Determining how SRSF2 mutation leads to MDS and CMML and its genetic co-operativities *in vivo* and *in vitro*

Author:
Jane Jialu XU
ORCID: 0000-0001-5065-4860

Supervisor:
Prof. Carl WALKLEY
Co-Supervisor:
Dr. Monique SMEETS
Dr. Alistair CHALK

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

in the

Cancer and RNA Biology Lab
St Vincent's Institute of Medical Research
&
Department of Medicine, St Vincent's Hospital
Faculty of Medicine, Dentistry and Health Sciences
University of Melbourne

September 2020

COPYRIGHT ©2020, BY JANE JIALU XU
ALL RIGHTS RESERVED.

Abstract

Myelodysplastic syndromes (MDS) are a group of heterogeneous clonal stem cell malignancies characterised by multilineage cytopenia and dysplasia. MDS mostly occurs in aged populations where there are limited therapeutic options. Compared to MDS, chronic myelomonocytic leukaemia (CMML) presents a monocytosis feature and has a poor survival of 16 months for high-risk patients. In the past decade, several sequencing studies have defined the complex molecular landscapes of MDS and CMML. SRSF2, a component of RNA splicing machinery, is one of the most frequent mutations in MDS and CMML. To understand the consequences and effects of SRSF2^{P95H} mutation on normal haematopoiesis, several groups, including our lab, have generated *in vivo* mouse models using various gene targeting strategies and Cre recombinases. These models demonstrated some effects of SRSF2^{P95H} mutation on haematopoiesis, and described certain mis-splicing changes in the *Srsf2*$^{P95H/+}$ cells. However, the mechanism and the role of SRSF2^{P95H} mutation in promoting and initiating MDS/CMML are still poorly defined. My thesis aimed to address key knowledge gaps by examining the cell of origin, transcriptomic/splicing changes, synthetic lethal genetic interactions, and the co-operative interactions of *Srsf2*$^{P95H/+}$ mutation in the initiation of MDS/CMML.

In the first part of my thesis, I assessed the cell of origin in SRSF2^{P95H} MDS by characterising a conditional knock-in *Srsf2*$^{P95H/+}$ mouse model, using *LysM*-Cre. After activating *Srsf2*$^{P95H/+}$ mutation in myeloid progenitors, I observed no development of MDS even after prolonged ageing (up to 52 weeks) and only mild changes in haematopoiesis. Compared to the stem cell activation model (h*Scl*-CreER^{T2}) that we reported, the results

of *LysM*-Cre demonstrated that the myeloid progenitor is not the cell of origin in SRSF2^{mut} MDS.

In the second part of my thesis, I analysed the transcriptomic and splicing changes of *Srsf2*$^{P95H/+}$ cells, using both purified stem and progenitor cell populations as well as Hoxb8 immortalised cell lines. The transcriptome analysis revealed up- and down-regulation of lineage associated genes and up-regulation of MDS associated pathways and the p38 MAPK kinase pathway. The splicing analysis demonstrated skipped exons as the most frequent alternative splicing event. In terms of specific mis-splicing targets, I examined exon inclusion in several reported transcripts and compared the most frequently mis-spliced genes across 12 human SRSF2^{mut}/murine Srsf2^{mut} datasets. Through this analysis, I found that mRNA processing and DNA repair represent the top mis-spliced pathways in *Srsf2*$^{P95H/+}$ cells. I also present a pilot study of single-cell RNA sequencing of *Srsf2*$^{P95H/+}$ stem and primitive progenitor populations, which unveiled a myeloid-biased signature and enhanced myeloid differentiation of the *Srsf2*$^{P95H/+}$ stem cells.

In the third part of my thesis, I explored the synthetic lethality of *Srsf2*$^{P95H/+}$ cells with a pooled CRISPR knock-out screen. I discovered that loss of DNA repair or Cell cycle pathways is synthetic lethal with *Srsf2*$^{P95H/+}$ mutation. Consistent with the genetic lethality, I demonstrated that Palbociclib, a CDK6 inhibitor, could preferentially target the *Srsf2*$^{P95H/+}$ cells. This finding opens up new therapeutic windows beyond known spliceosome inhibitors for SRSF2mut MDS.

In the fourth and last part of my thesis, I generated and characterised two co-mutation models: *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ and *Srsf2*$^{P95H/+}$ *Cbl*$^{-/-}$. In the *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ model, I observed profound myeloid bias, B

lymphoid suppression and increased ST-HSC percentages in the stem cell compartment after 52 weeks of activating/deleting mutations in the haematopoietic stem cells. Within the *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ cohort, I also observed development of CMML in both native haematopoiesis and transplantation settings. So far, this is the first model to demonstrate synergistic interactions between *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ mutation, as well as initiation of CMML *in vivo*. For the *Srsf2*$^{P95H/+}$ *Cbl*$^{-/-}$ model, I characterised a small cohort of mice due to prolonged breeding difficulties. Nevertheless, I discovered increased myeloid proliferation in the double *Srsf2*$^{P95H/+}$ *Cbl*$^{-/-}$ and *Srsf2*$^{P95H/+}$ *Cbl*$^{+/-}$ mutants.

Collectively, the work included in this thesis creates an original contribution to understanding the role of SRSF2^{P95H} mutation in MDS, its synthetic lethal genetic interactions, potential therapeutic targeting of SRSF2^{P95H} cells, and how it co-operates with other recurrent mutations in initiation of CMML.

Declaration

This is to certify that:

1. This thesis comprises only my original work towards the degree Doctor of Philosophy except where indicated in the Preface and acknowledged in the text.

2. Due acknowledgement has been made in the text to all other material used.

3. The thesis is fewer than 100,000 words in length, exclusive of tables, figures, bibliographies and appendices.

Jane Jialu Xu

September 2020

Preface

In accordance with the regulations of The University of Melbourne, my contribution to each chapter was as follows:

- Chapter 3: 95%
- Chapter 4: 80%
- Chapter 5: 85%
- Chapter 6: 95%
- Overall contribution to thesis: 90%

I acknowledge the help of others and contributions of data, which I have presented in this thesis:

- The section 1.10 and Figure 1.9 in Chapter 1 of this thesis includes part of our published review (Xu et al., 2019).
- Retro-orbital bleeding of mice was performed by Prof Louise Purton or Prof Carl Walkley.
- The bone marrow transplantation was performed by Prof Carl Walkley.
- Dr. Monique Smeets (St Vincent's Institute) contributed to data acquisition and analysis presented in Figure 4.3C-D and Figure 4.6.
- Dr. Alistair Chalk (St Vincent's Institute) performed bioinformatic analysis presented in Figure 4.2, 4.3, 4.4, 5.3, 5.4, 5.5, 5.6 and Table 5.1.

- Dr Davis McCarthy and Ms Ruqian Lyu (St Vincent's Institute) performed bioinformatic analysis presented in Figure 4.7, 4.8, 4.9 and Table 4.3, 4.4.

- Mr Scott Taylor contributed to the RNA seq library preparation and optimisation of the CRISPR screen.

- Fluorescence-activated cell sorting was performed by Michael Thompson and Dijana Miljkovic.

List of Publications

- Smeets, M. F., Tan, S. Y., **Xu, J. J.**, Anande, G., Unnikrishnan, A., Chalk, A. M., Taylor, S. R., Pimanda, J. E., Wall, M., Purton, L. E., and Walkley, C. R. (2018). Srsf2 P95H initiates myeloid bias and myelodysplastic/myeloproliferative syndrome from hemopoietic stem cells. Blood 132, 608-621.

- **Xu, J. J.**, Smeets, M. F., Tan, S. Y., Wall, M., Purton, L. E., and Walkley, C. R. (2019). Modeling human RNA spliceosome mutations in the mouse: not all mice were created equal. Experimental Hematology 70, 10-23.

Acknowledgements

This thesis and my PhD journey would have not been possible without the support and help of many people. This acknowledgement is an attribute to all those that assisted me intellectually and technically, whether big or small.

First and foremost, I would like to thank my principle supervisor Prof Carl Walkley. It was a hard decision to pursue a research career after I completed my Master of Laboratory Medicine degree and qualified to become a medical scientist. The prospect of another 5 years of student life certainly is daunting, but I am glad that I made the right decision and joined Carl's lab to start my honours and PhD studies. His guidance, advice and passion towards research encouraged me every step along the way. It would be impossible to conduct my PhD study without his immense help, both financially and intellectually. I am very grateful for all the opportunities he provided for me to attend conferences and network with fellow researchers.

Secondly, I would like to thank my co-supervisors: Dr. Monique Smeets and Dr. Alistair Chalk. It was stressful to learn all the experimental techniques and protocols at the beginning of my training. Monique's patience and thorough explanations came to aid every time. Thanks for all the wonderful discussions and advice. For Alistair, thank you for all the assistance in the bioinformatic analysis and intellectual input in the data interpretation. I am grateful to have both of you as my co-supervisors.

To the rest member of the Cancer and RNA Biology lab, I would

like to thank the inspiring Dr. Jacki Heraud-Farlow, my PhD buddy Dr. Wilson Castillo-Tandazo, wonderful research assistants Ankita Goradia and Scott Taylor, for your help both in and out of my research life. I would also like to thank Prof Louise Purton, for her help in the lab and also entrusting me as a dog/house sitter. To the past and present of the Stem Cell Regulation lab: Dr. Jess Holien, Dr. Gavin Tjin, Ms Lenny Straszkowski, Diannita Kwang, Kelli Schleibs, Dr. ShuhYing Tan, Dr. Slavisa Ninkovic and Clea Grace. Thank you for all the delighted conversations in the office, which provided necessary entertainment and stress relief during my PhD studies.

To my fellow PhD students: Dr. Winnie Tan, Martha Blank, Ms Yao Sun, Dr. Tingting Ge, Vanessa Tsui, and Ashley Ovens, thank you for the wonderful times during student events and lunch talks.

To St Vincent's Institute, thank you for providing me the PhD top up scholarship. It was very generous and I am very grateful for this support.

Last but not least, I would like to thank my parents and my partner, William. Your unconditional love and support have always carried me through hard times. Thank you all for being there for me. I love you all.

Contents

Abstract i

Declaration iv

Preface v

List of Publications vii

Acknowledgements viii

List of Figures xvi

List of Tables xxii

List of Abbreviations xxiv

1 Introduction and background 1

1.1 Haematopoiesis 2

1.2 HSC ageing and clonal haematopoiesis of indeterminate potential (CHIP) 5

1.3 Myelodysplastic syndromes and Chronic myelomonocytic leukaemia 8

1.3.1 Myelodysplastic syndromes (MDS) 8

1.3.2 Myelodysplastic/myeloproliferative syndromes (MDS/MPN) and Chronic myelomonocytic leukaemia (CMML) 13

1.3.3 Treatment options for MDS and CMML 14
1.4 Molecular abnormalities in MDS and CMML 19
1.4.1 Genetic landscape and clonal evolution of MDS . . 19
1.4.2 Genetic landscape and clonal evolution of CMML . 22
1.4.3 Genetic co-operativity in MDS and CMML 23
1.5 RNA splicing and cancer 25
1.5.1 RNA splicing and the spliceosome 25
1.5.2 Regulation of RNA splicing 25
1.5.3 Aberrant alternative RNA splicing in cancers 28
1.6 Serine-arginine rich splicing factor 2 (SRSF2) 30
1.6.1 SRSF2 function in haematopoiesis 31
1.6.2 *SRSF2* mutation leads to a change in RNA binding affinity 31
1.6.3 Mechanism of *SRSF2* P95H mutation in MDS pathogenesis: poison exon? 34
1.6.4 Targeted therapy for RNA splicing mutated MDS/CMML 36
1.7 DNA methylation and TET2 38
1.7.1 DNA methylation in haematopoiesis 38
1.7.2 Deregulation of DNA methylation in haematological malignancies 41
1.7.3 *TET2* mutation in haematological malignancies . . 42
1.7.4 Targeted epigenetic therapy for MDS/CMML . . . 43
1.8 CBL in haematopoiesis and leukaemogenesis 47
1.9 Mouse models of MDS and CMML 49
1.10 Mouse models of *Srsf2* and *Tet2* mutation 50
1.10.1 Mouse models of *Srsf2* mutation 50
1.10.2 Mouse models of *Tet2* mutation 56

1.11 Mouse models of co-operative mutations in MDS/MPN . 61
1.12 Hypothesis 62
1.13 Aims 64

2 Materials and Methods **65**
2.1 Materials and Reagents 66
2.2 Mice 66
2.3 DNA Isolation and Genotyping 71
2.4 Flow Cytometry and Fluorescent Activated Cell Sorting and Cell Analysis 75
2.4.1 Tissue Collection and Preparation 75
2.4.2 Peripheral Blood Analysis 75
2.4.3 Bone Marrow Isolation 76
2.4.4 May-Grünwald Giemsa Staining of the blood films and bone marrow cytospins 77
2.4.5 Spleen and Thymus Isolation 77
2.4.6 Flow Cytometry and Fluorescent Activated Cell Sorting 78
2.5 Bone Marrow Transplantation 79
2.6 RNA Extraction and cDNA Synthesis 82
2.7 RT-qPCR 82
2.8 Bulk and Single-Cell RNA-Sequencing 83
2.9 Bioinformatic Analysis 90
2.10 Cell Culture 91
2.10.1 General Cell Culture Technique and Handling . . . 91
2.10.2 Generation of Lentivirus and Retrovirus 92
2.10.3 Generation of Hoxb8 *R26*-CreER^{T2} *Srsf2*$^{P95H/+}$ and *Srsf2*$^{+/+}$ Cell lines 93
2.11 CRISPR Screen 95

2.11.1 Generation and Validation of Cas9-Blasticidin Hoxb8 *Srsf2*$^{P95H/+}$ And *Srsf2*$^{+/+}$ Cell Lines 95
2.11.2 Titration of Mouse Brie Library 96
2.11.3 Pooled CRISPR knock-out screen and parallel culture for RNA-Seq 96
2.11.4 Genomic DNA Preparation for CRISPR Screen . . . 97
2.11.5 Library Preparation for Illumina Sequencing 98
2.11.6 Bioinformatic Analysis for CRISPR Screen 102
2.12 ATP-Lite Cell Viability Assay – Endpoint *In Vitro* Drug Assay 103
2.13 Hoxb8 Cell Line Proliferation Assay Continuous *In Vitro* Drug Sensitivity Assay 104
2.14 Statistical Analysis 104

3 *Srsf2*$^{P95H/+}$ mutation cannot lead to fully penetrate MDS when activated in myeloid progenitors 105
3.1 Introduction 106
3.2 Results 108
3.2.1 Activation of *Srsf2*$^{P95H/+}$ mutation in myeloid progenitors is well tolerated in 20-30 week old animals 108
3.2.2 *Srsf2*$^{P95H/+}$ mutant cells show a differentiation bias towards the myeloid lineage 113
3.2.3 Activation of *Srsf2*$^{P95H/+}$ mutation in myeloid progenitors shows a mild myeloid bias and lymphoid suppression of haematopoiesis upon ageing 117
3.2.4 *Srsf2*$^{P95H/+}$ mutant cells display a myeloid bias and B lymphoid suppression in aged mice 122
3.3 Discussion 125

4 Transcriptome change of *Srsf2*$^{P95H/+}$ mutant cells 130

4.1 Introduction . . . 131

4.2 Results . . . 133

4.2.1 *Srsf2*$^{P95H/+}$ mutant stem cells exhibit a myeloid biased transcriptional signature . . . 133

4.2.2 mRNA processing and DNA repair are the top pathways represented in the mis-spliced transcripts in *SRSF2* mutant cells . . . 143

4.2.3 Single cell sequencing of *Srsf2*$^{P95H/+}$ stem and progenitor populations reveals a unique signature in the LT-HSC . . . 147

4.3 Discussion . . . 156

5 *In vitro* CRISPR screen identifies *Srsf2*$^{P95H/+}$ specific synthetic lethal pathways 161

5.1 Introduction . . . 162

5.2 Results . . . 164

5.2.1 Generation and characterization of *in vitro Srsf2* $^{P95H/+}$ cell lines . . . 164

5.2.2 Pooled CRISPRko negative selection screen in *Srsf2*$^{P95H/+}$ mutant cells . . . 167

5.2.3 CRISPR screen quality control and data analysis . . 170

5.2.4 Identification of RNA metabolism, Cell cycle and DNA damage response pathways as genetic vulnerabilities in *Srsf2*$^{P95H/+}$ mutant cells . . . 180

5.2.5 Pharmacological validation of *Srsf2* CRISPR Screen results . . . 188

5.3 Discussion . . . 192

6 *Srsf2*$^{P95H/+}$ co-operates with *Tet2* and *Cbl* mutations to promote myeloid bias and initiation of CMML **199**

6.1 Introduction 200

6.2 Results 202

6.2.1 h*Scl*-CreER^{T} *R26*eYFP *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ compound mutant mice show modest haematological change at 20 weeks post mutation activation 202

6.2.2 h*Scl*-CreER^{T} *R26*eYFP *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ mice develop more evident myeloid expansion and lymphoid suppression with time 209

6.2.3 h*Scl*-CreER^{T} *R26*eYFP *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ cells stably engraft but cannot compete with wild-type bone marrow in transplantation assays 220

6.2.3.1 Double mutant cells undergo myeloid biased differentiation at the expense of the lymphoid lineage 220

6.2.3.2 Double mutant cells still exhibit a myeloid bias but erythropoiesis is improved in aged stress haematopoiesis 226

6.2.4 *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ leads to development of CMML *in vivo* 234

6.2.5 *Srsf2*$^{P95H/+}$ *Cbl*$^{-/-}$ mutation leads to acceleration of myeloid proliferation 260

6.3 Discussion 269

7 Final discussion and conclusion **274**

A Publications associated with this thesis **281**

Bibiography **311**

List of Figures

1.1 The classical and newly proposed models of haematopoiesis. 4
1.2 Molecular changes underlying progression of CHIP in aged individuals 7
1.3 Current treatment roadmap for MDS (adapted from (Bejar and Steensma, 2014)) 18
1.4 Frequency of recurrent mutations and their correlations in MDS and CMML 24
1.5 Schematic of RNA splicing and aberrant splicing events . . 27
1.6 SRSF2 mutation frequently occurs at P95 and leads to a change in RNA binding site recognition 33
1.7 DNA methylation and demethylation by DNMT3A and TET2 40
1.8 CBL mutations occur at the linker-RING finger domain . . 48
1.9 Detailed genomic architecture of the *Srsf2*$^{P95H/+}$ knock-in mutant models. 56

3.1 *LysM*-Cre *R26*eYFP *Srsf2*$^{P95H/+}$ show no haematological change at 20 to 30 weeks of age. 110
3.2 *LysM*-Cre *R26*eYFP *Srsf2*$^{P95H/+}$ show limited lymphoid change in the bone marrow at 20 to 30 weeks of age 111
3.3 *LysM*-Cre *R26*eYFP *Srsf2*$^{P95H/+}$ have a decreased spleen cellularity due to loss of lymphoid populations at 20 to 30 weeks of age 112

3.4 *Srsf2*$^{P95H/+}$ expressing (eYFP positive) cells display a mild myeloid bias at the expense of the B lymphoid population 116
3.5 Aged *LysM*-Cre *R26*eYFP *Srsf2*$^{P95H/+}$ mice develop macrocytic anaemia with increased granulocytes in the peripheral blood . . . 119
3.6 Aged *LysM*-Cre *R26*eYFP *Srsf2*$^{P95H/+}$ mice have decreased lymphoid and erythroid populations in the bone marrow . 120
3.7 Aged *LysM*-Cre *R26*eYFP *Srsf2*$^{P95H/+}$ mice display a mild myeloid bias with lymphoid and erythroid suppression in the spleen. . . . 121
3.8 *Srsf2*$^{P95H/+}$ expressing (eYFP positive) cells in aged mice display a profound myeloid bias in bone marrow and spleen 124

4.1 The sorting strategy of *Srsf2*$^{P95H/+}$ stem and progenitor cells for h*Scl*-CreER$^{T\ Tg+}$ R26eYFP *Srsf2*$^{P95H/+}$ and transplant recipients of *R26*-CreER^{T2} *Srsf2*$^{P95H/+}$ . . . 135
4.2 Transcriptome analysis demonstrates an up-regulation of myeloid specific transcripts and down regulation of lymphoid associated genes in *Srsf2*$^{P95H/+}$ stem and progenitor cell . . . 137
4.3 Skipped exons are the most frequently mis-spliced event in *Srsf2*$^{P95H/+}$ cells . . . 141
4.4 *In vitro* splicing changes caused by *Srsf2*$^{P95H/+}$ expression mirror the *in vivo* changes . . . 142
4.5 mRNA processing and DNA repair are the top mis-spliced pathways caused by *SRSF2* mutation . . . 146
4.6 The sorting strategy of long-term haematopoietic stem cells and progenitor populations for single-cell sequencing . . . 149

4.7 UMAP plots of all cells sequenced (A) and separate samples of *Srsf2*$^{+/+}$ (LT-HSC/Prog) and *Srsf2*$^{P95H/+}$ (LT-HSC/Prog) (B) 151
4.8 Cluster comparison identifies a myeloid biased signature in *Srsf2*$^{P95H/+}$ haematopoietic stem cells 154
4.9 The trajectory plot of *Srsf2*$^{+/+}$ and *Srsf2*$^{P95H/+}$ samples reveals increased myeloid differentiation in the mutant cells 155

5.1 Generation and characterization of Hoxb8 immortalized cell lines and Cas9-blasticidin resistant cell lines 166
5.2 Outline of the pooled CRISPR knock-out *Srsf2* screen . . . 169
5.3 The sgRNA distribution and dynamics of the sgRNA pool in the *Srsf2* CRISPR screen 174
5.4 The correlation between the six Cas9 *Srsf2* cell lines across the time course of the screen 175
5.5 Depletion of the CEG essential genes at day 4 of the CRISPR screen 176
5.6 Visualisation of MDS related genes that are positively and negatively selected at day 4 (A), day 11 (B) and day 18 (C) of *Srsf2* CRISPR screen. 179
5.7 Overview of *Srsf2* negative selection CRISPR screen results 184
5.8 The number of negatively selected genes using different FDR cut-offs 185
5.9 Disruption of DNA repair, DNA replication, cell cycle and RNA metabolic pathways leads to synthetic lethality in *Srsf2*$^{P95H/+}$ mutant cells 186
5.10 Cell cycle genes, *Pkmyt1* and *Cdk6* are among the top 30 genes that are only negatively selected in the *Srsf2*$^{P95H/+}$ cell lines at day 18 187

5.11 *Srsf2*$^{P95H/+}$ cells are sensitive to the CDK6 inhibitor Palbociclib, as well as to the spliceosome inhibitor Pladienolide B 190
5.12 CDK6 inhibitor Palbociclib reduces the proliferation of *Srsf2*$^{P95H/+}$ cells while spliceosome inhibitor Pladienolide B is lethal for both genotypes in long-term *in vitro* culture. 191

6.1 Co-expression of *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ leads to macrocytic changes rapidly after gene mutation 204
6.2 Co-expression of *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ leads to hypocellular bone marrow at early time points post gene mutation . 206
6.3 Co-expression of *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ has no effect on stem and progenitor populations after 20 weeks of mutation activation 207
6.4 Co-expression of *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ leads to mild myeloid bias in the spleen after 20 weeks of mutation activation . . 208
6.5 Co-expression of *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ leads to macrocytic red blood cells and lymphoid suppression in the peripheral blood 52 weeks after mutation activation 212
6.6 Co-expression of *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ leads to significant myeloid bias and B cell suppression in the bone marrow 52 weeks after gene mutation 214
6.7 The eYFP positive *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ bone marrow cells are significantly biased towards the myeloid lineage instead of B lymphoid lineage 215
6.8 Co-expression of *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ leads to an increased proportion of phenotypic short-term haematopoietic stem cells 217
6.9 *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ mutation leads to myeloid bias and B cell suppression in the spleen 219

6.10 *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ cells have decreased engraftment capacity and strong myeloid bias with B cell suppression 223
6.11 *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ cells are out-competed by wild-type cells upon competitive transplantation 225
6.12 *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ cells can stably engraft haematopoietic tissues 52 weeks post transplantation 229
6.13 *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ cells are out-competed by wild-type cells and fail to efficiently engraft haematopoietic tissues at 52 weeks post transplantation 230
6.14 PB indices and lineage distribution of non-competitive transplant recipients at 52 weeks post transplantation . . . 231
6.15 *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ non-competitive transplant recipients show alterations in myeloid, lymphoid and erythroid lineages in the bone marrow at 52 weeks post transplantation 233
6.16 Sick *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ mice display features consistent with CMML including splenomegaly, dysplastic neutrophils and an increased number of immature monocytes 236
6.17 *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ mouse #359 developed myeloid bias and significant B cell suppression at 52 weeks after tamoxifen treatment 238
6.18 Moribund *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ mice displayed severe disruption of lymphoid and erythroid lineages, with myeloid bias in the bone marrow 241
6.19 The myeloid bias and lymphoid suppression is evident in the eYFP positive fraction of peripheral blood and bone marrow in #359 243
6.20 Sick mouse #359 has splenomegaly with significant myeloid expansion, erythroid compensation and near absence of B cells 245

6.21 *Srsf2*$^{P95H/+}$*Tet2*$^{-/-}$ non-competitive transplant recipients #835 and #836 developed leukocytosis, with granulocytosis and monocytosis in the peripheral blood 248
6.22 *Srsf2*$^{P95H/+}$*Tet2*$^{-/-}$ leads to hypercellularity and myeloid bias in the bone marrow and spleen of sick non-competitive transplant recipients 250
6.23 Cells expressing *Srsf2*$^{P95H/+}$*Tet2*$^{-/-}$ mutation drive the myeloid bias and B cell suppression in the non-competitive transplant recipients #835 and #836 253
6.24 FACS plots from the moribund *Srsf2*$^{P95H/+}$*Tet2*$^{-/-}$ mice showing myeloid bias, monocytosis and B cell suppression in the bone marrow (A) and spleen (B-C) 256
6.25 FACS plots of the eYFP positive population from the sick *Srsf2*$^{P95H/+}$*Tet2*$^{-/-}$ mice showing myeloid bias, monocytosis and B cell suppression in the bone marrow (A) and spleen (B-C) 259
6.26 Peripheral blood analysis of Srsf2$^{P95H/+}$ *Cbl*$^{-/-}$ compound mutant mice 262
6.27 Bone marrow analysis of *Srsf2*$^{P95H/+}$ *Cbl*$^{-/-}$ compound mutant mice 264
6.28 Spleen analysis of Srsf2$^{P95H/+}$ *Cbl*$^{-/-}$ compound mutant mice 265
6.29 Myeloid-hyperplasia in both the bone marrow and peripheral blood of sick Srsf2$^{P95H/+}$ *Cbl*$^{-/-}$ mice, including within the eYFP positive fraction 267
6.30 FACS plots showing a mild increase in the macrophage population in both bone marrow and peripheral blood of sick mice, including within the eYFP positive fraction. Time labels indicate weeks post tamoxifen cessation 268

List of Tables

1.1 Diagnostic criteria for MDS (Adapted from (Arber et al., 2016)) . . . 10
1.2 Risk assessment for MDS based on IPSS-2012 (Adapted from (Greenberg et al., 2012; Steensma, 2018)) . . . 12
1.3 Risk group and their scores (Adapted from (Greenberg et al., 2012; Steensma, 2018) . . . 12
1.4 Frequency of RNA splicing mutations in haematological malignancies . . . 29
1.5 Frequency of DNA methylation mutations in adult haematological malignancies (Adapted from (Brunetti et al., 2017; Guillamot et al., 2016)) . . . 44
1.6 Summary of the phenotypes of the *Srsf2* and *Tet2* mutant mouse models . . . 58

2.1 Reagents and materials used in experiments and the source 67
2.2 List of solutions and buffers prepared for use in experiments 69
2.3 List of equipment and their manufacture used in the experiments . . . 70
2.4 Genotyping Oligonucleotide sequences and product size . 73
2.5 Genotyping PCR mixture . . . 74
2.6 Genotyping PCR cycling conditions . . . 74
2.7 List of FACS antibodies used in the experiments . . . 80
2.8 RT-qPCR oligonucleotide list . . . 84

2.9 Oligonucleotide list for exon exclusion RT-qPCR 85
2.10 RT-qPCR cycling conditions 85
2.11 rRNA depletion-Hybridisation mix 88
2.12 rRNA depletion-Hybridisation cycling conditions 88
2.13 rRNA depletion-RNase H reaction mix 88
2.14 rRNA depletion-DNase digestion mix 89
2.15 Sequences of Multiplexing PCR primers used to generate RNA-seq libraries 89
2.16 Lentiviral transfection mix 93
2.17 Retroviral transfection mix 93
2.18 P5 primer sequences 99
2.19 P7 primer sequences 100
2.20 CRISPR DNA library thermocycling parameters 102

4.1 Datasets used for the comparison of *SRSF2*/*Srsf2* mis-spliced genes 144
4.2 List of top mis-spliced genes ranked by the number of datasets the transcript was identified in 147
4.3 Estimated number of cells processed for each sample . . . 150
4.4 The number of different cell types retrieved from each sample 150

5.1 Mapped read count and coverage for 24 libraries in *Srsf2* screen 172

List of Abbreviations

2HG	2-hydroxyglutarate
AGRF	Australian Genome Research Facility
AlloSCT	Allogenic haematopoietic stem cell transplantation
AML	Acute myeloid leukaemia
ARCH	Age-related clonal haematopoiesis
aUPD	acquired uniparental disomy
bcCML	Blast-crisis chronic myeloid leukaemia
BER	Base excision repair
BM	Bone marrow
BMMNC	Bone marrow mononuclear cell
CB2	CRISPRBetaBinomial package
CDR	Critical deleted region
CEG	Core essential gene
CFU-E	Colony-forming unit erythroid
CHIP	Clonal haematopoiesis of intermediate potential
CLP	Common lymphoid progenitor
CMML	Chronic myelomonocytic leukaemia
CMP	Common myeloid progenitor
CPM	Counts per million
CPSF	Cleavage and polyadenylation specificity factor
CR	Complete remission
Cre	Cre recombinase
CRISPR	Clustered regularly interspaced short palindromic repeats

CSTF	Cleavage stimulating factor
CTD	Carboxyl-terminal domain
DBA	Diamond-Blackfan Anaemia
DGIdb	Drug Gene Interaction database
DMR	Differentially methylated region
DMSO	Dimethylsulfoxide
DN	Double negative
DNMT	DNA methyltransferase
DP	Double positive
DSB	Double-strand break
DSBH	Double-stranded β-helix
EDTA	Ethylenediaminetetraacetic acid
ESA	Erythropoiesis-stimulating agent
ESE	Exonic splicing enhancer
ESS	Exonic splicing silencer
eYFP	Enhanced yellow fluorescent protein
FACS	Fluorescent activated cell sorting
FBS	Fetal bovine serum
FDR	False discovery rate
gDNA	Genomic DNA
GeCKO	Genome-wide CRISPRko library
GFP	Green fluorescent protein
GM-CSF	granulocyte-macrophage colony-stimulating factor
GMP	Granulocyte-macrophage progenitor
GO	Gene ontology
HBS	Hank's balanced salt solution
HMA	Hypomethylating agent
hnRNP	Heterogeneous nuclear ribonucleoprotein
HSC	Haematopoietic stem cell

IDT	Integrated DNA Technology
IFN	interferon
IgM	Immunoglobin M
IL	Interleukin
IMDM	Iscove Modified Dulbecco Media
ISE	Intronic splicing enhancer
ISS	Intronic splicing silencer
JMML	Juvenile myelomonocytic leukaemia
LMPP	Lymphoid-primed multipotent progenitor
LOH	Loss of heterozygosity
LT-HSC	Long-term haematopoietic stem cell
MAPK	Mitogen-activated protein kinase
MCV	Mean cell volume
MDS	Myelodysplastic syndromes
MDS/MPN	Myelodysplastic/myeloproliferative syndromes
MEP	Megakaryocyte-erythroid progenitor
MKP	Megakaryocyte progenitor
MLL	Mixed lineage leukaemia
MLP	Myeloid/lymphoid progenitor
mLT-HSC	myeloid biased LT-HSC
MOI	Multiplicity of infection
MP	Myeloid progenitor
MPN	Myeloproliferative neoplasm
MPP	Multipotent progenitor
NCBP	Nuclear cap binding complex
NGS	Next-generation sequencing
NHEJ	Non homologous end-joining
NMD	Non-sense mediated decay
PB	Peripheral blood

PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDX	Patient-derived xenograft
PFA	Paraformaldehyde
pIpC	Polyinosinic:polycytidylic acid
PLB	Pladienolide B
PLT	Platelet
Pre-CFU-E	Pre colony-forming unit erythroid
Pre-GM	Pre-granulocyte-monocyte progenitor
Pre-MegE	Pre-megakaryocyte progenitor
Pre-mRNA	Pre-messenger RNA
PTB	Polypyrimidine tract binding
QuSage	Quantitative Set Analysis for Gene Expression Analysis
RBC	Red blood cell
ROC	Receiver-operating characteristic (ROC)
Rpm	Revolutions per minute
RRM	RNA recognition motif
rRNA	Ribosomal RNA
RS	Arginine/serine-rich domain
RT	Room temperature
RT-qPCR	Real-time quantitative polymerase chain reaction
RTK	Receptor tyrosine kinase
SCF	Stem cell factor
scRNA-seq	Single cell RNA sequencing
sgRNA	Single-guide RNA
snoRNA	small nucleolar RNA
SR	Serine/arginine-rich
SRE	Splicing regulatory element
ST-HSC	Short-term haematopoietic stem cell

T-ALL	T-cell acute lymphoblastic leukaemia
TET	Tet-eleven translocation
TKB	tyrosine kinase-binding domain
TKO	Toronto KnockOut library
UBCs	Ubiquitin-carrier enzyme
UMAP	Uniform Manifold Approximation and Projection
UMI	Unique Molecular Identifiers
UTR	Untranslated region
VAF	Variant allele frequency
WES	Whole exome sequencing
WHO	World Health Organisation
WT	Wild type

Chapter 1

Introduction and background

1.1 Haematopoiesis

Haematopoiesis, or blood cell production, is a highly regulated process generating all cells across different lineages, including myeloid, lymphoid, erythroid and megakaryocyte lineages. The process starts from haematopoietic stem cells (HSCs), which then differentiate to lineage-specific progenitors and subsequently mature blood cells (Figure 1.1). Studies over the years have established and characterised a comprehensive list of phenotypically defined populations in both human and mouse, which primarily have used cell surface markers to distinguish discrete cell populations (Adolfsson et al., 2005; Kiel et al., 2005; Osawa et al., 1996; Pronk et al., 2007). HSCs are the most primitive population of blood-forming cells, which hold both self-renewal and differentiation capacity. Within the phenotypically defined populations, long-term haematopoietic stem cells (LT-HSC) ($LKS^+CD150^+CD48^-$ or $LKS^+CD34^-Flt3^-$) are the most quiescent population and have serial long-term repopulating capacity as demonstrated in transplantation analysis. Upon differentiation, HSCs lose self-renewal capacity and mature into multipotent progenitors (MPPs) ($LKS^+CD34^+Flt3^+$) and then subsequently progress to lineage-restricted oligopotent progenitors (LKS^-), such as common myeloid progenitors (CMP; $LKS^-CD34^+FcyRII^-$), granulocyte-macrophage progenitors (GMP; $LKS^-CD34^+FcyRII^+$) and megakaryocyte-erythroid progenitors (MEP; $LKS^-CD34^-FcyRII^-$). The overall view of haematopoiesis is progressively evolving, however the general populations and lineage commitments described above provide a good general understanding of the lineage commitment and differentiation that underlies the formation of the blood system.

