotech)
IL-6	10 ng/mL	200-06 (Peprotech)
G-CSF	10 ng/mL	300-23 (Peprotech)
GM-CSF	10 ng/mL	300-03 (Peprotech)
IL-3	5 ng/mL	200-03 (Peprotech)
EPO	1 IU/mL	Janssen, erythropoietin alpha, clinical grade

Table 2.26. Media conditions for single-cell colony assays

Four-days after sorting, wells were scored for the presence or absence of colony growth. Wells showing colony formation were replenished by removing 8uL media and adding 10uL 2X media described above. Colonies were scored again on day-9 and media replenished. Colonies were collected 13-14 days post-sort. Briefly, each colony was collected and transferred to a 96-well PCR plate to be washed with an equal volume of PBS. Plates were centrifuged at 500 x *g* for 5-minutes and supernatant carefully removed. Colonies were then resuspended in 20uL lysis buffer (**Table 2.28**) and incubated at 37°C for 2-hours in a ProFlex PCR System (ThermoFisher) thermocycler followed by 10-minutes at 95°C to inactivate proteinase K. DNA was then stored at -20°C prior to colony genotyping.

Component	Concentration	Catalog Number
1M Tris pH 8.0	50 mM	J62780.K3 (Thermo Scientific Chemicals)
0.5M EDTA pH 8.0	1 mM	324506 (Millipore)
Tween 20	0.50%	P1379 (Sigma-Aldrich)
20mg/mL Proteinase K	0.03 mg/mL	19131 (Qiagen)
RNase/DNase-free H ₂ O	up to final volume	10977035 (Thermo Scientific)

Table 2.27. Buffer for single-cell colony lysis

To determine whether *PCBP2*-KO or *AAVS1*-KO single cells could support colony growth, colonies were genotyped to confirm KO of the appropriate gene. PCR master mix was prepared as described (**Table 2.28**) and 1uL DNA added per reaction with thermocycler conditions as described in **Table 2.29**. PCR primers have been described previously in **Table 2.16**. Following PCR, colonies were cleaned-up using a 1:1 ratio of AmPure XP beads, as per the manufacturer's instructions, and resuspended in 10-20uL RNase/DNase-free H₂O for in-house Sanger sequencing. *PCBP2* and *AAVS1* wild-type samples were also Sanger sequenced to provide a wild-type chromatogram trace for inferring percentage editing via Synthego ICE tool (Synthego Performance Analysis, ICE Analysis, 2019. v3.0. Synthego).

Reagent	Volume per Reaction	Catalog Number
Platinum PCR SuperMix High Fidelity	22 uL	12532016 (Invitrogen)
10uM Forward Primer	1 uL	Custom oligonucleotides (Invitrogen)
10uM Reverse Primer	1 uL	

Table 2.28 Master mix for single cell colony genotyping

Step	Temperature (°C)	Time	Cycles
Initial Denaturation	94	2 min	1
Denaturation	94	30 sec	35
Annealing	63 (<i>PCBP2</i>) // 56 (<i>AAVS1</i>)	30 sec	
Extension	72	30 sec	
Final Extension	72	5 min	1
Cooling	8	∞	1

Table 2.29. Thermocycler conditions for single cell colony genotyping

2.2.9 Primary patient cell engraftment in human bone marrow organoids

Organoids for primary cell engraftment were kindly generated by Dr Abdullah Khan, Weatherall Institute of Molecular Medicine, University of Oxford.

Generation of human bone marrow organoids

Human bone marrow organoids were generated using a modified protocol based on the original publications^{136,137}. Briefly, human iPSCs (Gibco Line) were dissociated into suspension aggregates in StemFlex supplemented with RevitaCell. The resulting

embryoid bodies (EBs) were then collected by centrifugation ($100 \times g$ for 3-minutes) and resuspended in phase I medium as previously described^{137,138}. Aggregates were cultured at 5% O₂ for 21-days, during which they were subject to a series of cytokine cocktail changes including VEGFA, VEGFC, FGF2, BMP4, SCF, FLT3L, IL3, IL6, IL7 and mCSF over the course of 21-days of differentiation.

At day-5 of the protocol mesodermal aggregates were embedded in a mixed matrix scaffold comprised of fibrin, collagens type I and IV, and Matrigel. This scaffold drove vascular sprouting and the generation of a lumen forming vessel architecture invested with stromal and haematopoietic cells.

Organoid engraftment with primary cell populations

Primary cells were engrafted into day-21 organoids. Prior to engraftment, primary cells (not for lentiviral transduction) were thawed and sorted via FACS or column-enriched for selection of CD34-positive populations before staining with CellTrace Violet (Invitrogen, C34557), as per the manufacturer's protocol. CellTrace allows primary patient cells to be discriminated from organoid populations via flow cytometry and/or imaging.

Transduced primary cells for engraftment were thawed and left to recover overnight in cytokine-rich media, as previously described (Section 2.2.1). The following day cells were stained with CellTrace Violet and column-enriched for CD34-positive populations prior to transduction as per Section 2.2.3. Column-enrichment in this case was preferred to minimise the harsh exposure of primary cells to both FACS and lentivirus. Transduced cells were left overnight and engrafted the following day as described below. Homogeneous transduced populations were sorted for *PCBP2* quantification as described in Section 2.2.4.

Following staining, cells were counted, centrifuged at $300 \times g$ for 3-minutes and resuspended in an amount of media correlating to the number of organoids to be engrafted. Per organoid, 2,000 or 5,000 CD34+ cells were engrafted in 50uL media. Number of cells engrafted was kept consistent for all samples throughout independent experiments, and variation in engrafted cell numbers between experiments was due to the unpredictability of primary cell work where cell numbers cannot be guaranteed post-

thaw. Two media conditions for supporting primary cell survival and proliferation were compared, the standard organoid media used throughout organoid production (as described above for organoid generation) or an AML-cell specific media (**Table 2.30**). Cytokines within AML-specific media were added 2X since organoids were originally plated in standard organoid media. To seed primary cells, 50uL cell suspension was added gently to each organoid well of a 96-well plate and centrifuged at 300 x *g* for 30-seconds. To compare primary cell proliferation in organoids to standard liquid culture conditions, 1,000 CD34+ cells were also seeded in 100uL corresponding media alongside organoids. Plates were then returned to a 37°C incubator with 5% CO₂.

Component	Concentration (2X)	Catalog number
StemPro-34 SFM	-	10639011 (ThermoFisher)
P/S	0.2%	-
SCF	20	300-07 (Peprotech)
FLT3-L	20	300-19 (Peprotech)
TPO	20	300-18-10 (Peprotech)
IL-6	20	200-06 (Peprotech)
G-CSF	20	300-23 (Peprotech)
GM-CSF	20	300-03 (Peprotech)
IL-3	10	200-03 (Peprotech)

Figure 2.30. AML-specific media for organoid engraftment

Three days post-engraftment, organoids were imaged using an Evos M5000 microscope (ThermoFisher, AMF5000SV) and media replenished. To replenish media, 40uL media was carefully removed from wells and 50uL prewarmed media was added. Organoids for flow cytometry analysis, on days 3 and 7, were collected by gravitation and pooled. Per condition, up to 3 organoids were collected for disassociation. Organoids were washed twice in PBS, resuspended in 2.5mg/mL Collagenase D (Roche, 11088858001) and incubated at 37°C for 5-minutes before trituration and a further-5 minutes incubation. Incubations were repeated for up to 30-minutes until organoids had fully disassociated. Cells were then centrifuged at 300 x *g* for 5-minutes and resuspended in PBS supplemented with 10% FBS to stop the disassociation reaction. Cells were stained with 1:50 CD34-APCef780 (eBioscience, 47-0349-42) and 1:10,000 7AAD for flow cytometry performed on a BD LSRFortessa analyser (BD Biosciences). Analysis included appropriate single colour and fluorescence-minus-one controls and a representative gating strategy can be found in **Figure 2.7**.

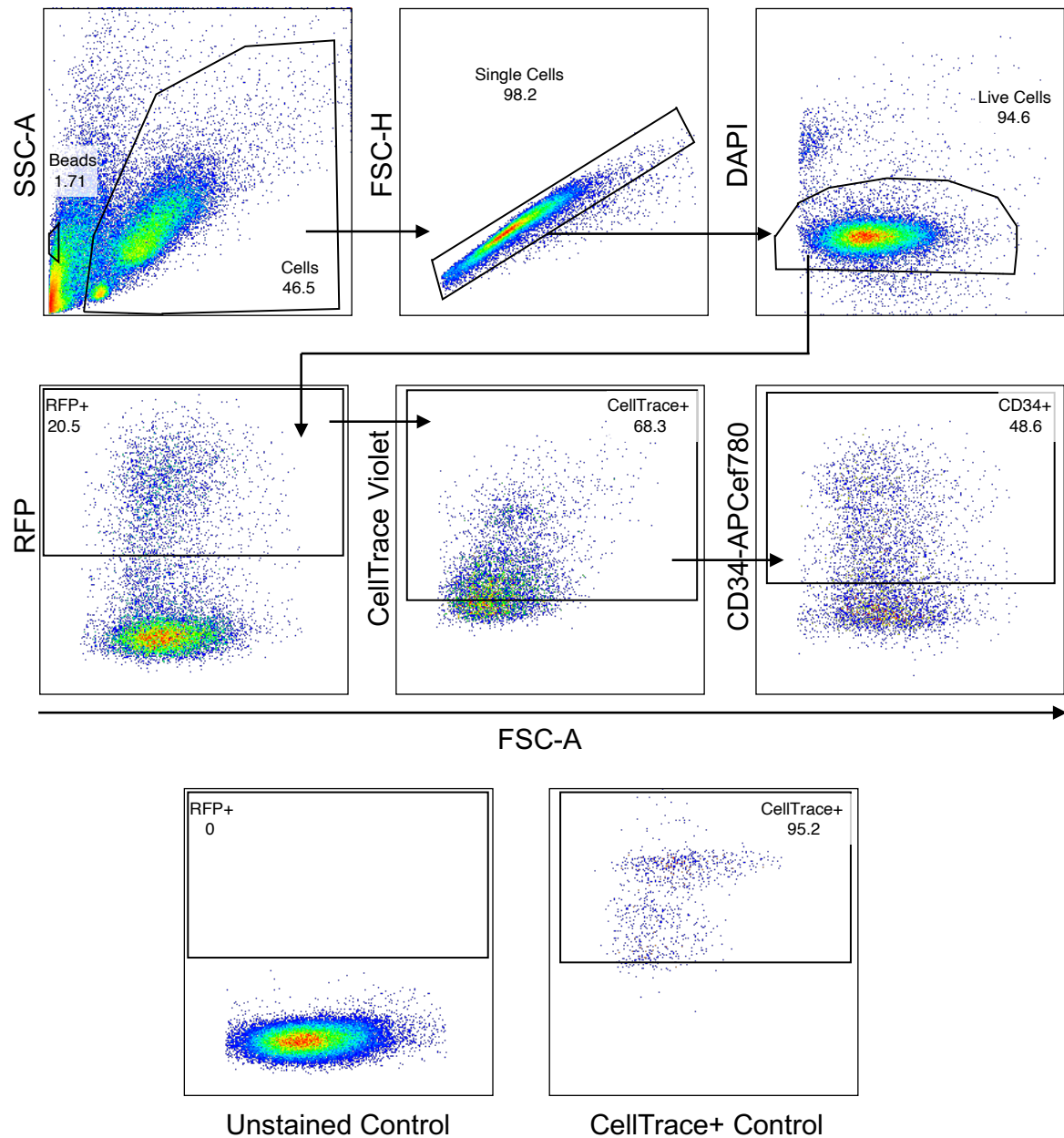


Figure 2.7. Flow cytometry gating strategy for assessment of transduced primary cell populations following organoid engraftment. Lentiviral vectors used for primary cell transduction contain an RFP reporter. Primary cells were stained with CellTrace Violet to discrimination from organoid-derived cell populations.

2.3 Bioinformatic analysis

2.3.1 TARGET-seq analysis

TARGET-seq libraries were prepared and targeted single-cell genotyping analysis, including data pre-processing, genotype assignment and computational reconstruction of clonal hierarchies performed as previously described^{63,139}.

Single-cell 3'-biased RNA-sequencing pre-processing and differential gene expression analysis for this thesis project were kindly performed by Dr Alba Rodriguez-Meira as previously described⁶³. Cells expressing >500 genes were included for differential gene expression analysis. Read counts for differential expression analysis were normalised and log2-transformed, including genes expressed in at least two cells. Differentially expressed genes were identified using a combination of nonparametric Wilcoxon test and Fisher's exact test, to compare expression values for each group and frequency of expression for each group, respectively. *P*-values generated from both tests were combined using the Fisher's exact test method and adjusted using the Benjamini-Hochberg method. Differentially expressed genes were those with a log2FoldChange > 0.5 and adjusted *p*-value < 0.1.

2.3.2 DepMap analysis

The DepMap database (<https://depmap.org/portal>) contains whole genome CRISPR loss-of-function screening data for over 1,000 cell line models, making it a valuable tool for assessing candidate genes targeting potential¹⁴⁰. Gene effect scores, accounting for copy number effects, were downloaded from DepMap data release 24Q2 (https://depmap.org/portal/data_page/?tab=allData&releasename=DepMap%20Public%2024Q2&filename=CRISPRGeneEffect.csv1)¹⁴¹. Gene effect data for HEL, HEL92.1.7 and SET-2 cell lines were extracted via their DepMap cell line IDs, ACH-000004, ACH-000005, ACH-000195, respectively. Essential genes were defined by a Gene Effect score < -0.5.

Gene expression in cancer cell lines

The DepMap database also contains candidate gene transcript and protein expression data across cancer cell line lineages. To investigate *PCBP2* expression data, *PCBP2*-specific data was downloaded from DepMap data release 24Q2 (<https://doi.org/10.25452/figshare.plus.25880521.v1>)¹². For *PCBP2* mRNA expression, values within were reported in transcripts per million (TPM) after log2 transformation, using a pseudo-count of 1; $\log_2(\text{TPM}+1)$ (https://depmap.org/portal/data_page/?tab=allData&releasename=DepMap%20Public%2024Q2&filename=OmicsExpressionProteinCodingGenesTPMLogp1.csv). *PCBP2* protein expression (dataset Q15366) was determined by mass spectrometry across 375 cancer cell lines by Nusinow *et al*¹⁴² and distributed via the DepMap database.

2.3.3 Gene ontology

Gene ontology (GO) analysis was complete using the Gene Ontology knowledge base (<http://geneontology.org/>)^{143,144}. Given a set of genes, enrichment analysis identifies GO terms overrepresented within the given gene set. GO enrichment analysis was complete within the GO knowledge base using PANTHER (Protein analysis through evolutionary relationships; v.17.0, <https://archive.pantherdb.usc.edu>)¹⁴⁵ and GO data accessed from 2023-05-10 release (10.5281/zenodo.7942786)¹⁴⁶. Overrepresented GO terms were identified via Fisher's Exact test.

2.3.4 Analysis of external AML datasets

Two external AML datasets including RNA-sequencing, *TP53*-mutation information and survival data, were analysed in this project, namely The Cancer Genome Atlas (TCGA) and BeatAML^{57,147,148}.

TCGA dataset

The TCGA-LAML (Acute Myeloid Leukaemia) dataset comprises 200 clinically annotated *de novo* AML cases plus whole-genome or whole-exome sequencing, RNA-sequencing and

DNA-methylation analysis⁵⁷. Clinical data and gene expression values were download from the National Cancer Institute GDC Data Portal (<https://portal.gdc.cancer.gov/>)¹⁴⁹. *TP53*-mutation profiles were downloaded from the cBioPortal for Cancer Genomics (cBioPortal; <http://cbioportal.org>)^{150,151}. *PCBP2* expression values were defined as reads per kilobase per million mapped reads (RPKM), and offset by a 1 and log2-transformed for plotting ($\log_2(\text{RPKM}+1)$). One-hundred and forty-five samples had *PCBP2* expression values and *TP53*-mutation information (10 *TP53*-mutant, 135 *TP53*-wildtype). Of these, 134 samples had the clinical data required for survival analysis.

BeatAML dataset

The BeatAML dataset comprises 672 biopsies from 562 AML patients including extensive clinical and genomic annotation^{147,148}. Clinical data and *TP53*-mutation profiles were downloaded from the cBioPortal for Cancer Genomics (cBioPortal; <http://cbioportal.org>)^{150,151}. Patients lacking *TP53*-mutation profiles were removed from further analysis. Gene expression values were downloaded from the BeatAML data viewer (www.vizome.org). *PCBP2* RPKM values were offset by a 1 and log2-transformed. Overall, 375 AML samples were available for analysis including *TP53*-mutation information and RNA-seq data. Fifty-nine of these patients were diagnosed with secondary AML, all of which had data available for survival analysis.

2.3.5 Identification of essential genes (CRISPR screening)

Illumina sequencing data was downloaded from Illumina BaseSpace in FASTQ format. FASTQ files were demultiplexed within the Illumina pipeline prior to downloading. Paired-end sequencing was utilised to maximise the number of multiplexed samples, however, after demultiplexing, only Read 1 FASTQ files were included for analysis. Initial quality of raw sequencing reads was assessed using FastQC v0.11.9 (<https://github.com/s-andrews/FastQC>). Nextera adapters were trimmed using Cutadapt v4.0¹⁵² (<https://github.com/marcelm/cutadapt>). Preceding 5' stagger sequences were also removed during this step. Three-prime adapters were not trimmed as testing only takes into consideration the first 20bp when counting guides.

The MAGeCK package v0.5.8 was used to determine gene essentiality^{153,154}. Using the `mageck count` command, sgRNA reads were mapped to corresponding genes. To account for differing sequencing depth between experimental replicates, mapped read counts were median normalised based on non-essential negative control sgRNA only (`--norm-method control`), rather than all sgRNA, which can result in overconservative estimates of essentiality. Once sgRNA counts were generated, the `mageck mle` command was used to perform maximum likelihood estimations (MLE) of gene essentiality scores. The `mageck mle` command output comprises a beta-score per gene and associated *p*-value, where beta-scores can be positive or negative reflecting the direction of gene selection.

2.3.6 Bulk RNA-seq analysis

Data pre-processing and read count quantification

Data pre-processing and read count quantification were complete by Antonio Rodriguez-Romera, current DPhil student at the Weatherall Institute of Molecular Medicine, University of Oxford.

Low quality reads and sequencing adaptors were removed from FASTQ files using TrimGalore v0.6.10. High quality transcripts were mapped to the human genome using STAR v2.7.8a (genome version GRCh38 v40). Lastly, reads mapping to coding genes were counted using feature counts from subread v2.0.3 (genome version GRCh38 v40). The Gene by Sample count matrix was imported in R for further processing.

Differential gene expression analysis

Differential gene expression was complete using the DESeq2 R package v1.44.0 (<https://github.com/thelovelab/DESeq2>)¹⁵⁵. The DESeq2 package uses Wald test pairwise comparisons to determine differentially expressed genes and *p*-values adjusted for multiple testing using the Benjamini-Hochberg method. Genes were filtered to exclude those with no expression across sample (zero read counts). Unless stated otherwise, significantly upregulated or downregulated genes were defined by a $\log_2\text{FoldChange} > 0.5$ or < -0.5 , respectively, and adjusted *p*-value < 0.05 .

Gene set enrichment analysis

Gene set enrichment analysis (GSEA) was performed using the GSEA software v4.3.2 (Broad Institute Inc, Massachusetts Institute of Technology)¹⁵⁶. Normalised count matrices of expressed genes were compared against defined gene sets derived from the Molecular Signatures Database (<https://www.gsea-msigdb.org/gsea/msigdb/>), primarily Hallmark (h.all.v2023.1)¹⁵⁷. Gene sets obtained from the literature¹⁵⁸ were first formatted into Gene Matrix Transposed file format (.gmt) before uploading to GSEA software. GSEA heatmap was generated using the R package, ComplexHeatmap (v.2.20.0; <https://github.com/jokergoo/ComplexHeatmap>).

2.3.7 Bulk-level splicing analysis

Splicing analysis, and description below, was kindly complete by Dr Sean Wen, previous DPhil student at the Weatherall Institute of Molecular Medicine, University of Oxford.

Processing of RNA-sequencing data

Raw RNA-sequencing reads in FASTQ format were trimmed with Trim Galore software (<https://github.com/FelixKrueger/TrimGalore>) to remove poor quality 3' bases defined with Phred score <20. Trimmed reads were aligned to the human reference genome (build 38) with STAR software in two-pass mode¹⁵⁹. Specifically, in the first round of alignment, the sequencing reads were aligned to the reference genome to obtain a list of splice junctions detected, but without generating the alignment output. In the second round of alignment, all splice junctions detected from the first round of alignment were used to increase the sensitive of splice junction detection. Additionally, the sequencing reads were aligned to the reference genome and the resulting alignment file in sequence alignment map (SAM) format generated. Samtools was then used to convert the alignment file from SAM to binary alignment map (BAM) format¹⁶⁰. The alignment file in BAM format was convert to bigwig format using deepTools¹⁶¹ for visualisation on the UCSC Genome Browser¹⁶² for *in silico* validation of candidate splicing events.

Isoform detection

Stringtie2 was used to detect isoforms across all samples¹⁶³. Specifically, the GENCODE database (version 31), which contains a catalogue of reported full-length isoforms¹⁶⁴, was used as a guide by Stringtie2 to detect both reported and novel isoforms in each sample. The isoforms detected at the sample-level were then merged into a gene transfer file (GTF) format to create a catalogue of all isoforms detected in this study. rMATS-turbo was then applied on this catalogue of isoforms to detect exon-level splicing events, namely skipped-exon (SE), mutually exclusive exons (MXE), retained intron (RI), alternative 5' splice site (A5SS), and alternative 3' splice site (A3SS). For RI splicing events, per-base intron coverage was computed with BEDTools software¹⁶⁵. The following files were subsequently used as input to create the MARVEL object for downstream analysis¹⁶⁶: (1) splicing junction expression from STAR aligner, (2) GENCODE (version 31) catalogue of reported isoforms, (3) exon-level splicing events detected by rMATS-turbo, (4) per-base intron coverage computed by BEDTools, and (5) normalised gene expression values in count per million (CPM) unit. After creation of MARVEL object, additional splicing event types, namely alternative first exon (AFE) and alternative last exon (ALE) were detected with the MARVEL software using the *DetectEvents* function. Cryptic A5SS and A3SS splicing events were also retrieved from the list of rMATS-turbo-detected A5SS and A3SS splicing events using the *SubsetCrypticSS* function.

Isoform quantification

The aforementioned seven types of exon-level splicing events were quantified with the *ComputePSI* and *ComputePSI.Posterior* function. The former function computes isoform expression based on the observed splice junction counts (frequentist approach). Splicing events with low coverage, defined as coverage of <10, were coded as missing values. The latter function imputes these missing values using a Bayesian approach¹⁶⁷. The expression values computed using the frequentist approach was used for differential splicing analysis while the expression values computed using the Bayesian approach was used for principal component analysis (PCA). The degree of splicing for each splicing event measured using either approach was represented in percent spliced-in (PSI) unit that ranges from 0 to 100%. A splicing event with PSI value of 0% indicates that the exon/intron is completely spliced-out whereas a splicing event with PSI value of 100%

indicates that the exon/intron is completed spliced-in. All subsequent analysis was performed with MARVEL.

Isoform analysis

Differential isoform analysis was performed between knock-down (KD) group 1 or KD group 2 versus control group using permutation test implemented by the *CompareValues* function. In this statistical approach, the sample group labels were shuffled several times to create the null distribution. The observed difference in PSI values of a given splicing event between two sample groups (Δ) is then compared against this null distribution to determine if the observed difference is greater or lesser than expected by chance. A splicing event is considered to be statistically significant if $P_{\text{permuted}} < 0.05$ and $|\Delta| > 5\%$. Statistically significant splicing events were examined for their impact on protein function, specifically using nonsense-mediated decay (NMD) prediction. For NMD prediction, any premature stop codon (PTC) introduced by the splicing event was predicted using the *FindPTC* function and whether the PTC led to any NMD was predicted using the *CompareExpr* function which examined if the PTC met the following criteria: (1) located >50bp away from the final splice junctions, (2) located within any open reading frame (ORF), and (3) at least one isoform met criteria (1) and (2).

2.3.8 Statistical analysis

Number of independent experiments, individual donors and replicates for each experimental are stated within respective result sections. For the comparison of parametric data from individual groups, two-way unpaired Student's *t*-tests were used. For larger sample groups, Fisher's exact tests were used instead of Student's *t*-test. Pearson correlation coefficient was used to evaluate linear correlations between datasets. Statistical analyses were performed using GraphPad Prism v10.1.0 (264) October 18, 2013 or R v4.4.1 (2024-06-14). Chi-squared tests were performed manually and used to analyse the non-parametric categorical data presented in Figure 5.5.

Chapter 3 RNA-binding proteins are overexpressed in *TP53*-sAML

3.1 Introduction

3.1.1 An introduction to RNA-binding proteins

Ribonucleic acid (RNA) is an essential cellular molecule required to translate genetic information encoded in DNA into functional proteins. Canonically, RNA takes three forms, all of which are crucial for protein synthesis and ultimately, cellular metabolism. Messenger RNA (mRNA) acts as a template for protein translation, transfer RNA (tRNA) decodes mRNA during translation and ribosomal RNA (rRNA) is the core structural component for the ribosome upon which protein synthesis takes place. More recently, RNA has been recognized as a highly versatile molecule with a multitude of additional regulatory roles within the cell through the discovery of catalytic RNA and non-coding RNA¹⁶⁸.

Newly synthesised mRNA molecules must be processed and undergo extensive post-transcriptional modifications (PTM) before reaching their functional form. Messenger RNA metabolism includes pre-mRNA processing (capping, splicing and polyadenylation), addition of RNA modifications, nuclear export of mature mRNA via the nuclear pore complex, transcript stability, protein translation and degradation (**Figure 3.1**). With many different stages of mRNA metabolism, each must be highly regulated to maintain normal cellular physiology. Post-transcriptional gene regulation (PTGR) is responsible for co-ordinating RNA metabolism and throughout its lifecycle, RNA associates directly with RNA-binding proteins (RBP) to form ribonuclear protein (RNP) complexes¹⁶⁸. RBP therefore contribute to all aspects of RNA metabolism including synthesis, trafficking, translation and degradation.

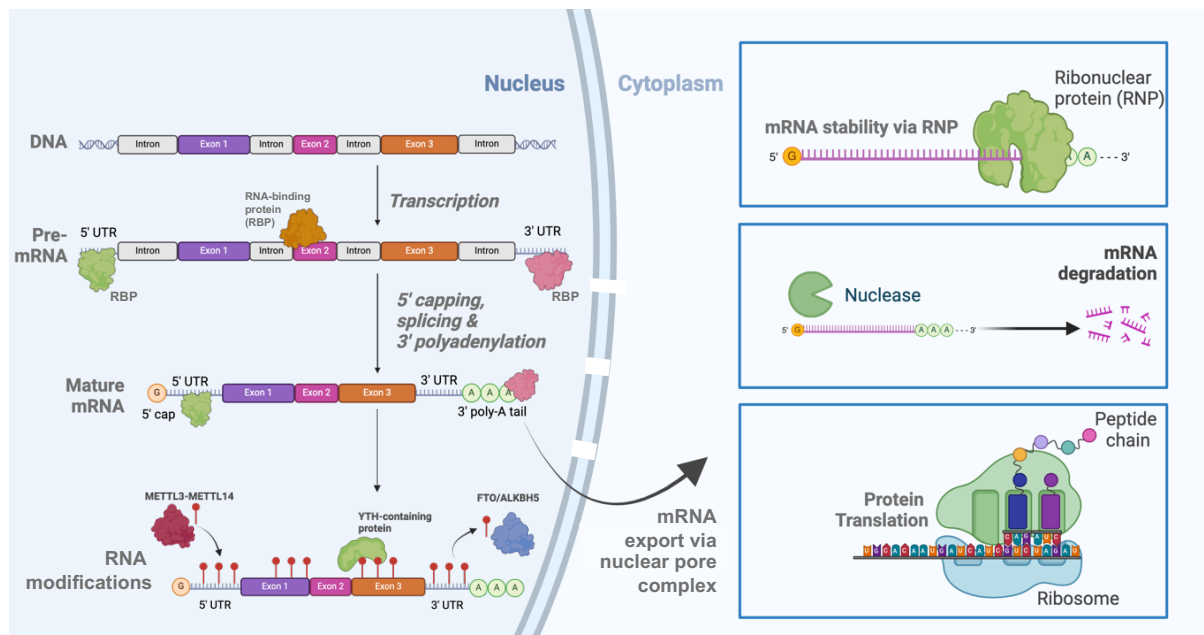


Figure 3.1. Processes involved in mRNA metabolism. Following transcription, nuclear pre-mRNA processing includes 5' capping, splicing and 3' polyadenylation. Mature mRNA will then be modified and exported to the cytoplasm via the nuclear core complex where it is stabilised. All aspects of RNA metabolism are regulated by RNA binding proteins (RBP). Illustration created with BioRender.com.

RBP are commonly defined by their RNA-binding domain (RBD), which are the structural domains responsible for conferring binding specificity to a specific cohort of mRNA transcripts. RBD bind RNA via *cis*-regulatory sequence-defined RNA elements. Approximately 600 RBD have been characterised with the most common domains – RRM (RNA-recognition motif), KH-domain (K-homology), DEAD-motif and DSRM (double-stranded RNA-recognition motif) - frequently found as repeats or in conjunction with other RBD to increase sequence specificity for transcript binding¹⁶⁸. Thus, to comprehensively understand the PTGR network requires identifying RBP, their RNP cofactors, target mRNA transcripts and mechanistic function(s) in RNA regulation. Such efforts require integrating multiple different investigative approaches and significant advances in high-throughput proteomic and transcriptomic technologies have paved the way for uncovering the complex roles of RBP in PTGR¹⁶⁹.

Traditionally, RBP were identified by UV-crosslinking or RNA affinity chromatography techniques followed by immunoprecipitation of single specific RBP proteins^{170,171}. Identifying mRNA-binding partners of early discovered RNP complexes involved protein immunoprecipitation followed by cDNA arrays¹⁷². Advances in proteomic techniques

have predicted the current census of RBP to be 1,562 RBP or more¹⁶⁸. Large cohorts of RBP have been identified via mass spectrometry of protein complexes associated with UV-linked polyadenylated RNA^{173,174}.

Next-generation sequencing (NGS) approaches have accelerated the discovery of RBP-RNA interactions. For example, RNA immunoprecipitation and crosslinking immunoprecipitation coupled to high-throughput sequencing, described as RIP-seq and CLIP-seq respectively, are primary techniques now adopted to investigate RBP-RNA interactions across the transcriptome¹⁷⁵. Such techniques have revealed the extent of RBP binding, with some RBP binding thousands of diverse transcripts and presenting as multifunctional entities in RNA regulation¹⁶⁸. Novel RNA-base editing approaches, such as HyperTRIBE, have also been developed to identify RBP target transcripts^{176,177}. Unlike CLIP-seq, which requires a large number of cells, the HyperTRIBE methodology can identify mRNA-protein interactions using low cell numbers (<500), allowing cell-specific interactions to be investigated in rare cell populations, such as HSC¹⁷⁸.

RBP are also responsible for the deposition, regulation and removal of RNA modifications. RNA modifications, such as the abundant internal N⁶-methyladenosine (m⁶A) modification can play diverse roles in RNA regulation. Whether an RBP-of-interest has any influence as a writer, reader, eraser or cofactor of RNA modifications can be investigated using techniques such as DART-seq or RNA modification mass spectrometry¹⁷⁹.

Integrative studies including multiple datatypes are key to comprehensively characterising the RBP-regulatory network, including moving from the discovery of RBP and their target sites to functional annotation of an RBP's role in transcriptional regulation¹⁶⁹. To date, the largest systematic effort to study the regulatory network and function of 356 RBP by Van Nostrand *et al*¹⁸⁰ integrates multiple datatypes including eCLIP-seq (enhanced CLIP-seq) to determine target transcripts, immunofluorescence techniques to determine subcellular localization and RBP perturbation followed by RNA-seq for functional assessment of RBP. Additional complexity in characterising the RBP-regulatory network is derived from the tissue and cell-type specific roles of RBP, which may have significant impacts on normal physiology and malignancy.

3.1.2 RBP in normal and malignant haematopoiesis

To maintain steady-state haematopoiesis, stem cell self-renewal and lineage differentiation must be tightly controlled with stage-specific transcriptomes defining cellular states. RBP play important mechanistic roles in the molecular regulation of haematopoietic processes, adaptively controlling transcript stability, isoform expression and protein abundance through PTGR. Dysregulation of RBP can occur via mutation, differential expression or PTM. Given their fundamental roles in haematopoiesis, dysregulation of RBP interactions can contribute to haematological malignancies. Encouragingly, through this dysregulation, RBP also present as therapeutic targets in myeloid malignancy. Below I present a variety of examples where dysregulation of RBP in haematopoiesis gives rise to malignancy.

RBP and splicing of haematopoietic transcripts

The importance of RBP in normal haematopoiesis can first be illustrated by their requirement in splicing as integral spliceosome components. The spliceosome is composed of >200 proteins, and recurrent mutations in a number of these spliceosome components – most commonly *SRSF2*, *SF3B1* and *U2AF1* – can be found in cases of myelodysplastic syndrome (MDS), clonal haematopoiesis and acute leukaemia^{181–183}. Particularly in MDS, >50% of cases examined by Yoshida *et al*¹⁸² presented with a missense mutation in splice factors. Mechanistically, dysregulation of spliceosome assembly can have a downstream impact on multiple transcripts important for haematopoiesis. For example, *SRSF2* is an important SR (serine/arginine-rich) spliceosome factor, required to protect exon inclusion during splicing of pre-mRNA, and has been reported as mutated in 20% of MDS cases¹⁸⁴. The most common *SRSF2* mutation at P95 alters *SRSF2* binding preferences and drives alternative splicing patterns of downstream transcripts, including *EZH2* – a key regulator of HSC self-renewal and differentiation. Alternative *EZH2* splicing leads to nonsense-mediated decay (NMD) due to exon exclusion forcing a frameshift mutation and incorporation of a premature stop codon¹⁸⁴. Thus, *SRSF2* impairs haematopoiesis via mis-splicing of key haematopoietic regulators, subsequently driving myelodysplasia and highlighting the functional significance of RBP in haematopoiesis.

Molecular vulnerabilities, are expressed genes which are required for cancer cell survival but do not depend on genetic mutations for their function in leukemogenesis. RBP expression is frequently dysregulated in cancer and using an RBD-targeted CRISPR-Cas9 screen, Wang *et al*¹⁸⁵ reported a number of upregulated RBP to be preferentially required for AML survival, including putative therapeutic target, RBM39. Immunoprecipitation mass spectrometry identified interacting partners of RBM39, including multiple members of the spliceosome complex and proteins involved in ribosome biogenesis. Importantly, RBP interacting partners of RBM39 were also identified as highly essential via RBD-targeted CRISPR screening¹⁸⁵. Thus, RBM39 was shown to interact within an RBP network which is necessary for AML survival, highlighting the complex interaction of RBP in malignant haematopoiesis.

RBP and translational regulation in haematopoiesis

Multiple examples of RBP supporting haematopoiesis via regulating mRNA translation exist. For example, PTBP1 is a member of the hnRNP (heterogeneous nuclear ribonucleoproteins) family and has been reported essential for HSC maintenance and erythropoiesis in part via its control over translation of mRNA encoding ribosomal proteins¹⁸⁶. In this regard, PTBP1 regulates the availability and effective translation of ribosomal proteins by preferentially binding the sequence-specific 5'UTR of ribosomal proteins transcripts, facilitating their protein translation. In the absence of *PTBP1*, mice have a global reduction in protein translation and suffer ribosomopathies-like phenotypes, disorders caused by defects in ribosome biogenesis or function, which predominately affect the haematopoietic system¹⁸⁶. This is concordant with the knowledge that HSC and erythrocytes are highly sensitive to perturbations in ribosomal function/protein synthesis rates¹⁸⁷, uncovering PTBP1 as an integral component of ribosome function in haematopoiesis. The same report also concluded PTBP1 to be important for the translation of iron-related erythropoiesis factors including haemoglobin and ferritin, highlighting PTBP1 as an important RBP in the translation of key erythroid factors¹⁸⁶.

A second RBP, YBX1, is essential for AML maintenance via translational regulation of oncogenic transcripts such as *c-Myc*¹⁸⁸. *YBX1* perturbation was found to have a minimal

effect on global transcript expression or splicing. However, *YBX1* depletion led to selective loss of *c-Myc* transcript expression in polysomal fractions containing translating mRNA. Furthermore, CLIP-seq data confirmed direct interactions between YBX1 and oncogenic transcripts, thus presenting YBX1 as a determinant of leukaemogenic translational regulation in AML¹⁸⁸. Interestingly, YBX1 has also been identified as an important mediator of *JAK2*-mutant MPN persistence via different mechanisms of PTGR¹⁸⁹. Research has shown *JAK2*V617F-mediated downstream signalling selectively activates YBX1 via phosphorylation in treatment-persistent MPN. Phospho-activated YBX1 subsequently regulates splicing of *Mknk1*, an integral component of the *JAK2*-ERK signalling pathway responsible for persistence of *JAK2*-mutant malignant clones¹⁸⁹.

RBP and RNA modifications

The N6-methyladenosine (m⁶A) modification is the most ubiquitous chemical RNA modification with the ability to influence mRNA stability, localisation and decay¹⁹⁰. A heterodimeric methyltransferase complex including METTL3 and METTL14 binds mRNA transcripts and is responsible for m⁶A deposition. Of interest to this project, METTL3 is an acknowledged regulator of both normal and malignant haematopoiesis¹⁹¹. In steady-state haematopoiesis, METTL3 is a key governor of adult HSC differentiation. *METTL3* depletion was found to drive HSC accumulation via a differentiation blockade and *in vivo*, differentiation defects led to marked reductions in reconstitution potential¹⁹². METTL3 was found to support HSC differentiation through depositing of m⁶A on target transcripts, such as *c-Myc*, to enhancing their translation and direct differentiation¹⁹². Yao *et al*¹⁹³ reported similar expansions of adult HSC in the bone marrow following *METTL3* perturbation and alternatively suggested loss of reconstitution capacity in long-term competitive reconstitution assays was due to METTL3 regulating the quiescent stem cell state. Alternative roles of METTL3 in normal HSPC populations have also been reported and depletion of *METTL3* shown to drive lineage differentiation¹⁹⁴. However, these experiments were performed on mixed HSPC populations, and thus the stage-specific role of METTL3 was not clarified in this report¹⁹⁴. Overall, METTL3 can be considered an important RBP for stem cell differentiation via m⁶A modification of target transcripts including *c-Myc*.

In contrast to the role of METTL3 in normal haematopoiesis, *METTL3* is overexpressed in AML and required for the maintenance of malignant leukaemic clones. In AML, METTL3 prevents differentiation of leukaemic cells, supporting their characteristic undifferentiated state. Multiple mechanisms of METTL3 action in this regard have been reported. Vu *et al*¹⁹⁴ reported that METTL3-dependent deposition of m⁶A modifications on key oncogenic transcripts including *c-Myc*, *BCL2* and *PTEN* promoted their translation in cancer cell line models. Alternatively, Barbieri *et al*¹⁹⁵, showed METTL3 associates with chromatin by binding sequence-specific promoter regions to co-transcriptionally deposit m⁶A-modifications onto actively transcribing target genes. Via this mechanism, leukaemic maintenance was shown to be dependent on the transcription factor, SP1, whose translation required co-transcriptional m⁶A deposition by METTL3. Thus, contrasting roles for METTL3 in m⁶A deposition during normal and malignant haematopoiesis exist, highlighting the cell-type specific roles of RBP.

RBP and mRNA decay

Degradation is the final regulatory step within the mRNA life cycle, and along with transcription, controls the cellular abundance of transcripts. Unsurprisingly, RBP are also crucial mediators of mRNA decay which involves decapping, deadenylation and exoribonucleolytic degradation¹⁹⁶.

ZFP36L2, is an RBP with defined roles in RNA decay and translational repression through the binding of *cis*-regulatory AU-rich elements in the 3' untranslated region (UTR) of target transcripts¹⁹⁷. Using surface-antigen guided CRISPR-Cas9 screening, ZFP36L2 has been identified as an essential regulator of leukaemic differentiation¹⁹⁷. Genetic ablation of *ZFP36L2* stabilised transcripts involved in monocytic/myeloid specification (*CD14*, *CSF1R*, *CEBPB* and *MPP9*) and triggered the terminal differentiation of AML cells¹⁹⁷. Mechanistically, ZFP36L2 was shown to bind 3'UTR AU-rich elements and recruit the deadenylase complex, CCR4-NOT, to deadenylate target transcripts within the myeloid differentiation network, promoting their degradation¹⁹⁷. In concordance, other members of the ZFP36 family have also been found to associate with the deadenylase CCR4-NOT complex to facilitate mRNA decay¹⁹⁸. Finally, the requirement of ZFP36L2 in AML cells was shown to be HOXA9-dependent. HOXA9 expression has been reported essential for

AML initiation and overexpression of this transcription factor is known to drive gene expression via leukemic-specific enhancers and thus, cell-type specific gene expression profiles¹⁹⁹. In this regard, HOXA9 upregulated *ZFP36L2* expression in leukaemic cell populations via an enhancer cluster downstream of *ZFP36L2* which was highly enriched in AML¹⁹⁷. While HOXA9-dependent expression suggests a cell-type specific requirement of *ZFP36L2* in leukaemic cell populations, a limitation of this work was the lack of functional studies in normal HSPC populations to determine the efficacy of targeting ZFP36-family proteins for differentiation-based therapies to treat AML¹⁹⁷.

3.1.3 RBP in *TP53*-mutant myeloid malignancies

At the time of writing, I was unaware of any publications relating to the role of RBP specifically within the context of *TP53*-mutant myeloid malignancies. Thus, while the role of RBP in myeloid malignancies is well reported, a knowledge gap remains regarding the role of RBP in the context of *TP53*-mutant myeloid malignancies.

3.1.4 Hypothesis

Genes, particularly RBP, overexpressed in *TP53*-mutant LSC and essential for cancer cell survival are molecular vulnerabilities with the potential to be harnessed as therapeutic targets

3.1.5 Summary and aims

Analysis of a multi-omic single-cell dataset including *TP53*-sAML, MPN and HD controls provides a unique opportunity to identify dysregulated gene sets specifically within disease-propagating *TP53*-mutant LSC populations. Utilising this dataset, and publicly available gene essentiality screens, I investigated whether any dysregulated genes identified in *TP53*-mutant LSC could be considered molecular vulnerabilities for potential therapeutic targeting.

Aims for this section are as follows:

1. To identify genes, particularly RBP, both overexpressed in *TP53*-mutant LSC and essential for cancer cell survival.
2. To validate the requirement of candidate genes via genetic perturbation screens in *JAK2/TP53*-mutant AML cell line models.

3.2 Results

3.2.1 Patient cohort

This project studied a multi-omic single-cell TARGET-seq⁶³ dataset comprising a cohort of 23 individuals, totalling 10,530 individual Lin-CD34+ HSPCs, including 7 *TP53*-sAML patients, 11 MPN patients and 5 healthy donor (HD) individuals derived from Rodriguez-Meira *et al*⁶³ (**Figure 3.2; Appendix 1**).

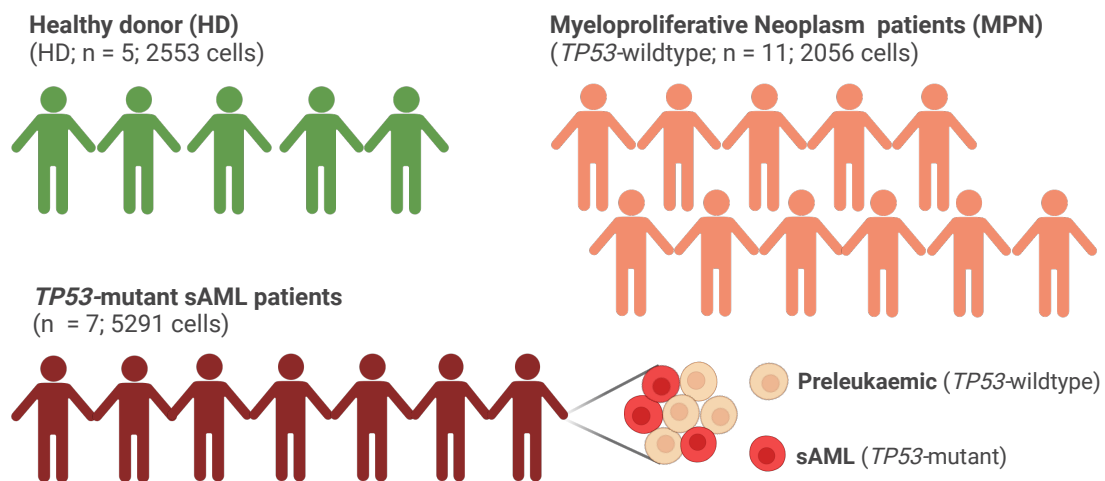


Figure 3.2. Patient cohort for TARGET-seq analysis. Includes 5 healthy donor (HD), 11 myeloproliferative neoplasm (MPN) patients and 7 *TP53*-sAML patients. Cells from *TP53*-sAML patients were stratified by *TP53*-mutation to be classified as either the preleukaemic *TP53*-WT clone (including both cells with or without an MPN-driver mutation) or the dominant *TP53*-mutant leukaemic clone.

All HD and MPN HSPC were determined to be *TP53*-WT, while HSPC derived from MPN patients carried mutations in the *JAK-STAT* signalling pathway. A proportion of HSPC found within *TP53*-sAML samples were *TP53*-WT, representing the antecedent clone. The majority of *TP53*-WT cells in *TP53*-sAML possessed mutations in *JAK2*, *CALR* or *MPL*. A small proportion of cells in *TP53*-sAML patients were WT for both *TP53* and MPN-driver

mutations. Together, all *TP53*-WT cells identified in *TP53*-sAML patients were defined as the preleukaemic clone for analysis. Integration of cellular phenotype via index sorting identified dominant *TP53*-mutant clones to be enriched in progenitor populations, as previously described in *de novo* AML²⁰⁰. Transcriptomic signatures showed *TP53*-mutant populations clustered independently from *TP53*-WT populations and were enriched for LSC and erythroid signatures⁶³. From here, dominant *TP53*-mutant clones were determined to be functional LSC via PDX of *TP53*-sAML samples, where leukaemic phenotypes *in vivo* recapitulated characteristics of the dominant clone from the primary patient sample⁶³.

Using this patient cohort, Rodriguez-Meira *et al*⁶³ previously deciphered the clonal architecture and genetically distinct subclones present in *TP53*-sAML. Following on, this project focused on identifying dysregulated genes in *TP53*-sAML, comparing *TP53*-mutant and non-mutant cells, including the inpatient comparison between *TP53*-WT HSPC (preleukaemic) and *TP53*-mutant LSC from *TP53*-sAML patients.

3.2.2 Identifying overexpressed genes in *TP53*-mutant LSC

To identify candidate genes which may present molecular vulnerabilities in *TP53*-sAML, genes overexpressed in *TP53*-mutant LSC were first identified. Three independent differential gene expression analysis were run, comparing gene expression in *TP53*-mutant LSC (sAML) to healthy donor controls (HD), MPN patients and *TP53*-WT HSPC derived from sAML patients (preleukaemic). Ninety-seven genes were identified to be significantly upregulated, defined as a Log₂FoldChange > 0.5 and adjusted *p*-value 0.1, in *TP53*-mutant LSC compared to all other cohorts (**Figure 3.3; Appendix 2**).

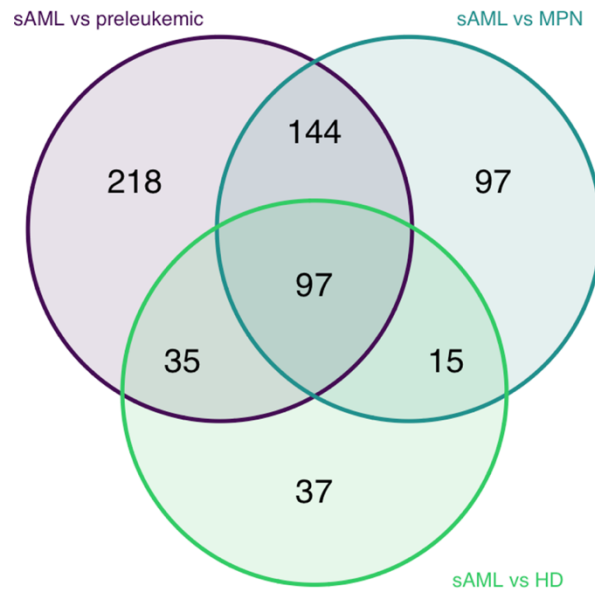


Figure 3.3. Venn diagram of commonly upregulated genes in *TP53*-mutant LSC. Differential gene expression analysis compared *TP53*-mutant LSC to healthy donor (HD) controls, MPN patients and *TP53*-WT HPSCs from sAML patients (preleukemic). Overexpressed genes defined at Log₂FoldChange > 0.5 and adjusted *p*-value < 0.1.

3.2.3 Identifying essential genes

To identify which overexpressed genes have potential for therapeutic targeting, upregulated genes were correlated with essentiality data from the DepMap database¹⁴⁰ [Chronos, 24Q2¹⁴¹]. DepMap contains whole-genome CRISPR-Cas9 loss-of-function screening data across over 1,000 cell line models, making it a valuable tool for assessing a target genes potential. CRISPR screens in the DepMap database were stratified to include only *JAK2/TP53*-mutant cell line models, specifically HEL, HEL2.9.1.7 and SET-2 cell lines, replicating the genetic background of *TP53*-sAML patient cohort. Approximately 9% of human genes (~1,800) have been determined essential for cancer cell survival using genome-wide CRISPR screens with enrichment for genes involved in fundamental processes such as DNA replication, RNA transcription and mRNA localisation²⁰¹. Twenty-one of the ninety-seven (21/97) genes overexpressed in *TP53*-sAML were identified to be essential genes for cancer cell survival in model cell lines, with a gene effect score < -0.5 (**Figure 3.4A**). Furthermore, >50% (12/21) of overexpressed and essential genes were identified to be RBP, as identified using RBP2GO, a comprehensive resource for assessing available RBP proteome datasets²⁰². This

enrichment of RBP was highly significant, with a p -value of $1.072e^{-08}$ (Fisher's exact test). Finally, gene ontology (GO) analysis confirmed a significant enrichment of genes with RNA-binding function in *TP53*-sAML (FDR-adjusted p -value < 0.05, Fisher's exact test; **Figure 3.4B**).

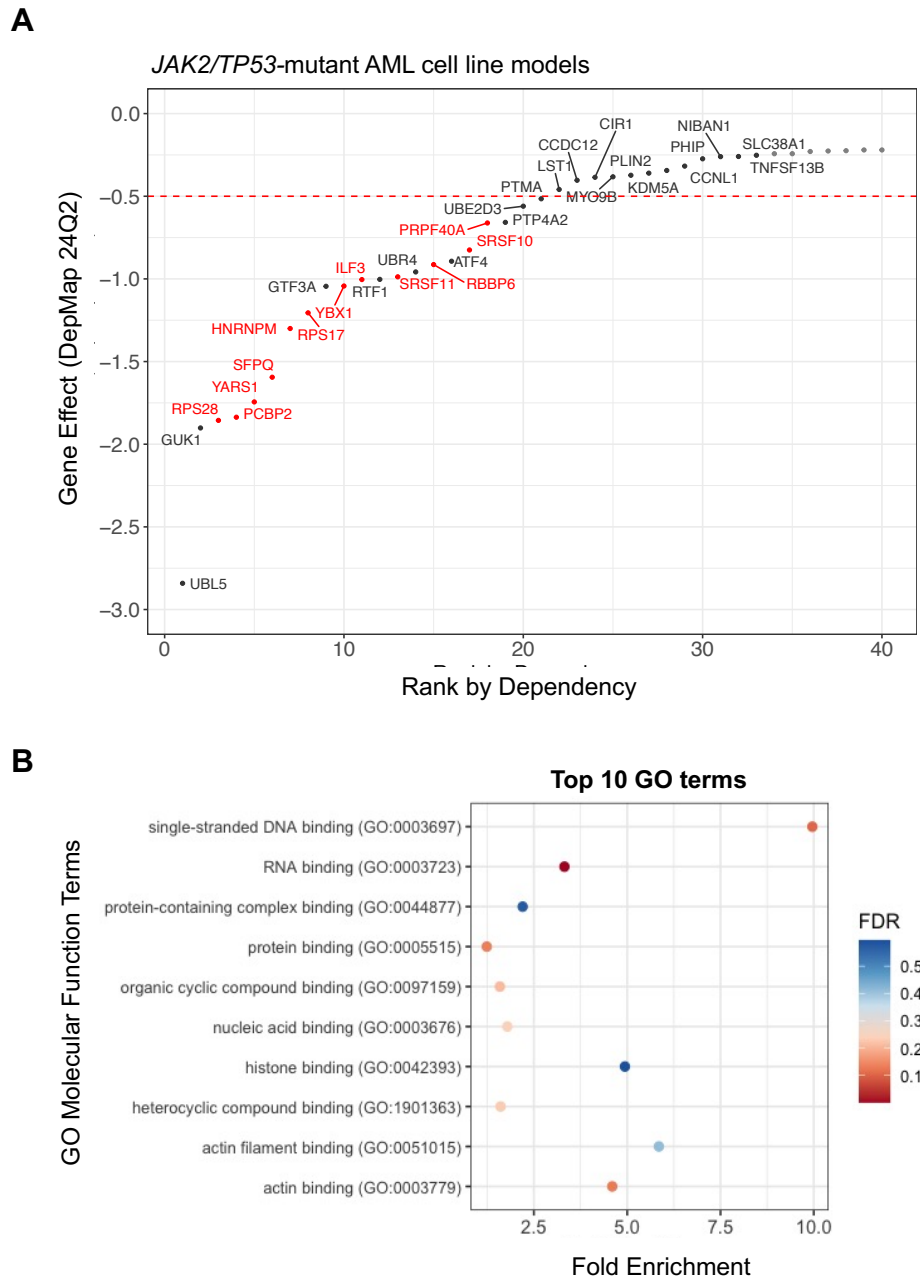


Figure 3.4. Investigating the cancer cell requirement of overexpressed genes (A) Essentiality of overexpressed genes in *JAK2/TP53*-mutant cell line models (HEL 92.1.7, HEL and SET-2) ranked by order. Essentiality data derived from the DepMap database (24Q2). Genes highlighted in red are RNA-binding proteins. Gene effect < -0.5 signifies essential genes. **(B)** Top 10 Molecular Function Gene Ontology (GO) terms associated with genes overexpressed in *TP53*-sAML. Adjusted p -values (FDR) calculated using Fisher's exact test with significance < 0.05.

3.2.4 Validating RNA-binding protein dependencies in *TP53*-sAML

Based on the DepMap analysis, dysregulated RBP are enriched in *TP53*-sAML and may present molecular vulnerabilities with the potential for therapeutic targeting. To validate the hypothesised requirement of RBP in *TP53*-sAML, a small-scale RBP-targeting CRISPR-Cas9 screen was designed including the 12 candidate RBP identified above and all appropriate positive, negative and non-targeting controls. In total, the RBP-targeted library comprised 61 sgRNA. *JAK2/TP53*-mutant AML cell lines, HEL and SET-2, were transduced with the sgRNA library and guide representation from this loss-of-function screen compared at day 0 and day 21 (**Figure 3.5**). Given the essentiality of candidate RBP, I expected sgRNA targeting RBP to be depleted from the screen overtime.

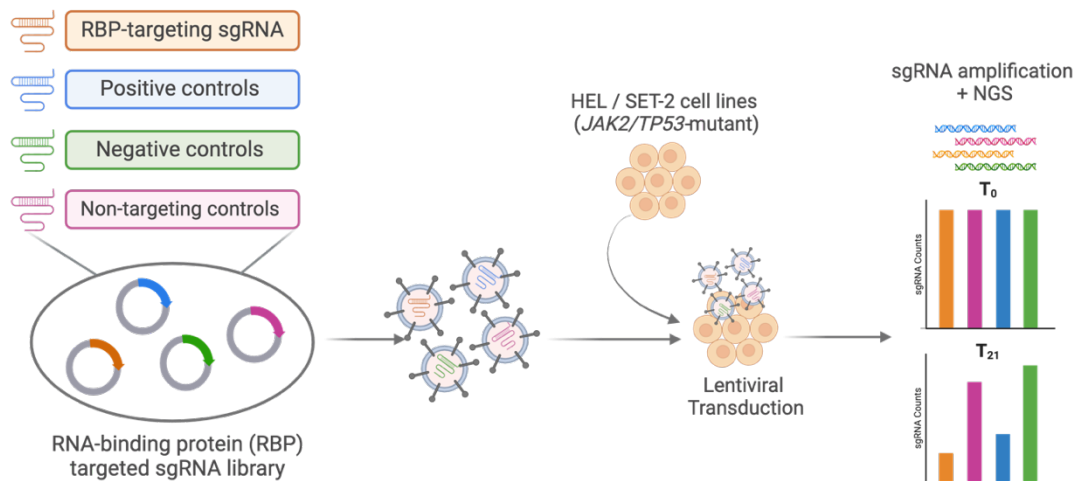


Figure 3.5. Schematic of CRISPR-Cas9 screening in *JAK2/TP53*-mutant AML cell line models

Guide representation at T_0 highly correlated with guide representation within the plasmid pool for both cell lines (**Figure 3.6A & 3.6C**) – an important quality control metric of lentiviral packaging and production. Essentiality was determined by beta-scores²⁰³, where a positive beta-score represents positive gene selection and a negative score represents negative selection. Negative selection indicates a gene is essential for cell survival. Guides targeting essential control genes – *TUBB*, *POLR2A* – involved in cellular structure, and transcription, respectively, had significant negative beta-scores in both HEL and SET-2 cell lines. Ribosomal genes overlapping as both target RBP and essential genes, *RPS28* and *RPS17*, also showed dependency across cell lines. Non-

targeting control guides with no sequence similarities to the human genome – *luciferase*, *aquaporin*, *yeast* and *LacZ* – had positive beta-scores with significance varying between genes and cell lines. In line with DepMap data, all targeted RBP were identified as essential, except *SFPQ*, which had an insignificant beta-score concordant across both cell lines (**Figure 3.6B & 3.6D**). Furthermore, comparing beta-scores between genes across screens showed a highly significant correlation (**Figure 3.6E**; Pearson's correlation $R^2 = 0.854$, $p\text{-value} = 3.163e^{-06}$) – validating the experimental integrity of the screening methodology and the requirement of target RBP in *JAK2/TP53*-mutant AML cell line models.

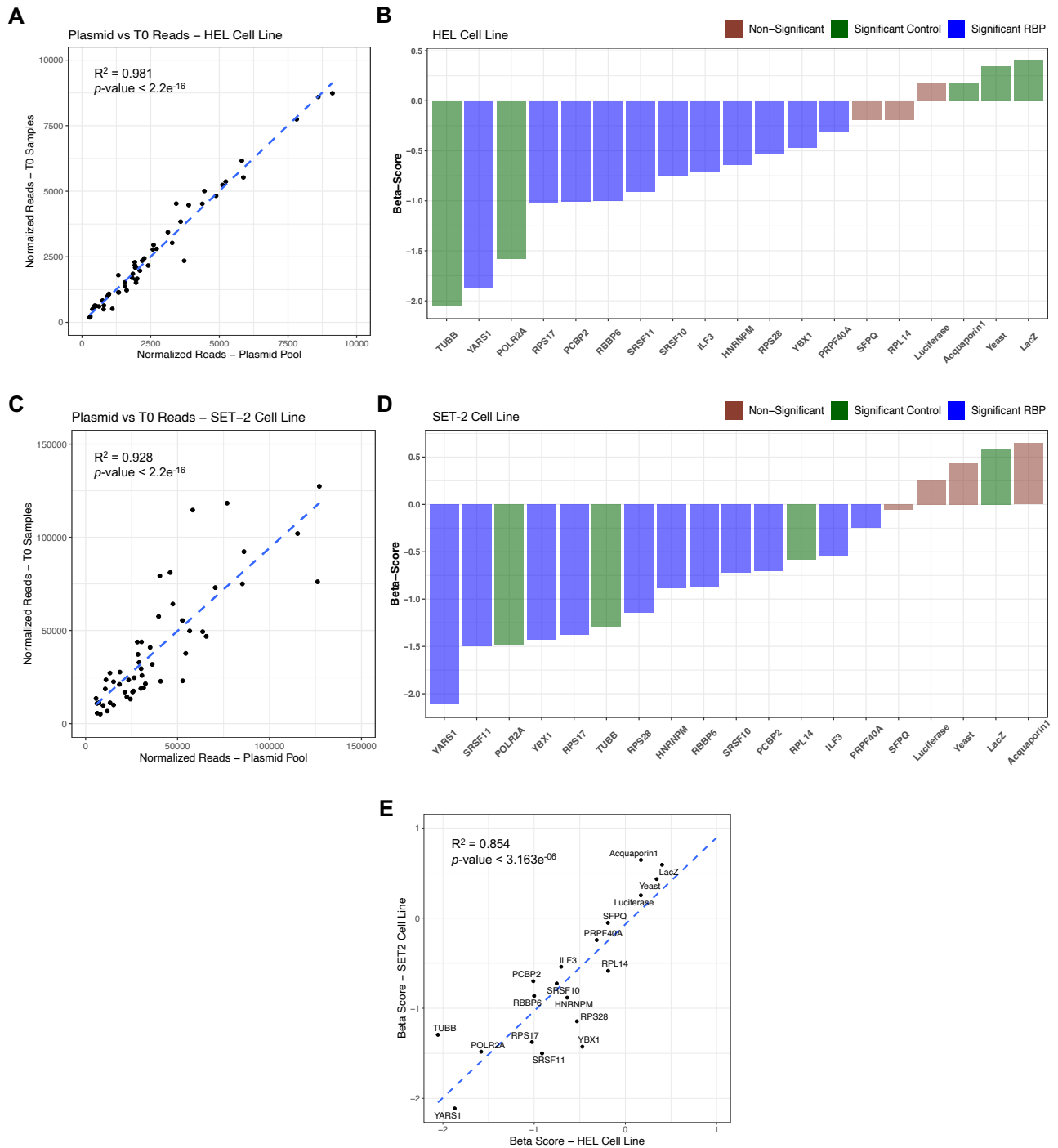


Figure 3.6. RBP-targeted CRISPR-Cas9 screening. (A+C) Pearson correlation co-efficient (R^2) and p -value comparing normalised guide representation at timepoint T₀ with the plasmid pool in (A) HEL and (C) SET-2 cell lines. (B+D) Gene essentiality scores (beta-scores) for target RBP and controls in (B) HEL and (D) SET-2 cell lines. Mageck-MLE algorithm for calculating beta-scores was reported for 6 technical replicates for HEL cells and 4 technical replicates for SET-2 cells (both split across two independent experiments). Blue and green bars represent significant RBP and controls, respectively. Significant genes determined where p -value < 0.05 using Wald Statistic Test with false discovery rate correction. Red bars represent non-statistically significant genes. (E) Pearson correlation co-efficient (R^2) and p -value comparing beta-scores between HEL and SET-2 cell lines.