In the last few decades, our understanding of haematopoiesis has

expanded dramatically from defining the key populations in blood production using cell surface markers to elucidating the lineage commitment and self-renewal using single-cell DNA/RNA sequencing (Buenrostro et al., 2018; Jacobsen and Nerlov, 2019; Laurenti and Göttgens, 2018; Nestorowa et al., 2016). Many studies performed transplantation of purified populations or single cells, or have used *in vivo* cell labelling to trace the lineage potential of each population. Instead of homogeneous populations as suggested by the phenotypically defined approaches, these highly purified cells have been found to have functionally distinct subsets that are heterogeneous in their repopulating behaviour (Biasco et al., 2016; Busch et al., 2015; Chapple et al., 2018; Pei et al., 2017; Sawai et al., 2016; Sawen et al., 2018; Yamamoto et al., 2013). Evidence of lineage priming in the HSCs also exits, with such lineage-restricted repopulating cells still harbouring self-renewal capacity (Carrelha et al., 2018). Nevertheless, most mature blood cells are produced by MPPs upon demand, in both native haematopoiesis and after bone marrow transplantation (Pietras et al., 2015). Through *in vivo* lineage tracing studies in the setting of native haematopoiesis, it has become apparent that homeostatic haematopoiesis is mainly driven by non-engraftable ST-HSC and progenitors (Busch et al., 2015; Christodoulou et al., 2020; Rodriguez-Fraticelli et al., 2018; Sun et al., 2014). These findings have deconvoluted the heterogeneity of HSCs and challenged the conventional hierarchical structure defined by cell surface markers. Instead of a stepwise bifurcation model of haematopoiesis, newer models propose that clones of lineage primed stem cells self-renew and initiate differentiation upon physiological demand (Klose et al., 2019; Velten et al., 2017). Further studies are needed to test this hypothesis and to understand the full complexity of haematopoiesis.

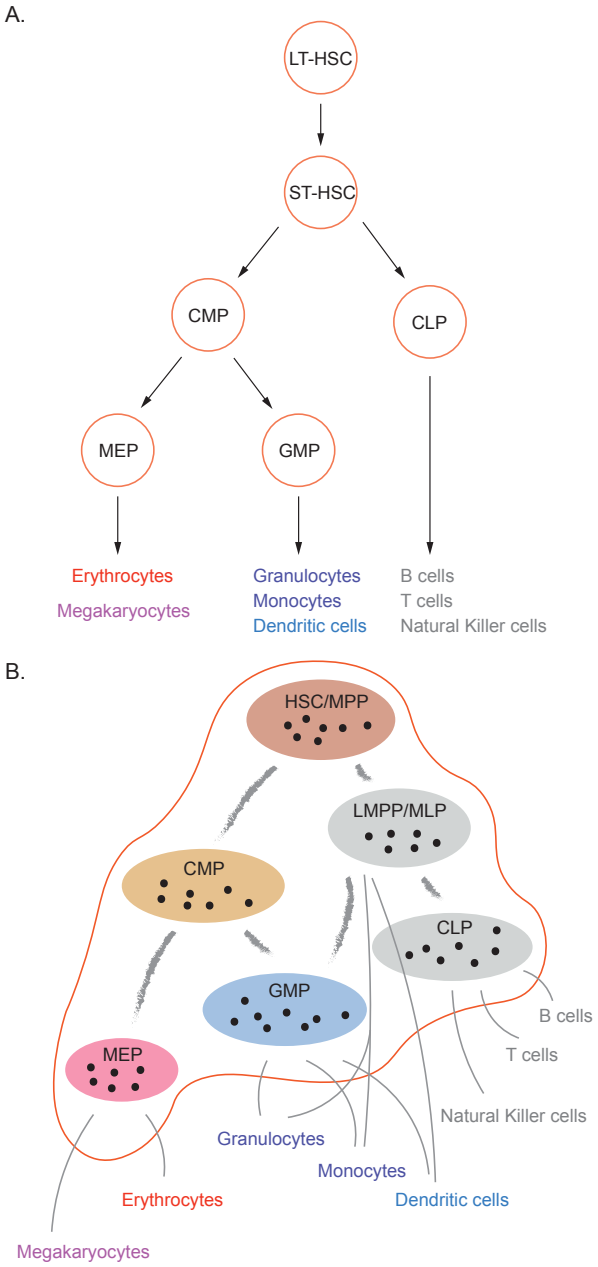


Figure 1.1: The classical and newly proposed models of haematopoiesis.
A. The classical hierarchy of haematopoiesis based on immunophenotyping.
B. The recently proposed model of haematopoiesis based on single-cell transcriptomes and single-cell transplantation. Connecting lines depict differentiation trajectories.
LT-HSC, long-term haematopoietic stem cells; ST-HSC, short-term haematopoietic stem cells; CMP, common myeloid progenitors; CLP, common lymphoid progenitors; LMPP, lymphoid-primed multipotent progenitors; MLP, myeloid/lymphoid progenitors; MEP, megakaryocyte-erythrocyte progenitors; GMP, granulocyte-macrophage progenitors.

1.2 HSC ageing and clonal haematopoiesis of indeterminate potential (CHIP)

Ageing of HSCs presents as a decrease in self-renewal capacity and lymphoid output (Young et al., 2016b). On one hand, ageing could be a cell autonomous behaviour, as demonstrated in the transplantation study of aged vs young stem cells (Verovskaya et al., 2013). On the other hand, cell non-autonomous factors, such as age-dependent DNA damage, epigenetic and micro-environment changes are linked with defects in the aged stem cells (Konieczny and Arranz, 2018). However, how these changes lead to specific phenotypes warrants further investigation.

A key phenotype linked to ageing is an increased bias towards the myeloid lineage at the expense of lymphoid lineages, which coincides with increased frequencies of myeloid malignancies in the elderly population (Beerman et al., 2010; Weiskopf et al., 2016). As mentioned before, HSCs are heterogeneous and composed of many clones with restricted lineage potential. Aged individuals are found to have a decreased diversity in the clonal population, with a specific skewing towards clones with myeloid potential. The shift in differentiation is mainly driven by molecular changes, with an upregulation of myeloid-associated genes and the downregulation of lymphoid-associated genes (Grover et al., 2016; Pang et al., 2011; Rundberg Nilsson et al., 2016). However, inflammation and ageing of the stem cell niche are also important co-factors in the modification of HSC function (Ho et al., 2019).

More recently, recurrent mutations in cancer-associated genes were discovered in the blood of aged individuals who did not have any known haematological diseases (McKerrell et al., 2015; Xie et al., 2014; Young et al., 2016a; Zink et al., 2017). Approximately 2% of the general population,

and increasing to 5-6% in people over the age of 70, are found to carry leukaemia/lymphoma associated mutations in such genes as *DNMT3A, TET2, JAK2, ASXL1, TP53, SF3B1 and SRSF2*. Many of these genes are known driver mutations in myeloid malignancies such as myelodysplastic syndromes (MDS), myeloproliferative neoplasm (MPN) and acute myeloid leukaemia (AML). In healthy individuals, these mutations drive clonal haematopoiesis, expansion of HSCs and progenitor clones, and lead to a significantly increased risk of developing haematological cancers at a later stage and with an overall higher mortality (Abelson et al., 2018; Genovese et al., 2014; Jaiswal et al., 2014). The presence of these mutations and the development of clonal haematopoiesis with age is now described as age-related clonal haematopoiesis (ARCH) or clonal haematopoiesis of indeterminate potential (CHIP). Like monoclonal gammopathy of undetermined significance and monoclonal B-cell lymphocytosis, CHIP is considered a precursor state of haematological malignancies, but unlike the B cell related disorders it does not often progress and does not require medical intervention (Steensma et al., 2015) (Figure 1.2).

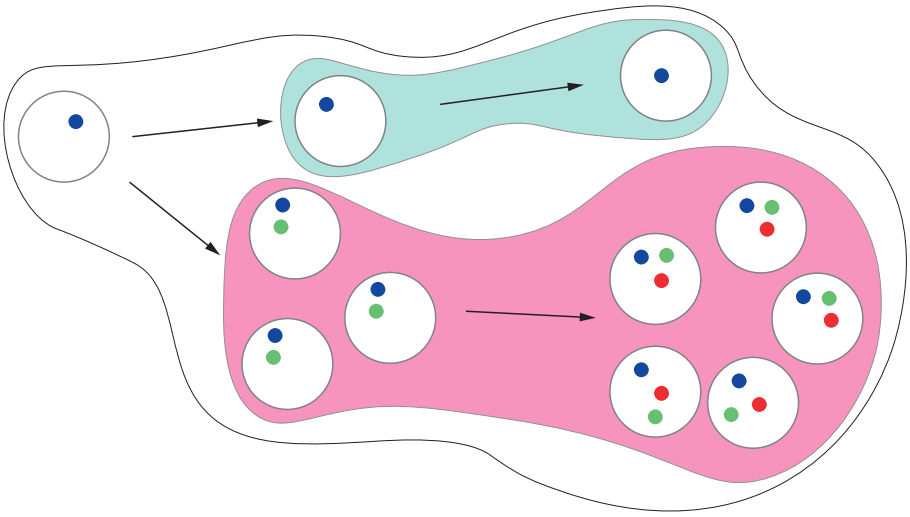


Figure 1.2: Molecular changes underlying progression of CHIP in aged individuals
Haematopoietic stem cells acquire mutations randomly during ageing. While some mutations do not have roles in disease (blue dots), certain mutations (green dots) initiate clonal expansion (pink shaded area) and lead to subsequent co-operating mutations (red dots) in the daughter cells. Light green shaded area represents cells without clonal expansion.

1.3 Myelodysplastic syndromes and Chronic myelomonocytic leukaemia

1.3.1 Myelodysplastic syndromes (MDS)

Myelodysplastic syndromes (MDS) are a group of heterogeneous clonal stem cell disorders. Pathology findings are cytopenia (reduction in cell number in the peripheral blood) and dysplasia (change of cell morphology) of one or multiple blood lineages (myeloid, erythroid/red blood cells and platelets) caused by ineffective haematopoiesis. Clinical manifestations are often mild or absent in the early stages, but it can progress to fatigue, recurrent infections and easy bruising in the later stages. The prevalence of MDS increases with age and has a higher incidence in men than women in the US (4.5 vs 2.7 per 100,000 per year) (Ma et al., 2007). In Victoria, the annual incidence of MDS in the aged population (65 and older) is 103 per 100,000, while the age-standardized incidence is 4 to 5.8 per 100,000 (Australian Institute of and Welfare, 2013; McQuilten et al., 2014). These statistics suggest a significant underrepresentation and unappreciation of the true extent of MDS in the community.

Because of its heterogeneous nature, MDS is categorised based on morphological appearance (dysplasia), cytogenetic abnormality (del(5q)), as well as blast cell percentage according to the World Health Organisation (WHO) and French-American-British (FAB) diagnosis criteria (Arber and Hasserjian, 2015; Arber et al., 2016; Vardiman, 2012). The latest revision of WHO diagnostic criteria acknowledged the mounting evidence of recurrent genetic mutations in MDS, which prompted a new category of MDS with ring sideroblasts (MDS-RS) and multilineage dysplasia, without excess blasts or del(5q) (Montalban-Bravo and Garcia-Manero, 2018) (Table 1.1). Patients can have a diagnosis of MDS-RS if an SF3B1

mutation is identified, even if only 5% ring sideroblasts are present. At the time of the diagnosis, the risk stratification is also conducted to predict the patient's outcome and determine treatment strategies. The most common tool used is the 2012 version of International Prognostic Scoring system (IPSS-2012) (Greenberg et al., 2012). IPSS-2012 stratifies patients into lower risk MDS (Very good and Good), intermediate or higher risk MDS (Poor and very Poor) based on the severity of cytopenias, bone marrow blast counts and cytogenetic abnormalities (Tables 1.2 and 1.3). More recently, the emerging connections between molecular abnormalities, disease progression and treatment outcome are being recognised in clinical diagnosis. Certain mutations are associated with poor outcomes, such as *TP53, NRAS, ASXL1, EZH2* and *SRSF2*, while *SF3B1* mutation is associated with more favourable outcomes (Bejar et al., 2011; Bejar et al., 2012; Papaemmanuil et al., 2013).

Table 1.1: Diagnostic criteria for MDS (Adapted from (Arber et al., 2016))

Category	Dysplastic lineages	Cytopenias*	Ring sideroblasts as % of marrow erythroid elements	BM and PB blasts	Cytogenetics by conventional karyotype analysis
MDS with single lineage dysplasia (MDS-SLD)	1	1 or 2	<15% or <5% if *SF3B1* mutation is present	BM <5%, PB <1%, no Auer rods	Any, unless fulfills all criteria for MDS with isolated del(5q)
MDS with multilineage dysplasia (MDS-MLD)	2 or 3	1-3	<15% or <5% if *SF3B1* mutation is present	BM <5%, PB <1%, no Auer rods	Any, unless fulfills all criteria for MDS with isolated del(5q)
MDS with ring sideroblasts (MDS-RS)					
MDS-RS with single lineage dysplasia (MDS-RS-SLD)	1	1 or 2	≥ 15% or ≥ 5% if *SF3B1* mutation is present	BM <5%, PB <1%, no Auer rods	Any, unless fulfills all criteria for MDS with isolated del(5q)
MDS-RS with multilineage dysplasia (MDS-RS-MLD)	2 or 3	1-3	≥ 15% or ≥ 5% if *SF3B1* mutation is present	BM <5%, PB <1%, no Auer rods	Any, unless fulfills all criteria for MDS with isolated del(5q)
MDS with isolated del(5q)	1-3	1-2	None or any	BM <5%, PB <1%, no Auer rods	Del(5q) alone or with 1 additional abnormality

					except -7 or del(7q)
MDS with excess blasts (MDS-EB)					
MDS-EB-1	0-3	1-3	None or any	BM 5-9% or PB 2-4%, no Auer rods	Any
MDS-EB-2	0-3	1-3	None or any	BM 10-19% or PB 5-19% or Auer rods	Any
MDS, unclassifiable (MDS-U)					
With 1% blood blasts	1-3	1-3	None or any	BM <5%, PB =1%†, no Auer rods	Any
With single lineage dysplasia and pancytopenia	1	3	None or any	BM <5%, PB <1%, no Auer rods	Any
Based on defining cytogenetic abnormality	0	1-3	<15%	BM <5%, PB <1%, no Auer rods	MDS-defining abnormality
Refractory cytopenia of childhood	1-3	1-3	None	BM <5%, PB <2%,	Any

*Cytopenias defined as: haemoglobin <10g/dL, platelet count<$100x10^9$/L, and absolute neutrophil count<$1.8 x10^9$/L.
†One percent PB blasts must be recorded on at least 2 separate occasions.
SLD, single lineage dysplasia; MLD, multilineage dysplasia.

Table 1.2: Risk assessment for MDS based on IPSS-2012 (Adapted from (Greenberg et al., 2012; Steensma, 2018))

Parameter	**Categories and Associated Scores (**Scores in bold**)**				
Cytogenetic risk group[a]	Very good	Good	Intermediate	Poor	Very poor
	0	**1**	**2**	**3**	**4**
Bone marrow blast proportion	≤ 2%	>2-5%	5-<10%	≥ 10%	
	0	**1**	**2**	**3**	
Haemoglobin	≥ 10 g/dL	8-<10 g/dL	<8 g/dL		
	0	**1**	**1.5**		
Absolute neutrophil count	≥ 0.8 x 10_9/L	<0.8 x 10_9/L			
	0	**0.5**			
Platelet count	≥ 100 x 10^9/L	50-100 x 10^9/L	<50 x 10^9/L		
	0	**0.5**	**1**		

[a]Cytogenetic risk group, very good: -Y, del(11q); good: normal; del(5q)±1 other abnormality del(20q), or del(12p); intermediate:+8, i(17q), del(7q),+19, any other abnormality not listed including the preceding with 1 other abnormality; poor: -7±del(7q), inv(3)/t(3q)/del(3q), any 3 separate abnormalities; very poor: more than 3 abnormalities, especially if 17p is deleted or rearranged

Table 1.3: Risk group and their scores (Adapted from (Greenberg et al., 2012; Steensma, 2018)

Risk group	**Total score**
Very low	0-1.0
Low	1.5-3.0
Intermediate	3.5-4.5
High	5.0-6.0
Very high	>6.0

*The score is on a 0-10 point scale

1.3.2 Myelodysplastic/myeloproliferative syndromes (MDS/MPN) and Chronic myelomonocytic leukaemia (CMML)

Myelodysplastic/myeloproliferative neoplasms (MDS/MPN) are a group of haematological malignancies that share clinical features of both MDS and MPN. These disorders include chronic myelomonocytic leukaemia (CMML), juvenile myelomonocytic leukaemia (JMML), atypical BCR-ABL negative chronic myeloid leukaemia (aCML), refractory anaemia with ring sideroblasts and thrombocytosis (RARS-T) as a provisional entry, and MDS/MPN unclassified (MDS/MPN-U). Each MDS/MPN category has its own unique features as opposed to those classified as pure MDS or MPN, such as monocytosis in CMML, thrombocytosis in RARS-T and dysplastic granulocytes in aCML (Thota and Gerds, 2018). It is a challenge to properly diagnose MDS/MPN, despite the revised WHO 2016 diagnostic criteria (Savona et al., 2015; Smith et al., 2019). However, recent advances in the identification of genetic abnormalities in MDS/MPN, such as the high prevalence of *SRSF2* and *TET2* mutations in CMML, have provided hope for better diagnostic and prognostic subgrouping of MDS/MPN through inclusion of genetic descriptors (Pati and Kundil Veetil, 2019).

Chronic myelomonocytic leukaemia (CMML) was previously classified as a subtype of MDS but in the latest WHO classification is now considered a separate entity within the myeloid neoplasms (Arber et al., 2016; Patnaik and Tefferi, 2018). It is one of the most aggressive chronic leukaemias with few treatment options. The incidence rate is 0.4 per 100,000 with most cases reported in older adults (Loghavi and Khoury, 2018). CMML is sub-categorised into CMML-1 (<5% peripheral blasts and <10% marrow blasts) or CMML-2 (5-19% peripheral blasts and 10-19% marrow blasts) based on the WHO classification. Currently, the FAB

classifies CMML into CMML-MD (white blood cell count < 13 x10 ^{9}/L) or CMML-MP (white blood cell count > 13 x10 ^{9}/L) (Bennett et al., 1994). Both classifications are utilised in the literature and clinical diagnosis. Compared to MDS, CMML patients have a poorer overall outcome and are more likely to transform to acute leukaemia (Parikh and Tefferi, 2013). Patients with CMML have many of the clinical presentations of MDS, but exhibit distinct proliferative features, in particular monocytosis. In fact, a diagnosis of CMML can be made if the patient presents with persistent unexplained monocytosis for more than 3 months, even in the absence of dysplastic features. More recently, Selimoglu-Buet et al. suggested that the increase in classical monocytes ($CD14^{+}CD16^{-}$), detected in CMML compared to reactive monocytosis or other haematological malignancies, could contribute to the diagnosis of CMML (Selimoglu-Buet et al., 2015). Other clinical manifestations, such as hepatosplenomegaly are more common in CMML than MDS as well.

1.3.3 Treatment options for MDS and CMML

Currently, the treatment approaches for MDS are largely based on the risk assessment and the clinical manifestation of each patient (Tables 1.2 and 1.3) (Figure 1.3). Clinical monitoring is recommended for low-risk patients, such as intermittent blood counting instead of active intervention if it is safe. Once patients develop severe cytopenia or an increased blast population, low intensity therapy, such as erythroid stimulating agents (Epoetin or Darbepoetin), myeloid growth factors (G-CSF) and thrombopoietin receptor agonist (TPO-RA), and anti-T cell immunosuppressive therapy (IST) can be given as supportive care. If deletion of the long arm of chromosome 5 is present, lenalidomide is also recommended (Fenaux and Adès, 2013; Steensma et al., 2008). However, if a patient becomes resistant to the treatment or is assessed as high-risk, the only remaining

options are hypomethylating agents (HMAs) or allogeneic haematopoietic stem cell transplantation (AlloSCT). I will discuss HMAs in more detail in Section 1.7.4. AlloSCT is the only curative treatment for MDS, but the number of patients that can undergo AlloSCT is very limited as a result of ineligibility due to the patients' age and comorbidities and limited donor resources. Most patients eventually progress to secondary leukaemia and succumb to the disease (Malcovati et al., 2013; Nachtkamp et al., 2016).

New effective agents are urgently needed for a better outcome for MDS patients. There are many challenges in developing such therapies, including clonal heterogeneity and a lack of understanding regarding how to effectively target the recurrent mutations. MDS is a disease of the stem cells (Woll et al., 2014). The recurrent mutations are thought to arise from the HSC pool, which then undergoes a complex clonal evolution process at the beginning of disease initiation and during disease progression (Papaemmanuil et al., 2013; Sperling et al., 2017). I will discuss the genetic landscape and clonal evolution of MDS in more detail in Section 1.4.

Patients that have undergone unsuccessful conventional treatments are recommended to enrol in clinical trials. Currently, many clinical trials are investigating the effect of combination therapy using multiple agents with or without AlloSCT in MDS patients. Some have shown to be effective in other myeloid neoplasms, such as kinase inhibitors, cytotoxic agents and immunomodulatory agents (Bejar and Steensma, 2014). Combination of azacitidine with lenalidomide or azacitidine with vorinostat, a deacetylase inhibitor, is also being tested (Platzbecker and Germing, 2013; Silverman et al., 2013). While many studies are currently ongoing, early results indicate that a combined cocktail may be more beneficial and produce longer lasting results for many MDS patients (Assi

et al., 2018; Kim et al., 2019; Voso et al., 2017). However, relapse remains a challenge and different therapeutic approaches are urgently needed.

For CMML, treatment is based on stratification of CMML subtypes when transplantation is unavailable. CMML-MD requires mostly treatment to alleviate symptom burden, such as erythropoiesis-stimulating agents (ESAs) for anaemia. Whereas, CMML-MP is managed more aggressively with cytoreductive agents such as hydroxyurea or HMAs in severe cases (Onida et al., 2013; Padron and Steensma, 2015; Solary and Itzykson, 2017). There is a huge gap to fill in terms of therapeutic options for CMML patients as currently only $\sim$20% of CMML-1 patients and $\sim$10% of CMML-2 patients survive longer than 5 years after initial diagnosis (Onida et al., 2013; Such et al., 2013).

Previous clinical trials were mostly reserved for high-risk MDS patients and limited numbers of CMML patients were included. Considering that CMML has established its own entity in the diagnostic guidelines, a more specific approach in managing CMML progression should be emphasized. One approach is to target the hypersensitivity to granulocyte-macrophage colony-stimulating factor (GM-CSF) with anti-GM-CSF monoclonal antibodies (lenzilumab) (Padron et al., 2013; Patnaik et al., 2019). Inhibition of the Ras signalling pathway, which contributes to the hypersensitivity, could become the new targeted therapy. Ruxolitinib, a JAK1/2 inhibitor commonly used for MPN has shown to relieve symptom burden in 7 patients (35%) in a Phase I clinical trial (Padron et al., 2016). Tipifarnib, a farnesyl transferase inhibitor, is currently being tested in clinical trials with outcomes pending (Patnaik et al., 2017). Besides RAS family inhibitors, novel agents, such as luspatercept and sotatercept could also be added into treatment regiments (Elmariah and DeZern, 2019). These agents trigger the haemoglobin response in MDS and beta thalassemia

patients, which could also alleviate anaemia and reduce red-cell transfusion dependencies in CMML patients (Carrancio et al., 2014; Fenaux et al., 2020; Komrokji et al., 2018; Platzbecker et al., 2016).

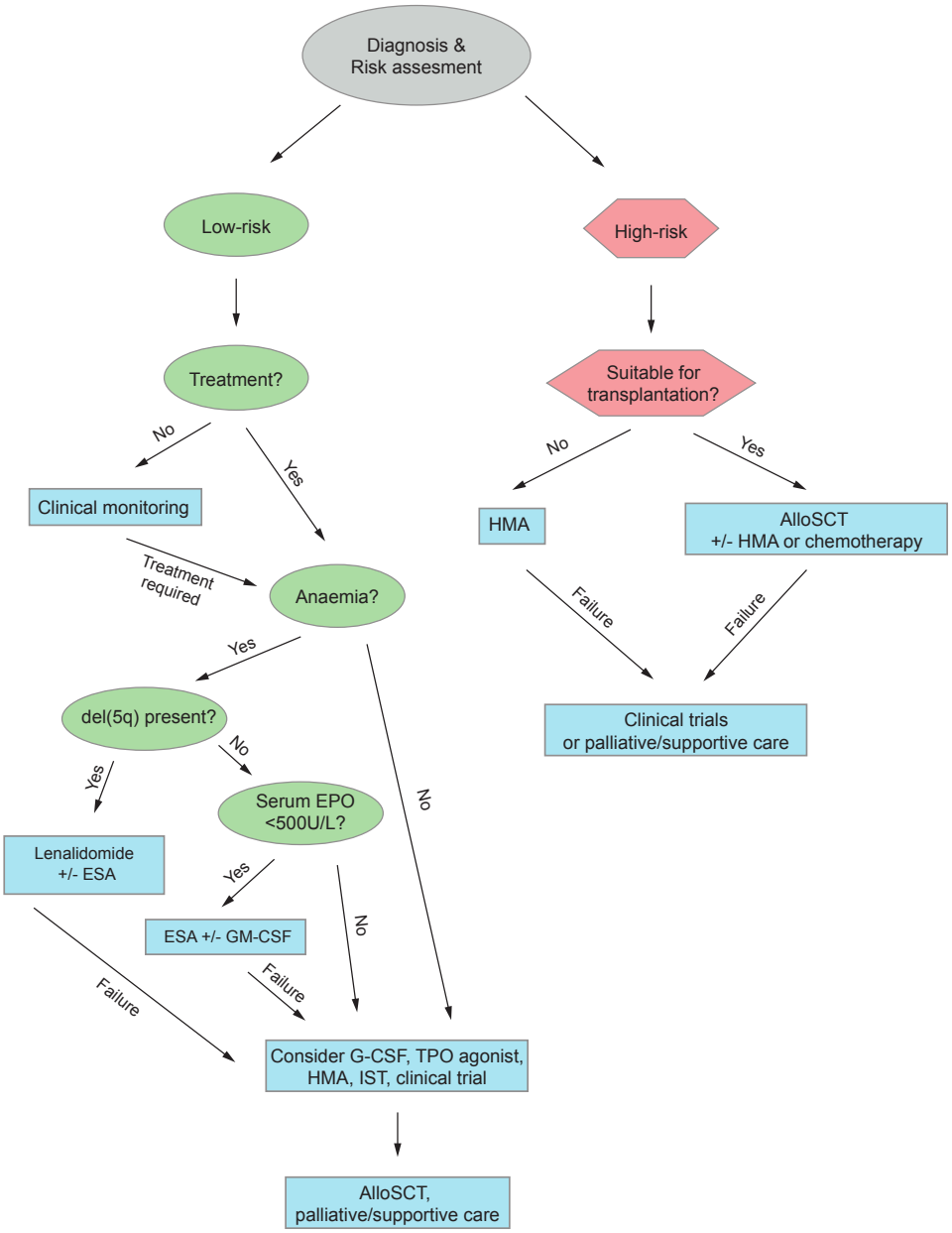


Figure 1.3: Current treatment roadmap for MDS (adapted from (Bejar and Steensma, 2014))
While low-risk patients have many potential therapies to choose from, only hypomethylating agents (HMA) and allogenic stem cell trans- plant (AlloSCT) are avaliable to high-risk patients. Allogenic stem cell transplant remains the only curative approach. ESA, erythropoiesis stimulating agent; EPO, erythopoietin; G-CSF, granulocyte colony-stimulating factor; TPO, thrombopoietin; IST, immunosuppressive therapy.

1.4 Molecular abnormalities in MDS and CMML

In the last decade, the description of the mutational pattern of both MDS and CMML has opened up new opportunities to understand these cancers. In agreement with the heterogeneous nature of MDS, sequencing studies have found mutations across several functional groups, such as histone modification, DNA methylation, RNA splicing, receptor kinases, transcription, DNA repair and RAS signalling (Haferlach et al., 2014). While some pathways overlap with known mutational pathways in AML, DNA methylation and RNA splicing were newly identified in MDS, demonstrating a specific association with MDS and CMML pathogenesis. Coincidentally, DNA methylation and RNA splicing mutations are often found in CHIP, as mentioned above, which suggests that CHIP could be a precursor state of MDS. However, this hypothesis would need further evaluation. Understanding the molecular abnormalities in MDS and CMML could pave the way to defining new disease entities, allowing for risk stratification and prediction of treatment outcome. Here, I present a summary of the latest research on the genetic landscape and clonal evolution of MDS/CMML.

1.4.1 Genetic landscape and clonal evolution of MDS

MDS is a heterogeneous disease, so is its genetic landscape. In one study, using targeted sequencing, nearly 90% of MDS patients were found to harbour at least one mutation with a median number of three mutations (Haferlach et al., 2014; Palomo et al., 2020). Forty-seven significantly mutated genes were then grouped based on functional pathways, disease subtypes and the correlation of gene mutations. Among the pathways, RNA splicing was the most frequently mutated (64%), followed by DNA methylation, chromatin modification, transcription and DNA

repair (Ogawa, 2019). Common recurrent mutations were *AXSL1, EZH2, DNMT3A, TET2, IDH1/2, SRSF2, SF3B1, CBL, RUNX1, RAS* and *TP53* (Haferlach et al., 2014; Sperling et al., 2017; Walter et al., 2013) (Figure 1.4A). Most genes were associated with high-risk MDS subtypes, such as MDS with excess blasts. Only *SF3B1* mutations had a favourable outcome and were associated with specific subtypes of MDS with ring siderob-lasts and MDS with isolated del(5q). In contrast, *SRSF2* mutation was commonly associated with shorter survival and advanced forms of MDS and CMML (Makishima et al., 2012). The significant correlation among different combinations of gene mutations suggests they may functionally interact in MDS pathogenesis. Some combinations are enriched, such as *TET2* with *SRSF2* and *ZRSR2* mutations, while some are mutually exclusive, such as *SRSF2* with *DNMT3A* and *EZH2* (Ochi et al., 2020).

The heterogeneity of MDS means that mutations could arise together in the same clone or in several separate clones. Understanding the clonal structure and relationship of genetic mutations is crucial for deciphering MDS pathogenesis and progression. MDS originates when a founding clone with somatically acquired mutations is positively selected because the mutations confer a growth advantage, driving further transformation. These driver mutations will be passed on to the descendent clones, along with 'passenger mutations' that are harmless but can serve as markers to determine clonal subpopulations. To study the clonal evolution of MDS, variant allele frequency (VAF) is calculated from whole exome sequencing or targeted sequencing data of bone marrow samples from MDS patients. The higher the VAF, the earlier the mutation appeared in the disease. Amongst the most frequently mutated genes, the DNA methylation and RNA splicing pathway mutations are thought to be founder events in MDS. Mutations in transcription factors, Ras signalling,

and DNA repair pathways appear to arise secondarily and often drive selection of dominant clones and progression to secondary leukaemia (Haferlach et al., 2014).

The process of clonal evolution is very dynamic. Analysis of longitudinally collected MDS samples revealed that there are two distinct evolution patterns (Makishima et al., 2012; Menssen and Walter, 2020). In the first, several subclones evolved in a linear fashion, until one dominant clone takes over. In the second, a new or pre-existing subclone becomes dominant and replaces an older subclone, and eventually one clone repopulates almost the entire tumour population. In both scenarios, new clones are constantly emerging but the competition between different subclones also intensifies. This is supported by observations that the variance of VAFs is high in the early stages of MDS but decreases with disease progression. Although the number of clones decreases, the number, diversity and size of the clonal mutations increases overtime (Chen et al., 2019; Makishima et al., 2012).

From a therapeutic point of view, characterising the clonal population not only can help to tailor the treatment plan for each individual patient, but also can identify therapy-resistant clones early on. Allogenic stem cell transplant or chemotherapy eradicates the majority of disease clones. However, the remaining clones evolve and or have subclonal selection leading to resistance to therapies (Ferrando and López-Otín, 2017). Recently, Chen et.al, reported a non-linear, parallel evolution pattern of MDS stem cell clones contributing to either MDS blast cell generation or progression to AML. Using single-cell sequencing of sorted cell populations from longitudinal samples of patients with MDS, Chen et.al. was able to identify small subclones that would be undetectable due to the technical limitations of bulk sequencing approaches (Chen et al.,

2019). More studies of MDS clonal dynamics are warranted in the era of single-cell sequencing.

1.4.2 Genetic landscape and clonal evolution of CMML

MDS and CMML share a similar genetic landscape with genes frequently mutated including *TET2* (60%) and *SRSF2* (50%) (Loghavi and Khoury, 2018; Palomo et al., 2020). Other common mutations are *ASXL1, RUNX1, CBL, U2AF1, KRAS, DNMT3A, EZH2* and *NRAS* (Patel et al., 2017). In total, mutations in 672 genes were detected in a cohort of 150 patients by whole exome sequencing (WES). Similarly, 90% of patients have at least one mutation with a medium number of three mutations per case, as determined by targeted sequencing (Palomo et al., 2020). The average number of mutations is 19 by WES. CMML-2, the advanced form, has a higher mutational burden than CMML-1 (Patel et al., 2017) (Figure 1.4B).

In terms of the pattern of clonal evolution, sequencing studies of CMML patients suggest an early clonal dominance at the stem cell stage, followed by a mostly linear progression and acquisition of secondary mutations in the same clone. Clonal heterogeneity is less common compared to MDS and most likely due to loss of heterozygosity (LOH) (Itzykson et al., 2013). When examining the order of mutation acquisition, *TET2* and *ASXL1* are the most common founder mutations. *SRSF2* mutation has an equal chance of presenting as founder or secondary hit. Although CMML shares a similar molecular profile with MDS, certain mutations are more common and associated with the monocytic phenotype in CMML, such as *TET2* and *SRSF2*, indicating a functional co-operation (Itzykson et al., 2013; Patel et al., 2017,Palomo, 2020).

1.4.3 Genetic co-operativity in MDS and CMML

Given the complexity of cancer mutation networks, it is unlikely that MDS or CMML pathogenesis are driven by a single genetic event.This multi-genic model of MDS and CMML pathogenesis has been confirmed by sequencing studies of both diseases (Haferlach et al., 2014). Analogous to the 'two hits' theory for acute leukaemia where a first hit stimulates cell proliferation and a second hit blocks haematopoietic differentiation, MDS and CMML are the result of multiple genetic alterations (Gilliland and Griffin, 2002; Menssen and Walter, 2020; Palomo et al., 2020). Even though MDS and CMML are both generally diseases of the elderly, genetic mutations responsible for the pathology may have been acquired decades prior to the patients being diagnosed (McKerrell et al., 2015). With the first somatic mutation being a randomly generated event, the second and third mutations are more likely to be a result of preferential selection. When comparing different pathway mutations in MDS, there is a correlation between certain functional pathways, such as the Ras pathway with splicing or receptor/kinase pathways with splicing (Haferlach et al., 2014) (Figure 1.4C-D). A specific combination of mutations in two pathways confers greater selective advantages over other possible mutational combinations. Makishima et al. and other groups demonstrated a dynamic clonal evolution of MDS to secondary AML (sAML) (Bejar, 2018; Makishima et al., 2017; Menssen and Walter, 2020). While many clones share the same set of mutations at the beginning, there is a distinct enrichment pattern for mutations in MDS phase (*TET2, ZRSR2* and *ASXL1*, enriched in high-risk MDS vs. low-risk MDS) or sAML phase (*FLT3, NRAS* and *IDH1*, enriched in sAML vs. high-risk MDS). Thus, understanding why certain mutations are enriched together or certain genetic combinations emerge as disease progresses is crucial for diagnosis, outcome prediction and identifying new treatment options.