3.3 Discussion

This project aimed to identify molecular vulnerabilities in a highly aggressive and poor prognosis cohort of transformed MPN patients harbouring *TP53*-mutations. Using an integrative analysis approach, I combined the analysis of a previously described TARGET-seq dataset with cancer vulnerabilities data from the publicly available DepMap database. I hypothesised genes upregulated in *TP53*-mutant LSC and essential for cancer cell survival are potential molecular vulnerabilities in *TP53*-sAML. Furthermore, I predicted such vulnerabilities have the potential to be harnessed as therapeutic targets. Overall, results showed a significant enrichment of RBP upregulated in *TP53*-mutant LSC, of which all but one RBP could be validated as essential for survival in model AML cell lines via RBP-targeted CRISPR-Cas9 loss-of-function screens.

This study was enabled by the allelic-resolution mutation detection of TARGET-seq technology. Resolving the genetic intratumoural heterogeneity of *TP53*-sAML was essential to confidently ascertain molecular vulnerabilities specific to *TP53*-mutant LSC populations. In the absence of such resolution, transcriptomic comparisons between *TP53*-WT and *TP53*-mutant populations, including intrapatient comparisons, would not have been possible. Therefore, TARGET-seq allowed me to confidently ascertain molecular vulnerabilities in *TP53*-mutant multi-hit LSC.

While I selected genes commonly dysregulated in *TP53*-mutant multi-hit LSC compared to all other cohorts for further analysis (**Figure 3.3**), all dysregulated genes within each individual comparison alone could also be investigated in future work. For example, the 504 differentially expressed genes identified in the sAML versus preleukaemic comparison may be of importance to revert *TP53*-sAML to the preceding *TP53*-WT MPN clone. In this thesis, I chose to investigate the 97 commonly dysregulated genes to refine my gene set of interest and maximise the possibility of targeting a gene which may be of importance in both MPN and *TP53*-sAML, while also limiting the effect of HD HSPC through differences in expression.

Following *TP53*-mutation and loss of the remaining WT-*TP53* allele, leukaemic transformation is also associated with complex karyotype and multiple copy number

alterations (CNA)⁶³. RBP loci did not coincide with characterised CNA events (data not shown), concluding a recurrent chromosomal amplification event is not responsible for the upregulation of individual or concurrent RBP in *TP53*-sAML.

Other mechanisms potentially driving upregulation of RBP in a *TP53*-mutant background include transcriptional regulation by gain-of-function (GOF) p53 mutants. While it remains to be wholly understood how mutant p53 regulates transcription, it is known GOF mutants are often stable and maintain their transcriptional activation domain, driving oncogenic transcriptional programmes²⁰⁴. Otherwise, GOF mutants can co-opt transcription factors, chromatin modifiers and/or transcription initiation complex components to promote networks of dysregulated gene expression²⁰⁵. Mechanistically, Loizou *et al*²⁰⁶ elegantly presented the GOF activities of a missense mutant *TP53* in AML. They showed the common *TP53*^{R175H} mutant is sufficient to drive dysregulated expression of the embryonic transcription factor, FOXH1, and its downstream target genes. Dysregulated *FOXH1* expression promoted aberrant stem cell self-renewal and leukemogenesis by enforcing expression of developmental stem cell and plasticity gene signatures²⁰⁶. Given the variety of missense mutations present in our *TP53*-sAML cohort (Appendix 1), how disparate mutations mechanistically converge on a single network of RBP remains unclear and is beyond the scope of this project. A second possibility is mutant p53 does not directly regulate RBP expression, and rather LSC which express members of the RBP cohort in a *TP53*-mutant context acquire a competitive advantage against clones lowly expressing candidate RBP. Hence, in this scenario, dominant *TP53*-mutant clones are those with dysregulated RBP expression. Furthermore, whether dysregulated RBP expression is also required for MPN transformation remains to be investigated.

The scope of this project did not cover the role of candidate RBP in *TP53*-mutant leukaemic transformation. *TP53*-mutant clones are present in 16% of patients with chronic-phase MPN and yet, many will not progress to leukaemic transformation¹²². Hence, additional genetic and molecular determinants must be required for leukaemic transformation. Recently, Rodriguez-Meira *et al*²⁰⁷ demonstrated a previously unappreciated role of chronic inflammation in MPN transformation. This work showed inflammation suppresses *TP53*-WT HSPC function in chronic-phase MPN while

concurrently fuelling the fitness advantage of *TP53*-mutant clones and promoting their genetic evolution²⁰⁷. Taking a similar approach in this project, the role of RBP in transformation could be investigated by comparing dysregulation of RBP in antecedent *TP53*-heterozygous mutant clones present in MPN-patients who undergo leukaemic transformation, to those who remain in the chronic-phase followed by *in vivo* validation. Identifying candidate RBP, independently or as a network, required for transformation would be a significant finding and if therapeutically targetable, *TP53*-mutant MPN transformation to aggressive myeloid malignancy could potentially be prevented.

CRISPR-Cas9 screening applications emphasising the importance of RBP in myeloid malignancies have previously been reported^{180,197,208}. Among the 12 RBP identified as overexpressed and essential in *TP53*-sAML, only SRSF10 coincides as a significantly dysregulated and essential hit in other screens examined¹⁸⁵. Furthermore, Wang *et al*¹⁸⁵ described SRSF10 to be an AML-specific dependency in the *TP53*-WT MOLM-13 AML cell line, when the requirement was compared across perturbation screens in a variety of cancer cell line lineages. This raises the question of whether the RBP dependencies identified in *TP53*-sAML are specific to 1) AML and/or 2) *TP53*-mutant backgrounds. All RBP identified in this project can be characterised as common essential genes, meaning they have pan-cancer essentiality and rank in the top percentile of depleted genes in 90% of DepMap dependency screens¹⁴¹. This suggests RBP identified in this project are likely required in other cancer contexts. However, whether the requirements of RBP also extends to *TP53*-WT myeloid malignancies, and other cancer types, is somewhat beyond the scope of this DPhil project, where the primary aim was to identify molecular vulnerabilities within a specific subset of aggressive *TP53*-mutant myeloid leukaemia.

Whether genes essential for cancer cell survival are also required in normal tissues is an important consideration in the selection of novel therapeutic targets. If essentiality translates to WT-HSPC, therapeutic targeting is likely to be cytotoxic to WT populations and disruptive to normal haematopoiesis. However, therapeutic targeting of essential genes has been shown possible in myeloid malignancies. Characteristics facilitating the development of treatments with a therapeutic index include preferential lethality within cancer cell-specific genetic backgrounds and distinct epigenetic and transcriptional profiles responsible for driving a cancer cell state. For example, RBM39 is a pan-essential

splice factor which shows preferential lethality in spliceosome mutant AML¹⁸⁵. Splice factor mutations are often mutually exclusive, requiring otherwise WT splice factors for leukaemic cell survival. Thus, pharmacological perturbation of RBM39 in combination with an additional spliceosome mutation accelerated splice dysregulation, including mis-splicing of other AML-essential RBP and oncogenic transcripts, inducing anti-leukaemic effects¹⁸⁵. Selective pharmacological inhibition of RBM39 is possible via sulfonamide treatment²⁰⁹. Sulfonamides act as a “molecular glue” between the E3-ubiquitinating ligase complex, CRL4-DCAF15 and RBM39, leading to CRL4-DCAF4-mediated ubiquitination and proteasomal degradation of RBM39. The therapeutic index for targeting RBM39 with sulfonamides is further increased by the requirement of DCAF15 expression for effect, which was found to be significantly overexpressed in AML compared to normal CD34+ HSPC¹⁸⁵. Based on these findings, patients likely to respond to treatment can be stratified and phase II clinical trials have been established to investigate the efficacy of sulfonamide treatment in patients with myeloid disease who harbour spliceosome mutations²¹⁰.

In this chapter's introduction, YBX1 was described as a pan-cancer essential RBP required for the persistence of treatment-resistant *JAK2*-mutant MPN and maintenance of AML^{189,211}. Importantly, heterozygous knockout of *YBX1* does not perturb normal haematopoiesis²¹¹ and the clinical application of targeting YBX1 for treatment of AML can be investigated using a novel small molecule inhibitor (SU056) which directly binds and inhibits YBX1²¹². Pharmacological YBX1 inhibition was found to specifically select against AML populations in PDX models, while normal HSPC populations remained competitive, emphasising the potential for therapeutically targeting a common essential RBP, such as YBX1²¹³.

Finally, METTL3 has been described as a key governor of RNA modifications in normal and malignant HSPC. However, despite these roles and known pan-cancer dependency, targeting METTL3 using small molecule inhibitors has been investigated as a treatment against AML²¹⁴. Results showed *in vivo* treatment using a METTL3-selective catalytic inhibitor, STM2457, targeted key leukaemic stem cell populations, impairing engraftment and prolonging survival in AML mouse models²¹⁴. Importantly, treatment showed no overt toxicity or lasting impacts on the normal haematopoietic compartment in pre-clinical models, including bone marrow-derived HSPC and peripheral blood constituents.

Overall, this work further highlights the potential pan-cancer essential RBP present as therapeutic targets for the treatment of AML.

Translation of pan-essential targets to clinically-applicable targeted therapeutics comes with a high attrition rate during early clinical trial phases²¹⁵. And more often than not, therapies targeting common essential genes more closely reflect the low therapeutic index of standard chemotherapies²¹⁶. However, as shown by the examples above, therapeutic targeting of pan-cancer essential genes is a surmountable challenge. Hence, while the effect of RBP perturbation on non-malignant cells will need to be extensively investigated, and high-levels of preclinical validation complete to prevent failure, previous work shows promise in targeting common essential genes for the treatment of myeloid malignancies.

In summary, this project identified RBP to be enriched and highly essential in *TP53*-sAML. While this project focussed on identifying genes upregulated in *TP53*-mutant LSC due to their targetability, downregulated genes should also be looked at in future work. The requirement of enriched RBP was validated using an RBP-targeted genetic perturbation screening in cancer cell line models, showing high concordance between different, but genetically similar, cell lines. Thus, RBP can be considered molecular vulnerabilities in *TP53*-sAML with the potential for therapeutic targeting. Next, careful consideration of each candidate RBP is required to select the candidate which presents most promise for therapeutic targeting for further functional validation. And finally, it is important to note that while RBP were identified to be upregulated at the mRNA level, future work will be required to validate upregulation of RBP-of-interest at the protein level in primary patient samples.

Chapter 4 PCBP2: a novel molecular vulnerability in *TP53*-sAML

4.1 Introduction

4.1.1 An introduction to PCBP2

The Poly(rC)-binding protein (PCBP) family of RBP comprises 5 members: PCBP1-4^{217,218} and hnRNP K²¹⁹. The PCBP family are ancient and highly conserved, with homologs throughout prokaryotic and eukaryotic evolution. All members of the PCBP family are defined by their triple KH-domain and affinity for poly(C)-rich DNA/RNA sequences. While sharing evolutionary origin, each PCBP has high specificity and affinity for a cohort of sequence-specific cytosine-rich motifs, leading PCBPs to be a functionally diverse family of RBP.

Human PCBP2, otherwise known as hnRNP-E2 or α CP-2, is located on chromosome 12q13.13 as a 15-exon entity, encoding a protein 365 amino acids in length. In normal tissues, *PCBP2* expression is ubiquitous with greatest expression in the bone marrow²²⁰ (**Figure 4.1A**). Multiple *PCBP2* isoforms are known to exist via alternative splicing, diversifying *PCBP2*'s transcript repertoire (**Figure 4.1B**). Two *PCBP2* pseudogenes, *PCBP2P1* and *PCBP2P2*, are derived from retrotransposition events and are located at dispersed loci²¹⁸.

Like all PCBP proteins, PCBP2 is comprised of 3 KH-domains (KHI, KHII, KHIII), each identified to be capable of nucleic acid binding²²¹⁻²²³, and encoded by a distinct subset of exons²²⁴ (**Figure 4.1B**). PCBP2 also contains two functionally independent nuclear localisation signals (NLS), facilitating PCBP2's movement between nuclear and cytoplasmic compartments, with NLS1 mapping between KH II-KH III domains and NLS2 located within KH III²²⁴. The linker domain between KH II and KH III also contains three WW-binding (WB) domains, which facilitate binding to WW-domain containing proteins²²⁵. PCBP2's KH-domains contain 65-70 amino acids with a defined secondary

structure characterised by three-stranded antiparallel beta-sheets packed against 3-alpha helices in the configuration, $\beta_1\text{-}\alpha_1\text{-}\alpha_2\text{-}\beta_2\text{-}\beta_3\text{-}\alpha_3$ ^{221-223,226} (**Figure 4.1C**). Recognition of poly(C)-rich motifs by the KH-domains occurs within a highly conserved *GXXG* loop located between alpha helices, $\alpha_1\text{-}\alpha_2$ ²²⁶. Protein-protein interactions are also facilitated through PCBP2's KH-domains and the dual function of domains KH I and KH II, but not KH III, have been reported²²². Since PCBP2 contains no additional functional protein motifs, following nucleic-acid binding, association with additional proteins via KH-domains and WB-domains is likely necessary for PCBP2 regulation and functional downstream consequences.

4.1.2 The role of PCBP2 in cancer

Oncogenic roles of PCBP2 have been described across a variety of cancers including hepatocellular carcinoma²²⁷, glioma²²⁸, gastric²²⁹ and pancreatic²³⁰ cancers. *PCBP2* overexpression associates with poorer survival outcomes in all described cancer types and in glioma and pancreatic cancer, *PCBP2* represents a biomarker for advanced disease stages^{227–230}.

In hepatocellular carcinoma, *PCBP2* is upregulated in human tissue and associated with poor prognosis²³¹. Furthermore, depletion of *PCBP2* by siRNA oligos was shown to inhibit proliferation of hepatocellular carcinoma cell line models, HepG2 and Huh7, leading to G₀/G₁ cell cycle arrest²³¹. *PCBP2* depletion was also found to enhance sensitivity to sorafenib treatment, and was associated with an upregulation of Bax and downregulation of Bcl-2, pro-apoptotic and anti-apoptotic proteins, respectively. Hence, this suggests *PCBP2* plays a role in regulating expression of apoptotic Bcl-2 family proteins in hepatocellular carcinoma²³¹.

PCBP2 has also been shown to support glioma cell proliferation, migration and invasion via a number of different mechanisms²³². For example, *PCBP2* was identified to enhance glioma growth by directly binding and negatively regulating the stability of pro-apoptotic *FHL3* transcripts²³³. In addition, *PCBP2* expression in glioma is negatively regulated by microRNA, miRNA-214²³⁴. *PCBP2* expression was shown to be inversely correlated to miRNA-214 expression and antisense oligos targeting miRNA-214, which bind and prevent miRNA function, led to upregulation of *PCBP2*. Finally, upregulation of *PCBP2* in the absence of tumour suppressive miRNA-214 regulation led to increased proliferation and invasion by glioma cell line models²³⁴.

Overall, functional studies implicate aberrant *PCBP2* expression in cancer, with overexpression contributing to enhanced cell proliferation and survival in a variety of cancer subtypes. As shown, the molecular mechanisms by which *PCBP2* exerts its oncogenic effects are diverse, highlighting the cell and tissue-specific context of *PCBP2* oncogenic functions. Overall, evidence from multiple carcinomas exemplify the potential

role of PCBP2 overexpression in the diagnosis, prognosis and treatment of various cancers.

4.1.3 PCBP2 in normal and malignant haematopoiesis

In the haematopoietic hierarchy, *PCBP2* expression is lowest in the HSC compartment and greatest in MEP populations²³⁵. Through multiple mechanisms of PTGR and protein-protein interactions, PCBP2 has become a defined regulator of haematopoiesis, particularly in erythropoiesis. *PCBP2*-null mice are non-viable, and fail to survive past mid-gestation (day 12.5-13.5), while *PCBP2* heterozygous mice develop normally²³⁶. Haematopoietic defects, specifically loss of the definitive erythroid lineage and impaired megakaryopoiesis, are responsible for the lack of survival in *PCBP2*-null mice, highlighting the importance of PCBP2 in murine development via its roles in haematopoiesis²³⁶. Mechanistically, PCBP2 was identified to regulate splicing of the transcription factor, *Runx1*, a key regulator of erythroid lineage commitment. Depletion of *PCBP2* led to the absence of PCBP2 at exon 6 splice acceptor sites within *Runx1* transcripts and enhanced exon 6 exclusion from *Runx1* transcripts, leading to loss of erythroid lineage commitment²³⁷. Similarly, mice homozygous for *Runx1* exon 6 deletion presented haematopoietic abnormalities mid-gestation, phenocopying *PCBP2*-null mice²³⁷.

In contrast to *PCBP2*-null models, conditional knockout of *PCBP2* in the erythroid lineage using *EpoR*-Cre, halting *PCBP2* expressing from day 12.5 onwards, has no effect on survival with adult mice presenting normal haematopoietic parameters²³⁸. In this conditional *EpoR*-Cre model, splicing of *Runx1* was observed in erythroid cells at day 14.5, however, did not lead to defective haematopoiesis²³⁸. This questions the relationship between PCBP2 and *Runx1* for definitive haematopoiesis, which is located in the murine bone marrow from day 15. One possibility is that *Runx1* is less required for definitive haematopoiesis at day 12.5 when *PCBP2* is conditionally deleted from the erythroid lineage. Another reason why we may not see a requirement for *PCBP2* in adult mice is that *PCBP2* deletion in null mouse models impacts other lineages required for development in addition to the erythroid lineage. Independent single-cell atlases of the murine haematopoietic and stromal compartments show *PCBP2* expression is ubiquitous, with

the exception of neutrophil and mature erythrocyte populations where expression appears to be reduced²³⁹. All adult blood cells express *Runx1*, except mature erythrocytes²⁴⁰, hence whether PCBP2 is required by other haematopoietic lineages in adult haematopoiesis and if so, whether this requirement is depending on *Runx1* splicing requires further investigation.

The role of PCBP2 in erythropoiesis is not limited to splicing of *Runx1*. PCBP2 is also important for the stability of mature α -globin mRNA during terminal erythroid differentiation through the formation of single PCBP2 protein RNP complexes at α -globin's poly(C)-rich 3'UTR, termed the α -complex^{241,242}. Independently, the α -complex has been identified to sequentially control both nuclear processing and cytoplasmic stability of α -globin transcripts. In the nucleus, PCBP2 co-transcriptionally associates with the poly(C)-rich 3' terminus of α -globin mRNA to enhance the recruitment of 3' processing machinery for pre-mRNA cleavage and polyadenylation²⁴³. While in the cytoplasm, α -complex formation within the 3'UTR is crucial for maintaining α -globin's exceptionally long-lasting transcript stability throughout terminal erythroid differentiation²⁴². In a similar manner, an α CP complex (with PCBP1 or PCBP2) has been reported to stabilise β -globin mRNA, although the involvement of specific PCBP family members remains to be fully described²⁴⁴.

Through protein-protein interactions, PCBP2 presents a gate keeper of cellular iron homeostasis. Erythroid progenitors require exceptional amounts of iron to be imported and localised to the mitochondria for heme biosynthesis. However, iron flux must be tightly regulated to coordinate efficient haemoglobin formation while preventing the toxic effects of excessive chemically reactive iron. DMT1, a transmembrane importer of ferrous iron, was identified as a protein binding partner of PCBP2²⁴⁵. The DMT1-PCBP2 interaction was identified to be iron-dependent and essential for the cellular uptake of iron²⁴⁵. Research showed ferrous iron was directly transferred from transmembrane DMT1 to cytosolic PCBP2, suggesting ferrous iron is not just liberated into the cytoplasm but transferred to chaperone proteins such as PCBP2²⁴⁵. PCBP2 was also identified to interact with the iron exporter, FPN1²⁴⁶. This interaction is similarly necessary for iron export and is iron-dependent, with iron-loaded, but not iron-depleted PCBP2, interacting with FPN1²⁴⁶. Finally, PCBP2 has been identified as an iron chaperone for the delivery of

iron to Ferritin, a cellular iron storage protein²⁴⁷. Overall, the importance of PCBP2 in iron transport and storage have been presented, however, structural investigation of iron-bound PCBP2 structures is required.

Despite the high expression of PCBP2 in the bone marrow (**Figure 4.1A**), and characterised PTGR roles within the erythroid lineage, the requirement of PCBP2 in other haematopoietic lineages is less known. Understanding PCBP2's roles in myeloid lineages predominantly comes from studies of myeloid malignancy, particularly blast-crisis CML (BC-CML). Differentiation arrest of myeloid progenitors is a common characteristic of blast-crisis progression in CML. Downregulation of C/EBP α , a master transcription factor of key granulocytic differentiation genes such as *CSF3R*, has been shown to drive the blast-crisis differentiation blockage in BC-CML²⁴⁸. In a BCR-ABL dependent manner, PCBP2 was identified to be upregulated in BC-CML and directly responsible for C/EBP α downregulation via translational suppression of *CEBPA* transcripts²⁴⁸. More specifically, PCBP2 expression was BCR-ABL dose and kinase dependent, requiring high-levels of BCR-ABL expression to drive MAPK^{ERK1/2} phosphorylation and stabilisation of PCBP2 via key EKR phosphorylation sites²⁴⁹. Thus, a BCR/ABL-MAPK^{ERK1/2}-PCBP2 axis exists to suppress myeloid differentiation in transformed CML. Given the highly-specific BCR-ABL dependent context described, the requirement of PCBP2 in differentiation of normal haematopoiesis and non-BCR-ABL driven myeloid malignancies requires further investigation.

4.1.1 Hypothesis

PCBP2 is an RBP whose overexpression in *TP53*-sAML promotes cancer cell maintenance and survival, the requirement of which can be functionally validated *in vitro* using model cancer cell lines.

4.1.2 Summary and aims

To consider a candidate RBP for therapeutic targeting in *TP53*-sAML, a selective requirement must exist between WT-HSPC and *TP53*-mutant LSC populations, to minimise undue treatment toxicity. Extensive literature review and access to external datasets can give an insight into the requirements of candidate genes. Based on this knowledge, in this thesis, I selected PCBP2 as a potential therapeutic target in *TP53*-sAML for further investigation.

Aims for this section are as follows:

1. To evaluate and select a promising candidate RBP for functional validation
2. To investigate the functional dependency of PCBP2 in *TP53*-sAML using cancer cell line models

4.2 Results

4.2.1 Investigating the requirement of candidate RBP in *TP53*-WT HSPC

To consider any candidate RBP for therapeutic targeting, a therapeutic index must exist between *TP53*-mutant LSC and *TP53*-WT HSPC populations. To assess the requirement of candidate RBP in *TP53*-WT HSPC, and identify a candidate for further functional characterisation, *ex vivo* genome-wide murine CRISPR-Cas9 screening data was shared by Professor Adam Wilkinson. Murine bone marrow HSPC (c-Kit+) were expanded for 3-weeks using a long-term *ex vivo* expansion protocol²⁵⁰. Fifty-million HSPC were then transduced with genome-wide CRISPR-screening library, including the candidate RBP identified in this thesis. Transduced HSPC were then selected using puromycin for 72-hours. Selected HSPC populations remained in *ex vivo* HSC expansion cultures for 21-days with c-Kit+Sca-1+Lineage- HSPC isolated by FACS to extract DNA and determine guide representation every 7-days. Access to this data allowed the requirement of each RBP in a *TP53*-WT HSPC setting to be assessed (**Figure 4.2A**). Any candidate showing high essentiality in this setting – for example *SFPQ* and *YARS1* – were disregarded at this time as potential targets.

Further considerations for selecting a therapeutic candidate included whether a target was druggable, the requirement in null and conditional mouse models, and previously reported roles in tumorigenesis. A comprehensive summary of each candidate RBP can be found in **Table 4.1** (at end of chapter).

As determined by differential gene expression analysis of the *TP53*-sAML TARGET-seq dataset, *PCBP2* expression was identified to be significantly upregulated in *TP53*-mutant LSC (*TP53*-sAML) compared to preleukaemic cells from the antecedent clone, *TP53*-WT MPN patients and HD controls (**Figure 4.2B**). Furthermore, utilising the BloodSpot database²³⁵ to compare *PCBP2* expression in a cohort of complex karyotype AML cases and healthy donor HSC showed a significant increase in *PCBP2* expression in complex karyotype AML (**Figure 4.2C**). Together, heightened expression of *PCBP2* in *TP53*-mutant LSC, differential *PCBP2* expression in an independent cohort of complex karyotype AML

compared to HD HSC and minimal requirement for *PCBP2* in murine *TP53*-WT HSPC suggests a potential selective requirement for *PCBP2* exists. Therefore, following an in-depth literature review, I determined *PCBP2* was a promising candidate for further functional validation.

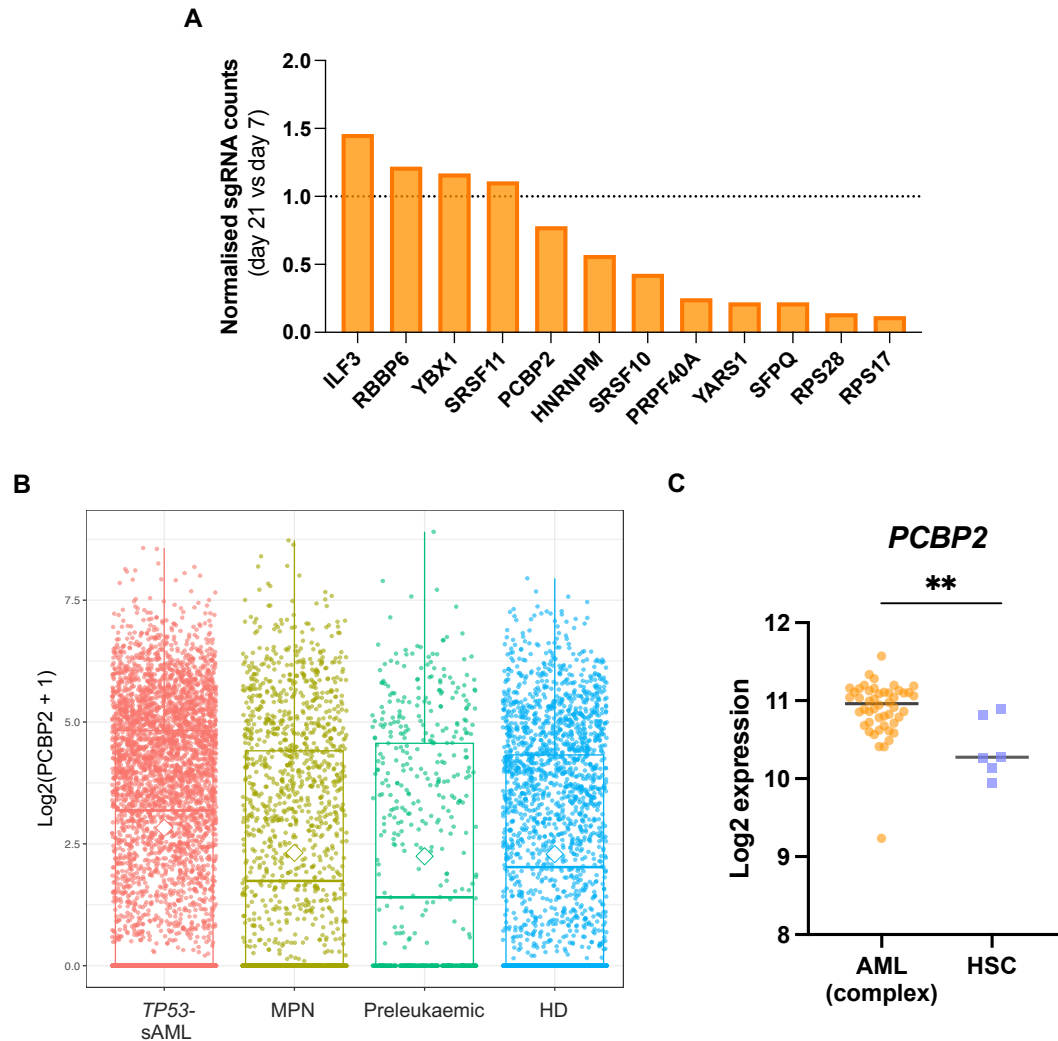


Figure 4.2. Requirement and expression of *PCBP2* in WT-HSPC **(A)** Requirement of candidate RBP in murine *ex vivo* CRISPR-Cas9 screen. Reads were median normalised and represent an average of two sgRNA/gene. Data shared by Professor Adam Wilkinson, Weatherall Institute of Molecular Medicine, University of Oxford. **(B)** *PCBP2* expression in single-cell TARGET-seq dataset. Each data point represents a single cell. Central diamonds represent mean $\log_2(PCBP2+1)$ expression values. Preleukemic cells represent the antecedent *TP53*-wildtype clone in *TP53*-sAML patients. Myeloproliferative neoplasm (MPN) and healthy donor (HD) controls are *TP53*-WT. **(C)** Expression of *PCBP2* in a cohort of complex karyotype AML cases compared to HD HSC derived from the BloodSpot database²³⁵. Significance calculated using a two-way unpaired Student *t*-test where $**p < 0.01$.

4.2.2 Evaluating *PCBP2* expression and effect in publicly available AML cohorts

Next, I evaluated whether the association between *TP53*-mutant LSC and *PCBP2* upregulation was observed more widely in publicly available AML cohorts. To do this, I utilised transcriptomic data from the BeatAML patient cohort, comprising 377 patient samples¹⁴⁷. Stratifying the BeatAML cohort by *TP53*-mutation showed a significant upregulation of *PCBP2* in *TP53*-WT AML patients (n=343) compared to *TP53*-mutant AML (n=32; **Figure 4.3A**). *PCBP2* expression was also analysed in available sAML samples where patients had a known precursor neoplasm, most commonly myelodysplastic syndrome (MDS). In this sAML cohort, a similar effect size between *TP53*-WT AML (n=49) and *TP53*-mutant AML cases (n=10) was observed (**Figure 4.3B**), suggesting *PCBP2* expression is high regardless of *TP53*-mutation and/or disease status.

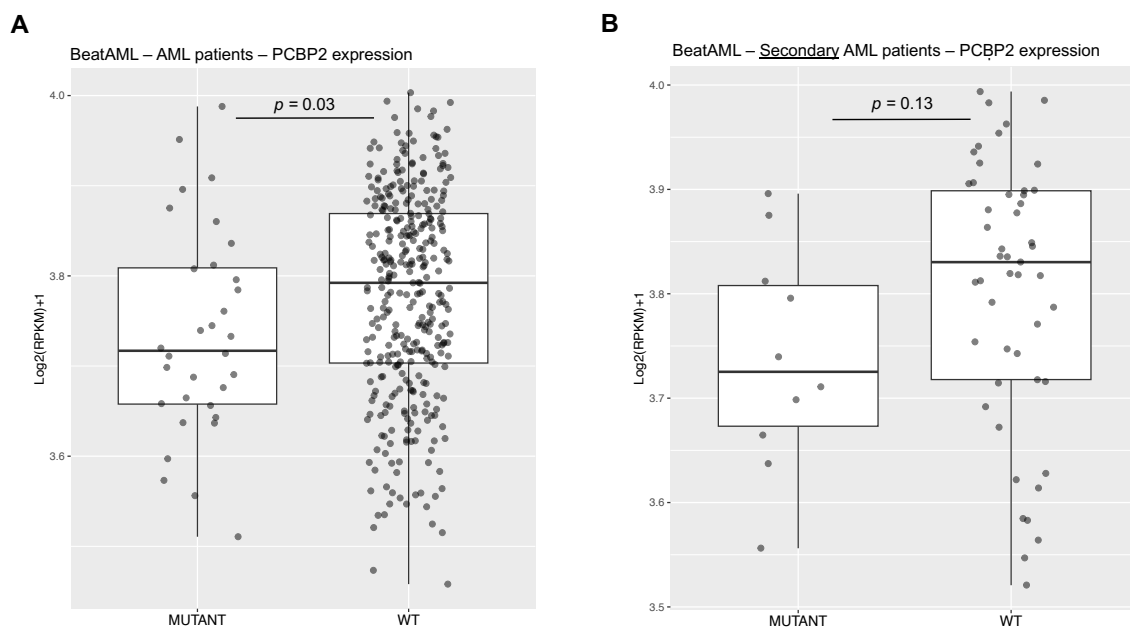


Figure 4.3. *PCBP2* expression in the BeatAML patient cohort stratified by *TP53*-mutation status. (A) *PCBP2* expression in *TP53*-WT (WT, n=343) AML patient cohort compared to *TP53*-mutant (MUTANT, n=32) AML patients. (B) *PCBP2* expression stratified by *TP53*-mutation status only in secondary AML cases (MUTANT n=10, WT n=49). Secondary AML refers to patients with a known precursor neoplasm, most commonly myelodysplastic syndrome. Each data point represents an individual patient, *p*-values calculated via unpaired two-way Student's *t*-test. RPKM; Reads per Kilobase per Million.

Progression to sAML presents patients with a particularly poor prognosis⁹. To understand whether high *PCBP2* expression associates with adverse survival outcomes in either *TP53*-mutant or *TP53*-WT sAML cases, BeatAML patients were stratified by median *PCBP2* expression into high and low *PCBP2* expressing cohorts (median $\log_2[\text{RPKM}]+1$ expression 3.817). Interestingly, in neither *TP53*-mutant or *TP53*-WT cohorts did high *PCBP2* expression significantly influence survival probability in sAML patients (**Figure 4.4**). Although, in *TP53*-mutant patients the association trended towards poorer overall survival outcomes in *PCBP2* high expressors (**Figure 4.4**; Cox regression analysis, *TP53*-Mutant $\text{HR}^{\text{PCBP2}(\text{LOW})}$ 0.48 [95% CI 0.08-2.65, *p*-value 0.4]; *TP53*-Wild type $\text{HR}^{\text{PCBP2}(\text{LOW})}$ 1.03 [95% CI 0.53-1.92, *p*-value 0.92]).

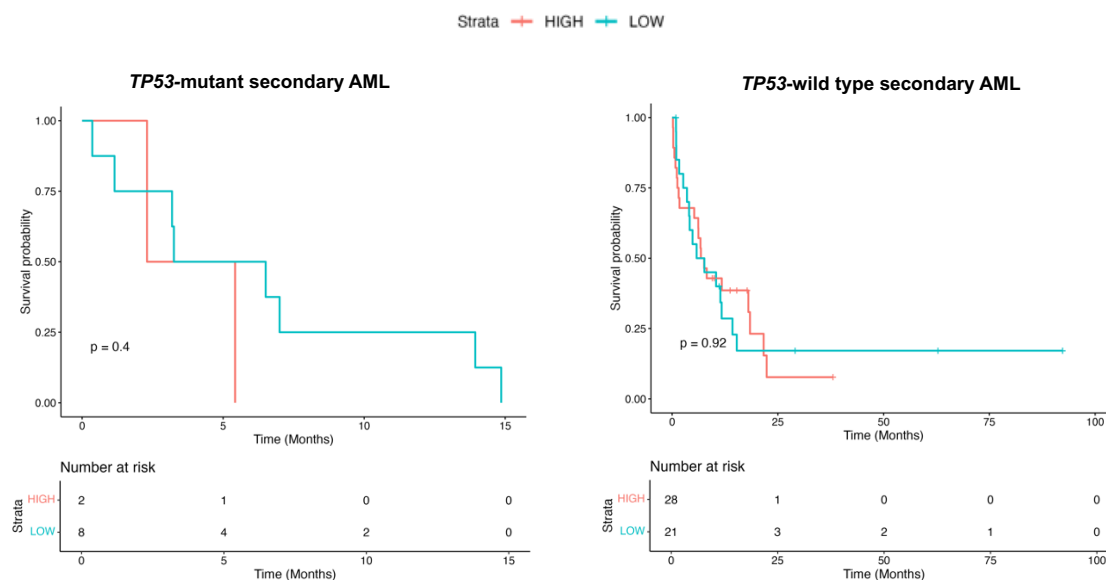


Figure 4.4. Survival probability by *PCBP2* expression in a *TP53*-stratified secondary AML from the BeatAML dataset. Survival outcomes of secondary AML patients from the BeatAML cohort by *PCBP2* expression when stratified by *TP53* mutation status. High and low *PCBP2* expressing patients determined via median *PCBP2* expression ($\log_2[\text{RPKM}]+1$ value of 3.817) across the BeatAML cohort. Cox regression analysis, (*TP53*-Mutant $\text{HR}^{\text{PCBP2}(\text{LOW})}$ 0.48 [95% CI 0.08-2.65, *p*-value 0.4]; *TP53*-WT $\text{HR}^{\text{PCBP2}(\text{LOW})}$ 1.03 [95% CI 0.52-1.98, *p*-value 0.92]). HR, hazard ratio; RPKM, Reads per Kilobase per Million mapped reads.

To corroborate analysis of the BeatAML dataset, *PCBP2* expression and its association with adverse survival outcomes was also evaluated in The Cancer Genome Atlas (TCGA) AML cohort (TCGA-LAML)²⁵¹. Both BeatAML and TCGA datasets include genetic sequencing (whole-genome, targeted and transcriptome), however, the BeatAML project is more clinically focussed, incorporating drug sensitivity data to accelerate precision

medicine in AML. In contrast, genomic profiling in the TCGA cohort is more specifically focussed on uncovering the underlying biological basis of AML. Expression analysis in the TCGA-LAML cohort reflected the trend observed in the BeatAML cohort, with elevated *PCBP2* expression in *TP53*-WT AML patients (n=79) compared to *TP53*-mutant AML patients (n=10) (**Figure 4.5A**). And while TCGA analysis could not be classified into *de novo* and sAML cohorts, as in the BeatAML analysis, when patients with survival data were stratified by *PCBP2* expression into high and low expressing cohorts (median $\log_2[\text{RPKM}]+1$ expression value 4.61), no significant differences in survival outcomes were observed between high and low *PCBP2* expressing cohorts in *TP53*-mutant AML patients (n=9) or in the *TP53*-WT cohort (**Figure 4.5B**; n=125; Cox regression analysis, *TP53*-Mutant $\text{HR}^{\text{PCBP2(LOW)}}$ 0.49 [95% CI 0.05-4.5, *p*-value 0.59]; *TP53*-WT $\text{HR}^{\text{PCBP2(LOW)}}$ 1.12 [95% CI 0.71-1.76, *p*-value 0.61]).

Overall, analysis of external AML datasets may suggest *PCBP2* is not important for survival or outcome in *TP53*-mutant AML, however, in contrast this analysis also emphasises the importance of single-cell approaches for the discovery of novel molecular vulnerabilities. Bulk transcriptomic analysis of whole bone marrow or peripheral blood samples, such as in TCGA and BeatAML datasets, fails to capture the considerable tumour heterogeneity present in AML patient cohorts. In external AML datasets, elevated *PCBP2* in *TP53*-WT cell populations compared to *TP53*-mutant populations, suggests *PCBP2* expression is upregulated independent of *TP53*-mutation status. Heightened *PCBP2* expression also appears to have no significant impact on survival in either *TP53*-WT or *TP53*-mutant cohorts. However, bulk sequencing analysis obscures interpretation of sample composition and cell proportions and is unable to isolate expression profiles from specific cell populations of interest. For example, bulk RNA-seq analysis does not take into account; (1) percentage of leukaemic blasts which will alter the relative proportion of remaining non-malignant cells and therefore expression values for any given gene, (2) *PCBP2* expression in residual healthy cells and 3) expression of *PCBP2* in the disease-propagating *TP53*-mutant LSC population. Thus, bulk RNA-seq analysis cannot delineate malignant and non-malignant cell populations and their associated transcriptomic profiles, and therefore tumour heterogeneity in publicly available bulk RNA-seq datasets cannot be accounted for.

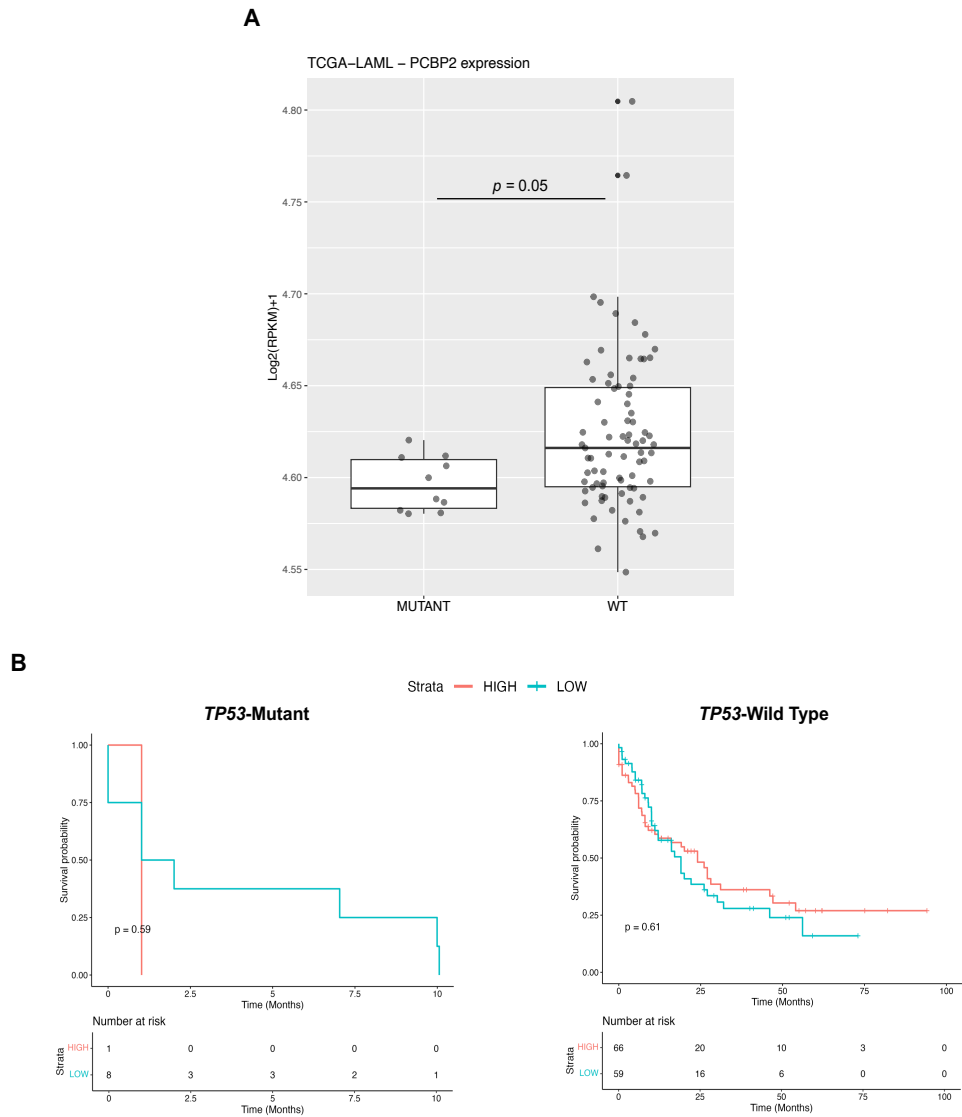


Figure 4.5. *PCBP2* expression and impact on survival in the TCGA-LAML cohort stratified by *TP53*-mutation status. (A) *PCBP2* expression in *TP53*-WT AML (n=79) compared to *TP53*-mutant patients (n=10). Each data point represents *PCBP2* expression in an individual patient with p -value calculated using unpaired two-tailed Student's t -test. (B) Survival outcomes in TCGA-LAML cohort by *PCBP2* expression when stratified on *TP53*-mutation status. High and low *PCBP2* expressors defined by median *PCBP2* expression value ($\log_2[\text{RPKM}]+1$ value of 4.61) across TCGA-LAML cohort. Cox regression analysis (TP53-Mutant $\text{HR}^{\text{PCBP2(LOW)}}$ 0.49 [95% CI 0.05-4.5, p -value 0.59]; TP53-WT $\text{HR}^{\text{PCBP2(LOW)}}$ 1.12 [95% CI 0.71-1.76, p -value 0.61]). HR, hazard ratio; RPKM, Reads per Kilobase per Million mapped reads.

Therefore, while I did not observe *PCBP2* to be specifically upregulated in *TP53*-mutant AML populations, this did not deter my selection of *PCBP2* as a candidate gene for further investigation. On the contrary, *PCBP2* is an increasingly interesting target following this analysis. Using a TARGET-seq approach, I have identified *PCBP2* to be specifically and

significantly upregulated in disease-propagating *TP53*-mutant LSC compared to preleukaemic *TP53*-WT cells using inpatient comparisons. From analysis of external datasets, it is clear this association between *TP53*-mutant LSC and *PCBP2* would not have been discovered. Thus, the requirement of *PCBP2* could have not been determined in *TP53*-sAML without utilisation of a single-cell multi-omic approach, such as TARGET-seq.

4.2.3 Investigating the expression and dependency of *PCBP2* across cancer cell line models

To further evaluate *PCBP2* as a worthwhile candidate gene for functional validation, I looked more closely at the expression and requirement of *PCBP2* in cancer cell line models using the DepMap database¹⁴¹. Firstly, myeloid cancer cell lines were amongst the highest expressors of *PCBP2* at the transcript and protein level (**Figure 4.6**).

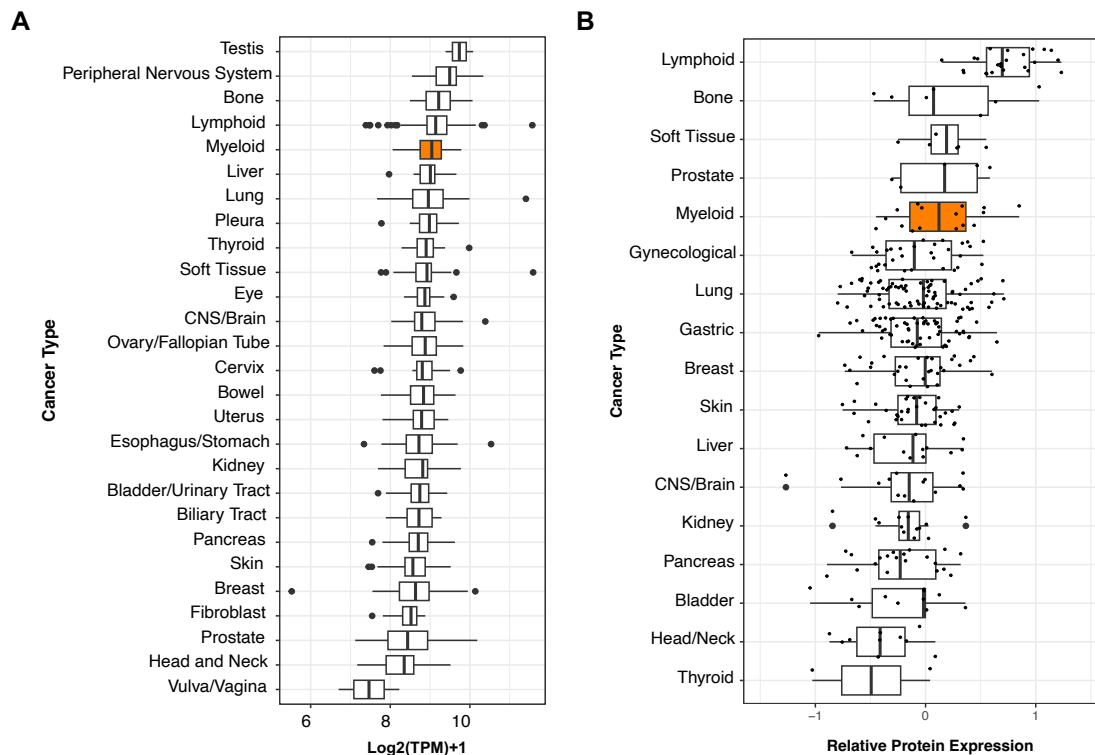


Figure 4.6. *PCBP2* expression across cancer types. (A) *PCBP2* mRNA expression, expressed as $\log_2(\text{TPM}) + 1$ and (B) protein expression across cancers. TPM; transcripts per million. Data downloaded from the DepMap database¹⁴¹.

Looking more closely at the functional requirement of *PCBP2*, *PCBP2* is a common essential gene whose essentiality is indifferent between *TP53*-mutant and *TP53*-WT cancer cell lines (**Figure 4.7A**). Focusing on AML cell lines, the requirement for *PCBP2* appears to increase in *TP53*-mutant AML cell lines compared to *TP53*-WT counterparts (**Figure 4.7B**), however limited *TP53*-WT cell lines were available for analysis. Furthermore, the sAML cell lines, HEL and SET-2, which harbour *JAK2* and *TP53* mutations, reflecting the genetic and phenotypic characteristics of our *TP53*-sAML cohort, show the highest expression and greatest dependency for *PCBP2* in the available AML cell line cohort (**Figure 4.7C**). Taken together, this data presents *PCBP2* as a novel candidate gene supporting leukaemic maintenance.

A therapeutic index for targeting candidate genes is essential and a selective requirement for *PCBP2* in *TP53*-mutant LSC populations is suggested by a minimal loss of *PCBP2* sgRNA in *ex vivo* murine *TP53*-WT HSPC loss-of-function screens. Furthermore, a significant correlation between high *PCBP2* expression and dependency is observed in cancer cell line models, with the greatest correlation in *JAK2/TP53*-mutant blast-phase MPN cell line models. Thus, *PCBP2* was taken forward as a candidate RBP for functional validation in *TP53*-sAML, utilising relevant HEL and SET-2 cell line models for *in vitro* investigations.

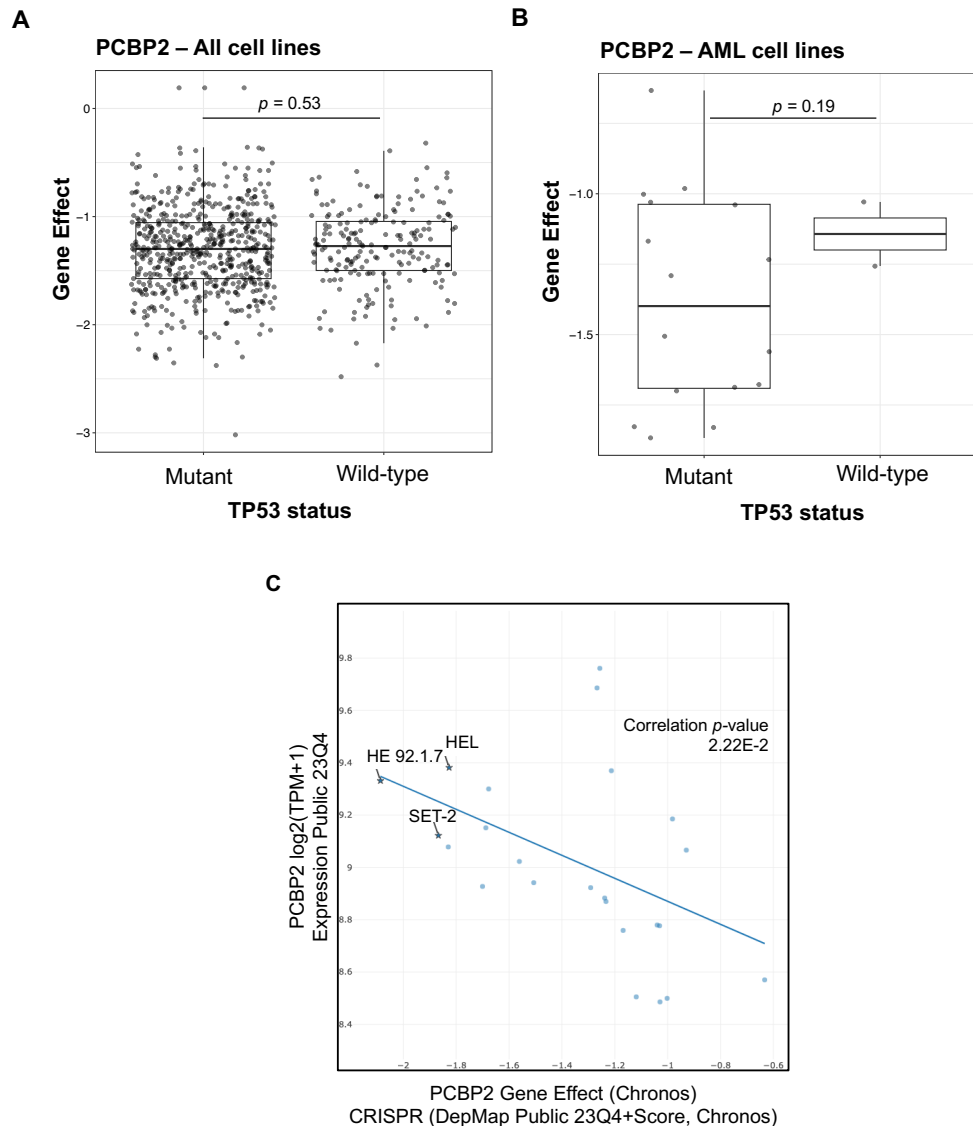


Figure 4.7. Interrogating expression and requirement of *PCBP2* in cancer cell lines using the DepMap database. (A) *PCBP2* essentiality across all DepMap cancer cell lines, stratified by *TP53* mutation status ($N = 586$, *TP53*-mutant; $N = 191$, *TP53*-WT) **(B)** *PCBP2* requirement confined to AML cell lines only and stratified by *TP53* mutation status ($N = 16$, *TP53*-mutant; $N = 2$, *TP53*-WT). Negative Gene Effect scores < -0.5 indicate gene essentiality. p -values calculated by unpaired two-way Student's t -test. **(C)** Linear regression analysis shows correlation between *PCBP2* expression and dependency in AML cell lines (Pearson correlation coefficient -0.465 , slope $-4.40E-1$, p -value < 0.05). *JAK2/TP53*-mutant sAML model cell lines are annotated and show highest *PCBP2* expression and greatest gene dependency. WT, wild-type.

4.2.4 PCBP2 presents a functional requirement in *JAK2/TP53*-mutant cell line models

To investigate the functional dependency of PCBP2 in *TP53*-sAML, a lentiviral short-hairpin RNA (shRNA) approach was adopted in *JAK2/TP53*-mutant HEL and SET-2 cell lines to knockdown (KD) *PCBP2* transcripts. A lentiviral shRNA approach was selected due to the utility of the RFP reporter included in the transfer plasmid backbone to track downstream functional consequences of *PCBP2*-KD. Using this approach, *PCBP2* depletion was highly effective at the mRNA and protein level in both cell lines using either PCBP2 KD #1 or PCBP2 KD #2 shRNA – targeting the open reading frame or 3'UTR, respectively – compared to scrambled shRNA control (**Figure 4.8**).

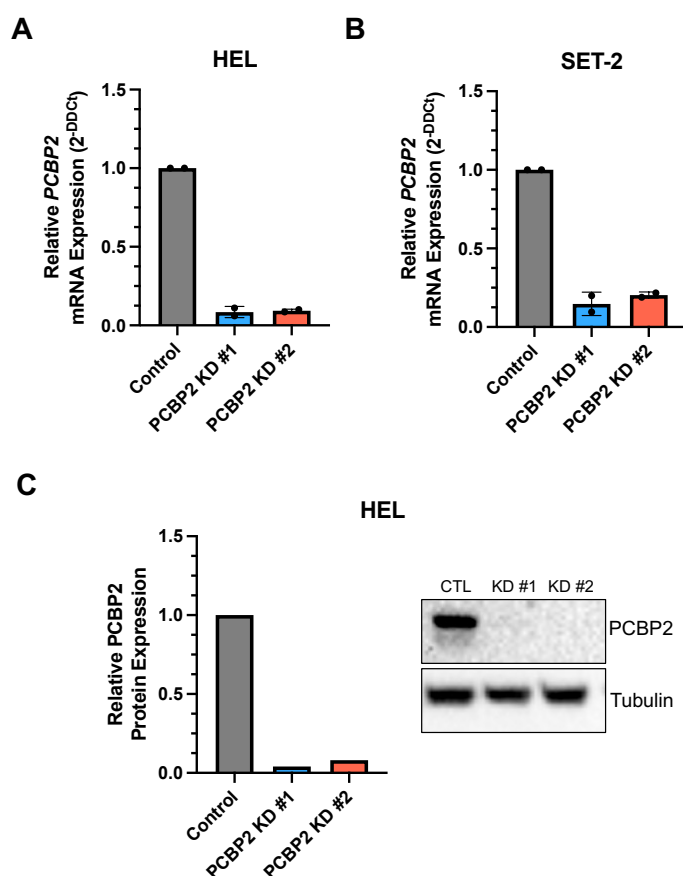


Figure 4.8. *PCBP2* depletion in cancer cell line models by lentiviral shRNA. Relative *PCBP2* mRNA expression following lentiviral transduction with either PCBP2 KD #1 or #2 shRNA compared to scrambled shRNA control determined by RT-qPCR in **(A)** HEL and **(B)** SET-2 cell lines. **(C)** Relative PCBP2 protein expression following lentiviral transduction as above determined by western blotting. Representative western blot in the HEL cell line. All samples were normalised to Tubulin loading controls before expression was compared relative to scrambled Control (CTL).

Having validated the knockdown capabilities of *PCBP2*-shRNA constructs, the functional requirement of *PCBP2* for cell line survival and proliferation was investigated. *PCBP2* depletion significantly increased the proportion of non-viable HEL and SET-2 cells compared to controls in Annexin V/DAPI apoptosis assays (**Figure 4.9**)

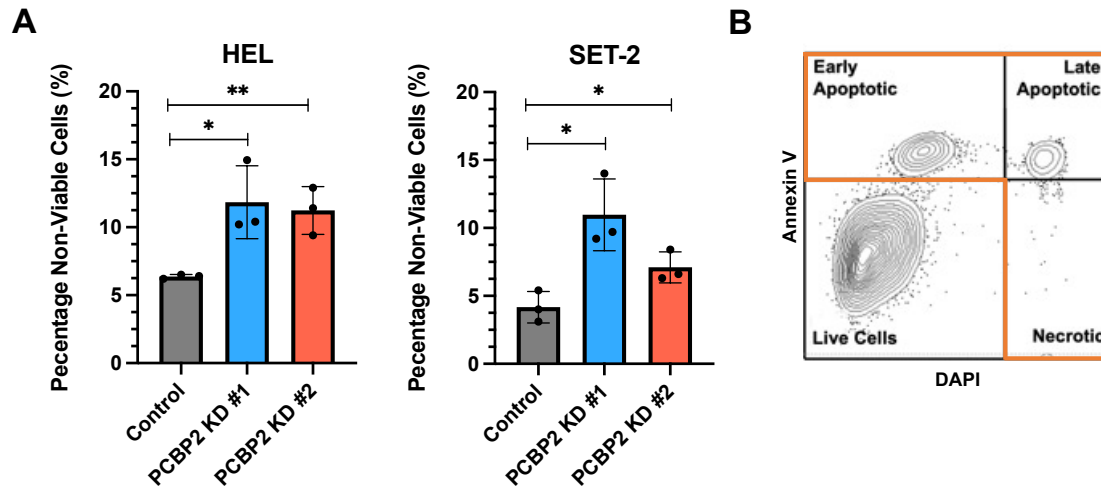


Figure 4.9. Effect of *PCBP2* depletion on model cell line survival. (A) HEL and SET-2 cell lines stained with Annexin V and DAPI to determine effect of *PCBP2*-KD on cell line survival. Graphs represent non-viable cells (early apoptotic, late apoptotic and necrotic) 5-days post-transduction. Data represents mean values $\pm$ SD of three independent experiments where $*p < 0.05$ and $**p < 0.01$ calculated using unpaired two-tailed Student's *t*-test. (B) Representative Annexin V/DAPI gating strategy. Non-viable populations bordered in orange. KD, knockdown; SD, standard deviation.

The effect of *PCBP2*-KD on cell proliferation was also investigated using a competitive cell culture (CCG) assay. Since lentiviral vectors harbouring shRNA sequences also include an RFP reporter, transduced *PCBP2*-KD populations can be distinguished from parental non-transduced counterparts. Following transduction, cells were maintained in mixed transduction conditions, including both RFP-positive and RFP-negative fractions, allowing competitive dynamics between the two populations to be assessed. During culture, total cell numbers were consistently maintained between passages. Overtime, results showed *PCBP2*-KD, with either shRNA construct, led to a gradual but significant decrease in RFP-positive populations overtime in HEL and SET-2 cell lines (**Figure 4.10A-B**). Despite a similar level of *PCBP2*-KD by both *PCBP2* KD #1 and KD #2 shRNA, the difference when targeting with *PCBP2* KD #2 more pronounced. Importantly, positive and negative fractions were sorted via fluorescently activated cell sorting (FACS) following the

final readout and *PCBP2* expression confirmed to be sufficiently maintained in RFP-negative populations compared to RFP-positive populations (**Figure 4.10C-D**). Hence, *PCBP2*-KD leads to loss of targeted cells in competitive culture conditions with parental WT counterparts, suggesting *PCBP2*-KD leads to loss of cells from culture by increased apoptosis as shown above but also through longer-term effects of cell proliferation. Given the limited effect of *PCBP2*-KD on apoptosis, and data (not shown) suggesting *PCBP2*-KD reduces metabolic activity in cell lines, it is likely the major effect of *PCBP2*-KD is driven by a reduction in cell proliferation over time.