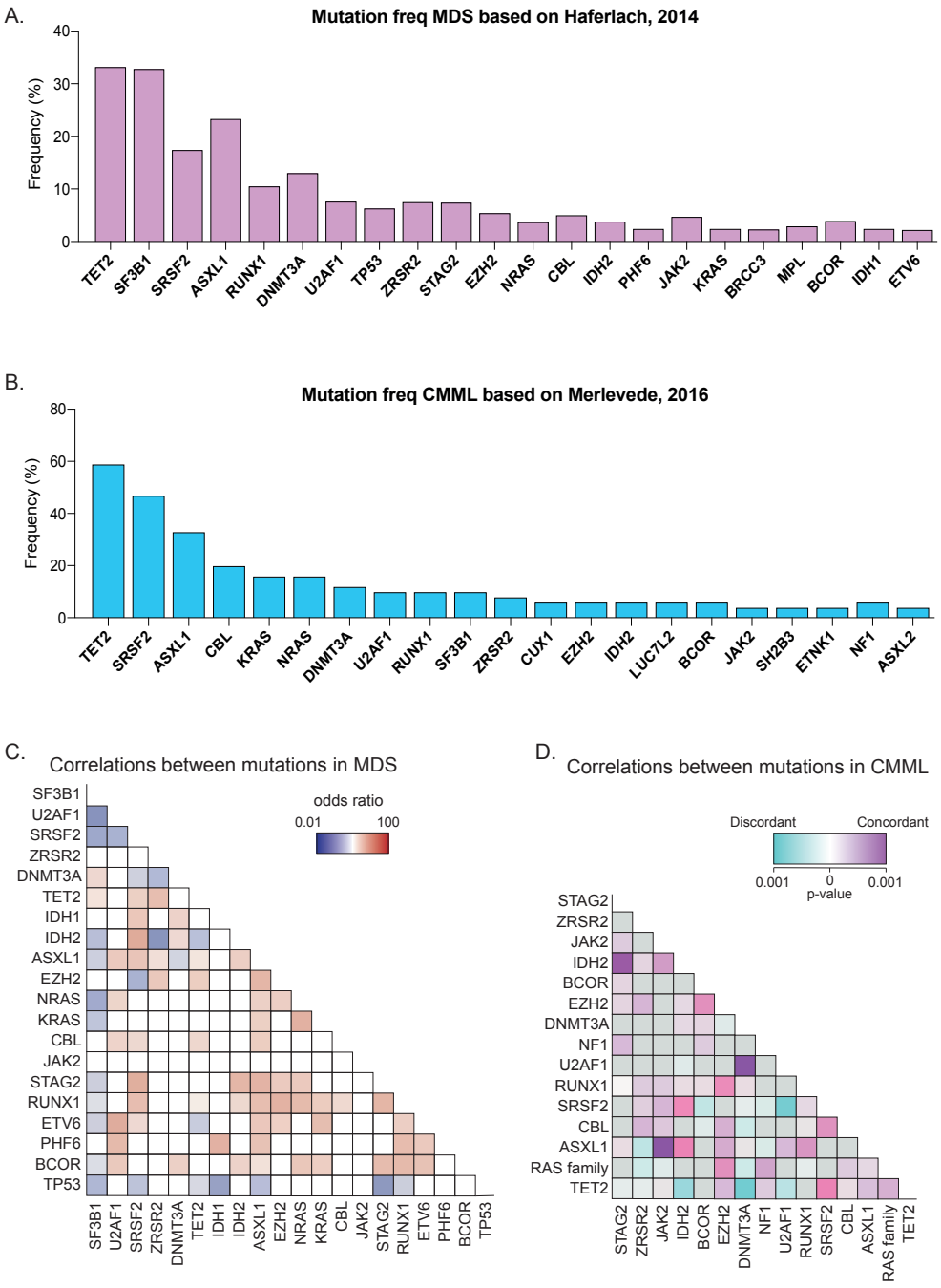


Figure 1.4: Frequency of recurrent mutations and their correlations in MDS and CMML

A. The frequency of recurrent mutations in MDS.
B. The frequency of recurrent mutations in CMML.
C. The correlations of MDS mutations illustrated in a staircase plot.
D. The correlations of CMML mutations illustrated in a staircase plot.
Mutation data from (Haferlach et al., 2014) for MDS and (Merlevede et al., 2016) for CMML. Correlation data from (Ogawa, 2019) for MDS and (Patel et al., 2017) for CMML.

1.5 RNA splicing and cancer

1.5.1 RNA splicing and the spliceosome

RNA splicing is an essential process that occurs in nearly all cellular transcripts. Splicing is carried out by the spliceosome, a complex molecular machine, which removes non-coding introns and joins the 5′ and 3′ splice sites together in the pre-messenger RNA (pre-mRNA) to form a multi-exonic mature mRNA. The "spliced" mRNA is then exported into the cytoplasm where it can be translated into protein. The first step of splicing involves the recognition of exon-intron boundaries by the U1 and U2 major spliceosome subunits. This is followed by assembly and disassembly of U2, U4 and U5 subunits, which catalyses the removal of introns and joining of two exons (Matera and Wang, 2014). Generally, pre-mRNA undergoes constitutive splicing, as well as alternative splicing which creates different mRNA isoforms, thus potentially multiple proteins, from the same gene. RNA splicing is a key step in increasing the diversity of the transcriptome, and subsequent proteome, that can be generated from a fixed number of genes within the genome. Alternative splicing occurs under normal physiological conditions as well as in diseases (Kelemen et al., 2013).

1.5.2 Regulation of RNA splicing

Splicing is tightly controlled and regulated to ensure its high-fidelity. The regulatory process mainly involves the cis-acting RNA elements and trans-acting factors. The cis-acting RNA elements include splicing signals, such as the 5′ and 3′ splicing sites and splicing regulatory elements (SREs) (Fu and Ares, 2014). SREs are short RNA motifs that regulate splicing site selection and include exonic splicing enhancers (ESEs), exonic splicing

silencers (ESSs), intronic splicing enhancers (ISEs) and intronic splicing silencers (ISSs). These motifs serve as the binding sites of trans-acting splicing factors (Figure 1.5A). Splicing signals serve as the navigator of the spliceosome machinery that enables splicing to take place at the correct location, which involves the definition of exon-intron junctions. There are many 'decoy' splice sites which contain sequences similar to the authentic splicing site in the intron. Despite the 'decoy' sites, the exon and intron definition complex - the machinery that marks the boundary of exons or introns - recognises the splicing site correctly. This suggests that additional structural information might be involved during the process (Matera and Wang, 2014).

The trans-acting splicing factors are RNA binding proteins that regulate splicing by interacting with cis-acting regulatory elements in splice sites during the early and intermediate steps of spliceosome assembly (Fu and Ares, 2014). The trans-acting splicing factors bind to short RNA sequences that lie within close proximity to splice sites in both exons and introns. One well-studied example is protein polypyrimidine tract binding (PTB), where a splicing factor binds to a short polypyrimidine track and inhibits splicing (Sharma et al., 2008). These enhancers or silencers are bound by trans-acting RNA binding proteins, such as serine/arginine-rich proteins (SR proteins) and heterogeneous nuclear ribonucleoproteins (hnRNPs). SR proteins often promote splicing while hnRNPs repress splicing. However, the exact mode of action is highly context-dependent and tightly controlled by complex regulatory mechanisms (Dvinge et al., 2016). Other factors that have been shown to influence splicing include RNA structure, transcription initiation and elongation, and chromatin structure (Matera and Wang, 2014).

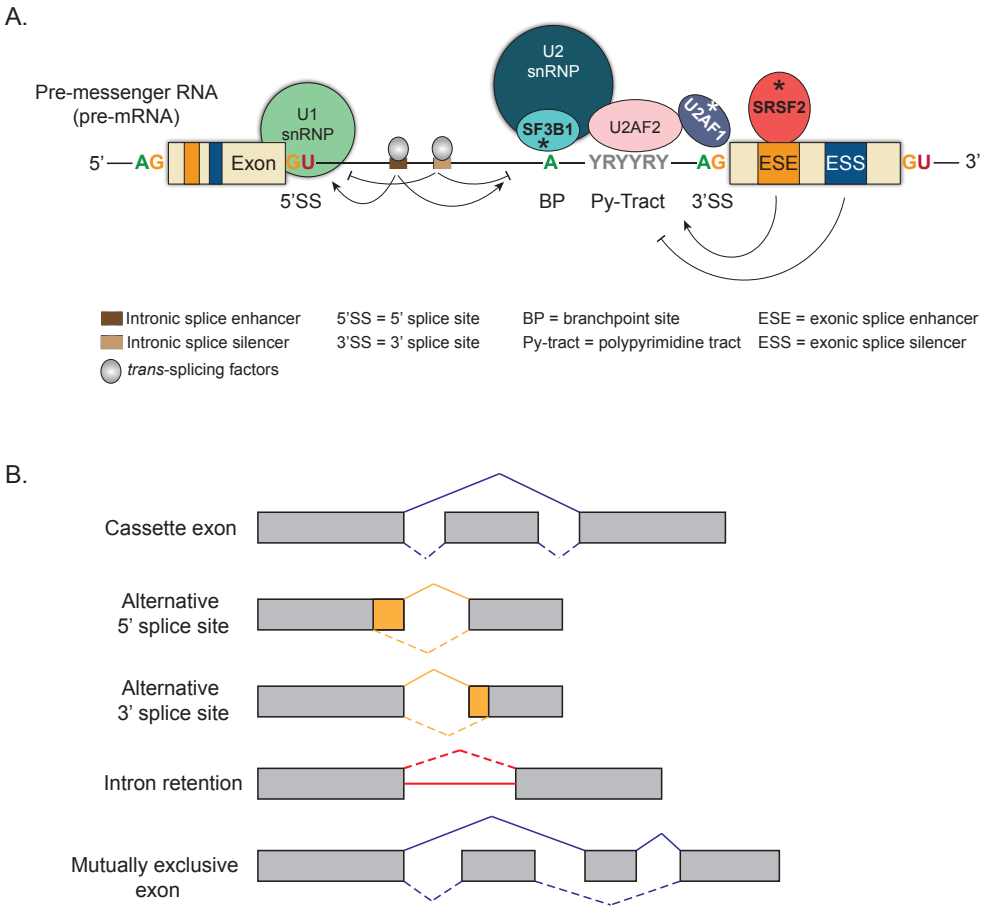


Figure 1.5: Schematic of RNA splicing and aberrant splicing events
A. Within the intron of the pre-mRNA, the U2 snRNP complex containing SF3B1 binds the branch-point site adenosine, U2AF2 binds to the polypyrimidine tract sequence, and U2AF1 binds the 3′ splice site; SRSF2 binds to the exonic splice enhancer sequence within the exonic sequence. Cis- and trans-acting factors can both positively or negative regulate splicing. Proteins which are frequently mutated are marked with an asterisk.
B. The different types of aberrant splicing events are illustrated. The solid lines represent aberrant splicing and dashed lines represent normal splicing events.

1.5.3 Aberrant alternative RNA splicing in cancers

Alternative splicing enables proteomic diversity from a fixed genome. Mis-regulation of alternative splicing enables cancer cells to create tumour-associated splice variants in almost every aspect of cancer biology, including apoptosis, cell-cycle control, differentiation, proliferation, invasion, and angiogenesis (Anczuków et al., 2012; Xu and Lee, 2003; Zhang and Manley, 2013). So far, more than 15,000 tumour-specific splice variants have been discovered across a wide range of cancers, with each type of cancer harbouring its own unique alternative splicing events (He et al., 2009; Shapiro et al., 2011; Venables et al., 2008). The types of aberrant alternative splicing are largely similar to those of normal differentiated cells and include cassette exons, alternative 5′ splice sites, alternative 3′ splice sites, intron retention and mutually exclusive exons (Figure 1.5B). Studies have shown that misregulation of alternative splicing is linked with altered levels of splicing factors (David and Manley, 2010). Overexpression of SR proteins, such as SRSF1, SRSF3 and SRSF6 are reported in lung, colon and breast cancer (Anczuków et al., 2012; David and Manley, 2010; Karni et al., 2007). The hnRNPs family and other splicing factors have also been implicated in both oncogenic and tumour suppressor roles. However, there is no evidence suggesting that altered levels of splicing factors alone can drive or maintain tumorigenesis. About a decade ago, the first direct link between RNA splicing and cancer was established with the discovery of splicing pathway mutations in haematological malignancies (Wang et al., 2011; Yoshida et al., 2011). The most frequent RNA splicing mutations in MDS include *SF3B1*, *SRSF2*, *U2AF1* and *ZRSR2* (Graubert et al., 2011; Yoshida et al., 2011) (Table 1.4 and Figure 1.5). Each splicing mutant is linked to a distinct subtype of MDS or MDS/MPN and mutations are mutually exclusive, although in rare cases two concomitant splicing mutations have been observed in the same cell (Taylor

et al., 2020). In myeloid leukaemia, SF3B1 mutations are predominantly (>90%) found in MDS with ring sideroblasts (MDS-RS) and have a good prognosis compared to other MDS subtypes (Papaemmanuil et al., 2011). When mutated, SF3B1 K700E leads to aberrant use of the 3′ splicing site due to mis-recognition of the branchpoint sequence (Carrocci et al., 2017; Darman et al., 2015; DeBoever et al., 2015). For U2AF1, hotspot mutation S34 leads to promotion of exon inclusion at the 3′ splicing site of the transcript (Brooks et al., 2014; Ilagan et al., 2015). Previously, I reviewed the phenotypes of the *in vivo* mouse models of splicing mutation, including *Sf3b1, U2af1* and *Srsf2* (Xu et al., 2019). I will summarise the current knowledge of SRSF2 and its function in normal and malignant haematopoiesis in Section 1.6.

Table 1.4: Frequency of RNA splicing mutations in haematological malignancies

Gene	Frequency (%)	Reference
SF3B1	MDS:32.9%	(Haferlach et al., 2014)
	MDS with ring sideroblasts: >90%	(Papaemmanuil et al., 2011)
	MDS/MPN:19.3%	(Malcovati et al., 2011)
	CLL: 9.7%-15%	(Quesada et al., 2012)
		(Wang et al., 2011)
SRSF2	MDS:17.6%	(Haferlach et al., 2014)
	CMML:50%	(Merlevede et al., 2016)
U2AF1	MDS:7.7%	(Haferlach et al., 2014)
ZRSR2	MDS: 7.9%	(Haferlach et al., 2014)

CLL, Chronic lymphocytic leukaemia

1.6 Serine-arginine rich splicing factor 2 (SRSF2)

Serine-arginine rich splicing factor 2 (SRSF2), previously known as SC35, is a member of the SR protein family that regulates constitutive and alternative splicing. The SRSF2 protein has two domains, the RNA binding domain and an arginine- and serine-rich (RS) domain. The RNA recognition motif binds to the ESEs elements in the exon, which are located close to the upstream 3' splicing site. After binding, SRSF2 interacts with both the U2 major spliceosome complex at the 3' splicing site and the U1 complex at the downstream 5' splicing site to promote exon inclusion. Structural studies have identified that CCNG or GGNG are the highly conserved consensus binding sequences for SRSF2 (Daubner et al., 2012).

While the role of SRSF2 in constitutive splicing is well-studied, its role in alternative splicing is yet to be fully understood (Fu, 1995). Early studies used artificial mini-gene constructs in transfected cells and found that alternative splicing might be regulated in a concentration-dependent manner (Fu et al., 1992; Krainer et al., 1990). However, there was limited understanding about the regulation of alternative splicing and its biological effect *in vivo*. Wang et al., first reported the role of SRSF2 in T cell development and alternative splicing of *Cd45* in a conditional knock-out murine model (Wang et al., 2001). Later Xiao et al., demonstrated that SRSF2 plays critical roles in cell proliferation and genome stability (Xiao et al., 2007). SRSF2 is also involved in other splicing-independent functions, such as transcriptional elongation and cell cycle control (Lin et al., 2008; Xiao et al., 2007).

1.6.1 SRSF2 function in haematopoiesis

The function of SRSF2 in haematopoiesis had not been extensively investigated except for its role in T cell development as mentioned earlier and it didn't come to light until Yoshida et al., revealed that SRSF2 mutations are frequently found in haematological malignancies such as MDS and CMML (Yoshida et al., 2011). To understand the role SRSF2 plays in haematopoiesis, Komeno et al., generated an *Srsf2* deficient mouse model (*Srsf2* $^{fl/fl}$ mice) using the constitutively active *Vav*-Cre, which deleted *Srsf2* in all haematopoietic tissues, and also the inducible *Mx1*-Cre, which deletes *Srsf2* somatically in many tissues upon poly(I);poly(C) injection. The study showed that deletion of *Srsf2* leads to pancytopenia and bone marrow failure in developing embryos, with *Srsf2* deficient embryos dying between E16.5 and E18.5, and when somatically deleted in adult mice (Komeno et al., 2015). The same phenotype was reproduced by Kim et al. (Kim et al., 2015). Whilst these studies have clearly defined an important role for SRSF2 in maintaining normal haematopoiesis, patient mutations are rarely null mutations.

1.6.2 *SRSF2* mutation leads to a change in RNA binding affinity

In MDS and CMML the mutations of *SRSF2* are heterozygous and the mutant protein has an altered function compared to normal SRSF2 protein. Homozygous mutations or compound heterozygosity with a mutant allele and a null allele have not been identified, suggesting that retention of a wild-type copy is necessary for cells to survive (Lee and Abdel-Wahab, 2016). Mutations in SRSF2 occur almost exclusively at proline 95 (p.Pro95), with the most frequent being a missense proline to histidine substitution (p.Pro95His; P95H) (Yoshida et al., 2011). P95 is located at a hinge region

between the RNA recognition motif (RRM) and arginine/serine-rich domain (RS) of SRSF2 protein (Figure 1.6A). Although P95H is not present in either RRM or RS, *in silico* analysis has predicted that it leads to a confirmation change in the SRSF2 protein that impacts the binding affinity of the RRM domain (Meggendorfer et al., 2012). Zhang et al., later demonstrated that P95H mutation leads to altered RNA binding affinities. Mutation of P95 to H95 changed the binding preference from an equal binding affinity to CCNG and GGNG motifs to an increased binding affinity for the CCNG motif and decreased binding to GGNG (Zhang et al., 2015) (Figure 1.6B). However, the altered binding affinity can additionally depend on the sequence context. The P95H mutation does not appear to affect the protein-protein interactions of SRSF2 with other spliceosome complexes, but as SRSF2 is a splicing regulator, any functional change could affect 3' splice-site recognition and result in aberrant splicing. Some of the proposed mis-spliced target transcripts include, *EZH2, BCOR, IKAROS, CASP8* and *FYN* (Armstrong et al., 2018; Kim et al., 2015; Zhang et al., 2015). The mis-splicing of SRSF2 P95H is discussed further in Section 1.6.3 and Chapter 4 of this thesis.

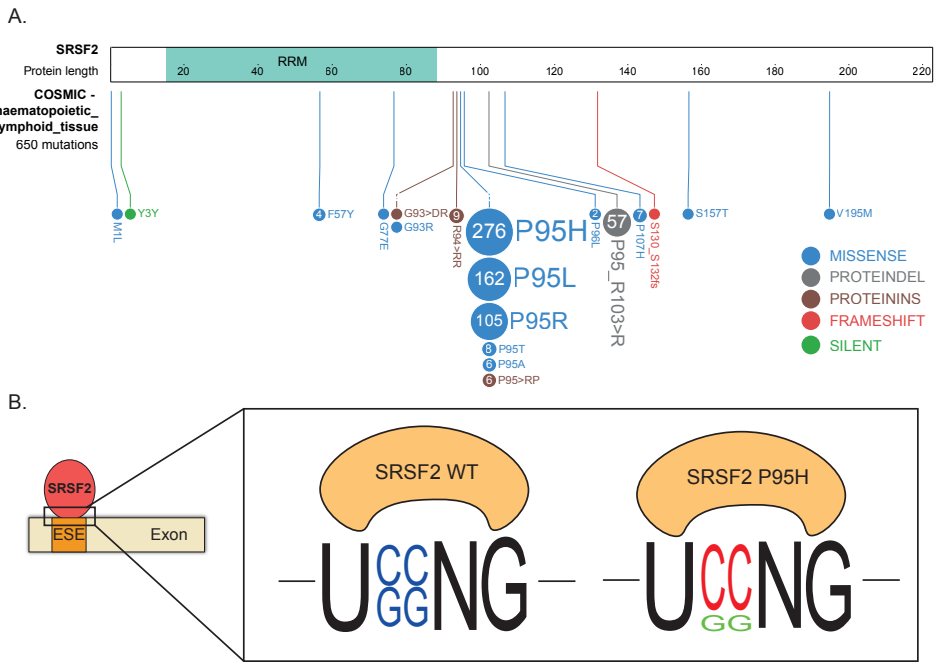


Figure 1.6: SRSF2 mutation frequently occurs at P95 and leads to a change in RNA binding site recognition
A. Mutations in SRSF2 mostly occur at the Proline (P) 95 site and lead to a substitution with Histidine (H), Leucine (L) or Arginine (R). The Proline to Histine change is the most frequent. RRM, RNA recognition motif. SRSF2 protein and mutation profile are produced using ProteinPaint.
B. SRSF2P95H protein has an altered binding affinity for the RNA sequence on the exonic splice enhancer (ESE). While WT SRSF2 recognises CCNG and GGNG equally well, SRSF2P95H has an increased binding affinity for CCNG (red color) and decreased affinity for GGNG (green color). N represents either A, U, C or G.

1.6.3 Mechanism of *SRSF2* P95H mutation in MDS pathogenesis: poison exon?

To determine how *Srsf2*^{P95H} induces MDS, Kim et al. analysed the binding preference of the RNA binding motif and found a similar change in binding affinity as was shown by Meggendorfer et al. and Zhang et al. (Meggendorfer et al., 2012; Zhang et al., 2015). The change in binding preference led to alternative splicing, with exon skipping being the most common event. Instead of solely a direct effect, alternative splicing is also likely to be an indirect consequence of *SRSF2* P95H mutation. Liang et al. demonstrated that most altered binding sites of SRSF2P95H are not located within the alternatively spliced exons, but in genes enriched in RNA processing and splicing. Alternative splicing is thus a result of a 'splicing cascade', leading to modification of the RNA splicing and processing network, set off by *SRSF2* P95H mutation (Liang et al., 2018). To identify common changes in mis-splicing events, Kim et al. compared sequencing data from purified murine lineage$^-$c-Kit$^+$Sca-1$^+$ (LKS$^+$) and myeloid progenitor (MP, lineage$^-$c-Kit$^+$Sca-1$^-$) cells, as well as CMML and AML samples with *SRSF2* mutation. They found that *EZH2*, a key regulator of haematopoiesis, showed altered splicing, with *SRSF2*^{P95H} mutation promoting the inclusion of a 'poison' exon (incorporation of a premature stop codon) of *EZH2*. *EZH2* is a gene with frequent loss-of-function mutations in MDS. The alternative exon usage was proposed to result in the degradation of *EZH2* mRNA via non-sense mediated decay (NMD), thus less functional EZH2 protein. However, the conclusion that *EZH2* is a critical mis-spliced target of SRSF2 P95H was challenged by Zhang et al. using human K562 *SRSF2*^{P95H} knock-in cells (Zhang et al., 2015) and also by Kon et al. in a murine *Srsf2*^{P95H} model (Kon et al., 2018),

where neither group was able to reproduce the original *EZH2/Ezh2* mis-splicing finding and, in contrast, found that the changes in *EZH2/Ezh2* splicing were not significant. Our group also examined the mis-splicing of *Ezh2* using a murine model. We found no change in transcript isoforms in total splenocytes from *Srsf2*$^{P95H/+}$ mice, nor in two other reported targets (*Bcor* or *Dot1l*). However, as a positive control, we could detect a change in exon inclusion of the *Cd45* transcript and the resultant CD45 protein isoforms in *Srsf2*$^{P95H/+}$ cells (Smeets et al., 2018).

Although few other groups also detected similar poison exon inclusion in *EZH2* in their studies, it is questionable whether a single transcript can account for the mechanism of SRSF2 P95H induced tumorigenesis (Bechara et al., 2013; Pellagatti et al., 2018). So far, none of the individual transcript targets discovered have been proven to be crucial pathogenic mis-splicing events caused by *SRSF2* P95H mutation. This is supported by an recent paper analyzing RNA-seq from > 1,200 patient samples (Hershberger et al., 2020). Thus, the question remains which transcript or set of transcripts are the pathogenic targets of splicing mutations. An alternative hypothesis suggests that MDS is more likely to be the result of multiple, potentially thousands of mis-splicing events, a phenomenon called 'Spliceosome Sickness'. The contribution of each individual mis-splicing event is insignificant when compared to the combined effect of all together (Darman et al., 2015). This is supported by evidence that *SRSF2* mutation induces a broad, but subtle, transcriptional and splicing change (Kim et al., 2015; Seiler et al., 2018). However, it is a technical challenge to confirm this hypothesis since the 'mini' mis-splicing events often fall below the cut-off range of commonly employed bioinformatics methods (Rahman et al., 2020; Zhang et al., 2015). More studies are needed to determine whether a 'specific mis-splicing target' model, a 'death by a

thousand cuts' model or maybe a combination of multiple mechanisms underlie the effects of spliceosome mutations.

1.6.4 Targeted therapy for RNA splicing mutated MDS/CMML

Spliceosome mutations are heterozygous and exhibit a mutually exclusive pattern between spliceosome mutations/disease subtype, which suggests a synergistic lethality or functional redundancy between the spliceosome compartments. The fact that spliceosome mutations exist in a heterozygous pattern also suggests that the remaining wild-type allele is crucial to sustain the essential splicing catalysis. Thus, cells carrying spliceosome mutations could be more susceptible to further spliceosome damage as there is only one copy of the wild-type allele compared to two copies in normal cells (Bonnal et al., 2012; Lee and Abdel-Wahab, 2016).

With this potential therapeutic targeting strategy in mind, the next step is to determine how to target the spliceosome and which drugs are available. Since splicing is a multi-step process, there are many choices for modulation, for example, disrupting spliceosome assembly, interfering with splicing regulatory elements or targeting the mis-spliced protein (Kim and Abdel-Wahab, 2017; Lee and Abdel-Wahab, 2016). For targeting splicing factors directly, in the context of *SRSF2* P95H, *SF3B1* K700E and *U2AF1* S34Y mutations, two groups of natural anti-tumour compounds and their derivatives have been identified. The first group is the Pladienolides, with Pladienolide B being the naturally occurring compound and E7107 a synthetic derivative (Kotake et al., 2007). The second group are the Spliceostatins, with FR901463, FR901464 and FR901465 as natural compounds and Meayamycins, Spliceostatin A and Sudemycins as derivatives (Kaida et al., 2007). Both classes of compounds disrupt the

binding of SF3B1 within the U2 complex to the 3' splicing site of the pre-mRNA. This results in an accumulation of unspliced or mis-spliced products in the cells, which ultimately leads to cell cycle arrest in the G1 and G2/M phases (Bonnal et al., 2012; Kaida et al., 2007).

Several pre-clinical and clinical trials have evaluated the efficacy of Pladienolides and Spliceostatins in haematological cancers. Lee et al. firstly tested E7107 in two models: MLL-AF9 transformed mouse leukaemia and acute myeloid leukaemia (AML) patient-derived xenografts (PDXs) (Lee and Abdel-Wahab, 2016). In both models, leukaemic cells with $Srsf2^{P95H/+}$ (mouse) or $SRSF2^{P95H/+}$ (human) showed a modest preferential sensitivity towards E7107 treatment. There was an increase in apoptosis in treated mutant cells and a slower leukaemia progression compared to vehicle. E7107 was also tested against $Sf3b1^{K700E/+}$ mutation bearing cells. In this mouse model, Obeng et al. showed a 2-fold difference between $Sf3b1^{+/+}$ (IC50 = 1.249nM) and $Sf3b1^{K700E/+}$ cells (IC50 = 0.619nM) (Obeng et al., 2016). For Spliceostatins, Shirai et al. treated U2AF1(S34F) expressing K562 and OCI-AML3 cells with Sudemycin D6 and showed a 2-fold difference in sensitivity between mutant and wild-type cells (Shirai et al., 2015). Aside from pre-clinical testing, E7107 was studied in two separate phase I clinical trials on locally advanced or metastatic solid tumours (NCT00459823, NCT00499499). However, both trials were terminated due to optical toxicity (blindness) (Hong et al., 2014). The more recent study introduced H3B-8800, another splicing modulating molecule based on Pladienolide. Seiler et al. treated a PDX model of AML with $SF3B1^{K700E}$ mutation and a PDX model of CMML with $SRSF2^{P95H}$ mutation with H3B-8800. In both models, there was a decrease in leukaemia burden in the splicing-mutant group compared to the non-splicing mutant counterpart. H3B-8800 also demonstrated a

higher efficacy than E7107 (Seiler et al., 2018). This compound is yet to undergo clinical trials, to test its on-target effect on spliceosome inhibition without potential excessive toxicity in wild-type cells.

1.7 DNA methylation and TET2

1.7.1 DNA methylation in haematopoiesis

DNA methylation is an important and widespread epigenetic modification of the genome. DNA methyltransferase (DNMT) adds a methyl group (m) to cytosine (C) and forms 5-methylcytosine (5mC). In contrast, the process of demethylation is catalysed by Tet-eleven translocation (TET) enzymes, which demethylate 5mC into 5-hydroxymethylcytosine (5-hmC), 5-formylcytosine (5-fmC) and 5-carboxylcytosine (5-CmC), in an α-ketoglutarate and Fe^{2+} dependent manner (Solary et al., 2014; Tahiliani et al., 2009) (Figure 1.7). Methylation of the promotor regions inhibits transcription and subsequently down-regulates the expression of affected transcripts, while demethylation up-regulates gene transcription.

Both self-renewal and differentiation of HSCs are tightly associated with changes in DNA methylation patterns. The loss of *Dnmt1* in HSCs severely impairs self-renewal capacity, whereas loss of *Dnmt3a* or *Dnmt3b*, the two de novo methyltransferases, enhances the self-renewal capacity (Challen et al., 2011; Trowbridge et al., 2009). In terms of differentiation, loss of *Dnmt3a* leads to a differentiation blockage, which is exaggerated by dual loss of *Dnmt3a* and *Dnmt3b* (Challen et al., 2011). In contrast, *Dnmt1*$^{-/-}$ HSCs lose commitment to lymphoid lineages while myeloid and erythroid lineages remain unaffected (Broske et al., 2009; Tadokoro et al., 2007). The phenotypic differences could be attributed to the different roles of the methyltransferases. DNMT1 performs constitutive methylation to

maintain the methylation pattern during replication, whereas DNMT3A and DNMT3B establish new methylation patterns on unmodified DNA (Okano et al., 1999). The changes in HSC self-renewal vs differentiation behaviour are accompanied by upregulation of multipotency genes, such as *Gata3, Runx1, Pbx1* and *Cdkn1a*, and downregulation of HSC differentiation genes, such as *Flk2, Ikaros, Sfpi1* and *Mef2c* (Challen et al., 2011). However, there is no clear correlation between differential gene expression and methylation status in *Dnmt3a*-null HSCs, which indicates an indirect regulatory mechanism.

Nevertheless, changes in DNA methylation and gene expression levels can be well-correlated in myeloid vs lymphoid differentiation. Ji et al. demonstrated a correlation between lineage-specific gene expression and differentially methylated regions (DMRs) (Ji et al., 2010). DMRs include both CpG islands, regions with higher CpG nucleotide density, and CpG shores, regions within 2 kilobases of a CpG island. Although most DNA methylation occurs on CpG islands, more differential methylation occurs on the CpG shores than islands during cellular differentiation (Doi et al., 2009). Genes that play important roles in lineage-fate decision, like *Lck* and *Mpo*, are gradually upregulated while DMRs on the gene are progressively demethylated (Ji et al., 2010). Overall, there is a steady increase in methylation upon lymphoid commitment relative to myeloid, which corresponds to the myeloid skewed phenotypes in *Dnmt1*-hypomorphic mice (Broske et al., 2009).

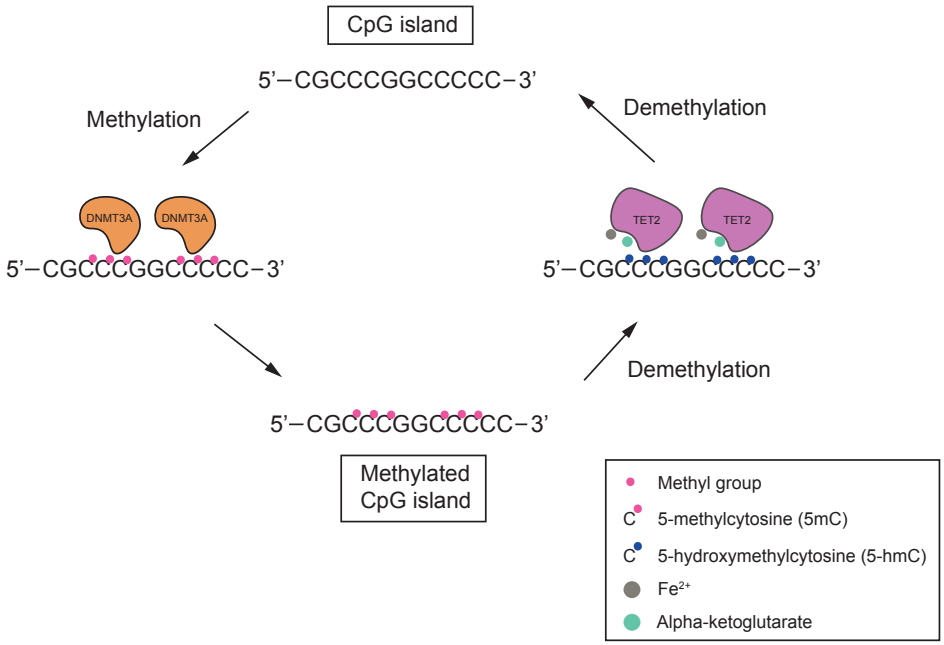


Figure 1.7: DNA methylation and demethylation by DNMT3A and TET2 DNA methylation and demethylation mostly occurs on the CpG islands of genome. Here in a simplified illustration, DNMT3A adds a methyl group to cytosine and forms 5-methylcytosine. TET2 demethylates by transforming 5-methylcytosine into 5-hydroxymethylcytosine, in an iron (Fe^{2+}) and alpha-ketoglutarate dependent manner. The 5hmC is then processed and eventually produces cytosine.

1.7.2 Deregulation of DNA methylation in haematological malignancies

Given the importance of DNA methylation in the regulation of normal haematopoiesis, it is no surprise that mutations in the DNA methylation pathway are frequently found in haematological malignancies. Recurrent mutations of *DNMT3A, TET2* and Isocitrate Dehydrogenase 1 and 2 (*IDH1/2*) have been implicated in myeloid malignancies, such as AML, MDS, MPN and MDS/MPN, as well as in lymphoid malignancies (Brunetti et al., 2017). I have summarized the relationship between mutations in the DNA methylation pathway and haematological malignancies in Table 1.5.

DNMT3A mutations were first reported in AML, followed by subsequent discoveries in MDS, MPN and MDS/MPN overlap syndromes (Ley et al., 2010; Shlush et al., 2014). Mutations are enriched at codon R882, which is predominantly a heterozygous mutation in AML and other myeloid neoplasms. The homozygous R882 mutation only occurs in T-cell acute lymphoblastic leukaemia (T-ALL), which suggest a gene-dosage effect on disease phenotype (Gaidzik et al., 2013; Gale et al., 2015). *DNMT3A* mutations are detectable in 20% of de novo AML, and frequently associated with *NPM1* and *FLT3* mutations. The prevalence of *DNMT3A* mutations in other myeloid neoplasms is significantly lower than in AML with 13.5% in MDS and 6.5% in MPN, respectively (Papaemmanuil et al., 2013; Stegelmann et al., 2011; Walter et al., 2011). Mechanistically, the R882 mutation has a dominant-negative effect on the homo-tetramerization of DNMT3A. This leads to formation of heterodimers instead of the catalytically active DNMT3A homo-tetramers, which reduces the methyltransferase activity.

IDH1/2 are metabolic enzymes that convert isocitrate to α-ketoglutarate (αKG) while producing NADPH from $NADP^+$ and CO_2 (Cairns and Mak, 2013). When mutated, IDH1 (R132) and IDH2 (R172 or R140) create neomorphic enzymes that convert αKG into a D-enantiomer of 2-hydroxyglutarate (2HG), which competitively inhibits the normal function of αKG. In cells, αKG acts as an important co-factor for enzymes of the TET family and for histone demethylases of the Jumonji class (Xu et al., 2011). Mutated *IDH1/IDH2* leads to a reduction of 5hmC levels, DNA hypermethylation and hypermethylation of histone markers, which phenocopies the effects of *TET2* mutation (Kernytsky et al., 2015). Although some phenotypes observed in *Idh1/Idh2* mutant mice are similar to *Tet2* knock-out mice, *Idh1/Idh2* mutation does not induce leukaemia (Kats et al., 2014; Sasaki et al., 2012). Co-operativity of *IDH1/2* mutation with other frequent mutations in AML is currently being investigated (Ogawara et al., 2015).