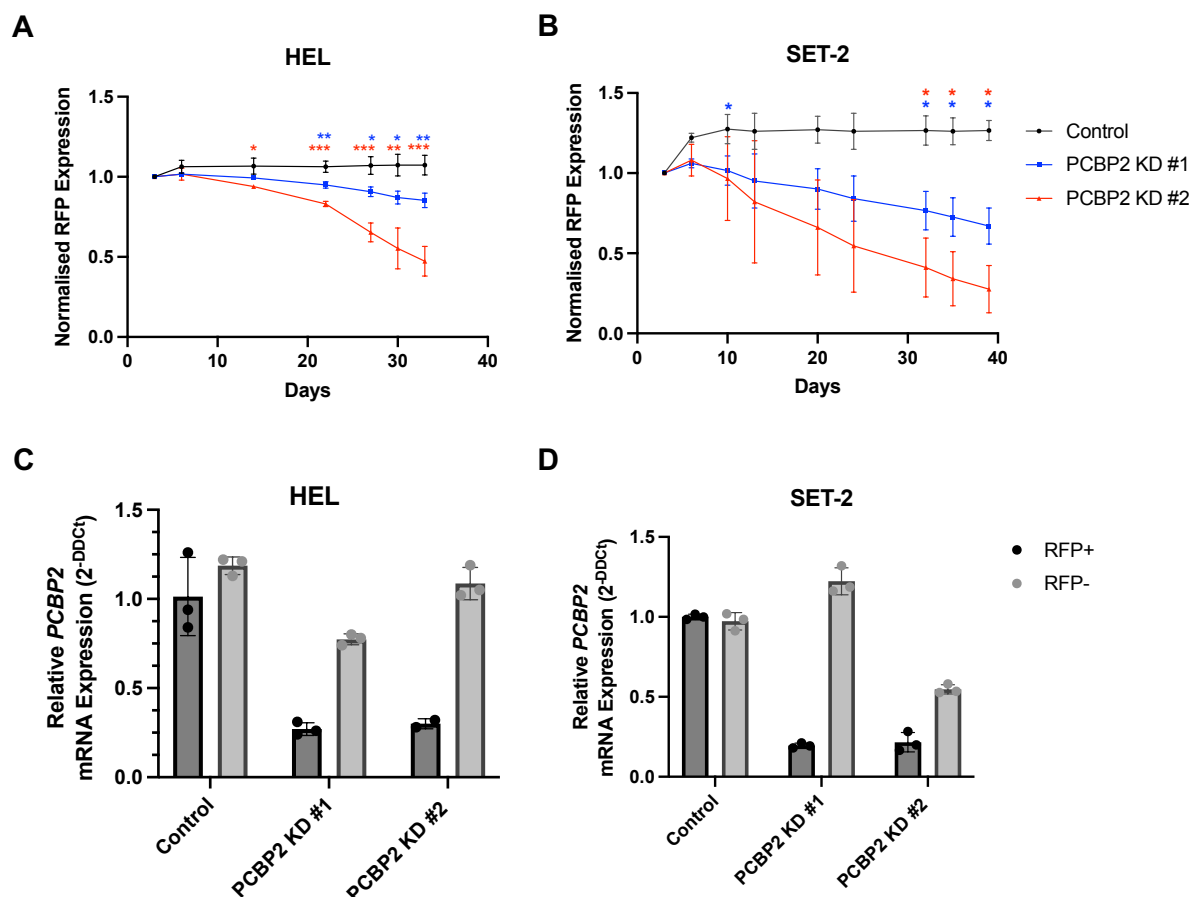


Figure 4.10. Competitive disadvantage of *PCBP2* depleted cell lines. (A-B) HEL and SET-2 competitive cell growth (CCG) assays. Cells were transduced and maintained in mixed transduction conditions with untransduced (RFP-) parental cells for up to 40-days. Total cell numbers maintained consistent throughout. Percentage RFP+ cells determined by flow cytometry at regular intervals. RFP percentage normalised to that of the first timepoint. Significance determined by two-way unpaired Student's *t*-test where * $p < 0.05$, ** $p < 0.001$, *** $p < 0.000.1$. Data points represent mean $\pm$ SD for three and two independent experiments, for HEL and SET-2 cell lines, respectively. **(C-D)** Relative *PCBP2* expression in RFP+ fractions compared to RFP- fractions assessed at experimental end-points. *PCBP2* expression for all samples normalised to RFP+ control per cell line. Data points represent three technical RT-qPCR replicates from a single experiment.

Given the difference in phenotypes resulting from the two *PCBP2*-targeting shRNA constructs and the known off-target effects of shRNA approaches²⁵², it is possible the enhanced phenotype observed in cells transduced with *PCBP2* KD #2 shRNA is due to off-target effects. Therefore, the requirement of *PCBP2* in model cell lines was validated using a CRISPR-Cas9 knockout (KO) approach. Here, the *PCBP2* loci was perturbed by one of two targeting sgRNA via RNP-nucleofection. DNA was extracted from cells 7-days post-editing and again after 21-days in culture. Similar to loss-of-function screens, editing at the *PCBP2* locus was significantly reduced over-time due to the loss of edited cells from culture, confirming the requirement for *PCBP2* expression in model cell lines (**Figure 4.11**). In contrast, editing of the control *AAVS1* safe harbour locus, showed no significant difference in indel frequency between day-7 and day-21 in either HEL or SET-2 cells. Confirmation of *PCBP2*-KO at the protein level at early time point was not complete, however, is important to firstly validate the functional effect of the KO and secondly, to determine whether non-edited cells compensate for reduced *PCBP2* expression in co-culture.

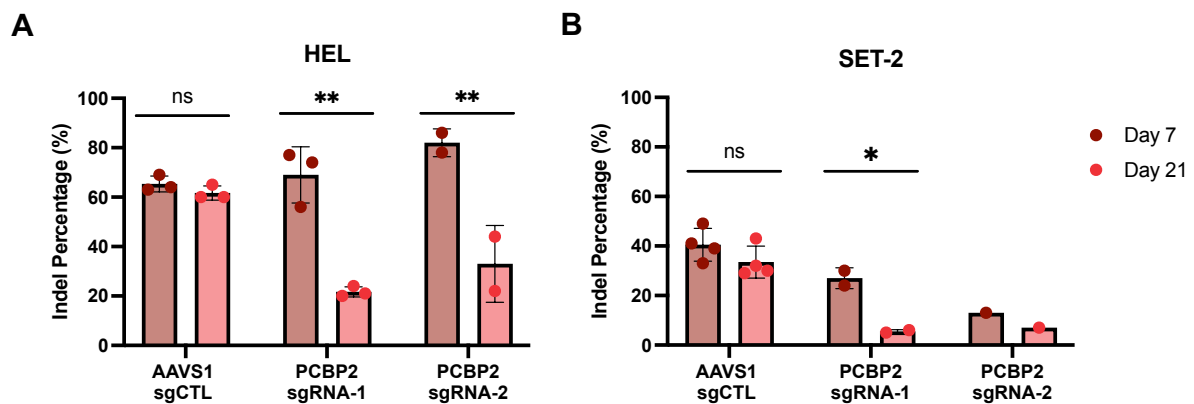


Figure 4.11. *PCBP2*-knockout cells are lost from culture overtime. *PCBP2*-edited (**A**) HEL and (**B**) SET-2 cells are lost from culture overtime as determined by indel frequency of edited sites at day-7 and day-21, compared to *AAVS1* control (sgCTL) edited cells. Data points represent mean +/- SD of independent experiments. Significance calculated via two-way unpaired Student's *t*-test where **p* < 0.05 and ***p* < 0.01. ns, non-significant.

4.3 Discussion

Having identified a cohort of RBP enriched in primary *TP53*-sAML LSC and required for model cell line survival, in this chapter I selected PCBP2 as a candidate RBP for further functional validation in cancer cell line models due to its potential for therapeutic targeting.

To be considered for therapeutic targeting, a candidate gene should meet the following criteria; (1) differentially expression between malignant and normal cell populations, (2) relatively low expression in other tissues, (3) present a selective requirement in the malignant cell population and (4) be a druggable target, or predicted as such. In addition, previous evidence a gene possesses oncogenic potential increases confidence during candidate gene selection. Therefore, to select the most promising candidate for functional validation studies, I comprehensively assessed the available literature and datasets.

As reported in the previous chapter of this thesis, *PCBP2* is significantly upregulated in *TP53*-mutant LSC compared to *TP53*-WT HSPC comparisons. To determine the impact of *PCBP2* depletion on *TP53*-WT haematopoietic cell populations, giving an indication of whether a targetable therapeutic index may exist, I first investigated the literature for relevant *PCBP2*-floxed mouse models. *PCBP2*-null mice are non-viable past mid-gestation with severe haematological defects responsible for their failure to survive²³⁶. In contrast, conditional knockout of *PCBP2* in the erythroid lineage has no impact on haematopoietic parameters in adult mice²³⁷. This suggests that while PCBP2 is essential for embryonic haematopoiesis, the requirement in adulthood is less pronounced. However, due to the erythroid-lineage specificity of the *EpoR* promoter, the requirement of PCBP2 in other adult haematopoietic lineages remains unknown. A *PCBP2^{fl/fl}/Vav-Cre* model, which inactivates genes-of-interest in adult haematopoietic and endothelial cells, would help to address the overall requirement of PCBP2 in adult haematopoiesis.

Comparing RBP requirements in a murine *ex vivo* genome-wide loss-of-function screen was an important reference for target selection. For example, *SFPQ* and *YARS1* sgRNA were dramatically lost over the screening period, while in contrast, *PCBP2* edited HSPC were not drastically depleted. This suggests the therapeutic index for targeting *YARS* or

SFPQ may be low and largely detrimental to WT-HSPC populations. In contrast, limited depletion of *PCBP2* sgRNA suggests *PCBP2* is not a required for WT-HSPC expansion and survival, increasing confidence in *PCBP2* as a therapeutic target in *TP53*-sAML.

Additional evidence suggesting *PCBP2* is an interesting candidate gene comes from *in silico* analysis of publicly available AML datasets. Without adopting a single-cell multi-omic approach, allowing gene expression signatures in *TP53*-mutant LSC and *TP53*-WT HSPC to be compared, the upregulation of *PCBP2* in *TP53*-mutant LSC would have unlikely been identified in bulk RNA-seq datasets. Thus, highlighting the importance of a single-cell multi-omics approach for identifying novel candidate genes for therapeutic targeting. Furthermore, analysis of the DepMap database showed *JAK2/TP53*-mutant AML cell lines, HEL and SET-2, have the highest *PCBP2* expression and greatest *PCBP2* dependency (**Figure 4.7**). Hence, *PCBP2* presents a promising selective molecular vulnerability in *TP53*-mutant transformed MPN.

To date, targeting RBP with conventional small molecule inhibitors has proven challenging, frequently due to their shared homologous domains and short degenerate sequences²⁵³. As such, the cellular phenotype of a small molecule inhibitor targeting the RBP, Musashi-2 (MSI-2), was discovered to be independent of MSI-2 inhibition, and instead via the inhibitors promiscuous binding to other RBP sharing the same RRM RBD²⁵⁴. While no drugs are commercially available for targeting *PCBP2*, predictions from the canSAR database – a comprehensive database for cancer drug discovery – suggest *PCBP2* has ligandable structures with the potential for targeting²⁵⁵. And aside from conventional small molecule inhibitors, new promising targeting mechanisms include decoy RNA oligonucleotides and RNA-PROTACs for the selective targeting of candidate RBP, such as *PCBP2*^{253,256}.

While I selected *PCBP2* as a molecular vulnerability for validation in this DPhil project other attractive candidates were present and worthy of future investigation. For example, *ILF3* was shown to be required for *TP53*-mutant model cell line survival via RBP-targeted CRISPR screening (**Figure 3.6**). *ILF3*-null mice are perinatal lethal with severe neuromuscular defects and blood counts at sacrifice from *ILF3*-null mice are suggestive of anaemia²⁵⁷. However, no conditional mouse models controlled by haematopoiesis-

specific promoters have been reported to assess haematopoietic phenotypes further. Interestingly, *ILF3* sgRNA were enriched in *ex vivo* murine HSPC screens, suggesting *ILF3*-KO provides a selective advantage to *TP53*-WT HSPC (**Figure 4.2**).

Mutually exclusive splicing of *ILF3* encodes NF90 and NF110. Both NF90 and NF110 bind double-stranded and single-stranded nucleic acids via double-stranded RNA-binding motifs and nucleic acid-binding arginine/glycine rich (RRG) domains, respectively. Through its dual RNA and DNA-binding capacities, multifunctional roles of NF90 and NF110 as regulators of gene transcription and transcript stability have been described. Using a cross-reactive NF90/NF110 antibody for ChIP-Seq studies, the association of NF90/NF110 with promoters of key genes for K562 cells erythroleukemia phenotype were identified²⁵⁸. Furthermore, knockdown of *ILF3* confirmed the NF90/NF110 transcription factor is responsible for driving expression of proliferation genes, *EGR1* and *c-Myc*, and repression of erythroid lineage determinant gene, *KLF2*²⁵⁸. This suggests NF90/NF110 regulates gene expression in a context-specific manner, depending on the promoter bound.

Roles of NF90/NF110 in PTGR have also been reported, such as the stabilisation of proangiogenic factors including *VEGF-A* and *CXCL-1* via binding of NF90/NF110 to the transcripts 3'UTR²⁵⁹. Thus, *ILF3* presents an attractive candidate gene due the increased likelihood of presenting a therapeutic index for targeting, given the selective advantage observed following *ILF3*-KO in murine *TP53*-WT HSPC. However, selection of *ILF3* for further investigation was hindered in this DPhil by limited resolution of NF90 and NF110 isoforms, their complex dual RNA and DNA-binding functions and lack of understanding on their physiological relevance in broad and haematopoiesis-related contexts. A similar approach was adopted to evaluate the potential of all candidate RBP for functional validation and ultimately, PCBP2 was selected based on its potential selectivity, reported roles in diverse malignancies and exciting uncharacterised requirement in AML.

In this chapter, the requirement of *PCBP2* in model cell lines was validated using complementary shRNA and CRISPR-Cas9 editing approaches. To conclusively show these results are a direct effect of *PCBP2* perturbation, future work should include rescue experiments where *PCBP2* expression is re-introduced and the *PCBP2*-KD/KO phenotype

subsequently reversed. In this project, the effects of *PCBP2* depletion were in line with those observed in solid cancer studies²²⁷⁻²³⁰, corroborating the oncogenic role of *PCBP2* in enhancing cancer cell proliferation and survival. In gastric cancer, *PCBP2* perturbation reduced cell proliferation via increased G₁ cell cycle arrest, where *PCBP2* directly stabilised key cell cycle progression gene, *cyclin-dependent kinase 2* (*CDK2*), via association with the transcripts 3'UTR²²⁹. *PCBP2* has also been reported to enhance cap-independent *c-Myc* translation via direct RNA-protein association, further highlighting the role of *PCBP2* in cell cycle regulation²⁶⁰. To confirm the reduced proliferation observed following *PCBP2*-KD in *JAK2/TP53*-mutant cell lines is also driven by cell cycle dysregulation, this project would benefit for cell cycle analysis following *PCBP2* depletion.

In summary, I have presented *PCBP2* as an attractive molecular vulnerability in *TP53*-sAML. Having determined a possible therapeutic index exists for selective targeting, through evaluation of cell line and primary data sets, I moved on to functionally validate the effect of *PCBP2* depletion on model cell line proliferation and survival. Results showed a significant loss of cell viability and reduced cell line proliferation following *PCBP2* perturbation, highlighting *PCBP2*'s oncogenic capabilities in *TP53*-mutant transformed MPN. However, transformed cell lines cannot replace or wholly reflect the effect of gene perturbation in primary samples or provide robust normal cell comparisons. Therefore, the following chapter details experimental efforts to functionally validate the requirement of *PCBP2* in primary *TP53*-mutant LSC and evaluate whether a differential requirement exists in healthy *TP53*-WT HSPC.

Gene	Canonical function(s)	RNA-binding domains	Protein-protein interactions	Mouse models	Cancer-related literature	Relevant functions in haematopoiesis	Enzymatic Function and Druggability
<i>HNRNP M</i>	Pre-mRNA splicing	RRM-domain (X3) and MR-repeat motif ²⁶⁸	Multimeric RBP complex, LASR (large assembly of splice regulators), also includes NF110 ²⁶⁶	<i>HnRNP m</i> -null model, preweaning lethality ²⁶⁵	hnRNP M required for transcriptome integrity, repressing cryptic splicing, dsRNA formation and tumours immune invasions ²⁶³	-	No enzymatic activity / Yes, known ligandable structures
<i>ILF3</i>	Transcriptional regulation via DNA-binding ²⁷⁰ . Transcript stability, translational regulation, miRNA/circRNA biogenesis ²⁶⁹ .	DFZ domain and DSRM domain (X2) ²⁶⁹	Heterodimer with ILF2 (NF45), dimerize through DZF domain. NF45 shows no RNA-binding ability in NF45/NF90 complex. Also, component of LASR spliceosome complex (with hnRNP M as above) ²⁶⁶	-	Upregulated NF90 in hepatocellular carcinoma promotes cyclin E1 mRNA stability driving cell proliferation and cancer progression ²⁶⁴	NF90/NF110 regulates transcription of key genes for erythroleukemia phenotype (<i>KLF1</i> , <i>MYC</i> , <i>ERG1</i>) via direct promoter binding ²⁶¹	No enzymatic activity. Ligandable structures in DSRM-domain
<i>PCBP2</i>	Post-transcriptional regulation	KH-domains (X3)	PCBP2 directly binds cGAS, preventing condensation necessary for cGAS-STING signalling ²⁶⁷ . PCBP2 acts as an adaptor between antiviral intermediary MAVS, and ubiquitin-ligase AIP4 ²²⁵	<i>PCBP2</i> -null, embryonic lethal with haematopoietic defects ²³⁶ <i>PCBP2^{fl/fl}/Epo^{R-Cre}</i> , viable with no haematopoietic defects in adulthood ²³⁸	Upregulated in various solid cancerous tissues where expression is associated with poor prognosis and advance disease stages ²²⁷⁻²³⁰	Regulates PTGR of key erythroid genes (<i>Runx1</i> , α -globin, β -globin) ^{237,241,242,262} . In blast-crisis CML, PCBP2 supports differentiation blockade via direct translational suppression of C/EBP α ^{248,249}	No enzymatic activity / Yes, known ligandable structures in KH domains

Table 4.1. Summary of candidate RNA-binding proteins. Part 1/4, continued on next page...

Gene	Canonical function(s)	RNA-binding domains	Protein-protein interactions	Mouse models	Cancer-related literature	Relevant functions in haematopoiesis	Enzymatic function/druggability
<i>PRPF40A</i>	Early factor in pre-mRNA splicing	WW-domain (X2), FF-domains (X6)	PRPF40A associates with the early U2 snRNP spliceosome complex via WW-domains ^{277,278}	<i>Prpf40a</i> -null, embryonic lethal. <i>Prpf40a</i> -heterozygotes show haematopoietic defects (reduced erythrocyte counts) ²⁶⁵	<i>PRPF40A</i> predicted a target of mutant/wild-type p53 binding ²⁷⁴	Regulates exon inclusion in transcripts required for myeloid differentiation ²⁷¹	No enzymatic activity / Yes, known ligandable domains
<i>RBBP6</i>	3' mRNA processing, E3-ubiquitin protein ligase	N-terminal DWNN domain, CHCC Zinc knuckle domain, and RING domains plus C-terminal RS domain, p53- and Rb-binding domain ²⁸²	Negative regulator of p53, acting as an adaptor for p53-/MDM2 interaction ²⁷⁶ . Interacts and ubiquitinates YBX1 for proteasomal degradation ²⁷⁹ . Interacts with Rb1 protein ²⁸⁰ .	<i>RBBP6</i> -null, embryonic lethal ²⁷⁶	Aberrant expression contributes to genomic instability through negative regulation of the mini-chromosome maintenance protein, MCM10 ²⁷⁵	<i>RBBP6</i> is a candidate gene in familial MPN susceptibility ²⁷²	E3-ubiquitin protein ligase / No known ligandable structures
<i>RPS17</i>	Component of ribosomal 40S subunit	DNA/RNA-binding 3-helical fold (similarities to human FF-domain)	Associates with other proteins of the 40S ribosomal subunit ²⁸¹	-	-	Mutations in <i>RPS17</i> cause Diamond-Blackfan Anaemia ribosomopathies ²⁷³	No enzyme activity / Non-specific to RPS17, Ataluren associates with ribosomes and facilitates readthrough past nonsense mutation sites

Table 4.1. Summary of candidate RNA-binding proteins. Part 2/4, continued on next page...

Gene	Canonical function(s)	RNA-binding domains	Protein-protein Interactions	Mouse models	Cancer-related literature	Relevant functions in haematopoietic context	Enzymatic function/Druggability
<i>RPS28</i>	Component of ribosomal 40S subunit	C4-ZFD	Associates with other ribosomal proteins of the 40S ribosomal subunit ²⁸¹	-	RPS28 controls MHC class I peptide synthesis by modulating non-canonical translation, exhibiting roles in cancer immunosurveillance ²⁸⁹	Mutations in <i>RPS28</i> cause Diamond-Blackfan Anaemia ribosomopathies ²⁸⁶	No enzyme activity / Non-specific to RPS28, Ataluren associates with ribosomes and facilitates readthrough past nonsense mutation sites
<i>SFPQ</i>	Early factor in pre-mRNA splicing, transcriptional regulation, DNA repair and a core paraspeckle protein	DBM, RRM domain (X2), NOPS domain, coiled-coil domain ²⁹⁴	Obligatory dimerization with family members (NONO and PSPC1) ²⁹⁴ . SFPQ binds RAD51D to regulate HDR and maintain genome integrity ²⁹⁵	Embryonic lethal with neurodegenerative defects ²⁶⁵	SFPQ regulates SRSF2-mediated response to platinum-based chemotherapy in ovarian cancer to modulate alternative splicing of caspase-9 ²⁹⁰	<i>SFPQ-ABLI</i> fusion gene drives B-cell ALL ²⁸⁷	No known enzymatic activity / N-3 and C-65 are small-molecule inhibitors against SFPQ ²⁸³
<i>SRSF10</i>	Splice factor regulating pre-mRNA processing	RRM-motif and RS domain (x2)	SRSF10 can directly bind YTHDC1 to regulate mRNA splicing ²⁹⁶	<i>SRSF10</i> -null mice die midgestation with severe cardiac defects ²⁹³	SRSF10 mediates splicing of pro-tumorigenic <i>BCLAF1-L</i> splice variant and expression associates with adverse tumour grades in colorectal cancer ²⁹¹ . Regulates splicing of oncogenic downstream effectors controlling NMD in hepatocellular carcinoma ²⁹²	<i>SRSF10</i> significantly upregulated in AML. Depletion drives cell cycle arrest and enhanced apoptosis ²⁸⁸	No enzymatic activity / 1C8 and GPS167 inhibit CLK kinases responsible for SRSF10 phosphorylation and subsequent activity ^{284,285}

Table 4.1. Summary of candidate RNA-binding proteins. Part 3/4, continued on next page...

Gene	Canonical function(s)	RNA-binding domains	Protein-protein Interactions	Mouse models	Cancer-related literature	Relevant functions in haematopoietic context	Enzymatic function/ Druggability
SRSF11	Roles in pre-mRNA processing (incompletely defined)	RRM-motif and RS domain	SRSF11 associates with PUF60 to facilitate splicing ³⁰⁶	<i>SRSF11</i> -null mice are embryonic lethal ²⁶⁵	SRSF11 is responsible for the recruitment of telomerase to nuclear speckles ³⁰⁰	<i>SRSF11</i> is upregulated in AML and associated with poor prognosis ²⁸⁸ . Inactivation of <i>SRSF11</i> sensitises AML cells to venetoclax treatment ²⁹⁸	No enzymatic activity/ known ligandable structures
YARS1	Catalyses attachment of tyrosine to tRNA. Proteolytic cleavage of full-length TyrRS results in two distinct cytokine mimics; 1) a N-terminal IL8-like derivative and 2) a C-terminal EMAPII-like derivative	tRNA synthetase class I domain and tRNA-binding domain	Homodimerizes for tRNA aminoacylation via N-terminal domain ³⁰⁷	<i>YARS1</i> -null model is embryonic lethal while <i>YARS1</i> -heterozygotes develop to adulthood with no haematopoietic defects ²⁶⁵	<i>YARS1</i> contributes to gastric cancer progression in a PI3K-Akt dependent manner ³⁰¹	-	tRNA synthetase enzyme with known ligandable structures
YBX1	Regulator of transcription and translation and pre-mRNA splicing	CSD- domain	RBBP6 binds and ubiquitinates YBX1 for proteasomal degradation ²⁷⁹ . YBX1 binds a selection of DNA-repair proteins (MSH2, Ku80, WRN) ³⁰⁸ . Associated with IGF2BP1 to mediate stability of <i>c-myc</i> transcripts ³⁰⁹	<i>YBX1</i> -null mice are perinatal lethal with severe growth retardation ³⁰⁴ . <i>YBX1</i> ^{fl/fl} <i>Mx1-Cre</i> mice show no obvious haematopoietic defects ³⁰⁵ . <i>YBX1</i> inactivation in a retroviral model of leukaemia transformation delays leukemia onset ²⁹⁹ .	Well characterised oncoprotein across cancer types ³⁰² . YBX1 is a transcriptional repressor of p53 ³⁰³ .	YBX1 mediates <i>JAK2</i> -mutant MPN persistence ¹⁸⁹ and AML maintenance ²⁹⁹ via different post-transcriptional gene regulation mechanisms	No enzymatic activity / Druggable (SU056) ^{213,297}

Table 4.1. Summary of candidate RNA-binding proteins. Part 4/4.

Chapter 5 Establishing functional assays to assess the role of PCBP2 in *TP53*-sAML

5.1 Introduction

5.1.1 Unknown functional role of PCBP2 in primary cells

Thus far, I have identified the requirement of PCBP2 in cancer cell line models through genome engineering techniques, namely shRNA and CRISPR-Cas9 approaches, coupled to downstream functional assays. To validate the role of PCBP2 in primary *TP53*-sAML presents additional challenges due to primary AML cells being notoriously difficult to grow *ex vivo* and a lack of available PCBP2-targeting inhibitors. Without available inhibitors, genome engineering techniques are required to perturb PCBP2 expression *in vitro*, extending experimental timelines and exposing fragile cell populations to additional stress. At present, the literature provides no functional evidence regarding the effects of PCBP2 perturbation in a primary human haematological context. The next steps in this project were to therefore establish assays that enable functional assessments of PCBP2's role in primary *TP53*-sAML.

5.1.2 Available models to study the primary cell requirement of PCBP2

Traditional methods for functional analysis of primary progenitor populations *in vitro* include short-term liquid culture assays, to investigate bulk HSPC proliferation and differentiation, and colony forming unit (CFU) assays. CFU assays are widely used to assess the functional capacity of individual haematopoietic progenitor cells to proliferate and differentiate³¹⁰. In addition, the self-renewal capacity of HSC populations can be determined by long-term culture initiating cell (LTC-IC) assays. Where an LTC-IC assay measures primitive HSC based on their capacity to produce myeloid progeny for up to 5-weeks with the support of a stromal feeder layer³¹¹.

Despite their frequent use, two-dimensional (2D) culture systems are highly artificial, leading to potentially irrelevant stem and progenitor cell phenotypes *in vitro*³¹². Furthermore, blood cancer cells can be notoriously difficult to recover following cryopreservation and 2D liquid culture systems lack the niche complexity to support their survival and proliferation *ex vivo*^{137,312}. PDX models can provide an *in vivo* environment for cell engraftment; however, malignant haematopoietic cell populations do still fail to engraft^{313–315}. And although humanised mouse models for AML cell engraftment are improving³¹³, for preliminary and high-throughput target validation studies, the utilisation of mouse models goes against ethical scientific efforts to reduce and replace experimental animal models. Hence, models which contain the relevant cell types to imitate the architecture and biological interactions of the bone marrow niche are important for disease modelling and functional analysis.

Haematopoietic stem cells reside in the highly specialised bone marrow niche, a complex microenvironment comprising both haematopoietic and non-haematopoietic mesenchymal stromal cell (MSC) and endothelial cell populations (**Figure 5.1**). Local interactions and cross-talk between cell populations within the bone marrow niche are crucial for stem cell function and haematopoietic homeostasis³¹⁶. Alterations to the bone marrow niche can contribute to the initiation and propagation of haematological malignancies, and similar to healthy HSC, LSC are dependent on interactions within the bone marrow microenvironment for their survival³¹⁶. It is therefore useful to consider the impact of niche factors for *in vitro* disease modelling.

Despite the well-acknowledged contribution of the niche to haematological malignancies, whether its main role is in disease initiation or progression remains unresolved³¹⁶. This is because mutations in non-haematopoietic cells or alterations to the niche can drive disease initiation, while transformed haematopoietic cells have the potential to remodel the niche and contribute to disease progression. Evidence of the niche as a predisposing factor for malignancy comes from allogeneic-HSCT, where malignant clones have been found derived from healthy donor haematopoietic cells, highlighting the selective pressures exerted by the recipient's malignant niche³¹⁷. Furthermore, non-conditional deletion of *Rb* in murine models leads to myeloproliferative disease, but conditional *Rb* deletion in haematopoietic populations does not lead to disease manifestation,

suggesting mutations with non-haematopoietic niche components are required for disease manifestation³¹⁸. Another important niche alteration contributing to disease initiation is inflammation. Given myeloid malignancies are most prevalent in elderly population, where inflammation is a hallmark of ageing, the proinflammatory niche has long been considered a predisposing factor for myeloid malignancies³¹⁶.

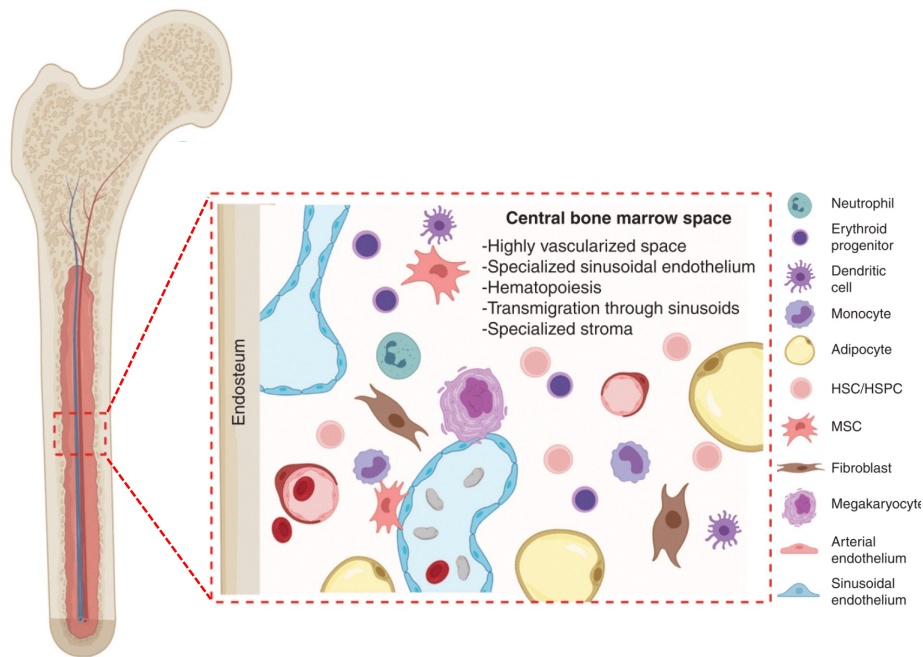


Figure 5.1. The bone marrow microenvironment. HSC, and their progenitors (HSPC), reside in the central bone marrow in concert with mesenchymal stromal cells (MSC), endothelial cells and myeloid and lymphoid subsets. Stem cell maintenance within the niche requires local interactions with MSC and endothelial subsets, providing HSC-supporting factors and maintenance of the bone marrow via vascular perfusion facilitating angiocrine signaling, nutrients supply and waste excretion. Illustration adapted from Khan *et al*, 2023¹³⁷.

Malignant haematopoietic cells can also transform the niche in support of their own survival. In MPN, both haematopoietic and non-haematopoietic constituents of the niche aberrantly secrete pro-inflammatory cytokines contributing to MPN pathogenesis³¹⁹. Furthermore, a single-cell atlas of the myelofibrotic murine bone marrow highlights previously unappreciated pathogenic-niche interactions contributing to myelofibrotic disease²³⁹. These included transcriptional reprogramming of MSC populations and a relative increase in inflammatory fibroblasts, reducing expression of haematopoietic niche factors and driving upregulation of proinflammatory signalling pathways,

respectively²³⁹. Thus, it is highly likely the niche has the potential to initiate disease, and be remodelled by malignant haematopoietic cells to support and accelerate disease progression. These events likely occur in a non-mutually exclusive manner, and overall highlight the importance of niche factors in the modelling of haematopoietic malignancies *in vitro*.

Early improvements to 2D culture systems include co-culturing HSPC populations with primary bone marrow-derived MSC. While improving proliferation, 2D co-culture systems fail to maintain HSC self-renewal capacity³¹². This is likely due to this simple co-culture approach failing to capture the complexity of the bone marrow niche. Improvements to HSPC-MNC co-culture systems were made with structural advancements, from 2D systems to three-dimensional (3D) systems. Natural or synthetic polymers, such as fibrin and hydrogels, can form porous extracellular matrixes which provide structural support, increased the area available for cell growth and facilitate cell-cell interactions^{312,320,321}. Co-culturing haematopoietic subsets and MSC within these 3D matrices increases the complexity of artificial stem cell niches, further fulfilling niche requirements^{312,320}.

To further advance the complexity and mimicry of the bone marrow niche *ex vivo*, systems require vascularisation and perfusion. The highly vascularised nature of the bone marrow niche is important for facilitating angiocrine signalling, supplying nutrients and extracting waste³²². Attempts to vascularise 3D models *in vitro* include the incorporation of laminar flow microchannels to create “biochips”³²². The biochip created by Chou *et al*³²³ comprises a top fibrin gel-filled channel where HSPC and MSC are co-cultured, and a bottom channel lined with human vascular endothelium. Perfusion of cytokine-supportive media moves exclusively through the vascular channel into the haematopoietic compartment via a porous membrane, recapitulating the vascular architecture of the bone marrow where nutrients and waste transit the niche through a network of bone marrow vessels³²³.

Hence, although significant improvements have been made in modelling the bone marrow niche, the systems described still present with limitations. For example, models rarely represent all bone marrow components, lack vascularisation and show limited

specialisation of stromal populations^{137,322}. In addition, many systems are not scalable and the use of human bone marrow-derived MSC in co-culture systems creates reproducibly issues due to finite sample availability and inter-donor variability¹³⁷.

Organoids are self-organising 3D structures which reflect the complex structural and functional characteristics of the corresponding tissue *in vivo*³²⁴. Organoids can be derived from induced pluripotent stem cell (iPSC)-lines or tissue-derived cells from patient biopsies and murine models. Compared to conventional 2D liquid cultures and mouse models, organoids provide a platform for increasing patient specificity and as such, have been utilised for disease modelling, drug discovery and personalised medicine³²⁴. In 2009, Sato *et al*³²⁵ reported the first organoid model, illustrating the potential of intestinal stem cells to form crypt-like structures *in vitro*. Since then, human heart-forming, pancreatic and liver organoids have been developed, to name a few³²⁴. And although organoids are proving transformative in other fields, relatively limited progress had been made in deriving human bone marrow organoids until recently.

The first iPSC-derived human bone marrow organoids have recently been reported¹³⁷. Derivation from iPSC-lines facilitates generation of mesenchymal, vascular and myelopoietic bone marrow components, resembling the cellular composition and spatial architecture of the bone marrow niche¹³⁷. A detailed protocol paper has also been released, enabling easy dispersion of this highly valuable resource to the haematology research community¹³⁸. In quick succession, a second iPSC-derived human bone marrow organoid model was also published³²⁶, considered to more closely resemble intraembryonic haematopoiesis. Khan *et al*¹³⁷ have shown human bone marrow organoids are capable of supporting the engraftment of adult donor-derived cells, including cell populations from myeloid and lymphoid malignancies which are notoriously difficult to sustain *ex vivo*, facilitating disease modelling¹³⁷. Thus, organoids present major advancements in modelling of the bone marrow niche and provide a platform to study complex haematopoietic malignancies and their therapeutic targeting *ex vivo*.

5.1.3 Hypothesis

Using suitable models for the investigation of candidate gene perturbation in primary cells *ex vivo*, I hypothesise a therapeutic index for targeting *PCBP2* between primary *TP53*-WT HSPC and *TP53*-mutant LSC can be identified.

5.1.4 Summary and aims

Cancer cell line models alone cannot justify selection of candidate genes as promising therapeutic targets. Results must be translatable to primary patient samples and a therapeutic index must exist between malignant and health cell populations. To validate the requirement of *PCBP2* in primary patient samples and investigate the selective requirement for *PCBP2* in *TP53*-mutant LSC versus *TP53*-WT HSPC, the aims for this Chapter are as follows:

1. Optimise cellular and molecular techniques to investigate the effects of *PCBP2* depletion in primary patient samples
2. Investigate the survival and proliferation of primary *TP53*-sAML cells in a human bone marrow organoid model
3. Determine the effect of *PCBP2* perturbation in primary *TP53*-WT HSPC and *TP53*-mutant LSC

5.2 Results

5.2.1 PCBP2 is dispensable for *TP53*-WT HSPC proliferation and differentiation

To investigate the effect of *PCBP2* depletion in *TP53*-WT HSPC, I used the same shRNA lentiviral constructs used to perturb *PCBP2* expression in cell line models and developed an optimised protocol for primary lentiviral transduction (described in Section 2.2.3). Briefly, two independent pooled-donor cord blood (CB) samples, enriched for CD34+ HSPC, were thawed and left to recover overnight in cytokine-rich media including high concentrations of SCF, FLT3-L and TPO. The following day, cells were transduced with scrambled shRNA control or either *PCBP2* KD #1 or *PCBP2* KD #2 shRNA. To aid transduction efficiency of primary cells, an optimised “double transduction” approach was taken where virus was bound to the RetroNectin-coated plate by centrifugation and additional virus added to the cell suspension for repeated centrifugation at an MOI 100. Seventy-two hours post-transduction, 500 viable CD34+RFP+ HSPC were sorted per technical replicate by flow cytometry for multilineage differentiation and CFU assays, as described in Section 2.2.5 and 2.2.6, respectively. An overview of the optimised workflow can be found in **Figure 5.2**.

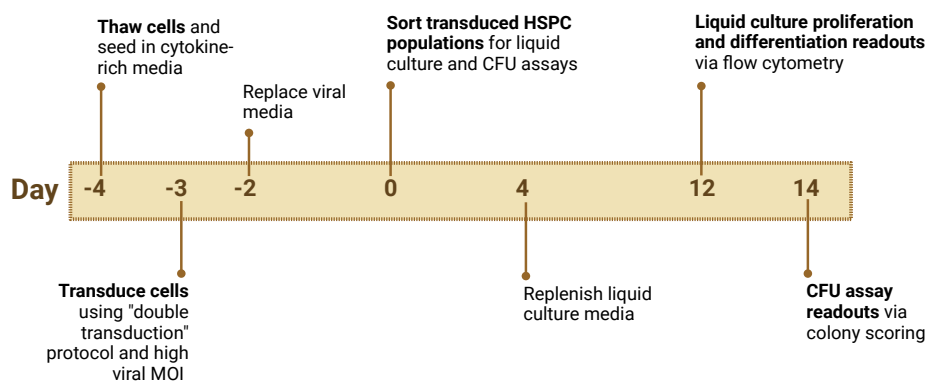


Figure 5.2. Workflow for primary HSPC transduction.

Illustration created with BioRender.com

Lentiviral transduction of CB CD34+ HSPC was highly effective with over 65% of all cells across samples expressing the lentiviral RFP reporter (**Figure 5.3A-B**). Upon sorting for

functional assays, *PCBP2*-KD was determined at the mRNA level in a homogenous RFP+ population, showing a 53% and 61% reduction in *PCBP2* expression using *PCBP2* KD #1 and *PCBP2* KD#2 shRNA constructs, respectively (**Figure 5.3C**). Interestingly, despite high transduction efficiencies, *PCBP2*-KD failed to reflect the nearly complete KD observed using identical shRNA constructs in cell line models. Bright RFP expression determined by flow cytometry suggests multiple shRNA copies integrated per cell, suggesting reduced KD observed in primary CB samples is not related to the transduction protocol.

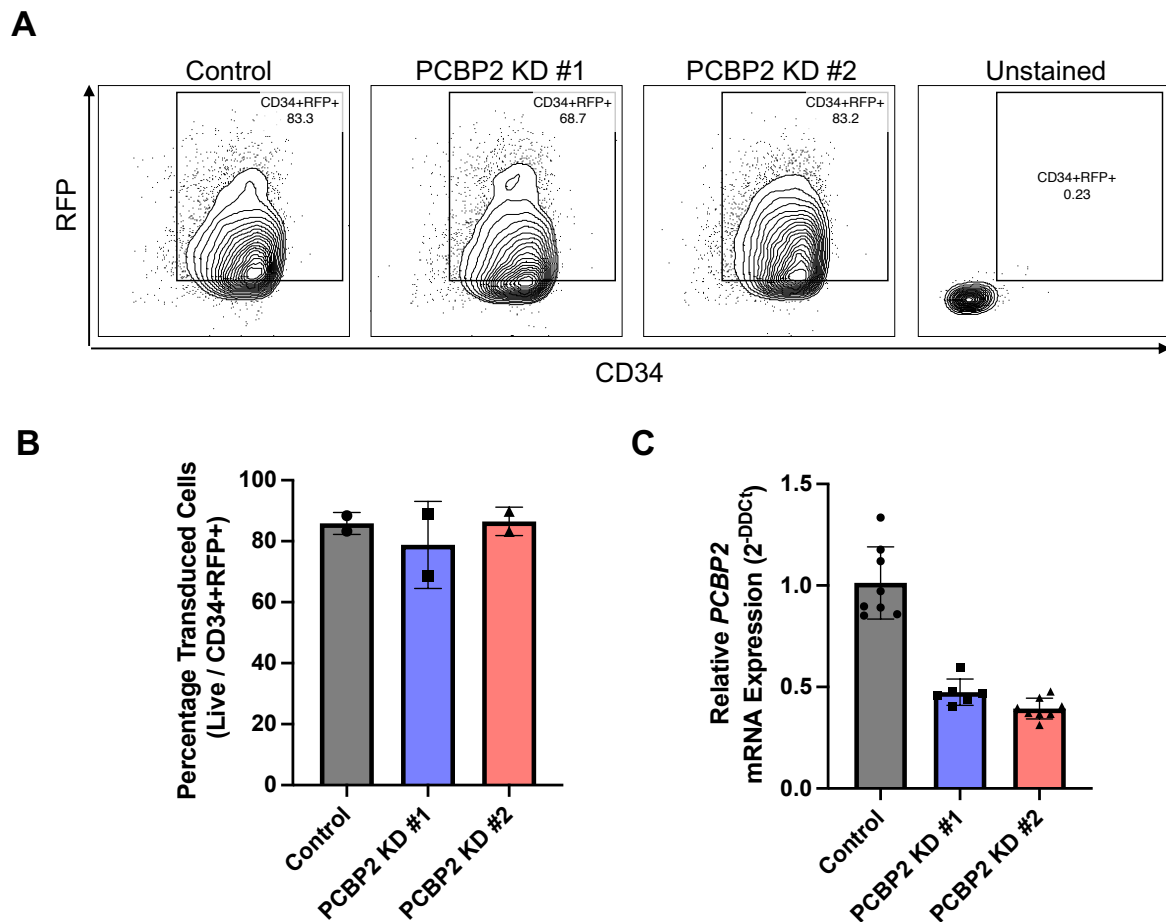


Figure 5.3. *PCBP2* depletion in *TP53*-WT HSPC via lentiviral shRNA. (A) Representative flow gating of CD34+RFP+ populations per shRNA condition. Gates were set on unstained sample plus appropriate single stains and fluorescence minus one controls. Parental population for flow cytometry gating was 7AAD- viable cells. **(B)** Percentage of viable CD34+ cord blood (CB) cells transduced with either scrambled shRNA control, *PCBP2* KD #1 or *PCBP2* KD #2 shRNA. Transduced cells express lentiviral RFP reporter. Data points represent mean percentage $\pm$ SD of transduced cells per condition from two independent experiments. **(C)** Relative *PCBP2* expression following lentiviral transduction with *PCBP2* KD #1 or KD #2 shRNA compared to scrambled shRNA control as determined by RT-qPCR. Data points represent RT-qPCR technical replicates from two independent experiments.

PCBP2 KD #2, although interpatient variability in proliferation potential was observed (**Figure 5.4A**).

Liquid culture multilineage differentiation readouts were conducted via flow cytometry for appropriate lineage markers. Analysis of CD34+ HSPC populations and key terminal haematopoietic lineages including megakaryocytic, myeloid and erythroid lineages showed no difference in differentiation output between scrambled shRNA control and PCBP2 KD #1 or PCBP2 KD #2 shRNA after 12-days in culture (**Figure 5.4B**).

CFU assays to assess the clonogenic potential of haematopoietic progenitor populations were concordant with liquid culture results, showing no significant difference in total colonies formed between KD and control constructs (**Figure 5.4C**). In addition, there was no statistical difference between colony types when targeted with either scrambled shRNA control or *PCBP2*-KD constructs (unpaired two-way Student's *t*-test). Overall, this data suggests a ~50% reduction in *PCBP2* expression is compatible with continued *TP53*-WT HSPC proliferation and differentiation *in vitro*.

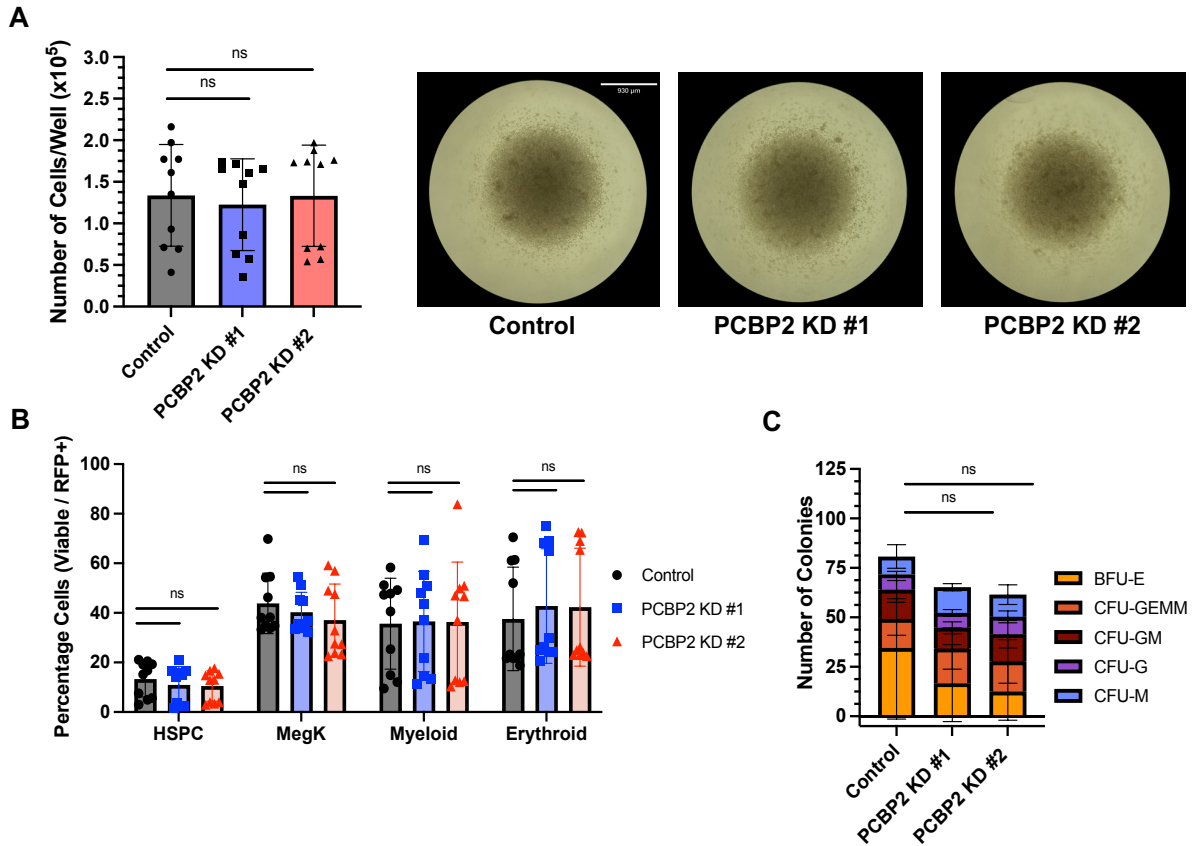


Figure 5.4. Effect of *PCBP2* depletion on cord blood HSPC. (A). Cell counts following *PCBP2* depletion in liquid culture via *PCBP2* KD #1 or *PCBP2* KD# 2 shRNA compared to scrambled control (left) Representative culture images per shRNA construct at day-12 (right). **(B)** Multi-lineage differentiation readouts via flow cytometry following 12-days in liquid culture. For (A) and (B) data points represent values per technical replicate from two independent experiments and mean $\pm$ SD where significance was determined via unpaired two-way Student's *t*-test. HSPC, Haematopoietic stem and progenitor cell (CD34); MegK, Megakaryocyte (CD41a, CD42b); Myeloid (CD11b, CD16); Erythroid (CD71). **(C)** Effect of *PCBP2*-KD on CB clonogenicity via 14-day CFU (colony-forming unit) assay. Data represented mean $\pm$ SD for two independent experiments and significance testing comparing total counts per construct determined via unpaired two-way Student's *t*-test. BFU-E, burst-forming unit erythroid; CFU-GEMM, granulocyte/erythroid/macrophage/megakaryocyte; CFU-GM, granulocyte/macrophage; CFU-G, granulocyte; CFU-M, macrophage.

To validate the outcome of *PCBP2* depletion in *TP53*-WT HSPC, I adopted a CRISPR-Cas9 editing approach coupled to single-cell colony genotyping. This single-cell colony approach was an attempt to overcome the limited capabilities of shRNA lentiviral vectors to completely deplete *PCBP2* expression in primary cells. Edited cells can either be heterozygous or homozygous for an indel at the target loci, therefore, this experiment investigated the potential of single *PCBP2*-heterozygous and -homozygous HSPC, WT for

TP53, to form colonies. Briefly, two CD34-enriched HD apheresis samples were thawed and left to recover overnight in cytokine-rich media as previously described. The following day, cells were edited via RNP nucleofection with either dual *PCBP2*-sgRNA or *AAVS1*-sgRNA. Twenty-four hours after editing, CD34+ HSPC were single-cell sorted into Terasaki plates with colonies maintained and scored over a period of 13-days. On day-13, colonies were collected and lysed for genotyping. A detailed methodology for primary cell editing and single-cell colony genotyping can be found described in Section 2.2.7 and Section 2.2.8, respectively.

The percentage of colonies present on day-13 relative to the number of single cells initially seeded per condition into Terasaki plates was low (**Figure 5.5A**). However, there was no significant difference in the percentage of colonies formed between *PCBP2* or *AAVS1* targeted colonies compared to Cas9-only controls. This suggests that indel mutations, independent of their target loci, are not detrimental to colony growth when compared to Cas9-only control colonies which lack sgRNA inclusion and therefore indel formation.

For each condition, a proportion of colonies were genotyped (*PCBP2* n=42; *AAVS1* n=17) and assigned with either wildtype (WT), heterozygous (HET) or homozygous (HOM) status at *PCBP2* and *AAVS1* loci. *AAVS1* editing frequencies give an indication of what can be expected when gene KO has no effect on cellular phenotype. Results showed minimal differences in the percentage of each genotype classification between *PCBP2* and *AAVS1* edited colonies (**Figure 5.5B**).

Furthermore, a Chi-Square test comparing the expected and observed frequencies of WT colonies between *PCBP2* and *AAVS1* colonies failed to reject the null hypothesis (WT vs mutant (HET+HOM) colonies, $X^2 = 0.64$, degrees of freedom = 1, p -value < 0.9) confirming there is no significant difference in the frequency of *PCBP2* or *AAVS1* WT colonies formed. These results suggest WT colonies do not dominate *PCBP2* genotyped colonies as you would expect if *PCBP2* was required for colony formation. The Chi-Square test was repeated to compare WT and homozygous KO clones only. This Chi-Square test was reflective of the combined mutant variable (HOM+HET), showing no difference in the frequency of WT clones in *PCBP2* or *AAVS1*-edited colonies (WT vs HOM colonies, $X^2 =$

0.998, degrees of freedom = 1, p -value < 0.9). Overall, these findings corroborate primary shRNA-KD data suggesting *PCBP2* is not required for *TP53*-WT HSPC growth and proliferation. Furthermore, results elaborate on previous findings, showing formation of *PCBP2* homozygous-KO colonies is feasible and *PCBP2*-WT colonies are present at a frequency comparable to *AAVS1*-WT clones.

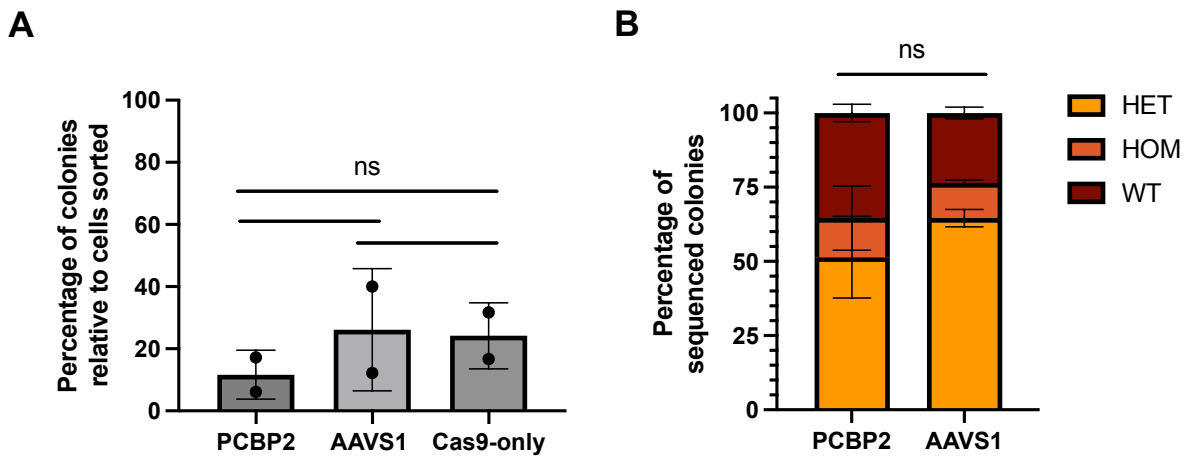


Figure 5.5. Effect of *PCBP2*-KO on HD HSPC single-cell colony formation. (A) Percentage of colonies formed on day 13 relative to the percentage of single cells sorted per construct. Data points represent percentage of colonies from two independent donors with mean $\pm$ SD. Significance calculated using unpaired two-way Student's t -test. **(B)** Percentage of sequenced colonies by genotype at target loci (*PCBP2* $n=42$; *AAVS1* $n=17$). Data represented mean $\pm$ SD for two independent donors. Significance calculated via Chi-Square test comparing WT vs mutant (HET+HOM) colonies ($X^2 = 0.64$, degrees of freedom = 1, p -value < 0.9). HET, heterozygous; HOM, homozygous; WT, wildtype

5.2.2 Conventional liquid culture conditions are insufficient to assess *PCBP2* depletion in primary *TP53*-sAML

No inhibitors are currently available to target *PCBP2* and therefore genome engineering techniques are required to perturb *PCBP2* expression in primary cells. These approaches can be cytotoxic as cells are subject to additional extrinsic stressors. In addition, LSC from primary AML samples are characteristically difficult to maintain *ex vivo* due to their rapid apoptosis, differentiation and loss of clonogenicity³²⁷. During this project I tested multiple techniques to perturb *PCBP2* in primary samples and found many to be insufficient for assessing the effect of gene perturbation of primary cell function. Below, the techniques

trialled to optimise primary cell gene perturbation and their pitfalls in sustaining primary cell survival and proliferation, specifically LSC from *TP53*-sAML patients, are described.

An all-in-one Cas9-PCBP2-sgRNA lentiviral vector is unsuitable for the transduction of primary cells

Primary cell transductions were first attempted using an all-in-one CRISPR-Cas9 vector, including Cas9, *PCBP2*-targeting sgRNA and GFP reporter. This Cas9-PCBP2-sgRNA plasmid had previously been used for RBP-targeted cell line screening (Chapter 3). This therefore seemed a good starting point for testing primary cell transduction and gene editing techniques for *PCBP2* depletion.

Given the limited availability of *TP53*-sAML samples for this project, this lentiviral gene editing approach was first attempted in HD HSPC. Briefly, cells were thawed, CD34-selected via column-enrichment and transduced. Unlike optimised transduction protocols described above in Section 5.2.1 for primary HD HSPC, here a MOI 50 was used and a “single transduction” approach where virus was only added to cell suspensions. Results showed Cas9-PCBP2-sgRNA vectors failed to transduce CD34+ HSPC populations (**Figure 5.6**). In comparison, a small GFP barcoding virus (5Kb) shared by Dr Ryan Beveridge at the WIMM Virus Screening Facility, was capable of transducing HSPC populations with a transduction efficiency of 17.3%.

While this experiment did not adopt the optimised “double transduction” protocol presented in Section 5.2.1, which may have improved Cas9-PCBP2-sgRNA transduction efficiency, the small barcoding control construct exemplifies the inherent difficulty of sizeable vectors to transduce primary cell populations. Hence, even if transduction using Cas9-PCBP2-sgRNA could be achieved in HD HSPC, improvements are unlikely to directly translate to LSC which are increasingly difficult to transduce. Therefore, I concluded an alternative technique with increased likelihood for success in both HD HSPC and *TP53*-mutant LSC was required.

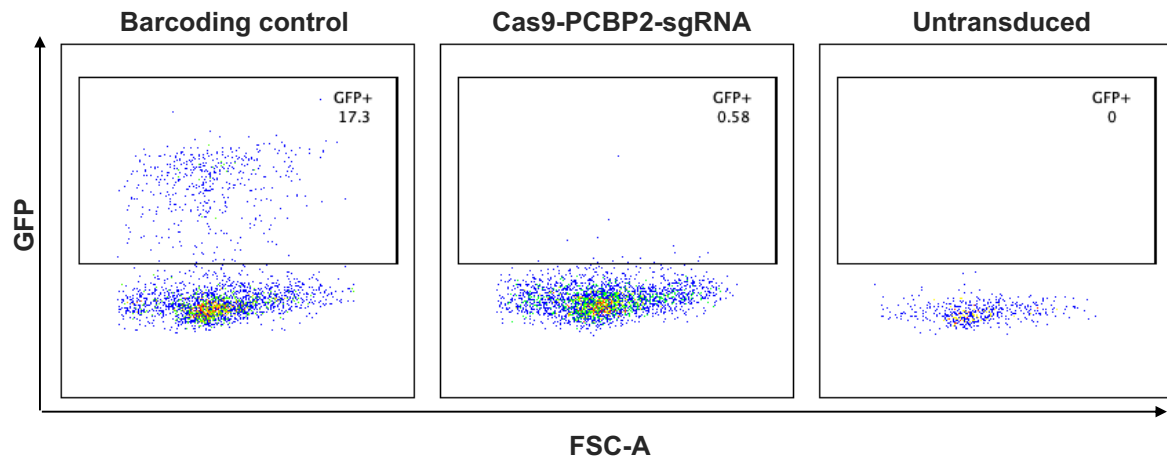


Figure 5.6. Transducing HD CD34+ HSPC with all-in-one Cas9-PCBP2-sgRNA lentiviral vector. Values represent percentage GFP-positive cells from parent DAPI-/CD34+ cell population. Gates set based on untransduced control. Barcoding control, ~5Kb; Cas9-PCBP2-sgRNA, 13.7Kb.

Primary AML transduction via lentiviral shRNA is limited by poor LSC survival in liquid culture and methylcellulose-based assays

Having shown primary cell transduction is possible with smaller sized vectors, I adopted an shRNA approach where constructs for lentiviral packaging are notably smaller than the previously described Cas9-PCBP2-sgRNA vector (9.4Kb vs 13.7Kb). Utilising a smaller lentiviral construct and optimising the transduction protocol markedly improved transduction efficiencies as previously shown in *TP53*-WT HSPC (**Figure 5.3**) and here in LSC from *TP53*-sAML patients (**Figure 5.7A**). Furthermore, cells were highly viable 3-days post-transduction during FACS for multilineage differentiation cultures and methylcellulose-based CFU assays (**Figure 5.7B**). Knockdown of *PCBP2* in *TP53*-sAML cells also reflected the KD in HD HSPC, with ~50% *PCBP2* depletion determined via RT-qPCR (**Figure 5.7C**).

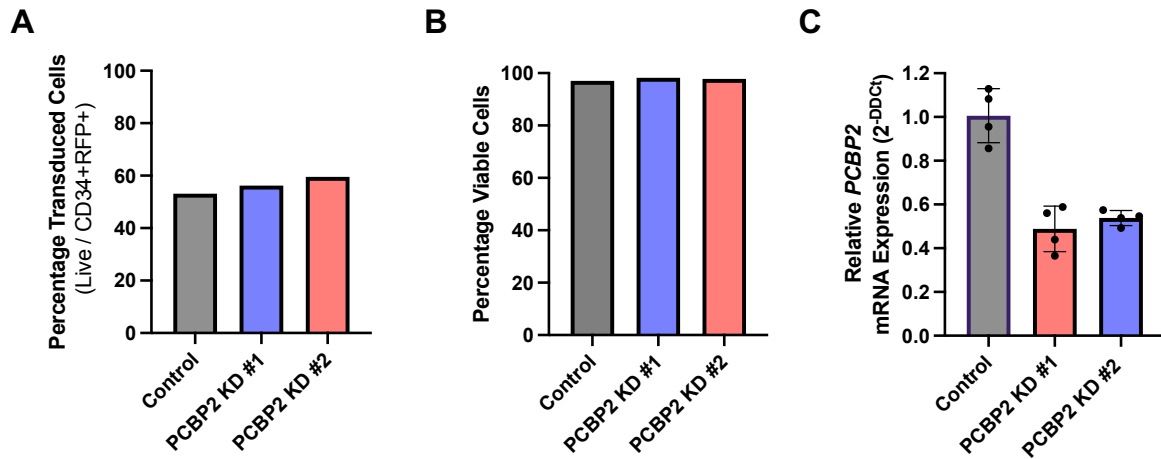


Figure 5.7. *PCBP2* depletion in *TP53*-sAML samples via lentiviral shRNA. (A) Percentage of transduced HSPC (viable, CD34+) with either scrambled shRNA control, shRNA *PCBP2* KD #1 or *PCBP2* KD #2. Transduced cells express lentiviral RFP reporter. Data represents values from one independent experiment. **(B)** Percentage total viable cells 3-days post-transduction with *PCBP2* KD #1/#2 shRNA or scrambled shRNA control constructs. Over 95% of viable cells per condition were CD34-positive. **(C)** Relative *PCBP2* expression following lentiviral transduction with *PCBP2* KD #1 or KD #2 shRNA compared to scrambled shRNA control as determined by RT-qPCR. Knockdown assessed in homogenous RFP+ populations enriched by FACS. Data points represent mean $\pm$ SD for RT-qPCR technical replicates from one independent experiment.

In contrast to the successful multilineage differentiation and CFU readouts from transduced HD HSPC (**Figure 5.4**), *TP53*-sAML samples failed to survive and proliferate in liquid culture or produce colonies in methylcellulose during CFU assays, independent of the shRNA construct present. Minimal cells (<5%) were acquired during flow cytometry for differentiation analysis, independent of condition, showing cells were unable to survive and expand in liquid culture conditions (**Figure 5.8A**). Similarly, 500 cells were plated per replicate for both HD HSPC and *TP53*-sAML CFU assays. Compared to HD HSPC which produced upwards of 50 colonies per replicate, *TP53*-sAML samples yielded minimal colonies (**Figure 5.8B**). Colonies which did form were also comparably smaller than colonies formed from HD HSPC and lacked mature morphologies. Any interpretation of CFU assays from *TP53*-sAML cells therefore would have been inaccurate.

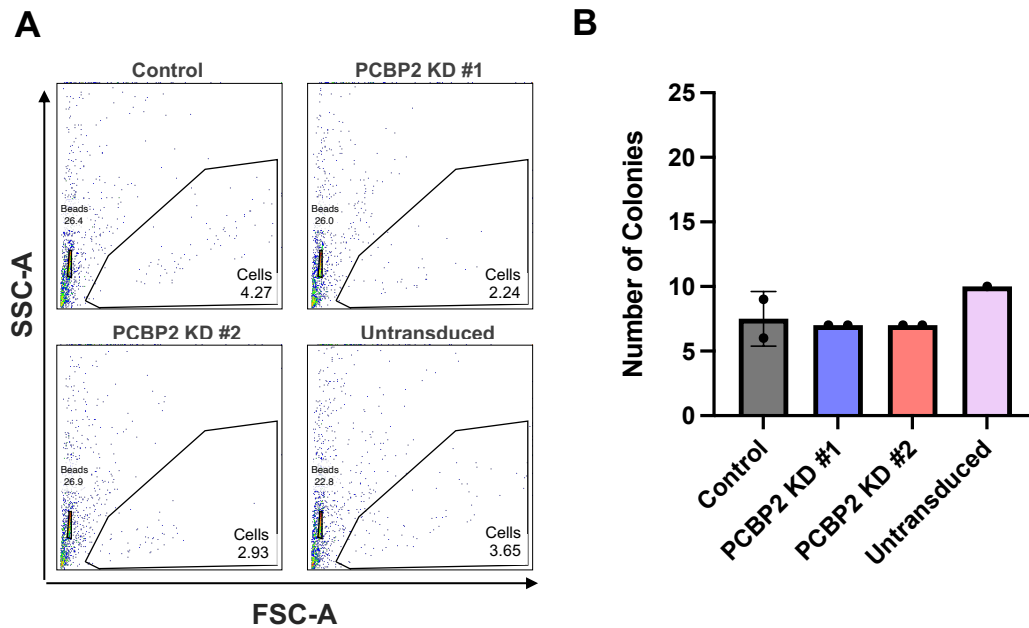


Figure 5.8. *TP53*-sAML cells fail to expand in multilineage differentiation cultures or generate colonies in methylcellulose-based CFU assays. (A) Cell gating from flow cytometry analysis of *TP53*-mutant sAML cells following a 12-day liquid culture multilineage differentiation protocol. **(B)** Number of CFU following transduction with scrambled shRNA control, either shRNA PCBP2 KD #1/KD #2 or cells left untransduced. 500 cells seeded per replicate for CFU assays. Data points represent technical replicates from a single experiment.

Given *TP53*-mutant sAML samples were run alongside HD HSPC samples described in Section 5.2.1, lack of cell survival and proliferation is specific to *TP53*-sAML samples. Results show transduction of LSC from *TP53*-sAML samples was improved via adoption of a smaller lentiviral shRNA construct and a “double transduction” protocol, however, neither liquid culture conditions or a methylcellulose-based CFU assays were supportive of transduced LSC populations survival, proliferation or clonogenic potential over 12-14 days. This work highlighted the need for alternative models for the investigation of *PCBP2* perturbation in *TP53*-sAML.

5.2.3 Engraftment of primary *TP53*-sAML samples into a human bone marrow organoid model enhances cell survival and proliferation

Given the difficulties experienced in manipulating gene expression and sustaining primary cells *in vitro* using conventional liquid culture systems, I adopted a human bone marrow organoid model to support primary cell proliferation and survival^{137,138}.

Previously, this human bone marrow model has been shown to support primary cells from various primary malignancies including MF, multiple myeloma and ALL¹³⁷. However thus far, this system has not been shown to support the proliferation and survival of primary AML, highlighting the importance of this work.

As a proof of principle, primary cell proliferation and survival were compared in organoid models and conventional liquid culture conditions (**Figure 5.9**). Briefly, *TP53*-sAML patient samples and HD controls were thawed, CD34+ HSPC enriched by FACS and stained with CellTrace to discriminate between engrafted primary cells and organoid-derived populations. CellTrace stained cells were then engrafted into organoids or seeded for liquid culture. Cultures were maintained for 7-days with engraftment, proliferation and viability assessed on day-3 and day-7. Media conditions supportive of the organoid bone marrow niche have previously been established (shared by Dr Abdullah Khan, unpublished data) and are capable of supporting myeloid malignancies *ex vivo*. Thus, in principle, these conditions should allow the organoid microenvironment to support engraftment of primary AML samples. However, primary AML cells can be uniquely challenging, and given the experienced difficulties in culturing *TP53*-mutant transformed MPN samples, I began by comparing standard organoid media conditions against AML-specific culture conditions. This comparison was important as little was known about the requirement of AML-specific cytokines, and their concentrations, on these transformed samples *ex vivo*. Due to limited HD HSPC for engraftment, organoid-standard and AML-specific media conditions were only compared for *TP53*-sAML samples.

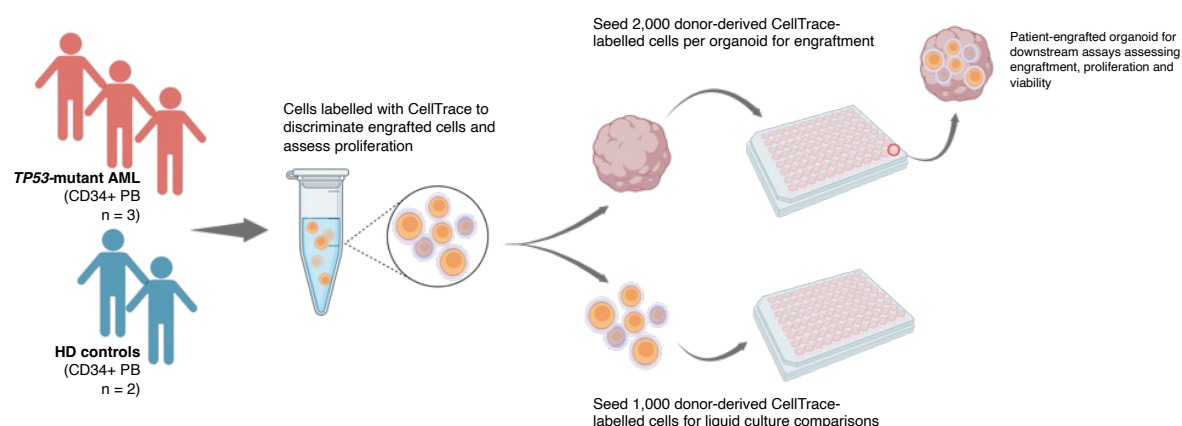


Figure 5.9. Schematic workflow of primary cell engraftment into human bone marrow derived organoids for comparison against conventional liquid culture conditions. Illustration created using BioRender.com

Three-days post-engraftment, distinct CellTrace-positive populations were detectable in bone marrow organoids, confirming successful primary cell engraftment (**Figure 5.10**). CellTrace dye becomes diluted with each cell division, allowing cell proliferation to be monitored over multiple generations. Hence, reduced CellTrace signal observed in day-7 images is a visual indication of primary cell proliferation. In contrast to organoid cultures, liquid culture conditions appeared to retain their bright CellTrace-positive signal and showed no visual signs of cell expansion over the analysis period.

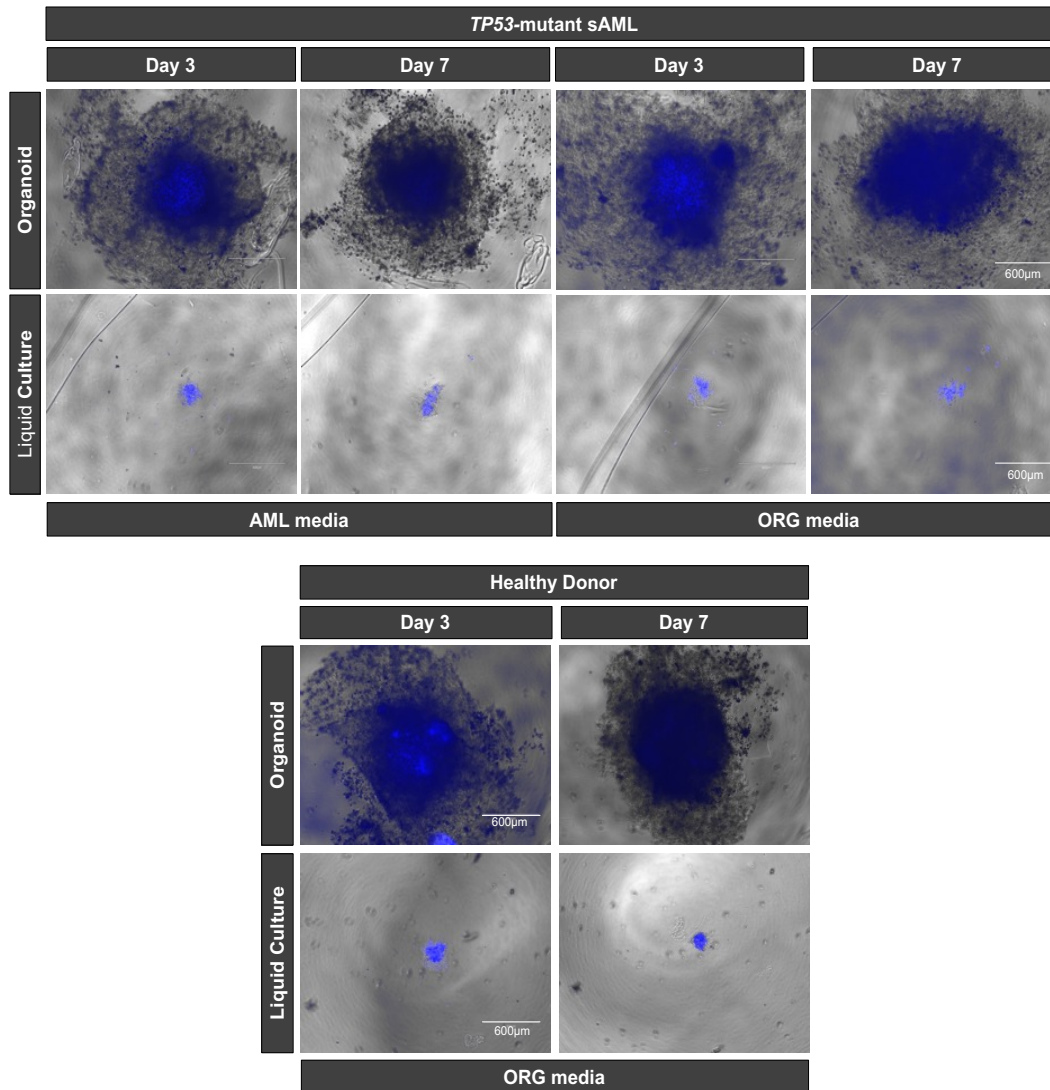


Figure 5.10. Human bone marrow organoids engrafted with CellTrace-stained *TP53*-sAML samples or HD controls compared to liquid culture conditions. Donor-derived cells are identifiable by violet CellTrace staining. Two media conditions containing different cytokine cocktails, an AML-specific and standard-organoid, were compared for culturing *TP53*-sAML cells in liquid culture and bone marrow organoid models.

In line with imaging, flow cytometry analysis confirmed the serial dilution of CellTrace dye from organoid engrafted primary patient samples, with an obvious shift in CellTrace peaks between day-3 and day-7 (**Figure 5.11A**). Furthermore, proliferation was significantly improved following organoid engraftment of *TP53*-mutant LSC in organoid-specific media, but not in AML-specific media conditions, compared to liquid culture counterparts (**Figure 5.11B**). In contrast to organoid engrafted patient-derived cells, patient cells seeded in liquid culture conditions showed very minimal shifts in CellTrace intensity, reflective of reduced cell proliferation.

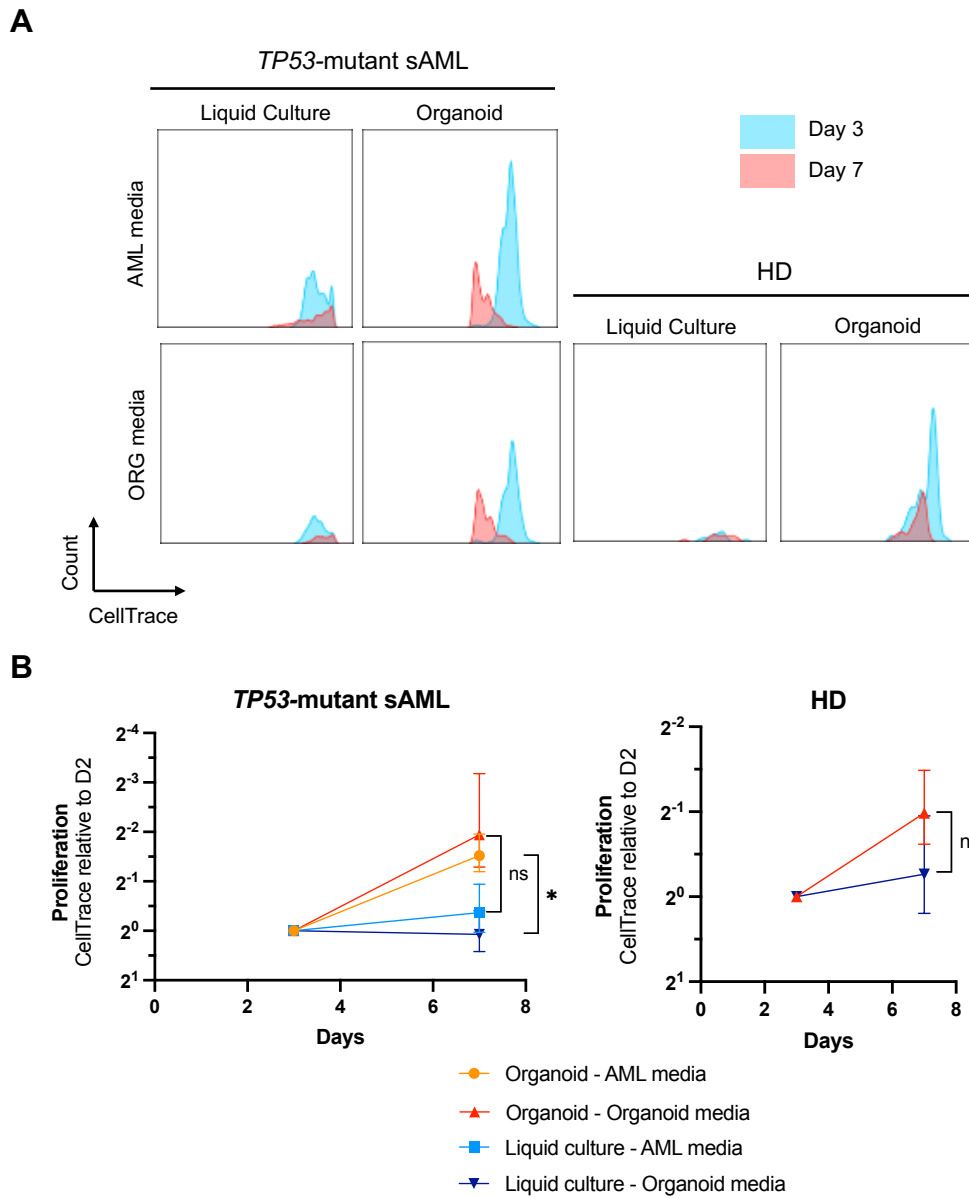


Figure 5.11. The effect of human bone marrow organoid engraftment on primary cell proliferation. (A) Serial dilution of CellTrace peaks indicative of cell proliferation, comparing *TP53*-mutant sAML patients and HD controls in bone marrow organoids and standard liquid culture conditions at day 3 and day 7 post- thawing/engraftment. Additional comparison included use of AML-specific or organoid-standard culture media for *TP53*-mutant sAML samples. **(B)** Proliferation of *TP53*-mutant sAML samples (n=3) and HD (n=2) controls following 3 and 7 days in liquid culture or organoids. Mean fluorescent intensity (geometric mean) at day-7 was normalised to day 2 and presented as reverse $\log_2()$. Flow cytometry gating was for single cells and then CellTrace-positive populations. Data represents mean $\pm$ SD with significance calculated using unpaired two-way Student's *t*-test where $*p < 0.05$.

Primary cell viability was also assessed via flow cytometry on day-3 and day-7. CellTrace-positive primary cells were found to be viable via exclusion of viability dye, 7AAD. Results showed primary cell viability was significantly improved by engraftment into bone

marrow organoids compared to 2D liquid culture, independent of the culture conditions used (**Figure 5.12**). Healthy donor samples also showed significantly improved viability following organoid engraftment on day-7.

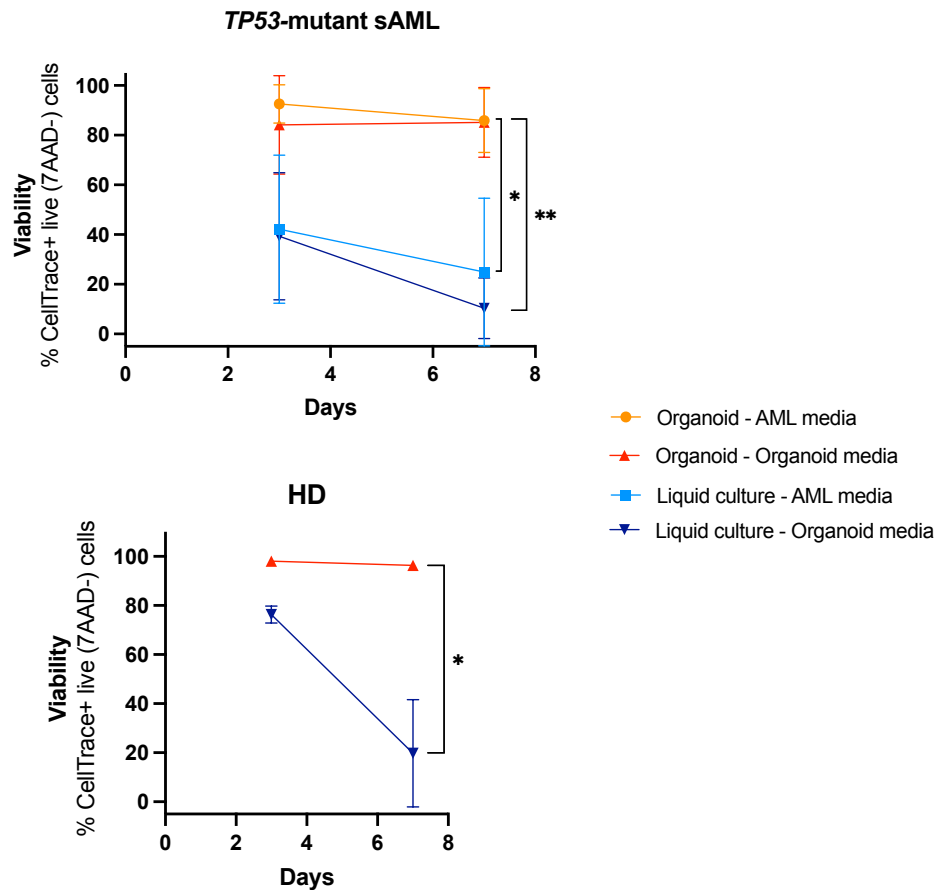


Figure 5.12. The effect of human bone marrow organoid engraftment on primary cell viability. Viability of *TP53*-sAML samples (n=3) and HD (n=2) controls in liquid culture and following organoid engraftment. Comparison includes use of AML-specific and organoid-standard media conditions for *TP53*-mutant sAML samples. Data represented mean +/- SD with significance calculated using unpair two-way Student's *t*-test where * $p < 0.05$ and ** $p < 0.01$.

Overall, results show engraftment of primary cells into a human bone marrow organoid model significantly improves primary cell proliferation and survival compared to conventional 2D liquid culture systems. Differences in performance of AML-specific and organoid-standard media in supporting primary patient samples was found to be minimal and standardised media conditions have previously been described to support the engraftment of disparate haematological malignancies into organoids¹³⁷. Therefore, in line with previously published work, future experiments in this thesis used organoid-

standard media. Finally, having validated a human bone marrow organoid model as a superior system for supporting primary *TP53*-sAML cells, its utility for assessing the effect of *PCBP2* depletion in primary *TP53*-mutant LSC was investigated.

5.2.4 A human bone marrow organoid model facilitates functional assessment of *PCBP2* knockdown in primary cells

Having shown above organoid engraftment facilitates the improved survival and proliferation of *TP53*-sAML cells, the next step was to investigate whether an organoid system could support primary samples following perturbation of *PCBP2* and facilitate subsequent functional readouts. To do this I returned to my optimised shRNA lentiviral transduction approach to deplete *PCBP2* in primary *TP53*-mutant LSC and control HD HSPC. If an organoid system was successful in supporting transduced samples, I hypothesised *PCBP2*-KD would selectively reduce survival and proliferation of LSC from *TP53*-sAML patients compared to HSPC from HD controls.

Patient-derived cells transduced with shRNA constructs will express RFP and hence, be distinguishable from organoid-derived cell populations and tractable in the organoid model. Cell numbers for primary experiments are often a limiting factor and through this experimental workflow including thawing, recovery and CellTrace staining, many cells can be lost. Therefore, to avoid limiting cell numbers per condition and to maximise transduction and engraftment efficiencies through high cell densities, only scrambled shRNA control and shRNA *PCBP2* KD #2 were used for the following experiments.

To increase the likelihood that any differential effects observed following *PCBP2*-KD were due to targeting of the *TP53*-mutant LSC clone, *TP53*-sAML samples were selected based on high blast counts and *TP53* VAF. For example, patient IF0394 had a blast count >80% and a p53 VAF of 77.78% for the variant c.440T>A p.Val147Asp. All samples selected had biallelic *TP53* mutations or expected LOH events (based on VAF exceeding 50%) in *TP53*, a *JAK2* or *CALR* mutation and a complex karyotype. While other co-mutations were likely present, they were not considered during sample selection.

As illustrated in **Figure 5.13**, cells were thawed and left to recover overnight in cytokine-rich media. The following day, cells were stained with CellTrace and column-enriched for CD34+ HSPC/LSC populations. *TP53*-sAML samples were selected based on high blast counts and *TP53* VAF, determined previously by bulk genotyping of MNC or CD34-enriched cells²⁰⁷. This was important to ensure any effect observed following *PCBP2* depletion was due to targeting within the dominant *TP53*-mutant LSC clone.

Cells were then transduced as described previously using the optimised “double transduction” protocol and afterwards, cells were left to recover in the incubator overnight. Co-staining transduced cells with CellTrace was an important experimental control, allowing any issues with engraftment and/or transduction to be disentangled from one another. For example, if cells were not RFP-positive upon imaging, CellTrace staining would allow us to determine whether samples failed to engraft or were not successfully transduced. The day following transduction, 5,000 cells were engrafted per organoid. A small proportion of cells were taken for FACS where a homogenous population of transduced RFP+CD34+ cells were collected to determine *PCBP2*-KD. After 7-days in culture, the effect of *PCBP2* depletion on primary cells was assessed via flow cytometry.

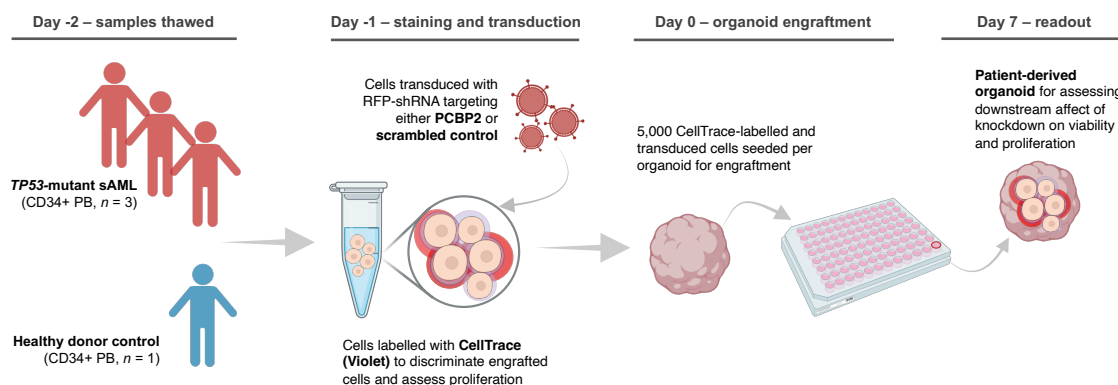


Figure 5.13. Workflow for human organoid engraftment with *PCBP2*-depleted primary cells. Samples from *TP53*-sAML (n=3) and HD (n=1) donors were thawed and left to recover overnight in cytokine-rich media. The following day, cells were transduced with lentiviral shRNA vectors, targeting *PCBP2* or a scrambled control. Cells were also stained with CellTrace Violet for discrimination following engraftment, and assessment of proliferation. Twenty-four hours after transduction, 5,000 CD34+ transduced HSPC were engrafted per organoid. Media was replenished 3-days post-engraftment and the effect of *PCBP2*-KD on patient-derived viability and proliferation assessed by flow cytometry on day-7. Illustration created using BioRender.com.

Primary samples from three *TP53*-sAML patients and one HD control were thawed for transduction with scrambled shRNA control or shRNA *PCBP2* KD #2 lentiviral constructs. Transduction efficiency was incredibly efficient 24-hours post-transduction with over 90% of all donor cells across conditions expressing RFP (**Figure 5.14A**). Highly efficient transduction was contrary to *PCBP2*-KD efficiencies. Twenty-four hours post-transduction patient samples, IF0394 and IF0944, showed no reduction in *PCBP2* expression (data not shown). A population of homogenous RFP+ HSPC were kept in culture for re-assessment of *PCBP2* expression at later timepoints. RT-qPCR results from 96-hours were concordant with 24-hour results, showing no reduction of *PCBP2* expression in either sample IF0394 or IF0944 (**Figure 5.14B**). At both timepoints, *PCBP2* expression in the remaining two AML samples, IF0318 and IF01377, was reduced by ~40%, suggesting KD occurs early post-transduction and is stable when present.

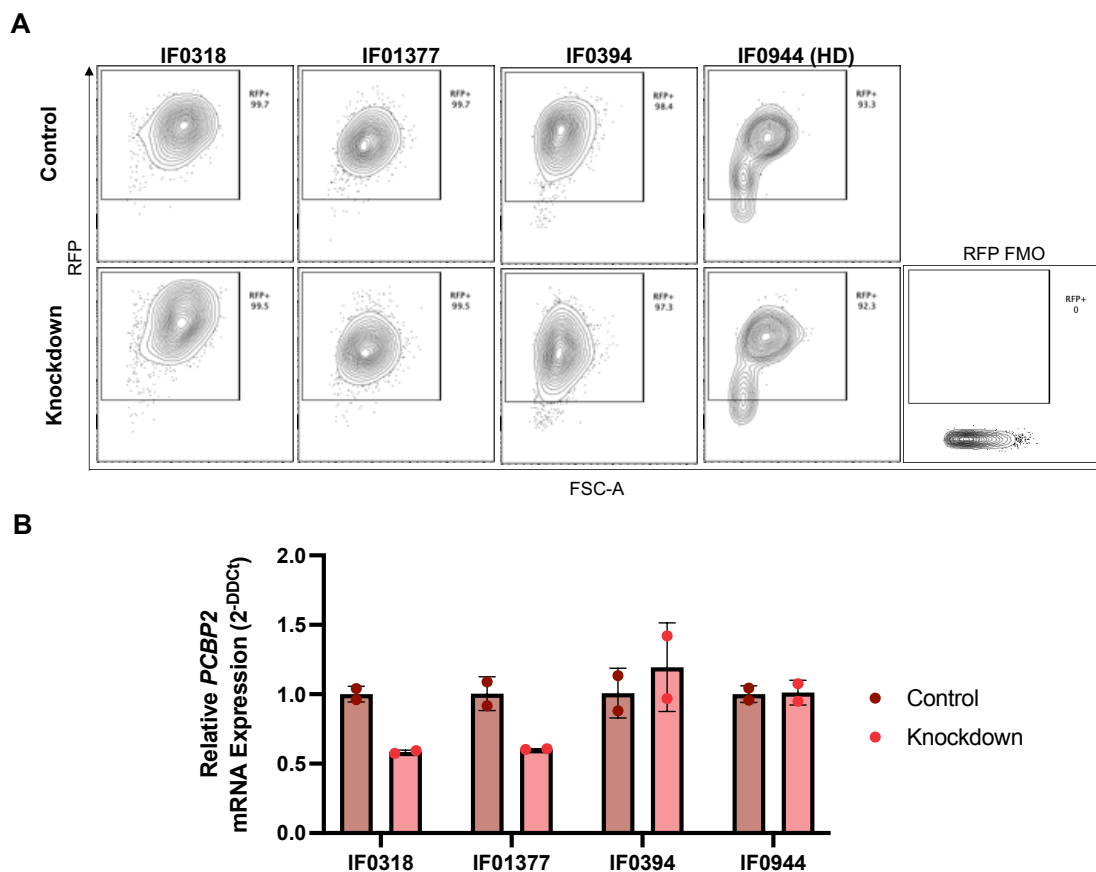


Figure 5.14. Lentiviral transduction of primary *TP53*-sAML samples with *PCBP2*-targeting shRNA. (A) Transduction efficiency of live CD34+ cells determine by RFP expression in *TP53*-sAML and HD samples using scrambled shRNA control or shRNA *PCBP2* KD #2. **(B)** Relative *PCBP2* mRNA expression following *PCBP2*-KD compared to scrambled control as determine by RT-qPCR. Knockdown is calculated relative to control per patient sample. Data represents mean +/- SD for two RT-qPCR technical replicates per sample. FMO; Fluorescence minus one.

Seven days post-engraftment both CellTrace and RFP signals could be identified within the organoids, corroborating transduction and confirming successful engraftment (**Figure 5.15**). RFP signals are more intense than CellTrace on day-7 as a result constitutive RFP expression following lentiviral transduction compared to the serial dilution of CellTrace with each cell division. This data provides a proof-of-principle that organoid models can provide a tractable system for investigating functional gene requirements *in vitro*.

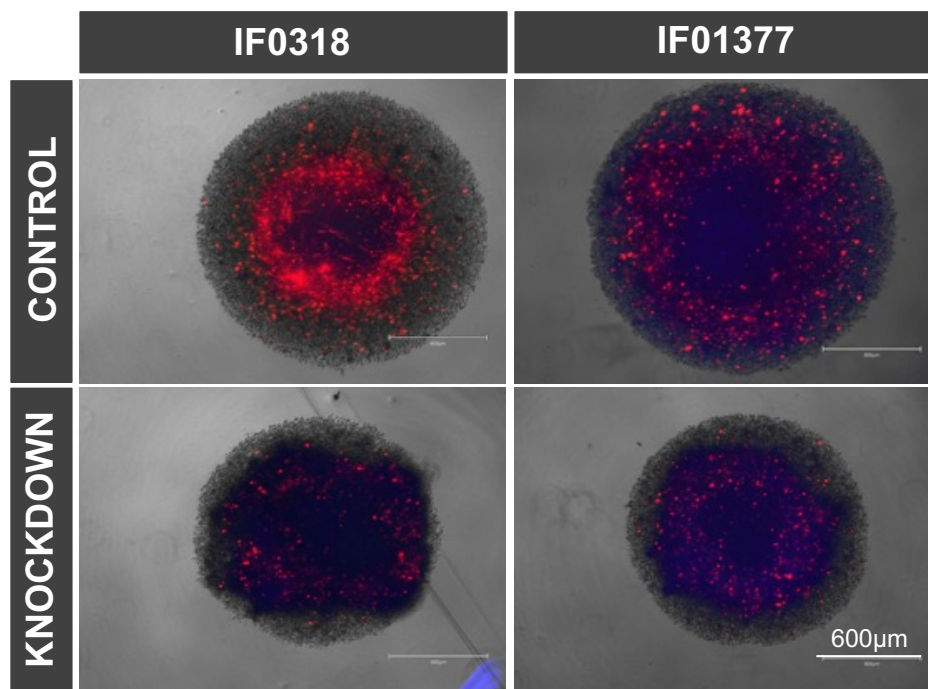


Figure 5.15. Transduced LSC from *TP53*-sAML patients are engraftable and tractable in human bone marrow organoids. Images taken 7-days post-engraftment. Cells transduced with shRNA lentiviral vectors and CellTrace Violet dye are identifiable by red and violet staining, respectively.

Next, I hypothesised *PCBP2*-KD would lead to a reduction in the viability, percentage and absolute number of RFP-expressing cells over 7-days compared to engrafted cells targeted with the scrambled control. After 7-days, organoids were harvested and pooled for flow cytometry analysis of RFP expressing primary populations. Per patient technical replicate, three organoids were dissociated and combined for flow cytometry analysis. No difference in viability was observed between control and knockdown conditions for either *TP53*-sAML sample engrafted, with all samples remaining highly viable (**Figure 5.16A**). Data showed a nominal increase in RFP expressing cells following *PCBP2*-KD

compared to scrambled shRNA control, however, this difference was not significant when patient data was combined for statistical testing (**Figure 5.16B**).

The inclusion of counting beads during flow cytometry acquisition also allowed an estimate of absolute viable RFP+ cells to be determined. In line with the increased percentage of RFP+ cells following *PCBP2*-KD, results showed an increase in absolute cell numbers in *PCBP2*-KD cell populations compared to scrambled controls. However, this increase in proliferation was determined to be non-significant upon statistical testing (**Figure 5.16C**).

Overall, results show *PCBP2* depletion, closely resembling a heterozygous genotype, has no significant effect on viability or proliferation of LSC from *TP53*-sAML patients compared to cells transduced with a scrambled shRNA control. Given the incomplete KD of *PCBP2* in this study, whether *PCBP2* is essential for primary *TP53*-mutant sAML cell survival remains inconclusive and requires further investigation. Despite this, in this chapter I have conclusively shown *PCBP2* depletion in HD HSPC has minimal effect on proliferation, differentiation and clonogenicity using both KO and KD approaches. Importantly, homozygous knockout of *PCBP2* did not reduce colony formation compared to *AAVS1* controls, a crucial finding in the validation of a *PCBP2* as a selective molecular vulnerability in *TP53*-sAML. Furthermore, this thesis presents the first proof-of-principle evidence that human bone marrow organoids can support primary AML, an important finding for the haematology field. Finally, I have presented a tractable system for investigating candidate gene perturbation in challenging patient samples utilising a human bone marrow organoid model.

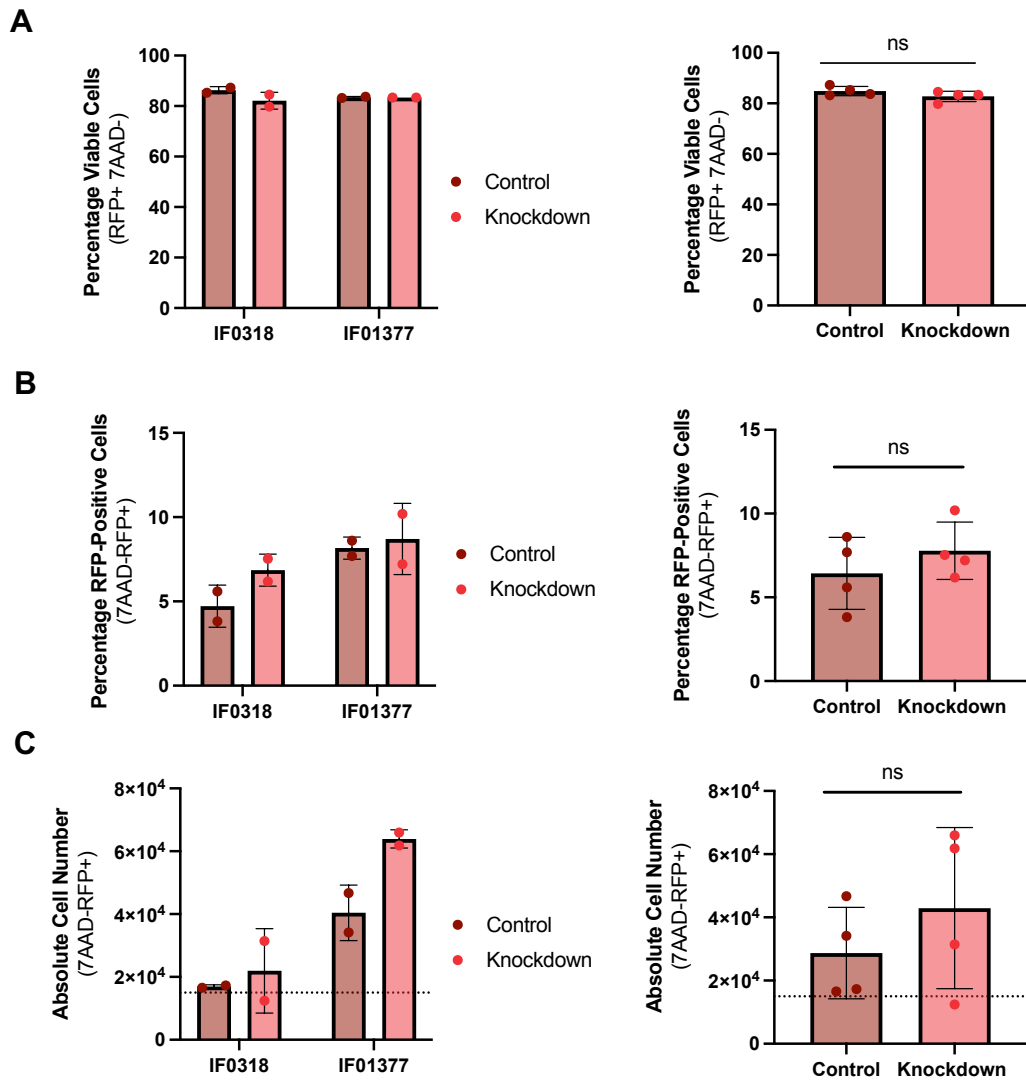


Figure 5.16. Effect of *PCBP2* depletion on viability and proliferation of LSC from *TP53*-sAML samples following 7-days organoid engraftment (A) Effect of *PCBP2*-KD on transduced cell viability compared to scrambled shRNA control in two independent *TP53*-sAML donors (B) Effect of *PCBP2*-KD on total percentage of live (7AAD-) RFP+ cells (C) Effect of *PCBP2*-KD on absolute cell number (7AAD-RFP+) as determine by counting beads via flow cytometry. Horizontal line highlights starting cell numbers engrafted for technical triplicates. For each sub-figure, independent patient data is presented (left) and combined for statistical testing (right). Data represents mean $\pm$ SD for technical replicates where three organoids were pooled per replicate with significance determined by two-way unpaired Student's *t*-test.

5.3 Discussion

This thesis has shown *PCBP2* is required for model cancer cell line survival and proliferation. Data mining and literary review suggested *PCBP2* is non-essential for *TP53*-WT HSPC function. Therefore, this chapter investigated whether *PCBP2* is selectively required by *TP53*-mutant LSC compared to *TP53*-WT HSPC, to determine whether an index exists to therapeutically target *PCBP2*.

While cancer cell line models are an invaluable research tool, their utility is limited and ultimately, any therapeutically translatable findings need to be validated in primary material. To investigate the requirement of *PCBP2* in primary *TP53*-sAML, called for an effective editing strategy to perturb *PCBP2* expression, in the absence of commercially-available inhibitors, and a system for effectively culturing *TP53*-sAML cells *ex vivo*. In this chapter, I utilised a highly effective lentiviral transduction protocol for gene perturbation in primary patient samples and adopted an iPSC-derived human bone marrow organoid model to support the growth and proliferation of primary *TP53*-sAML samples.

To begin investigating whether results from cell line models are translatable to primary patient samples, and whether a therapeutic index exists, an effective methodology for gene perturbation in primary HSPC was required. Compared to lentiviral transduction of cell lines, transduction of HSPC population is increasingly difficult and requires pre-stimulation of HSPC in tailored cytokine conditions, high cell densities, exposure to higher lentiviral MOI and spinoculation^{328–331}. Initial attempts to transduce HSPC population used an all-in-one Cas9-sgRNA-GFP lentiviral construct. However, results showed complete absence of lentiviral transduction with this construct which exceeded 10Kb in size (**Figure 5.6**). This result was not entirely unexpected, since lentiviral packaging and transduction efficiency are known to reduce inversely with transgene size^{331–333}. Given the difficulty in transducing HD HSPC with this vector, no attempts were made in LSC populations and alternative approaches explored.

A smaller RFP-shRNA lentiviral vector (<10Kb), was found to be highly effective in the transduction of both HSPC and LSC (**Figure 5.3 & 5.7**). Using this approach, the effect of *PCBP2* depletion in *TP53*-WT HD HSPC was determined. Results showed no difference in

proliferation, differentiation or clonogenicity of HD HSPC suggesting minimal effect of *PCBP2*-depletion on *TP53*-WT HSPC function (**Figure 5.4**). However, an important caveat of these results was the incomplete KD of *PCBP2* (**Figure 5.3**). Transduction with either *PCBP2* KD #1 and #2 shRNA depleted *PCBP2* expression by ~50% relative to scrambled control shRNA, mimicking a *PCBP2* heterozygous genotype. This was an interesting finding considering the high transduction efficiency achieved in primary cells and the almost complete KD observed in cell lines using identical constructs. A lower KD efficiency in primary cells could be explained by different promoter usage or efficiency of the shRNA processing machinery in primary cells compared to cell lines. Hence, while these experiments cannot determine the effect of complete *PCBP2*-KD on *TP53*-WT HSPC function, we can conclude that up to a 50% depletion in *PCBP2* expression does not detrimentally affect *TP53*-WT HD HSPC.

Transduction of *TP53*-sAML samples with RFP-shRNA lentiviral vectors was variable across experiments, ranging between ~55-95% (**Figure 5.7A & 5.14A**), highlighting the impact of experimental and interpatient variability on transduction efficiency. In transduced LSC populations, *PCBP2*-KD was reflective of efficiencies observed in *TP53*-WT HSPC (**Figure 5.7C**). However, samples failed to survive in liquid culture conditions and thus, results of functional assays could not be assessed. This could be considered unexpected as samples showed high viability during FACS for functional assays three days post-transduction in liquid culture (**Figure 5.7B**). Untransduced control populations also failed to survive, clarifying survival was independent of viral exposure and gene perturbation (**Figure 5.8**). Furthermore, HD and *TP53*-sAML samples were run in parallel, largely ruling out experimental error contributing to lack of primary AML cell survival.

Exposure to specific cytokines is important for HSPC proliferation, survival and instructive control over cell lineage fate, therefore it is important to consider whether cytokine conditions contributed to lack of *TP53*-sAML cell survival¹⁷. The cytokine cocktail used for multilineage liquid culture proliferation and differentiation assays following shRNA-KD (conditions described in Section 2.2.5) was primarily based on cytokine combinations and concentrations described in Psaila *et al*⁶⁹ for the differentiation of myelofibrosis samples. Given many samples utilised in this project were

derived from patients with primary MF who subsequently transformed, and higher cytokine concentrations can be required to encourage differentiation, these conditions were deemed suitable for examining proliferation and differentiation of *TP53*-mutant LSC. An alternative “AML-specific” cytokine cocktail, described by Rodriguez-Meira *et al*⁶³, included lower concentration of key cytokines (SCF, TPO, IL-6), and was used for testing AML survival following engraftment into organoids and parallel liquid culture conditions (conditions described in Section 2.2.9). Despite alterations to cytokine cocktails, when *TP53*-sAML cells were cultured in either medium described above, *TP53*-mutant LSC failed to grow in liquid culture conditions, highlighting the extreme difficulties faced when culturing AML cells *ex vivo*.

Evaluation of the literature concluded no standardised approach has been established for the culture of AML cells following lentiviral transduction. Transduced AML cells are often placed in MSC co-cultures or PDX engraftment models which provide additional factors resembling the bone marrow niche^{299,334–336}. Transduced cells can also be seeded into methylcellulose-based cultures, although experiments are often limited to 7-days rather than the standard 14-days adopted for HD cell CFU assays^{334,337}. Similarly, shorter experimental readouts have been described in 2D co-cultures, limiting time in culture to 6-days post-transduction³³⁶. In this thesis, functional effects of *PCBP2*-KD were assessed after 12-14 days, not including the 72-hours required post-thawing for cell recovery, lentiviral transduction and visualisation of RFP expression. Therefore, I expect extended experimental timelines likely contributed to the liquid culture failings in this project.

CFU assays also used lower seeding densities than reported elsewhere, an important consideration given AML cells present reduced colony output potential compared to HD HSPC with to the majority of CD34+ cells being leukaemic blasts, rather than “true” stem and progenitor populations^{299,327,337}. Thus, CFU assays using *TP53*-sAML samples may have been improved with higher seeding densities and whether this would increase colony formation in semi-solid media remains an open question and requires further investigation. Overall, evaluation of the data and literature suggests 2D liquid cultures and methylcellulose-based systems adopted in this thesis lacked the niche complexity required for LSC proliferation and survival over an extended time period.

Having presented evidence showing AML samples are notoriously difficult to culture *in vitro*, I adopted a human bone marrow organoid model to better support the survival and proliferation of *TP53*-sAML cells. Given *TP53*-sAML samples for this project are finite, I bypassed testing and optimisation of 3D HSPC-MSC co-cultures, since organoids are a superior model in supporting haematological malignancies, including MF, ALL and multiple myeloma¹³⁷. However, while a human bone marrow organoid model was adopted in this DPhil project, it cannot be dismissed that a co-culture system may be a simpler alternative to organoids in maintaining the viability of *TP53*-sAML samples and requires validation. Results showed engraftment of patient samples significantly improved survival and proliferation over 7-days compared to conventional liquid culture conditions (**Figure 5.11 & 5.12**). This result was independent of the organoid-standard or AML-specific media conditions used, emphasising the utility of a standardised organoid system to support a range of haematological malignancies. A limitation of this work was the inability to track viability and proliferation of engrafted cells long-term due to the dilution of CellTrace with each cell division. Therefore, future work may use an iPSC cell line with a stable fluorescent marker to discriminate patient and organoid-derived cells.

Adopting a human bone marrow organoid model supportive of primary haematological malignancies gave me an opportunity to validate *PCBP2* as a molecular vulnerability in primary *TP53*-sAML. To do so, I utilised the optimised lentiviral shRNA approach to perturb *PCBP2* in *TP53*-sAML LSC prior to organoid engraftment. Transduced cells expressed RFP and therefore provided a tractable system for functional interrogation following *PCBP2* depletion. Despite transduction efficiencies for all samples exceeding 90% 24-hours post-transduction, two samples (IF0394 and IF0944) showed no evidence of *PCBP2*-KD compared to scrambled control shRNA (**Figure 5.14**). This result could not be explained by a rapid loss of the disadvantaged *PCBP2*-KD clone or delayed depletion of *PCBP2* transcripts, since *PCBP2* expression remained consistent at 24- and 72-hours post-transduction. Samples IF0394 and IF0944 were not included in the TARGET-seq dataset and expression of *PCBP2*, which may have influenced the knockdown efficient, was therefore not known. However, *PCBP2* expression via RT-qPCR data for both samples were found to be comparable to samples where a KD was achieved suggesting low *PCBP2* expression was not responsible for the limited knockdown in these samples. Thus, the

absence of *PCBP2*-KD in some patient samples but not others, despite high levels of RFP expression post-transduction, is peculiar and likely a donor-dependent effect of lentiviral transduction.

Having a tractable organoid model system for the functional assessment of *PCBP2*-KD in LSC from *TP53*-sAML patients, I assessed viability and proliferation of patient samples 7-days post-engraftment following *PCBP2*-KD. Five-thousand cells were engrafted per organoid, and assuming uniform engraftment across conditions, results were largely inconclusive. Results showed no significant difference in viability of transduced populations or changes in the percentage nor absolute number of RFP+ cells following *PCBP2*-KD compared to cells transduced with scrambled control (**Figure 5.16**). This result is not too concerning, considering a 50% reduction in gene expression is unlikely to show a clear phenotype for many essential candidate genes. However, to confidently conclude the effect of *PCBP2* in *TP53*-sAML, *PCBP2* depletion needs to be closer to representing a homozygous deletion.

Short hairpin-RNA expression is dependent on promoter strength for high expression and efficiency; hence promoter selection is an important consideration for lentiviral transduction. A number of promoters have previously been adopted for the transduction of human HSPC including EF1 α ³³⁸⁻³⁴⁰, CMV^{328,338,339}, SFFV³³¹, MSCV^{329,330,338} and SV40³³⁹. The EF1 α promoter has been utilised for transduction of HSPC for autologous haematopoietic stem cell gene therapies³⁴⁰, justifying the utility of the EF1 α promoter for transduction of HSPC in this project. Alternative promoters may enhance shRNA-KD of *PCBP2* and attempts to identify optimal promoters for transgene expression in HSPC have been made, although comparisons and findings are variable. One comparative study showed CMV and EF1 α promoters drove higher GFP expression compared to UBC and SV40 promoters in CB HSPC³³⁹. In contrast, a second study suggested CMV is most suitable for transgene expression in cell lines with other promoters, namely MSCV and PGK, outperforming CMV in CB HSPC³³⁸. Hence, no promoter can be clearly justified as superior for transgene expression in HSPC. With regards to this thesis project, future experiments will consider adopting alternative promoters to improve expression of shRNA constructs and therefore the *PCBP2*-KD in primary HSPC/LSC populations.

Another factor possibly limiting the depletion of *PCBP2* in HSPC/LSC populations is the expression of the RNase III enzyme, Dicer. Recognition of shRNA duplex by Dicer, processes shRNA into small interfering RNA (siRNA) intermediates, lacking the shRNA terminal loop^{341,342}. Processed siRNA duplexes are subsequently incorporated into RNA-induced silencing complex (RISC), a multicomponent nuclease complex which identifies target mRNA transcripts through complementary base pairing, initiating mRNA degradation^{341,342}. Thus, inherent expression and activity of Dicer in HSPC populations could be a limiting gene silencing via shRNA. This hypothesis is corroborated by evidence in ovarian cancer, where reduced Dicer expression in primary patient samples leads to limited gene silencing following shRNA transduction³⁴³. In future experiments, assessing Dicer expression and including a positive control shRNA known to have a high KD efficiency in other primary cell types will be important troubleshooting steps.

Interestingly, *PCBP2* has been shown to associate with Dicer and enhance miRNA/shRNA processing³⁴⁴. Depletion of *PCBP2* was found to reduce the efficiency of shRNA-mediated KD, by inhibiting processing of shRNA precursors³⁴⁴. *PCBP2* expression was also shown to enhance endogenous microRNA processing, which also required Dicer activity, but had no effect on the silencing potential of siRNA which do not require processing by Dicer. These findings suggest targeting *PCBP2* via shRNA leads to reduced shRNA processing efficiency, which in turn limits the processing of *PCBP2*-targeting shRNA. Whether this is the mechanism by which *PCBP2*-KD is limited in the primary cell data presented in this thesis requires further investigation. In this regard, it would be informative to compare *PCBP2*-KD knockdown efficiency to that of other genes not expected to influence shRNA/miRNA processing. Finally, these findings suggest alternative promoters driving shRNA expression, as described above, may not be an appropriate solution to enhance *PCBP2*-KD efficiency, however both possibilities will need to be investigated.

Alternative strategies to deplete primary cell *PCBP2* expression prior to organoid engraftment include CRISPR-editing techniques. To adopt traditional CRISPR-KO techniques, such as the RNP approach previously described, would require a marker for the discrimination of transfected donor-derived cells within engrafted organoids. Fluorescently labelled Cas9 is commercially available and could be utilised to purify and/or discriminate Cas9-expressing cell populations from organoid populations.

Similarly, tracrRNA, which anneals target-specific crRNA forming sgRNA duplexes which complex with Cas9, can also be fluorescently labelled. However, transfection efficiency with Cas9-sgRNA RNP complexes does not directly correlate to gene editing efficiencies, and prior to organoid engraftment experiments, high transfection and editing efficiencies in primary patient cells needs to be achieved. An alternative strategy is CRISPR interference (CRISPRi), a technique which controls sequence-specific repression of gene expression³⁴⁵. In this repurposed CRISPR system, Cas9 is rendered catalytically inactive (dCas9) and lacks DNA cleavage capabilities. Binding of dCas9-sgRNA to DNA creates a steric block which prevents transcriptional initiation or elongation by RNA polymerase, depending on the position of the target loci³⁴⁵. Hence, other options exist for the perturbation of target genes in primary patient samples and the applicability of each will be considered for future experiments.

In summary, work described in this chapter optimises lentiviral shRNA transduction parameters to investigate the effects of *PCBP2* depletion in primary patient samples. A human bone marrow organoid model was shown to be superior in supporting *TP53*-sAML cells compared to conventional liquid culture conditions and presents a tractable model for functional interrogation of candidate genes in primary patient samples *ex vivo*. The inconclusive results following *PCBP2*-KD in *TP53*-sAML should not dissuade from *PCBP2*'s potential as a therapeutic target in *TP53*-sAML. I have presented evidence suggesting a minimal role of *PCBP2* in *TP53*-WT HSPC, through shRNA and CRISPR-Cas9 approaches and strong evidence in cancer cell lines models showing an important requirement of *PCBP2* in *JAK2/TP53*-mutant AML cell lines. Thus, in this thesis I have made significant progress in developing a tractable system for investigating gene perturbations in *TP53*-sAML and future work optimising gene depletion to reflect a homozygous genotype will allow me to generate irrefutable conclusions on the role of *PCBP2* in *TP53*-sAML.

Chapter 6 Investigating PCBP2 mechanisms of action in *TP53*-sAML

6.1 Introduction

6.1.1 PCBP2 and regulation of gene transcription

Historic data identified PCBP2 as a single-stranded DNA binding factor with affinity for the positive regulatory region of the *BRCA1* promoter³⁴⁶. Other evidence supporting the role of PCBP2 in gene transcription comes from murine models, where a collection of PCBP proteins (PCBP1, PCBP2 and hnRNP K) were identified to regulate transcription of the murine μ opioid receptor gene (*Oprm1*)³⁴⁷.

Van Nostrand *et al*¹⁸⁰ performed ChIP-seq on a cohort of RBP, including PCBP2, to determine the interaction of specific RBP with DNA. In K562 and HepG2 cells, 1842 and 3657 PCBP2 ChIP-seq peaks were identified respectively, highlighting the specific association of PCBP2 with chromatin¹⁸⁰. Most RBP associated with actively transcribed euchromatin, and this remained the case for PCBP2. ChIP-seq peaks for PCBP2 centred around gene bodies, and frequently overlapped with eCLIP-seq peaks¹⁸⁰. Overlap between the two datasets was a rare occurrence, limited to only a small cohort of investigated RBP. Furthermore, while overlapping CLIP-seq and ChIP-seq peaks at gene promoters can simply reflect co-transcriptional association, concurrent peaks within gene bodies likely represent an uncharacterised and distinct recruitment mechanism and requirement of PCBP2 to chromatin¹⁸⁰.

6.1.2 PCBP2 and post transcriptional gene regulation

PCBP2 and regulation of transcript stability and translation

As previously described in Chapter 4, early studies identified PCBP2 as a regulator of mature α -globin stability during terminal erythroid differentiation^{241,242}. Mechanistically, PCBP2 was identified to bind cytidine-rich sequences (-CCUCC-) within the 3'UTR of the

α -globin transcript. Three cytidine-rich regions were mapped within the 109bp 3'UTR and mutations in any single C-rich segment had the capacity to destabilise α -globin mRNA in a translation independent manner²⁴². Thus, *cis*-regulatory poly(C)-rich elements were identified to be non-redundant in the stabilisation of α -globin transcripts by PCBP2. Closely related poly(C)-rich 3'UTR sequences have been identified to stabilise a subset of erythroid transcripts including β -globin and erythropoietin, highlighting the important contribution of PCBP2 to erythroid differentiation and terminal function^{262,348}.

PCBP2 was also shown to co-transcriptionally enhance pre-processing of α -globin transcripts²⁴³. Nuclear 3' cleavage and polyadenylation of pre-mRNA is a crucial determinant of transcript stability and function. PCBP2 was identified to bind poly(C)-rich regions within α -globin 3'UTR and associate directly with 3' processing machinery to enhance both cleavage and subsequent polyadenylation of transcripts²⁴³. Thus, PCBP2 forms a common RNP-complex linking nuclear processing and cytoplasmic stability crucial for the sustained expression of α -globin in erythroid lineages.

PCBP2 is a well-described host factor contributing to the replication and translation of RNA viruses. RNA viruses form extensive secondary structures, including major stem loop structures at their 5' and 3' terminus which are essential for their viral replication cycle. Poliovirus replication requires the formation of a circular RNP complex between the 5' and 3' ends of poliovirus RNA. PCBP2 was identified to be essential for poliovirus RNA replication³⁴⁹. PCBP2 was shown to bind stem loop I within the 5' non-coding regions (NCR), in complex with viral protein 3DC, leading to circularization through association with the 3' bound poly(A)-binding protein, facilitating viral replication³⁴⁹. Host factors are also known to participate in viral translation, recruiting ribosomes to internal ribosome entry sites (IRES) within the 5'NCR of viral transcripts. PCBP2 was identified to bind stem loop IV within the 5'NCR of viral transcripts, containing IRES elements, and promote efficient translation of poliovirus RNA³⁵⁰.

In contrast to the roles of PCBP2 in promoting viral RNA replication and translation, PCBP2 is important for the antiviral activity of IFN- α in the treatment of chronic hepatitis C virus (HCV) infection³⁵¹. PCBP2 was identified to be downregulated in response to chronic HCV infection. However, PCBP2 overexpression was found to enhance the efficacy

of IFN- α treatment for HCV infection by stabilising downstream effectors of IFN- α , *STAT1* and *STAT2*. Mechanistically, PCBP2 stabilised transcripts by binding poly(C)-rich regions within the 3'UTR of *STAT1* and *STAT2*³⁵¹. Thus, assessing PCBP2 expression may help predict response to IFN- α treatment in chronically infected HCV patients. Furthermore, the stabilisation of *STAT* transcripts by PCBP2 has important implications for MPN biology. Overall, PCBP2 presents diverse contributions to viral biology, with roles in both facilitating and perturbing the viral life cycle through a variety of mechanistic pathways and interactions.

In a p53-dependent context, PCBP2 has been reported to regulate the translational response to the small-molecule inhibitor of MDM2, Nutlin-3a⁹⁶. Activation of WT p53 leads to two major cellular responses, cell cycle arrest or apoptosis. Tissues undergoing prolonged cell cycle arrest are more likely to experience disease recurrence, hence, predicting outcomes of Nutlin-3a exposure is important for treatment outcomes. At a transcriptional level, predicting cellular responses to p53-activation is difficult, since a conserved core of p53 target genes are transcribed in cells undergoing both cell cycle arrest and apoptosis. Through sequencing of polysomal fractions, Rizzotto *et al*⁹⁶ discovered translational control of p53 target transcripts, mediated by PCBP2 and DHX90, was responsible for alternative cell cycle arrest and apoptosis responses to Nutlin-3a treatment. In cells undergoing Nutlin-3a induced cell cycle arrest, PCBP2 and DHX90 were found to bind the CG-rich 3'UTR motif of apoptosis-inducing transcripts and suppress their translation, thus favouring translation of transcripts associated with p53-dependent cell cycle arrest⁹⁶. This work highlights the role of PCBP2 in transcript translation, and furthermore, it emphasises the association of PCBP2 with p53-related apoptotic pathways which may be of important relevance to this project.