1.7.3 *TET2* mutation in haematological malignancies

TET2 is part of the TET family of dioxygenases, which contains two other members, TET1 and TET3. TET1 was discovered as a fusion protein in the mixed lineage leukaemia (MLL) gene translocation in AML, followed by identification of TET2 and TET3 by homology search (Lorsbach et al., 2003). TET enzymes exhibit distinct expression patterns in various tissues, with *TET1* transcripts expressed mostly in embryonic stem cells and *TET2* transcripts predominantly in haematological and neuronal tissues (Tahiliani et al., 2009). The catalytic domain of TET2 is located at the C-terminus, which is a double-stranded β-helix (DSBH) fold. The DSBH fold contains motifs that assist in the binding of Fe^{2+} and 2-oxoglutarate (also known as α-KG) (Pastor et al., 2013). TET-mediated demethylation is performed in both passive and active forms by the enzyme, and is

closely co-ordinated with DNA replication and base excision repair (BER) (Hajkova et al., 2010; Mayer et al., 2000).

Since 2009, *TET2* mutations have been reported in a range of haematological malignancies, such as AML, MDS, MPN, MDS/MPN and lymphoid malignancies (Abdel-Wahab et al., 2012; Delhommeau et al., 2009; Langemeijer et al., 2009; Quivoron et al., 2011) (summarised in Table 1.5). The frequency of *TET2* mutation ranges from 10 to 50%, with the highest rate observed in CMML. *TET2* mutations impact all domains of the TET2 protein and result in a partial or total loss of TET2 activity, and then a global reduction of 5hmC in TET2 mutant CMML patients (Ko et al., 2015). *TET2* mutation is an early genetic event and occurs before transformation, as alterations in TET2 can be found in the healthy elderly population as well as in individuals with CHIP (Ferrando and López-Otín, 2017).

1.7.4 Targeted epigenetic therapy for MDS/CMML

The plasticity of DNA methylation patterns and its frequent involvement in haematological malignancies paved the way for DNA hypomethylating agents (HMA). HMAs are thought to work in part by causing the re-expression of genes that were silenced during malignant transformation. 5-azacytidine (Azacitidine) and 5-aza-2-deoxycytidine (Decitabine) are the two FDA-approved HMAs used widely in the treatment of MDS, MDS/MPN and secondary acute myeloid leukaemia (sAML). Both agents act as DNMT1 inhibitors that stop the aberrant methylation pattern being passed on to daughter cells during DNA replication (Uy et al., 2017). In the cell, azacitidine is incorporated into RNA as well as DNA at a 2:1 ratio, whereas decitabine is only incorporated into DNA. The difference in pharmacokinetics could account for the distinct effects of the two agents on malignant cells. A study by Flotho et al. showed that cells treated with

Table 1.5: Frequency of DNA methylation mutations in adult haematological malignancies (Adapted from (Brunetti et al., 2017; Guillamot et al., 2016))

Gene	Haematological malignancy	Frequency (%)	Reference
DNMT3A	AML	De novo AML 22.1% Intermediate risk AML 29.8% De novo AML, s-AML and t-AML: 20.9%	(Gaidzik et al., 2013; Gale et al., 2015; Ley et al., 2010)
	T-ALL	19.3%	(Grossmann et al., 2013)
	MPAL	T/myeloid AML 55.6%	(Wolach and Stone, 2015)
	MDS	13.5%	(Haferlach et al., 2014)
	MPN	6.5%	(Stegelmann et al., 2011)
	CMML	CMML-1: 3.1% CMML-2: 13.2%	(Roller et al., 2013)
	T-cell lymphoma	AITL and PTCL-NOS: 26.6% CTCL: 42.5%	(Choi et al., 2015; Sakata-Yanagimoto et al., 2014)
IDH1/2	AML	16-19%	(Mardis et al., 2009) (Marcucci et al., 2010)
	MPN	2-5%	
	MDS	3%	

	AITL	20-45%	(Odejide et al., 2014)
TET2	AML	7-23%	(Abdel-Wahab et al., 2009)
	MDS	10-20%	(Langemeijer et al., 2009)
	CMML	50%	(Delhommeau et al., 2009; Tefferi et al., 2009)
	DLBCL	6-12%	(Quivoron et al., 2011) (Asmar et al., 2013)
	PTCL	19-51%	(Palomero et al., 2014) (Sakata-Yanagimoto et al., 2014)
	AITL	Up to 76%	(Asmar et al., 2013; Cairns et al., 2012; Couronne et al., 2012; Odejide et al., 2014; Quivoron et al., 2011)

AML, acute myeloid leukemia; t-AML, therapy-related AML; s-AML, secondary AML; T-ALL, T-cell acute lymphoblastic leukemia; MPAL, mixed phenotype acute leukemia; MDS, myelodysplastic syndromes; MPN, myeloproliferative neoplasms; CMML, chronic myelomonocytic leukemia; AITL, angioimmunoblastic lymphoma; DLBCL, diffuse large B-cell lymphoma; PTCL, peripheral T-cell lymphoma; PTCL-NOS, peripheral T-cell lymphomas not otherwise specified; CTCL, cutaneous T-cell lymphoma.

azacitidine exhibit different gene expression profiles compared to those treated with decitabine (Flotho et al., 2009). However, both azacitidine and decitabine are used as frontline treatment for MDS and show no significant difference in clinical response (Roberts and Steensma, 2015).

Even though the HMAs are widely used therapies for MDS and MDS/MPN, the overall response rate and response duration is low and variable. A meta-analysis showed that only 12% of patients achieve complete remission (CR). In addition, the median response duration for CR patients is only about one year, and the median survival is less than 6 months once patients relapse (Xie et al., 2015). Some clinical factors, such as a younger age, favourable cytogenetics and lower bone marrow blasts are associated with higher response rates. In contrast, prior treatment with cytarabine, high bone marrow blast counts (>15%) and complex karyotypes are associated with poor response rates (Itzykson and Fenaux, 2011; Roberts and Steensma, 2015). In terms of molecular markers, *TET2* mutation has been linked with a higher response rate (Bejar and Steensma, 2014; Itzykson and Fenaux, 2011). The presence of other DNA methylation mutations, such as *DNMT3A* and *IDH1/2*, also predict a better response to HMAs (Roberts and Steensma, 2015; Traina et al., 2014). The mechanism behind HMA resistance is largely unknown. Recent studies of HMA responders vs non-responders indicate that HMA non-responders have more quiescent HSCs before treatment. Since active DNA replication is required for incorporation of azacytidine or decitabine into the DNA, this might explain the lack of response in these patients (Meldi and Figueroa, 2015; Unnikrishnan et al., 2017). Overall, these results highlight the fact that alternative therapeutic options are urgently needed to improve the survival rate of high-risk MDS and MDS/MPN patients.

1.8 CBL in haematopoiesis and leukaemogenesis

CBL (also known as c-CBL) is part of a protein family that is involved in negative regulation of activated receptor tyrosine kinases (RTKs). The main domains of the CBL proteins are a tyrosine kinase-binding domain (TKB), a linker region and RING finger domain, followed by a proline-rich region and the carboxyl-terminal domain. Activated RTK binds to CBL, which becomes phosphorylated, then recruits ubiquitin-conjugating enzymes (E2s) or ubiquitin-carrier enzymes (UBCs) and directs the degradation of RTKs. Besides regulating RTKs, CBL also positively regulates SRC-family kinases independent of its E3 ubiquitin activity (Thien and Langdon, 2001).

CBL is a proto-oncogene which drives tumorigenesis when mutated. In AML (1-3%), CMML (13-15%) and Juvenile myelomonocytic leukaemia (JMML, 11-17%) patients, *CBL* mutations mostly occur in the linker and RING finger domain, principally as missense mutations and deletions (Sanada et al., 2009; Klampfl et al., 2013; Niemeyer, 2018; Palomo et al., 2020) (Figure 1.8). *CBL* mutations are highly correlated with acquired uniparental disomy (aUPD), where cells acquire two copies of the same parental allele instead of one from each parental chromosome, leading to loss of heterozygosity. When *Cbl* was deleted in mice with AML, Sananda et al. found an acceleration of AML progression into the blast-crisis phase, indicating Cbl functions as a tumour suppressor *in vivo*. However, *CBL* mutations are not loss-of-function mutations as occur in *TET2*, but rather represent a gain-of function. When $Cbl^{-/-}$ cells were reconstituted with *CBL* mutations, the cells became hypersensitive to cytokines such as SCF, IL-3, TPO and FLT3. There was also an increased activation of

downstream signalling targets (STAT5, Akt), indicating *CBL* mutations confer a proliferative advantage (Sanada et al., 2009).

Although *Cbl* mutations were known to promote aberrant cell signalling, the exact effect of *Cbl* mutation on haematopoiesis was unclear. Rathinam et al. first characterised a Cbl RING finger mutant, *Cbl*(C379A), mouse model. As most homozygous RING finger mutant mice died *in utero*, mice were analysed as a single C379A (called A from here on) hemizygous allele ($Cbl^{A/-}$). Unlike control genotypes, $Cbl^{A/-}$ mice developed lethal myeloproliferative leukaemia (Rathinam et al., 2010). $Cbl^{A/-}$ haematopoietic stem cells showed increased proliferation and reconstitution ability. Mutant cells were hypersensitive to cytokine stimulation and had increased activation of Akt. These phenotypes were attenuated when Flt3 ligand was deleted, indicating FLT3 was one of the main targets of the CBL protein. Increasing clinical evidence indicates that *CBL* mutation is one of the key mutations in myeloid leukaemia. The latest WHO classification of myeloid leukaemia additionally recommends germline *CBL* mutation testing as one of the diagnostic criteria for JMML (Arber and Hasserjian, 2015).

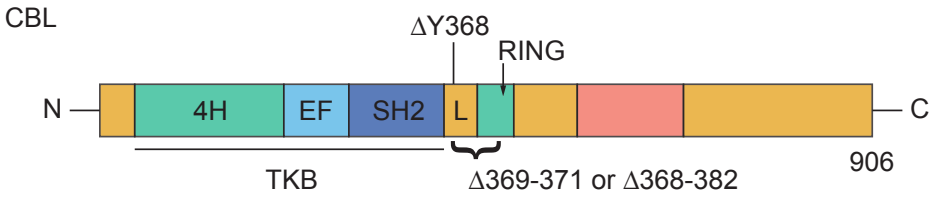


Figure 1.8: CBL mutations occur at the linker-RING finger domain
CBL mutations are mostly missense mutations, such as ΔY368, and occur at the linker-RING finger domain, which is central to its E3 ubiquitin ligase activity. Larger deletions can also occur in the linker-RING finger domain. TKB, tyrosine kinase binding domain; 4H, four-helix bundle; EF, Ca^{2+}-binding EF hand; SH2, Src homology region 2.

1.9 Mouse models of MDS and CMML

Mouse models have served as a valuable tool for understanding the pathogenesis of haematological malignancies. In MDS, mouse models were generated by targeting putative mutations that may be responsible for MDS, of which the number was limited before the discovery of recurrent mutations by Next-Generation sequencing (NGS) technology. The first model to be generated was a mouse model of del(5q), since del(5q) was one of the most common cytogenetic abnormalities in MDS with well characterised clinical manifestations. Barlow et al., first generated a 5q deletion model by large-scale deletion of the chromosomal region corresponding to the human 5q critical deleted region (CDR) (Barlow et al., 2010). However, a large-scale deletion could not pinpoint the role of each specific gene that could be responsible for the 5q- syndromes, as the CDR spans over 1.5 megabase and contains at least 40 genes that might account for the disease phenotype. Thus, new mouse models of 5q-syndromes focused on genes located on the 5q CDR that were identified through functional screens or frequently mutated in AML. Other MDS models were generated based on translocations found in MDS patients following transformation to AML (NUP98-HOXD13) or on genes that had been implicated in MDS pathogenesis through bone osteoprogenitor dysfunction, such as *Dicer1* and *Sbds* deletion (Pineault et al., 2003; Raaijmakers et al., 2010; Wegrzyn et al., 2011).

Another model that has been utilised is that of patient-derived xenografts of MDS bone marrow into immunodeficient mice, which enables researchers to study human MDS in a controllable *in vivo* system. However, this approach has had many obstacles mostly due to low engraftment/propagation of MDS-derived clones. So far, only a handful of groups have successfully developed xenograft models that can engraft

both low-risk and high-risk MDS samples as well as CMML samples and replicate patients' disease features (Rouault-Pierre et al., 2017; Song et al., 2019; Yoshimi et al., 2017).

1.10 Mouse models of *Srsf2* and *Tet2* mutation

1.10.1 Mouse models of *Srsf2* mutation

Kim et al., reported the first *Srsf2* P95H conditional knock-in model (Kim et al., 2015). The allele described approximated a gene-duplication strategy (Figure 1.9) (Table 1.6). Very little data was presented describing the phenotypic effects of this *Srsf2*$^{P95H/+}$ model on native haematopoiesis (i.e. not from transplant recipients). The authors state that none of the *Srsf2*$^{P95H/+}$ mice developed AML out to 70 weeks of monitoring, although it is not certain if this refers to *de novo* mutated animals or transplant recipients.

The authors principally reported analysis from non-competitive bone marrow transplant assays for the remaining studies. Subsequent studies with this *Srsf2* allele from the same laboratory have also used this approach (Lee et al., 2018; Lee and Abdel-Wahab, 2016). At 18 weeks post pIpC, the *Srsf2*$^{P95H/+}$ recipients developed macrocytic anaemia and leukopenia due to decreased B cell numbers, while the platelet numbers remained normal. There was evidence of myeloid (hypolobated and hypogranulated neutrophils) and erythroid dysplasia in the peripheral blood/bone marrow but the estimated percentage of affected cells was not reported. Within the bone marrow of the recipients, Kim et al., reported that a major effect of the *Srsf2*$^{P95H/+}$ mutation was an expansion of the HPC-2 phenotype (LKS^{+}CD150^{+}CD48^{+}) within the hematopoietic stem and progenitor population. The authors reported an increased frequency

of cells in S-phase and early apoptosis within the bone marrow. There is an inconsistency of the reported phenotype presentation and penetrance across publications from the same laboratory using this mouse model that has not been addressed, but potentially indicates that the phenotype is not fully penetrant. Through analysis of RNA-seq from both human patients and murine samples the authors identified a group of mis-spliced candidate genes, with *EZH2* proposed as a key mis-spliced gene. This is reviewed in Section 1.6.2 of this thesis.

Kon et al., reported an *Srsf2* P95H conditional knock-in model developed using an inverted exon strategy (termed FLEx switch inversion technique; Figure 1.9) (Table 1.6) (Kon et al., 2018). RNA sequencing showed no expression from the modified allele in Cre-negative heterozygous animals, indicating that the floxed allele was likely null when not recombined. This was consistent with the failure to obtain animals that had the non-recombined P95H allele homozygosed even in a Cre-negative background (Kon et al., 2018). Why this is the case is unclear at present. The targeting strategy itself may be the most significant contributor in this particular instance. The separation of the 5'UTR from the first coding exon, in effect exonisation of the UTR, and the numerous modifications around the coding exon and the sequence insertion within the first intron may have had a significant effect on endogenous gene regulation.

The *Srsf2* P95H conditional knock-in mice were crossed to *Vav*-Cre mice to assess the effects of the heterozygous knock-in allele on haematopoiesis. Unlike the other described models, *Vav*-Cre is a constitutive Cre and is active in the hematopoietic cells without the addition of an inducer such as pIpC/IFNα (*Mx1*-Cre) or tamoxifen (CreER models). At around 15 weeks of age, the *Srsf2*$^{P95H/+}$ mice developed macrocytic anaemia, which remained stable. After observation out to 90 weeks of age, none of the

Srsf2$^{P95H/+}$ mice developed MDS or leukemia. A caveat with this model, as was described for the previous models describing inverted exon-based targeting methods, was that the mutant allele was expressed at 27-35% (mean 31%) based on RNA-seq analysis of the *Vav*-Cre *Srsf2*$^{P95H/+}$ mice.

Upon analysis of the bone marrow under native conditions, there was a reduction in the number of LT-HSC ($LKS^+CD150^+CD48^-$), MPP ($LKS^+CD150^-CD48^-$) and HPC-1 ($LKS^+CD150^-CD48^+$) populations, associated with increased cell cycling. There was no change in the cellularity or lineage distribution of the bone marrow and spleen, with the exception of a mild B cell differentiation defect in the bone marrow. There was no evidence of dysplasia reported in the bone marrow under native settings. The authors then performed non-competitive and competitive bone marrow transplant assays of the *Srsf2*$^{P95H/+}$ cells. There was engraftment in the non-competitive transplant settings, however the recipients of competitive bone marrow transplants had a significant reduction in chimerism with either whole bone marrow or HSCs (LKS^+CD34^-) as the donor. The impaired competitive transplant potential is a common feature of all the conditional splicing mutants to date (Kim et al., 2015; Kon et al., 2018; Smeets et al., 2018). Extensive work demonstrated an engraftment defect of the *Srsf2*$^{P95H/+}$ HSCs. The *Srsf2*$^{P95H/+}$ recipients developed macrocytic anaemia and leukopenia. In the bone marrow, there was a significant reduction in LT-HSC and most of the progenitor populations (MPP, HPC-1, CMP, CLP). In addition, *Srsf2*$^{P95H/+}$ recipients had a significantly higher proportion of cycling and apoptotic HSCs than controls. The lineage distribution was skewed towards myeloid at the expense of B cell differentiation and dysplastic erythroid cells as well as dys-megakaryopoiesis were evident in the bone marrow.

Analysis of RNA splicing using multiple purified hematopoietic

populations from both native haematopoiesis and cells isolated from transplant recipients was described. From these analyses the authors could identify mis-splicing of several genes that had been identified in human *SRSF2* mutant patient samples, including *Csf3r, Gnas, Hnrnpa2b1*, and several novel differentially spliced genes such as *Atrx* and *Mllt10* both of which are implicated in haematological malignancies. Of note, despite using numerous approaches the authors could not find evidence to support *Ezh2* as a differentially spliced candidate in the mouse. It was proposed that this may be due to species differences, with human *EZH2* having two CCNG motifs whilst mouse has a single CCNG motif (Kon et al., 2018).

Last but not least, my laboratory generated a conditional mouse model of *Srsf2* P95H mutation using a gene duplication method, independently of the two models described above (Figure 1.9) (Table 1.6) (Smeets et al., 2018). Upon Cre-mediated excision, the endogenous *Srsf2* locus was deleted and the duplicated locus containing the P95H mutation was retained. RNA sequencing confirmed the heterozygous expression at both the genomic and transcriptional level. Analysis of Cre negative animals and animals not treated with tamoxifen (the CreER inducer used in these studies) demonstrated a low level of transcriptional read-through in the absence of activation of the P95H bearing allele. Long-term monitoring of Cre-ve and non-tamoxifen treated/recombined Cre+ve animals demonstrated that this "leaky" expression was not pathogenic.

Multiple Cre lines were applied to understand the cell population in which P95H was required to be able to impact normal haematopoiesis: whole body/broadly expressed (*R26*-CreER^{T2}), more specific to the HSC and primitive progenitor populations (h*Scl*-CreER$^{T\,Tg/+}$) and a myeloid progenitor targeted constitutive Cre line (*Lys*M-Cre$^{Ki/+}$; described in

Chapter 3) (Clausen et al., 1999; Gothert et al., 2005; Ventura et al., 2007). We analysed the effect of *Srsf2*$^{P95H/+}$ on native haematopoiesis at 20 weeks post activation of the P95H mutation. In all cases, Cre+ve wild-type animals were used as controls and for CreER models these were tamoxifen treated CreER+ve *Srsf2* wild-type littermate controls. At 20 weeks post *Srsf2*$^{P95H/+}$ activation, *Srsf2*$^{P95H/+}$ mice developed macrocytic anaemia and increased myeloid cells in the peripheral blood. The myeloid bias was more evident in the bone marrow and was accompanied by compromised B lymphopoiesis. There was a reduction in erythropoiesis in the bone marrow accompanied by increased splenic erythropoiesis. Within the phenotypic stem and progenitor populations, there was a reduction or trend to reduction in the number of stem cell populations using two different phenotypic methods (LKS^{+}CD34/Flt3 or LKS^{+}CD105/CD150), while the more mature myeloid progenitor populations remained largely unchanged (Adolfsson et al., 2005; Kiel et al., 2005; Osawa et al., 1996; Pronk et al., 2007; Yang et al., 2005). The level of CD45RB, a previously characterized splicing target of SRSF2, was reduced in *Srsf2*$^{P95H/+}$ splenocytes, indicating that altered RNA splicing was being reflected in the proteome (Wang et al., 2001). RNA-seq analysis of lin^{-}cKit^{+}eYFP^{+} cells (where eYFP marked cells and their daughters/progeny in which Cre was activated) isolated 20 weeks after activation of P95H demonstrated that the *Srsf2*$^{P95H/+}$ mutation induced gene expression changes consistent with myeloid bias, loss of lymphoid potential and transcriptional signatures found in MDS. We identified similarly mis-spliced transcripts as Kon et al., and also could not define mis-splicing of *Ezh2* in the murine setting.

In agreement with the previous knock-in models of spliceosome mutations, the *Srsf2*$^{P95H/+}$ cells exhibited a poor competitive engraftment potential. This appears to be a universal feature of the models described

to date, with the work of Kon et al, best characterizing this phenotype and demonstrating an engraftment defect in the mutant cells (Kon et al., 2018). This experimental result is confounding, as *SRSF2* mutations are subjected to positive selection and are implicated in age-related clonal haematopoiesis (Genovese et al., 2014; Jaiswal et al., 2014; Martincorena et al., 2017). However, we observed that the poor engraftment could be modified by altering the nature of the bone marrow competitor used. Using the h*Scl*-CreER$^{T\,Tg/+}$ model, we found that *Srsf2*$^{P95H/+}$ cells could competitively engraft recipient animals and expanded when transplanted with the age/microenvironment matched competitor cells (i.e. cells taken from the same bone marrow but where Cre was not activated, based on *R26*-eYFP reporter marking of the cells).

Upon ageing, non-transplanted *Srsf2*$^{P95H/+}$ mice developed fatal MDS by 12 months after activating the P95H mutation in both the *R26*-CreER^{T2} and h*Scl*-CreER$^{T\,Tg/+}$ model. The disease in both models was highly comparable. When moribund, the mice presented with macrocytic anaemia, myeloid bias (granulocytosis and monocytosis) and morphological dysplasia of myeloid and erythroid lineages in both the peripheral blood (>10-50% of cells) and bone marrow, all characteristics of MDS/MPN. Analysis of a small cohort by exome capture demonstrated that there was a subclonal accumulation of mutations associated with human *SRSF2* mutant disease, including *Dnmt3a, Tet2, Phf6*, and *Ras* members, in the bone marrow of the moribund animals (Makishima et al., 2012; Smeets et al., 2018). Non-competitive transplant of the moribund bone marrow recapitulated the MDS/MPN but there was no evidence of progression to acute leukemia even with long term monitoring of secondary recipients or with concurrent *p53* deletion. Of the currently described *Srsf2* point mutant models, this is the only model to develop monocytosis and MDS/MPN in

the setting of native haematopoiesis. The review of *Srsf2* mutant mouse models and other *in vivo* mouse models of splicing mutation, including *Sf3b1* and *U2af1* has been published (Xu et al., 2019).

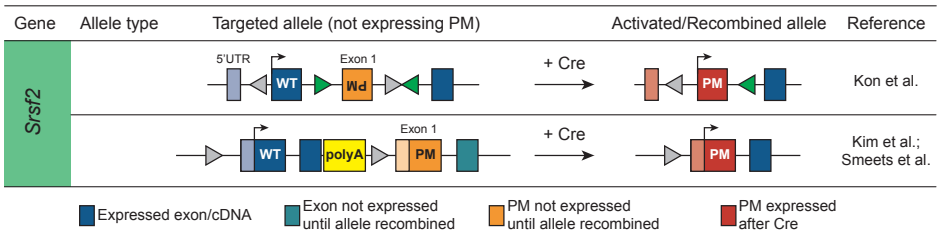


Figure 1.9: Detailed genomic architecture of the *Srsf2*$^{P95H/+}$ knock-in mutant models.
LoxP elements are indicated as grey or green triangles indicating different types of loxP elements. Expressed exons are shown in blue, exons not expressed until after locus recombination are shown in teal, and mutation-bearing exons are shown in orange before recombination (silent) and in red when recombined and expressed.

1.10.2 Mouse models of *Tet2* mutation

Delhommeau et al. firstly demonstrated the functional consequences of *TET2* mutation with cells from a myeloproliferative neoplasm (MPN) patient. There was an increase in the engraftment of *TET2* mutant MPN cells in immunodeficient mice (Delhommeau et al., 2009). Following this finding, knock-out mouse models were generated to understand the consequences of *TET2* loss upon haematopoietic development (Table 1.6). In these models, *Tet2*$^{-/-}$ mice developed a CMML-like disease with leukocytosis and monocytosis (Ko et al., 2011; Moran-Crusio et al., 2011). *Tet2*$^{-/-}$ haematopoietic stem cells showed an increased self-renewal and a skewed differentiation towards myeloid lineages (Quivoron et al., 2011). As expected, methylation analysis showed a dramatic reduction in total 5hmC in *Tet2* mutant mice. Compared to *Tet2*$^{-/-}$, *Tet2*$^{+/-}$ mice also displayed similar phenotypes demonstrating that *Tet2* haploinsufficiency

is sufficient to cause aberrant haematopoiesis. The murine models (both germline and conditional models) resemble the human mutation pattern, as most patients carry monoallelic *TET2* mutation rather than biallelic mutation (Abdel-Wahab et al., 2009; Jankowska et al., 2009). Recently, Awada et al. demonstrated that biallelic *TET2* mutation has a stronger association with monocytosis, hypercellularity and profound myeloid dysplasia than monoallelic *TET2* mutation. This indicates that the different types of *TET2* mutation might have diagnostic value to identify specific morphologic sub-entities in myeloid neoplasm (Awada et al., 2019).

So far, one study has attempted to reverse the effect of *TET2* mutation by restoring its function using *in vivo* shRNA transgenic models. With a reversible *Tet2* knock down model, Cimmino et al. showed that restoration of *Tet2* expression stopped aberrant self-renewal and proliferation of stem cells and promoted DNA demethylation and differentiation of myeloid lineage cells (Cimmino et al., 2017). To restore TET2 function in patients, Cimmino et al. proposed to augment TET2 activity pharmacologically as re-introduction of *TET2* genetically is less feasible. Alpha-ketoglutarate and Fe^{2+} are both co-factors in the TET2 demethylation reaction. Vitamin C, as an antioxidant, was found to increase TET2 activity via reducing Fe^{3+} to Fe^{2+}. When Vitamin C was administered, it could mimic the genetic restoration of *Tet2* in mice and block aberrant self-renewal and progression of myeloid disease (Cimmino et al., 2017). It is a promising therapeutic strategy. However, the augmentation would only apply to monoallelic *TET2* mutation, as the prerequisite for increasing the function of a protein is to have an endogenous wild-type *TET2* allele present, which is around 55% of patients with *TET2* mutations (Delhommeau et al., 2009). Further studies are required to understand the pharmacokinetics of Vitamin C to make this treatment feasible.

Table 1.6: Summary of the phenotypes of the *Srsf2* and *Tet2* mutant mouse models

Gene	Model	rtTA/Cre	VAF	Native hematopoiesis			Non-competitive transplant			Competitive transplant
				Peripheral Blood/Spleen	BM	Dysplasia/ Disease	Peripheral Blood/Spleen	BM	Dysplasia/ Disease	
Srsf2	Duplication (Kim et al.)	*Mx1*-Cre	~50 %	Not described			Macrocytic anemia Leukopenia: ↓B cells	↑LKS; ↑LKS in S phase; ↑HPC2 ↓B cells	Neutrophil and erythroid dysplasia in PB; No AML (70wks)	↓↓Engraftment↑LKS ↑HPC-2
	Inversion (Kon et al.)	*Vav1*-Cre	~31 %	Macrocytic anemia	↓LTHSC ↓MPP ↓HPC-1	No dysplasia /No MDS or AML (90 weeks)	Macrocytic anemia Leukopenia ↑Grans ↓B cell	↓LTHSC; ↓MPP; ↓HPC1; ↓CMP; ↓CLP;	Dysplastic erythroid and ↑megakaryo-cytes in BM /No MDS or AML	↓↓Engraftment; defect in homing
	Duplication (Smeets et al.)	*R26*-CreER^{T2}; h*Scl*-CreER^{T} $^{Tg/+}$; *Lys*M-Cre	~50 %	Macrocytic anemia; ↑Grans; ↑Monocytes.	↓LTHSC; ↑%Grans; ↓B cells; ↓erythroid.	MDS/MPN by ~12 mths; no AML; ↑ and dysplastic myeloid cells (10-50% PB), monocytosis in PB / transplantable to 2^{ndry} recipients	Macrocytic anemia Leukopenia ↑%Grans & monocytes ↓B cell; Spleen - ↑% monocytes	↑%Grans; ↑%monocytes.	No AML	↓↓Engraftment in *R26*-CreER^{T2} model; myeloid bias; engraftment and ↑ mutant cells in h*Scl*CreER^{T} $^{Tg/+}$ model.
Tet2	Germ-line deletion (Ko et al.)	N/A	N/A	↑Spleen cellularity	↑Bone marrow cellularity ↑%LKS/LK ↑CMP	Not described	Splenomegaly in some recipients	↑LKS/LK	Not described	↑↑Engraftment

Tet2	Conditional deletion (Moran-Crusio et al.)	*Vav1*-Cre *Mx1*-Cre	N/A	Splenomegaly ↑%CD11b+	↑LKS ↑GMP	CMML by 20 weeks; Leukocytosis with neutrophilia and monocytosis in PB; myeloid dysplasia; splenomegaly	Not described			↑↑Engraftment
	Haploinsufficient conditional deletion (Moran-Crusio et al.)	*Vav1*-Cre	N/A	Not described			↑Monocytes; extramedullary haematopoiesis	↑LK and LKS cells	Not described	↑Engraftment (slower kinetics than Tet2 null cells)
	Hypomorphic deletion (Quivoron et al.)	*Mx1*-Cre	~20-50 %	Similar to conditional deletion model (Quivoron et al.)	Similar to conditional deletion model (Quivoron et al.)	Leukocytosis, aneamia, thrombocytopenia, splenomegaly; spleen fibrosis; myeloid/erythroid hyperplasia in the spleen /transplantable to 2^{ndry} recipients	Not described			↓↓Engraftment in *R26*-$CreER^{T2}$ model; myeloid bias; engraftment and ↑ mutant cells in h*Sc*/$CreER^{T\ Tg/+}$ model.

Tet2	Conditional deletion (Quivoron et al.)	*Mx1*-Cre	~20 - 50 %	↑WBC ↓RBC ↓Platelets; hepatosplenomegaly; ↑Grans in PB and spleen ↑early erythroblast in the spleen	↑early erythroblast ↓late erythroblast ↓Pre-, Pro- and mature B cells	No disease development up to 15 months	Not described	↑Engraftment ↑LKS Similar myeloid/lymphoid lineage change to native animals

The immunophenotypic population listed in this table is defined as following: LKS^+, $linage^-/c\text{-}Kit^+/Sca\text{-}1^+$; LKS^-, $lineage^-/c\text{-}Kit^+/Sca\text{-}1^-$; LT-HSC, $LKS^+/CD150^+/CD48^-$; MPP, $LKS^+/CD150^-/CD48^-$; HPC-1, $LKS^+/CD150^-/CD48^+$; HPC-2, $LKS^+/CD150^+/CD48^+$; HSC (Mupo et al.), $LKS^+/CD34^-/flk2^-$; pre-MegE, $LKS^-/CD41^-/Fc\gamma R^-/CD150^+/CD105^-$; pre-CFUE, $LKS^-/CD41^-/Fc\gamma R^-/CD150^+/CD105^+$. % = the percentage of this population.

1.11 Mouse models of co-operative mutations in MDS/MPN

Considering the complex genetic landscape in MDS and MDS/MPN, studies have started to focus more on elucidating the mechanism of potential co-operative mutations in disease initiation. So far, models have looked at co-occurring mutations in DNA methylation (*TET2*), chromatin remodelling (*EZH2*), Ras signalling (*N-RAS*) and spliceosome (*SRSF2*). Muto et al., initially reported a model of co-deletion of *Ezh2* and *Tet2*, where loss of both genes significantly shortened the latency of MDS and where mice developed a more aggressive form of MDS or CMML (Muto et al., 2013). The enhanced phenotypes indicated that there is a co-operative effect of concurrent *Ezh2* and *Tet2* mutations (Muto et al., 2013). A follow-up study using the same model analysed the epigenetic changes in the MDS mice, since both *Ezh2* and *Tet2* are epigenetic regulators of the genome. In double knock-out cells, there was an increase in aberrant DNA methylation and a loss of H3K27me3 marks compared to single knock-out cells. However, the mechanism of their co-operativity is still unclear (Hasegawa et al., 2017). Another double mutant model, characterised by Kuminoto et al., proposed a mechanistic co-operativity of *Tet2* and *NRas* mutation (Kunimoto et al., 2018). While *Tet2* mutation destabilises the methylome, an *NRas* mutation is acquired and leads to a decrease in negative regulators of mitogen-activated protein kinase (MAPK) activation. The aberrant activation is augmented by *Tet2* mutation and then leads to clonal expansion and MDS. The double mutant cells are hypersensitive to MEK inhibitors, which further supports the hypothesis of cooperativity (Kunimoto et al., 2018).

Yoshimi et al., investigated cooperativity of *SRSF2* and *IDH2* mutation (Yoshimi et al., 2019). Overexpression of human *IDH2* mutant (*IDH2*^{R140Q} or *IDH2*^{R172K}) in *Vav*-Cre *Srsf2*$^{P95H/+}$ cells led to significant myeloid skewing, macrocytic leukaemia and thrombocytopenia. Transplant recipients of *Srsf2* and *IDH2* double mutant cells developed lethal MDS with blast-like BM cells. Similar phenotypes were reported when both mutants were expressed from endogenous loci in a *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Idh2*^{R140Q} model. There are few caveats in this study. To begin, authors only reported phenotypes in a transplantation setting but not in native haematopoiesis in *Mx1*-Cre *Srsf2*$^{P95H/+}$ *Idh2*^{R140Q} models. Furthermore, deletion of *Tet2* in *Srsf2*$^{P95H/+}$ cells was unable to replicate some phenotypes seen in *Srsf2* and *Idh2* double mutant cells, even though TET2 inhibition is a recognised mechanism of IDH2 mutant leukaemogenesis. Yoshimi et al., finally proposed that *INT3* is at the centre of IDH2 and SRSF2 mutant co-operativity. *SRSF2* mutation leads to aberrant splicing of *INT3* and subsequent downregulation of gene expression, which is further promoted by *IDH2* mutation. Restoration of *INT3* expression in *SRSF2* and *IDH2* double mutant HL-60 cells released the differentiation block in a xenograft model.

1.12 Hypothesis

MDS are heterogeneous haematological malignancies characterised by multilineage dysplasia, cytopenia and ineffective haematopoiesis, while CMML is characterised by monocytosis and multilineage dysplasia. Until recently, the diagnostic criteria for MDS and CMML have relied heavily on morphological findings and cytogenetic profiles and with limited treatment options, the prognoses for MDS and CMML have improved

very little over the last 10 years, despite the introduction of DNA hypomethylating agents. In the past decade, the description of their genetic landscapes has opened new vistas for understanding the pathogenesis of MDS and CMML. Novel recurrent mutations in the spliceosome and DNA methylation pathways have been found to be associated with different disease phenotypes. By studying the effect of these recurrent genetic mutations in murine models, scientist have established a greater understanding of the role of each mutation and how it disrupts normal haematopoiesis.