Related to activation of p53 pathways, co-depletion of PCBP1 and PCBP2 leads to p53-independent activation of p21 and subsequent cell cycle G₁ arrest³⁵². Individual depletion of neither PCBP1 nor PCBP2 induced p21 stabilisation or G₁ arrest, suggesting in this context, PCBP1 and PCBP2 play overlapping and/or redundant roles. PCBP1 and PCBP2 were found to independently associate with a 127bp region within the 3' UTR of *CDKN1A* (p21) transcripts³⁵². This region contains three CU-rich regions, closely related to those identified in α -globin and erythropoietin mRNA. Although, in contrast to the stabilising

effect exerted by PCBP2 in regulation of α -globin and erythropoietin, here and in the previous example, PCBP-family members destabilise transcript expression, highlighting the transcript-dependent and redundant context of PCBP1/2 function.

PCBP2 as a regulator of mRNA splicing

Splicing is the process by which non-coding introns are removed from pre-mRNA transcripts to generate mature mRNA for protein translation. Alternative splicing occurs when multiple mRNA isoforms are identified to be derived from the same gene. Alternative splicing therefore increases the complexity of gene expression and is a crucial component in cellular diversity³⁵³.

The major U2-dependent spliceosome is responsible for 99% of splicing events and is composed of 5 small nuclear ribonuclear proteins (snRNP); U1, U2, U4, U5 and U6 and other non-snRNP. The major spliceosome assembles through sequential recognition of *cis*-acting mRNA encoded sequences. Spliceosome assembly starts with binding of the U1 snRNP to the 5' donor splice site within the 5' intron region, followed by the U2 snRNP binding to the 3' splice acceptor site and upstream adjacent branch site sequence. At this point, U5 and U4/6 are recruited, leading to a series of confirmation changes and the subsequent disassociation of U1 and U4 snRNP. Finally, a catalytically active spliceosome complex has formed, leading to intron excision and exon linkage³⁵³ (**Figure 6.1**).

Alternative splicing is usually accompanied by sequence-specific RBP binding *cis*-acting pre-mRNA sequences to regulate splice events. In this regard, PCBP2 was found to regulate splice activity. PCBP2 binds a C-rich polypyrimidine tract (PTT) upstream of the major splice acceptor site, where PCBP2 interacts with components of the U2 snRNP splice complex, specifically U2AF65³⁵⁴. Importantly, U2AF65 is not binding its common U-rich PPT, replaced in this context by the C-rich PPT bound by PCBP2. Instead, U2AF65 association with the spliceosome machinery is mediated directly via PCBP2³⁵⁴. Thus, via this mechanism, PCBP2 binds the C-rich PPT and regulates the initial stages of spliceosome assembly, deviating from the canonical U2AF65-mediated model to influence global splicing events.

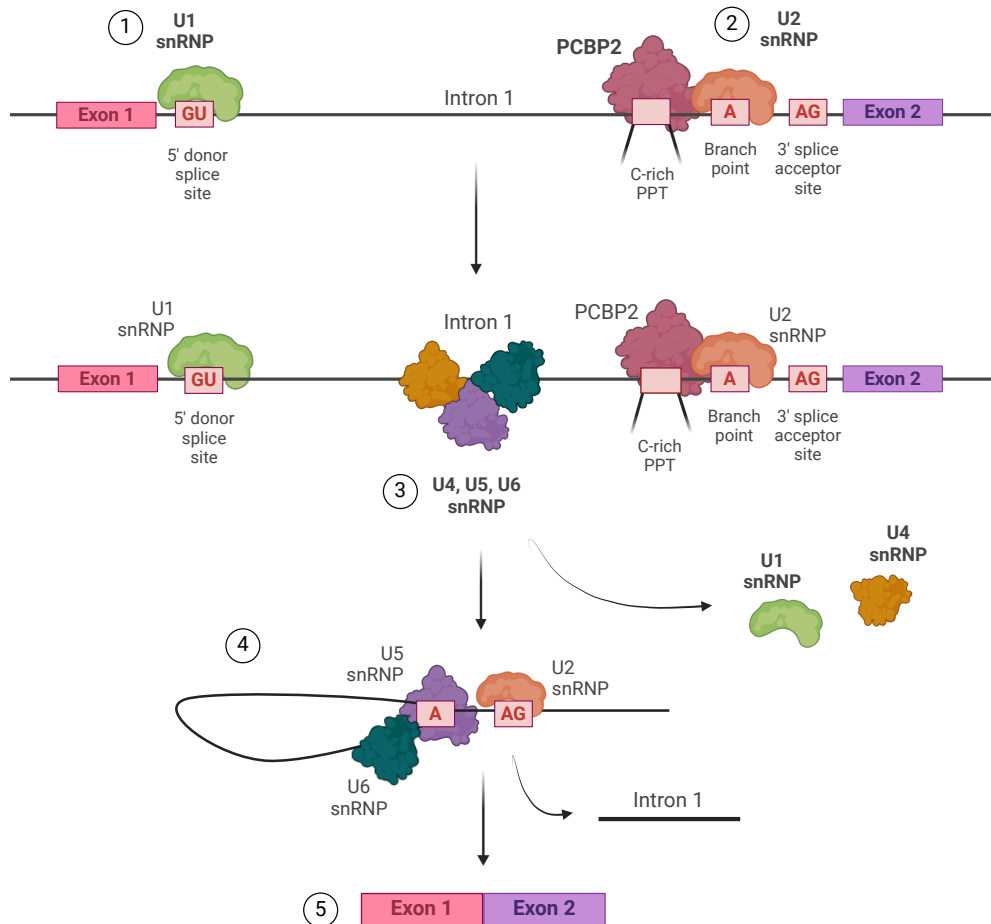


Figure 6.1. Major steps in the U2 spliceosome pathway. The U2 spliceosome is composed of 5 main small nuclear ribonucleoproteins (snRNP). (1) U1 snRNP binds the pre-mRNA 5' donor splice site. (2) U2 snRNP then binds the branch point and downstream 3' splice acceptor site. (3) U4, U5 and U6 snRNP are recruited and the complex undergoes a conformational change. U1 and U4 leave the complex. (4) The final catalytically active spliceosome is formed, (5) leading to intron exclusion and exon linkage. PCBP2 is shown binding C-rich polypyrimidine tract (PPT) upstream of the branch point to mediate splicing. Image created with BioRender.com

6.1.3 PCBP2 and protein-protein interactions

Despite canonical roles as an RBP in PTGR, PCBP2 can also participate in protein-protein interactions. Through these interactions, PCBP2 maintains immune homeostasis via the regulation of antiviral signalling pathways. The cytoplasmic viral-RNA sensor, RIG-I, triggers host innate immune responses through the binding of viral RNA released during viral replication. The interaction of activated RIG-I with adaptor protein, MAVS, leads to production of type I interferons and inflammatory cytokines via transcription factors, IRF3 and NF- κ B, initiating an antiviral response. Fine-tuning of this response is essential to prevent the detrimental effects of excessive inflammatory stimulus. PCBP2 was

identified to negatively regulate the RIG-I-MAVS signalling pathway via associating with MAVS and acting as a scaffold for the recruitment of the HECT domain-containing E3 ubiquitin ligase, AIP4²²⁵. Thus, PCBP2 attenuates RIG-I-MAVS antiviral signalling by mediating AIP4-driven ubiquitination of MAVS, leading to the adaptor proteins proteasomal degradation²²⁵.

Similarly, PCBP2 is also important in the antiviral response to DNA viruses via the cGAS-STING pathway²⁶⁷. Cyclic GMP-AMP synthase (cGAS) is a cytosolic DNA sensor which following DNA-binding, aggregates to form liquid droplets and catalyse the conversion of ATP/GTP to cGAMP. cGAMP is an important secondary messenger which binds STING to form a trimeric complex comprising STING-TBK1-IRF3. Complex formation is required for nuclear translocation and transcriptional activation of IRF3 to facilitate downstream type I interferon and inflammatory signalling pathways. PCBP2 was identified to directly interact with cGAS, antagonising its condensation and abrogating downstream signalling²⁶⁷.

Interestingly, mutant p53 can also suppress cGAS-STING-mediated immune signalling by interacting with TBK1 and preventing STING-TBK1-IRF3 trimer formation, attenuating activation of the downstream type I interferon response³⁵⁵. In tumours, where chromosomal instability leads to the formation of micronuclei prone to rupture and release of dsDNA, mutant p53 suppresses the innate immune response leading to immune evasion via cell-autonomous and non-cell autonomous mechanisms³⁵⁵. In contrast, chronic activation of the cGAS-STING signalling pathway can also promote rewiring of downstream signalling to promote a pro-metastatic tumour environment^{356,357}. Given *TP53*-sAML is characterised by CNA and complex karyotypes, there is evidence to support the role of PCBP2 in downregulation of excessive proinflammatory cGAS-STING signalling to promote leukaemic survival.

6.1.4 Hypothesis

PCBP2 is an RBP with diverse roles in PTGR and protein-protein interactions. I hypothesise interrogation of PCBP2's molecular mechanisms in relevant cell line models will identify potential pathways by which PCBP2 is required for *TP53*-sAML maintenance.

6.1.5 Summary and aims

In previous chapters, I identified PCBP2 as an RBP with the potential for therapeutic targeting in *TP53*-sAML, validated the requirement of *PCBP2* in appropriate cancer cell line models and made significant progress in translating these findings to primary patient samples. PCBP2 is a functionally diverse RBP with roles in PTGR and pathway regulation via direct protein-protein interactions (**Figure 6.2**). With numerous mechanistic roles in cancer biology and haematological contexts, an obvious role of PCBP2 in *TP53*-sAML is unknown. Therefore, in this chapter I sought to investigate the mechanisms by which PCBP2 is required to support *TP53*-sAML.

Aims for this section are as follows:

1. Investigate the impact of *PCBP2* perturbation on transcript stability
2. Determine the effect of *PCBP2* knockdown on splicing
3. Evaluate mechanistic hypotheses gained from the literature

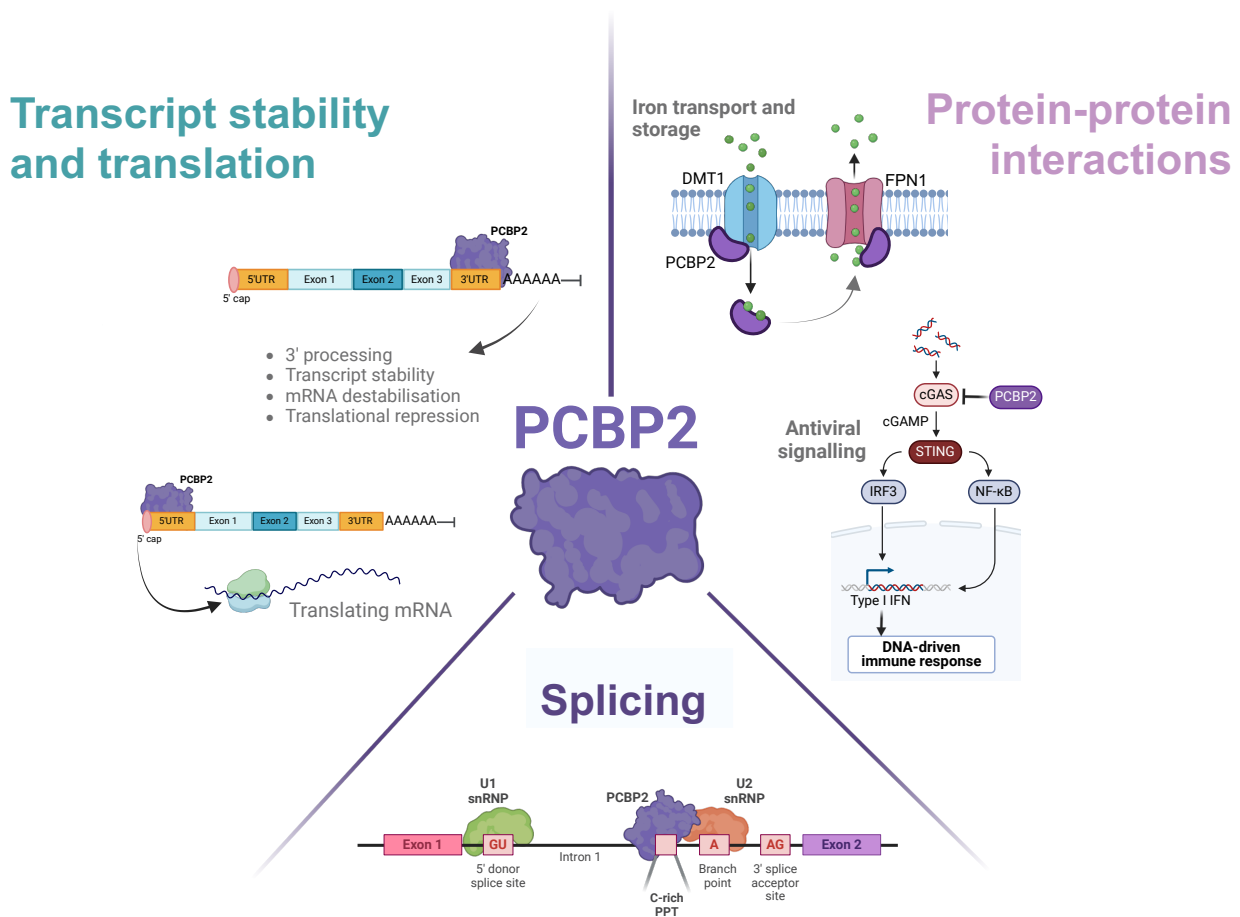


Figure 6.2. Mechanistic roles of PCBP2. As an RBP PCBP2 has a number of roles in post-transcriptional gene regulation, including regulation of transcript processing, splicing, stability and translation. Most frequently PCBP2 carries out these roles by binding C-rich RNA sequences within the 5' or 3'UTR or C-rich polypyrimidine tracts (PPT) upstream of splice acceptor sites. PCBP2 also regulates processes such as iron transport and antiviral innate immune signalling via protein-protein interactions. Illustration generated using BioRender.com

6.2 Results

6.2.1 Investigating the impact of PCBP2 on transcript stability and splicing in *TP53*-sAML

Having identified PCBP2 as a potential therapeutic vulnerability in *TP53*-sAML, I next investigated the molecular mechanisms by which PCBP2 is required for the survival and maintenance of *TP53*-sAML. To investigate the impact of PCBP2 on transcript stability and splicing, I performed bulk RNA-sequencing in the *JAK/TP53*-mutant HEL cell line following transduction with either scrambled control shRNA or PCBP2 KD #1 or KD #2 shRNA.

RNA-seq processing and quality control

To ensure total RNA for library preparation was collected at a timepoint which ensured maximal depletion of PCBP2 protein with minimal secondary effects, *PCBP2*-KD was compared at 72-hours and 96-hours post-transduction via western blot. Results showed increased knockdown at 96-hours, compared to 72-hours and therefore, this timepoint was used for total RNA extraction following lentiviral transduction (**Figure 6.3**).

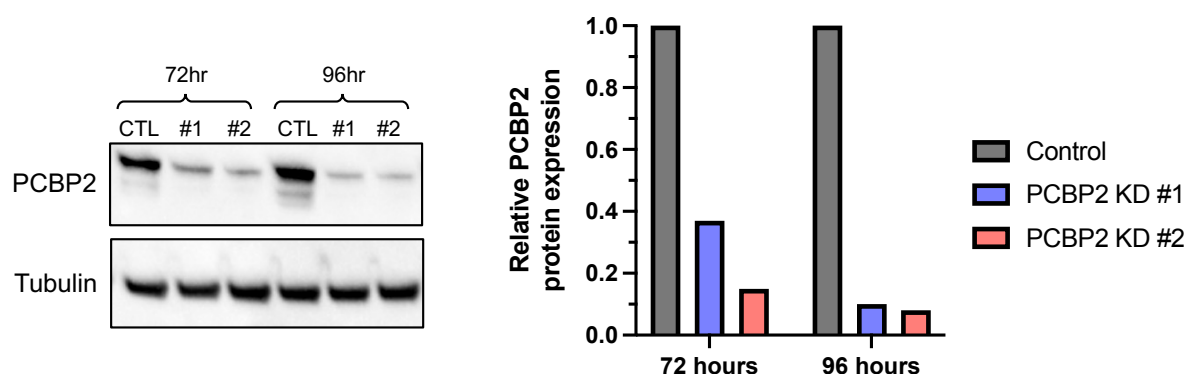


Figure 6.3. Comparing PCBP2 knockdown in HEL cells at 72-hours and 96-hours post-transduction. All samples were normalised to endogenous Tubulin expression. Per timepoint, normalised PCBP2 expression was then calculated relative to Control samples. Data represents one independent experiment. HEL cells were transduced with lentiviral vectors at MOI 5.

Details regarding library preparation, sequencing and analysis are detailed in Section 2.1.13 and Section 2.3.6, respectively. RNA-seq pre-processing and baseline quality

control were kindly performed by Antonio Rodriguez-Romera. Four technical replicates were included per shRNA construct and all samples passed baseline quality control for further downstream analysis. The average sequencing depth per sample was 60-million reads, sufficient for subsequent analysis of alternative splicing events.

Unsupervised principal component analysis (PCA) using the most variable genes from the normalised count matrix was applied to assess the relationship between samples based on their overall gene expression profiles. Here, samples clustered by experimental condition with the majority of variation accounted for in principle components 1 and 2 (PC1 and PC2; **Figure 6.4A**). Normalised gene counts were also used to verify *PCBP2* depletion in appropriate samples (**Figure 6.4B**).

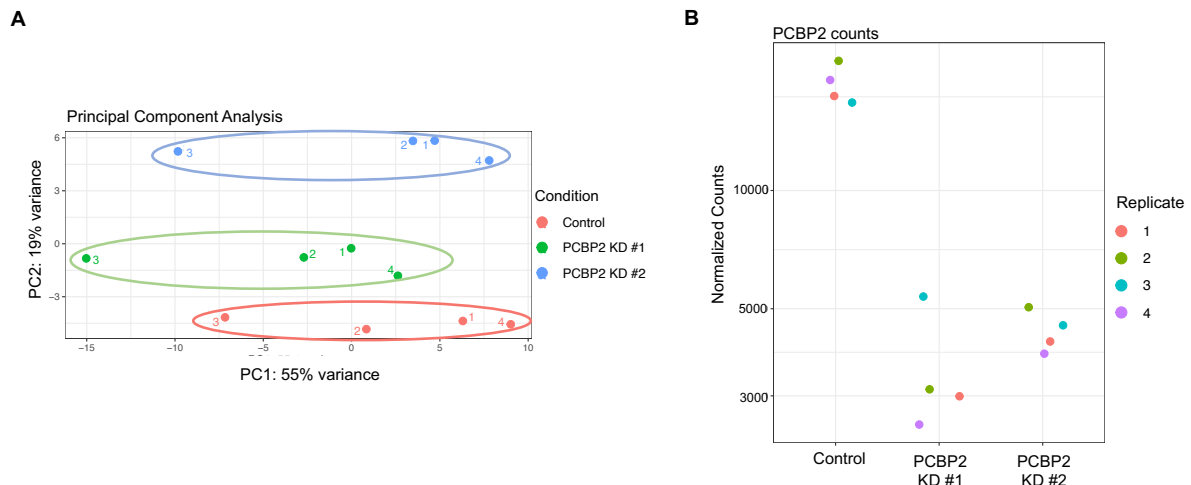


Figure 6.4. Principal component analysis (PCA) and *PCBP2* counts. (A) PCA using the most highly variable genes from a normalised count matrix. Control clusters are illustrated with red ovals, PCBP2 KD #1 samples with green and PCBP2 KD #2 with blue. **(B)** Normalised *PCBP2* counts in Control, PCBP2 KD #1 and PCBP2 KD #2 conditions. Each data point represents an independent technical replicate.

Investigating the effect of *PCBP2* depletion on differential gene expression

DESeq2 was used to identify differentially expressed genes (DEG) by applying Wald-test pairwise comparisons¹⁵⁵. I applied the DESeq2 differential gene expression pipeline to PCBP2 KD #1 (KD1) and PCBP2 KD #2 (KD2) shRNA compared to control independently, rather than combined (KDcomb). This was to reduce the inclusion of genes not identified to be differentially expressed in either single comparison or specific to only one shRNA-

PCBP2 construct, as illustrated in **Figure 6.5A**. Differentially expressed genes were defined as $\log_2\text{FoldChange} > 0.5$ with an adjusted p -value < 0.05 . DESeq2 analysis identified 213 DEG between KD1 and Control, and 417 DEG when comparing KD2 and Control. Of these, 76 genes were commonly dysregulated between KD1 and KD2 (**Figure 6.5B-C; Appendix 5**). This overlap in differentially expressed genes between KD1 and KD2 was statistically significant when determined by Fisher's exact test (p -value = $2.2e^{-16}$).

Identifying commonly dysregulated genes across two DESeq2 comparisons means some DEG may be under the cut-off for inclusion in one comparison and therefore be missed, despite significant expression in the other comparison. For example, interesting candidates may have a p -value of 0.049 in one comparison and 0.051 in the other, overall leading to them being excluded from analysis. To ensure no genes were excluded from being identified as commonly dysregulated because they missed the cut-off for significance in one KD comparison by a small margin, I investigated how lowering the p -value for one of the two KD comparisons changes gene inclusion. Analysis showed a more lenient p -value of 0.1 in KD1 comparisons led to inclusion of an additional 8 commonly dysregulated genes. Similarly, in KD2 comparisons, the same approach led to inclusion of an additional 11 genes. Individual investigation of each excluded gene concluded this analysis did not need to be adjusted for their inclusion as none could be determined as potential genes of interest. For example, genes lacked sufficient supporting literature to be interesting candidate genes or presented highly tissue-specific expression unrelated to the haematopoietic compartment. Hence, while an important consideration for maximising identification of candidate genes, on this occasion a p -value of 0.05 was maintained for both comparisons to identify commonly dysregulated genes.

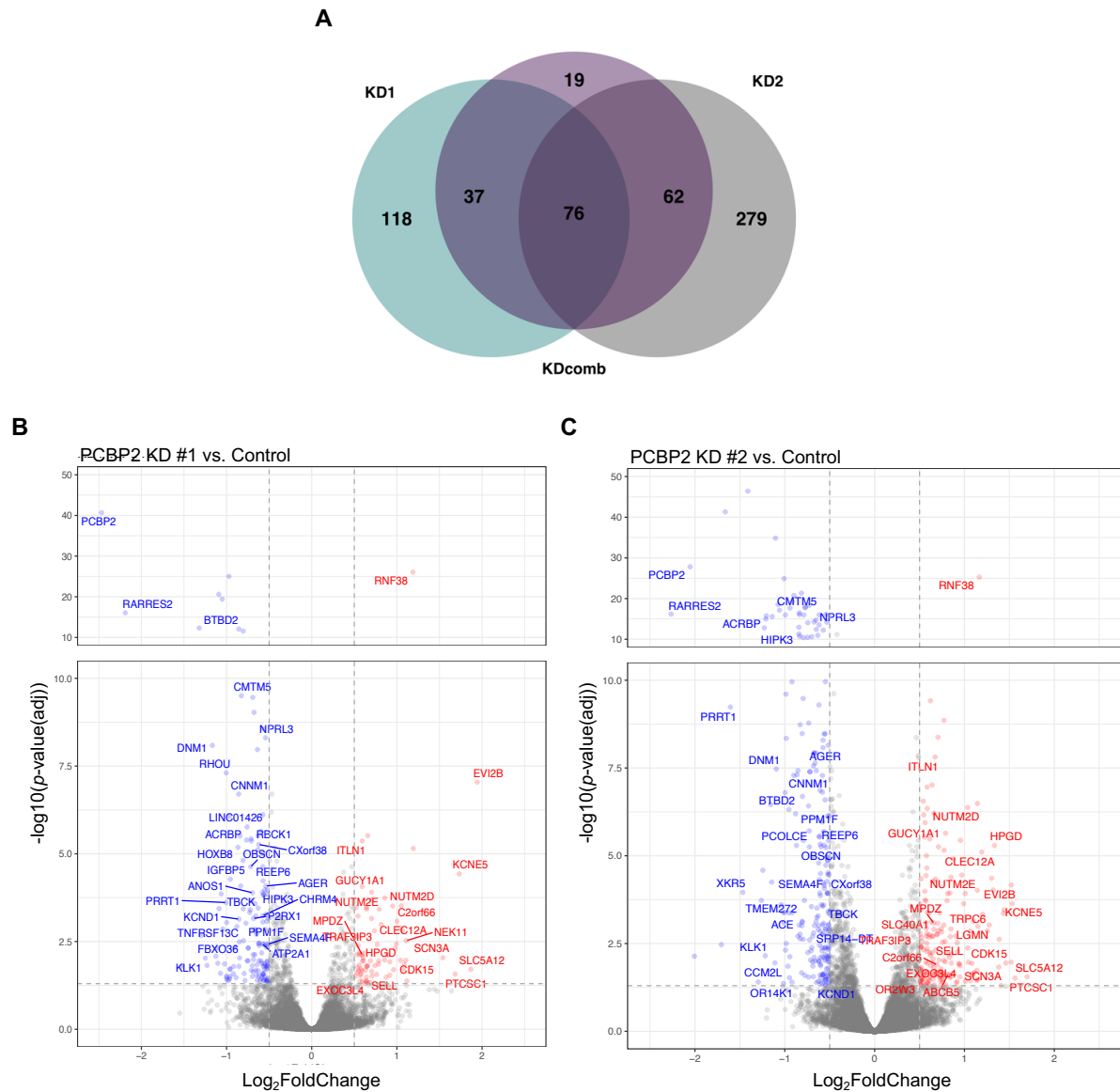


Figure 6.5. Identifying differentially expressed genes using DESeq2. (A). Venn diagram comparing differentially expressed genes (DEG; $\log_2\text{FoldChange} > 0.5$ and < -0.5 , adjusted p -value < 0.05) identified when running DESeq2 analysis combining *PCBP2*-knockdown conditions (KDcomb) or running *PCBP2* KD #1 (KD1) and *PCBP2* KD #2 (KD2) comparisons independently. Overlap between common KD1 and KD2 DEG was statistically significant as determined by Fisher's exact test p -value = $2.2e^{-16}$. **(B-C).** Volcano plots of common differentially expressed genes. Red and blue labels represent differentially upregulated and downregulated genes, respectively. DEG determined by Wald test pair-wise comparison and Benjamini-Hochberg correction for multiple-testing using the DESeq2 pipeline.

To explore the impact of *PCBP2* depletion on transcriptional profiles, gene set enrichment analysis (GSEA) was conducted. Top dysregulated pathways included hypoxia and cholesterol homeostasis pathways, both of which were downregulated in *PCBP2*-depleted samples (**Figure 6.6**). However, no other Hallmark pathways were commonly dysregulated in both KD1 and KD2 samples to a significant degree ($\text{FDR} < 0.25$).

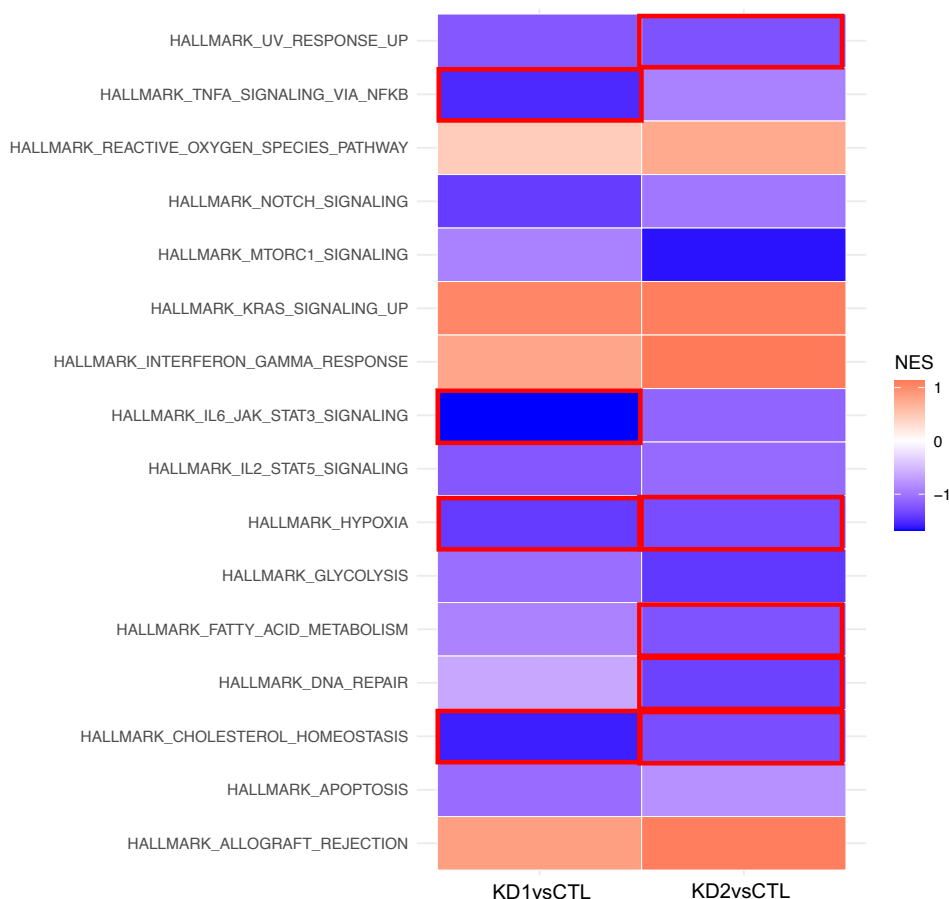


Figure 6.6. Gene set enrichment analysis (GSEA) from RNA-seq data. GSEA of selected Hallmark pathways (h.all.v2023.1) displaying normalised enrichment scores (NES) for both PCBP2 KD #1 (KD1) and PCBP2 KD #2 (KD2) vs. control comparisons. Red boxes represent significantly dysregulated pathways where adjusted p -value < 0.25.

I then investigated whether *PCBP2*-KD effects the expression of interferon stimulated genes (ISG), given the role of *PCBP2* in antiviral innate immune pathways^{225,267}. To do this, GSEA was run on a previously published list of 379 annotated ISG¹⁵⁸. Results showed no significant enrichment of ISG expression in *PCBP2*-KD cells (NES = -1.08, p -value 0.58; **Figure 6.7A**). A core set of ten ISG have been identified to be universally upregulated across cell types in response to type I interferon³⁵⁸. I compared expression of these core ISG across control and KD conditions. Results corroborated GSEA analysis showing no difference in core ISG expression between *PCBP2*-KD and control samples (**Figure 6.7B**). This is not completely unexpected as previous reports show *PCBP2* regulation of cGAS-STING and RIG-I-MAVS antiviral signalling pathways requires exposure to dsDNA or dsRNA, respectively^{225,267}.

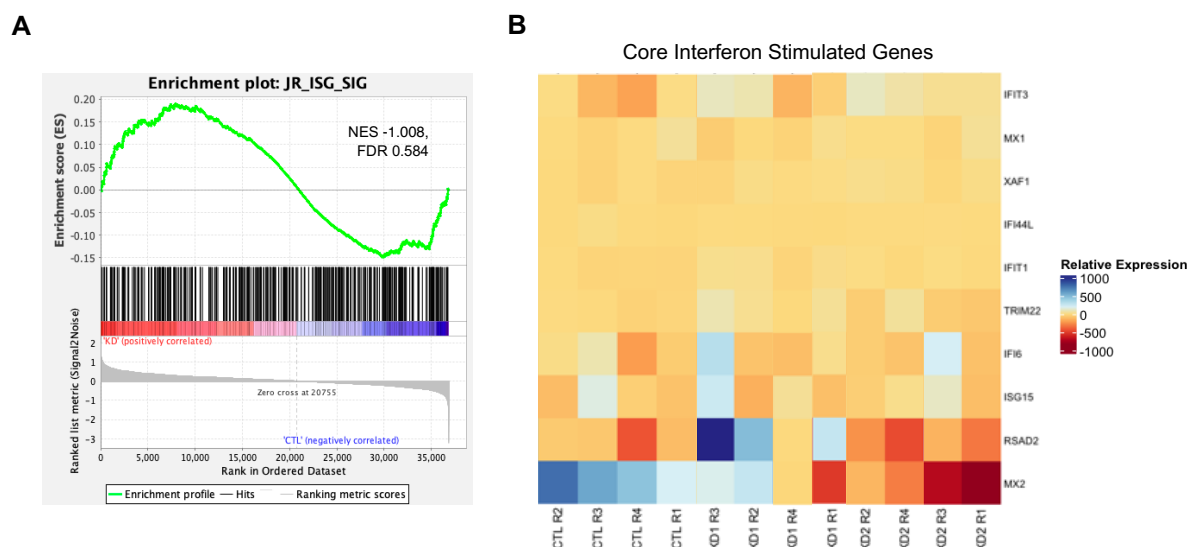


Figure 6.7. Investigating the requirement of PCBP2 in regulation of interferon stimulated genes. (A) GSEA for 379 annotated interferon stimulated genes (ISG)¹⁵⁸. Normalised enrichment score for combined PCBP2 KD #1 and PCBP2 KD #2 vs control comparison. **(B)** Expression of core ISG universally expressed across cell-types in PCBP2 KD #1 or KD #2 samples compared to controls. Core ISG were defined by Rigby *et al*³⁵⁸.

Next, to determine whether *PCBP2* perturbation had any impact of the transcription of genes essential for cancer cell survival, the requirement of dysregulated genes was assessed using the DepMap database^{140,141}. Surprisingly, none of the 76 differentially expressed genes identified were considered essential for cancer cell survival in model HEL and SET-2 cell lines (**Figure 6.8**). This finding is interesting given *PCBP2* is known to stabilise essential genes such as *c-Myc*, in other cancer cell line models^{230,309}. Having shown *PCBP2* does not regulate stability of transcripts essential for *JAK2/TP53*-mutant cancer cell survival, I moved onto investigate whether *PCBP2* regulates alternative PTGR processes, such as splicing.

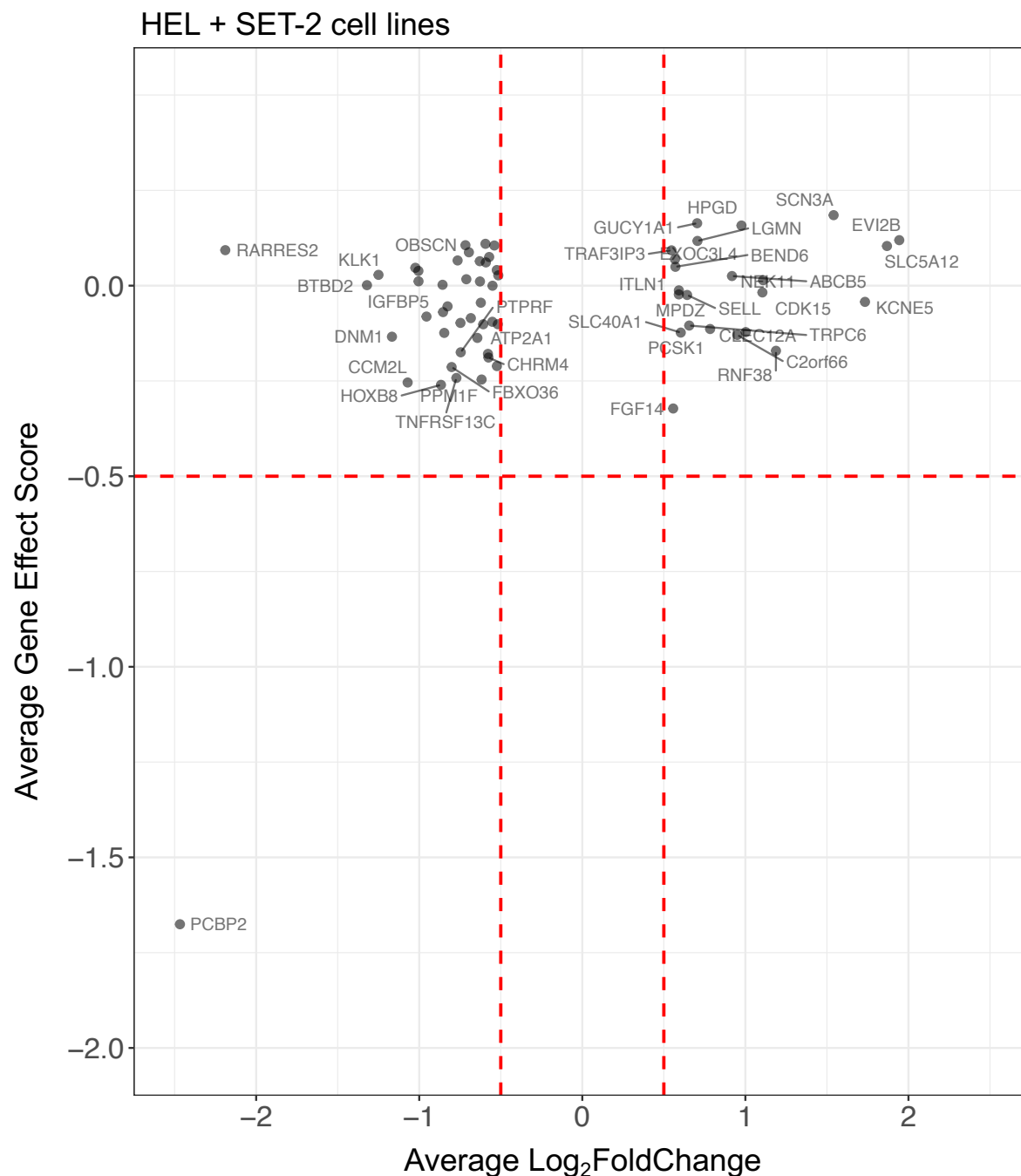


Figure 6.8. Investigating the essentiality of differentially expressed genes in *JAK2/TP53*-mutant cell line models. Average log₂FoldChange from PCBP2 KD #1 and PCBP2 KD #2 DESeq2 comparisons were calculated. Data points represent differentially expressed genes commonly dysregulated in both comparisons. Average gene effect scores from the DepMap database (24Q2)¹⁴¹ were determined for differentially expressed genes in HEL, HEL92.1.7 and SET-2 cell lines. Vertical red lines represent a Log₂FoldChange > 0.5 and < -0.5 and horizontal lines represent threshold below which genes are considered essential (< -0.5).

Investigating effect of *PCBP2* depletion on alternative splicing

As described in this Chapter's introduction, PCBP2 has known roles in alternative splicing. Therefore, I next investigated whether alternative splice events (ASE) are observed following *PCBP2* knockdown in the HEL cell line. Alternative splice analysis was computed using MARVEL¹⁶⁶, a package originally developed for splice analysis at single-cell resolution. Here, the MARVEL package was applied to the same bulk RNA-Seq dataset described above. Pre-processing for identification of alternative splice events and differential splice analysis were kindly complete by Dr Sean Wen, previously Weatherall Institute of Molecular Medicine, University of Oxford. The MARVEL workflow determines the percentage spliced-in (PSI) value of alternative splicing events. PSI values are calculated as total reads supporting the alternative exon, divided by the total reads supporting both the alternative exon and constitutive exon¹⁶⁶. PSI values are then stratified into 7 discrete exon-level alternative splice events namely, skipped exon (SE), mutually exclusive exons (MXE), retained intron (RI), alternative 5' splice site (A5SS), alternative 3' splice site (A3SS), alternative first exon (AFE) and alternative last exon (ALE) (**Figure 6.9**).

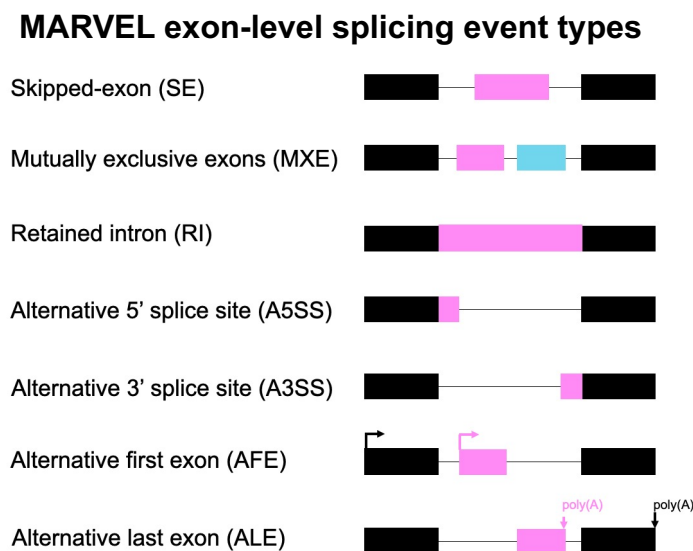


Figure 6.9. MARVEL exon-level splicing event types. The MARVEL package stratifies percentage spliced-in (PSI) values into 7 discrete exon-level splice events. Black boxes represent exons. Pink and blue boxes represent the alternative splicing events.

Similar to differential gene expression analysis, alternative splice analysis was run independently for KD1 and KD2 constructs. ASE were defined as $\Delta\text{PSI} > 5\%$ and $p\text{-value} < 0.05$. Results showed a greater effect of KD2 on splicing with 1649 differential events identified compared to Control, in contrast to the 954 events observed in the KD1 comparison (**Figure 6.10A**). Of these ASE, 509 showed significant overlap between KD1 and KD2 (Fisher's exact test, $p\text{-value} = 2.2e^{-16}$), with SE events being most frequent followed by RI and AFE events (**Figure 6.10B**). Furthermore, upregulated and downregulated ASE contributed equally to dysregulated splicing in both PCBP2 knockdown comparisons (**Figure 6.10C; Appendix 6**).

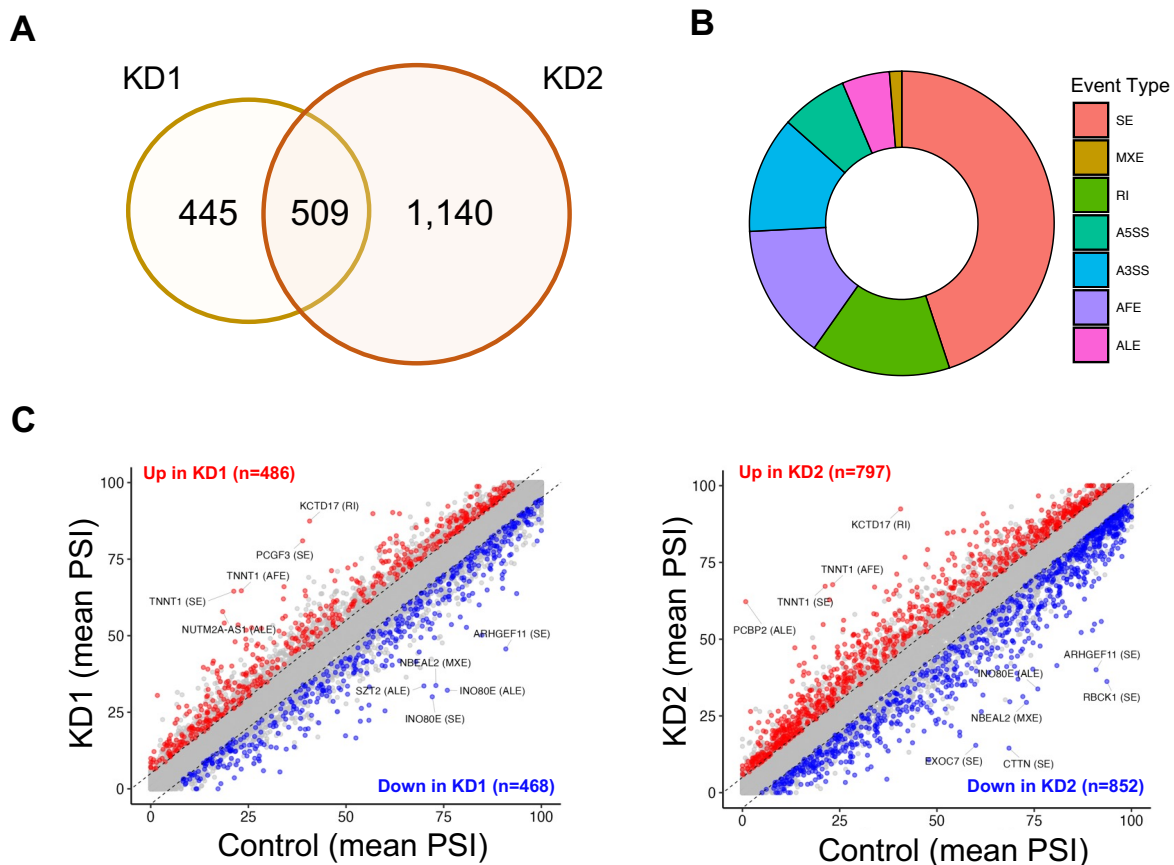


Figure 6.10. Identifying alternative splice events (ASE) in PCBP2-depleted cells. (A). Venn diagram identifying common ASE between PCBP2 KD #1 (KD1) and PCBP2 KD #2 (KD2) comparisons. ASE defined as delta-percentage spliced-in (PSI) $> 5\%$ and $p\text{-value} < 0.05$. Significant overlap of ASE was observed between KD1 and KD2 comparisons calculated by Fisher's exact test ($p\text{-value} = 2.2e^{-16}$). **(B)** Pie chart depicting the proportion of each different ASE occurring in common ASE. **(C)** Scatterplot comparing mean PSI values between KD constructs (KD1 or KD2, y-axis) and Control (x-axis). Events significantly upregulated and downregulated in KD constructs are highlighted in red and blue, respectively. The top 10 ASE per KD construct are annotated. MARVEL analysis and associated figures generated by Dr Sean Wen, previously Weatherall Institute of Molecular Medicine, University of Oxford.

Curiously, an ALE event in *PCBP2* appeared to be significantly enriched in *PCBP2*-KD samples. To investigate this ALE event in more detail, reads were aligned to the reference genome (hg38) and ASE visualised using the UCSC genome browser³⁵⁹. This showed the inclusion of an alternative last exon within the *PCBP2* region defined as intron 13 (**Figure 6.11**). This ASE was highly prevalent in *PCBP2* KD #2 samples ($\Delta\text{PSI}=61.3$), and present to a lesser extent in *PCBP2* KD #1 samples ($\Delta\text{PSI} = 11.3$). Increased prevalence of this ALE event in *PCBP2* KD #2, compared to *PCBP2* KD #1 samples, may be explained as a by-product of *PCBP2* KD #2 shRNA targeting exon 14 downstream of the ALE. Although, how targeting with *PCBP2* KD #1 shRNA could also generate this ASE to a certain degree requires explanation. Whether this event occurs and influences the limited KD observed in primary cells also requires further consideration. However, given this event is far more prevalent in *PCBP2* KD #2 samples, how this ASE could limit KD to a similar degree in both *PCBP2* KD #1 and *PCBP2* KD #2 primary samples needs to be investigated. Overall, while an interesting occurrence, it is clear there is significant depletion of all *PCBP2* transcripts in *PCBP2* KD #1 and *PCBP2* KD #2 HEL cell line samples, corroborating the strong knockdown observed by RT-qPCR.

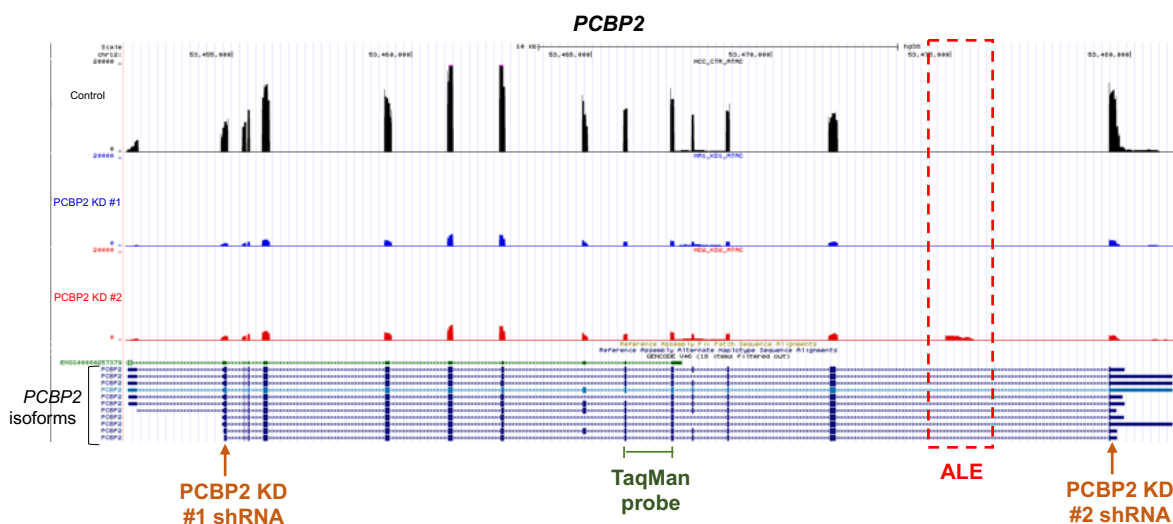


Figure 6.11. Visualisation of *PCBP2* transcripts alternative last exon (ALE) event. Reads were aligned to the human genome (hg38) and visualised using the UCSC Genome Browser³⁵⁹. A prominent ALE event frequently occurred in *PCBP2* intron 13 in *PCBP2* KD #2. This ALE was also present in *PCBP2* KD #1 samples to a lesser extent, however, due to scaling could not be visualised in this image. *PCBP2* ALE event highlighted by red dashed box. Regions bound by TaqMan probe and shRNA are shown. Peaks are all to the same scale for visual comparison.

Next, gene expression-splicing dynamics were investigated by integrating ASE and gene expression analysis. Gene expression-splicing relationships can either be coordinated, opposing or isoform switching. In coordinated relationships, mean PSI and gene expression values move in the same direction within a cell population, or condition. Opposing relationships are driven by changes in PSI and gene expression values in opposite directions, while isoform switching occurs when PSI distributions are significantly different but gene expression values remain consistent between populations. The 509 common ASE identified in KD1 and KD2 (**Figure 6.10A**), occurred in 412 independent genes, highlighting the potential for certain genes to be targeted for multiple ASE. Results showed minimum coordinated and opposing events, with the majority (99%) of ASE being isoform switching events (**Figure 6.12A**). The only coordinated event observed was within *RBCK1*, while 3 opposing events were identified including *PTPRF*, *MPDZ* and *PCBP2*. Isoform switching events were identified in genes known to contribute to myeloid malignancies, including *MEIS1*, *RUNX1* and *STAT3* (**Figure 6.12B**). A complete list of gene-expression-splicing relationships identified can be found in **Appendix 7**. Nonsense-mediated decay (NMD) predictions reflect the majority of ASE being isoform switching events, as minimal NMD events were identified (<5 per KD construct). Finally, gene ontology analysis did not identify any enriched pathways among the 412 differentially spliced genes.

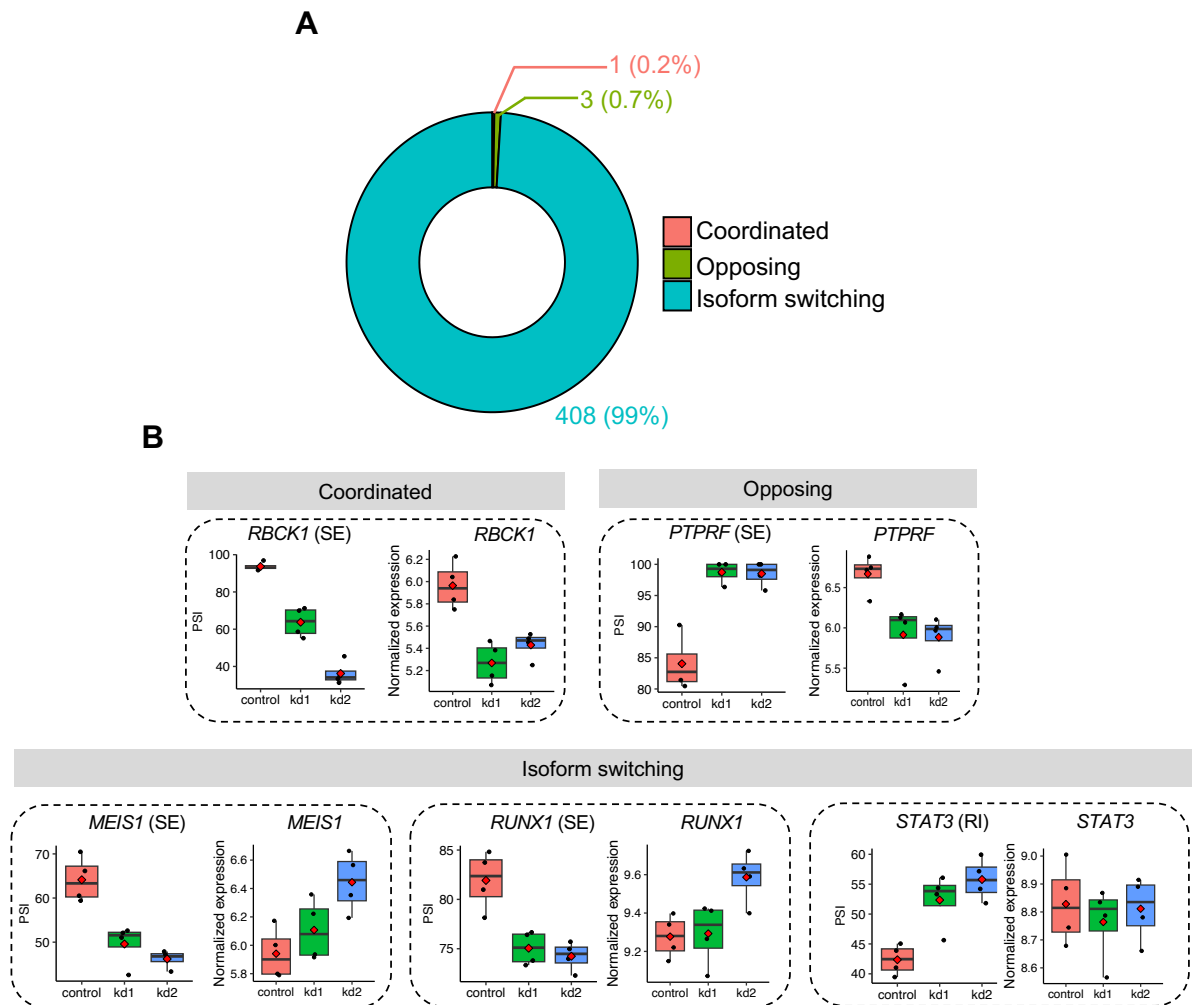


Figure 6.12. Investigating gene expression-splicing dynamics. (A) Pie-chart illustrating the proportion of each gene expression-splicing relationship. In coordinated relationships, mean gene expression and percentage-spliced in (PSI) values move in the same direction within a condition. In opposing relationships PSI and gene expression move in the opposite direction and isoform switching occurs when a significant PSI event occurs but gene expression remains consistent between comparisons. **(B)** Examples of coordinated, opposing and isoform switching dynamics in PCBP2 KD #1 and PCBP2 KD #2 conditions compared to control. Significant ASE defined as $\Delta\text{PSI} > 5\%$ and $p\text{-value} < 0.05$, DEG defined as $\log_2\text{FoldChange} > 0.5$ and < -0.5 and $p\text{-value} < 0.05$ (in both KD1 and KD2). RI, retained intron; SE, skipped exon. *MARVEL analysis and associated figures generated by Dr Sean Wen, previously Weatherall Institute of Molecular Medicine, University of Oxford.*

Using the UCSC genome browser, I visualised the SE and A3SS events occurring in *RUNX1* and *STAT3*, respectively, given their importance in haematological malignancies. Highly conserved cytosine-rich PPT sequences upstream of 3' splice acceptor site are the *cis*-regulatory regions reported to be responsible for PCBP2-mediated splicing³⁵⁴. Furthermore, *Runx1* exon 6 SE events have been observed in PCBP2-null mouse models²³⁷. This ASE was also observed in PCBP2-KD HEL cells, with *RUNX1* exon 6 more

frequently skipped in PCBP2 KD #1 and PCBP2 KD #2 compared to control samples, with Δ PSI -6.86 and -7.68, respectively (**Figure 6.13A**). Thus, identification of previously reported ASE mediated by PCBP2 validates the integrity of this RNA-seq dataset.

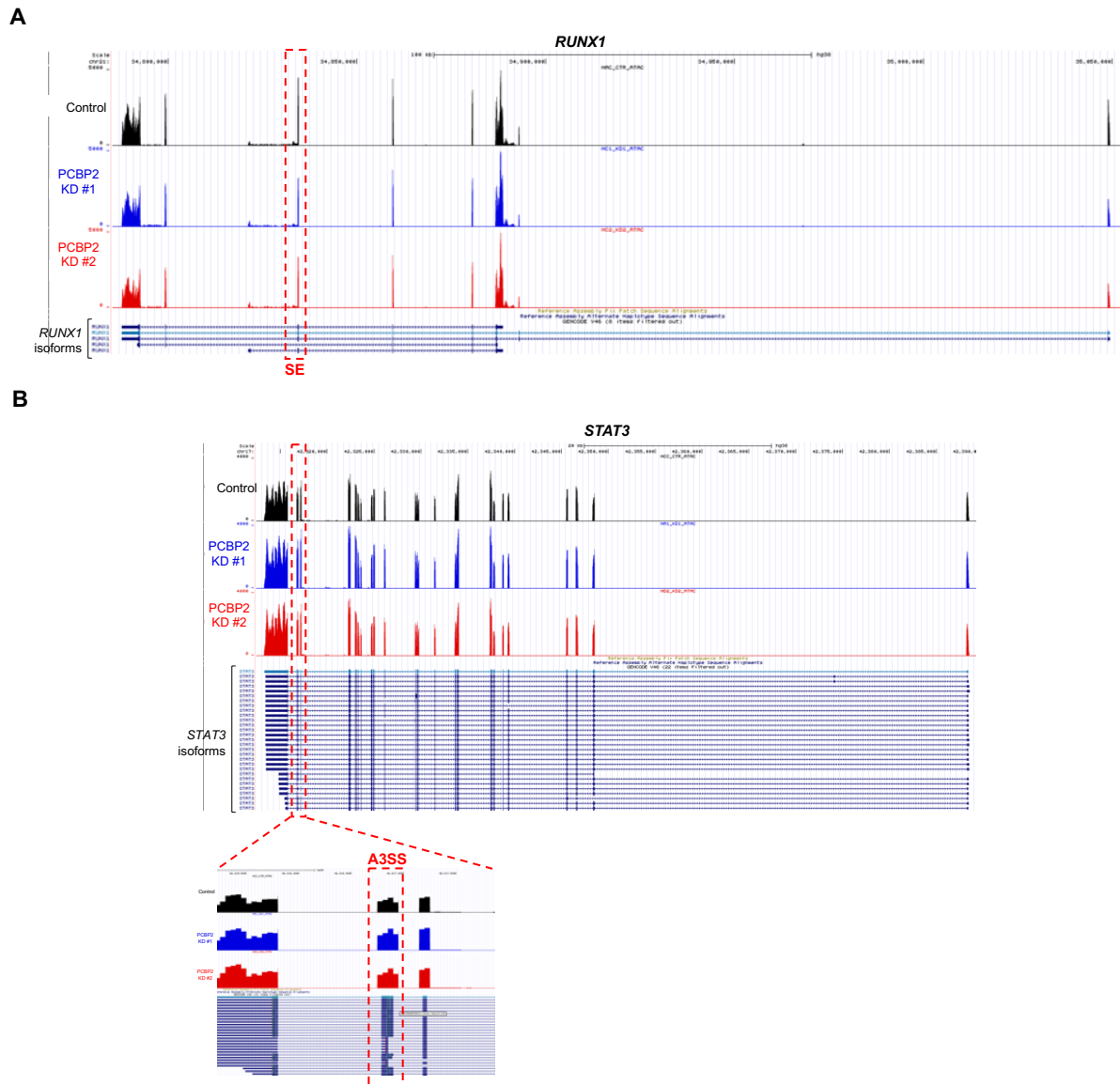


Figure 6.13. Visualisation of *RUNX1* skipped exon and *STAT3* alternative 3' splice site events. Reads were aligned to the human genome (hg38) and visualised using the UCSC Genome Browser for **(A)** *RUNX1* skipped exon event (SE) and **(B)** *STAT3* alternative 3' splice site (A3SS) event. Alternative splice events are highlighted by red dashed boxes. *STAT3* A3SS event is enlarged to show alternative isoforms. Peaks were scaled for comparison.

Alternative splicing of *STAT3* is also of particularly interest since JAK-STAT signalling drives the myeloproliferative disease preceding *TP53*-sAML transformation. The identified *STAT3* A3SS event appears to alter the ratio of two common *STAT3* isoforms,

STAT3α and *STAT3β* (**Figure 6.13B**). *STAT3α* is more often associated with proliferation and oncogenic activities, while *STAT3β* associates with differentiation pathways. Thus *STAT3α* and *STAT3β* are regarded to work as tumour promoters and suppressors, respectively³⁶⁰. *PCBP2*-KD leads to increased usage of *STAT3* exon 23 upstream 3' splice acceptor site, increasing the frequency of *STAT3β* isoforms (Δ PSI KD1 = 5.48, Δ PSI KD2 = 7.2, p -value < 0.05). Hence, *PCBP2*-KD may influence survival and maintenance of *TP53*-sAML through regulation of *STAT3* signalling pathways and this ASE requires further investigation to uncover its importance in *TP53*-sAML.

Finally, the requirement of alternative spliced genes in cancer cell line models was assessed using the DepMap database^{140,141}. Differential splicing can often produce non-functional transcripts or transcripts with alternative or opposing functions. Considering this, I hypothesised that alternatively splicing may lead to an increased proportion of non-functional/alternative function transcripts from genes known to be essential for cancer cell survival. Such transcripts may be responsible for driving the reduced proliferation and increased apoptosis observed in *PCBP2*-KD cell line models (Chapter 4). A number of alternative spliced genes were identified to be essential for cancer cell survival, including genes relevant to cancer biology (*BCL2L1*, *RUNX1*, *CDK2*), splicing (*SF1*, *PUF60*) and candidate RBP (*HNRNP M*) identified in Chapter 3 of this project (**Figure 6.14**).