In particular, *Srsf2* mutation is the most frequent spliceosome mutation in CMML and in high-risk MDS, such as MDS with excess blasts (Mian et al., 2013; Patel et al., 2017). Although several models have explored the role of *Srsf2* mutation in MDS, a faithful model that can reproduce MDS phenotypes in native haematopoiesis is still lacking (Kim et al., 2015; Kon et al., 2018). To address this, my lab generated a conditional *Srsf2* mutant model that developed lethal MDS in native haematopoiesis (Smeets et al., 2018).

Following up on this study, I will further examine the effects of $Srsf2^{P95H/+}$ mutation when activated in myeloid progenitors. Mechanistically, I will examine the molecular consequences of $Srsf2^{P95H/+}$ mutation on mis-splicing and gene transcription and the functional genetic interactions with $Srsf2^{P95H/+}$ through a synthetic lethal CRISPR knock-out screen. At last, I will examine the co-operativity of *Srsf2* with *Tet2* (*Srsf2* x *Tet2*) and *Srsf2* with *Cbl* (*Srsf2* x *Cbl*) mutations and how these co-occurring events lead to CMML.

List of Hypothesis

1. Activating *Srsf2*$^{P95H/+}$ in haematopoietic stem cells is required to initiate MDS.

2. *Srsf2*$^{P95H/+}$ mutation results in specific transcriptomic and splicing changes that drive the disease.

3. *Srsf2*$^{P95H/+}$ mutation leads to genetic vulnerabilities in pathways other than mRNA splicing.

4. *Srsf2*$^{P95H/+}$ mutation co-operates with *Tet2* or *Cbl* mutation to initiate MDS/MPN.

1.13 Aims

The aims of my thesis are:

1. To determine the effect of activating *Srsf2*$^{P95H/+}$ mutation in myeloid progenitors *in vivo*.

2. To define the effect of *Srsf2*$^{P95H/+}$ mutation on the transcriptome and mis-splicing of haematopoietic stem cells.

3. To determine the negative genetic interactions of *Srsf2*$^{P95H/+}$ mutation *in vitro*.

4. To understand the co-operativity of *Srsf2*$^{P95H/+}$ with *Tet2* mutation and *Srsf2*$^{P95H/+}$ with *Cbl* mutation *in vivo*.

Chapter 2

Materials and Methods

2.1 Materials and Reagents

All materials, reagents and equipment used in experiments and the source/suppliers or manufacturers are listed in Table 2.1, 2.2 and 2.3.

2.2 Mice

All mice experiments were approved by the Animal Ethics Committee, St. Vincent's Hospital, Melbourne, Australia (AEC#001/16 and 007/19). Mice used in the experiments were housed at the Bioresource Centre (BRC) located at St. Vincent's Hospital.

The *Srsf2*$^{P95H/+}$ (C57BL/6NTac-*Srsf2*$^{tm2874(P95H)Arte}$), *Rosa26*-CreER^{T2} (strain #008463, Jackson Laboratory), h*Scl*-CreER^{T}, *Rosa26*-eYFP (strain #006148, Jackson Laboratory) and *LysM*-Cre knock-in mice (strain #004781; Jackson Laboratory) have been previously described (Smeets et al., 2018; Xu et al., 2019). The *Tet2*$^{fl/fl}$ mice (B6;129S-*Tet2*$^{tm1.1laai}$/J; strain #017573, Jackson Laboratory), *c-Cbl*$^{-/-}$ mice and *c-Cbl*$^{A/-}$ mice (*Cbl*^{tm2Wlan}) were previously described as well (Moran-Crusio et al., 2011; Murphy et al., 1998; Rathinam et al., 2010). The *c-Cbl*$^{-/-}$ mice and *c-Cbl*$^{A/+}$ mice were kindly provided by Professor Wallace Y. Langdon, University of Western Australia. The *c-Cbl*$^{-/-}$ and *c-Cbl*$^{A/-}$ mice were on a mixed C57BL/6J x 129Sv background. All the other mouse strains are on a C57BL/6J background. h*Scl*-CreER^{T} $^{Tg+}$ *Rosa26*-eYFP$^{ki/ki}$ *Srsf2*$^{P95H/+}$ mice were crossed to *Tet2*$^{fl/fl}$ mice to generate h*Scl*-CreER^{T} $^{Tg+}$ *Rosa26*-eYFP$^{ki/ki}$ *Srsf2*$^{P95H/+}$ *Tet2*$^{fl/fl}$ mice. h*Scl*-CreER^{T} $^{Tg+}$ *Rosa26*-eYFP$^{ki/ki}$ *Srsf2*$^{P95H/+}$ mice were crossed to the *c-Cbl*$^{-/-}$ mice to generate h*Scl*-CreER^{T} $^{Tg+}$ *Rosa26*-eYFP$^{ki/ki}$ *Srsf2*$^{P95H/+}$ *c-Cbl*$^{-/-}$ mice. h*Scl*-CreER^{T} $^{Tg+}$ *Rosa26*-eYFP$^{ki/ki}$ *Srsf2*$^{P95H/+}$ *c-Cbl*$^{-/-}$ mice were crossed to the *c-Cbl*$^{A/+}$ mice to generate h*Scl*-CreER^{T} $^{Tg+}$ *Rosa26*-eYFP$^{ki/ki}$

Table 2.1: Reagents and materials used in experiments and the source

Reagent	Supplier
23-gauge needles	BD Biosciences (San Jose, CA, USA)
4-Hydroxytamoxifen	Merck Millipore (Billerica, MA, USA)
Agarose tablets	Bioline (London, UK)
ATPlite Luminescence Assay System	Perkin Elmer (Waltham, MA, USA)
BD microtainer K2E tubes	BD Biosciences (San Jose, CA, USA)
BD vacutainer capillary tubes	BD Biosciences (San Jose, CA, USA)
Blasticidin	Sigma-Aldrich (St. Louis, MO, USA)
Calcium chloride ($CaCl_2$)	Sigma-Aldrich (St. Louis, MO, USA)
Cell counter chamber slide (Countess)	BioTools (Keperra, QLD, Australia)
Cell strainer (40µm)	Corning (Tewksbury, MA, USA)
Chromium™ Chip Kit	10x Genomics (California, USA)
Chromium™ Gel Bead Kit & Multiplex Kit	10x Genomics (California, USA)
Cryotube vials	Thermo Fisher Scientific (Waltham, MA, USA)
CutSmart buffer	New England Biolabs (Ipswich, MA, USA)
Cytology funnels	BMP Medical (Sterling, MA, USA)
Dimethyl sulfoxide (DMSO)	Sigma-Aldrich (St. Louis, MO, USA)
DirectZol Microprep kits	ZYMO Research (Califormia, USA)
Disposable Pipettes (5, 10, 25mL)	Corning (Tewksbury, MA, USA)
Disposable scalpel	Swann Morton (Sheffield, England)
DNA molecular weight marker 50bp (HyperLadder)	Bioline (London, UK)
DPX mounting media	Sigma-Aldrich (St. Louis, MO, USA)
Dulbecco's Modified Eagle Medium (DMEM)	Sigma-Aldrich (St. Louis, MO, USA)
Dulbecco's Phosphate Buffered Saline (PBS)	Sigma-Aldrich (St. Louis, MO, USA)
Eppendorf tubes	Corning (Tewksbury, MA, USA)
Ethanol (EtOH)	Thermo Fisher Scientific Australia Pty Ltd (Scoresby, VIC, Australia)
Ethylenediaminetetra acetic acid (EDTA)	Astral Scientific (Taren Point, NSW, Australia)
Falcon centrifuge tubes (5, 10, 50mL)	Thermo Fisher Scientific (Waltham, MA, USA)
Ficoll-Paque™ PLUS	GE Healthcare (Buckinghamshire, UK)
Foetal bovine serum (FBS)	Assay Matrix (Melbourne, VIC, Australia)
Gel Loading Dye	New England Biolabs (Ipswich, MA, USA)
Gentra Puregene Cell Kit	Qiagen (Hilden, Germany)
Giemsa solution	Amber Scientific (Midvale, WA, Australia)
HybridaseTM Thermostable RNase H (5U/µL)	Illumina (California, USA)
Insulin syringes	BD Biosciences (San Jose, CA, USA)
Iscove's Modified Dulbecco's Medium (IMDM)	Sigma-Aldrich (St. Louis, MO, USA)
Isolate II Genomic DNA kit	Bioline (London, UK)

Isolate II PCR and Gel kit	Bioline (London, UK)
Isopropanol	Merck Pty Ltd (NSW, Australia)
Kapa Stranded RNA-seq Library Preparation Kit	Kapa Biosystem/Roche (Basel, Switzerland)
L-Alanyl-L-Glutamine	Sigma-Aldrich (St. Louis, MO, USA)
May-Grünwald solution	Amber Scientific (Midvale, WA, Australia)
Methanol (MeOH)	Thermo Fisher Scientific Australia Pty Ltd (Scoresby, VIC, Australia)
Microscope slides and coverslips	Thermo Fisher Scientific Australia Pty Ltd (Scoresby, VIC, Australia)
Mr Frosty™ Freezing Container	Thermo Fisher Scientific (Waltham, MA, USA)
MyTaq HS DNA polymerase	Bioline (London, UK)
MyTaq Red Reaction Buffer	Bioline (London, UK)
Nunc™ EasYFlask™ Cell Culture T75 Flasks	Thermo Fisher Scientific (Waltham, MA, USA)
Oligonucleotides PCR Primers	Integrated DNA Technologies (IDT; Coralville, Iowa, USA)
Palbociclib	Selleckchem (Houston, Texas, USA)
Paraformaldehyde (PFA) aqueous solution (16%)	Proscitech (QLD, Australia)
PCR 8-strip Tubes with flat-caps	SSIbio (Lodi, CA, USA)
PD0166285	Selleckchem (Houston, Texas, USA)
Penicillin/Streptomycin	Sigma-Aldrich (St. Louis, MO, USA)
Pladienolide B	Santa Cruz Biotechnology (Dallas, USA)
Polybrene (Hexadimethrine bromide)	Sigma-Aldrich (St. Louis, MO, USA)
Potassium chloride (KCl)	Sigma-Aldrich (St. Louis, MO, USA)
Puromycin Dihydrochloride	Merck (Darmstadt, Germany)
qPCR Master Mix	Agilent Technologies (Santa Clara, CA, USA)
Recombinant human IL-6	Amgen (Thousand Oaks, CA, USA) (gift)
Recombinant murine IL-3	Peprotech (Rocky Hill, NJ, USA)
Recombinant murine Stem Cell Factor (mSCF)	Peprotech (Rocky Hill, NJ, USA)
Restriction enzymes (various)	New England Biolabs (NEB, Ipswich, Massachusetts, USA)
RiboSafe Murine RNase Inhibitor (40U/µL)	Bioline (London, UK)
Roscovitine	Selleckchem (Houston, Texas, USA)
Skim milk powder	Devondale (VIC, Australia)
Sodium chloride (NaCl)	Sigma-Aldrich (St. Louis, MO, USA)
Sodium hydroxide (NaOH) tablets	Merck (Darmstadt, Germany)
SYBR Safe™ DNA gel stain	Thermo Fisher Scientific (Waltham, MA, USA)

Syringes (1mL, 5mL)	BD Biosciences (San Jose, CA, USA)
TE Buffer	Invitrogen (Carlsbad, CA, USA)
Tetro cDNA synthesis kit	Bioline (London, UK)
Tissue culture 10-cm dishes	Corning (Tewksbury, MA, USA)
Tissue culture flask (T25, T75, T175)	Corning (Tewksbury, MA, USA)
Tissue culture plates (6, 12, 24, 48, 96 wells)	Corning (Tewksbury, MA, USA)
Tris	Astral Scientific (Taren Point, NSW, Australia)
TRIzol™ reagents	Thermo Fisher Scientific (Waltham, MA, USA)
Trypan blue solution (0.4%)	NanoEnTek (Seoul, South Korea)
Trypsin-EDTA solution	Sigma-Aldrich (St. Louis, MO, USA)
TURBO DNA-free™ Kit (Includes 10x TURBO DNase buffer and Turbo DNase)	Life Technologies (California, USA)
UltraPure DNAse/RNAse-free water	Life Technologies (Carlsbad, CA, USA)

Table 2.2: List of solutions and buffers prepared for use in experiments

Solutions	**Composition**
5x RNaseH(-) Mg buffer	500mM Tris pH 7.4, 1M NaCl
Agarose gel 3%	3g agarose tablets in 100mL TBE 1x
Cell Culture Medium Supplements	10% (v/v) FSB, 1% (v/v) Penicillin-Streptomycin, 1% (v/v) Glutamine, 1% (v/v) GM-CSF; GM-CSF conditioned media collected from BHK-HM5 cells (Gift from Prof Steve Collins, Fred Hutchinson Cancer Research Centre)
Cell freezing medium	90% (v/v) FBS, 10% (v/v) DMSO
FACS buffer	2% FBS (v/v) in PBS
Fixing solution (for tissues)	2% PFA in PBS from 16% Paraformaldehyde(v/v)
HEPES buffered saline (HBS)	140mM NaCl, 0.75mM Na_2HPO_4, 5mM KCl, 6mM Dextrose, 25mM HEPES, pH 7.05
Red blood cell (RBC) lysis buffer	150mM NH_4Cl, 10mM $KHCO_3$, 0.1mM Na_2 EDTA, pH7.3
TAE buffer	40 mM Tris, 40 mM Acetate, 1 mM EDTA, pH 8.0.
Stimulating medium for culturing bone marrow cells	IMDM, 20% (v/v) FBS, 1% (v/v) Penicillin-Streptomycin, 1% (v/v) Glutamine, 1% (v/v), 50ng/mL rmSCF, 10ng/mL rmIL3, 10ng/mL rhIL6

Table 2.3: List of equipment and their manufacture used in the experiments

Equipment	Manufacturer
Allegra X-12 Centrifuge	Beckman Coulter (Brea, CA, USA)
Analytical balances UW220H	Shimadzu (Sydney, NSW, Australia)
Automatic Pipette S1 (with filter)	Thermo Fisher Scientific (Waltham, MA, USA)
BD FACSAria	BD Biosciences (San Jose, CA, USA)
BD FACSInflux	BD Biosciences (San Jose, CA, USA)
BD LSRII Fortessa	BD Biosciences (San Jose, CA, USA)
CO_2 Incubator HeraCell 240i	Thermo Fisher Scientific (Waltham, MA, USA)
Countess II cell counter	Thermo Fisher Scientific (Waltham, MA, USA)
Cytospin™4 Cytocentrifuge machine	Thermo Fisher Scientific (Waltham, MA, USA)
Enspire plate reader	PerkinElmer (Melbourne, VIC, Australia)
EPSON Perfection V700 photo scanner	EPSON (Suwa, Nagano, Japan)
Freezer (-80°C)	Thermo Fisher Scientific (Waltham, MA, USA)
Gelaire Class II Biological Safety Cabinet	Gelaire Pty Ltd (Sydney, NSW, Australia)
Heraeus Multifuge 3SR+ Centrifuge	Thermo Fisher Scientific (Waltham, MA, USA)
Heraeus Pico 17 Centrifuge	Thermo Fisher Scientific (Waltham, MA, USA)
LabChip GX Touch Nucleic Acid Analyzer	PerkinElmer (Melbourne, Victoria)
Leica DMRB fluorescence microscope	Leica (Wetzlar, Germany)
Mastercycler Pro PCR machine	Eppendorf (Hamburg, Germany)
Medium Orbital Shaker	Ratek (Boronia, VIC, Australia)
Milli-Q™ ultra-pure water	Thermo Fisher Scientific (Waltham, MA, USA)
PowerEase 90W	Life Technologies (Carlsbad, CA, USA)
Powerpac Electrophoresis System	Bio-Rad Laboratories (Hercules, CA, USA)
Retiga-EXi 12 Bit CCD camera	QImaging (Tucson, AZ, USA)
Rotary Tube Mixer with Disc	Ratek (Boronia, VIC, Australia)
Spectrophotometer ND – 1000 (Nanodrop)	Thermo Fisher Scientific (Waltham, MA, USA)
Stratagene Mx3000P qPCR machine	Agilent Technologies (Santa Clara, Ca, USA)
Sysmex KX21 Haematology Analyser	Sysmex Corporation (Kobe, Hyogo, Japan)
T3000 Thermocycler	Biometra (Göttingen, Germany)
Ultra Basic pH Meter	Denver Instrument (Bohemia, NY, USA)
VersaDoc Imaging System	Bio-Rad Laboratories (Hercules, CA, USA)
Vortemp 56 Shaking Incubator	Labnet International Inc. (Edison, NJ, USA)
Vortex Mixer SA8	Stuart Equipment (Staffordshire, UK)
Water bath WB7	Ratek (Boronia, VIC, Australia)

Srsf2$^{P95H/+}$ *c-Cbl*$^{A/-}$ mice. Congenic B6.SJL-Ptprc$^{aPep3b/BoyJArc}$ were purchased from the Animal Resources Centre (Canning Vale, Western Australia). Heterozygous CD45.1/CD45.2, derived from the intercross of C57BL/6 and B6.SJL-Ptprc$^{aPep3b/BoyJArc}$ mice, were bred at St. Vincent's Hospital. Tamoxifen containing food was prepared containing 400mg/kg tamoxifen citrate (Sigma or Selleckchem) in standard mouse chow (Specialty Feeds, Western Australia) and irradiated. The tamoxifen chow was fed ad *libitum* during the treatment.

2.3 DNA Isolation and Genotyping

For genotyping, toe clips from 7-10 day old mice were collected at numbering in 1.5mL microcentrifuge tubes. The tubes were spun briefly and digested in 300μL 50mM NaOH for 20 minutes at 95^{o}C and 1,200 revolutions per minute (rpm) in a Vortemp 56 shaking incubator. After digestion, 100μL of 1M Tris pH8.0 was added and vortexed immediately for 5 seconds. The tube was spun at 17,000g for 5 minutes in a Heraeus Multifuge 3SR+ centrifuge. The soluble gDNA was used for genotyping. For experimental mice and cell lines, gDNA was extracted from cell pellets using a Bioline Genomic DNA Kit as per the manufacturer's instructions.

Genotyping was performed by polymerase chain reaction (PCR) in the Mastercycler Pro PCR machines (Eppendorf). All oligonucleotides were ordered from Integrated DNA Technology (IDT) and are listed in Table 2.4. Reactions were set up in PCR tubes as described in Table 2.5. For the Srsf2 6036 PM reaction, 0.5μL of dimethylsulfoxide (DMSO) was added and the nuclease-free water was reduced to 5.4μL.

All genotyping reactions were performed on the program described

in Table 2.6. PCR products from the 6036 reaction require further restriction enzyme digestion to confirm the nucleotide change. The digestion mixture contained 2μL 10x CutSmart buffer (New England Biolab), 0.15μL Hpy188III restriction enzyme (New England Biolab), 7.85μL nuclease-free water, and 10μL of the genotyping PCR reaction. This was mixed by pipetting and incubated at 37^{o}C for 45 minutes to allow enzymatic digestion. Details of the PCR products for wild type (WT) and point mutant (PM) alleles and the digested products are described in Table 2.4. PCR products, including digested PCR products, and a standard DNA 50bp ladder (Hyperladder, Bioline), were loaded into 3% agarose gels containing SYBR DNA gel stain (1:20,000, Bioline) and electrophoresed using a Powerpac Electrophoresis (Bio-Rad) system at 120V for 50 minutes. After electrophoresis, the gel was imaged with a VersaDoc imaging system (Bio-Rad).

Table 2.4: Genotyping Oligonucleotide sequences and product size

Genotyping allele	Oligonucleotide Sequence (5'-3')	Detected allele	Product size (bp)
h*Scl*-CreER^{T}	CAACAACAACCGGGTGAAGA	Transgenic	209
	CTTGCACCATGCCGCCCACGAC		
	GAGACTCTGGCTACTCATCC	Internal control	585
	CCTTCAGCAAGAGCTGGGGAC		
Rosa26-CreER^{T2}	CCTGATCCTGGCAATTTCG	WT	650
	GGAGCGGGAGAAATGGATATG	KI	825
	AAGTCGCTCTGAGTTGTTAT		
Rosa26-eYFP	CCTGATCCTGGCAATTTCG	WT	500
	GGAGCGGGAGAAATGGATATG	KI	250
	AAAGTCGCTCTGAGTTGTTAT		
Srsf2 6034 5'loxP	GTTATGATCCACACCTCTCACC	WT	327
	ATAAACGTTTATGTCCGCTACC	Conditional KI	446
Srsf2 6035 3'loxP	ACGTCCAGACACAGCATAGG	Conditional KI	364
	GGCTTGCAATTTGAGTATGG		
	GAGACTCTGGCTACTCATCC	Internal control	585
	CCTTCAGCAAGAGCTGGGGAC		
Srsf2 6036 PM	ACGTGAACGAAGCGACAGG	WT & PM	495
		WT post digestion	487+8
	GCGCGTCTTCGAGAAATACG	PM post digestion	303+184+8
Tet2 fl	AAGAATTGCTACAGGCCTGC	WT	249
	TTCTTTAGCCCTTGCTGAGC	Floxed allele	427
c-Cbl knock-out	TAGGCGAAACCTGACCAAAT	WT	456
	CTCGC CAGTCCCTGCTTAC	KO	288
	TGCTACTTCCATTTGTCACG		

Table 2.5: Genotyping PCR mixture

Reagent	Volume (µL)
gDNA	1
5x MyTaq red reaction buffer	2
pooled oligonucleotides (10µM)	1
MyTaq HS DNA polymerase	0.1
Nuclease-free water*	5.9
Total Volume	**10**

*For Srsf2 6036 PM reaction, add 0.5μL of DMSO and reduce nuclease-free water to 5.4μL

Table 2.6: Genotyping PCR cycling conditions

Temperature	Duration	Cycle Number
95°C	5 minutes	x1
95°C	20 seconds	x35
60°C	30 seconds	
72°C	45 seconds	
72°C	3 minutes	x1
20°C	Hold	

2.4 Flow Cytometry and Fluorescent Activated Cell Sorting and Cell Analysis

2.4.1 Tissue Collection and Preparation

For regular monitoring of animals, peripheral blood samples (100-150μL) were obtained by retro-orbital eye-bleeding into BD microtainer K2E tubes using BD vacutainer capillary tubes. For analysis at the end of the experiment or when animals were moribund, peripheral blood samples were obtained by the same method. Mice were then euthanised by CO^2 asphyxiation or cervical dislocation. All blood sample collection and euthanasia procedures were performed by Prof Carl Walkley. At the time of analysis, both ears were collected in 1.5mL microcentrifuge tubes and stored in a -80^{o}C freezer for re-genotyping if required. Tibia and sterna were fixed in 2% Paraformaldehyde (PFA) in PBS overnight. The bones were then washed 3 times with PBS and transferred into 70% ethanol in a 10mL tube and then stored at 4^{o}C.

2.4.2 Peripheral Blood Analysis

For full blood count, 25μL of peripheral blood was mixed with 75μL PBS and differentially counted on a Sysmex KX21 haematological analyser. Blood films were prepared by adding 2μL of peripheral blood onto a clean glass microscope slide. A second clean slide was then used to disperse the blood over the slide's length in a 45 degree angle. The slide was allowed to air-dry for at least an hour and then fixed in 100% methanol for 10 minutes. The blood film was then allowed to air-dry and was stored at room temperature (RT). For flow cytometry analysis, the remaining peripheral blood was depleted of red blood cells by adding 1mL red blood cell lysis buffer (150mM NH_4Cl, 10mM $KHCO_3$, 0.1mM Na_2EDTA,

pH 7.3). The sample was mixed by inversion and then centrifuged at 400g for 5 minutes. After centrifuging, the supernatant was discarded, and the lysis repeated. The cell pellet was washed with 1mL PBS and centrifuged before being re-suspended in 150μL of FACS buffer (PBS, 2% FBS). Samples were stored on ice until stained or used for analysis.

2.4.3 Bone Marrow Isolation

Bone marrow was collected by flushing two femurs with 2mL of FACS buffer using a 1mL syringe (BD) and 23-gauge needle (BD). The bones were flushed until the bone turned pale (usually flushed 3-4 times). Next, the bone marrow suspension was filtered into a 5mL round-bottom tube (Falcon) through a 40μm cell strainer cap (Falcon). For cell counting 100μL of the sample was counted on a Sysmex KX21 haematological analyser. The remaining cells were stored on ice until required for staining. Once analysis was completed and depending upon the genotype, 20 million whole bone marrow cells were cryopreserved in 1mL freezing medium (FBS with 10% DMSO). When limited number of cells were left after FACS analysis, 10 million whole bone marrow cells were cryopreserved in 1mL freezing medium in cryogenic tubes (Nunc, ThermoScientific). The tubes were placed in a Mr. Frosty^{TM} freezing container in a -80^{o}C freezer for 24 hours before transferring into storage boxes. Cells were subsequently transferred into liquid nitrogen for long-term storage. Concurrently, 5 million whole bone marrow cells were kept as a cell pellet in a 1.5mL microcentrifuge tube for future gDNA isolation if required. For morphology analysis, a bone marrow cytospin slide was prepared. A total of 1x10^5 whole bone marrow cells in 100μL PBS were spun onto charged microscope slides using cytology funnels (BMP Medical) in a CytospinTM4 Cytocentrifuge (Thermofisher). The cells were centrifuged at 800rpm for

7 mins and then air-dried, before fixing in 100% methanol for 10 minutes. The cytospin slides were allowed to air-dry and stored at RT.

2.4.4 May-Grünwald Giemsa Staining of the blood films and bone marrow cytospins

Fixed blood films and bone marrow cytospins were stained in undiluted May-Grünwald solution (Amber Scientific) for 5 minutes. Then, the slides were washed briefly in tap water before staining in 3% Giemsa solution (Amber Scientific) diluted in tap water for 20 minutes. The slides were washed again in tap water and the intensity of the staining assessed. If the staining was unsatisfactory, the slide was re-stained in 3% Giemsa solution for 20 minutes, with a visual check after 10 minutes. If a good/sufficient intensity was achieved, the slides were air dried before mounting with DPX mounting media (Sigma-Aldrich) and a coverslip. Images were captured on a Leica DMRB fluorescence microscope at 20x or 40x magnification and processed with cellSens software (Olympus).

2.4.5 Spleen and Thymus Isolation

Spleen and thymus were collected during autopsy. Both organs were weighed on an analytical balance (Shimadzu) before being crushed through a 40μm cell strainer using the back of a 5mL syringe plunger. The spleen was crushed into 5mL FACS buffer while the thymus was crushed into 2mL FACS buffer. The spleen from diseased animals was photographed using iPhone X before creating the single cell suspension. 100μL of cell suspension from the spleen or thymus was counted on a Sysmex KX21 haematological analyser. The remaining cells were stored on ice until required for staining or analysis.

2.4.6 Flow Cytometry and Fluorescent Activated Cell Sorting

The antibodies and their conjugates, dilutions, catalogue number and suppliers are listed in Table 2.7 (Liddicoat et al., 2015; Singbrant et al., 2014; Smeets et al., 2018). Antibody cocktails were freshly prepared in FACS buffer at the 2x concentration determined previously by antibody titration. For staining, 30μL of cell suspension (around 0.5-1 million cells) was added to 30μL of 2x antibody cocktail and mixed well. The sample was incubated on ice for 30 minutes then washed with 150μL FACS buffer. For biotinylated antibodies, the cells were stained as described for the primary antibody, then washed and incubated with 30μL of streptavidin-conjugated Brilliant Violet-605 (Biolegend) for another 30 minutes. Next, the cells were washed with 150μL FACS buffer before resuspending in 100μL FACS buffer. Samples were analysed on a BD LSRII Fortessa. Results were analysed with FlowJo software version 10.4.2 (Treestar). Different cell populations were identified based on immunophenotyping and gated. The absolute number of a cell population was then calculated based on the percentage of this population (as determined by gating) multiplied by the cellularity obtained by Sysmex counting.

To determine the activation of *LysM*-Cre in various cell populations, the following haematopoietic fractions were purified from bone marrow of *LysM*-Cre *R26*eYFP *Srsf2*$^{P95H/+}$ mice without lineage depletion according to the immunophenotypes: CD2$^-$CD3$^-$CD4$^-$CD5$^-$CD8a$^-$CD11b$^-$Gr1$^-$B220$^-$Ter119$^-$ (lineage$^-$) c-Kit$^+$ Sca-1$^-$ (LKS$^-$) CD34$^+$FcyRII$^-$, CMP; LKS$^-$CD34$^+$FcyRII$^+$, GMP; Gr1$^+$CD11b$^+$, granulocytes; B220$^+$CD11b$^-$, B lymphocytes. For bulk population RNA-seq, the following haematopoietic fraction was purified from ficolled bone marrow (refer to Chapter 2.10.3) of h*Scl*-CreER^T *R26*eYFP *Srsf2*$^{P95H/+}$ and *Srsf2*$^{+/+}$ mice without lineage depletion

according to the immunophenotypes: lineage$^-$c-Kit$^+$eYFP$^+$ (eYFP positive LK$^+$ cells). For RT-qPCR validation of differentially expressed genes, the following haematopoietic fraction was purified from ficolled bone marrow of non-competitive transplantation recipients of *R26*-Cre *Srsf2*$^{P95H/+}$ and *Srsf2*$^{+/+}$ mice without lineage depletion according to the immunophenotypes: lineage$^-$c-Kit$^+$CD45.2$^+$ (CD45.2$^+$ LK$^+$ cells). For single-cell RNA sequencing, the following haematopoietic fractions were purified from ficolled bone marrow of *R26*-Cre *Srsf2*$^{P95H/+}$ and *Srsf2*$^{+/+}$ mice without lineage depletion according to the immunophenotypes: LKS$^+$CD34$^-$CD135$^-$ (LT-HSC); LKS$^+$CD34$^+$CD135$^{+/-}$ (ST-HSC/MPP). For *in vitro* drug assays, the following haematopoietic fractions were purified from ficolled bone marrow of *R26*-Cre *Srsf2*$^{P95H/+}$ and *Srsf2*$^{+/+}$ mice without lineage depletion according to the immunophenotypes: LKS$^+$. Cells were sorted with either BD FACSAria or BDInflux (BD Biosciences).

2.5 Bone Marrow Transplantation

All transplants were performed into lethally irradiated (2x 5Gy split dose 3 hours apart; total of 10Gy, GammaCell40 irradiator) recipients (B6.SJL-Ptprc $^{aPep3b/BoyJArc}$; CD45.1; purchased from ARC, Perth) via intravenous injections. All recipients received antibiotics (Baytril) in the drinking water for 3 weeks after transplant. A total of 3 recipients/cohort were transplanted. For non-competitive transplantation, a total of 2x10^6 unfractionated whole bone marrow cells were collected from donor mice (CD45.2) at 20 weeks post tamoxifen cessation. The bone marrow cells were injected into lethally irradiated recipients (CD45.1). For competitive transplantation, a total of 1x10^6 unfractionated whole bone marrow cells from donor mice (CD45.2, 20 weeks post tamoxifen cessation) were transplanted with 1x10^6 whole bone marrow competitor cells (CD45.1/CD45.2)

Table 2.7: List of FACS antibodies used in the experiments

Antibody (clone)	Conjugate	Dilution	Catalogue #	Supplier
B220	PerCP-Cy5.5	1:100	45-0452-80	eBioscience
B220	APC-eFlour 780	1:200	47-0452-82	eBioscience
B220	biotin	1:400	13-0452-85	eBioscience
B220/CD45R (RA3-6B2)	APC	1:200	17-0452-83	Life Technologies Australia Pty Ltd/Thermo Fisher
c-Kit/CD117 (2B8)	APC-eFluor-780	1:200	47-1171-82	Life Technologies Australia Pty Ltd/Thermo Fisher
CD105 (MJ7/18)	PE-Cy7	1:400	120409	Biolegend
CD11b	biotin	1:400	13-0112-85	eBioscience
CD11b/Mac1	PE-Cy7	1:100	25-0112-82	eBioscience
CD11b/Mac1	eFluor-450	1:200	48-0112-82	Life Technologies Australia Pty Ltd/Thermo Fisher
CD11b/Mac1 (M1/70)	PE	1:400	12-0112-83	Life Technologies Australia Pty Ltd/Thermo Fisher
CD135/Flk2/Flt3 (A2F10)	PE	1:100	12-1351-82	Life Technologies Australia Pty Ltd/Thermo Fisher
CD150 (TC15-12F12.2)	PE	1:100	115904	Biolegend
CD19 (1D3)	PerCP-Cy5.5	1:150	45-0193-82	Life Technologies Australia Pty Ltd/Thermo Fisher
CD2	biotin	1:400	13-0021-85	eBioscience
CD25 (PC61.5)	PE	1:400	12-0251-83	Life Technologies Australia Pty Ltd/Thermo Fisher
CD34 (RAM34)	eFluor-660	1:25	50-0341-82	Life Technologies Australia Pty Ltd/Thermo Fisher
CD3e	biotin	1:400	13-0031-85	eBioscience
CD4	biotin	1:400	13-0041-85	eBioscience
CD4 (RM4-5)	eFluor-450	1:400	48-0042-82	Life Technologies Australia Pty Ltd/Thermo Fisher
CD41 (MWReg30)	eFluor-450	1:200	48-0411-82	Life Technologies Australia Pty Ltd/Thermo Fisher
CD43 (S7)	PE	1:200	553271	BD Pharmingen

CD44	PE-Cy7	1:400	25-0441-82	eBioscience
CD44 (IM7)	APC	1:400	17-0441-83	Life Technologies Australia Pty Ltd/Thermo Fisher
CD45.1	PE	1:100	12-0453-82	eBioscience
CD45.2	APC	1:100	17-0454-82	BD Pharmingen
CD45RB	biotin	1:400	553098	BD Pharmingen
CD5a	biotin	1:400	13-0051-85	eBioscience
CD71 (RI7 217.1.4)	APC	1:400	17-0711-82	Life Technologies Australia Pty Ltd/Thermo Fisher
CD8a	eFluor-450	1:100	48-0081-82	eBioscience
CD8a	biotin	1:400	13-0081-85	eBioscience
CD8a/Ly-2 (53-6.7)	APC-eFluor 780	1:400	47-0081-82	Life Technologies Australia Pty Ltd/Thermo Fisher
F4/80 (BM8.1)	APC	1:250	20-4801-U100	Tonbo Biosciences
CD16/32 (FcγRII, 93)	PerCP-Cy5.5	1:300	45-0161-82	Life Technologies Australia Pty Ltd/Thermo Fisher
CD16/32 (FcγRII, 93)	eFluor-450	1:400	48-0161-82	eBioscience
Gr-1	APC-eFlour 780	1:600	47-5931-82	eBioscience
Gr-1	biotin	1:400	13-5931-85	eBioscience
Gr-1/Ly6G (RB6-8C5)	PE-Cy7	1:1000	25-5931-82	Life Technologies Australia Pty Ltd/Thermo Fisher
IgM (II/41)	Biotin	1:400	13-5790-82	Life Technologies Australia Pty Ltd/Thermo Fisher
Ly6C	PE-Cy7	1:400	25-5932-80	Life Technologies Australia Pty Ltd/Thermo Fisher
Ly6G	PE-ef610	1:400	61-9668-80	eBioscience
Rat anti-mouse IgG1-APC	APC	1:100	550874	BD Pharmingen
Sca-1	PE-Cy7	1:100	25-5981-82	eBioscience
Sca-1 (D7)	APC	1:100	17-5981-82	Life Technologies Australia Pty Ltd/Thermo Fisher
Streptavidin	BV 605	1:400	563260	BD Pharmingen

TCRβ (H57-597)	PE	1:400	12-5961-83	Life Technologies Australia Pty Ltd/Thermo Fisher
Ter119	biotin	1:400	13-5921-85	eBioscience
Ter119	PE	1:400	12-5921-83	Life Technologies Australia Pty Ltd/Thermo Fisher

into lethally irradiated recipients (CD45.1). Peripheral blood was taken at 4, 12, 20 and 52 weeks post transplantation to monitor multi-lineage donor reconstitution. The bone marrow, spleen and thymus of recipients were analysed at 52 weeks post transplantation. The experiments were performed in duplicate.

2.6 RNA Extraction and cDNA Synthesis

For FACS sorted cell populations, total RNA was extracted with TRIzol (Thermo Fisher Scientific) and DirectZol Microprep kits (Zymo Research) as per manufacturers' instructions. For Cas9 Hoxb8 immortalised cell lines, RNA was extracted with the ISOLATE II RNA mini kit (Bioline) as per manufacturers' instructions. cDNA was synthesised from RNA using a Tetro cDNA synthesis kit (Bioline) as per manufacturer's instructions. An equivalent amount of RNA was used for cDNA synthesis of all samples.

2.7 RT-qPCR

RT-qPCR was performed in duplicates using Brilliant II SYBR Green QPCR Master Mix (Agilent Technologies) as per manufacturer's instructions on the Stratagene Mx3000P qPCR machine. Briefly, a master mix without cDNA was made by combining 2x Brilliant II SYBR Green QPCR Master Mix, pooled oligonucleotides (5μM) from IDT (Table 2.8) and

nuclease-free water. The master mix was then mixed thoroughly and aliquoted into designated wells. Then the diluted cDNA was added. The total volume per well was 20μL. The plate was then run on the program described in Table 2.10. SYBR green fluorescence was measured during the annealing/extension step. A dissociation curve was performed after RT-qPCR to rule out primer dimer formation. Data was then processed using MxPro software (Agilent Technologies)

For exon inclusion/exclusion analysis, isoform specific oligonucleotides were used as described previously (Kon et al., 2018). Briefly, oligo (dT) cDNA was generated from splenocytes of tamoxifen treated h*Scl*-CreER^{T} *R26*eYFP *Srsf2*$^{P95H/+}$ and h*Scl*-CreER^{T} *R26*eYFP *Srsf2*$^{+/+}$ mice and tamoxifen treated Hoxb8 immortalised *R26*-CreER^{T2} *Srsf2*$^{+/+}$ and *R26*-CreER^{T2} *Srsf2*$^{P95H/+}$ cell lines. The exon inclusion/exclusion levels were then measured using RT-qPCR. Primer sequences are included in Table 2.9.

2.8 Bulk and Single-Cell RNA-Sequencing

For *Srsf2*$^{P95H/+}$ RNA-seq, total RNA was extracted from FACS sorted haematopoietic populations as described above. For Hoxb8 immortalised cell lines, total RNA was extracted from cell pellets stored in TRIzol. Sequins were added to the total RNA prior to ribosome RNA (rRNA) depletion to provide an internal control across samples (Hardwick et al., 2016). To remove rRNA, the total RNA was processed through hybridisation, RNAse H reaction and DNase digestion on a Biometra T3000 thermocycler. Briefly, 1μg total RNA was mixed with 0.25μL of 1mM EDTA (pH 7.4), 1μL of 5x RNaseH(-)Mg buffer (500mM Tris pH 7.4, 1M NaCl), 0.8μL of rRNA pooled oligo (100μM) (Adiconis et al., 2013; Morlan

Table 2.8: RT-qPCR oligonucleotide list

Gene	Oligonucleotide sequence (5'-3')	Direction	Product Size (bp)
Lck	GACGATCTCGGGGATCATGG	Forward	122
	GCGAGATCTTGCTGTCCAGT	Reverse	
Ly6d	CAAAACCGTCACCTCAGTGGAG	Forward	153
	AGCCTCTCGTTGCATAGGTC	Reverse	
Blk	GGCGAAGTCTGGATGGGTTA	Forward	145
	CGTAGAGACGAACCAGCCTC	Reverse	
Mcl1	AGAGGCTGGGATGGGTTTGT	Forward	145
	CCCTATTGCACTCACAAGGC	Reverse	
Bcl2l11 (Bim)	GTTCGATCGGCGCAACAC	Forward	189
	CTCCTGTCTTGCGGTTCTGT	Reverse	
Socs3	TAGACTTCACGGCTGCCAAC	Forward	122
	CGGGGAGCTAGTCCCGAA	Reverse	
Dach2	GTGGGAGGGTTGCACACAG	Forward	189
	CCTTGAACTGGCGTTGGTG	Reverse	
Ncam1	CCATCATGGGCCTGAAACCT	Forward	123
	CACTGGGTTCCCGGACTG	Reverse	
Ifih1 (Mda5)	TCACTGATCTGCCCTCTCCT	Forward	129
	CCTTCTCGAAGCAAGTGTCC	Reverse	

Table 2.9: Oligonucleotide list for exon exclusion RT-qPCR

Gene	Oligonucleotide sequence (5'-3')	Direction
Hnrnpa2b1 exon 8	GACCTGGAGGTGGCAATTTT	Inclusion forward
	TCCTCCTCCATAGTTGTCATAACC	Inclusion reverse
	GAGGAGGACCTGGAGGAAAT	Exclusion forward
	ACCCCCAAAGTTTCCACTCT-	Exclusion reverse
Trp53bp1 exon 5	CAAGCAGTGCTCTGGCAGT	Inclusion forward
	GCAGATGCTCTTGGTTGACA	Inclusion reverse
	CCAAACAGGACAAGCAGATTC	Exclusion forward
	TCGCATCCACATCAGATAGC	Exclusion reverse
Usp4 exon 20	TGGAGTTCCCAGTCAGAGCT	Inclusion forward
	CCATTTCCCGTTCAGTCTGT	Inclusion reverse
	TGTCCCACCTGTAAGAAGCA	Exclusion forward
	TGCAGTGTCTGACTGGGAAC	Exclusion reverse
Ezh2 Murine counterparts of human poison exon	CAGCATCTGCCACTCCTAC	Inclusion forward
	AGAGCAGCAGCAAACTCCTTA	Inclusion reverse
	AGCATCTGGAGGGAGCTA	Exclusion forward
	GGGTCTGCTACTGTTATTC	Exclusion reverse
Ezh2 Reported murine cassette exon	AGGACGGCTCCTCTAACCAT	Inclusion forward
	TTGGTGTTGCATTGTGCTTT	Inclusion reverse
	CAACTGAAAAAGGGTCAAAACC	Exclusion forward
	TTTTTAGAGCCCCGCTGAAT	Exclusion reverse

Table 2.10: RT-qPCR cycling conditions

Temperature	Duration	Cycle Number
95°C	3 minutes	x1
95°C	20 seconds	x40
60°C	1 minute	

et al., 2012) and made up with nuclease-free water to 5μL (Table 2.11). The hybridisation reaction was run on the cycling program described in Table 2.12. Two samples were processed in parallel to measure the rRNA depletion later. After hybridisation, RNA samples were mixed with 1μL 5x RNaseH(-)Mg buffer, 0.2μL 1M $MgCl_2$, 0.1μL murine RNase Inhibitor (40U/μL, Bioline), 2μL Hybridase (Thermostable RNase H; 5U/μL, Illlumina) and 1.7μL nuclease-free water (Table 2.13). For the two additional samples, one was mixed as the RNase reaction described above, while no Hybridase was added to the other sample (1μL 5x RNaseH(-)Mg buffer, 0.2μL 1M $MgCl_2$, 0.1μL murine RNase Inhibitor (40U/μL) and 3.7μL nuclease-free water). This Hybridase-free sample served as a rRNA non-depleted control and the other sample as the rRNA depleted control. The samples were incubated with RNase H at 45^{o}C for 30 minutes. After the RNase H reaction, the samples were mixed with 7μL of 10x TURBO DNase buffer (Life Technologies), 4μL of TURBO DNase (2U/μL, Life Technologies) and 59μL nuclease-free water (Table 2.14) and then incubated at 37 ^{o}C for 30 minutes. After rRNA depletion, the products were purified with an RNA Clean & Concentrator Kit. Once purified, the rRNA depletion level was measured by RT-qPCR using the rRNA depleted and non-depleted control. The *Ppia* level was measured as an internal control (Forward primer: 5'-GTCAACCCCACCGTGTTCTT-3', reverse primer: 5'-CTGCTGTCTTTGGAACTTTG-3'). A 20-cycle difference between the rRNA depleted and non-depleted was considered as an efficient rRNA depletion. The purified, rRNA depleted RNA was then subjected to indexing and library preparation using the Kapa Stranded RNA-seq Library Preparation Kit (Kapa Biosystems/Roche) and sequenced with 150bp paired end reads on the Illumina platform by Novogene (China or Singapore). The sequence of the forward indexing primer was: 5'-

AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGAC-3′. The sequences of the various sample specific indexed primers used are listed in Table 2.15. The *Srsf2*$^{P95H/+}$ RNA-seq library was prepared by Mr. Scott Taylor.

For *Srsf2*$^{P95H/+}$ single-cell RNA-seq (scRNA-seq), libraries were prepared from the FACS sorted haematopoietic fractions using the ChromiumTM Singe Cell Library Kit V2: Chip Kit and Gel Bead & Multiplex Kit (equipment and service provided by the Australian Genome Research Facility (AGRF)). The resulting libraries were pooled and sequenced using Illumina chemistry with 100bp paired end reads. The libraries for scRNA-seq were generated and sequenced by AGRF.

Table 2.11: rRNA depletion-Hybridisation mix

Hybridisation mix	Volume (µL)
Input RNA	1µg
1mM EDTA (pH 7.4)	0.25
5x RNaseH(-) Mg buffer	1
rRNA oligo pool (100uM)	0.8
Nuclease-free water	Up to 3µL
Total volume	**5**

Table 2.12: rRNA depletion-Hybridisation cycling conditions

Temperature	Duration	Cycle Number
95°C	2 minutes	x1
95°C to 45°C at 0.1 °C/second		x1
45˚C	Hold	

Table 2.13: rRNA depletion-RNase H reaction mix

Hybridisation mix	Volume (µL)
1M $MgCl_2$	0.2
5x RNaseH(-) Mg buffer	1
Murine RNase Inhibitor (40U/µL)	0.1
Hybridase (5U/µL)	2
Nuclease-free water	1.7
Total volume	**5**

Table 2.14: rRNA depletion-DNase digestion mix

Hybridisation mix	Volume (µL)
10x TURBO DNase buffer	7
TURBO DNase (2U/µL)	4
Nuclease-free water	59
Total volume	**70**

Table 2.15: Sequences of Multiplexing PCR primers used to generate RNA-seq libraries

Name	Sequence to include in P7 primer, 5'-3'	Index read (rev comp sequence)	Full sequence
PCR Primer Index 1	CGTGAT	ATCACG	CAAGCAGAAGACGGCATACGAGAT CGTGATGTGACTGGAGTTCAGACG TG
PCR Primer Index 2	ACATCG	CGATGT	CAAGCAGAAGACGGCATACGAGAT ACATCGGTGACTGGAGTTCAGACG TG
PCR Primer Index 3	GCCTAA	TTAGGC	CAAGCAGAAGACGGCATACGAGAT GCCTAAGTGACTGGAGTTCAGACG TG
PCR Primer Index 4	TGGTCA	TGACCA	CAAGCAGAAGACGGCATACGAGAT TGGTCAGTGACTGGAGTTCAGACG TG
PCR Primer Index 5	CACTGT	ACAGTG	CAAGCAGAAGACGGCATACGAGAT CACTGTGTGACTGGAGTTCAGACG TG
PCR Primer Index 6	ATTGGC	GCCAAT	CAAGCAGAAGACGGCATACGAGAT ATTGGCGTGACTGGAGTTCAGACG TG
PCR Primer Index 7	GATCTG	CAGATC	CAAGCAGAAGACGGCATACGAGAT GATCTGGTGACTGGAGTTCAGACG TG
PCR Primer Index 8	TCAAGT	ACTTGA	CAAGCAGAAGACGGCATACGAGAT TCAAGTGTGACTGGAGTTCAGACG TG
PCR Primer Index 9	CTGATC	GATCAG	CAAGCAGAAGACGGCATACGAGAT CTGATCGTGACTGGAGTTCAGACG TG
PCR Primer Index 10	AAGCTA	TAGCTT	CAAGCAGAAGACGGCATACGAGAT AAGCTAGTGACTGGAGTTCAGACG TG
PCR Primer Index 11	GTAGCC	GGCTAC	CAAGCAGAAGACGGCATACGAGAT GTAGCCGTGACTGGAGTTCAGACG TG
PCR Primer Index 12	TACAAG	CTTGTA	CAAGCAGAAGACGGCATACGAGAT TACAAGGTGACTGGAGTTCAGACG TG

2.9 Bioinformatic Analysis

For *Srsf2*$^{P95H/+}$ RNA-seq and Hoxb8 cell line RNA-seq, remaining ribosomal reads were first removed using BBmap (bbsplit, v37.5). Reads were then trimmed for adaptor sequences using Trim_galore (v0.4.4). Reads were then aligned to the mm10 mouse reference genome (Annotation: gencode.mm10.vM14.annotation.gtf) using STAR (STAR_2.5.3a). Duplicate reads were marked using Picard (v2.2.2). Read counting was performed using featureCounts (subread-1.5.1) ignoring duplicate reads. Filtering and differential expression was performed using degust (http://degust.erc.monash.edu/). Data was filtered to keep only genes with: a) Min gene read count: 5, and b) min gene CPM: 1 (in at least 2 samples) resulting in 15,466 expressed genes. Differential expression was performed using voom with sample weighting (Law et al., 2014). The bioinformatic analysis for RNA-seq was performed by Dr. Alistair Chalk.

For *Srsf2*$^{P95H/+}$ scRNA-seq, sequencing data was pre-processed on CellRanger software (10x Genomics) to perform alignment, filtering, barcode counting and Unique Molecular Identifiers (UMI) counting. Then, cells were filtered by their library size, number of genes expressed, and mitochondrial percentage. The DoubletDetection method (https://github.com/JonathanShor/DoubletDetection) was used to predict and remove potential doublets from the dataset. Following this, gene expression was normalised using sctransform and visualised with Uniform Manifold Approximation and Projection (UMAP). Marker genes reported previously (Cheng et al., 2019) were used to annotate the cell clusters. Pseudotime analysis was performed using monocle2 (Trapnell et al., 2014). The bioinformatic analysis for scRNA-seq was performed by Ms Ruqian Lyu and Dr. Davis McCarthy.

2.10 Cell Culture

2.10.1 General Cell Culture Technique and Handling

All cell culture experiments were performed in a Gelaire class II biological safety cabinet. Unless otherwise stated, cell culture medium and other solutions were pre-warmed in a WB7 water bath at 37°C before use. The safety cabinet, media and equipment used were sprayed with 70% ethanol to reduce the risk of contamination. Cells were incubated in a CO_2 HeraCell 240i incubator at 5% CO_2 and 37°C

Cell passaging and media replacement were performed every 2-3 days. For adherent cells, the supernatant was discarded, and cells were washed once with PBS before adding trypsin (0.025% in PBS, 1mM EDTA). The cells were incubated for 3-5 minutes at 37°C, then PBS was added to dilute the trypsin and collect the cells. The cell suspension was centrifuged at 400g for 5 mins to collect the cells and then resuspended in 1mL of tissue culture medium. The cells were counted and re-seeded at the desired cell density in pre-warmed medium. For suspension cells, these were mixed thoroughly in the tissue culture flasks/plates before cell counting and re-seeding at the desired cell density in pre-warmed medium. Cell counting was performed by mixing 10μL of cells with 10μL trypan blue solution (ThermoFisher Scientific) in a round-bottom 96 well non-tissue culture treated plate (Falcon), then 10μL of mixture was transferred into the chamber of a cell counting slide (BioTools) and counted on a Countess II cell counter (Invitrogen).

To cryopreserve cells, approximately $5x10^6$ cells were collected and pelted by centrifuging at 400g for 5 minutes. The cell pellet was then resuspended in 1mL of freezing medium (FBS with 10% DMSO) and stored in cryogenic vials (Nunc, ThermoScientific). The tubes were put in

a Mr. Frosty^{TM} freezing container in the -80°C freezer for 24 hours before being transferred into storage boxes. Cells were subsequently transferred into liquid nitrogen for long-term storage. To thaw cells, the cryogenic vial was placed directly into a 37°C water bath until all frozen content was thawed. The content was transferred into 10mL PBS in a 12mL tube and the cells were pelleted by centrifugation at 400g for 5 minutes. The supernatant was discarded, and the cells were seeded at a density and in medium appropriate for each cell line/type.

2.10.2 Generation of Lentivirus and Retrovirus

For virus production, 2.5x10^6 HEK293T cells were seeded in a 10cm tissue culture plate (Falcon) in 10mL complete DMEM and incubated overnight. When the cells reached 70-80% confluence, and transfection was carried out using the calcium-phosphate precipitation method. All solutions used in this procedure were at room-temperature. For lentiviral production, 10μg of Cas9-blasticidin plasmid DNA (pDNA) (Addgene, #52962), 3.8μg pVSV-G pDNA and 6.3μg psPax2 pDNA (Addgene, #12260) were added to 61μL 2M $CaCl_2$ and topped up to 500μL with sterile nuclease-free water (Table 2.16). For retroviral production, 10μg of MSCV-Hoxb8 pDNA and 10μg of EcoPac pDNA were added to 61μL 2M $CaCl_2$ and topped up to 500μL with sterile nuclease-free water (Table 2.17) in a 15mL tube (Falcon). For both virus types, 500μL 2x HEPES buffered saline (2xHBS) was then added dropwise into the plasmid mixture while vortexing (1,800rpm). The calcium phosphate-DNA precipitates were then added dropwise onto the HEK239T culture and incubated overnight (12-16hrs). The precipitate was checked the next day. The medium was then carefully removed and replaced with fresh medium. The cells were incubated for a further 24-28hrs and the virus containing supernatant was collected at 48hrs and 72hrs post transfection. The virus containing supernatant was centrifuged

at 400g for 5 minutes to remove any cell debris and stored as 1mL aliquots at -80ºC.

Table 2.16: Lentiviral transfection mix

Lentiviral mix	Amount/Volume
Cas9-blasticidin plasmid DNA (pDNA) (Addgene, #52962)	10µg
pVSV-G pDNA	3.8µg
psPax2 (Addgene, #12260)	6.3µg
2M $CaCl_2$	61µL
Sterile nuclease-free water	Up to 439µL
Total volume	**500µL**

Table 2.17: Retroviral transfection mix

Retroviral mix	Amount/Volume
MSCV-Hoxb8 pDNA	10µg
EcoPac pDNA	10µg
2M $CaCl_2$	61µL
Sterile nuclease-free water	Up to 439µL
Total volume	**500µL**

2.10.3 Generation of Hoxb8 *R26*-CreER^{T2} *Srsf2*$^{P95H/+}$ and *Srsf2*$^{+/+}$ Cell lines

Bone marrow cells were collected from *R26*-CreER^{T2} *Srsf2*$^{P95H/+}$ and *R26*-CreER^{T2} *Srsf2*$^{+/+}$ mice by crushing both femurs, tibiae and iliac crests with a mortar and pestle. A total of 40mL of ice-cold FACS buffer was added into the mortar to flush the bone marrow cells by gentle pipetting. The cell suspension was then filtered through a 40μm cell strainer and pelleted. The supernatant was discarded, and the cells were resuspended in 4mL FACS buffer. Using a 15mL falcon tube, 3mL of Ficoll-Pague*TM* PLUS media (GE healthcare, density 1.078 g/mL) was added and then

the 4mL bone marrow cell suspension was carefully overlaid on top. The cells/ficoll were centrifuged at 400g for 10 minutes without applying the brake setting on the centrifuge. After centrifugation, the mononuclear cells were carefully aspirated from the interphase of the ficoll and PBS and the cells were washed with 10mL PBS. 100μL from this cell suspension was used to count on Sysmex KX21 haematological analyser. After washing, the cells were resuspended in 1mL PBS.

The ficoll-purified bone marrow cells (around 1x10^6/mL) were stimulated for 48 hours in complete IMDM (20% FBS, 1% Penicillin/Streptomycin, 1% glutamine) supplemented with recombinant mouse stem cell factor (rmSCF, 50ng/mL), recombinant mouse interleukin 3 (rmIL-3, 10ng/mL) and recombinant human interleukin 6 (rhIL-6, 10ng/mL). After stimulation, 1x10_6 cells in 1mL of complete IMDM were mixed with 1mL of Hoxb8 retrovirus containing medium. Additional cytokines (rmSCF, rmIL-3 and rmIL-6) were added, as well as polybrene (8μg/mL). The cells were spin-infected for 90 minutes at 1,100g. The cells were then incubated for 18-24 hours and then the medium was replaced with fresh IMDM with cytokines. After another 24 hours, the suspension cells were passaged into complete IMDM (10% FBS, 1% Penicillin/Streptomycin, 1% glutamine) with 1% granulocyte-macrophage colony-stimulating factor (GM-CSF) conditioned media (from BHK-HM5 cell conditioned media). After 1 to 2 passages, the cells were cultured in IMDM containing 10% FBS with 1% GM-CSF only. The cells were fully immortalised after four weeks.

2.11 CRISPR Screen

2.11.1 Generation and Validation of Cas9-Blasticidin Hoxb8 *Srsf2*$^{P95H/+}$ And *Srsf2*$^{+/+}$ Cell Lines

To generate Cas9-expressing Hoxb8 GM-CSF *R26*-CreER^{T2} *Srsf2*$^{P95H/+}$ and *R26*-CreER^{T2} *Srsf2*$^{+/+}$ cell lines, the Hoxb8 cell lines were infected with Cas9-blasticidin lentivirus by spin-infection at 1,100g for 90 minutes. The infected cells were then cultured in IMDM-Cas9 media (IMDM, 10% FBS, 1% GM-CSF and 3μg/mL blasticidin) for two weeks to select for a Cas9-expressing population. The Cas9-blasticidin *R26*-CreER^{T2} *Srsf2*$^{P95H/+}$ and *R26*-CreER^{T2} *Srsf2*$^{+/+}$ cell lines were maintained in IMDM-Cas9 media to maintain positive selection pressure.

To check the editing activity of the Cas9-blasticidin Hoxb8 *R26*-CreER^{T2} *Srsf2*$^{P95H/+}$ and *R26*-CreER^{T2} *Srsf2*$^{+/+}$ cell lines, stable Cas9 cells were infected with pLentiGuide (provided by A/Prof Kaylene Simpson and Dr Iva Nikolic, Victorian Cancer and Functional Genomics (VCFG); originally sourced from Addgene) expressing sgRNA targeting the *Ifih1* gene (5'-CACCGGTTTGACATAGCGCGCGGCT-3'). Puromycin (0.5μg/mL) was added to the IMDM-Cas9 media after 48 hours. Six-days later, the cells were collected, and RNA extracted to determine the expression of *Ifih1* via RT-qPCR (Table 2.8 and 2.9).

To determine the proliferation kinetics of the Cas9-blasticidin Hoxb8 *R26*-CreER^{T2} *Srsf2*$^{P95H/+}$ and *R26*-CreER^{T2} *Srsf2*$^{+/+}$ cell lines, the cells were seeded at 1x10^5 cells/mL in IMDM-Cas9 media. On day 0, 400nM 4-hydroxy tamoxifen was added. After 4 days of tamoxifen treatment, the cells were taken off tamoxifen. The cells were then counted and passaged at 1x10^5 cells/mL every 2 to 3 days for up to 14 days.

2.11.2 Titration of Mouse Brie Library

Aliquots of concentrated Mouse Brie sgRNA library (Doench et al., 2016) were provided by VCFG. To determine the optimal amount of library to achieve a Multiplicity of Infection (MOI) of 0.3, 1.25 x 10^6 stable Cas9-blasticidin Hoxb8 cells were seeded in 1mL of IMDM-Cas9 media in a 12-well plate (Falcon). The cells were spin-infected with increasing amounts of concentrated virus (0, 2.5, 5, 7.5, 10μL per mL). 1mL of IMDM-Cas9 media was added after 4-6 hours, which was then replaced with 2mL of fresh IMDM-Cas9 media after 24 hours. Puromycin (0.5μg/mL) was added to the culture 24 hours after infection and cell viability was assessed 72 hours later with the Countess II cell counter using trypan blue staining. The viral concentration that resulted in $\sim$30% of cell survival (MOI=0.3) at 72hrs was the concentration used for the screen.

2.11.3 Pooled CRISPR knock-out screen and parallel culture for RNA-Seq

At the start of the CRISPR screen, 130x10^6 Cas9 cells from each of the cell lines (n=3 independently derived *R26*-CreER^{T2} *Srsf2*$^{P95H/+}$ and *R26*-CreER^{T2} *Srsf2*$^{+/+}$ cell lines) were infected with the Brie library. The number of infected cells was calculated by multiplying 78,637 (the number of guides in the library) by 500 (representing the desired number of cells for each guide at the start of the experiment), divided by 0.3 (MOI). The 130x10^6 cells were separated into four Nunc^{TM} cell culture T75 flasks to ensure efficient spin-infection and avoid nutrient deprivation (previously optimised using GFP infection and culture after infection by Mr Scott Taylor). In each spin-infection flask, 35x10^6 cells were mixed with 122μL of concentrated lentivirus containing the Brie Library and 35μL of polybrene (8μg/mL) in 35mL of IMDM-Cas9 media. The infection mix

was then spun at 1,100g for 90 minutes and the cells were re-incubated immediately after the spin-infection. Around 4 hours later, the mixture was transferred into two T175 flasks (Falcon) with the addition of 65mL of IMDM-Cas9 media and cells were re-incubated. After 48 hours, cells were counted and diluted to $8x10^5$ cells/mL before adding puromycin ($0.5\mu g$/mL). After 4 days of puromycin selection, the puromycin and blasticidin was washed off and the cells were cultured in IMDM with 10%FBS, 1% GM-CSF and 400nM 4-hydroxy tamoxifen. This timepoint is labelled day 0 of the screen. The tamoxifen was removed from the cells after 4 days and the cells were then passaged at $2x10^5$/mL every 2 days (Monday and Wednesday) and $1x10^5$/mL for a 3-day interval (Friday) for up to 14 days. A minimum of $40x10^6$ cells was passaged during each split to maintain the initial targeted representation of an average of 500 copies per guide. On day 0, day 4, day 11 and day 18 of the screen, pellets of $40x10^6$ cells were collected for gDNA isolation and library preparation. On day 0, day 2 and day 4 of the screen, pellets of $2x10^6$ cells were collected for DNA recombination analysis.

In parallel to the CRISPR knock-out screen, $1x10^5$ Cas9 cells were treated with 400nM 4-hydroxy tamoxifen in a 6-well plate on day 0 of the screen. On day 4, tamoxifen was removed from the cells and cells were cultured in IMDM-Cas9 media. On day 11 (7 days post tamoxifen cessation), $1x10^6$ cells were collected for RNA extraction and whole transcriptome sequencing.

2.11.4 Genomic DNA Preparation for CRISPR Screen

Genomic DNA was extracted from the cell pellets using the Gentra Puregene kit (Qiagen) according to manufacturer's instructions. Cell pellets were resuspended in 3mL of lysis buffer on the day of collection and

stored at -20°C. To extract gDNA, the lysate was thawed and vortexed at maximum speed for 15 seconds. The rest of the procedure was performed as per manufacturer's instructions. Due to large number of cells collected, the DNA pellet yielded from the extraction process was suspended in 1mL of DNA hydration solution and incubated at 65°C overnight to ensure that the DNA was fully dissolved. Next, the DNA was quantified using the Nanodrop spectrophotometer. To check the integrity of the extracted DNA, 1μL of the extracted DNA (around 550μg) was electrophoresed on a 0.8% agarose gel at 120V for 1 hour. After electrophoresis, the gel was imaged with a VersaDoc imaging system (Bio-Rad).

2.11.5 Library Preparation for Illumina Sequencing

DNA libraries were generated by PCR amplification of the integrated sgRNA constructs with a barcode sequence attached (following the protocol provided by the Broad Institute; https://portals.broadinstitute.org/gpp/public/resources/protocols). In total, 24 samples from four time points (day 0, 4, 11 and 18) and six cell lines (three *R26*-CreER^{T2} *Srsf2*$^{+/+}$ and three *R26*-CreER^{T2} *Srsf2*$^{P95H/+}$) were processed. To ensure full representation of the sgRNAs in the samples, all of the DNA collected from 40x10^6 cells was amplified, which resulted in 58-93 PCR reactions per sample. Each 100μL PCR reaction contained 10μg of gDNA, 10μL 10x ExTaq reaction buffer, 8μL 2.5mM dNTP mixture, 0.75μL ExTaq polymerase, 0.5μL P5 primer mix (100μM; Table 2.18), and, for each of the 24 samples, 10μL of a unique P7 index primer (5μM; Table 2.19) and nuclease-free water to a total volume of 100μL. For each sample, the PCR reactions were prepared in a 96-well PCR plate (Thermo Scientfic) and cycled with the parameters described in Table 2.20. 15μL of the PCR product was then pooled and mixed well. 100μL of the pooled PCR product was mixed with 24μL 5x DNA loading dye (Bioline). Then, 30μL of the mixture was loaded

onto a 1% agarose gel and run for 50 minutes at 120V. The 400bp PCR products were excised from the gel and the DNA was extracted using a PCR & Gel purification Kit (Bioline). The purified pooled library was first quantified with a Nanodrop spectrophotometer and the concentration then confirmed using a LabChip GX Touch Nucleic Acid Analyzer (PerkinElmer). Finally, 18μL of purified pooled library containing at least 2ng/μL of DNA was sent for sequencing on the Illumina platform using 150bp paired end reads by Novogene (China).

Table 2.18: P5 primer sequences

Name	Sequence
P5 0 nt stagger	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTTTGTGGAAAGGACGAAACACCG
P5 1nt stagger	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTCTTGTGGAAAGGACGAAACACCG
P5 2nt stagger	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTGCTTGTGGAAAGGACGAAACACCG
P5 3nt stagger	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTAGCTTGTGGAAAGGACGAAACACCG
P5 4nt stagger	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTCAACTTGTGGAAAGGACGAAACACCG
P5 5nt stagger	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTTGCACCTTGTGGAAAGGACGAAACACCG
P5 7nt stagger	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTACGCAACTTGTGGAAAGGACGAAACACCG
P5 8nt stagger	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTGAAGACCCTTGTGGAAAGGACGAAACACCG

Table 2.19: P7 primer sequences

Name	Sequence to include in P7 primer, 5'-3'	Index read (rev comp sequence)	Full sequence
P7_index01 _lentiGuide	CGGTTCAA	TTGAACCG	CAAGCAGAAGACGGCATACGAGAT CGGTTCAAGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTA TTCTTTCCCCTGCACT*G*T
P7_index02 _lentiGuide	GCTGGATT	AATCCAGC	CAAGCAGAAGACGGCATACGAGAT GCTGGATTGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTA TTCTTTCCCCTGCACT*G*T
P7_index03 _lentiGuide	TAACTCGG	CCGAGTTA	CAAGCAGAAGACGGCATACGAGAT TAACTCGGGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTA TTCTTTCCCCTGCACT*G*T
P7_index04 _lentiGuide	TAACAGTT	AACTGTTA	CAAGCAGAAGACGGCATACGAGAT TAACAGTTGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCTTCTACTAT TCTTTCCCCTGCACT*G*T
P7_index05 _lentiGuide	ATACTCAA	TTGAGTAT	CAAGCAGAAGACGGCATACGAGAT ATACTCAAGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCTTCTACTAT TCTTTCCCCTGCACT*G*T
P7_index06 _lentiGuide	GCTGAGAA	TTCTCAGC	CAAGCAGAAGACGGCATACGAGAT GCTGAGAAGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTA TTCTTTCCCCTGCACT*G*T
P7_index07 _lentiGuide	ATTGGAGG	CCTCCAAT	CAAGCAGAAGACGGCATACGAGAT ATTGGAGGGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTA TTCTTTCCCCTGCACT*G*T
P7_index08 _lentiGuide	TAGTCTAA	TTAGACTA	CAAGCAGAAGACGGCATACGAGAT TAGTCTAAGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCTTCTACTAT TCTTTCCCCTGCACT*G*T
P7_index09 _lentiGuide	CGGTGACC	GGTCACCG	CAAGCAGAAGACGGCATACGAGAT CGGTGACCGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTA TTCTTTCCCCTGCACT*G*T
P7_index10 _lentiGuide	TACAGAGG	CCTCTGTA	CAAGCAGAAGACGGCATACGAGAT TACAGAGGGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTA TTCTTTCCCCTGCACT*G*T
P7_index11 _lentiGuide	ATTGTCAA	TTGACAAT	CAAGCAGAAGACGGCATACGAGAT ATTGTCAAGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCTTCTACTAT TCTTTCCCCTGCACT*G*T
P7_index12 _lentiGuide	TATGTCTT	AAGACATA	CAAGCAGAAGACGGCATACGAGAT TATGTCTTGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCTTCTACTAT TCTTTCCCCTGCACT*G*T

P7_index13 _lentiGuide	ATTGGATT	AATCCAAT	CAAGCAGAAGACGGCATACGAGAT ATTGGATTGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCTTCTACTAT TCTTTCCCCTGCACT*G*T
P7_index14 _lentiGuide	ATACTCGG	CCGAGTAT	CAAGCAGAAGACGGCATACGAGAT ATACTCGGGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTA TTCTTTCCCCTGCACT*G*T
P7_index15 _lentiGuide	TATGAGAA	TTCTCATA	CAAGCAGAAGACGGCATACGAGAT TATGAGAAGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCTTCTACTAT TCTTTCCCCTGCACT*G*T
P7_index16 _lentiGuide	GCACAGTT	AACTGTGC	CAAGCAGAAGACGGCATACGAGAT GCACAGTTGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTA TTCTTTCCCCTGCACT*G*T
P7_index17 _lentiGuide	CGTGGATT	AATCCACG	CAAGCAGAAGACGGCATACGAGAT CGTGGATTGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTA TTCTTTCCCCTGCACT*G*T
P7_index18 _lentiGuide	TAGTAGAA	TTCTACTA	CAAGCAGAAGACGGCATACGAGAT TAGTAGAAGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCTTCTACTAT TCTTTCCCCTGCACT*G*T
P7_index19 _lentiGuide	GCACGATT	AATCGTGC	CAAGCAGAAGACGGCATACGAGAT GCACGATTGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTA TTCTTTCCCCTGCACT*G*T
P7_index20 _lentiGuide	CGGTAGCC	GGCTACCG	CAAGCAGAAGACGGCATACGAGAT CGGTAGCCGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTA TTCTTTCCCCTGCACT*G*T
P7_index21 _lentiGuide	TAGTTCTT	AAGAACTA	CAAGCAGAAGACGGCATACGAGAT TAGTTCTTGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCTTCTACTAT TCTTTCCCCTGCACT*G*T
P7_index22 _lentiGuide	TACAAGTT	AACTTGTA	CAAGCAGAAGACGGCATACGAGAT TACAAGTTGTGACTGGAGTTCAGAC GTGTGCTCTTCCGATCTTCTACTAT TCTTTCCCCTGCACT*G*T
P7_index23 _lentiGuide	ATCACTGG	CCAGTGAT	CAAGCAGAAGACGGCATACGAGAT ATCACTGGGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTA TTCTTTCCCCTGCACT*G*T
P7_index24 _lentiGuide	CGCATCAA	TTGATGCG	CAAGCAGAAGACGGCATACGAGAT CGCATCAAGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTCTACTA TTCTTTCCCCTGCACT*G*T

Table 2.20: CRISPR DNA library thermocycling parameters

Temperature	Duration	Cycle Number
95˚C	Hot plate start	
95˚C	1 minute	x1
95˚C	30 seconds	x28
53°C	30 seconds	
72˚C	30 seconds	
72˚C	10 minutes	x1
4˚C	On Hold	

2.11.6 Bioinformatic Analysis for CRISPR Screen

CRISPR screen analysis was performed using MaGeCK, MaGeck-VISPR and CRISPRBetaBinomial package (CB2) (v 1.3.0) by Dr Alistair Chalk (Li et al., 2014) (Li et al., 2015) (Jeong et al., 2019). Raw reads were mapped to the broadgpp_brie_crispr library with run_sgrna_quant() and read counts were normalised with get_CPM(). Annotation with ENSEMBL was performed with biomaRt (Durinck et al., 2009). To measure differences in sgRNA levels, CB2 combines a beta-binomial model with a modified Student's t-test. Next, to estimate gene-level significance, a Fisher's combined probability test was conducted. The sgRNA distribution, Gini index and heatmap were generated using the ggplot2 package (Wickham, 2016). The Venn diagrams were generated with the VennDiagram package (Chen and Boutros, 2011). The heatmap for negatively selected pathway analysis was generated with Metascape (Zhou et al., 2019).