BCL2L1 is an attractive candidate given alternative splicing leads to two isoforms with different roles in apoptosis. Alternative usage of a 5' splice site within exon 2 of *BCL2L1* encodes either anti-apoptotic, BCL-X_L, or pro-apoptotic, BCL-X_S. Hence, dysregulated ratios of BCL-X_L and BCL-X_S isoforms could enhance or dampen apoptotic phenotypes. However, the event observed in this dataset is a distinct A3SS within exon 2 and upon examination in the UCSC genome browser, reflects a common splice alteration, leading to the addition of 18 extra base pairs within the *BCL2L1* 5'UTR. This A3SS event is distinct from the A5SS event responsible for alternative BCL-X_L and BCL-X_S isoform switching and overall, was considered unlikely to have any functional implications.



Figure 6.14. Investigating the essentiality of alternatively spliced genes in *JAK2/TP53*-mutant cancer cell line models. Average gene effect score calculated for HEL, HEL92.1.7 and SET-2 cell lines and average percentage spliced-in (PSI) calculated using values from KD1 and KD2 vs Control comparisons. Alternatively spliced genes identified as $\Delta\text{PSI} > 0.5$ and < -0.5 (vertical red lines) and $p\text{-value} < 0.05$. Horizontal red lines indicate gene scores < -0.5 , below which genes are considered essential. A3SS, alternative 3' splice site; A5SS, alternative 5' splice site; AFE, alternative first exon; ALE, alternative last exon, MXE, mutually exclusive exon; RI, retained intron; SE, skipped exon.

Other attractive candidate genes within this alternative splicing dataset still require further investigation, including *HNRNP C*, an RBP which was recently reported to be indispensable for AML³⁶¹. In conclusion, through this splice analysis I have identified a number of splice events in genes known to contribute to myeloid malignancies and an

additional cohort of putative candidates by which PCBP2 could mediate *TP53*-sAML survival and proliferation.

6.2.2 Investigating the requirement of PCBP2 in *BCL2L1* regulation

Despite concluding alternative splicing of *BCL2L1* has no important functional implications, it does corroborate data showing the potential for PCBP2 to bind *BCL2L1* transcripts⁹⁶. Interestingly, erythroid and megakaryocytic leukaemia are highly dependent on anti-apoptotic BCL-XL, and sensitive to treatment with the BCL-XL inhibitor, A1331853³⁶². In contrast, erythroid/megakaryocytic leukaemia are not dependent on pro-apoptotic BCL-2 and are resistant to BCL-2 inhibition. Furthermore, primary patient erythroid/megakaryocytic AML samples carrying *TP53*-mutations showed persistent sensitivity to BCL-XL inhibition³⁶².

TP53-mutant multi-hit HSPC in *TP53*-sAML display an erythroid-associated transcriptional phenotype and a propensity to differentiate towards the erythroid lineage²⁰⁷. Therefore, I investigated the effect of PCBP2 on erythroleukemia sensitivity to BCL-XL, hypothesising PCBP2 may regulate *BCL2L1* on a translational level. This work was complete in collaboration with Dr Shady Awad, postdoctoral researcher at the Weatherall Institute for Molecular Medicine, University of Oxford.

HEL cells transduced with either scrambled control or PCBP2 KD #2 shRNA were seeded at 1.25×10^6 /50uL media and treated with either A-1331853, Navitoclax or Venetoclax. A1331853 is a BCL-XL-specific inhibitor and Venetoclax a BCL-2-specific inhibitor, while Navitoclax targets both pro-apoptotic proteins. Drug concentrations were as recommended by the manufacturers with 10-fold dilutions from 1,000nM to 1nM for Navitoclax and A1331853, and 10,000nM to 1nM for Venetoclax. Seventy-two hours after treatment, viability was determined by CellTitreGlo assay. Concordant with the literature, HEL cells were resistant to Venetoclax and showed the highest sensitivity to A1331853 (**Figure 6.15A**). However, no significant difference was observed between Control and PCBP2-KD cells for any treatment, refuting any involvement of PCBP2 in the sensitivity of *TP53*-mutant erythroleukemia to BCL-XL inhibition.

In parallel, I also investigated the effect of PCBP2 on BCL-XL translation, given PCBP2 is known to bind *BCL2L1* 3' UTR, a mechanism which has been implication in translational regulation⁹⁶. Results showed no change in BCL-XL expression at the protein level in the absence of PCBP2 (**Figure 6.15B**), suggesting after rigorous testing at the transcriptional, splicing and translational level, PCBP2 unlikely plays a role in the PTGR of *BCL2L1* expression in *TP53*-sAML.

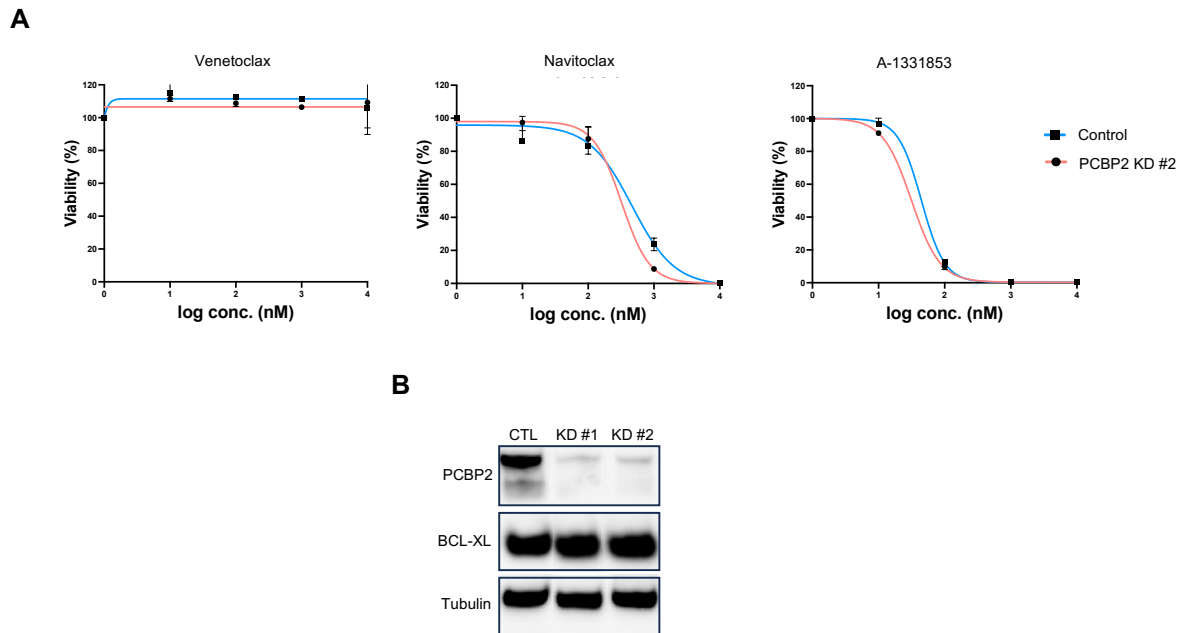


Figure 6.15. Dependency of *JAK2/TP53*-mutant erythroleukemia HEL cells on anti-apoptotic proteins following *PCBP2*-KD. (A) Viability of HEL cells transduced with either Control and *PCBP2* KD #2 shRNA, prior to treatment with BCL-2 specific inhibitor (Venetoclax), BCL-2/BCL-XL inhibitor Navitoclax) or BCL-2 specific inhibitor (A-1331853). Experiment was run in triplicate, representing one independent experiment. Drug concentrations were diluted 10-fold from 10,000nM to 1nM for Venetoclax and from 1,000nM to 1nM for Navitoclax and A1331853. **(B)** Western blot assaying BCL-XL expression in the presence or absence of *PCBP2* expression. Image representative of one independent experiment. *Viability assays and western blots were complete by Dr Shady Awad, postdoctoral researcher at the Weatherall Institute of Molecular Medicine, with input on experimental design and interpretation by myself.*

6.2.3 Investigation the potential role of PCBP2 in R-loop biology

A number of potential mechanisms by which PCBP2 may mediate *TP53*-sAML came from review of the literature. For example, SFPQ, one of the twelve candidate RBP identified in this DPhil project, was shown to prevent nascent transcripts forming dsRNA: DNA hybrids³⁶³. In this regard, PCBP2 has also been identified to associate with dsRNA: DNA

hybrids, suggesting a role of PCBP2 in R-loop prevention or resolution similar to SFPQ^{364,365}. R-loops are dsRNA: DNA hybrid structures which form during transcription when nascent RNA transcripts anneal to the template DNA strand and displace the non-template ssDNA strand. Formation of R-loops can be non-pathological, for example by forming intermediates during immunoglobulin class switching recombination³⁶⁶. However, aberrant R-loop formation can lead to replicative stress, DNA breaks and genome instability³⁶⁷. I therefore hypothesised PCBP2 associates with dsRNA: DNA hybrids to resolve excessive R-loop formation. Excessive R-loop formation can exacerbate DNA damage and heighten sensitivity of malignant cells to DNA-damaging agents. This is increasingly relevant in *TP53*-sAML where CNA and genomic instability are characteristic of leukaemic transformation.

To investigate this hypothesis, I collaborated with Dr Chiara Beghè, a post-doctoral researcher in Professor Natalie Gromak's laboratory at the Sir William Dunn School of Pathology, University of Oxford, where they are experts in R-loop biology. Following *PCBP2*-KD in the HEL cell line using *PCBP2*-KD #1, *PCBP2*-KD #2 shRNA or scrambled control, the presence of dsRNA: DNA hybrids was assessed using the highly-specific S9.6 antibody via slot blot.

Probing for dsRNA: DNA hybrids using the S9.6 antibody via slot blot requires a number of controls. RNase H cleaves R-loops, hence lack of signal in RNase H treated (+ RNase H) lanes confirms the sensitivity and specificity of the S9.6 antibody for dsRNA: DNA hybrids. RNase H treatment requires a 150-minute incubation at 37°C, and to control for these conditions, an additional RNase H-negative (- RNase H) condition is included. Untreated samples are not exposed to this incubation and therefore, discrepancies between untreated and RNase H-negative conditions can be expected. Overall, results showed highly specific detection of dsRNA: DNA hybrids across all untreated samples, however, limited differences in dsRNA: DNA hybrids between *PCBP2*-KD samples and controls (**Figure 6.16**). This suggests PCBP2 does not play prominent role in the resolution of R-loops in *TP53*-sAML.

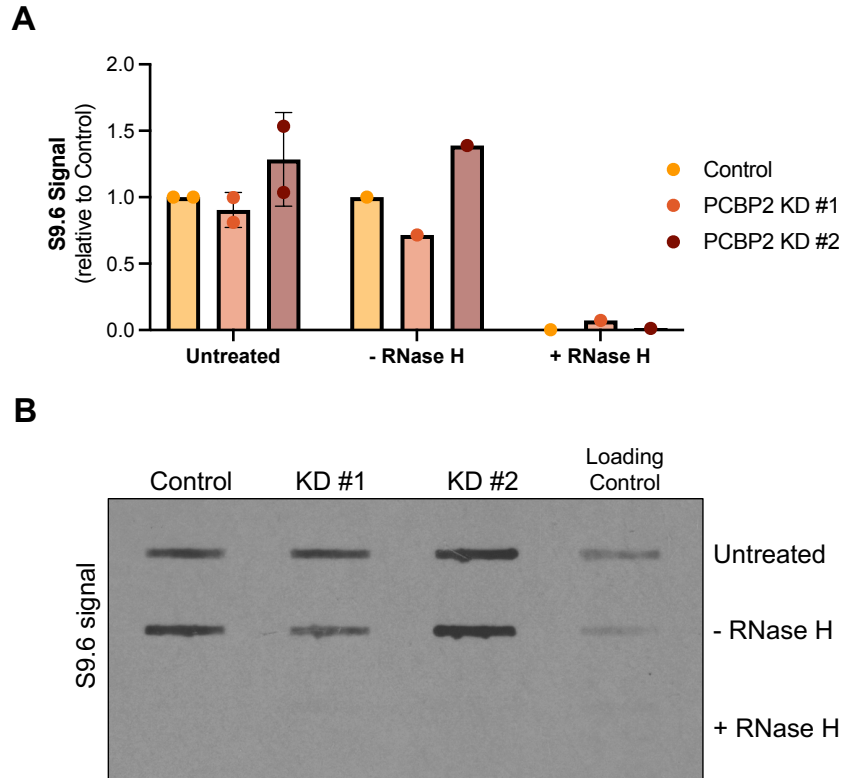


Figure 6.16. Investigating R-loop formation following *PCBP2* depletion using dsRNA: DNA specific S9.6 antibody. (A) Relative S9.6 signal in Control conditions compared to PCBP2 KD #1 or PCBP2 KD #2. Untreated samples were not exposed to RNase H or the 150-minute incubation at 37°C. RNase H-negative (-RNase H) samples underwent 150-minute 37°C incubation without RNase H. **(B)** Representative slot blot membrane. DNA loading controls were 1.5ug, 1ug and 0.5ug from top to bottom of the membrane. Two independent slot blots were run for untreated samples providing technical replicates. *Experiments were performed in collaboration with Dr Chiara Beghè, Sir William Dunn School of Pathology, University of Oxford.*

In parallel, I also investigated the effect of PCBP2 on DNA-damage since accumulation of R-loops is known to drive transcription-replication conflicts and consequently, double strand breaks (DSB)³⁶⁷. Upon formation of DSB, the nucleosome histone, H2AX, is phosphorylated by DNA-damage response kinases, ATR or ATM³⁶⁸. Phosphorylated H2AX, γ H2AX, is the first step in the DSB DNA-damage response and therefore γ H2AX loci can be assayed for determination of DSB. I compared the formation of γ H2AX foci in *PCBP2*-KD #2 cells to scrambled control conditions. Interestingly, γ H2AX foci per nucleus were significantly increased in PCBP2 KD #2 HEL cells compared to those transduced with the scrambled control (**Figure 6.17**). In agreement with an increase in DSB, GSEA also identified a DNA damage response gene set to be significantly downregulated following *PCBP2*-KD (**Figure 6.6**). Overall, these results suggest PCBP2 regulates DNA-damage via

a mechanism independent of R-loop resolution. In this experiment, significant differences in γ H2AX foci between KD and control conditions may be driven by the large number of cells imaged for analysis. Thus, this experiment requires repeating to provide independent replicates and further investigation will be necessary to determine the specific role of PCBP2 in response to DNA-damage, such a PCBP2-mediated regulation or interaction with specific DNA-repair proteins.

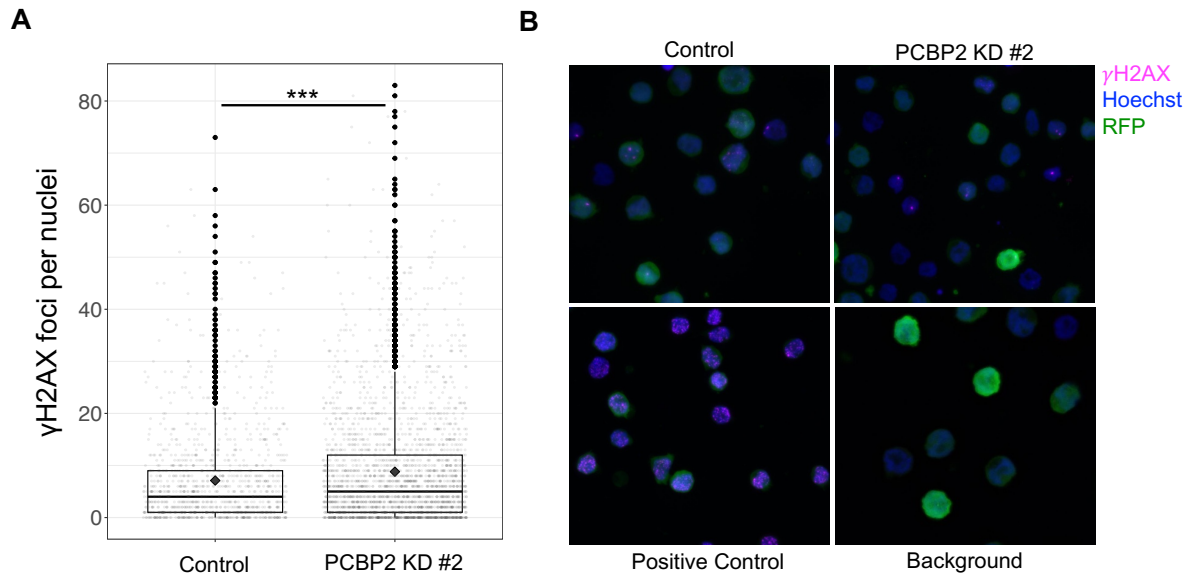


Figure 6.17. Detecting double-strand breaks via γ H2AX staining in PCBP2-KD HEL cells. **(A)** Number of γ H2AX foci per nucleus. All conditions included ≥ 329 nuclei. Data represents one independent experiment using only PCBP2 KD #2 or scrambled control shRNA. Significance determined by one-way ANOVA where *** $p < 0.0001$. **(B)** Representative γ H2AX immunofluorescence images. RFP identified cells transduced with either scrambled-control or PCBP2 KD #2 shRNA. Hoechst stains nuclear DNA and γ H2AX staining identifies double-strand breaks. Background image contains no γ H2AX antibody and thus no signal. *This experiment was performed in collaboration with Antonio Rodriguez-Romera, DPhil student at the Weatherall Institute of Molecular Medicine, University of Oxford.*

6.3 Ongoing mechanistic work

At the time of thesis submission, a number of experiments were still ongoing in a continued attempt to determine the mechanistic role of PCBP2 in *TP53*-sAML. This work is therefore described below.

Identifying PCBP2 protein interactions via co-immunoprecipitation

A number of PCBP2 protein-protein interactions have been reported, including interactions with innate immunity signalling proteins (cGAS²⁶⁷, MAVS²²⁵ and AIP4²²⁵) and iron-transporter proteins (DMT²⁴⁵ and FPN1²⁴⁶). Since PCBP2 contains no motifs capable of enzymatic functions, association with additional proteins via KH and WW-binding domains is likely necessary for PCBP2's regulatory functions. For example, PCBP2 acts as an anchor protein for AIP4-mediated ubiquitination of MAVS to facilitate negative regulation of antiviral signalling via PCBP2 WW-binding domains association with AIP4 WW-domains²²⁵.

To determine interactions important for *TP53*-sAML, co-immunoprecipitation experiments followed by mass spectrometry to identify interacting proteins are in progress. In the HEL cell line, *PCBP2* expression was deleted by shRNA as previously described and cells transduced with an exogenous FLAG-tagged *PCBP2* construct. The exogenous *PCBP2* coding sequence has been codon-optimised to prevent binding of the constitutively expressed *PCBP2*-targeting shRNA. Alongside collecting protein lysates for co-immunoprecipitation, rescue experiments are ongoing to confirm expression of exogenous *PCBP2* can rescue the knockdown phenotypes observed in Chapter 4. If *PCBP2* re-expression can rescue the knockdown phenotypes, I hypothesis any protein-protein interactions present may be important for *TP53*-sAML maintenance.

Identifying PCBP2 target transcripts via HyperTribe

Given the primary role of PCBP2 as an RBP, comprehensive assessment of the PCBP2 bound transcriptome in *TP53*-sAML is a crucial component of this mechanistic investigation. Furthermore, identifying PCBP2-bound transcripts will complement RNA-seq data, confirming whether changes in gene expression or splice events are direct PCBP2 interactions or in fact, secondary effects. Traditional methods, such as CLIP-seq

can be laborious and technically challenging. For this reason, I investigated PCBP2-binding sites across the transcriptome using a HyperTRIBE approach^{176,177}.

HyperTRIBE is a method for identifying an RBP's target transcripts, derived from the original TRIBE (targets of RNA-binding proteins identified by editing) methodology^{177,369}. Both methods utilise the editing ability of the *Drosophila* RNA-editing enzyme, ADAR, to identify RBP target transcripts. ADAR contains two domains a dsRNA-binding motif and a catalytic domain, responsible for deamination of adenosine-to-inosine, which is read by ribosomes and reverse transcriptase as guanosine. The HyperTRIBE methodology fuses the coding sequence of an RBP-of-interest to the catalytic domain of ADAR, which subsequently performs adenosine-to-inosine editing on RBP-target transcripts¹⁷⁷. Hence, the ADAR fusion protein leaves a “fingerprint” of A-to-G editing events on RBP-target transcripts which can be identified by high-throughput sequencing. Compared to TRIBE, the HyperTRIBE methodology used a hyperactive ADAR mutant to improve editing efficiencies¹⁷⁷.

In this project, PCBP2 coding sequence was fused to the catalytic domain of *Drosophila* ADAR, separated by a small linker. ADAR carried the E488Q hyperactivation mutation and was codon optimisation to maximise translation in human cells. The PCBP2-ADAR sequence was cloned into an SFFV-eGFP lentiviral transfer plasmid. Similar to the rationale for co-immunoprecipitation experiments explained above, *PCBP2* was codon-optimised and overexpressed in HEL cells depleted of endogenous *PCBP2*. Thus, any transcripts exogenous PCBP2 binds are potentially relevant for *TP53*-sAML maintenance and survival. To control for background editing events, PCBP2 was also fused to a catalytic dead version of ADAR (PCBP2-ADAR_{DC}) and ADAR-only was used as a further control.

At the time of writing, sequencing for this HyperTRIBE project was complete and awaiting analysis. The main advantage of HyperTRIBE over CLIP-Seq is the low cell input required, requiring < 500 cells compared to the millions necessary for conventional CLIP-Seq. Thus, HyperTRIBE allows cell-type specific interactions to be investigated in rare cell populations. This makes HyperTRIBE highly applicable to haematology research for the study of RBP interactomes in rare HSC populations, as previously shown by Nguyen *et al*¹⁷⁸.

6.4 Discussion

In this thesis chapter, I investigated the molecular mechanism by which PCBP2 facilitates *TP53*-sAML maintenance and survival. Given the multiple divergent and cell-specific roles of PCBP2 in PTGR and protein-protein interactions, I was presented with many avenues for exploration.

Analysis of RNA-seq data showed a surprisingly limited number of differentially regulated genes following *PCBP2* depletion, suggesting PCBP2 primarily functions by alternate mechanisms in *TP53*-sAML. Interestingly, none of the commonly dysregulated genes between PCBP2 KD #1 and PCBP2 KD #2 were essential for cancer cell survival (**Figure 6.8**). This made candidate target validation from differential gene expression analysis increasingly difficult.

I next hypothesised PCBP2 overexpression in *TP53*-sAML may alter target transcript splicing, favouring generation of mRNA isoforms implicated in cancer evolution and progression. To investigate the extent to which PCBP2 influences alternative splicing, I analysed ASE in HEL cells transduced with *PCBP2*-targeting shRNA compared to scrambled controls. The MARVEL splice analysis pipeline was kindly run by Dr Sean Wen¹⁶⁶. In contrast to other available pipelines, MARVEL allows comprehensive analysis of the splicing landscape, including PSI-values for RI, A5SS and A3SS ALE and AFE, functionalities lacking from other pipelines¹⁶⁶. Skipped exon events were the most commonly identified ASE (**Figure 6.10**). Furthermore, the integrity of this dataset was corroborated by the identification of previously reported splice targets, such as *RUNX1* and *CDK2*^{237,370}. The majority of alternatively spliced genes underwent isoform switching (**Figure 6.12**); and crucially, would not have been identified by differential gene expression analysis alone. Hence, this analysis highlights the important role of PCBP2 in splicing and the breadth of PTGR mechanisms which require investigation to identify RBP-mediated, and disease-relevant, molecular mechanisms.

In this project, a number of ASE were of interest for further investigation. These included *STAT3*, whose A3SS altered the ratio of *STAT3* α and *STAT3* β transcripts, leading to increased *STAT* β expression and potentially associated tumour suppressive phenotypes.

CDK2 is another alternatively spliced gene-of-interest where PCBP2 has previously been shown to mediate exon inclusion³⁷⁰. PCBP2 ensures retention of exon 5 in *CDK2* transcripts, without which cell cycle kinetics are dysregulated, leading to an enrichment of G₀/G₁ phase cells³⁷⁰. Thus, PCBP2-mediated splicing of *CDK2* may explain reduced competition of *PCBP2* depleted cells in co-culture with parental WT cells in Chapter 4. *CDK2* was also identified to be essential for *JAK2/TP53*-mutant cell line survival (**Figure 6.14**), making it an attractive candidate. However, cell cycle analysis in *PCBP2*-KD versus control cells is required to build on this hypothesis.

While PCBP2 was not determined to play a role in the resolution of R-loops in *TP53*-sAML, DNA-damage was significantly enhanced in the *PCBP2*-KD HEL cell line (**Figure 6.17**). I expect, if PCBP2 plays a role in protecting from cells from excessive DNA-damage, in response to environmental stressors, DNA-damage observed in *PCBP2*-KD cells will be accentuated and cells increasingly sensitive to cell death. This role of PCBP2 in DNA-damage will need to be studied extensively to determine whether the association is direct or indirectly related to DNA-damage response pathways. One possible indirect mechanism is via the regulation of reactive oxygen species (ROS) levels. Loss of PCBP2 is known to sensitise cells to ROS^{371,372}. At the time of writing, preliminary data, generated by Dr Shady Awad (not shown), supports these findings showing *PCBP2*-KD cell line models were increasingly sensitive to H₂O₂ exposure compared to controls. Furthermore, ROS were identified to be upregulated by GSEA in *PCBP2*-KD cells (not significant; **Figure 6.6**). ROS is a well-characterised mediator of DNA-damage³⁷³, however even in the absence of ROS stimulus, I observe an increase in DNA-damage in *PCBP2*-KD cells. Thus, in this context, a number of avenues remain to be explored including, 1) whether *PCBP2*-KD phenotypes can be accentuated by environmental stimuli, 2) whether increasing ROS levels elevates DNA-damage in *PCBP2*-KD cells and 3) what molecular mechanisms are directly mediated by PCBP2?

Overall, PCBP2 is a multifunctional RBP and through analysis of RNA-seq data to identify differentially expressed genes and alternative splicing events, I have presented a variety of mechanisms by which PCBP2 mediates PTGR. Given this, it is likely more than one mechanism contributes to PCBP2-mediated *TP53*-sAML pathogenesis and further investigation is required to elucidate the molecular mechanisms by which PCBP2 is

required in *TP53*-sAML. An important consideration for future mechanistic experiments will be to determine whether PCBP1 expression compensates for the perturbation of *PCBP2* given PCBP1 and PCBP2 possess functional redundancy in a number of contexts, including in iron metabolism. Thus, ongoing experiments, including identification of PCBP2-target transcripts and protein-binding partners will increase confidence in current mechanistic hypothesis and potentially unearth previously unconsidered molecular mechanisms for future investigation.

Chapter 7 Conclusions and future work

7.1 Conclusions

7.1.1 A cohort of RBP, essential for cancer cell survival, are enriched in *TP53*-sAML

Patients with *TP53*-mutant transformed MPN have a dire prognosis with lack of access to effective therapies. In this thesis I investigated the molecular vulnerabilities present in a *TP53*-mutant transformed MPN (*TP53*-sAML) patient cohort, with the aim of identifying candidates for therapeutic targeting. Through analysis of a single-cell multi-omic TARGET-seq dataset, I identified a cohort of RBP significantly upregulated in *TP53*-sAML. Enrichment was specific to the *TP53*-mutant disease-propagating LSC population and importantly, candidate RBP were identified to be essential in *JAK2/TP53*-mutant cell line models via loss-of-function CRISPR-Cas9 screening. While the importance of RBP in AML has previously been reported^{197,299,361}, this work is the first to identify enrichment of an essential RBP cohort specifically in a *TP53*-sAML background. Given the requirement of the identified RBP cohort for model cancer cell line survival, I concluded these RBP are novel molecular vulnerabilities with the potential for therapeutic targeting in *TP53*-sAML.

7.1.2 PCBP2 is a molecular vulnerability with potential for therapeutic targeting in *TP53*-sAML

Having identified a cohort of upregulated RBP as novel molecular vulnerabilities in *TP53*-sAML, I next selected a promising candidate RBP for further functional validation. Through literature review and analysis of external datasets, I provided evidence to suggest PCBP2 is a selective vulnerability, specifically upregulated in the *TP53*-mutant LSC compartment and previously unreported in *TP53*-sAML. I showed depletion of *PCBP2* leads to increased apoptosis and reduced cell proliferation, functionally validating the requirement of PCBP2 in *JAK2/TP53*-mutant AML cell line models. In primary cells, I showed there is a minimal requirement of PCBP2 in WT-HSPC, highlighting the potential therapeutic index for targeting PCBP2 in *TP53*-sAML. Furthermore, while I selected

PCBP2 for functional validation, other RBP within the cohort also presented promise as selective molecular vulnerabilities and can be investigated during future work.

7.1.3 Human bone marrow organoids provide a tractable system for investigating gene perturbations in *TP53*-sAML patient samples

In this thesis, I adopted an iPSC-derived human bone marrow organoid model to overcome the limitations of conventional liquid culture conditions in supporting the survival and proliferation of *TP53*-sAML patient samples. Importantly, this work was the first to show the utility of human bone marrow organoids in supporting primary AML samples.

Having provided a proof-of-principle that human bone marrow organoids are a superior model system for *ex vivo* culture of *TP53*-sAML samples, I used this human bone marrow model to investigate *PCBP2* perturbation in primary patient samples. I transduced *TP53*-sAML samples with *PCBP2*-targeting shRNA and engrafted primary cells into human bone marrow organoids. However, insufficient *PCBP2*-KD efficiencies in primary cells led to inconclusive results regarding the role of *PCBP2* in *TP53*-sAML. Despite this, I have made significant progress in generating a tractable system to investigate gene perturbation in primary cells which are notoriously difficult to grow *ex vivo*. Future work optimising gene depletion will provide conclusive evidence of *PCBP2*'s requirement in *TP53*-sAML.

7.1.4 *PCBP2* is a multifunctional RBP with various roles in post-transcriptional gene regulation

While mechanistic work remains ongoing, I have illustrated the multifunctional roles of *PCBP2* in PTGR. Through RNA-seq analysis, I have shown *PCBP2* influences transcript stability using differential gene expression analysis and GSEA. In addition, *PCBP2* perturbation leads to alternative splicing of key genes involved in myeloid malignancy and worthy of further investigation. Future work to decipher the target transcripts and protein-binding partners of *PCBP2* will help to finalise mechanistic endeavours.

7.2 Wider relevance of this work

Treatments for *TP53*-sAML are inadequate and patients who transform are presented with a dire prognosis. This thesis characterised molecular vulnerabilities in *TP53*-sAML and presented RBP as candidates with the potential for therapeutic targeting. A RBP splicing network has previously been identified as essential for AML maintenance and pharmacological inhibition of this network can lead to broad anti-leukaemic effects¹⁸⁵. Hence, this work corroborates previous findings, emphasising the role of RBP in AML, and extends findings specifically within a *TP53*-mutant background. Furthermore, this thesis adds to the cohort of previously reported RBP implicated in AML.

The successful engraftment of primary AML samples into human bone marrow organoids presented in this thesis is of importance to the wider haematological community. AML samples are notoriously difficult to grow *ex vivo* and frequently fail to engraft murine models. This proof-of-principle work therefore gives researchers a platform to model disease, study gene perturbations and investigate niche-interactions using primary AML samples.

7.3 Limitations of this work

A major limitation of this work was the incomplete knockdown achieved using an shRNA approach in primary patient samples. Despite the same shRNA constructs achieving a KD efficiency resembling a homozygous deletion in cell line models, this efficiency did not translate to primary cells. With only a 50% KD efficiency in primary cells, I could not reliably conclude whether *PCBP2* was selectively required in *TP53*-mutant LSC. Work to overcome this limitation is ongoing, including troubleshooting the shRNA-KD methodology and considering other genome editing approaches.

This work would have benefited from investigating the effect of extrinsic stressors, such as Ruxolitinib treatment, DNA-damage agents or ROS on *PCBP2*-KD phenotypes in functional and mechanistic assays. Physiologically, *PCBP2*-KD in cell line models does not accurately imitate the clinical treatments cells would be exposed to or the bone marrow microenvironment within which *TP53*-mutant LSC reside. Thus, I expect extrinsic

stressors will accentuate the phenotypes observed in functional and mechanistic studies, further emphasising the sensitivity of *TP53*-sAML to *PCBP2* depletion.

7.4 Future work

This thesis made major progress in presenting *PCBP2* as a molecular vulnerability in *TP53*-sAML. However, some research questions still remain unanswered and will need to be addressed in future work.

7.4.1 What is the role of other candidate RBP in *TP53*-sAML?

While I selected *PCBP2* for further functional investigation, other candidate RBP identified also presented promising therapeutic targets. For example, *ILF3* was another candidate of interest, particularly due to its positive selection in *ex vivo* murine WT-HSPC screens. Through this DPhil project, I have generated a general pipeline for the investigation of additional candidate RBP. Importantly, major hurdles in investigating gene perturbation in primary patient samples have been overcome through the application of a human bone marrow organoid model. Hence, this work has accelerated the future investigation of other candidate RBP-of-interest.

7.4.2 Is *PCBP2* required for *TP53*-mutant leukaemic transformation?

Whether *PCBP2* expression is required for *TP53*-mutant MPN transformation was beyond the scope of this project, however, is an important consideration in the prevention of disease progression. Future work can investigate *PCBP2* expression in serial samples from MPN patients who do and do not transform, allowing initial conclusions on *PCBP2* requirement in leukaemic progression to be assessed. If *PCBP2* expression associates with progression, further work would include investigating whether overexpression of *PCBP2* leads to enhanced leukaemogenic phenotypes *in vitro*. Ultimately, the impact of *PCBP2* expression on disease progression will subsequently need to be assessed in MPN-mouse models or PDX models to accurately recapitulate disease.

7.4.3 Is PCBP2 essential for *TP53*-sAML survival?

While this project made significant progress in evaluating the requirement of PCBP2 in *TP53*-sAML, experiments investigating the requirement of PCBP2 in primary *TP53*-sAML samples were inconclusive. A ~50% KD in *PCBP2* expression appeared to have limited effect on *TP53*-sAML proliferation or survival. However, it is impossible to decipher whether this result is due to the limited KD efficiency, or a true lack of requirement by *TP53*-sAML LSC for PCBP2. Thus, ongoing and future work must optimise *PCBP2* gene perturbation to reflect a homozygous deletion.

To do this I am considering a number of options, including different promoters for optimal shRNA expression in primary cells. However, evidence also suggests depletion of *PCBP2* may directly influence the function of miRNA/shRNA processing enzyme, Dicer³⁴⁴. Inherent expression of Dicer in primary cells can also limit KD efficiency³⁴³. Thus, additional experimental controls are required to comprehend the limited KD of *PCBP2* in primary cells including a highly efficient shRNA which does not have any association with the miRNA/shRNA processing pathway.

7.4.4 What are the molecular mechanisms through which PCBP2 supports *TP53*-sAML survival and proliferation?

Differential gene expression and splice analysis in *PCBP2*-KD cell lines begins to unravel the PTGR network of PCBP2. This work will greatly benefit from PCBP2 co-immunoprecipitation experiments to identify PCBP2-interacting proteins and identification of PCBP2-bound transcripts by HyperTribe (Section 6.3). Firstly, these experiments will validate molecular mechanisms identified through RNA-seq analysis, confirming candidate dysregulation is a primary, and not secondary effect, of *PCBP2*-KD. Secondly, they will allow additional targets to be uncovered and previously unconsidered mechanisms investigated. Thorough exploration of PCBP2's multifunction potential as an RBP would also investigate PCBP2 association with chromatin via ChIP-seq, and the impact of PCBP2 on translation through sequencing of polysomal fractions or mass spectrometry of control versus *PCBP2*-KD cell lysates. Future work could also include

mechanistic studies using *TP53*-mutant iPSC or a *p53*^{-/-}-JAK2V617F mouse model to investigate the effects of *PCBP2* perturbation.

At the time of writing, preliminary data showed *PCBP2*-KD cells were increasingly sensitive to H₂O₂ exposure compared to controls, emphasising the potential sensitivity of *PCBP2*-KD cells to exogenous stimuli. With this in mind, applying external stimuli to *PCBP2*-KD cells before executing mechanistic experiments may enhance the effect of important molecular mechanisms. I hypothesise this approach will separate important mechanisms for *TP53*-sAML maintenance from background mechanisms with limited effect on *TP53*-sAML proliferation and survival.

7.5 Final summary

In summary, this thesis presents a cohort of RBP as novel molecular vulnerabilities in *TP53*-sAML, and functionally validates the RBP, *PCBP2*, as a candidate with potential for selective therapeutic targeting. This work shows a minimal requirement of *PCBP2* in *TP53*-WT HSPC and makes significant progress in the development of techniques required to validate the role of *PCBP2* in primary *TP53*-sAML samples. Finally, mechanistic studies exemplified the multifunctional roles of *PCBP2* as an RBP in PTGR and point towards interesting molecular mechanisms for *TP53*-sAML survival and proliferation.

Appendix

Appendix 1. Primary patient cohort

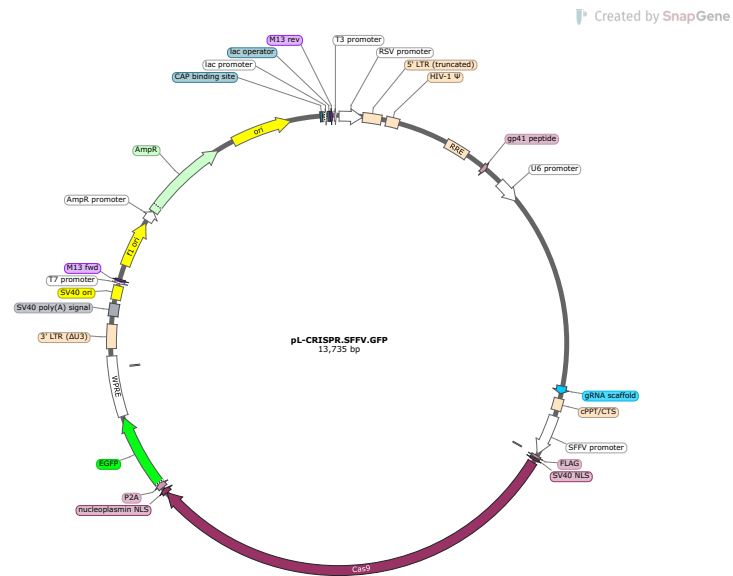
Patient ID	Sample usage	Disease Stage	Source (MNCs, PB/ BM)	MPN subtype and diagnosis age	MPN latency	Mutation at MPN stage (<i>JAK2/CALR</i>)	AML subtype diagnosis	Blast Percentage (%)	Survival (days)	<i>TP53</i> - mutations (number) and type	<i>TP53</i> mutation VAF	chr17p13.1 CNA	Karyotype
IF0392	TARGET-seq	<i>TP53</i> -mutant sAML	MNCs (PB)	ET (age 38)	27 years	<i>JAK2</i> V617F	MPN-AP	17% (BM aspirate)	deceased (134 days after diagnosis)	1: c647T>G p.Val121Gly	92.88	Loss	2 (del5q; del17p)
IF0308	TARGET-seq	<i>TP53</i> -mutant sAML	MNCs (PB & BM)	ET (age 64); MF (age 71)	10 years	<i>CALR</i> type 1	AML-NOS	75% (BM aspirate)	deceased (309 days after diagnosis)	1: c742G>T p.Arg248Trp	99.21	Loss	Complex (>3)
IF0393	TARGET-seq & Organoid engraftment	<i>TP53</i> -mutant sAML	MNCs (PB)	ET (age 66); MF (age 71)	6 years	<i>JAK2</i> V617F	MPN-BP	20% (BM aspirate)	deceased (181 days after diagnosis)	1: c641A>G p.His214Arg	98.04	Normal	1 (chr20q11.22-13.2 loss)
GR001	TARGET-seq	<i>TP53</i> -mutant sAML	NGS on MNCs (PB)	ET (date unknown, before age 72)	NA	<i>CALR</i> type 2	AML	49% (BM aspirate)	deceased (158 days after diagnosis)	2: c477483dupCATGCC p.Ile162Hisfs21; c833C>T p.Pro278Leu	37; 28	NA	NA
IF0391	TARGET-seq	<i>TP53</i> -mutant sAML	MNCs (PB)	ET (age 66)	2 years	<i>JAK2</i> V617F	MPN-BP	20% (BM aspirate)	deceased (22 days after diagnosis)	1: c524G>A p.Arg175His	40.24	Loss	Complex (>3)
IF0131	TARGET-Seq & Single-colony genotyping & Organoid engraftment	<i>TP53</i> -mutant sAML	MNCs (PB)	ET (age 53); MF (age 60)	7 years	<i>JAK2</i> V617F	MPN-AP	10-15% (trephine); CD34+CD117+CD68+	deceased (162 days from MPN-AP diagnosis)	2:c814G>A p.Val272Met; c672+1 G>C	1356; 1181	Normal	Complex (>3)
IF0318	TARGET-Seq & Single-colony genotyping & Organoid engraftment	<i>TP53</i> -mutant sAML	MNCs (PB)	NA	NA	<i>CALR</i> type 2	MPN-BP	60% (trephine)	deceased (175 days after diagnosis)	2: c722C>G p.Ser241Cys; c818G>A p.Arg273His	3934; 38.12	Normal	Complex (>3)
IF0394	Organoid engraftment	<i>TP53</i> -mutant sAML	MNCs (PB)	PV (age 70)	9 years	<i>JAK2</i> V617F	MPN-BP	80% (BM aspirate)	deceased (19 days after diagnosis)	1: c640T>A p.Val147Asp	77.78	Normal	Complex (>3)
001/377	Organoid engraftment	<i>TP53</i> -mutant sAML	MNC (PB)	ET		<i>JAK2</i> V617F	MPN-BP	25% (BM aspirate)	unknown	3: c817G>T p.Arg273Cys; c701A>G p.(Tyr234Cys); c594G>A p.(Arg175His)	<10; unknown; <15	Loss	Complex (>3)
Aph1	TARGET-seq	MPN	MNCs (mobilized PB)	NA	NA	NA	NA	NA	NA	0: NA	NA	NA	NA
HD85	TARGET-seq	MPN	MNCs (mobilized PB)	NA	NA	NA	NA	NA	NA	0: NA	NA	NA	NA
IF0140	TARGET-seq	MPN	MNCs (PB)	MF (post-PV)	NA	<i>JAK2</i> V617F	NA	Yes (4%)	NA	0: NA	NA	NA	NA
IF0101	TARGET-seq	MPN	MNCs (PB)	PMF	NA	<i>JAK2</i> V617F	NA	Yes (<5%)	NA	0: NA	NA	NA	Normal
IF0123	TARGET-seq	MPN	MNCs (PB)	PMF	NA	<i>JAK2</i> V617F	NA	Yes (26; BM aspirate)	NA	0: NA	NA	NA	NA
IF0138	TARGET-seq	MPN	MNCs (PB)	MF (post-PV)	NA	<i>JAK2</i> V617F	NA	NA	NA	0: NA	NA	NA	Normal
IF0137	TARGET-seq	MPN	MNCs (PB)	PMF	NA	<i>JAK2</i> V617F	NA	Yes (1-2%)	NA	0: NA	NA	NA	Normal
IF0131	TARGET-seq	MPN	MNCs (BM)	ET (age 69); MF (age 80)	13 years	<i>JAK2</i> V617F	NA	NA	alive 1087 days post-sampling	1: c725G>A p.Cys242Tyr	16.98	NA	NA
IF0602	TARGET-seq	MPN	MNCs (PB)	PMF	NA	<i>JAK2</i> V617F	NA	Yes (<5%)	NA	0: NA	NA	NA	Trisomy 8; Trisomy 9
IF0155	TARGET-seq	MPN	MNCs (PB)	MF (post-ET)	NA	<i>JAK2</i> V617F	NA	NA	NA	0: NA	NA	NA	Normal
IF0157	TARGET-seq	MPN	MNCs (PB)	MF (post-PV)	NA	<i>JAK2</i> V617F	NA	NA	NA	0: NA	NA	NA	NA
HD15 8650	TARGET-seq	Healthy donor	MNCs (mobilized PB)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
IF0704	TARGET-seq	Healthy donor	MNCs (BM)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
IF0902	TARGET-seq	Healthy donor	MNCs (mobilized PB)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
IF0903	TARGET-seq	Healthy donor	MNCs (mobilized PB)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
IF0905	TARGET-seq	Healthy donor	MNCs (mobilized PB)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
IF0944	Transduction & Organoid engraftment	Healthy donor	MNCs (mobilized PB)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
IF0947	Single cell colony genotyping & Organoid engraftment	Healthy donor	MNCs (mobilized PB)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
IF0948	Transduction & Organoid engraftment	Healthy donor	MNCs (mobilized PB)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
IF0951	Single cell colony genotyping & Organoid engraftment	Healthy donor	MNCs (mobilized PB)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

Appendix 2. List of 97 commonly upregulated genes in *TP53*-mutant LSC.

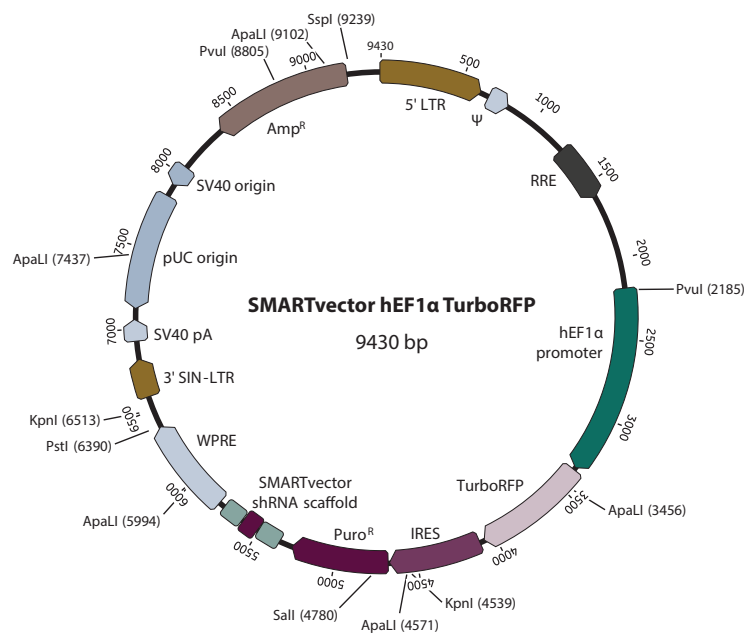
Expression in *TP53*-mutant LSC compared to all other patient cohorts (HD, MPN, preleukaemic).

C1orf228	YARS	PHIP
PLIN2	H3F3A	PRPF40A
MIR181A1HG	BIRC7	MYO9B
COTL1	UBR4	PTMA
HCST	CTSH	PRSS21
TNFSF13B	REL	PRKCB
PMAIP1	DDX39A	RBBP6
LSP1	FAM46A	RPS17L
SLC38A1	MT-ND1	CAPNS1
LGALS1	MTDH	KDM5A
MYO1F	SRSF10	NECAP1
SRSF11	NBEAL2	HNRNPM
MIR4435-1HG	ETS2	PTP4A2
ITM2A	CD36	UBE2D3
TMSB4X	SLC25A37	CCDC12
ATF4	CCND2	UBL5
CCNL1	SUB1	IFIT3
IDH2	IFITM2	IKZF1
PTPRE	TMIGD2	PCBP2
PDLIM1	TIMP1	PELI1
ILF3	GLUL	ARPC5L
TP53INP1	CIR1	TMSB10
TPM4	GPX1	TP53
ATP5E	CYTH4	GTF3A
FAM129A	UBE2J1	SLC25A36
SFPQ	ZC3H15	PAPOLA
SMARCA4	CD44	ARF1
CD97	RTF1	HDC
RPS28	IFIT2	ISG15
YBX1	IREB2	HBG1
LST1	SLC2A3	GUK1
DENND3	ZNF292	AIF1
TMA7	-	-

Appendix 3. Cas9-sgRNA-GFP all-in-one vector lentiviral vector



Appendix 4. shRNA-RFP lentiviral vector



Appendix 5. DESeq2 results for 76 common differentially expressed genes.

Gene expression and *p*-values for both PCBP2 KD #1 and PCBP2 KD #2 vs Control comparisons in the HEL cell line

PCBP2 KD #1 vs Control						
	baseMean	log2FoldChange	lfcSE	pvalue	padj	diffexpressed
PCBP2	8217.8455	-2.467248957	0.1740338	1.27E-45	2.05E-41	DOWN
RNF38	2069.273	1.187647597	0.10316222	1.14E-30	9.21E-27	UP
RARRES2	54.688334	-2.18925381	0.23774556	3.31E-20	8.89E-17	DOWN
PCBP2P2	33.902115	-2.631932922	0.31359924	4.75E-17	1.09E-13	DOWN
BTBD2	2895.2838	-1.320437	0.16099624	2.37E-16	4.78E-13	DOWN
CMTM5	2865.0359	-0.826166645	0.1125576	2.14E-13	3.13E-10	DOWN
NPRL3	4253.8136	-0.541850363	0.07832847	4.59E-12	4.93E-09	DOWN
DNM1	1345.0768	-1.167201955	0.17071422	8.08E-12	8.14E-09	DOWN
RHOJ	775.80982	-1.004245536	0.15317009	5.51E-11	4.94E-08	DOWN
EVI2B	44.614158	1.94370626	0.30108112	1.08E-10	9.14E-08	UP
CNNM1	461.36299	-0.856963601	0.13556573	2.59E-10	1.99E-07	DOWN
LINC01426	260.03144	-0.758357039	0.12743513	2.67E-09	1.72E-06	DOWN
RBCK1	2806.5806	-0.711164321	0.12275669	6.90E-09	3.71E-06	DOWN
ACRBP	815.91998	-0.764632565	0.13245598	7.80E-09	4.06E-06	DOWN
ITLN1	916.59905	0.592199271	0.10286453	8.56E-09	4.31E-06	UP
CXorf38	1594.5515	-0.622274459	0.10901993	1.14E-08	5.42E-06	DOWN
HOXB8	421.82079	-0.866639953	0.15283415	1.42E-08	6.56E-06	DOWN
OBSCN	1622.694	-0.716637626	0.13262902	6.54E-08	2.34E-05	DOWN
KCNE5	79.646619	1.733533684	0.32666248	1.12E-07	3.75E-05	UP
IGFBP5	19851.725	-0.955431598	0.182347	1.61E-07	5.29E-05	DOWN
REEP6	1528.9319	-0.570951557	0.10946694	1.83E-07	5.90E-05	DOWN
AGER	955.28658	-0.539447659	0.10484327	2.67E-07	8.28E-05	DOWN
HIPK3	1522.3956	-0.55066242	0.10828041	3.67E-07	0.00010745	DOWN
GUCY1A1	685.954	0.704293142	0.13947141	4.42E-07	0.00012512	UP
ANOS1	392.3649	-0.695318668	0.13805106	4.74E-07	0.00012941	DOWN
NUTM2D	239.81517	0.857821179	0.17359394	7.75E-07	0.00018373	UP
PRRT1	51.266461	-1.004966988	0.20630037	1.11E-06	0.00024471	DOWN
C2orf66	172.86755	0.950531874	0.19751534	1.49E-06	0.00030811	UP
TBCK	805.17174	-0.628556816	0.13175791	1.84E-06	0.00035256	DOWN
NUTM2E	97.676865	0.792236952	0.16697943	2.09E-06	0.00037853	UP
CHRM4	249.12054	-0.575469337	0.12406363	3.51E-06	0.00058318	DOWN
P2RX1	1320.3546	-0.683360134	0.14912804	4.60E-06	0.0006991	DOWN
KCND1	205.7758	-0.853864321	0.18692783	4.93E-06	0.00072857	DOWN
CLEC12A	356.75092	1.002016935	0.22094677	5.76E-06	0.00083608	UP
TNFRSF13C	129.59258	-0.772093604	0.17195001	7.11E-06	0.00094781	DOWN
FBXO36	71.830554	-0.801271904	0.19160625	2.89E-05	0.00298744	DOWN
NEK11	75.048976	1.10881305	0.26555617	2.97E-05	0.00300252	UP
PPM1F	11446.629	-0.617502831	0.14965375	3.69E-05	0.00353841	DOWN
ATP2A1	161.00472	-0.578540695	0.14070704	3.93E-05	0.00370275	DOWN
SEMA4F	576.58921	-0.51505234	0.12563228	4.14E-05	0.00378906	DOWN
TRAF3IP3	493.34251	0.547062102	0.13446243	4.73E-05	0.00414488	UP
PTPRF	4385.6559	-0.745444453	0.18552637	5.87E-05	0.00490209	DOWN
HPGD	3850.8455	0.975976566	0.24340314	6.08E-05	0.00501614	UP
MPDZ	1221.1171	0.591848953	0.15235896	0.00010251	0.00763835	UP
SLC40A1	10481.501	0.603752532	0.15759685	0.00012763	0.00906263	UP
SCN3A	63.863918	1.541033226	0.40249959	0.00012884	0.00910893	UP
ACE	385.25914	-0.846553369	0.22152121	0.00013262	0.00929422	DOWN
KLK1	42.921841	-1.249713824	0.32737835	0.0001349	0.0093726	DOWN
CDK15	40.407286	1.104407674	0.29257604	0.00016015	0.01084676	UP
GJB2	690.1928	-0.74705821	0.1989469	0.0001733	0.01121841	DOWN
DNHD1	797.40101	-0.593970138	0.15833096	0.00017582	0.01129787	DOWN
TPP1	1719.1788	-0.524592397	0.14007608	0.00018036	0.01149071	DOWN
PAQR6	298.7743	-0.523665165	0.14032521	0.00019012	0.01187801	DOWN
ABCB5	41.744789	0.917822967	0.24619933	0.00019303	0.01201336	UP
TEX19	53.921711	-1.02391351	0.27765115	0.00022623	0.01376065	DOWN
PCSK1	75.298389	0.783851866	0.21307966	0.00023444	0.01399632	UP
PALM	503.50014	-0.643581695	0.17562564	0.00024781	0.01457851	DOWN
FGF14	142.18044	0.557284601	0.15227629	0.00025252	0.01474753	UP
OR14K1	23.212608	-1.229815291	0.33711704	0.00026425	0.01521242	DOWN
FGFR4	622.94255	-0.551784059	0.15146164	0.00026941	0.0153452	DOWN
SRP14-DT	252.25443	-0.522711023	0.14357533	0.00027192	0.01543366	DOWN
TRPC6	375.78701	0.655576286	0.18015478	0.00027374	0.01548215	UP
TMEM272	39.086528	-0.916817699	0.25342734	0.00029726	0.01624257	DOWN
SLC5A12	16.692403	1.868280121	0.52695709	0.000392	0.01974575	UP
OR2W3	1107.3107	0.591573458	0.16863068	0.00045132	0.0217158	UP
COL11A2	224.45078	-0.627531677	0.17885605	0.0004505	0.0217158	DOWN
LGMN	247.61304	0.706203719	0.20189383	0.00046894	0.02236866	UP
C2CM2L	43.290437	-1.070977611	0.31029084	0.00055742	0.02546077	DOWN
PTCS2L	16.665855	1.684152633	0.49098383	0.00060323	0.02700951	UP
BEND6	209.24825	0.570851293	0.16830204	0.00069428	0.02968479	UP
PCOLCE	363.88331	-0.518139943	0.1562838	0.00091519	0.03624564	DOWN
FAM120C	132.97059	-0.591198784	0.17845355	0.00092334	0.03647871	DOWN
RAB3B	1171.7944	-0.607526476	0.1856277	0.00106479	0.04072902	DOWN
XKR5	114.42405	-0.974558109	0.29808028	0.00107757	0.04092534	DOWN
SELL	257.62928	0.642920821	0.19803486	0.00116832	0.04323349	UP
EXOC3L4	615.39267	0.568275534	0.17505671	0.00116942	0.04323349	UP

PCBP2 KD #2 vs Control						
Gene	baseMean	log2FoldChange	lfcSE	pvalue	padj	diffexpressed
PCBP2	8217.846	-2.053805225	0.173997	3.74E-32	1.51E-28	DOWN
RNF38	2069.273	1.166091653	0.103371	1.63E-29	5.27E-26	UP
CMTM5	2865.036	-1.059144645	0.112934	6.70E-21	7.20E-18	DOWN
RARRES2	54.68833	-2.265729003	0.24825	7.05E-20	6.69E-17	DOWN
ACRBP	815.92	-1.205269881	0.13406	2.46E-19	1.72E-16	DOWN
NPRL3	4253.814	-0.67410949	0.078591	9.70E-18	5.79E-15	DOWN
PCBP2P2	33.90211	-2.614398204	0.325025	8.72E-16	4.13E-13	DOWN
HIPK3	1522.396	-0.842081981	0.109037	1.14E-14	5.09E-12	DOWN
PRRT1	51.26646	-1.605062104	0.227709	1.81E-12	5.82E-10	DOWN
ITLN1	916.5991	0.673992319	0.1032	6.54E-11	1.55E-08	UP
AGER	955.2866	-0.680684296	0.105731	1.21E-10	2.75E-08	DOWN
DNM1	1345.077	-1.094445777	0.170944	1.53E-10	3.38E-08	DOWN
CNNM1	461.363	-0.865034954	0.136616	2.42E-10	4.94E-08	DOWN
NUTM2D	239.8152	1.035770005	0.173968	2.62E-09	4.27E-07	UP
BTBD2	2895.284	-0.954405416	0.160952	3.03E-09	4.84E-07	DOWN
PPM1F	11446.63	-0.879778275	0.149724	4.20E-09	6.39E-07	DOWN
GUCY1A1	685.954	0.789218605	0.139792	1.65E-08	2.29E-06	UP
REEP6	1528.932	-0.604600305	0.109862	3.73E-08	4.69E-06	DOWN
PCOLCE	363.8833	-0.87037168	0.158696	4.15E-08	5.10E-06	DOWN
HPGD	3850.846	1.334389373	0.243391	4.19E-08	5.12E-06	UP
OBSCN	1622.694	-0.725861651	0.132925	4.74E-08	5.58E-06	DOWN
CLEC12A	356.7509	1.192582735	0.22112	6.91E-08	7.83E-06	UP
EVI2B	44.61416	1.522372281	0.308286	7.89E-07	6.83E-05	UP
XKR5	114.4241	-1.467818534	0.303992	1.38E-06	0.00011	DOWN
CXorf38	1594.551	-0.526354713	0.109244	1.45E-06	0.000115	DOWN
SEMA4F	576.5892	-0.609983702	0.126708	1.48E-06	0.000115	DOWN
NUTM2E	97.67686	0.812998683	0.169371	1.59E-06	0.00012	UP
TMEM272	39.08653	-1.260219154	0.268457	2.68E-06	0.000187	DOWN
KCNE5	79.64662	1.530609364	0.329256	3.34E-06	0.000229	UP
LINC01426	260.0314	-0.592874233	0.128166	3.73E-06	0.000247	DOWN
RBCK1	2806.581	-0.567367546	0.122843	3.86E-06	0.000253	DOWN
CHRM4	249.1205	-0.579165549	0.125951	4.26E-06	0.000276	DOWN
HOXB8	421.8208	-0.703947716	0.153243	4.36E-06	0.000281	DOWN
TPP1	1719.179	-0.641870056	0.140424	4.86E-06	0.000309	DOWN
ACE	385.2591	-1.002449891	0.222718	6.76E-06	0.000409	DOWN
PAQR6	298.7743	-0.630415089	0.142238	9.33E-06	0.000517	DOWN
PTPRF	4385.656	-0.806249797	0.18562	1.40E-05	0.000739	DOWN
MPDZ	1221.117	0.65770959	0.152544	1.62E-05	0.000822	UP
TBCK	805.1717	-0.569428526	0.132199	1.65E-05	0.000835	DOWN
TRPC6	375.787	0.772030591	0.180537	1.90E-05	0.000934	UP
FBXO36	71.83055	-0.840237554	0.196687	1.94E-05	0.000938	DOWN
LGMN	247.613	0.849866223	0.202321	2.66E-05	0.001216	UP
SRP14-DT	252.2544	-0.603887643	0.14563	3.37E-05	0.001427	DOWN
SLC40A1	10481.5	0.63819181	0.157621	5.15E-05	0.001985	UP
FAM120C	132.9706	-0.725433572	0.182143	6.81E-05	0.00249	DOWN
SELL	257.6293	0.784147085	0.198465	7.78E-05	0.002769	UP
TRAF3IP3	493.3425	0.521160427	0.135206	0.000116	0.003853	UP
IGFBP5	19851.72	-0.698443533	0.182349	0.000128	0.004143	DOWN
FGFR4	622.9426	-0.582137998	0.15215	0.00013	0.004206	DOWN
FGF14	142.1804	0.586834591	0.154098	0.00014	0.004451	UP
ANOS1	392.3649	-0.525943472	0.138484	0.000146	0.004595	DOWN
GJB2	690.1928	-0.744391843	0.199395	0.000189	0.005706	DOWN
P2RX1	1320.355	-0.550876864	0.149289	0.000224	0.006537	DOWN
BEND6	209.2482	0.622175584	0.169334	0.000239	0.006891	UP
KLK1	42.92184	-1.217045726	0.331874	0.000245	0.007049	DOWN
PCSK1	75.29839	0.769874212	0.215874	0.000362	0.009598	UP
RAB3B	1171.794	-0.66069592	0.185956	0.000381	0.009967	DOWN
ATP2A1	161.0047	-0.504918011	0.142547	0.000397	0.010286	DOWN
CDK15	40.40729	1.045523178	0.296768	0.000427	0.010864	UP
CCM2L	43.29044	-1.109715062	0.316026	0.000446	0.011295	DOWN
NEK11	75.04898	0.937049297	0.268929	0.000493	0.012195	UP
TNFRSF13C	129.5926	-0.602264551	0.172993	0.000499	0.012311	DOWN
C2orf66	172.8676	0.693502471	0.199938	0.000523	0.012759	UP
TEX19	53.92171	-0.965503333	0.281118	0.000594	0.013989	DOWN
SCN3A	63.86392	1.382779746	0.404926	0.000638	0.014841	UP
RHOU	775.8098	-0.517495609	0.152788	0.000707	0.015973	DOWN
DNHD1	797.401	-0.535350304	0.158699	0.000743	0.01657	DOWN
ABCB5	41.74479	0.817011046	0.25171	0.001171	0.023773	UP
PALM	503.5001	-0.569460659	0.176121	0.001223	0.024559	DOWN
COL11A2	224.4508	-0.581867392	0.180083	0.001233	0.02466	DOWN
SLC5A12	16.6924	1.698740842	0.533966	0.001466	0.027995	UP
PTCSC1	16.66586	1.563676606	0.497708	0.001679	0.030559	UP
EXOC3L4	615.3927	0.54694076	0.175509	0.001831	0.032612	UP
OR2W3	1107.311	0.51107145	0.168943	0.002485	0.040387	UP
OR14K1	23.21261	-1.015276338	0.339039	0.002748	0.043602	DOWN
KCND1	205.7758	-0.552102972	0.18681	0.003122	0.047968	DOWN