2.12 ATP-Lite Cell Viability Assay – Endpoint *In Vitro* Drug Assay

Hoxb8 immortalised *R26*-CreER^{T2} *Srsf2*$^{P95H/+}$ and *R26*-CreER^{T2} *Srsf2*$^{+/+}$ cell lines were treated with 400nM 4-hydroxy tamoxifen for four days to activate the Cre-mediated recombination. At 48 hours post tamoxifen cessation, 1,000 cells were plated in 50μL of IMDM with 10% FBS and 1% GM-CSF media in 96-well tissue culture plates, then 50μL of a 2x concentration of drugs was added in triplicates per dose (Dose range: Palbociclib 1nM-10μM, Roscovitine 100nM-50μM, PD0166285 1nM-1μM, Pladienolide B 0.1nM-10nM). A DMSO or water control was included for all analysis. Cells were cultured with drugs for 4 days before assessing the cell viability using the ATP-Lite Luminescence Assay System as per manufacturer's instructions. The luminescence was measured using an Enspire plate reader.

For LKS$^+$ in vitro drug assays, the sorted LKS$^+$ cells (described in 2.4.6) were cultured for 24 hrs in complete IMDM (20% FBS, 1% Penicillin/Streptomycin, 1% glutamine) supplemented with recombinant mouse stem cell factor (rmSCF, 50ng/mL), recombinant mouse interleukin 3 (rmIL-3, 10ng/mL) and recombinant human interleukin 6 (rhIL-6, 10ng/mL). After stimulation, 400nM 4-hydroxy tamoxifen was added. After 4 days of tamoxifen treatment, the cells were taken off tamoxifen. Then, 1,000 cells were plated in 50μL of complete IMDM supplemented with cytokines (50ng/mL rmSCF; 10ng/mL rmIL-3; 10ng/mL rhIL-6) in a 96-well tissure culture plate, then 50μL of a 2x concentration of drugs was added in triplicates per dose (Dose range: Palbociclib 1nM-10μM, Roscovitine 100nM-50μM, PD0166285 1nM-1μM, Pladienolide B 0.1nM-10nM). A DMSO or water control was included for all analysis. Cells were cultured

with drugs for 4 days before assessing the cell viability using the ATP-Lite Luminescence Assay System as per manufacturer's instructions. The luminescence was measured using an Enspire plate reader. The IC50 of was calculated in Prism 8 (Graphpad) using log(inhibitor) vs. normalized response- variable slope.

2.13 Hoxb8 Cell Line Proliferation Assay Continuous *In Vitro* Drug Sensitivity Assay

Hoxb8 immortalised *R26*-CreER^{T2} *Srsf2*$^{P95H/+}$ and *R26*-CreER^{T2} *Srsf2*$^{+/+}$ cell lines were treated with 400nM 4-hydroxy tamoxifen for four days to activate the Cre-mediated recombination. At 48 hours post tamoxifen cessation, 1x10^5 cells were seeded in 5mL IMDM with 10%FBS and 1% GM-CSF media in T-25 flasks (Falcon). Cells were then treated with either Palbociclib (57μM) or Pladienolide B (0.85nM) for 14 days. Cell were counted on the Countess II cell counter and passaged every two- three days. A DMSO (0.1%) control was included as well.

2.14 Statistical Analysis

Statistical significance tests were conducted in Prism software version 8 (GraphPad), Tests used include unpaired two-tailed Student t-tests, ordinary one-way ANOVA and Receiver-operating characteristic (ROC) curves. All data are depicted as mean $\pm$ standard error of the mean. Significance is indicated using the following convention: *P<.05; **P<.01; ***P<.001; ****P<.0001.

Chapter 3

Srsf2$^{P95H/+}$ mutation cannot lead to fully penetrate MDS when activated in myeloid progenitors

3.1 Introduction

Genetically engineered animal models are an important tool for understanding the consequences of gene mutation on normal haematopoiesis. To characterise the effect of *Srsf2*$^{P95H/+}$ mutation *in vivo*, numerous groups have established their own models using different gene-targeting strategies and the LoxP-Cre system, which I reviewed in Chapter 1 of my thesis. One of the advantages of the LoxP-Cre system is that it enables either constitutive or inducible expression of the mutation in a specific cell type. This is important for identification and characterisation of the cell of origin in *Srsf2*$^{P95H/+}$ mutated MDS. Previously, a specific group of MDS stem cells (MDS-SCs) was identified in low-risk MDS patients (Woll et al., 2014). The MDS-SCs propagate *in vivo* and produce MDS-progenitor cells. Further mutation acquisition in MDS progenitors confers self-renewal potential to the progenitor compartment, which enables disease progression towards secondary AML (sAML) (Woll et al., 2014). In acute myeloid leukaemia (AML), both long-term haematopoietic stem cells (LT-HSCs) and granulocyte-monocyte progenitors (GMPs) are able to initiate and sustain experimental AML driven by the oncogenic MLL-AF9 fusion. However, the LT-HSC-origin AML cells are more aggressive and more potent in inducing leukaemia than the GMP-origin AML cells (Chen et al., 2008; Stavropoulou et al., 2016). To investigate if *Srsf2*$^{P95H/+}$ mutation could initiate MDS from a myeloid progenitor population, I took advantage of the *LysM*-Cre model to activate *Srsf2*$^{P95H/+}$ in myeloid cells.

The *LysM*-Cre mice were generated by inserting the cDNA of Cre recombinase behind the promotor region of the *Lysozyme 2* gene, which enables constitutive Cre expression in cells expressing *LysM* (Clausen et al., 1999). *LysM*-Cre is predominantly expressed in cells committed to the myeloid lineage and when cells differentiate into myeloid progenitors. To

enable the tracking of cells where Cre has been expressed, I utilised the *R26*-LSL-eYFP reporter system. Here *EYFP* was inserted into the Rosa26 (Gt(ROSA)26Sor) locus behind a *loxP* flanked STOP cassette. Upon Cre-mediated excision, the stop cassette is deleted, and the yellow fluorescent protein is expressed for the lifetime of that cell and any daughter cells that arise from it. This can be visualized through FACS (Srinivas et al., 2001). Based on eYFP expression, *LysM*-Cre can achieve 50-70% recombination efficiency in granulocytes, monocytes and spleen dendritic cells (Abram et al., 2014; Ye et al., 2003). The eYFP labelling therefore enabled the tracing of *Srsf2*$^{P95H/+}$ mutant cells in unperturbed haematopoiesis from the time of *LysM* expression and concurrent induction of *SRSF2* mutation.

To understand the pathogenesis of MDS, it is important to fully explore the implications of *Srsf2*$^{P95H/+}$ mutation in the myeloid progenitor populations. The *LysM*-Cre model was in part described in the publication reporting the *R26*-CreER^{T2} and h*Scl*-CreER^{T} *Srsf2*$^{P95H/+}$ models, on which I am a co-author (Smeets et al., 2018). Here, I present the full characterisation and comparison of the short-term and long-term effects of *Srsf2*$^{P95H/+}$ mutation on haematopoiesis when activated in myeloid progenitors. This will provide useful information regarding both the prognosis and the therapeutic options for patients with MDS.

3.2 Results

3.2.1 Activation of *Srsf2*$^{P95H/+}$ mutation in myeloid progenitors is well tolerated in 20-30 week old animals

To determine if the *Srsf2*$^{P95H/+}$ mutation would have an effect on haematopoiesis when activated in myeloid progenitors, I established cohorts of *LysM*-Cre *R26*eYFP *Srsf2*$^{+/+}$ and *LysM*-Cre *R26*eYFP *Srsf2*$^{P95H/+}$ mice and allowed them to age. I then performed analysis of peripheral blood (PB), bone marrow (BM), spleen and thymus at 20 to 30 weeks of age (Figure 3.1A). PCR analysis of sorted cells indicated that at this age approximately 50% recombination of the *Srsf2*$^{P95H/+}$ allele was achieved in the common myeloid progenitor (CMP) and GMP populations and this rose to more than 90% in the granulocyte population (Figure 3.1B). The recombination efficiency is well-correlated with the eYFP positive percentage as determined via FACS.

In the peripheral blood, *LysM*-Cre$^{KI/+}$ *R26*eYFP *Srsf2*$^{P95H/+}$ mice (designated as *LysM*-Cre *Srsf2*$^{P95H/+}$) showed a significant decrease in red blood cells (RBC) and increased platelet numbers (PLT) (Figure 3.1C). There was no change in haemoglobin or mean cell volume (MCV) and no increase in the number of leukocytes . There was no change in the myeloid or lymphoid compartments in terms of total cell number, when compared to age matched *LysM*-Cre$^{KI/+}$ *R26*eYFP *Srsf2*$^{+/+}$ (*LysM*-Cre *Srsf2*$^{+/+}$) littermates (Figure 3.1D).

Within the bone marrow, *LysM*-Cre *Srsf2*$^{P95H/+}$ mice had a comparable bone marrow cellularity to age matched *LysM*-Cre *Srsf2*$^{+/+}$ mice (Figure 3.2A). In the more differentiated lineages, there were no changes

in the myeloid and erythroid lineages (Figure 3.2B). However, cell numbers were significantly decreased for the Pre-B, immature B ($B220^{+}IgM^{-}$) and mature B cell populations ($B220^{+}IgM^{+}$), as well as the $CD4^{+}$ T cell populations (Figure 3.2C-D). There were no changes in the stem and progenitor populations, except for decreased numbers of the pre-granulocyte-monocyte progenitors (Pre-GM) (Figure 3.2E-G).

In the spleen, the *LysM*-Cre *Srsf2*$^{P95H/+}$ mice had a decreased cellularity, primarily due to a significant decrease in both B and T lymphoid cells (Figure 3.3A). There was no change in either the erythroid or myeloid compartments (Figure 3.3B). In the thymus, *LysM*-Cre *Srsf2*$^{P95H/+}$ mice had a significantly lower cellularity (Figure 3.3C). The number of T cells, both mature and immature, including double-positive (DP), $CD4^{+}$ and $CD8^{+}$ T cells were significantly lower than in age matched *LysM*-Cre *Srsf2*$^{+/+}$ mice (Figure 3.3D).

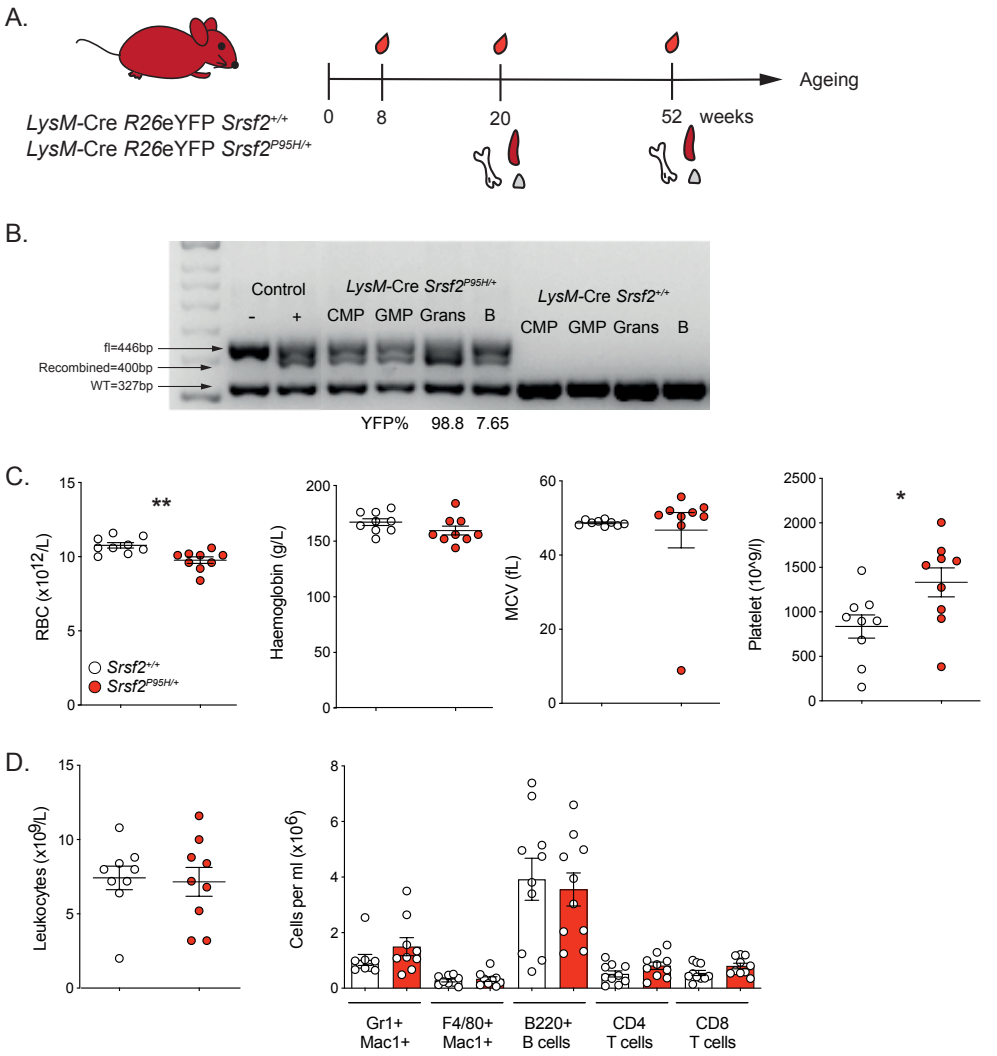


Figure 3.1: *LysM*-Cre *R26*eYFP *Srsf2*$^{P95H/+}$ show no haematological change at 20 to 30 weeks of age.

A. Schematic of *LysM*-Cre *R26*eYFP *Srsf2*$^{P95H/+}$ experiments.

B. Gel image showing that recombination mostly occurs within the myeloid population with granulocytes achieving the highest recombination rate. The eYFP positive percentage for granulocytes and B cells for the same animal is labelled underneath.

C. PB indices of *LysM*-Cre *Srsf2*$^{P95H/+}$ mice around 20 weeks of age.

D. PB leukocyte counts and lineage counts of *LysM*-Cre *Srsf2*$^{P95H/+}$ mice around 20 weeks of age.

CMP, common myeloid progenitor; GMP, granulocyte-macrophage progenitor; Grans, granulocyte. n=9 per genotype. Presented as mean ± standard error of mean. Unpaired student t test. *P<.05, **P<.01.

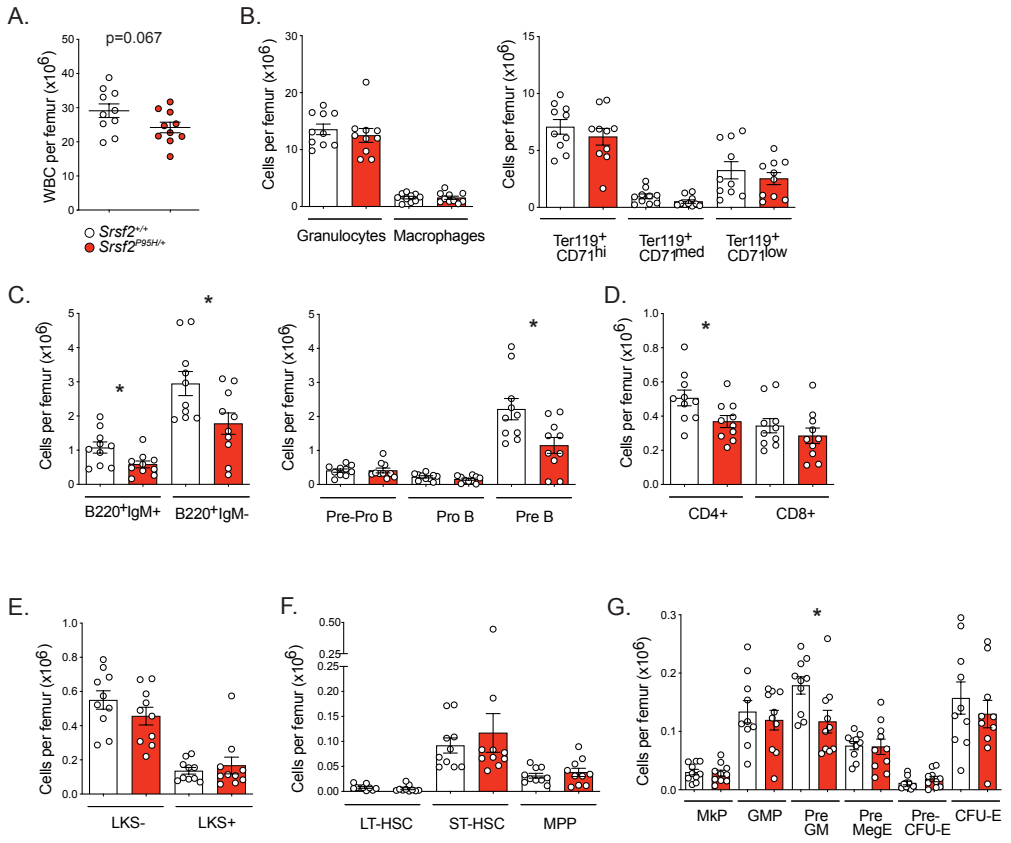


Figure 3.2: *LysM*-Cre *R26*eYFP *Srsf2*$^{P95H/+}$ show limited lymphoid change in the bone marrow at 20 to 30 weeks of age

A. Bone marrow cellularity.

B. No changes in myeloid or erythroid lineages in young mice.

C. Pre B, immature B cell and mature B cell populations are decreased in the *Srsf2*$^{P95H/+}$ mice compared to *Srsf2*$^{+/+}$.

D. CD4^{+} T cells are decreased in the bone marrow of *Srsf2*$^{P95H/+}$ mice.

E. The LKS^{-} and LKS^{+} populations remain unchanged.

F. The number of LT-HSC, ST-HSC and MPP remain unchanged.

G. The progenitor populations remain normal except for a loss in the Pre-GM population in *Srsf2*$^{P95H/+}$ mice,

LKS^{-}, lin^{-}c-Kit^{+}Sca-1^{-}, LKS^{+}, lin^{-}c-Kit^{+}Sca-1^{+}. LT-HSC, long-term haematopoietic stem cell; ST-HSC, short-term haematopoietic stem cell; MPP, multipotent progenitor; Pre-GM, pre-granulocyte-macrophage; MkP, megakaryocyte progenitor; Pre-MegE, pre-megakaryocyte erythroid progenitor; Pre-CFU-E, pre-colony-forming unit erythroid; CFU-E, colony-forming unit erythroid; IgM, immunoglobin M. n=9 per genotype. Presented as mean ± standard error of mean. Unpaired student t test. *P<.05.

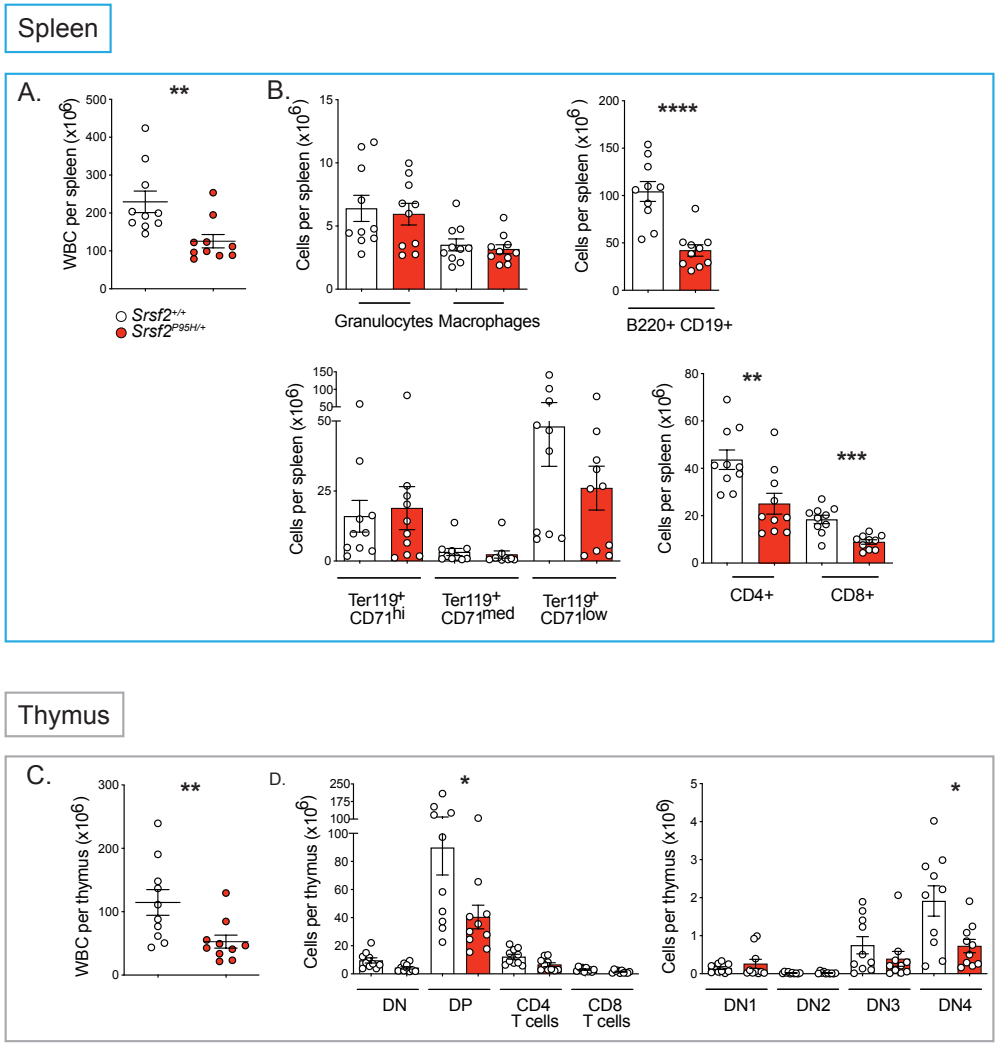


Figure 3.3: *LysM*-Cre *R26*eYFP *Srsf2*$^{P95H/+}$ have a decreased spleen cellularity due to loss of lymphoid populations at 20 to 30 weeks of age
A. Spleen cellularity.
B. B and CD4$^+$ T cell populations are decreased in the spleen of *Srsf2*$^{P95H/+}$ mice
C. Thymus cellularity.
D. No difference detected in the thymic subpopulations, except for a decrease in double-positive (DP) and double-negative 4 (DN4) T cells in *Srsf2*$^{P95H/+}$ mice.
n=9 per genotype. Presented as mean $\pm$ standard error of mean. Unpaired student t test. *P<.05, **P<.01, ***P<.001, ****P<.0001.

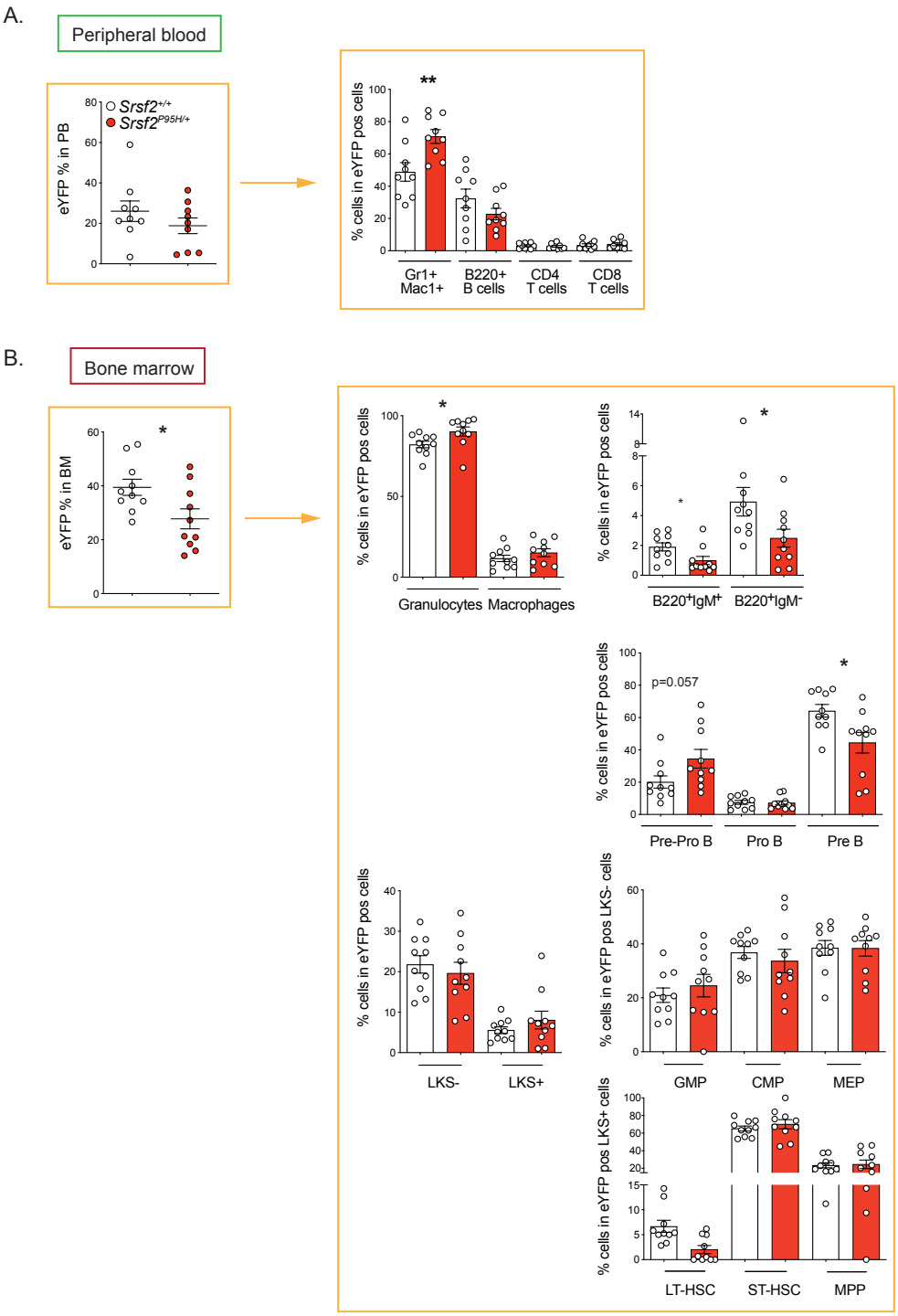
3.2.2 *Srsf2*$^{P95H/+}$ mutant cells show a differentiation bias towards the myeloid lineage

In this murine model, eYFP expression identifies cells that have activated Cre, and by extension have also activated expression of the *Srsf2*$^{P95H/+}$ mutation. The lineage distribution (%) of eYFP positive populations can be used as an indication of any biased differentiation of the *Srsf2*$^{P95H/+}$ mutated cells. Additionally, eYFP labelling creates a chimeric system that provides additional information on the performance of the WT (eYFP negative) population that co-habits the same microenvironment as *Srsf2*$^{P95H/+}$ mutated cells in a native state. In *LysM*-Cre *Srsf2*$^{P95H/+}$ mice, the eYFP positive percentages in all haematopoietic compartments (bone marrow, spleen and thymus), except peripheral blood, were all significantly lower than the eYFP positive percentage in age matched *LysM*-Cre *Srsf2*$^{+/+}$ mice (Figure 3.4A). The eYFP is also activated in lymphoid compartments because *LysM* is expressed at a low level in the stem cell compartment (Clausen et al., 1999). Therefore, the eYFP might be activated in non-myeloid populations (Laiosa et al., 2006). This indicates that *Srsf2*$^{P95H/+}$ mutation has no selective advantage when expressed in the committed myeloid cells.

Within the eYFP positive cells, *Srsf2*$^{P95H/+}$ mutation led to a significant 1.4-fold increase in granulocytes in the peripheral blood when assessed by percentages. Other lineages, as well as the WT (eYFP negative) population (not shown), displayed no difference between *LysM*-Cre *Srsf2*$^{P95H/+}$ and *Srsf2*$^{+/+}$ (Figure 3.4B). In the bone marrow, the percentage of granulocytes within the eYFP positive population was mildly increased (1.06 fold) while the B cells were suppressed at the Pre-B, immature B and mature B cell stages. There was also an increased Pre-Pro B population, indicating a differentiation block (Figure 3.4B). The long-term

haematopoietic stem cells (LT-HSC) were significantly decreased based on analysis using CD34/Flt3 phenotypic populations (Figure 3.4B). In contrast there were no changes in the progenitor populations, including GMP, CMP and myelo-erythroid progenitor (MEP).

In the spleen, however, the myeloid percentages (granulocytes and macrophages) within the eYFP positive fraction were significantly increased. B cells were suppressed, as was observed in the peripheral blood and bone marrow (Figure 3.4C). There were no changes in the T cells, in neither spleen nor thymus (Figure 3.4C-D).


A.
Peripheral blood
eYFP % in PB
Srsf2+/+
Srsf2P95H/+
% cells in eYFP pos cells
Gr1+ Mac1+
B220+ B cells
CD4 T cells
CD8 T cells
B.
Bone marrow
eYFP % in BM
Granulocytes
Macrophages
B220+IgM+
B220+IgM-
p=0.057
Pre-Pro B
Pro B
Pre B
LKS-
LKS+
% cells in eYFP pos LKS- cells
GMP
CMP
MEP
% cells in eYFP pos LKS+ cells
LT-HSC
ST-HSC
MPP

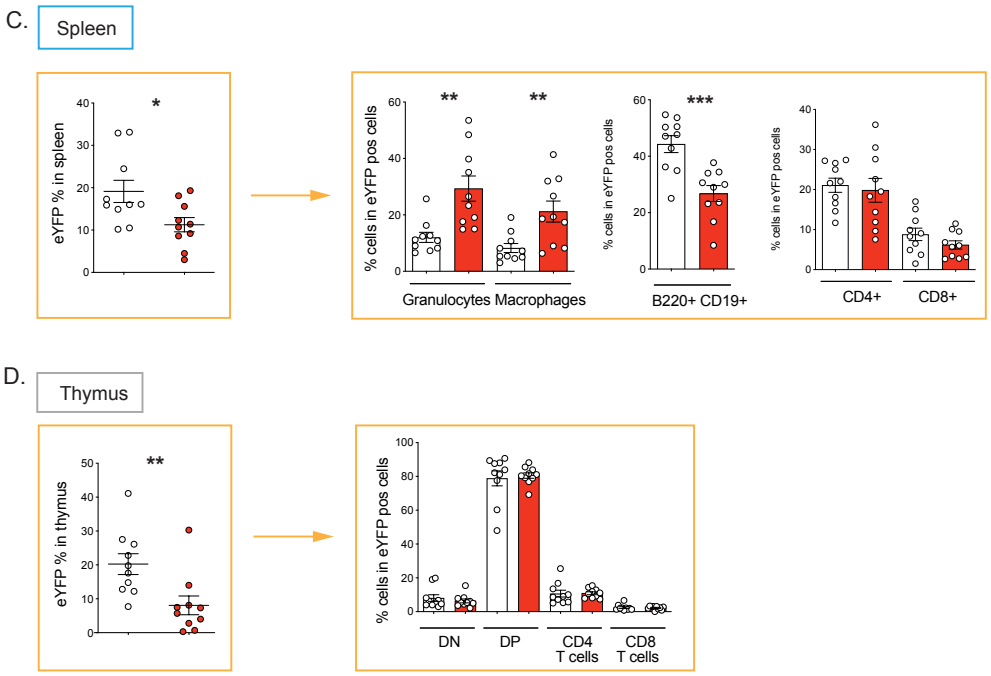


Figure 3.4: *Srsf2*$^{P95H/+}$ expressing (eYFP positive) cells display a mild myeloid bias at the expense of the B lymphoid population
A. PB eYFP percentage and the lineage distribution of PB leukocytes within the eYFP positive fraction.
B. *Srsf2*$^{P95H/+}$ bone marrow cells display a mild biased differentiation towards myeloid, instead of B cells.
C. The myeloid bias is more evident in the eYFP positive spleen cells of *Srsf2*$^{P95H/+}$.
D. No changes in the thymic population within the eYFP positive fraction.
LKS$^-$, lin$^-$c-Kit$^+$Sca-1$^-$, LKS$^+$, lin$^-$c-Kit$^+$Sca-1$^+$. LT-HSC, long-term haematopoietic stem cell; ST-HSC, short-term haematopoietic stem cell; MPP, multipotent progenitor; GMP, granulocyte-macrophage progenitor; CMP, common myeloid progenitor; MEP, myeloerythroid progenitor; IgM, immunoglobin M. n=9 per genotype. Presented as mean ± standard error of mean. Unpaired student t test. **P*<.05, ***P*<.01, ****P*<.001.

3.2.3 Activation of *Srsf2*$^{P95H/+}$ mutation in myeloid progenitors shows a mild myeloid bias and lymphoid suppression of haematopoiesis upon ageing

To determine if activation of the *Srsf2*$^{P95H/+}$ mutation in the myeloid progenitor population has an effect on haematopoiesis with extended ageing, I aged cohorts of *LysM*-Cre *Srsf2*$^{P95H/+}$ mice, along with age matched control *LysM*-Cre *Srsf2*$^{+/+}$ mice and performed full analysis on peripheral blood, bone marrow, spleen and thymus at 50-70 weeks of age. This time point was chosen as our analysis of HSC restricted expression of *Srsf2*$^{P95H/+}$ showed that these mice succumb to MDS/MPN by 50-60 weeks after activation of the *Srsf2* mutation.

In the peripheral blood, the *LysM*-Cre *Srsf2*$^{P95H/+}$ mice presented with macrocytic anaemia (Figure 3.5A). The aged *Srsf2* mutant mice had an increased MCV compared to younger *Srsf2* mutant mice. Lineage analysis showed increased myeloid cells, indicative of myeloid bias as I had observed in 20-week-old animals. Interestingly though, there were also increased CD4 and CD8 mature T cells in the aged mice. The total number of B cells were unchanged (Figure 3.5B). Morphological examination of blood films showed no dysplasia in cells in the peripheral blood (Figure 3.5C)

In contrast to the 20-30 weeks old cohorts, the bone marrow cellularity of 50-70-week-old *LysM*-Cre *Srsf2*$^{P95H/+}$ mice was not significantly different from that of age matched controls (Figure 3.6A). Granulocytes were elevated in percentage but not in absolute number (Figure 3.6B). Conversely, there were significant decreases in most of the B lymphoid populations, from Pre-Pro B cells to immature B cells, as well as CD4$^+$ T cells (Figure 3.6C-D). There was evidence of ineffective erythropoiesis,

supported by a decrease in the number of late-stage erythroid progenitors (Ter119^{+}CD71^{-} or Ter119^{+}CD44^{-}) (Figure 3.6E). In the stem and progenitor populations, LT-HSCs were decreased compared to age matched controls. Progenitor populations showed a trend toward decreased numbers in the *Srsf2*$^{P95H/+}$ bone marrow, but the change was not significant (Figure 3.6F-G).

In the spleen, the *LysM*-Cre *Srsf2*$^{P95H/+}$ mice had a normal cellularity (Figure 3.7A). The spleen, however, showed a mild myeloid bias and an elevated number of early-stage erythroid precursors in *Srsf2*$^{P95H/+}$, which is likely to be a compensatory response to the ineffective erythropoiesis in the bone marrow and low red blood cells in the peripheral blood (Figure 3.7B-C). The B cells and CD4^{+} T cells remained suppressed but to a lesser extent compared to the younger age cohorts that were assessed (Figure 3.7D). In the thymus, double positive (DP) and CD8^{+} T cells were decreased in the *LysM*-Cre *Srsf2*$^{P95H/+}$ mice. The thymus weight was slightly decreased but the cellularity was comparable to age matched control mice (Figure 3.7E-F).