Appendix 6. Top 50 alternative splice events in HEL cells following PCBP2-KD

PCBP2 KD #1 vs Control									
tran_id	event_type	gene_short_name	mean.g1 (Control)	mean.g2 (KD)	mean.diff	statistic	p.val	p.val.adj	direction
chr22:37060823:37061629:+@chr22:37062525:37064296	RI	KCTD17	40.6635	87.4314	46.7680	-2.5639	0.0104	0.0104	up
chr1:156939548:156939910:@chr1:156938418:156938513:@chr1:156937249:156937496	SE	ARHGEF11	90.8982	45.6656	-45.2326	2.5125	0.0120	0.0120	down
chr16:30000929:30001040:+@chr16:30001414:30001575 30005221:30005717	ALE	INO80E	75.9243	32.1768	-43.7475	2.5961	0.0094	0.0094	down
chr19:55146681:55146705:@chr19:55146434:55146466:@chr19:55145400:55145565	SE	TNNT1	21.2606	64.5139	43.2533	-2.6266	0.0086	0.0086	up
chr4:730970:731110:+@chr4:732400:732498:@chr4:733672:733789	SE	PCGF3	38.8851	80.9803	42.0952	-2.6091	0.0091	0.0091	up
chr16:30000929:30001040:+@chr16:30001414:30001530:@chr16:30005221:30005793	SE	INO80E	72.0701	30.0577	-42.0123	2.6011	0.0093	0.0093	down
chr19:55146681:55146942 55146434:55146454:@chr19:55145544:55145565	ALE	TNNT1	23.1885	64.6669	41.4784	-2.6288	0.0086	0.0086	up
chr10:87288442:87288495:@chr10:87238671:87238849 87201647:87203936	ALE	NUTM2A-AS1	18.4219	57.8871	39.4652	-2.1444	0.0320	0.0320	up
6991840:46991946:@chr3:46992475:46992555:@chr3:46992926:46993119:@chr3:46993937:46993937	MXE	NBEAL2	72.9994	33.7246	-39.2749	2.3582	0.0184	0.0184	down
chr1:43404380:43404550:@chr1:43406283:43406894 43409493:43410577	ALE	SZT2	69.9351	33.5897	-36.3453	2.2132	0.0269	0.0269	down
chr19:42403031:42402607:@chr19:42402075:42401816	RI	LIPE	18.8313	54.1246	35.2933	-2.6075	0.0091	0.0091	up
chr17:76094414:76094581:@chr17:76091143:76091235:@chr17:76089175:76089320	SE	EXOC7	59.9687	25.4529	-34.5159	2.4515	0.0142	0.0142	down
chr9:129924960:129925157:@chr9:129923996 129924026:129923844	A3SS	FNBP1	55.8482	22.3720	-33.4762	2.5953	0.0095	0.0095	down
chr5:141655621:141655758:@chr5:141655362:141655400:@chr5:141653323:141654435	SE	ARAP3	56.8680	89.8612	32.9932	-2.6224	0.0087	0.0087	up
chr16:70676881:70676978:@chr16:70674432 70674528:70674306	A3SS	MTSS2	49.8618	17.3226	-32.5392	2.4905	0.0128	0.0128	down
chr16:30000929:30001040:+@chr16:30001212:30001530:@chr16:30005221:30005367	SE	INO80E	53.6583	22.0645	-31.5938	2.3783	0.0174	0.0174	down
chr1:243293091:243293219:@chr1:243293399:243293494:@chr1:243304713:243304777	SE	SDCCAG8	22.3721	53.5228	31.1508	-2.5183	0.0118	0.0118	up
chr20:419558419731:@chr20:420871421031:@chr20:422127:422238	SE	RBCK1	93.6729	63.7565	-29.9164	2.5031	0.0123	0.0123	down
chr14:64737356:64737375:@chr14:64737862:64738224:@chr14:64738742:64738855	SE	PLEKHG3	55.8743	27.3227	-28.5516	2.3321	0.0197	0.0197	down
chr4:730970:731110:+@chr4:732400 732444:732498	A3SS	PCGF3	61.9419	90.2276	28.2857	-2.6165	0.0089	0.0089	up
chr19:55456349:55456489:@chr19:5545636:55455845:@chr19:55455260:55455330	SE	ISOC2	80.8207	52.7255	-28.0952	2.4302	0.0151	0.0151	down
chr2:127059011:127059155:@chr2:127057473:127057601:@chr2:127051154:127051243	SE	BIN1	33.9649	62.0125	28.0476	-2.4829	0.0130	0.0130	up
chr6:157653525:157653627:@chr6:157672724 157672769:157678157	A3SS	ZDHHC14	33.0729	5.5145	-27.5583	2.5961	0.0094	0.0094	down
chr20:35618450:35618475:@chr20:35618612:35618720	RI	SPAG4	24.4261	51.8161	27.3900	-2.2838	0.0224	0.0224	up
chr11:67289348:67289550:@chr11:67289829:67289965:@chr11:67290078:67290213	SE	ANKRD13D	21.5905	48.0201	26.4297	-2.6083	0.0091	0.0091	up
chr20:34860254:34860321:@chr20:34860042 34860130:34859969	A3SS	GGT7	13.4711	39.6454	26.1743	-2.4139	0.0158	0.0158	up
chrX:48706498:48706627:@chrX:48707437:48709012	RI	SUV39H1	16.0520	41.7045	25.6525	-2.5079	0.0121	0.0121	up
chrX:130339911:130339976 130340134:130340353:@chrX:130345182:130345275	APE	SLC25A14	39.4916	64.6804	25.1888	-2.0518	0.0402	0.0402	up
chr2:178537361:178537447:@chr2:178537604:178537691	RI	TTN-AS1	15.8803	40.9755	25.0952	-2.4547	0.0141	0.0141	up
chr12:122992499:122992669:@chr12:122991755:122991898:@chr12:122990545:122990709	SE	PITPNM2	24.3288	48.9003	24.5715	-2.5635	0.0104	0.0104	up
chr9:129924960:129925157:@chr9:129923844:129923996:@chr9:129908890:129908999	SE	FNBP1	14.6186	39.0682	24.4497	-2.5665	0.0103	0.0103	up
chr8:124562782:124562992:@chr8:124557681:124557875:@chr8:124556232:124556405	SE	MTSS1	68.7243	44.6981	-24.0262	2.5887	0.0096	0.0096	down
chr2:1219443108:1219443150:@chr2:1219444025:1219444059:@chr2:1219444653:1219444742	SE	SPEG	84.0757	60.0516	-24.0240	2.1546	0.0312	0.0312	down
chr22:37060823:37060922:@chr22:37061104:37064296	RI	KCTD17	39.6585	63.0168	23.3583	-2.4919	0.0127	0.0127	up
chr14:68879962:68880108:@chr14:68878458:68878523:@chr14:68877082:68877240	SE	ACTN1	72.6114	49.3416	-23.2697	2.3288	0.0199	0.0199	down
chr2:73946719:73946906:@chr2:73950585:73950732:@chr2:73958146:73958245	SE	DGUOK	36.7274	13.6891	-23.0384	2.6247	0.0087	0.0087	down
chr9:35684246:35684315:@chr9:35682937:35683241 35682014:35682163	ALE	TPM2	28.8563	5.9115	-22.9448	2.4763	0.0133	0.0133	down
chr1:111126227:111126294:@chr1:111125389:111125650 111120516:111120693	ALE	DRAM2	63.6834	41.7895	-21.8939	2.0657	0.0389	0.0389	down
chr16:28835933:28836122:@chr16:28836326:28836485:@chr16:28836728:28837436	SE	ATXN2L	34.4694	56.3539	21.8845	-2.3959	0.0166	0.0166	up
chr3:197194425:197194589:@chr3:197161640:197161738:@chr3:197149743:197149796	SE	DLG1	33.7615	55.6282	21.8667	-2.4609	0.0139	0.0139	up
chr2:30447257:30447404:@chr2:30459552:30459722:@chr2:30525587:30525755	SE	LCLAT1	34.6814	12.9539	-21.7274	2.1709	0.0299	0.0299	down
chr8:143684479:143684542:@chr8:143689204:143689301:@chr8:143690057:143690123	SE	ZNF707	17.0106	38.5372	21.5266	-2.3461	0.0190	0.0190	up
chr16:28835933:28836485:@chr16:28836728:28837436	RI	ATXN2L	61.8200	40.3908	-21.4292	2.5125	0.0120	0.0120	down
chr8:143684507:143684542 143688659:143689301:@chr8:143690057:143690123	APE	ZNF707	24.4042	45.6845	21.2803	-2.2742	0.0230	0.0230	up
chr19:14161312:14161626:@chr19:14160593:14160696:@chr19:14160112:14160297	SE	ADGR1L	86.1088	64.9214	-21.1874	2.4840	0.0130	0.0130	down
chr9:128771815:128771941 128771581:128771927:@chr9:128755408:128755659	APE	ZER1	47.1403	67.9412	20.8009	-2.1700	0.0300	0.0300	up
chr9:128273850:128273972:@chr9:128272785:128272865:@chr9:128268420:128268524	SE	GOLGA2	40.3209	19.6333	-20.6876	2.2530	0.0243	0.0243	down
chr3:47853163:47853352:@chr3:47851868:47853039 47850690:47852938	ALE	MAP4	61.2504	81.6913	20.4409	-2.5523	0.0107	0.0107	up
chr2:32433644:32433804:@chr2:32435496:32435585:@chr2:32436053:32436184	SE	BIRC6	58.8803	79.0825	20.2022	-2.2099	0.0271	0.0271	up
chr12:6333069:6333105 6330853:6331381:@chr12:6330598:6330711	APE	TNFRSF1A	89.9975	69.9488	-20.0487	2.3504	0.0188	0.0188	down

Top 50 alternative splice events flowing PCBP2-KD vs Control in HEL cells continued...

PCBP2 KD #2 vs Control									
tran_id	event_type	gene_short_name	mean.g1 (Control)	mean.g2 (KD)	mean.diff	statistic	p.val	p.valadj	direction
chr2:73946719:73946906+@chr2:73950585:73950732+@chr2:73958146:73958245	SE	DGUOK	36.7274	13.6891	-23.0384	2.6247	0.0087	0.0087	down
chr16:30000929:30001040+@chr16:30001414:30001530+@chr16:30005221:30005793	SE	INO80E	72.0701	30.0577	-42.0123	2.6011	0.0093	0.0093	down
chr1:243293091:243293219+@chr1:243293399:243293494+@chr1:243304713:243304777	SE	SDCCAG8	22.3721	53.5228	31.1508	-25183	0.0118	0.0118	up
chr19:40341890:40341786 40341839+@chr19:40336320:40336407	A5SS	C19orf47	24.5242	37.8378	13.3136	-24793	0.0132	0.0132	up
chr4:730970:731110+@chr4:732400:732498+@chr4:733672:733789	SE	PCGF3	38.8851	80.9803	42.0952	-26091	0.0091	0.0091	up
chr11:85996826:85996929+@chr11:85990250:85990399+@chr11:85983866:85983973	SE	PICALM	45.0683	30.3348	-14.7335	2.6083	0.0091	0.0091	down
chr17:16439045:16439414+@chr17:16439660:16439703+@chr17:16440185:16440253	SE	SNHG29	13.6160	18.7391	5.1231	-24461	0.0144	0.0144	up
chr3:78651732:78651929+@chr3:78647629:78647655+@chr3:78646148:78646190	SE	ROBO1	73.9472	56.6615	-17.2857	2.5243	0.0116	0.0116	down
chr15:73289074:73289238+@chr15:73293390:73293548+@chr15:73298348:73298611	SE	NEO1	88.1999	94.9665	6.7665	-25890	0.0096	0.0096	up
chr9:129924960:129925157+@chr9:129923996 129924026:129923844	A3SS	FNBP1	55.8482	22.3720	-33.4762	2.5953	0.0095	0.0095	down
chr20:419558:419731+@chr20:420871421031+@chr20:422127:422238	SE	RBCK1	93.6729	63.7565	-29.9164	2.5031	0.0123	0.0123	down
chr1:15719734:15719920+@chr1:15721329:15721388+@chr1:15725317:15725545	SE	PLEKHM2	44.0283	32.0976	-11.9307	2.4940	0.0126	0.0126	down
chr17:44322808:44322852+@chr17:44322528 44322552:44322419	A3SS	SLC25A39	20.2284	7.9601	-12.2682	2.6264	0.0086	0.0086	down
chr5:181233413:181233435+@chr5:181233636:181233763	RI	TRIM41	19.7588	28.5446	8.7858	-23932	0.0167	0.0167	up
chr22:42510631:42510724+@chr22:42511816:42511927	RI	SERHL	63.4460	75.1439	11.6980	-25269	0.0115	0.0115	up
chrX:48706498:48706627+@chrX:48707437:48709012	RI	SUV39H1	16.0520	41.7045	25.6525	-25079	0.0121	0.0121	up
chr21:45283527:45283352 45283383+@chr21:45282349:45282459	A5SS	POFUT2	3.5399	14.5368	10.9968	-26000	0.0093	0.0093	up
chr21:39180285:39180436+@chr21:39179924:39179986+@chr21:39178449:39178647	SE	PSMG1	66.6596	73.2038	6.5442	-25218	0.0117	0.0117	up
chr9:33960826:33960878+@chr9:33953285:33953474+@chr9:33948374:33948587	SE	UBAP2	86.5201	80.9513	-5.5688	2.2040	0.0275	0.0275	down
chr17:16439078:16439414+@chr17:16439581:16439703+@chr17:16440185:16440253	SE	SNHG29	19.5565	26.4342	6.8778	-24698	0.0135	0.0135	up
chr19:39383208:39383291+@chr19:39383499:39384895 39383708:39384188	ALE	SAMD4B	96.6160	86.2135	-10.4025	2.2950	0.0217	0.0217	down
chr5:17753802:17753497+@chr5:17753350:17753168	RI	FAM193B	65.6066	54.1563	-11.4503	2.2719	0.0231	0.0231	down
chr19:40705447:40705325+@chr19:40705181:40705096	RI	COQ8B	14.0533	27.6304	13.5771	-25642	0.0103	0.0103	up
chr4:73258276:73258322 73222974:73223039+@chr4:73177380:73177533	AFE	ANKRD17	16.4973	22.6464	6.1491	-22564	0.0240	0.0240	up
chr8:143829280:143829303 143824313:143824399+@chr8:143821818:143821913	AFE	PUF60	83.4160	77.5779	-5.8380	2.1285	0.0333	0.0333	down
chr5:141655621:141655758+@chr5:141655362:141655400+@chr5:141653323:141654435	SE	ARAP3	56.8680	89.8612	32.9932	-26224	0.0087	0.0087	up
chr5:181232659:181232889+@chr5:181233413:181233435+@chr5:181233636:181233763	SE	TRIM41	95.7295	87.6953	-8.0342	2.5190	0.0118	0.0118	down
chr11:800244:800131+@chr11:800014:799815	RI	PIDD1	37.0751	45.0859	8.0108	-22433	0.0249	0.0249	up
chr2:101867212:101867309+@chr2:101868029:101868037+@chr2:101869622:101869797	SE	MAP4K4	38.2618	24.9373	-13.3245	2.5828	0.0098	0.0098	down
chr9:35684246:35684315+@chr9:35681992:35682904 35682014:35682163	ALE	TPM2	17.6712	7.9203	-9.7510	1.9837	0.0473	0.0473	down
chr7:90405519:90405608+@chr7:90412007:90412049+@chr7:90412644:90415954	SE	CLDN12	11.7929	0.0000	-11.7929	2.0256	0.0428	0.0428	down
chr16:28981455:28981615+@chr16:28981901:28982056+@chr16:28982356:28982545	SE	SPNS1	86.7747	95.9879	9.2132	-25092	0.0121	0.0121	up
chr1:161176434:161176579+@chr1:161176074 161176095:161175808	A3SS	B4GALT3	27.7912	10.7410	-17.0502	2.4523	0.0142	0.0142	down
chr17:75289354:75289510+@chr17:75288388:75288591+@chr17:75288633:75288602	SE	SLC25A19	40.0758	49.1641	9.0883	-23243	0.0201	0.0201	up
chr18:908265:908363+@chr18:909446:912172	RI	ADCYAP1	15.1827	21.1281	5.9454	-23077	0.0210	0.0210	up
chr16:28835933:28836122+@chr16:28836326:28836485+@chr16:28836728:28837436	SE	ATXN2L	34.4694	56.3539	21.8845	-23959	0.0166	0.0166	up
chr6:157653525:157653627+@chr6:157672724 157672769:157678157	A3SS	ZDHHC14	33.0729	5.5145	-27.5583	2.5961	0.0094	0.0094	down
chr1:155060763:155060832+@chr1:155061904:155061975+@chr1:155062245:155062369	SE	ADAM15	14.7091	32.4377	17.7286	-26035	0.0092	0.0092	up
chrX:135052196:135051917+@chrX:135051519:135050932	RI	RTL8A	58.0822	69.4108	11.3286	-22490	0.0245	0.0245	up
chr5:177492405:177492435+@chr5:177491807:177491995+@chr5:177491404:177491420	SE	PDLM7	19.7910	35.1137	15.3227	-25924	0.0095	0.0095	up
chr2:178478279:178478600+@chr2:178477062:178477213+@chr2:178469652:178469785	SE	FKBP7	62.5136	70.1890	7.6754	-20997	0.0358	0.0358	up
chr11:31802704:31802834+@chr11:31801871:31801912+@chr11:31800691:31801776	SE	PAX6	26.2921	39.5875	13.2955	-22098	0.0271	0.0271	up
chr9:35684246:35684315+@chr9:35682937:35683241 35682014:35682163	ALE	TPM2	28.8563	5.9115	-22.9448	2.4763	0.0133	0.0133	down
chr11:910775:911479+@chr11:908542:908645+@chr11:904706:904859	SE	CHID1	36.1856	28.2539	-7.9316	2.3377	0.0194	0.0194	down
chrX:130901684:130901731+@chrX:130783547:130783690+@chrX:130703120:130703254	SE	ENOX2	68.5314	75.0967	6.5653	-20442	0.0409	0.0409	up
chr10:110207684:110208082 110226084:110226283+@chr10:110228189:110228321	AFE	MXI1	76.4673	87.1289	10.6617	-23139	0.0207	0.0207	up
chr17:16439063:16439414+@chr17:16439528:16439703+@chr17:16440185:16440253	SE	SNHG29	18.9916	24.7926	5.8010	-24785	0.0132	0.0132	up
chr17:47808399:47808191+@chr17:47807836:47807373	RI	AC003665.1	94.1119	81.5952	-12.5167	2.1581	0.0309	0.0309	down
chr7:142952509:142952638+@chr7:142946207:142946317+@chr7:142944643:142944741	SE	KEL	90.3084	74.4647	-15.8436	2.4946	0.0126	0.0126	down
chr3:49016340:49016158+@chr3:49016060:49015973	RI	DALRD3	67.9539	77.7224	9.7686	-21454	0.0319	0.0319	up

Appendix 7. Complete list of differential gene expression-splice relationships

Gene ID	Gene Name	Gene Type	Dynamic	Gene ID	Gene Name	Gene Type	Dynamic
ENSG00000013306.16	SLC25A39	protein_coding	Iso-Switch	ENSG00000116786.13	PLEKHM2	protein_coding	Iso-Switch
ENSG00000172932.14	ANKRD13D	protein_coding	Iso-Switch	ENSG00000144893.12	MED12L	protein_coding	Iso-Switch
ENSG00000073921.18	PICALM	protein_coding	Iso-Switch	ENSG00000115904.12	SOS1	protein_coding	Iso-Switch
ENSG00000079435.10	LIPE	protein_coding	Iso-Switch	ENSG00000171843.16	MLLT3	protein_coding	Iso-Switch
ENSG00000175048.17	ZDHHC14	protein_coding	Iso-Switch	ENSG00000110665.11	C11orf21	protein_coding	Iso-Switch
ENSG00000134283.17	PPHLN1	protein_coding	Iso-Switch	ENSG00000100650.15	SRSF5	protein_coding	Iso-Switch
ENSG00000204463.12	BAG6	protein_coding	Iso-Switch	ENSG00000093144.19	ECHDC1	protein_coding	Iso-Switch
ENSG00000067141.17	NEO1	protein_coding	Iso-Switch	ENSG00000072071.16	ADGRL1	protein_coding	Iso-Switch
ENSG00000170873.18	MTSS1	protein_coding	Iso-Switch	ENSG00000160392.13	C19orf47	protein_coding	Iso-Switch
ENSG00000007264.15	MATK	protein_coding	Iso-Switch	ENSG00000134072.11	CAMK1	protein_coding	Iso-Switch
ENSG00000101040.19	ZMYND8	protein_coding	Iso-Switch	ENSG00000117569.18	PTBP2	protein_coding	Iso-Switch
ENSG000000092199.17	HNRNPC	protein_coding	Iso-Switch	ENSG00000183496.6	MEX3B	protein_coding	Iso-Switch
ENSG00000123815.12	COQ8B	protein_coding	Iso-Switch	ENSG00000169398.19	PTK2	protein_coding	Iso-Switch
ENSG00000090975.12	PITPNM2	protein_coding	Iso-Switch	ENSG00000142949.17	PTPRF	protein_coding	Opposing
ENSG00000107186.16	MPDZ	protein_coding	Opposing	ENSG00000170004.16	CHD3	protein_coding	Iso-Switch
ENSG00000165480.16	SKA3	protein_coding	Iso-Switch	ENSG00000011600.11	TYROBP	protein_coding	Iso-Switch
ENSG00000266173.7	STRADA	protein_coding	Iso-Switch	ENSG00000062194.16	GPBP1	protein_coding	Iso-Switch
ENSG00000145907.15	G3BP1	protein_coding	Iso-Switch	ENSG00000156515.23	HK1	protein_coding	Iso-Switch
ENSG00000036672.16	USP2	protein_coding	Iso-Switch	ENSG00000237298.10	TTN-AS1	lncRNA	Iso-Switch
ENSG00000120611.5	ERCC1	protein_coding	Iso-Switch	ENSG00000105327.17	BBC3	protein_coding	Iso-Switch
ENSG00000047849.21	MAP4	protein_coding	Iso-Switch	ENSG00000111642.15	CHD4	protein_coding	Iso-Switch
ENSG00000152601.17	MBNL1	protein_coding	Iso-Switch	ENSG00000182473.22	EXOC7	protein_coding	Iso-Switch
ENSG00000077235.18	GTFC1	protein_coding	Iso-Switch	ENSG00000153250.20	RBMS1	protein_coding	Iso-Switch
ENSG00000099999.14	RNF215	protein_coding	Iso-Switch	ENSG00000196924.17	FLNA	protein_coding	Iso-Switch
ENSG00000116871.15	MAP7D1	protein_coding	Iso-Switch	ENSG00000105492.16	SIGLEC6	protein_coding	Iso-Switch
ENSG00000100813.14	ACIN1	protein_coding	Iso-Switch	ENSG00000064393.16	HIPK2	protein_coding	Iso-Switch
ENSG00000197283.16	SYNGAP1	protein_coding	Iso-Switch	ENSG00000058799.15	YIPF1	protein_coding	Iso-Switch
ENSG00000154310.17	TNIK	protein_coding	Iso-Switch	ENSG00000149532.15	CPSF7	protein_coding	Iso-Switch
ENSG00000172250.16	SERHL	ribed_unitary_pseud	Iso-Switch	ENSG00000088833.17	NSFL1C	protein_coding	Iso-Switch
ENSG00000169855.20	ROBO1	protein_coding	Iso-Switch	ENSG00000154222.14	CC2D1B	protein_coding	Iso-Switch
ENSG00000183527.11	PSMG1	protein_coding	Iso-Switch	ENSG00000063241.8	ISOC2	protein_coding	Iso-Switch
ENSG00000122390.19	NAA60	protein_coding	Iso-Switch	ENSG00000157193.17	LRP8	protein_coding	Iso-Switch
ENSG00000054282.16	SDCCAG8	protein_coding	Iso-Switch	ENSG00000110497.14	AMBRA1	protein_coding	Iso-Switch
ENSG00000132694.18	ARHGEF11	protein_coding	Iso-Switch	ENSG00000146094.14	DOK3	protein_coding	Iso-Switch
ENSG00000169682.18	SPNS1	protein_coding	Iso-Switch	ENSG00000117481.10	NSUN4	protein_coding	Iso-Switch
ENSG00000157350.13	ST3GAL2	protein_coding	Iso-Switch	ENSG00000105176.18	URI1	protein_coding	Iso-Switch
ENSG00000163528.13	CHCHD4	protein_coding	Iso-Switch	ENSG00000197256.10	KANK2	protein_coding	Iso-Switch
ENSG00000101945.16	SUV39H1	protein_coding	Iso-Switch	ENSG00000164088.18	PPM1M	protein_coding	Iso-Switch
ENSG00000139219.19	COL2A1	protein_coding	Iso-Switch	ENSG00000070882.13	OSBPL3	protein_coding	Iso-Switch
ENSG00000112514.16	CUTA	protein_coding	Iso-Switch	ENSG00000085511.20	MAP3K4	protein_coding	Iso-Switch
ENSG00000125826.20	RBCK1	protein_coding	Coordinated	ENSG00000086758.16	HUWE1	protein_coding	Iso-Switch
ENSG00000166200.15	COPS2	protein_coding	Iso-Switch	ENSG00000182534.13	MXRA7	protein_coding	Iso-Switch
ENSG00000131473.17	ACLY	protein_coding	Iso-Switch	ENSG00000166169.17	POLL	protein_coding	Iso-Switch
ENSG00000243156.8	MICAL3	protein_coding	Iso-Switch	ENSG00000144283.21	PKP4	protein_coding	Iso-Switch
ENSG00000138668.19	HNRNPD	protein_coding	Iso-Switch	ENSG00000084411.6	NFIX	protein_coding	Iso-Switch
ENSG00000095319.14	NUP188	protein_coding	Iso-Switch	ENSG00000122481.17	RWDD3	protein_coding	Iso-Switch
ENSG00000068903.20	SIRT2	protein_coding	Iso-Switch	ENSG00000196498.13	NCOR2	protein_coding	Iso-Switch
ENSG00000197993.9	KEL	protein_coding	Iso-Switch	ENSG00000115275.13	MOGS	protein_coding	Iso-Switch
ENSG00000197111.15	PCBP2	protein_coding	Opposing	ENSG00000189306.11	RRP7A	protein_coding	Iso-Switch
ENSG00000273611.5	ZNHIT3	protein_coding	Iso-Switch	ENSG00000163520.14	FBLN2	protein_coding	Iso-Switch
ENSG00000049618.23	ARID1B	protein_coding	Iso-Switch	ENSG00000119950.21	MXI1	protein_coding	Iso-Switch
ENSG00000132305.20	IMMT	protein_coding	Iso-Switch	ENSG00000244486.8	SCARF2	protein_coding	Iso-Switch
ENSG00000083544.14	TDIRD3	protein_coding	Iso-Switch	ENSG00000165233.18	CARD19	protein_coding	Iso-Switch
ENSG00000112983.17	BRD8	protein_coding	Iso-Switch	ENSG00000187838.17	PLSCR3	protein_coding	Iso-Switch
ENSG00000124782.20	RREB1	protein_coding	Iso-Switch	ENSG00000157873.17	TNFRSF14	protein_coding	Iso-Switch
ENSG00000120645.11	IQSEC3	protein_coding	Iso-Switch	ENSG00000141433.13	ADCYAP1	protein_coding	Iso-Switch
ENSG00000185187.13	SIGIRR	protein_coding	Iso-Switch	ENSG00000167904.15	TMEM68	protein_coding	Iso-Switch
ENSG00000163918.10	RFC4	protein_coding	Iso-Switch	ENSG00000145335.15	SNCA	protein_coding	Iso-Switch
ENSG00000140416.21	TPM1	protein_coding	Iso-Switch	ENSG00000166004.15	CEP295	protein_coding	Iso-Switch
ENSG00000091527.15	CDV3	protein_coding	Iso-Switch	ENSG00000115170.13	ACVR1	protein_coding	Iso-Switch
ENSG00000198563.14	DDX39B	protein_coding	Iso-Switch	ENSG00000129810.15	SGO1	protein_coding	Iso-Switch
ENSG00000275052.5	PPP4R3B	protein_coding	Iso-Switch	ENSG00000161265.15	U2AF1L4	protein_coding	Iso-Switch
ENSG00000141504.11	SAT2	protein_coding	Iso-Switch	ENSG00000163156.11	SCNM1	protein_coding	Iso-Switch
ENSG00000261408.7	TEN1-CDK3	protein_coding	Iso-Switch	ENSG00000137776.17	SLTM	protein_coding	Iso-Switch
ENSG0000029534.20	ANK1	protein_coding	Iso-Switch	ENSG00000182087.14	TMEM259	protein_coding	Iso-Switch
ENSG00000113108.19	APBB3	protein_coding	Iso-Switch	ENSG00000158805.12	ZNF276	protein_coding	Iso-Switch
ENSG00000160058.18	BSDC1	protein_coding	Iso-Switch	ENSG00000061656.9	SPAG4	protein_coding	Iso-Switch
ENSG00000169951.9	ZNF764	protein_coding	Iso-Switch	ENSG00000144366.16	GULP1	protein_coding	Iso-Switch
ENSG00000171552.13	BCL2L1	protein_coding	Iso-Switch	ENSG00000088451.11	TGDS	protein_coding	Iso-Switch
ENSG00000072110.13	ACTN1	protein_coding	Iso-Switch	ENSG00000143226.13	FCGR2A	protein_coding	Iso-Switch
ENSG00000133619.17	KRBA1	protein_coding	Iso-Switch	ENSG00000124596.17	OARD1	protein_coding	Iso-Switch
ENSG00000121644.19	DES12	protein_coding	Iso-Switch	ENSG00000143995.20	MEIS1	protein_coding	Iso-Switch
ENSG00000139083.10	ETV6	protein_coding	Iso-Switch	ENSG00000126522.16	ASL	protein_coding	Iso-Switch
ENSG00000196670.13	ZFP62	protein_coding	Iso-Switch	ENSG00000163681.15	SLMAP	protein_coding	Iso-Switch

Appendix 7 continued...

Gene ID	Gene Name	Gene Type	Dynamic	Gene ID	Gene Name	Gene Type	Dynamic
ENSG00000110274.16	CEP164	protein_coding	Iso-Switch	ENSG00000115020.17	PIKFYVE	protein_coding	Iso-Switch
ENSG00000102710.20	SUPT20H	protein_coding	Iso-Switch	ENSG00000100241.21	SBF1	protein_coding	Iso-Switch
ENSG00000103168.16	TAF1C	protein_coding	Iso-Switch	ENSG00000084774.14	CAD	protein_coding	Iso-Switch
ENSG00000115073.7	ACTR1B	protein_coding	Iso-Switch	ENSG00000137074.18	APTX	protein_coding	Iso-Switch
ENSG00000257267.3	ZNF271P	transcribed_unitary_pseudogene	Iso-Switch	ENSG00000090905.19	TNRC6A	protein_coding	Iso-Switch
ENSG00000154723.12	ATP5PF	protein_coding	Iso-Switch	ENSG00000171307.19	ZDHH16	protein_coding	Iso-Switch
ENSG00000131504.17	DIAPH1	protein_coding	Iso-Switch	ENSG00000105669.14	COPE	protein_coding	Iso-Switch
ENSG00000131626.18	PPFIA1	protein_coding	Iso-Switch	ENSG00000198431.16	TXNRD1	protein_coding	Iso-Switch
ENSG00000103326.12	CAPN15	protein_coding	Iso-Switch	ENSG00000011132.12	AC005954.1	protein_coding	Iso-Switch
ENSG00000132466.18	ANKRD17	protein_coding	Iso-Switch	ENSG00000074054.18	CLASP1	protein_coding	Iso-Switch
ENSG00000167548.15	KMT2D	protein_coding	Iso-Switch	ENSG00000005810.18	MYCBP2	protein_coding	Iso-Switch
ENSG00000149476.15	TKFC	protein_coding	Iso-Switch	ENSG00000125458.7	NTSC	protein_coding	Iso-Switch
ENSG00000167110.17	GOLGA2	protein_coding	Iso-Switch	ENSG00000135597.18	REPS1	protein_coding	Iso-Switch
ENSG00000159216.18	RUNX1	protein_coding	Iso-Switch	ENSG00000119487.17	MAPKAP1	protein_coding	Iso-Switch
ENSG00000256683.7	ZNF350	protein_coding	Iso-Switch	ENSG00000136861.18	CDK5RAP2	protein_coding	Iso-Switch
ENSG00000203950.6	RTL8A	protein_coding	Iso-Switch	ENSG00000089053.13	ANAPC5	protein_coding	Iso-Switch
ENSG00000068831.19	RASGRP2	protein_coding	Iso-Switch	ENSG00000102858.13	MGRN1	protein_coding	Iso-Switch
ENSG00000253394.6	LINC00534	lncRNA	Iso-Switch	ENSG00000254093.9	PINX1	protein_coding	Iso-Switch
ENSG00000156304.14	SCAF4	protein_coding	Iso-Switch	ENSG00000142694.6	EVA1B	protein_coding	Iso-Switch
ENSG00000099783.12	HNRNPM	protein_coding	Iso-Switch	ENSG00000185630.18	PBX1	protein_coding	Iso-Switch
ENSG00000108474.16	PIGL	protein_coding	Iso-Switch	ENSG00000071564.16	TCF3	protein_coding	Iso-Switch
ENSG00000169733.12	RFNG	protein_coding	Iso-Switch	ENSG00000133739.16	LRRCC1	protein_coding	Iso-Switch
ENSG00000110455.13	ACCS	protein_coding	Iso-Switch	ENSG00000142856.16	ITGB3BP	protein_coding	Iso-Switch
ENSG00000134874.17	DZIP1	protein_coding	Iso-Switch	ENSG00000170364.12	SETMAR	protein_coding	Iso-Switch
ENSG00000171206.14	TRIM8	protein_coding	Iso-Switch	ENSG00000079150.18	FKBP7	protein_coding	Iso-Switch
ENSG00000104765.16	BNIP3L	protein_coding	Iso-Switch	ENSG00000172889.16	EGFL7	protein_coding	Iso-Switch
ENSG00000112305.14	SMAP1	protein_coding	Iso-Switch	ENSG00000213024.11	NUP62	protein_coding	Iso-Switch
ENSG00000184792.16	OSBP2	protein_coding	Iso-Switch	ENSG00000140987.20	ZSCAN32	protein_coding	Iso-Switch
ENSG00000153037.14	SRP19	protein_coding	Iso-Switch	ENSG00000154380.17	ENAH	protein_coding	Iso-Switch
ENSG00000227507.2	LTB	protein_coding	Iso-Switch	ENSG00000181350.11	LRR75A	protein_coding	Iso-Switch
ENSG00000168675.18	LDLRAD4	protein_coding	Iso-Switch	ENSG00000138674.17	SEC31A	protein_coding	Iso-Switch
ENSG00000151914.20	DST	protein_coding	Iso-Switch	ENSG00000231074.9	HCG18	lncRNA	Iso-Switch
ENSG00000012822.16	CALCOCO1	protein_coding	Iso-Switch	ENSG00000079156.17	OSBP6	protein_coding	Iso-Switch
ENSG00000133612.19	AGAP3	protein_coding	Iso-Switch	ENSG00000188690.15	UROS	protein_coding	Iso-Switch
ENSG00000198198.16	SZT2	protein_coding	Iso-Switch	ENSG00000099875.14	MKNK2	protein_coding	Iso-Switch
ENSG00000122484.9	RPAP2	protein_coding	Iso-Switch	ENSG00000008179.19	DELE1	protein_coding	Iso-Switch
ENSG00000144057.15	STGAL2	protein_coding	Iso-Switch	ENSG00000175054.16	ATR	protein_coding	Iso-Switch
ENSG00000115760.14	BIRC6	protein_coding	Iso-Switch	ENSG00000167380.16	ZNF226	protein_coding	Iso-Switch
ENSG00000081760.17	AACS	protein_coding	Iso-Switch	ENSG00000154832.14	CXCC1	protein_coding	Iso-Switch
ENSG00000227403.2	LINC01806	lncRNA	Iso-Switch	ENSG00000249859.11	PVT1	lncRNA	Iso-Switch
ENSG00000104164.10	BLOC1S6	protein_coding	Iso-Switch	ENSG00000100284.21	TOM1	protein_coding	Iso-Switch
ENSG00000137073.21	UBAP2	protein_coding	Iso-Switch	ENSG00000176978.14	DPP7	protein_coding	Iso-Switch
ENSG00000168137.16	SETD5	protein_coding	Iso-Switch	ENSG00000160948.14	VPS28	protein_coding	Iso-Switch
ENSG00000100296.13	THOC5	protein_coding	Iso-Switch	ENSG00000234420.7	ZNF37BP	ed_processed_pse	Iso-Switch
ENSG00000104825.17	NFKBIB	protein_coding	Iso-Switch	ENSG00000131503.21	ANKHD1	protein_coding	Iso-Switch
ENSG00000173933.20	RBM4	protein_coding	Iso-Switch	ENSG00000044090.8	CUL7	protein_coding	Iso-Switch
ENSG00000174669.12	SLC29A2	protein_coding	Iso-Switch	ENSG00000138111.14	MFS13A	protein_coding	Iso-Switch
ENSG00000136247.14	ZDHH4	protein_coding	Iso-Switch	ENSG00000125967.16	NECAB3	protein_coding	Iso-Switch
ENSG00000028116.17	VRK2	protein_coding	Iso-Switch	ENSG00000101294.17	HM13	protein_coding	Iso-Switch
ENSG00000197253.13	TPSB2	protein_coding	Iso-Switch	ENSG00000099957.16	P2RX6	protein_coding	Iso-Switch
ENSG00000154889.16	MPPE1	protein_coding	Iso-Switch	ENSG00000217555.12	CKLF	protein_coding	Iso-Switch
ENSG00000158411.11	MITD1	protein_coding	Iso-Switch	ENSG00000196275.14	GTF2IRD2	protein_coding	Iso-Switch
ENSG00000213983.11	AP1G2	protein_coding	Iso-Switch	ENSG00000089335.21	ZNF302	protein_coding	Iso-Switch
ENSG00000176124.14	DLEU1	lncRNA	Iso-Switch	ENSG00000137700.18	SLC37A4	protein_coding	Iso-Switch
ENSG00000183793.13	NPIPA5	protein_coding	Iso-Switch	ENSG00000135441.7	BLOC1S1	protein_coding	Iso-Switch
ENSG00000109466.14	KLHL2	protein_coding	Iso-Switch	ENSG00000164751.15	PEX2	protein_coding	Iso-Switch
ENSG00000125354.23	Sep-06	protein_coding	Iso-Switch	ENSG00000184584.13	TMEM173	protein_coding	Iso-Switch
ENSG00000005486.17	RHBDD2	protein_coding	Iso-Switch	ENSG00000111271.14	ACAD10	protein_coding	Iso-Switch
ENSG00000132376.20	INPP5K	protein_coding	Iso-Switch	ENSG00000171055.15	FEZ2	protein_coding	Iso-Switch
ENSG00000278535.5	DHRS11	protein_coding	Iso-Switch	ENSG00000108387.14	Sep-04	protein_coding	Iso-Switch
ENSG00000172954.13	LCLAT1	protein_coding	Iso-Switch	ENSG00000140545.15	MFGE8	protein_coding	Iso-Switch
ENSG00000160445.11	ZER1	protein_coding	Iso-Switch	ENSG00000156171.14	DRAM2	protein_coding	Iso-Switch
ENSG00000264538.6	SUZ12P1	transcribed_unproc essed_pseudogene	Iso-Switch	ENSG00000126351.12	THRA	protein_coding	Iso-Switch
ENSG00000187244.11	BCAM	protein_coding	Iso-Switch	ENSG00000162910.19	MRPL55	protein_coding	Iso-Switch
ENSG00000175556.16	LONRF3	protein_coding	Iso-Switch	ENSG00000139428.12	MMAB	protein_coding	Iso-Switch
ENSG00000137177.20	KIF13A	protein_coding	Iso-Switch	ENSG00000171603.17	CLSTN1	protein_coding	Iso-Switch
ENSG00000110400.11	NECTIN1	protein_coding	Iso-Switch	ENSG00000140632.17	GLYR1	protein_coding	Iso-Switch
ENSG00000186088.16	GSAP	protein_coding	Iso-Switch	ENSG00000198169.9	ZNF251	protein_coding	Iso-Switch
ENSG00000072195.15	SPEG	protein_coding	Iso-Switch	ENSG00000102078.16	SLC25A14	protein_coding	Iso-Switch
ENSG00000150455.13	TIRAP	protein_coding	Iso-Switch	ENSG00000102287.19	GABRE	protein_coding	Iso-Switch
ENSG00000138767.13	CNOT6L	protein_coding	Iso-Switch	ENSG00000096093.16	EFHC1	protein_coding	Iso-Switch
ENSG00000205937.12	RNPS1	protein_coding	Iso-Switch	ENSG00000143569.19	UBAP2L	protein_coding	Iso-Switch
ENSG00000178149.17	DALRD3	protein_coding	Iso-Switch	ENSG00000205309.14	NTSM	protein_coding	Iso-Switch
ENSG00000205464.12	ATP6AP1L	protein_coding	Iso-Switch	ENSG00000165675.16	ENOX2	protein_coding	Iso-Switch

Appendix 7 continued...

Gene ID	Gene Name	Gene Type	Dynamic	Gene ID	Gene Name	Gene Type	Dynamic
ENSG00000147854.17	UHRF2	protein_coding	Iso-Switch	ENSG00000118495.19	PLAGL1	protein_coding	Iso-Switch
ENSG00000166025.18	AMOTL1	protein_coding	Iso-Switch	ENSG00000120318.16	ARAP3	protein_coding	Iso-Switch
ENSG00000089041.17	P2RX7	protein_coding	Iso-Switch	ENSG00000123200.16	ZC3H13	protein_coding	Iso-Switch
ENSG00000123636.18	BAZ2B	protein_coding	Iso-Switch	ENSG00000123374.11	CDK2	protein_coding	Iso-Switch
ENSG00000143416.21	SELENBP1	protein_coding	Iso-Switch	ENSG00000123562.17	MORF4L2	protein_coding	Iso-Switch
ENSG00000170776.21	AKAP13	protein_coding	Iso-Switch	ENSG00000125454.12	SLC25A19	protein_coding	Iso-Switch
ENSG00000185215.9	TNFAIP2	protein_coding	Iso-Switch	ENSG00000126453.9	BCL2L12	protein_coding	Iso-Switch
ENSG00000179632.10	MAF1	protein_coding	Iso-Switch	ENSG00000126822.17	PLEKHG3	protein_coding	Iso-Switch
ENSG00000137817.17	PARP6	protein_coding	Iso-Switch	ENSG00000129187.14	DCTD	protein_coding	Iso-Switch
ENSG00000157224.16	CLDN12	protein_coding	Iso-Switch	ENSG00000130731.16	METTL26	protein_coding	Iso-Switch
ENSG00000196792.12	STRN3	protein_coding	Iso-Switch	ENSG00000131067.17	GGT7	protein_coding	Iso-Switch
ENSG00000168228.15	ZCCHC4	protein_coding	Iso-Switch	ENSG00000132613.15	MTSS2	protein_coding	Iso-Switch
ENSG00000225697.13	SLC26A6	protein_coding	Iso-Switch	ENSG00000135426.16	TESPA1	protein_coding	Iso-Switch
ENSG00000110011.13	DNAJC4	protein_coding	Iso-Switch	ENSG00000136717.15	BIN1	protein_coding	Iso-Switch
ENSG00000048405.10	ZNF800	protein_coding	Iso-Switch	ENSG00000141068.14	KSR1	protein_coding	Iso-Switch
ENSG00000157045.9	NTAN1	protein_coding	Iso-Switch	ENSG00000142453.11	CARM1	protein_coding	Iso-Switch
ENSG00000033800.13	PIAS1	protein_coding	Iso-Switch	ENSG00000143393.16	PI4KB	protein_coding	Iso-Switch
ENSG00000109572.13	CLCN3	protein_coding	Iso-Switch	ENSG00000143537.13	ADAM15	protein_coding	Iso-Switch
ENSG00000167414.4	AC093503.1	protein_coding	Iso-Switch	ENSG00000146063.20	TRIM41	protein_coding	Iso-Switch
ENSG00000125637.15	PSD4	protein_coding	Iso-Switch	ENSG00000146067.16	FAM193B	protein_coding	Iso-Switch
ENSG00000103381.12	CPPED1	protein_coding	Iso-Switch	ENSG00000154134.15	ROBO3	protein_coding	Iso-Switch
ENSG00000197119.13	SLC25A29	protein_coding	Iso-Switch	ENSG00000157540.21	DYRK1A	protein_coding	Iso-Switch
ENSG00000059377.17	TBXAS1	protein_coding	Iso-Switch	ENSG00000158850.15	B4GALT3	protein_coding	Iso-Switch
ENSG00000127663.15	KDM4B	protein_coding	Iso-Switch	ENSG00000160256.13	FAM207A	protein_coding	Iso-Switch
ENSG00000009830.11	POMT2	protein_coding	Iso-Switch	ENSG00000160703.16	NLRX1	protein_coding	Iso-Switch
ENSG00000005243.10	COPZ2	protein_coding	Iso-Switch	ENSG00000160796.17	NBEAL2	protein_coding	Iso-Switch
ENSG00000144560.15	VGLL4	protein_coding	Iso-Switch	ENSG00000164649.20	CDCA7L	protein_coding	Iso-Switch
ENSG00000122515.15	ZMIZ2	protein_coding	Iso-Switch	ENSG00000165699.14	TSC1	protein_coding	Iso-Switch
ENSG00000142046.15	TMEM91	protein_coding	Iso-Switch	ENSG00000167766.18	ZNF83	protein_coding	Iso-Switch
ENSG00000160193.11	WDR4	protein_coding	Iso-Switch	ENSG00000168066.20	SF1	protein_coding	Iso-Switch
ENSG00000162390.17	ACOT11	protein_coding	Iso-Switch	ENSG00000168488.18	ATXN2L	protein_coding	Iso-Switch
ENSG00000253352.9	TUG1	lncRNA	Iso-Switch	ENSG00000168610.14	STAT3	protein_coding	Iso-Switch
ENSG00000108509.21	CAMTA2	protein_coding	Iso-Switch	ENSG00000169592.15	INO80E	protein_coding	Iso-Switch
ENSG00000108556.10	CHRNE	protein_coding	Iso-Switch	ENSG00000173264.14	GPR137	protein_coding	Iso-Switch
ENSG00000127483.18	HP1BP3	protein_coding	Iso-Switch	ENSG00000174373.16	RALGAP1	protein_coding	Iso-Switch
ENSG00000140553.17	UNC45A	protein_coding	Iso-Switch	ENSG00000175061.18	SNHG29	lncRNA	Iso-Switch
ENSG00000124155.18	PIGT	protein_coding	Iso-Switch	ENSG00000175387.16	SMAD2	protein_coding	Iso-Switch
ENSG00000214279.13	SCART1	protein_coding	Iso-Switch	ENSG00000175538.10	KCNE3	protein_coding	Iso-Switch
ENSG00000167703.14	SLC43A2	protein_coding	Iso-Switch	ENSG00000177565.16	TBL1XR1	protein_coding	Iso-Switch
ENSG00000151276.23	MAG11	protein_coding	Iso-Switch	ENSG00000177595.18	PIDD1	protein_coding	Iso-Switch
ENSG00000167613.16	LAIR1	protein_coding	Iso-Switch	ENSG00000177830.17	CHID1	protein_coding	Iso-Switch
ENSG00000119729.12	RHOQ	protein_coding	Iso-Switch	ENSG00000178802.18	MPI	protein_coding	Iso-Switch
ENSG00000006025.12	AC003665.1	protein_coding	Iso-Switch	ENSG00000179134.15	SAMD4B	protein_coding	Iso-Switch
ENSG00000006282.21	SPATA20	protein_coding	Iso-Switch	ENSG00000179950.13	PUF60	protein_coding	Iso-Switch
ENSG00000007372.23	PAX6	protein_coding	Iso-Switch	ENSG00000180902.18	D2HGDH	protein_coding	Iso-Switch
ENSG00000048991.16	R3HDM1	protein_coding	Iso-Switch	ENSG00000181135.16	ZNF707	protein_coding	Iso-Switch
ENSG000000067182.7	TNFRSF1A	protein_coding	Iso-Switch	ENSG00000182179.13	UBA7	protein_coding	Iso-Switch
ENSG000000071054.16	MAP4K4	protein_coding	Iso-Switch	ENSG00000183963.18	SMTN	protein_coding	Iso-Switch
ENSG000000075711.21	DLG1	protein_coding	Iso-Switch	ENSG00000185619.18	PCGF3	protein_coding	Iso-Switch
ENSG000000080822.17	CLDND1	protein_coding	Iso-Switch	ENSG00000186866.16	POFUT2	protein_coding	Iso-Switch
ENSG000000084234.17	APLP2	protein_coding	Iso-Switch	ENSG00000187239.17	FNBP1	protein_coding	Iso-Switch
ENSG00000099949.21	LZTR1	protein_coding	Iso-Switch	ENSG00000188517.16	COL25A1	protein_coding	Iso-Switch
ENSG00000100023.19	PPIL2	protein_coding	Iso-Switch	ENSG00000196923.14	PDLIM7	protein_coding	Iso-Switch
ENSG00000100379.17	KCTD17	protein_coding	Iso-Switch	ENSG00000198467.15	TPM2	protein_coding	Iso-Switch
ENSG00000105048.17	TNNT1	protein_coding	Iso-Switch	ENSG00000198836.10	OPA1	protein_coding	Iso-Switch
ENSG00000105983.21	LMBR1	protein_coding	Iso-Switch	ENSG00000204482.10	LST1	protein_coding	Iso-Switch
ENSG00000109332.20	UBE2D3	protein_coding	Iso-Switch	ENSG00000223482.8	NUTM2A-AS1	lncRNA	Iso-Switch
ENSG00000114956.20	DGUOK	protein_coding	Iso-Switch	ENSG00000250571.7	GLI4	protein_coding	Iso-Switch

References

1. Zhang, Y., Gao, S., Xia, J. & Liu, F. Hematopoietic Hierarchy – An Updated Roadmap. *Trends in Cell Biology* vol. 28 976–986 Preprint at <https://doi.org/10.1016/j.tcb.2018.06.001> (2018).
2. Psaila, B. & Mead, A. J. Single-cell approaches reveal novel cellular pathways for megakaryocyte and erythroid differentiation. *Blood* **133**, 1427–1435 (2019).
3. Khoury, J. D. *et al.* The 5th edition of the World Health Organization Classification of Haematolymphoid Tumours: Myeloid and Histiocytic/Dendritic Neoplasms. *Leukemia* vol. 36 1703–1719 Preprint at <https://doi.org/10.1038/s41375-022-01613-1> (2022).
4. Bavaro, L., Martelli, M., Cavo, M. & Soverini, S. Mechanisms of disease progression and resistance to tyrosine kinase inhibitor therapy in chronic myeloid leukemia: An update. *International Journal of Molecular Sciences* vol. 20 Preprint at <https://doi.org/10.3390/ijms20246141> (2019).
5. Passamonti, F. & Maffioli, M. *Update from the Latest WHO Classification of MPNs: A User's Manual.* (2016).
6. Szuber, N. *et al.* 3023 Mayo Clinic Patients With Myeloproliferative Neoplasms: Risk-Stratified Comparison of Survival and Outcomes Data Among Disease Subgroups. *Mayo Clin Proc* **94**, 599–610 (2019).
7. Titmarsh, G. J. *et al.* How common are myeloproliferative neoplasms? A systematic review and meta-analysis. *Am J Hematol* **89**, 581–587 (2014).
8. Tefferi, A. *et al.* Long-term survival and blast transformation in molecularly annotated essential thrombocythemia, polycythemia vera, and myelofibrosis. *Blood* **124**, 2507–2513 (2014).
9. Dunbar, A. J., Rampal, R. K. & Levine, R. Leukemia secondary to myeloproliferative neoplasms. *Blood* **136**, 61–70 (2020).
10. Mead, A. J. & Mullally, A. Myeloproliferative neoplasm stem cells. *Blood* **129**; **12**, (2017).
11. Sanjuan-Pla, A. *et al.* Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature* **502**, 232–236 (2013).
12. Yamamoto, R. *et al.* Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. *Cell* **154**, (2013).
13. Royston, D., Mead, A. J. & Psaila, B. Application of Single-Cell Approaches to Study Myeloproliferative Neoplasm Biology. *Hematology/Oncology Clinics of North America* vol. 35 279–293 Preprint at <https://doi.org/10.1016/j.hoc.2021.01.002> (2021).
14. Hu, X., li, J., Fu, M., Zhao, X. & Wang, W. The JAK/STAT signaling pathway: from bench to clinic. *Signal Transduction and Targeted Therapy* vol. 6 Preprint at <https://doi.org/10.1038/s41392-021-00791-1> (2021).
15. Akada, H. *et al.* Critical role of Jak2 in the maintenance and function of adult hematopoietic stem cells. *Stem Cells* **32**, 1878–1889 (2014).
16. Radosevic, N. *et al.* JAK2 contributes to the intrinsic capacity of primary hematopoietic cells to respond to stem cell factor. *Exp Hematol* **32**, 149–156 (2004).

17. Oded. Livnah *et al.* Crystallographic Evidence for Preformed Dimers of Erythropoietin Receptor Before Ligand Activation. *Science* (1979) 986–990 (1999).
18. Fasouli, E. S. & Katsantoni, E. JAK-STAT in Early Hematopoiesis and Leukemia. *Frontiers in Cell and Developmental Biology* vol. 9 Preprint at <https://doi.org/10.3389/fcell.2021.669363> (2021).
19. Cabagnols, X. *et al.* Presence of atypical thrombopoietin receptor (MPL) mutations in triple-negative essential thrombocythemia patients. *Blood* **127**, 333–342 (2016).
20. Chen, E. & Mullally, A. How does JAK2V617F contribute to the pathogenesis of myeloproliferative neoplasms? *Hematology Am Soc Hematol Educ Program* 268–276 (2014).
21. Klampfl, T. *et al.* Somatic Mutations of Calreticulin in Myeloproliferative Neoplasms. *New England Journal of Medicine* **369**, 2379–2390 (2013).
22. Nangalia, J. *et al.* Somatic CALR Mutations in Myeloproliferative Neoplasms with Nonmutated JAK2. *New England Journal of Medicine* **369**, 2391–2405 (2013).
23. Elf, S. *et al.* Mutant calreticulin requires both its mutant C-terminus and the thrombopoietin receptor for oncogenic transformation. *Cancer Discov* **6**, 368–381 (2016).
24. Chachoua, I. *et al.* Thrombopoietin receptor activation by myeloproliferative neoplasm associated calreticulin mutants. *Blood* **127**; **10**, 1325–1335 (2016).
25. Qian, H. *et al.* Critical Role of Thrombopoietin in Maintaining Adult Quiescent Hematopoietic Stem Cells. *Cell Stem Cell* **1**, 671–684 (2007).
26. Kaushansky, K. & St, D. *Lineage-Specific Hematopoietic Growth Factors*. *N Engl J Med* vol. 354 www.nejm.org (2006).
27. Feenstra, J. D. M. *et al.* Whole-exome sequencing identifies novel MPL and JAK2 mutations in triple-negative myeloproliferative neoplasms. *Blood* **127**; **3**, (2016).
28. Rampal, R. *et al.* Integrated genomic analysis illustrates the central role of JAK-STAT pathway activation in myeloproliferative neoplasm pathogenesis. *Blood* **123**, (2014).
29. Ortmann, C. A. *et al.* Effect of Mutation Order on Myeloproliferative Neoplasms. *New England Journal of Medicine* **372**, 601–612 (2015).
30. Grinfeld, J. *et al.* Classification and Personalized Prognosis in Myeloproliferative Neoplasms. *New England Journal of Medicine* **379**, 1416–1430 (2018).
31. Leroy, B., Anderson, M. & Soussi, T. TP53 mutations in human cancer: Database reassessment and prospects for the next decade. *Human Mutation* vol. 35 672–688 Preprint at <https://doi.org/10.1002/humu.22552> (2014).
32. Prior, I. A., Hood, F. E. & Hartley, J. L. The frequency of ras mutations in cancer. *Cancer Res* **80**, 2669–2974 (2020).
33. Jyoti Nangalia *et al.* DNMT3A mutations occur early or late in patients with myeloproliferative neoplasms and mutation order influences phenotype. *Haematologica* **100**, e438–e442 (2015).
34. Klampfl, T. *et al.* Genome integrity of myeloproliferative neoplasms in chronic phase and during disease progression. *Blood* **118**, 167–176 (2011).
35. Rumi, E. *et al.* Identification of genomic aberrations associated with disease transformation by means of high-resolution SNP array analysis in patients with myeloproliferative neoplasm. *Am J Hematol* **86**, 974–979 (2011).

36. Wang, L., Wheeler, D. A. & Prchal, J. T. Acquired uniparental disomy of chromosome 9p in hematologic malignancies. *Experimental Hematology* vol. 44 644–652 Preprint at <https://doi.org/10.1016/j.exphem.2015.11.005> (2016).
37. Kralovics, R., Guan, Y. & Prchal, J. T. Acquired uniparental disomy of chromosome 9p is a frequent stem cell defect in polycythemia vera. *Exp Hematol* **30**, 229–236 (2002).
38. Landgren, O. *et al.* Increased risks of polycythemia vera, essential thrombocythemia, and myelofibrosis among 24 577 first-degree relatives of 11 039 patients with myeloproliferative neoplasms in Sweden. *Blood* **112**, 2199–2204 (2008).
39. Jones, A. V. *et al.* JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. *Nat Genet* **41**, 446–449 (2009).
40. Lancman, G., Brunner, A., Hoffman, R., Mascarenhas, J. & Hobbs, G. Outcomes and predictors of survival in blast phase myeloproliferative neoplasms. *Leuk Res* **70**, 49–55 (2018).
41. Tefferi, A. *et al.* Blast phase myeloproliferative neoplasm: Mayo-AGIMM study of 410 patients from two separate cohorts. *Leukemia* **32**, 1200–1210 (2018).
42. Bewersdorf, J. P. & Rampal, R. K. Hitting the brakes on accelerated and blast-phase myeloproliferative neoplasms: current and emerging concepts. *Hematology Am Soc Hematol Educ Program* 218–224 (2022).
43. Guglielmelli, P. *et al.* JOURNAL OF CLINICAL ONCOLOGY MIPSS70: Mutation-Enhanced International Prognostic Score System for Transplantation-Age Patients With Primary Myelofibrosis. *J Clin Oncol* **36**, 310–318 (2017).
44. Caramazza, D. *et al.* Refined cytogenetic-risk categorization for overall and leukemia-free survival in primary myelofibrosis: A single center study of 433 patients. *Leukemia* **25**, 82–88 (2011).
45. Vannucchi, A. M. *et al.* Mutations and prognosis in primary myelofibrosis. *Leukemia* **27**, 1861–1869 (2013).
46. Tefferi, A. *et al.* Mutation-enhanced international prognostic systems for essential thrombocythaemia and polycythaemia vera. *Br J Haematol* **189**, 291–302 (2020).
47. Tefferi, A. *et al.* Survival and prognosis among 1545 patients with contemporary polycythemia vera: An international study. *Leukemia* **27**, 1874–1881 (2013).
48. Barraco, D. *et al.* Cytogenetic findings in WHO-defined polycythaemia vera and their prognostic relevance. *British Journal of Haematology* vol. 182 437–440 Preprint at <https://doi.org/10.1111/bjh.14798> (2018).
49. Gangat, N. *et al.* Cytogenetic abnormalities in essential thrombocythemia: Clinical and molecular correlates and prognostic relevance in 809 informative cases. *Blood Cancer J* **12**, (2022).
50. O'sullivan, J. M. *et al.* The poor outcome in high molecular risk, hydroxycarbamide-resistant/intolerant ET is not ameliorated by ruxolitinib. *Blood* **134**, 2107–2111 (2019).
51. Berk, P. D., Wasserman, L. R., Fruchtman, S. M. & Goldberg, J. D. *Treatment of Polycythemia Vera: A Summary of Clinical Trials Conducted by the Polycythemia Vera Study Group*. (WB Saunders Co Ltd (17 Oct. 1994), Philadelphia, PA, USA, 1995).
52. Kiladjian, J. J., Chevret, S., Dosquet, C., Chomienne, C. & Rain, J. D. Treatment of polycythemia vera with hydroxyurea and pipobroman: Final results of a randomized trial initiated in 1980. *Journal of Clinical Oncology* **29**, 3907–3913 (2011).

53. Nand, S., Stock, W., Godwin, J. & Fisher, S. G. Leukemogenic risk of hydroxyurea therapy in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Am J Hematol* **52**, 42–46 (1996).
54. Wang, R. *et al.* Second malignancies among older patients with classical myeloproliferative neoplasms treated with hydroxyurea. *Blood Adv* **7**, 734–743 (2023).
55. Björkholm, M. *et al.* Treatment-related risk factors for transformation to acute myeloid leukemia and myelodysplastic syndromes in myeloproliferative neoplasms. *Journal of Clinical Oncology* **29**, 2410–2415 (2011).
56. Guido Finazzi, Marco Ruggeri, Francesco Rodeghiero & Tiziano Barbui. Efficacy and safety of long-term use of hydroxyurea in young patients with essential thrombocythemia and a high risk of thrombosis. *Blood* **101**, (2003).
57. Timothy J. Ley, M. D. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *N ENGL J MED* **368**, 2059–2074 (2013).
58. Papaemmanuil, E. *et al.* Genomic Classification and Prognosis in Acute Myeloid Leukemia. *New England Journal of Medicine* **374**, 2209–2221 (2016).
59. McNamara, C. J. *et al.* The mutational landscape of accelerated- and blast-phase myeloproliferative neoplasms impacts patient outcomes. *Blood Adv* **2**, 2658–2671 (2018).
60. Rumi, E. *et al.* Identification of genomic aberrations associated with disease transformation by means of high-resolution SNP array analysis in patients with myeloproliferative neoplasm. *Am J Hematol* **86**, 974–979 (2011).
61. Klampfl, T. *et al.* Genome integrity of myeloproliferative neoplasms in chronic phase and during disease progression. *Blood* **118**, 167–176 (2011).
62. Thoennissen, N. H. *et al.* Prevalence and prognostic impact of allelic imbalances associated with leukemic transformation of Philadelphia chromosome-negative myeloproliferative neoplasms. *Blood* **115**, 2882–2890 (2010).
63. Rodriguez-Meira, A. *et al.* Single-cell multi-omics identifies chronic inflammation as a driver of TP53-mutant leukemic evolution. *Nat Genet* **55**, 1531–1541 (2023).
64. Rodriguez-Meira, A. *et al.* Unravelling Intratumoral Heterogeneity through High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing. *Mol Cell* **73**, 1292-1305.e8 (2019).
65. Nam, A. S. *et al.* Somatic mutations and cell identity linked by Genotyping of Transcriptomes. *Nature* **571**, 355–360 (2019).
66. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* **14**, 865–868 (2017).
67. Turkalj, S. *et al.* GTAC enables parallel genotyping of multiple genomic loci with chromatin accessibility profiling in single cells. *Cell Stem Cell* **30**, 722-740.e11 (2023).
68. Izzo, F. *et al.* Mapping genotypes to chromatin accessibility profiles in single cells. *Nature* **629**, 1149–1157 (2024).
69. Psaila, B. *et al.* Single-Cell Analyses Reveal Megakaryocyte-Biased Hematopoiesis in Myelofibrosis and Identify Mutant Clone-Specific Targets. *Mol Cell* **78**, 477-492.e8 (2020).
70. Cahu, X. *et al.* Allo-SCT for Philadelphia-negative myeloproliferative neoplasms in blast phase: A study from the Societe Française de Greffe de Moelle et de Therapie Cellulaire (SFGM-TC). *Bone Marrow Transplant* **49**, 756–760 (2014).

71. McNamara, C. J. *et al.* The mutational landscape of accelerated- and blast-phase myeloproliferative neoplasms impacts patient outcomes. *Blood Adv* **2**, 2658–2671 (2018).
72. Thepot, S. *et al.* Treatment of progression of Philadelphia-negative myeloproliferative neoplasms to myelodysplastic syndrome or acute myeloid leukemia by azacitidine: A report on 54 cases on the behalf of the Groupe Francophone des Myelodysplasies (GFM). *Blood* **116**, 3735–3742 (2010).
73. Badar, T. *et al.* Therapeutic benefit of decitabine, a hypomethylating agent, in patients with high-risk primary myelofibrosis and myeloproliferative neoplasm in accelerated or blastic/acute myeloid leukemia phase. *Leuk Res* **39**, 950–956 (2015).
74. Venton, G. *et al.* Impact of gene mutations on treatment response and prognosis of acute myeloid leukemia secondary to myeloproliferative neoplasms. *Am J Hematol* **93**, 330–338 (2018).
75. Dinardo, C. D. *et al.* Venetoclax combined with decitabine or azacitidine in treatment-naïve, elderly patients with acute myeloid leukemia. *Blood* **133**; 1, (2019).
76. Pollyea, D. A. *et al.* Venetoclax with azacitidine disrupts energy metabolism and targets leukemia stem cells in patients with acute myeloid leukemia. *Nat Med* **24**, 1859–1866 (2018).
77. Gangat, N. *et al.* Venetoclax with azacitidine or decitabine in blast-phase myeloproliferative neoplasm: A multicenter series of 32 consecutive cases. *Am J Hematol* **96**, 781–789 (2021).
78. Pemmaraju, N. *et al.* Ten years after ruxolitinib approval for myelofibrosis: a review of clinical efficacy. *Leukemia and Lymphoma* vol. 64 1063–1081 Preprint at <https://doi.org/10.1080/10428194.2023.2196593> (2023).
79. Eghtedar, A. *et al.* Phase 2 study of the JAK kinase inhibitor ruxolitinib in patients with refractory leukemias, including postmyeloproliferative neoplasm acute myeloid leukemia. *Blood* **119**, 4614–4618 (2012).
80. Rampal, R. *et al.* Genomic and functional analysis of leukemic transformation of myeloproliferative neoplasms. *Proc Natl Acad Sci U S A* **111**, E5401–E5410 (2014).
81. Mascarenhas, J. O. *et al.* Phase 2 study of ruxolitinib and decitabine in patients with myeloproliferative neoplasm in accelerated and blast phase. *Blood Adv* **4**, 5246–5256 (2020).
82. Rampal, R. K. *et al.* Safety and efficacy of combined ruxolitinib and decitabine in accelerated and blast-phase myeloproliferative neoplasms. *Blood Adv* **2**, 3572–3580 (2018).
83. Rampal, R. *et al.* An open-label, multicenter, phase 1b/2 study of navtemadlin (KRT-232) in patients with relapsed/refractory acute myeloid leukemia secondary to myeloproliferative neoplasms. *Journal of Clinical Oncology* **40**, TPS7063–TPS7063 (2022).
84. Wang, X., Hu, C. S., Gillespie, V., Krejsa, C. M. & Hoffman, R. Navtemadlin (KRT-232), a Small Molecule MDM2 Inhibitor, Is More Effective Than Decitabine Against Myeloproliferative Neoplasm-Blast Phase in a Patient-Derived Xenograft Model. *Blood* **138**, 3591 (2021).
85. Rampal, R. K. *et al.* LSD1 Inhibitor CPI-482 Shows Efficacy and Prolongs Survival in Mouse Models of AML and Post-MPN AML in the Context of Constitutive JAK-STAT Pathway Activation. *Blood* **136**, 50–51 (2020).

86. Levine, A. J. & Oren, M. The first 30 years of p53: Growing ever more complex. *Nature Reviews Cancer* vol. 9 749–758 Preprint at <https://doi.org/10.1038/nrc2723> (2009).
87. Shiv Srivastava, Zhiqiang Zou, Kathleen Pirollo, William Blattner & Esther H. Chang. Germ-line transmission of a mutated p53 gene in a cancer-prone family with Li-Fraumeni syndrome. *Nature* **348**, 747–749 (1990).
88. Lawrence A. Donehower *et al.* Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* **356**, 215–221 (1992).
89. Soussi, T. & Wiman, K. G. TP53: An oncogene in disguise. *Cell Death and Differentiation* vol. 22 1239–1249 Preprint at <https://doi.org/10.1038/cdd.2015.53> (2015).
90. Kasthuber, E. R. & Lowe, S. W. Putting p53 in Context. *Cell* vol. 170 1062–1078 Preprint at <https://doi.org/10.1016/j.cell.2017.08.028> (2017).
91. Aubrey, B. J., Kelly, G. L., Janic, A., Herold, M. J. & Strasser, A. How does p53 induce apoptosis and how does this relate to p53-mediated tumour suppression? *Cell Death and Differentiation* vol. 25 104–113 Preprint at <https://doi.org/10.1038/cdd.2017.169> (2018).
92. Abuetaf, Y. *et al.* DNA damage response revisited: the p53 family and its regulators provide endless cancer therapy opportunities. *Experimental and Molecular Medicine* vol. 54 1658–1669 Preprint at <https://doi.org/10.1038/s12276-022-00863-4> (2022).
93. Chen, J. The cell-cycle arrest and apoptotic functions of p53 in tumor initiation and progression. *Cold Spring Harb Perspect Med* **6**, (2016).
94. Williams, A. B. & Schumacher, B. p53 in the DNA-damage-repair process. *Cold Spring Harb Perspect Med* **6**, (2016).
95. Aubrey, B. J., Kelly, G. L., Janic, A., Herold, M. J. & Strasser, A. How does p53 induce apoptosis and how does this relate to p53-mediated tumour suppression? *Cell Death and Differentiation* vol. 25 104–113 Preprint at <https://doi.org/10.1038/cdd.2017.169> (2018).
96. Rizzotto, D. *et al.* Nutlin-Induced Apoptosis Is Specified by a Translation Program Regulated by PCBP2 and DHX30. *Cell Rep* **30**, 4355–4369.e6 (2020).
97. Liu, Y., Su, Z., Tavana, O. & Gu, W. Understanding the complexity of p53 in a new era of tumor suppression. *Cancer Cell* vol. 42 946–967 Preprint at <https://doi.org/10.1016/j.ccell.2024.04.009> (2024).
98. Giacomelli, A. O. *et al.* Mutational processes shape the landscape of TP53 mutations in human cancer. *Nature Genetics* vol. 50 1381–1387 Preprint at <https://doi.org/10.1038/s41588-018-0204-y> (2018).
99. Pfister, N. T. & Prives, C. Transcriptional regulation by wild-type and cancer-related mutant forms of p53. *Cold Spring Harb Perspect Med* **7**, (2017).
100. Chen, X. *et al.* Mutant p53 in cancer: from molecular mechanism to therapeutic modulation. *Cell Death and Disease* vol. 13 Preprint at <https://doi.org/10.1038/s41419-022-05408-1> (2022).
101. Ventura, A. *et al.* Restoration of p53 function leads to tumour regression in vivo. *Nature* **445**, 661–665 (2007).
102. Hassin, O. & Oren, M. Drugging p53 in cancer: one protein, many targets. *Nature Reviews Drug Discovery* vol. 22 127–144 Preprint at <https://doi.org/10.1038/s41573-022-00571-8> (2023).
103. Yu, X., Vazquez, A., Levine, A. J. & Carpizo, D. R. Allele-Specific p53 Mutant Reactivation. *Cancer Cell* **21**, 614–625 (2012).

104. Lawrence A. Donehower *et al.* Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* 215–221 (1992).
105. Liu, Y. *et al.* p53 Regulates Hematopoietic Stem Cell Quiescence. *Cell Stem Cell* **4**, 37–48 (2009).
106. Chen, J. *et al.* Enrichment of hematopoietic stem cells with SLAM and LSK markers for the detection of hematopoietic stem cell function in normal and Trp53 null mice. *Exp Hematol* **36**, 1236–1243 (2008).
107. TeKippe, M., Harrison, D. E. & Chen, J. Expansion of hematopoietic stem cell phenotype and activity in Trp53-null mice. in *Experimental Hematology* vol. 31 521–527 (Elsevier Inc., 2003).
108. Liu, D. *et al.* Puma is required for p53-induced depletion of adult stem cells. *Nat Cell Biol* **12**, 993–998 (2010).
109. Dumble, M. *et al.* The impact of altered p53 dosage on hematopoietic stem cell dynamics during aging. *Blood* **109**, 1736–1742 (2007).
110. Wang, Y. V. *et al.* Fine-tuning p53 activity through C-terminal modification significantly contributes to HSC homeostasis and mouse radiosensitivity. *Genes Dev* **25**, 1426–1438 (2011).
111. Nahi, H. *et al.* Mutated and non-mutated TP53 as targets in the treatment of leukaemia. *Br J Haematol* **141**, 445–453 (2008).
112. Agirre, X. *et al.* TP53 Is Frequently Altered by Methylation, Mutation, and/or Deletion in Acute Lymphoblastic Leukaemia. *Mol Carcinog* **38**, 201–208 (2003).
113. Avet-Loiseau, H. *et al.* P53 deletion is not a frequent event in multiple myeloma. *Br J Haematol* **106**, 717–719 (1999).
114. Arber, D. A. *et al.* International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. *Blood* **140**, 1200–1228 (2022).
115. Haferlach, C. *et al.* Mutations of the TP53 gene in acute myeloid leukemia are strongly associated with a complex aberrant karyotype. *Leukemia* **22**, 1539–1541 (2008).
116. Abel, H. J. *et al.* Genomic landscape of TP53-mutated myeloid malignancies. *Blood Adv* **7**, 4586–4598 (2023).
117. Santini, V., Stahl, M. & Sallman, D. A. TP53 Mutations in Acute Leukemias and Myelodysplastic Syndromes: Insights and Treatment Updates . *American Society of Clinical Oncology Educational Book* **44**, (2024).
118. Lindsley, R. C. *et al.* Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. *Blood* **125**, 1367–1376 (2015).
119. Bernard, E. *et al.* Implications of TP53 allelic state for genome stability, clinical presentation and outcomes in myelodysplastic syndromes. *Nat Med* **26**, 1549–1556 (2020).
120. Harutyunyan Ashot, Klampfl Thorsten, Cazzola Mario & Kralovics Robert. p53 Lesions in Leukemic Transformation. *N ENGL J MED* **364**, 488–490 (2011).
121. Paz, D. L. *et al.* Leukemic evolution of polycythemia vera and essential thrombocythemia: Genomic profiles predict time to transformation. *Blood Adv* **4**, 4887–4897 (2020).
122. Kubesova, B. *et al.* Low-burden TP53 mutations in chronic phase of myeloproliferative neoplasms: Association with age, hydroxyurea administration, disease type and JAK2 mutational status. *Leukemia* **32**, 450–461 (2018).
123. Morita, K. *et al.* Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics. *Nat Commun* **11**, (2020).

124. Miles, L. A. *et al.* Single-cell mutation analysis of clonal evolution in myeloid malignancies. *Nature* **587**, 477–482 (2020).
125. Jacobsen, F. W., Stokke, T. & Jacobsen, S. E. W. Transforming Growth Factor- β Potently Inhibits the Viability-Promoting Activity of Stem Cell Factor and Other Cytokines and Induces Apoptosis of Primitive Murine Hematopoietic Progenitor Cells. *Blood* **86**, (1995).
126. Bryder, D. *et al.* Self-Renewal of Multipotent Long-Term Repopulating Hematopoietic Stem Cells Is Negatively Regulated by Fas and Tumor Necrosis Factor Receptor Activation. *J. Exp. Med* **194**, 941–952 (2001).
127. Essers, M. A. G. *et al.* IFN α activates dormant haematopoietic stem cells in vivo. *Nature* **458**, 904–908 (2009).
128. Walter, D. *et al.* Exit from dormancy provokes DNA-damage-induced attrition in haematopoietic stem cells. *Nature* **520**, 549–552 (2015).
129. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative CT method. *Nat Protoc* **3**, 1101–1108 (2008).
130. Doudna, J. A. & Charpentier, E. *The New Frontier of Genome Engineering with CRISPR-Cas9*. <https://www.science.org>.
131. Gonçalves, E. *et al.* Minimal genome-wide human CRISPR-Cas9 library. *Genome Biol* **22**, (2021).
132. Heckl, D. *et al.* Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nat Biotechnol* **32**, 941–946 (2014).
133. Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera, M. D. C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat Biotechnol* **32**, 267–273 (2014).
134. Henkel, L., Rauscher, B., Schmitt, B., Winter, J. & Boutros, M. Genome-scale CRISPR screening at high sensitivity with an empirically designed sgRNA library. *BMC Biol* **18**, (2020).
135. Eekels, J. J. M. *et al.* A competitive cell growth assay for the detection of subtle effects of gene transduction on cell proliferation. *Gene Ther* **19**, 1058–1064 (2012).
136. Olijnik, A. A. *et al.* Generating human bone marrow organoids for disease modeling and drug discovery. *Nat Protoc* **19**, 2117–2146 (2024).
137. Khan, A. O. *et al.* Human Bone Marrow Organoids for Disease Modeling, Discovery, and Validation of Therapeutic Targets in Hematologic Malignancies. *Cancer Discov* **13**, 364–385 (2023).
138. Olijnik, A. A. *et al.* Generating human bone marrow organoids for disease modeling and drug discovery. *Nat Protoc* **19**, 2117–2146 (2024).
139. Rodriguez-Meira, A., O’Sullivan, J., Rahman, H. & Mead, A. J. TARGET-Seq: A Protocol for High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing. *STAR Protoc* **1**, (2020).
140. Tsherniak, A. *et al.* Defining a Cancer Dependency Map. *Cell* **170**, 564–576.e16 (2017).
141. DepMap, B. DepMap 24Q2 Public. (2024) doi:10.25452/figshare.plus.25880521.v1.
142. Nusinow, D. P. *et al.* Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell* **180**, 387–402.e16 (2020).
143. Aleksander, S. A. *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224**, (2023).

144. Ashburner, M. *et al.* *Gene Ontology: Tool for the Unification of Biology The Gene Ontology Consortium**. <http://www.flybase.bio.indiana.edu> (2000).
145. Thomas, P. D. *et al.* PANTHER: Making genome-scale phylogenetics accessible to all. *Protein Science* vol. 31 8–22 Preprint at <https://doi.org/10.1002/pro.4218> (2022).
146. Carbon, S. & Mungall, C. Gene Ontology Data Archive. Preprint at <https://doi.org/10.5281/zenodo.7942786> (2023).
147. Tyner, J. W. *et al.* Functional genomic landscape of acute myeloid leukaemia. *Nature* **562**, 526–531 (2018).
148. Bottomly, D. *et al.* Integrative analysis of drug response and clinical outcome in acute myeloid leukemia. *Cancer Cell* **40**, 850–864.e9 (2022).
149. Grossman, R. L. *et al.* Toward a Shared Vision for Cancer Genomic Data. *New England Journal of Medicine* **375**, 1109–1112 (2016).
150. de Bruijn, I. *et al.* Analysis and Visualization of Longitudinal Genomic and Clinical Data from the AACR Project GENIE Biopharma Collaborative in cBioPortal. *Cancer Res* **83**, 3861–3867 (2023).
151. Cerami, E. *et al.* The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401–404 (2012).
152. Martin, M. Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads. *EMBnet Journal*. **17**, 10–12 (2011).
153. Li, W. *et al.* Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biol* **16**, (2015).
154. Li, W. *et al.* MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol* **15**, 554 (2014).
155. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, (2014).
156. Subramanian, A. *et al.* *Gene Set Enrichment Analysis: A Knowledge-Based Approach for Interpreting Genome-Wide Expression Profiles*. www.pnas.org/cgi/doi/10.1073/pnas.0506580102 (2005).
157. Liberzon, A. *et al.* The Molecular Signatures Database Hallmark Gene Set Collection. *Cell Syst* **1**, 417–425 (2015).
158. Schoggins, J. W. *et al.* A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* **472**, 481–485 (2011).
159. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
160. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
161. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. DeepTools: A flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**, (2014).
162. Nassar, L. R. *et al.* The UCSC Genome Browser database: 2023 update. *Nucleic Acids Res* **51**, D1188–D1195 (2023).
163. Kovaka, S. *et al.* Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol* **20**, (2019).
164. Harrow, J. *et al.* GENCODE: The reference human genome annotation for the ENCODE project. *Genome Res* **22**, 1760–1774 (2012).
165. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
166. Wen, W. X., Mead, A. J. & Thongjuea, S. MARVEL: an integrated alternative splicing analysis platform for single-cell RNA sequencing data. *Nucleic Acids Res* **51**, (2023).

167. Wen, S. BRACE: A novel Bayesian-based imputation approach for dimension reduction analysis of alternative splicing at single-cell resolution. *bioRxiv* (2024) doi:10.1101/2024.08.01.606201.
168. Gerstberger, S., Hafner, M. & Tuschl, T. A census of human RNA-binding proteins. *Nat Rev Genet* **15**, 829–845 (2014).
169. Ascano, M., Gerstberger, S. & Tuschl, T. Multi-disciplinary methods to define RNA-protein interactions and regulatory networks. *Current Opinion in Genetics and Development* vol. 23 20–28 Preprint at <https://doi.org/10.1016/j.gde.2013.01.003> (2013).
170. And, M. & Biology, C. *Characterization of Heterogeneous Nuclear RNA-Protein Complexes In Vivo with Monoclonal Antibodies*. (1984).
171. Pifiol-Roma, S., Choi, Y. Do, Matunis, M. J. & Dreyfuss, G. Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. *Genes Dev* **2**, 215–227 (1988).
172. Tenenbaum, S. A., Carson, C. C., Lager, P. J. & Keene, J. D. Identifying mRNA subsets in messenger ribonucleoprotein complexes by using cDNA arrays. *PNAS* **97**; **26**, 14085–14090 (2000).
173. Castello, A. *et al.* Insights into RNA Biology from an Atlas of Mammalian mRNA-Binding Proteins. *Cell* **149**, 1393–1406 (2012).
174. Baltz, A. G. *et al.* The mRNA-Bound Proteome and Its Global Occupancy Profile on Protein-Coding Transcripts. *Mol Cell* **46**, 674–690 (2012).
175. Smith, J. M., Sandow, J. J. & Webb, A. I. The search for RNA-binding proteins: A technical and interdisciplinary challenge. *Biochemical Society Transactions* vol. 49 393–403 Preprint at <https://doi.org/10.1042/BST20200688> (2021).
176. Rahman, R., Xu, W., Jin, H. & Rosbash, M. Identification of RNA-binding protein targets with HyperTRIBE. *Nat Protoc* **13**, 1829–1849 (2018).
177. Xu, W., Rahman, R. & Rosbash, M. Mechanistic implications of enhanced editing by a HyperTRIBE RNA-binding protein. *RNA* **24**, 173–182 (2018).
178. Nguyen, D. T. T. *et al.* HyperTRIBE uncovers increased MUSASHI-2 RNA binding activity and differential regulation in leukemic stem cells. *Nat Commun* **11**, (2020).
179. Meyer, K. D. DART-seq: an antibody-free method for global m6A detection. *Nat Methods* **16**, 1275–1280 (2019).
180. Van Nostrand, E. L. *et al.* A large-scale binding and functional map of human RNA-binding proteins. *Nature* **583**, 711–719 (2020).
181. McKerrell, T. *et al.* Leukemia-Associated Somatic Mutations Drive Distinct Patterns of Age-Related Clonal Hemopoiesis. *Cell Rep* **10**, 1239–1245 (2015).
182. Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64–69 (2011).
183. Bamopoulos, S. A. *et al.* Clinical presentation and differential splicing of SRSF2, U2AF1 and SF3B1 mutations in patients with acute myeloid leukemia. *Leukemia* **34**, 2621–2634 (2020).
184. Kim, E. *et al.* SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. *Cancer Cell* **27**, 617–630 (2015).
185. Wang, E. *et al.* Targeting an RNA-Binding Protein Network in Acute Myeloid Leukemia. *Cancer Cell* **35**, 369–384.e7 (2019).
186. Rehn, M. *et al.* PTBP1 promotes hematopoietic stem cell maintenance and red blood cell development by ensuring sufficient availability of ribosomal constituents. *Cell Rep* **39**, (2022).

187. Signer, R. A. J., Magee, J. A., Salic, A. & Morrison, S. J. Haematopoietic stem cells require a highly regulated protein synthesis rate. *Nature* **508**, 49–54 (2014).
188. Perner, F. *et al.* YBX1 mediates translation of oncogenic transcripts to control cell competition in AML. *Leukemia* **36**, 426–437 (2022).
189. Jayavelu, A. K. *et al.* Splicing factor YBX1 mediates persistence of JAK2-mutated neoplasms. *Nature* **588**, 157–163 (2020).
190. Fu, Y., Dominissini, D., Rechavi, G. & He, C. Gene expression regulation mediated through reversible m⁶A RNA methylation. *Nature Reviews Genetics* vol. 15 293–306 Preprint at <https://doi.org/10.1038/nrg3724> (2014).
191. Hodson, D. J., Screen, M. & Turner, M. *RNA-Binding Proteins in Hematopoiesis and Hematological Malignancy*. <http://ashpublications.org/blood/article-pdf/133/22/2365/1553512/blood839985.pdf> (2019).
192. Lee, H. *et al.* Stage-specific requirement for Mettl3-dependent m⁶A mRNA methylation during haematopoietic stem cell differentiation. *Nat Cell Biol* **21**, 700–709 (2019).
193. Jason Yao, Q. *et al.* Mettl3-Mettl14 methyltransferase complex regulates the quiescence of adult hematopoietic stem cells. *Cell Res* **28**, 952–954 (2018).
194. Vu, L. P. *et al.* The N⁶-methyladenosine (m⁶A)-forming enzyme METTL3 controls myeloid differentiation of normal hematopoietic and leukemia cells. *Nat Med* **23**, 1369–1376 (2017).
195. Barbieri, I. *et al.* Promoter-bound METTL3 maintains myeloid leukaemia by m⁶A-dependent translation control. *Nature* **552**, 126–131 (2017).
196. Garneau, N. L., Wilusz, J. & Wilusz, C. J. The highways and byways of mRNA decay. *Nature Reviews Molecular Cell Biology* vol. 8 113–126 Preprint at <https://doi.org/10.1038/nrm2104> (2007).
197. Wang, E. *et al.* Surface antigen-guided CRISPR screens identify regulators of myeloid leukemia differentiation. *Cell Stem Cell* **28**, 718–731.e6 (2021).
198. Fabian, M. R. *et al.* Structural basis for the recruitment of the human CCR4-NOT deadenylase complex by tristetraprolin. *Nat Struct Mol Biol* **20**, 735–739 (2013).
199. Collins, C. T. & Hess, J. L. Role of HOXA9 in leukemia: Dysregulation, cofactors and essential targets. *Oncogene* vol. 35 1090–1098 Preprint at <https://doi.org/10.1038/onc.2015.174> (2016).
200. Goardon, N. *et al.* Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer Cell* **19**, 138–152 (2011).
201. Tim Wang, K. B. *et al.* Identification and characterization of essential genes in the human genome. *Science (1979)* **350**, 1096–1101 (2015).
202. Caudron-Herger, M., Jansen, R. E., Wassmer, E. & Diederichs, S. RBP2GO: A comprehensive pan-species database on RNA-binding proteins, their interactions and functions. *Nucleic Acids Res* **49**, D425–D436 (2021).
203. Li, W. *et al.* Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. *Genome Biol* **16**, (2015).
204. Weisz, L., Oren, M. & Rotter, V. Transcription regulation by mutant p53. *Oncogene* vol. 26 2202–2211 Preprint at <https://doi.org/10.1038/sj.onc.1210294> (2007).
205. Pfister, N. T. & Prives, C. Transcriptional regulation by wild-type and cancer-related mutant forms of p53. *Cold Spring Harb Perspect Med* **7**, (2017).
206. Loizou, E. *et al.* A gain-of-function p53-mutant oncogene promotes cell fate plasticity and myeloid leukemia through the pluripotency factor foxh1. *Cancer Discov* **9**, 962–979 (2019).

207. Rodriguez-Meira, A. *et al.* Single-cell multi-omics identifies chronic inflammation as a driver of TP53-mutant leukemic evolution. *Nat Genet* **55**, 1531–1541 (2023).
208. Vujovic, A. *et al.* In Vivo Screening Unveils Pervasive RNA-Binding Protein Dependencies in Leukemic Stem Cells and Identifies ELAVL1 as a Therapeutic Target. *Blood Cancer Discov* **4**, 181–207 (2023).
209. Han, T. *et al.* Anticancer sulfonamides target splicing by inducing RBM39 degradation via recruitment to DCAF15. *Science (1979)* **356**, (2017).
210. Bewersdorf, J. P. *et al.* E7820, an anti-cancer sulfonamide, degrades RBM39 in patients with splicing factor mutant myeloid malignancies: a phase II clinical trial. *Leukemia* **37**, 2512–2516 (2023).
211. Perner, F. *et al.* YBX1 mediates translation of oncogenic transcripts to control cell competition in AML. *Leukemia* **36**, 426–437 (2022).
212. Tailor, D. *et al.* Y box binding protein 1 inhibition as a targeted therapy for ovarian cancer. *Cell Chem Biol* **28**, 1206-1220.e6 (2021).
213. Schnoeder, T. M. *et al.* Pre-Clinical Investigation of a Novel Small Molecule Inhibitor Targeting YBX1 in AML. *Blood* **140**, 491–492 (2022).
214. Yankova, E. *et al.* Small-molecule inhibition of METTL3 as a strategy against myeloid leukaemia. *Nature* **593**, 597–601 (2021).
215. Rudd, S. G. Targeting pan-essential pathways in cancer with cytotoxic chemotherapy: challenges and opportunities. *Cancer Chemotherapy and Pharmacology* vol. 92 241–251 Preprint at <https://doi.org/10.1007/s00280-023-04562-3> (2023).
216. Chang, L., Ruiz, P., Ito, T. & Sellers, W. R. Targeting pan-essential genes in cancer: Challenges and opportunities. *Cancer Cell* vol. 39 466–479 Preprint at <https://doi.org/10.1016/j.ccell.2020.12.008> (2021).
217. Leffers, H., Dejgaard, K. & Celis, J. E. Characterisation of Two Major Cellular Poly(rC)-Binding Human Proteins, Each Containing Three K-homologous (KH) Domains. *Eur J Biochem* **230**, 447–453 (1995).
218. Makeyev, A. V. & Liebhaber, S. A. Identification of two novel mammalian genes establishes a subfamily of KH-domain RNA-binding proteins. *Genomics* **67**, 301–316 (2000).
219. Siomi, H., Matunis, M. J., Michael, W. M. & Dreyfuss, G. *The Pre-mRNA Binding K Protein Contains a Novel Evolutionary Conserved Motif. Nucleic Acids Research* vol. 21 <https://academic.oup.com/nar/article/21/5/1193/1027307> (1993).
220. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science (1979)* **347**, (2015).
221. Du, Z., Yu, J., Chen, Y., Andino, R. & James, T. L. Specific recognition of the C-rich strand of human telomeric DNA and the RNA template of human telomerase by the first KH domain of human poly(C)-binding protein-2. *Journal of Biological Chemistry* **279**, 48126–48134 (2004).
222. Du, Z., Fenn, S., Tjhen, R. & James, T. L. Structure of a construct of a human poly(C)-binding protein containing the first and second KH domains reveals insights into its regulatory mechanisms. *Journal of Biological Chemistry* **283**, 28757–28766 (2008).
223. Sidiqi, M. *et al.* Structure and RNA binding of the third KH domain of poly(C)-binding protein 1. *Nucleic Acids Res* **33**, 1213–1221 (2005).
224. Makeyev, A. V. & Liebhaber, S. A. *The Poly(C)-Binding Proteins: A Multiplicity of Functions and a Search for Mechanisms.* (2002).

225. You, F. *et al.* PCBP2 mediates degradation of the adaptor MAVS via the HECT ubiquitin ligase ALP4. *Nat Immunol* **10**, 1300–1308 (2009).
226. Musco, G., Stier, G. & Joseph, C. *Three-Dimensional Structure and Stability of the KH Domain: Molecular Insights into the Fragile X Syndrome*. *Cell* vol. 85 (1996).
227. Zhang, X. *et al.* Overexpression of PCBP2 contributes to poor prognosis and enhanced cell growth in human hepatocellular carcinoma. *Oncol Rep* **36**, 3456–3464 (2016).
228. Mao, J. *et al.* PCBP2 promotes the development of glioma by regulating FHL3/TGF- β /Smad signaling pathway. *J Cell Physiol* **235**, 3280–3291 (2020).
229. Chen, C. *et al.* Poly(rC) binding protein 2 (PCBP2) promotes the viability of human gastric cancer cells by regulating CDK2. *FEBS Open Bio* **8**, 764–773 (2018).
230. Wan, C. *et al.* β 2-adrenergic receptor signaling promotes pancreatic ductal adenocarcinoma (PDAC) progression through facilitating PCBP2-dependent c-myc expression. *Cancer Lett* **373**, 67–76 (2016).
231. Zhang, X. *et al.* Overexpression of PCBP2 contributes to poor prognosis and enhanced cell growth in human hepatocellular carcinoma. *Oncol Rep* **36**, 3456–3464 (2016).
232. Yuan, C., Chen, M. & Cai, X. Advances in poly(rC)-binding protein 2: Structure, molecular function, and roles in cancer. *Biomedicine and Pharmacotherapy* vol. 139 Preprint at <https://doi.org/10.1016/j.biopha.2021.111719> (2021).
233. Han, W. *et al.* RNA-binding protein PCBP2 modulates glioma growth by regulating FHL3. *Journal of Clinical Investigation* **123**, 2103–2118 (2013).
234. Tang, S.-L., Gao, Y.-L. & Chen, X.-B. *MicroRNA-214 Targets PCBP2 to Suppress the Proliferation and Growth of Glioma Cells*. *Int J Clin Exp Pathol* vol. 8 www.ijcep.com/ (2015).
235. Gíslason, M. H. *et al.* BloodSpot 3.0: a database of gene and protein expression data in normal and malignant haematopoiesis. *Nucleic Acids Res* **52**, D1138–D1142 (2024).
236. Ghanem, L. R. *et al.* The Poly(C) Binding Protein Pcbp2 and Its Retrotransposed Derivative Pcbp1 Are Independently Essential to Mouse Development. *Mol Cell Biol* **36**, 304–319 (2016).
237. Ghanem, L. R. *et al.* Poly(C)-Binding Protein Pcbp2 Enables Differentiation of Definitive Erythropoiesis by Directing Functional Splicing of the Runx1 Transcript. *Mol Cell Biol* **38**, (2018).
238. Ji, X. *et al.* RNA-Binding Proteins PCBP1 and PCBP2 Are Critical Determinants of Murine Erythropoiesis. *Mol Cell Biol* **41**, (2021).
239. Li, R. *et al.* A proinflammatory stem cell niche drives myelofibrosis through a targetable galectin-1 axis. *Sci. Transl. Med* **16**, 7552 (2024).
240. De Bruijn, M. & Dzierzak, E. Runx transcription factors in the development and function of the definitive hematopoietic system. *Blood* **15**, 2061–2069 (2017).
241. Chkheidze, A. N. *et al.* *Assembly of The-Globin mRNA Stability Complex Reflects Binary Interaction between the Pyrimidine-Rich 3 Untranslated Region Determinant and Poly(C) Binding Protein CP*. *MOLECULAR AND CELLULAR BIOLOGY* vol. 19 (1999).
242. Wang, X., Kiledjian, M., Weiss, I. M. & Liebhaber, S. A. *Detection and Characterization of a 3 Untranslated Region Ribonucleoprotein Complex Associated with Human-Globin mRNA Stability*. *MOLECULAR AND CELLULAR BIOLOGY* vol. 15 (1995).

243. Ji, X., Kong, J. & Liebhaber, S. A. An RNA-protein complex links enhanced nuclear 3' processing with cytoplasmic mRNA stabilization. *EMBO Journal* **30**, 2622–2633 (2011).
244. Yu, J. & Russell, J. E. Structural and Functional Analysis of an mRNP Complex That Mediates the High Stability of Human β -Globin mRNA. *Mol Cell Biol* **21**, 5879–5888 (2001).
245. Yanatori, I., Yasui, Y., Tabuchi, M. & Kishi, F. Chaperone protein involved in transmembrane transport of iron. *Biochemical Journal* **462**, 25–37 (2014).
246. Yanatori, I., Richardson, D. R., Imada, K. & Kishi, F. Iron export through the transporter ferroportin 1 is modulated by the iron chaperone PCBP2. *Journal of Biological Chemistry* **291**, 17303–17318 (2016).
247. Leidgens, S. *et al.* Each member of the poly-r(C)-binding protein 1 (PCBP) family exhibits iron chaperone activity toward ferritin. *Journal of Biological Chemistry* **288**, 17791–17802 (2013).
248. Perrotti, D. *et al.* BCR-ABL suppresses C/EBP α expression through inhibitory action of hnRNP E2. *Nat Genet* **30**, 48–58 (2002).
249. Ji, S. C. *et al.* High levels of the BCR/ABL oncoprotein are required for the MAPK-hnRNP-E2-dependent suppression of C/EBP α -driven myeloid differentiation. *Blood* **110**, 994–1003 (2007).
250. Wilkinson, A. C. *et al.* Long-term ex vivo haematopoietic-stem-cell expansion allows nonconditioned transplantation. *Nature* **571**, 117–121 (2019).
251. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *New England Journal of Medicine* **368**, 2059–2074 (2013).
252. Cullen, B. R. Enhancing and confirming the specificity of RNAi experiments. *Nature Methods* vol. 3 677–681 Preprint at <https://doi.org/10.1038/nmeth913> (2006).
253. Ghidini, A., Cléry, A., Halloy, F., Allain, F. H. T. & Hall, J. RNA-PROTACs: Degradars of RNA-Binding Proteins. *Angewandte Chemie - International Edition* **60**, 3163–3169 (2021).
254. Walters, K. *et al.* Small-molecule Ro-08-2750 interacts with many RNA-binding proteins and elicits MUSASH12-independent phenotypes. *RNA* **29**, 1458–1470 (2023).
255. Di Micco, P. *et al.* canSAR: update to the cancer translational research and drug discovery knowledgebase. *Nucleic Acids Res* **51**, D1212–D1219 (2023).
256. Denichenko, P. *et al.* Specific inhibition of splicing factor activity by decoy RNA oligonucleotides. *Nat Commun* **10**, (2019).
257. Shi, L. *et al.* NF90 regulates cell cycle exit and terminal myogenic differentiation by direct binding to the 3'-untranslated region of MyoD and p21 WAF1/CIP1 mRNAs. *Journal of Biological Chemistry* **280**, 18981–18989 (2005).
258. Wu, T. H. *et al.* NF90/ILF3 is a transcription factor that promotes proliferation over differentiation by hierarchical regulation in K562 erythroleukemia cells. *PLoS One* **13**, (2018).
259. Vrakas, C. N. *et al.* RNA stability protein ILF3 mediates cytokine-induced angiogenesis. *FASEB Journal* **33**, 3304–3316 (2019).
260. Evans, J. R. *et al.* Members of the poly (rC) binding protein family stimulate the activity of the c-myc internal ribosome entry segment in vitro and in vivo. *Oncogene* **22**, 8012–8020 (2003).
261. Wu, T. H. *et al.* NF90/ILF3 is a transcription factor that promotes proliferation over differentiation by hierarchical regulation in K562 erythroleukemia cells. *PLoS One* **13**, (2018).

262. Yu, J. & Russell, J. E. Structural and Functional Analysis of an mRNP Complex That Mediates the High Stability of Human β -Globin mRNA. *Mol Cell Biol* **21**, 5879–5888 (2001).
263. Zheng, R. *et al.* hnRNPM protects against the dsRNA-mediated interferon response by repressing LINE-associated cryptic splicing. *Mol Cell* **84**, 2087–2103.e8 (2024).
264. Jiang, W. *et al.* Regulation of cell cycle of hepatocellular carcinoma by NF90 through modulation of cyclin E1 mRNA stability. *Oncogene* **34**, 4460–4470 (2015).
265. Groza, T. *et al.* The International Mouse Phenotyping Consortium: comprehensive knockout phenotyping underpinning the study of human disease. *Nucleic Acids Res* **51**, D1038–D1045 (2023).
266. Damianov, A. *et al.* Rbfox Proteins Regulate Splicing as Part of a Large Multiprotein Complex LASR. *Cell* **165**, 606–619 (2016).
267. Gu, H. *et al.* PCBP2 maintains antiviral signaling homeostasis by regulating cGAS enzymatic activity via antagonizing its condensation. *Nat Commun* **13**, (2022).
268. Datar, K. V, Dreyfuss, G. & Swanson, M. S. *The Human HnRNP M Proteins: Identification of a Methionine/Arginine-Rich Repeat Motif in Ribonucleoproteins.* *Nucleic Acids Research* vol. 21 (1993).
269. Grasso, G. & Kiernan, R. The Polyvalent Role of NF90 in RNA Biology. *Int J Mol Sci* **23**; **13584**, (2022).
270. Wu, T. H. *et al.* NF90/ILF3 is a transcription factor that promotes proliferation over differentiation by hierarchical regulation in K562 erythroleukemia cells. *PLoS One* **13**, (2018).
271. Tan, C. W. *et al.* PRPF40A induces inclusion of exons in GC-rich regions important for human myeloid cell differentiation. *Nucleic Acids Res* **52**, 8800–8814 (2024).
272. Harutyunyan, A. S. *et al.* Germline RBBP6 mutations in familial myeloproliferative neoplasms. *Blood* **127**, 362–365 (2016).
273. Cmejla R, Cmejlova J, Handrkova H, Petrak J & Pospisilova D. Ribosomal protein S17 gene (RPS17) is mutated in Diamond-Blackfan anemia. *Hum Mutat.* 1178–82 (2007).
274. Garritano, S., Inga, A., Gemignani, F. & Landi, S. More targets, more pathways and more clues for mutant p53. *Oncogenesis* vol. 2 Preprint at <https://doi.org/10.1038/oncsis.2013.15> (2013).
275. Miotto, B. *et al.* The RBBP6/ZBTB38/MCM10 Axis Regulates DNA Replication and Common Fragile Site Stability. *Cell Rep* **7**, 575–587 (2014).
276. Li, L. *et al.* PACT is a negative regulator of p53 and essential for cell growth and embryonic development. *PNAS* **104**; **19**, 7951–7956 (2007).
277. Lin, K.-T., Lu, R.-M. & Tarn, W.-Y. The WW Domain-Containing Proteins Interact with the Early Spliceosome and Participate in Pre-mRNA Splicing In Vivo. *Mol Cell Biol* **24**, 9176–9185 (2004).
278. Makarov, E. M., Owen, N., Bottrill, A. & Makarova, O. V. Functional mammalian spliceosomal complex e contains SMN complex proteins in addition to U1 and U2 snRNPs. *Nucleic Acids Res* **40**, 2639–2652 (2012).
279. Chibi, M. *et al.* RBBP6 Interacts with Multifunctional Protein YB-1 through Its RING Finger Domain, Leading to Ubiquitination and Proteosomal Degradation of YB-1. *J Mol Biol* **384**, 908–916 (2008).
280. Witte, M. M. & Scott, R. E. *The Proliferation Potential Protein-Related (P2P-R) Gene with Domains Encoding Heterogeneous Nuclear Ribonucleoprotein Association and Rb1 Binding Shows Repressed Expression during Terminal Differentiation.* *Cell Biology* vol. 94 www.pnas.org. (1997).

281. Bommer, U. A., Stahl, J., Henske, A., Lutsch, G. & Bielka, H. Identification of proteins of the 40 S ribosomal subunit involved in interaction with initiation factor eIF-2 in the quaternary initiation complex by means of monospecific antibodies. *FEBS Lett* **233**, 114–118 (1988).
282. Di Giammartino, D. C. *et al.* RBBP6 isoforms regulate the human polyadenylation machinery and modulate expression of mRNAs with AU-rich 3' UTRs. *Genes Dev* **28**, 2248–2260 (2014).
283. Takayama, K. I. *et al.* Identification of small-molecule inhibitors against the interaction of RNA-binding protein PSF and its target RNA for cancer treatment. *PNAS Nexus* **2**, (2023).
284. Shkreta, L. *et al.* Modulation of the splicing regulatory function of SRSF10 by a novel compound that impairs HIV-1 replication. *Nucleic Acids Res* **45**, 4051–4067 (2017).
285. Sohail, M. *et al.* A novel class of inhibitors that target SRSF10 and promote p53-mediated cytotoxicity on human colorectal cancer cells. *NAR Cancer* **3**, 1–22 (2021).
286. Gripp, K. W. *et al.* Diamond-Blackfan anemia with mandibulofacial dystostosis is heterogeneous, including the novel DBA genes TSR2 and RPS28. *Am J Med Genet A* **164**, 2240–2249 (2014).
287. Brown, L. M. *et al.* SFPQ-ABL1 and BCR-ABL1 use different signaling networks to drive B-cell acute lymphoblastic leukemia. *Blood Adv* **6**, 2373–2387 (2022).
288. Zhong, F. M. *et al.* Splicing factor-mediated regulation patterns reveals biological characteristics and aid in predicting prognosis in acute myeloid leukemia. *J Transl Med* **21**, (2023).
289. Wei, J. *et al.* Ribosomal Proteins Regulate MHC Class I Peptide Generation for Immunosurveillance. *Mol Cell* **73**, 1162–1173.e5 (2019).
290. Pellarin, I. *et al.* Splicing factor proline- and glutamine-rich (SFPQ) protein regulates platinum response in ovarian cancer-modulating SRSF2 activity. *Oncogene* **39**, 4390–4403 (2020).
291. Zhou, X. *et al.* BCLAF1 and its splicing regulator SRSF10 regulate the tumorigenic potential of colon cancer cells. *Nat Commun* **5**, (2014).
292. Chang, C. *et al.* The aberrant upregulation of exon 10-inclusive SREK1 through SRSF10 acts as an oncogenic driver in human hepatocellular carcinoma. *Nat Commun* **13**, (2022).
293. Feng, Y. *et al.* SRp38 Regulates Alternative Splicing and Is Required for Ca²⁺ Handling in the Embryonic Heart. *Dev Cell* **16**, 528–538 (2009).
294. Knott, G. J., Bond, C. S. & Fox, A. H. The DBHS proteins SFPQ, NONO and PSPC1: A multipurpose molecular scaffold. *Nucleic Acids Research* vol. 44 3989–4004 Preprint at <https://doi.org/10.1093/nar/gkw271> (2016).
295. Rajesh, C., Baker, D. K., Pierce, A. J. & Pittman, D. L. The splicing-factor related protein SFPQ/PSF interacts with RAD51D and is necessary for homology-directed repair and sister chromatid cohesion. *Nucleic Acids Res* **39**, 132–145 (2011).
296. Xiao, W. *et al.* Nuclear m6A Reader YTHDC1 Regulates mRNA Splicing. *Mol Cell* **61**, 507–519 (2016).
297. Tailor, D. *et al.* Y box binding protein 1 inhibition as a targeted therapy for ovarian cancer. *Cell Chem Biol* **28**, 1206–1220.e6 (2021).
298. Wang, E. *et al.* Modulation of RNA splicing enhances response to BCL2 inhibition in leukemia. *Cancer Cell* **41**, 164–180.e8 (2023).
299. Perner, F. *et al.* YBX1 mediates translation of oncogenic transcripts to control cell competition in AML. *Leukemia* **36**, 426–437 (2022).

300. Lee, J. H., Jeong, S. A., Khadka, P., Hong, J. & Chung, I. K. Involvement of SRSF11 in cell cycle-specific recruitment of telomerase to telomeres at nuclear speckles. *Nucleic Acids Res* **43**, 8435–8451 (2015).
301. Zhang, C. *et al.* YARS as an oncogenic protein that promotes gastric cancer progression through activating PI3K-Akt signaling. *J Cancer Res Clin Oncol* **146**, 329–342 (2020).
302. Lasham, A., Print, C. G., Woolley, A. G., Dunn, S. E. & Braithwaite, A. W. YB-1: Oncoprotein, prognostic marker and therapeutic target? *Biochemical Journal* vol. 449 11–23 Preprint at <https://doi.org/10.1042/BJ20121323> (2013).
303. Lasham, A. *et al.* The Y-box-binding protein, YB1, is a potential negative regulator of the p53 tumor suppressor. *Journal of Biological Chemistry* **278**, 35516–35523 (2003).
304. Zhi Hong Lu, J. T. B. & Ley, T. J. YB-1 Is Important for Late-Stage Embryonic Development, Optimal Cellular Stress Responses, and the Prevention of Premature Senescence. *Mol Cell Biol* **25**, 4625–4637 (2005).
305. Feng, M. *et al.* YBX1 Is Required for Maintaining Myeloid Leukemia Cell Survival by Regulating BCL2 Stability in an m 6 A-Dependent Manner. <http://ashpublications.org/blood/article-pdf/138/1/71/1813020/bloodbld2020009676.pdf> (2021).
306. Schonleber Page-Mccaw, P., Amonlirdviman, K. & Sharp, P. A. PUF60: A Novel U2AF65-Related Splicing Activity. (1999).
307. Averdunk, L. *et al.* The recurrent missense mutation p.(Arg367Trp) in YARS1 causes a distinct neurodevelopmental phenotype. *J Mol Med* (2021) doi:10.1007/s00109-021-02124-9.
308. Gaudreault, I., Guay, D. & Lebel, M. YB-1 promotes strand separation in vitro of duplex DNA containing either mispaired bases or cisplatin modifications, exhibits endonucleolytic activities and binds several DNA repair proteins. *Nucleic Acids Res* **32**, 316–327 (2004).
309. Weidensdorfer, D. *et al.* Control of c-myc mRNA stability by IGF2BP1-associated cytoplasmic RNPs. *RNA* **15**, 104–115 (2009).
310. Frisch, B. J. & Calvi, L. M. Hematopoietic stem cell cultures and assays. *Methods in Molecular Biology* **1130**, 315–324 (2014).
311. Perry, J. M. & Li, L. Functional assays for hematopoietic stem cell self-renewal. *Methods Mol Biol* **636**, 45–54 (2010).
312. Raic, A., Naolou, T., Mohra, A., Chatterjee, C. & Lee-Thedieck, C. 3D models of the bone marrow in health and disease: Yesterday, today, and tomorrow. *MRS Communications* vol. 9 37–52 Preprint at <https://doi.org/10.1557/mrc.2018.203> (2019).
313. Wunderlich, M. *et al.* AML xenograft efficiency is significantly improved in NOD/SCID-IL2RG mice constitutively expressing human SCF, GM-CSF and IL-3. *Leukemia* vol. 24 1785–1788 Preprint at <https://doi.org/10.1038/leu.2010.158> (2010).
314. Celik, H. *et al.* A Humanized Animal Model Predicts Clonal Evolution and Therapeutic Vulnerabilities in Myeloproliferative Neoplasms. *Cancer Discov* **11**, 3126–3141 (2021).
315. Lysenko, V. *et al.* Enhanced engraftment of human myelofibrosis stem and progenitor cells in MISTRG mice. *Blood Adv* **4**, 2477–2488 (2020).

316. Méndez-Ferrer, S. *et al.* Bone marrow niches in haematological malignancies. *Nature Reviews Cancer* vol. 20 285–298 Preprint at <https://doi.org/10.1038/s41568-020-0245-2> (2020).
317. Sala-Torra, O. *et al.* Evidence of Donor-Derived Hematologic Malignancies after Hematopoietic Stem Cell Transplantation. *Biology of Blood and Marrow Transplantation* **12**, 511–517 (2006).
318. Walkley, C. R., Shea, J. M., Sims, N. A., Purton, L. E. & Orkin, S. H. Rb Regulates Interactions between Hematopoietic Stem Cells and Their Bone Marrow Microenvironment. *Cell* **129**, 1081–1095 (2007).
319. Kleppe, M. *et al.* JAK-STAT pathway activation in malignant and nonmalignant cells contributes to MPN pathogenesis and therapeutic response. *Cancer Discov* **5**, 316–331 (2015).
320. Raic, A., Rödling, L., Kalbacher, H. & Lee-Thedieck, C. Biomimetic macroporous PEG hydrogels as 3D scaffolds for the multiplication of human hematopoietic stem and progenitor cells. *Biomaterials* **35**, 929–940 (2014).
321. Ventura Ferreira, M. S. *et al.* Cord blood-hematopoietic stem cell expansion in 3D fibrin scaffolds with stromal support. *Biomaterials* **33**, 6987–6997 (2012).
322. Bessy, T., Itkin, T. & Passaro, D. Bioengineering the Bone Marrow Vascular Niche. *Frontiers in Cell and Developmental Biology* vol. 9 Preprint at <https://doi.org/10.3389/fcell.2021.645496> (2021).
323. Chou, D. B. *et al.* On-chip recapitulation of clinical bone marrow toxicities and patient-specific pathophysiology. *Nat Biomed Eng* **4**, 394–406 (2020).
324. Zhao, Z. *et al.* Organoids. *Nature Reviews Methods Primers* **2**, (2022).
325. Sato, T. *et al.* Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* **459**, 262–265 (2009).
326. Frenz-Wiessner, S. *et al.* Generation of complex bone marrow organoids from human induced pluripotent stem cells. *Nat Methods* **21**, 868–881 (2024).
327. Mayani, H., Flores-Figueroa, E. & Chávez-González, A. In vitro biology of human myeloid leukemia. *Leukemia Research* vol. 33 624–637 Preprint at <https://doi.org/10.1016/j.leukres.2008.11.011> (2009).
328. Millington, M., Arndt, A., Boyd, M., Applegate, T. & Shen, S. Towards a clinically relevant lentiviral transduction protocol for primary human CD34+ hematopoietic stem/progenitor cells. *PLoS One* **4**, (2009).
329. Uchida, N. *et al.* High-Efficiency Lentiviral Transduction of Human CD34+ Cells in High-Density Culture with Poloxamer and Prostaglandin E2. *Mol Ther Methods Clin Dev* **13**, 187–196 (2019).
330. Uchida, N. *et al.* Optimal conditions for lentiviral transduction of engrafting human CD34 + cells. *Gene Ther* **18**, 1078–1086 (2011).
331. Canté-Barrett, K. *et al.* Lentiviral gene transfer into human and murine hematopoietic stem cells: Size matters. *BMC Res Notes* **9**, (2016).
332. Kumar, M., Keller, B., Makalou, N. & Sutton, R. E. *Systematic Determination of the Packaging Limit of Lentiviral Vectors*. *HUMAN GENE THERAPY* vol. 12 www.liebertpub.com (2001).
333. Sweeney, N. P. & Vink, C. A. The impact of lentiviral vector genome size and producer cell genomic to gag-pol mRNA ratios on packaging efficiency and titre. *Mol Ther Methods Clin Dev* **21**, 574–584 (2021).
334. Bajaj, J. *et al.* An in vivo genome-wide CRISPR screen identifies the RNA-binding protein Staufen2 as a key regulator of myeloid leukemia. *Nat Cancer* **1**, 410–422 (2020).

335. Park, S. M. *et al.* IKZF2 Drives Leukemia Stem Cell Self-Renewal and Inhibits Myeloid Differentiation. *Cell Stem Cell* **24**, 153-165.e7 (2019).
336. Woodley, K. *et al.* Mannose metabolism inhibition sensitizes acute myeloid leukaemia cells to therapy by driving ferroptotic cell death. *Nat Commun* **14**, (2023).
337. Paris, J. *et al.* Targeting the RNA m6A Reader YTHDF2 Selectively Compromises Cancer Stem Cells in Acute Myeloid Leukemia. *Cell Stem Cell* **25**, 137-148.e6 (2019).
338. Ramezani, A., Hawley, T. S. & Hawley, R. G. Lentiviral vectors for enhanced gene expression in human hematopoietic cells. *Molecular Therapy* **2**, 458-469 (2000).
339. Ravi, N. *et al.* Original Article Lentiviral Based Gene Transduction and Promoter Studies in Human Hematopoietic Stem Cells (hHSCs). *JSRM* **7**;1, 41-53 (2011).
340. Kohn, D. B. *et al.* Autologous Ex Vivo Lentiviral Gene Therapy for Adenosine Deaminase Deficiency. *New England Journal of Medicine* **384**, 2002-2013 (2021).
341. Paddison, P. J., Caudy, A. A., Bernstein, E., Hannon, G. J. & Conklin, D. S. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev* **16**, 948-958 (2002).
342. Moore, C. B., Guthrie, E. H., Huang, M. T. H. & Taxman, D. J. Short hairpin RNA (shRNA): design, delivery, and assessment of gene knockdown. *Methods Mol Biol* **629**, 141-158 (2010).
343. Merritt, W. M. *et al.* Dicer, Drosha, and Outcomes in Patients with Ovarian Cancer. *New England Journal of Medicine* **359**, 2641-2650 (2008).
344. Li, Y. *et al.* Iron homeostasis regulates the activity of the microRNA pathway through poly(C)-binding protein 2. *Cell Metab* **15**, 895-904 (2012).
345. Larson, M. H. *et al.* CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nat Protoc* **8**, 2180-2196 (2013).
346. Thakur, S. *et al.* Regulation of BRCA1 Transcription by Specific Single-Stranded DNA Binding Factors. *Mol Cell Biol* **23**, 3774-3787 (2003).
347. Choi, H. S. *et al.* A proteomics approach for identification of single strand DNA-binding proteins involved in transcriptional regulation of mouse μ opioid receptor gene. *Molecular and Cellular Proteomics* **7**, 1517-1529 (2008).
348. Czyzyk-Krzeska, M. F. & Bendixen, A. C. Identification of the Poly(C) Binding Protein in the Complex Associated With the 3' Untranslated Region of Erythropoietin Messenger RNA. *Blood* **93**(6), 2111-2120 (1999).
349. Spear, A., Sharma, N. & Flanagan, J. B. Protein-RNA tethering: The role of poly(C) binding protein 2 in poliovirus RNA replication. *Virology* **374**, 280-291 (2008).
350. Blyn, L. B., Towner, J. S., Semler, B. L. & Ehrenfeld, E. Requirement of Poly(RC) Binding Protein 2 for Translation of Poliovirus RNA. *JOURNAL OF VIROLOGY* vol. 71 (1997).
351. Xin, Z. *et al.* Pcbp2 enhances the antiviral activity of ifn- α against hcv by stabilizing the mrna of stat1 and stat2. *PLoS One* **6**, (2011).
352. Waggoner, S. A., Johannes, G. J. & Liebhauer, S. A. Depletion of the poly(C)-binding proteins α CP1 and α CP2 from K562 cells leads to p53-independent induction of cyclin-dependent kinase inhibitor (CDKN1A) and G1 arrest. *Journal of Biological Chemistry* **284**, 9039-9049 (2009).
353. Zhang, Y., Qian, J., Gu, C. & Yang, Y. Alternative splicing and cancer: a systematic review. *Signal Transduction and Targeted Therapy* vol. 6 Preprint at <https://doi.org/10.1038/s41392-021-00486-7> (2021).
354. Ji, X. *et al.* α CP binding to a cytosine-rich subset of polypyrimidine tracts drives a novel pathway of cassette exon splicing in the mammalian transcriptome. *Nucleic Acids Res* **44**, 2283-2297 (2016).

355. Ghosh, M. *et al.* Mutant p53 suppresses innate immune signaling to promote tumorigenesis. *Cancer Cell* **39**, 494-508.e5 (2021).
356. Li, J. *et al.* Non-cell-autonomous cancer progression from chromosomal instability. *Nature* **620**, 1080–1088 (2023).
357. Hong, C. *et al.* cGAS–STING drives the IL-6-dependent survival of chromosomally instable cancers. *Nature* **607**, 366–373 (2022).
358. Rigby, R. E., Rue-Albrecht, K., Sims, D. & Rehwinkel, J. Single-cell analysis of signalling and transcriptional 1 responses to type I interferons. *bioRxiv* (2023) doi:10.1101/2023.07.03.547491.
359. Nassar, L. R. *et al.* The UCSC Genome Browser database: 2023 update. *Nucleic Acids Res* **51**, D1188–D1195 (2023).
360. Aigner, P., Just, V. & Stoiber, D. STAT3 isoforms: Alternative fates in cancer? *Cytokine* vol. 118 27–34 Preprint at <https://doi.org/10.1016/j.cyto.2018.07.014> (2019).
361. Vijay, V. *et al.* Heterogeneous Nuclear Ribonucleoprotein C is an Indispensable Target in Acute Myeloid Leukemia. *bioRxiv* (2023) doi:10.1101/2023.10.02.559746.
362. Kuusanmäki, H. *et al.* Erythroid/megakaryocytic differentiation confers BCL-XL dependency and venetoclax resistance in acute myeloid leukemia. *Blood* **141**;13, (2023).
363. Chakraborty, P., Huang, J. T. J. & Hiom, K. DHX9 helicase promotes R-loop formation in cells with impaired RNA splicing. *Nat Commun* **9**, (2018).
364. Cristini, A., Groh, M., Kristiansen, M. S. & Gromak, N. RNA/DNA Hybrid Interactome Identifies DXH9 as a Molecular Player in Transcriptional Termination and R-Loop-Associated DNA Damage. *Cell Rep* **23**, 1891–1905 (2018).
365. Wang, I. X. *et al.* Human proteins that interact with RNA/DNA hybrids. *Genome Res* **28**, 1405–1414 (2018).
366. García-Muse, T. & Aguilera, A. R Loops: From Physiological to Pathological Roles. *Cell* vol. 179 604–618 Preprint at <https://doi.org/10.1016/j.cell.2019.08.055> (2019).
367. Brickner, J. R., Garzon, J. L. & Cimprich, K. A. Walking a tightrope: The complex balancing act of R-loops in genome stability. *Molecular Cell* vol. 82 2267–2297 Preprint at <https://doi.org/10.1016/j.molcel.2022.04.014> (2022).
368. Kuo, Linda. J. & Yang, L.-X. γ -H2AX – A Novel Biomarker for DNA Double-strand Breaks. *in vivo* **22**, 305–310 (2008).
369. McMahon, A. C. *et al.* TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins. *Cell* **165**, 742–753 (2016).
370. Ji, X., Humenik, J., Yang, D. & Liebhaber, S. A. PolyC-binding proteins enhance expression of the CDK2 cell cycle regulatory protein via alternative splicing. *Nucleic Acids Res* **46**, 2030–2044 (2018).
371. Dubreuil, M. M. *et al.* Systematic Identification of Regulators of Oxidative Stress Reveals Non-canonical Roles for Peroxisomal Import and the Pentose Phosphate Pathway. *Cell Rep* **30**, 1417-1433.e7 (2020).
372. Ishii, T. *et al.* PCBP1 and PCBP2 both bind heavily oxidized RNA but cause opposing outcomes, suppressing or increasing apoptosis under oxidative conditions. *Journal of Biological Chemistry* **295**, 12247–12261 (2020).
373. Srinivas, U. S., Tan, B. W. Q., Vellayappan, B. A. & Jeyasekharan, A. D. ROS and the DNA damage response in cancer. *Redox Biology* vol. 25 Preprint at <https://doi.org/10.1016/j.redox.2018.101084> (2019).

- 374. Psaila, B. & Mead, A. J. Single-cell approaches reveal novel cellular pathways for megakaryocyte and erythroid differentiation. *Blood* **133**(13), 1427–1435 (2019).
- 375. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
- 376. Varadi, M. *et al.* AlphaFold Protein Structure Database: Massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res* **50**, D439–D444 (2022).