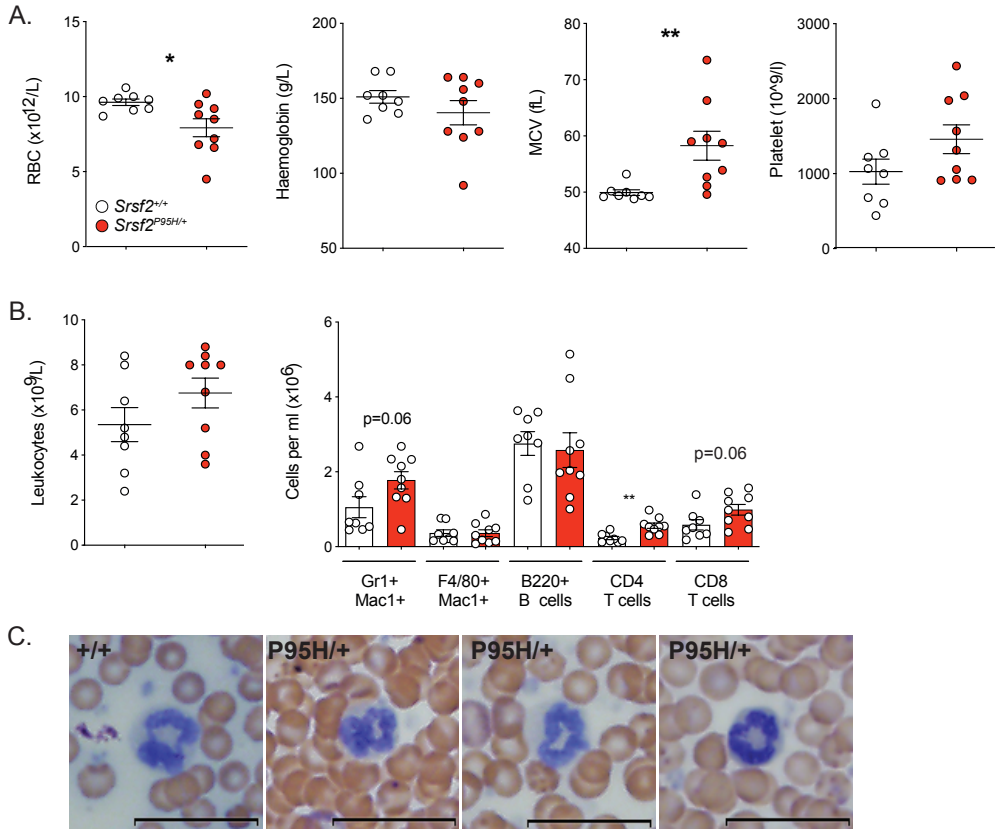


Figure 3.5: Aged *LysM*-Cre *R26*eYFP *Srsf2*$^{P95H/+}$ mice develop macrocytic anaemia with increased granulocytes in the peripheral blood
A. PB indices of the aged cohorts.
B. PB leukocyte count and lineage distribution of cells in aged mice.
C. PB blood film showing no dysplastic changes in the leukocytes of aged *LysM*-Cre *R26*eYFP *Srsf2*$^{P95H/+}$ mice.
n=8 *Srsf2*$^{+/+}$, n=9 *Srsf2*$^{P95H/+}$. May-Grünwald Giemsa staining. Image taken at 40x. Presented as mean $\pm$ standard error of mean. Unpaired student t test. *P<.05, **P<.01. Scale bar 20μm.

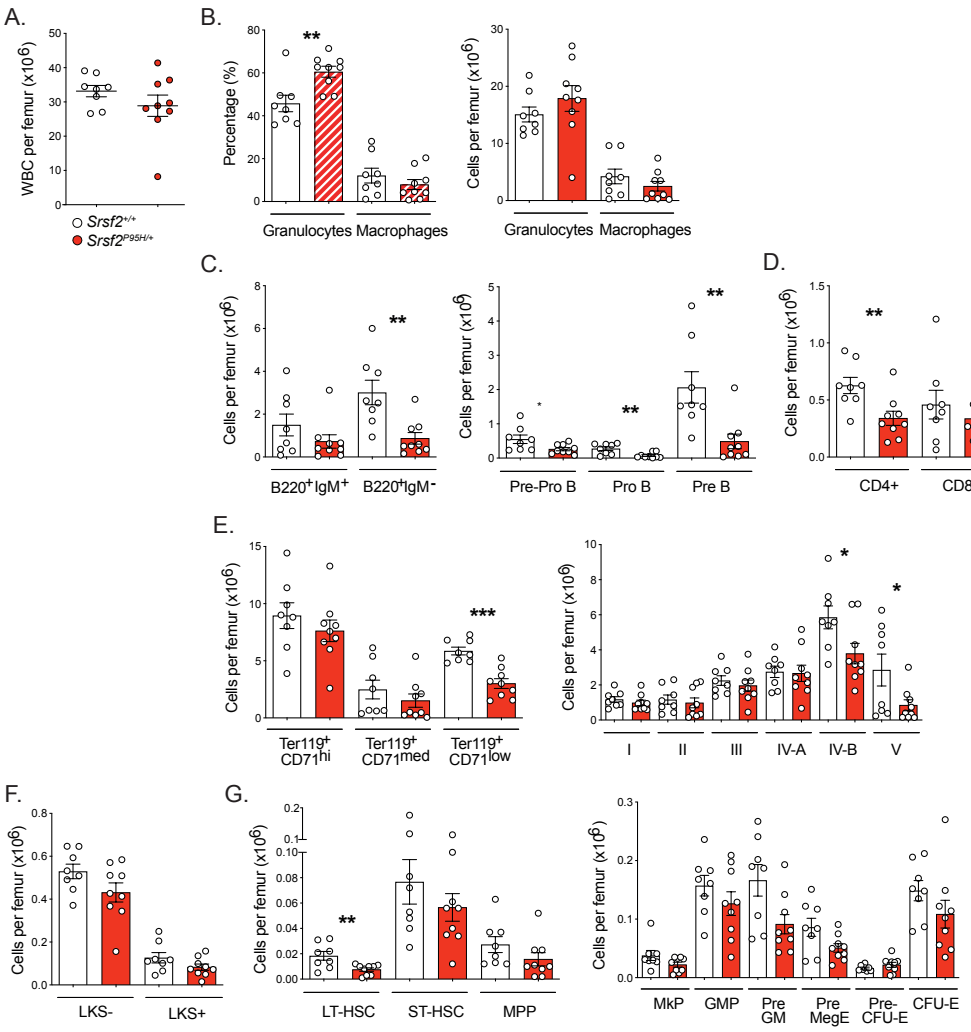


Figure 3.6: Aged *LysM*-Cre *R26*eYFP *Srsf2*$^{P95H/+}$ mice have decreased lymphoid and erythroid populations in the bone marrow

A. Bone marrow cellularity.

B. Increased percentage of granulocytes in the *Srsf2*$^{P95H/+}$ indicating myeloid expansion, although not significant in absolute number.

C. B lymphopoiesis is severely compromised in aged *Srsf2*$^{P95H/+}$ bone marrow.

D. CD4^{+} T cells are decreased in the aged *Srsf2*$^{P95H/+}$ bone marrow.

E. The late erythroid precursor population is decreased in the aged *Srsf2*$^{P95H/+}$ bone marrow.

F. No differences in the number of LKS^{-} and LKS^{+} cells between genotypes.

G. LT-HSC population is decreased in the aged *Srsf2*$^{P95H/+}$ bone marrow, while other stem and progenitor populations remain unchanged.

LKS^{-}, lin^{-}c-Kit^{+}Sca-1^{-}, LKS^{+}, lin^{-}c-Kit^{+}Sca-1^{+}. LT-HSC, long-term haematopoietic stem cell; ST-HSC, short-term haematopoietic stem cell; MPP, multipotent progenitor; GMP, granulocyte-macrophage progenitor; CMP, common myeloid progenitor; MEP, myeloerythroid progenitor; IgM, immunoglobin M. n=8 *Srsf2*$^{P+/+}$, n=9 *Srsf2*$^{P95H/+}$. Presented as mean ± standard error of mean. Unpaired student t test. **P*<.05, ***P*<.01, ****P*<.001.

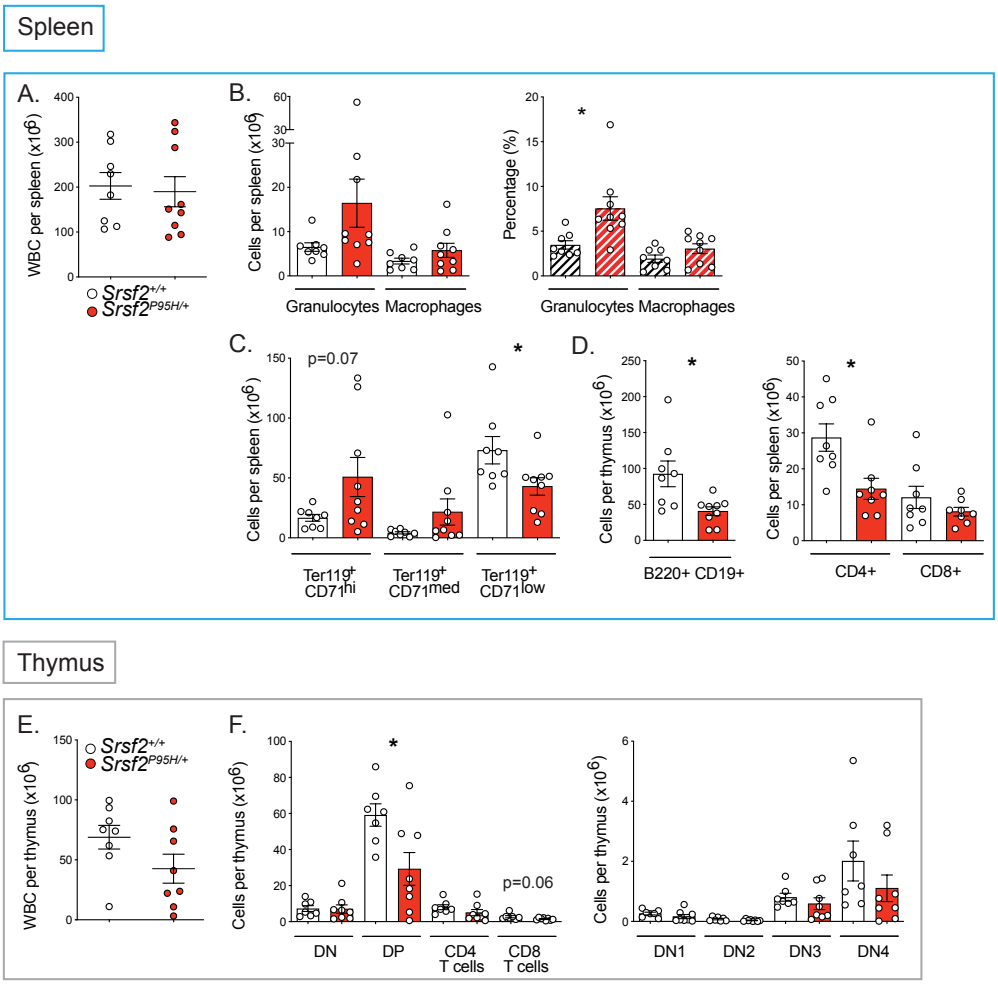


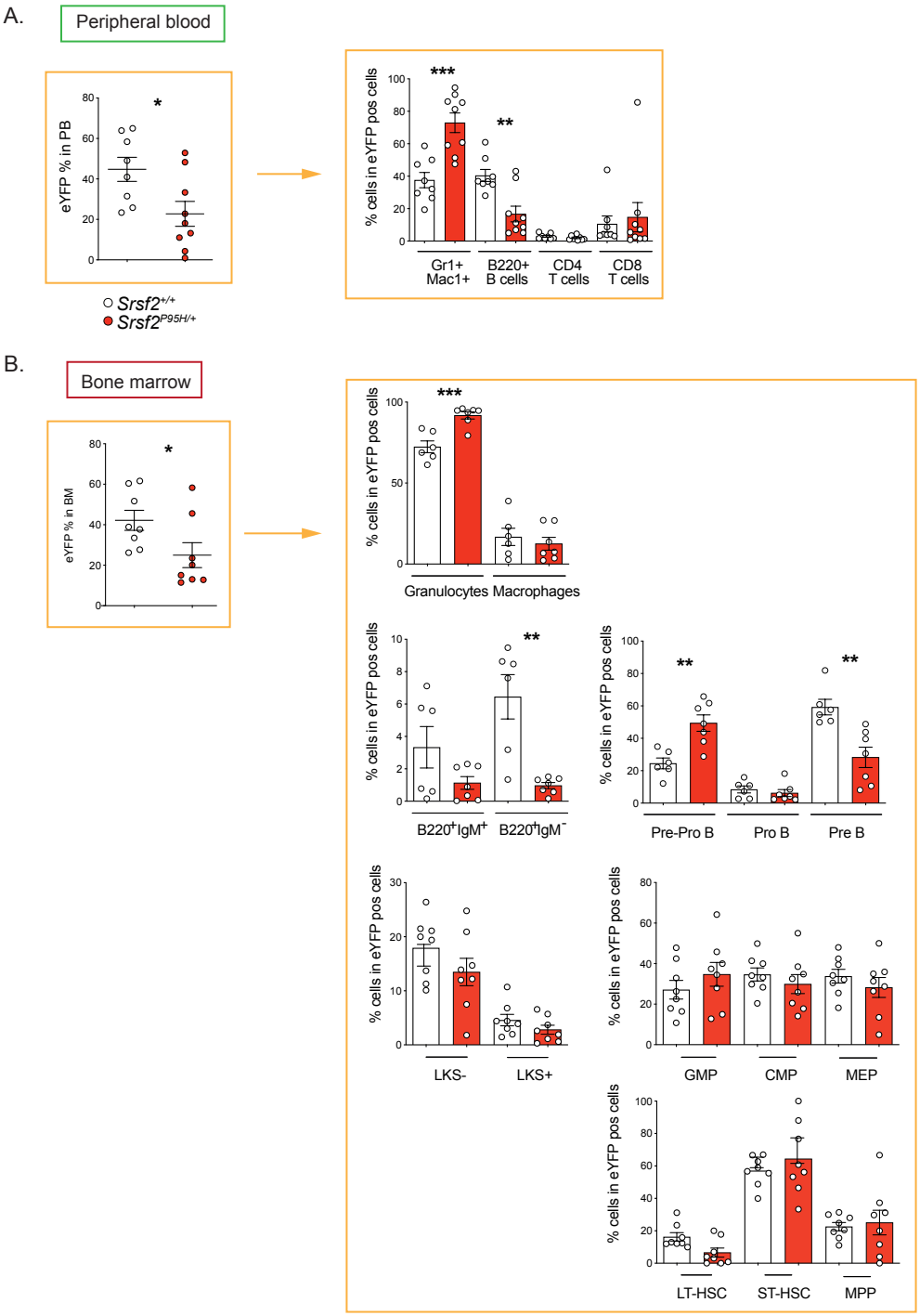
Figure 3.7: Aged *LysM*-Cre *R26*eYFP *Srsf2*$^{P95H/+}$ mice display a mild myeloid bias with lymphoid and erythroid suppression in the spleen.
A. Spleen cellularity.
B. Myeloid expansion is mild in the spleen of aged *Srsf2*$^{P95H/+}$ mice.
C. The late erythroid precursors are decreased in the spleen of aged *Srsf2*$^{P95H/+}$ mice.
D. B cells and CD4^{+} T cells are decreased in the spleen of aged *Srsf2*$^{P95H/+}$ mice.
E. Thymus cellularity.
F. Double positive (DP) population is significantly lower in *Srsf2*$^{P95H/+}$ thymus while other populations remain normal.
n=8 *Srsf2*$^{+/+}$, n=9 *Srsf2*$^{P95H/+}$. Presented as mean $\pm$ standard error of mean. Unpaired student t test. *P<.05.

3.2.4 *Srsf2*$^{P95H/+}$ mutant cells display a myeloid bias and B lymphoid suppression in aged mice

Compared to the younger cohort, the eYFP positive percentages in the peripheral blood, bone marrow, spleen and thymus stayed low in the aged *LysM*-Cre *Srsf2*$^{P95H/+}$ mice, while the age matched control maintained a higher level of eYFP positivity.

In the peripheral blood, the eYFP positive *Srsf2*$^{P95H/+}$ mutant cells produced a significantly (1.75 fold) increased percentage of granulocytes and decreased percentages of B cell as was observed in the young cohorts, and no change in CD4 or CD8 positive T cells (Figure 3.8A). In the bone marrow, the differentiation of *Srsf2*$^{P95H/+}$ mutated cells was still biased towards granulocytes away from B cells. The block in B cell development was more prominent as the Pre-pro B population from *Srsf2*$^{P95H/+}$ cells further increased and Pre-B and immature B cells further decreased, compared to the young cohorts (Figure 3.8B). No changes were detected in the stem and progenitor population in the eYFP positive fraction (Figure 3.8B).

In contrast to the bone marrow, the myeloid bias in the spleen (both granulocytes and macrophages) was more severe than observed in the young cohorts and the B cell suppression was still present. There were no obvious changes in the T cell populations in the spleen (Figure 3.8C) or thymus (Figure 3.8D).


A.
Peripheral blood
eYFP % in PB
Srsf2+/+
Srsf2P95H/+
% cells in eYFP pos cells
Gr1+ Mac1+
B220+ B cells
CD4 T cells
CD8 T cells
B.
Bone marrow
eYFP % in BM
Granulocytes
Macrophages
B220+IgM+
B220+IgM-
Pre-Pro B
Pro B
Pre B
LKS-
LKS+
GMP
CMP
MEP
LT-HSC
ST-HSC
MPP

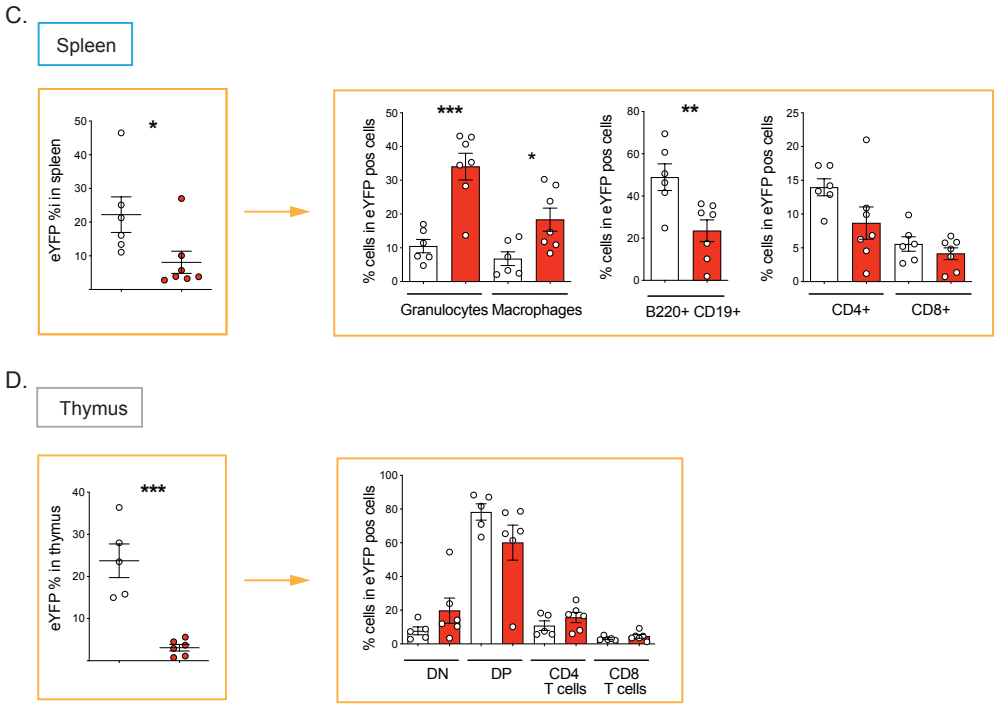


Figure 3.8: *Srsf2*$^{P95H/+}$ expressing (eYFP positive) cells in aged mice display a profound myeloid bias in bone marrow and spleen
A. The lineage distribution of PB leukocytes within the eYFP positive fraction.
B. Aged *Srsf2*$^{P95H/+}$ bone marrow cells display a biased differentiation towards myeloid, instead of B cells.
C. The myeloid bias is significant in the eYFP positive spleen cells of aged *Srsf2*$^{P95H/+}$ (n=6 *Srsf2*$^{+/+}$, n=7 *Srsf2*$^{P95H/+}$).
D. No changes in the thymic population within the eYFP positive fraction (n=5 *Srsf2*$^{+/+}$, n=6 *Srsf2*$^{P95H/+}$).
LKS^{-}, lin^{-}c-Kit^{+}Sca-1^{-}, LKS^{+}, lin^{-}c-Kit^{+}Sca-1^{+}. LT-HSC, long-term haematopoietic stem cell; ST-HSC, short-term haematopoietic stem cell; MPP, multipotent progenitor; GMP, granulocyte-macrophage progenitor; CMP, common myeloid progenitor; MEP, myeloerythroid progenitor; IgM, immunoglobin M. n=8 *Srsf2*$^{+/+}$, n=9 *Srsf2*$^{P95H/+}$ unless otherwise specified. Presented as mean $\pm$ standard error of mean. Unpaired student t test. **P*<.05, ***P*<.01, ****P*<.001.

3.3 Discussion

Understanding how *Srsf2*$^{P95H/+}$ mutation disrupts normal haematopoiesis is important for elucidating the role of this RNA splicing mutation in MDS/CMML pathogenesis. Genetically engineered murine models have provided a great opportunity to trace the origin of MDS initiation. For *Srsf2*$^{P95H/+}$ mutation, various conditional knock-in models were generated, driven by Cre recombinase under the control of different promoters, such as *Mx1*-Cre and *Vav1*-Cre. Both *Mx1*-Cre and *Vav1*-Cre models activate *Srsf2*$^{P95H/+}$ mutation in all haematopoietic tissues, including the bone marrow stroma (De Boer et al., 2003; Joseph et al., 2013). Although this approach explores the full effects of *Srsf2*$^{P95H/+}$ mutation on the haematopoietic system, the cells of origin that are responsible for RNA splicing mutant MDS initiation remained undetermined. To understand which cell population is crucial for MDS initiation, our lab utilised three different Cre recombinase systems: *R26*-CreER^{T2}, h*Scl*-CreER^{T}and *LysM*-Cre. *R26*-CreER^{T2} activates the mutation widely across different tissues including non-haematopoietic tissues, while h*Scl*-CreER^{T}activates the mutation in haematopoietic stem cells. Both *R26*-CreER^{T2} and h*Scl*-CreER^{T}*Srsf2*$^{P95H/+}$ mice developed MDS/MPN after activating *Srsf2*$^{P95H/+}$ mutation, while initial analysis of *LysM*-Cre *Srsf2*$^{P95H/+}$ mice suggested that *Srsf2*$^{P95H/+}$ mutation is required in the stem cells for MDS development (Smeets et al., 2018).

Here I report the phenotypes of *LysM*-Cre *R26*eYFP *Srsf2*$^{P95H/+}$ mice in settings of both early and aged haematopoiesis. At 20 to 30 weeks of age, the *LysM*-Cre *R26*eYFP *Srsf2*$^{P95H/+}$ mice presented with largely normal haematopoietic development. A suppression of B lymphopoiesis was evident in spleen and bone marrow in *Srsf2*$^{P95H/+}$ cells, while only subtle myeloid and erythroid changes were apparent in the spleen and

bone marrow of *Srsf2*$^{P95H/+}$ mice. There was reduced thymic cellularity in *Srsf2*$^{P95H/+}$ mice as well. The peripheral blood (PB) showed a decrease in red blood cells but no macrocytic changes were detected as were apparent in the *R26*-CreER^{T2} and h*Scl*-CreER^{T}directed expression of *Srsf2*. At 50 to 70 weeks of age, the *LysM*-Cre *R26*eYFP *Srsf2*$^{P95H/+}$ mice showed a more pronounced myeloid bias and B lymphoid suppression across all haematopoietic tissues. The peripheral blood revealed macrocytic anaemia and elevated white blood cells. The long-term haematopoietic stem cell (LT-HSC) population was decreased, and ineffective erythropoiesis occurred in the bone marrow while compensatory stress erythropoiesis was present in the spleen. It has been reported that in *LysM*-Cre mice eYFP labelling can be detected in haematopoietic stem cells or even in lymphoid and erythroid lineages. This is due to the *lysozyme* gene being expressed at a low level in a subset of haematopoietic stem cells, thereby activating the Cre (Ye et al., 2003). Therefore, low levels of *Srsf2*$^{P95H/+}$ mutation might be expressed in the B and T lymphoid progenitors, which might partially account for the reduced thymic population and B cells. However, for the purpose of assessing the role of *Srsf2*$^{P95H/+}$ in myeloid progenitors, this is still a valid system as the mutation is primarily activated within the myeloid lineage (50-70% determined by eYFP labelling).

While the phenotypes of *LysM*-Cre *R26*eYFP *Srsf2*$^{P95H/+}$ mice included some features of MDS, they did not satisfy the criteria for MDS diagnosis. First of all, there were no dysplastic changes detected in the peripheral blood, even in the aged animals. Secondly, the myeloid bias (60% of granulocytes in bone marrow) is only mild compared to *R26*-CreER^{T2} or h*Scl*-CreER^{T} *Srsf2*$^{P95H/+}$ mice in the MDS state (75% of granulocytes in bone marrow). Comparing the aged and young (20-30 weeks) *LysM*-Cre mice, there is little change in the eYFP positivity, which marks the

Srsf2$^{P95H/+}$ expressing cells. This is in contrast to an expansion of eYFP positivity seen in the h*Scl*-CreER^{T} *R26*eYFP *Srsf2*$^{P95H/+}$ mice (Smeets et al., 2018). In Clonal haematopoiesis of indeterminate potential (CHIP), *SRSF2* mutation is positively selected, which argues for a likely HSC origin, as only stem cells self-renew and propagate (Genovese et al., 2014; Martincorena et al., 2018). This allows maintenance of *SRSF2* mutation and accumulation of co-operative mutations. The increased myeloid bias is likely due to a continuous push for differentiation from within the *Srsf2*$^{P95H/+}$ mutant myeloid progenitors. It would be worthwhile to test the functional capacity of the mutant myeloid cells to determine if they are normal. However, these splicing mutant progenitors don't have the capacity for self-renewal or transformation. Therefore, activating *Srsf2*$^{P95H/+}$ mutation in myeloid progenitors is not sufficient for MDS initiation and subsequent development. Our model is the first to investigate the cell of origin in MDS using myeloid specific Cre activation. Other MDS murine models, such as *NUP98-Hox13* and *NPM1* mutation, are germ-line models and cannot address the question about cell-of-origin in MDS (Grisendi et al., 2006; Lin et al., 2005; Sportoletti et al., 2008). Consistent with human studies, the stem cell compartment is more likely to be the initiating MDS clone (Woll et al., 2014). It is possible that progenitor cells carrying the same genetic lesion are insufficient to propagate and result in MDS.

These results raise the important question: what is the role/s of myeloid progenitors in MDS pathogenesis? As demonstrated experimentally, myeloid progenitors are not able to act as a cell of origin for MDS, when using SRSF2^{P95H} as the initiating lesion, but how much they contribute to the development of MDS warrants further investigation. The disturbed haematopoiesis shown in the aged *LysM*-Cre *R26*eYFP *Srsf2*$^{P95H/+}$ mice did overlap with the phenotypes of *Srsf2*$^{P95H/+}$ mice that

developed MDS (using *R26*-CreER^{T2} or h*Scl*-CreER^{T}), except for the lack of dysplasia. It is unclear why no dysplasia was found in the *LysM*-Cre model. It is possible that splicing changes of particular transcripts are required in the stem cell population to induce the morphological abnormality. This argues that *Srsf2* mutant myeloid progenitors may be the source of excessive myeloid lineage production, which then suppresses B cell production in the bone marrow and spleen. When the myeloid bias persists and is sustained for long periods, the erythroid lineage becomes affected. This can lead to the ineffective erythropoiesis and macrocytic anaemia observed in the old mice. However, unlike the *Rps14* insufficiency, no changes were observed in the early stages of erythropoiesis or haemoglobin levels of *Srsf2* mutant mice (Schneider et al., 2016). There was no splenic enlargement that would be indicative of extramedullary erythropoiesis. This series of events is seen in the *LysM*-Cre *Srsf2*$^{P95H/+}$ mice, which suggests that myeloid progenitors might contribute to the myeloid bias and subsequent lineage disruption. Since the *Srsf2*$^{P95H/+}$ mutation is acquired in the myeloid progenitor stage, the lymphoid and erythroid progenitors should not carry the mutation. Thus, the effect of *Srsf2*$^{P95H/+}$ mutation seen on B and erythroid lineages is an indirect result of myeloid bias. There are a few hypotheses about the mechanism of lineage disruption, with microenvironmental changes being the favourite candidate (Raaijmakers et al., 2010; Singbrant et al., 2011; Walkley et al., 2007a; Walkley et al., 2007b). However, the phenotypes observed in the *LysM*-Cre model are relatively mild compared to the *R26*-CreER^{T2} or h*Scl*-CreER^{T} *Srsf2*$^{P95H/+}$ mice. Therefore, the myeloid progenitor population is unlikely to be the only contributor to the myeloid bias and lineage disruption. More important, the fact that no dysplasia was observed in the *LysM*-Cre model suggests that myeloid bias and dysplastic changes are likely the result of separate molecular changes in *Srsf2*$^{P95H/+}$ mutant

cells. These results reflect the complexity of RNA splicing and emphasise its role in many of the fundamental cellular pathways.

Chapter 4

Transcriptome change of *Srsf2*$^{P95H/+}$ mutant cells

4.1 Introduction

In the effort of understanding the pathogenesis of MDS more completely, several studies have performed transcriptional profiling and pathway analysis on human or murine MDS samples. So far, several dysregulated pathways have been identified. Pellagatti et al. analysed the transcriptional profile of $CD34^+$ cells from 183 MDS patients (Pellagatti et al., 2018). A total of 35 genes were up-regulated and 139 were down-regulated in at least half of the samples. Pathways analysis identified Interferon signalling, thrombopoietin signalling and IL-2 signalling as the top up-regulated pathways, while primary immunodeficiency signalling, B-cell receptor signalling and cell cycle regulation by BTG family proteins were the top down-regulate pathways (Pellagatti et al., 2018). For $SRSF2^{P95H}$ mutant MDS, however, there were few transcriptional changes in $SRSF2^{P95H}$ cells compared to $SRSF2^{WT}$ control in human $CD34^+$ cells (Pellagatti et al., 2018; Shiozawa et al., 2018). In murine models, Kim et al. reported up-regulation of *Gfi1*, *Cebpe* and *Hoxb2* in the $lin^-c\text{-}Kit^+Sca\text{-}1^+$ (LKS^+) population and down-regulation of *Gata1* and *Gata2* in progenitors ($lin^-c\text{-}Kit^+Sca\text{-}1^-$; LKS^-) and *Cdkn1a* in both LKS^+ and LKS^-. Gene ontology (GO) revealed that down-regulated genes were involved in cell cycle regulation, apoptosis, proliferation and differentiation (Kim et al., 2015). Similar changes in expression were reported by Kon et al., with additional up-regulated genes including *Hnrnpa2b1* and down-regulated genes such as *Mpl, Zrf4, Csf3r* and *Trp53bp1* (Kon et al., 2018).

A comparison of the dysregulated gene expression and the transcripts subjected to altered splicing in $SRSF2^{P95H}$ mutant cells resulted in a more dynamic and complex picture emerging. From eight $SRSF2^{P95H}$ mutant MDS $CD34^+$ samples, Pellagatti et al. found 132 mis-splicing events, with skipped exons being the most frequent aberrant splicing

type. The top alternatively spliced genes were *PKFM*, *METTL17*, *STAG2* and *EZH2*. GO analysis found that messenger RNA metabolism (RNA splicing, processing and translation) and ribonucleoprotein complex biogenesis (mitochondrial organisation and translation) associate with the mis-spliced events (FDR <0.05) (Pellagatti et al., 2018). Subsequent to this finding, Shiozawa et al. analysed thirty SRSF2^{P95H} mutant MDS patient samples. A total of 599 differentially spliced events were found in CD34^{+} cells and 1,155 in bone marrow mononuclear cells (BMMNCs), with 303 events overlapping between the two populations. Among these events in CD34^{+} cells, mis-splicing of cassette exons (also known as exon skipping) was the most frequent (n=380), followed by changes in alternative 3′ splicing (n=76), alternative 5′ splicing (n=64) and intron retention (n=39). Most alternative splicing events were predicted to be in coding variants of transcripts, with 46% predicted to produce a truncated transcript. Truncated transcripts would have deleterious effects on gene expression. However, there was no significant reduction in the expression level of the identified transcripts, suggesting that the excess truncated transcripts do not disrupt canonical splicing (Shiozawa et al., 2018).

Here, I have examined the transcriptome changes in purified *Srsf2*$^{P95H/+}$ cells from the h*Scl*-CreER$^{T\ Tg+}$ *R26*eYFP *Srsf2* murine model 20 weeks after activation of *Srsf2*$^{P95H/+}$ mutation. At this time point, the myeloid bias is apparent, but the animals have not yet developed overt MDS/MPN. This enabled me to understand the effects of the *Srsf2* mutation without the possibly confounding effects of additional mutations which may contribute to the manifestations of the full MDS/MPN phenotype at later points. I then compared the splicing changes identified from the *in vivo* model, as well as *in vitro* cell lines, with data from published studies. This was to allow the identification of the most frequent and species

conserved splicing changes in *Srsf2*$^{P95H/+}$ cells. Finally, I present the result of a preliminary single cell sequencing study of *Srsf2*$^{P95H/+}$ mutated haematopoietic stem and progenitor cells.

4.2 Results

4.2.1 *Srsf2*$^{P95H/+}$ mutant stem cells exhibit a myeloid biased transcriptional signature

To understand the transcriptome changes in *Srsf2*$^{P95H/+}$ mutant cells, I activated the *Srsf2*$^{P95H/+}$ mutation in h*Scl*-CreER$^{T\ Tg+}$ *R26*eYFP *Srsf2*$^{P95H/+}$ mice by tamoxifen treatment (n=3). I also tamoxifen-treated the h*Scl*-CreER$^{T\ Tg+}$ *R26*eYFP *Srsf2*$^{+/+}$ mice at the same time as control (n=3). After 20 weeks of mutation activation, I sorted the Lin^{-} c-Kit^{+} eYFP^{+} fraction, which contains the haematopoietic stem and progenitor cells, and performed RNA sequencing (Figure 4.1). This phenotypic population was chosen because at the time of the experiment comparable phenotypic populations had been used to assess gene expression and splicing changes in an alternative *Srsf2*$^{P95H/+}$ model (Kim et al., 2015) and in a *Sf3b1*$^{K700E/+}$ mutant model (Obeng et al., 2016). There was no evidence to suggest any changes of cell surface phenotypes which might impact the sorting of phenotypically defined Lin^{-} c-Kit^{+} eYFP^{+} population. All the bioinformatic analysis post RNA-seq was performed by Dr. Alistair Chalk.

Transcriptome analysis detected both up- and down-regulation of lineage-specific transcripts. In total, 195 genes were down-regulated and 45 up-regulated (at an FDR <0.05). Among the down-regulated genes, several lymphoid-associated transcripts, such as *Lck*, *Ly6d* and *Blk*, exhibited a 1.3 to 2-fold decreased expression in *Srsf2*$^{P95H/+}$ compared to *Srsf2*$^{+/+}$ cells. In contrast, genes associated with myeloid expansion or

differentiation and limiting HSC self-renewal, such as *Hmga2*, *Ncam1*, *Aff2* and *Grb10*, were up-regulated (Lam et al., 2014; Sasca et al., 2019; Yan et al., 2016). Genes negatively regulating myeloid life span or function, such as *Bcl2l11* (*Bim*) and *Socs3*, were down-regulated in *Srsf2*$^{P95H/+}$ cells (Kimura et al., 2004; Kirschnek et al., 2011) (Figure 4.2A). Analysis of pathway enrichment using Quantitative Set Analysis for Gene Expression (QuSage) (Yaari et al., 2013) revealed activation of the P38 mitogen-activated protein kinase (MAPK) pathways and gene expression signatures associated with human MDS in the *Srsf2*$^{P95H/+}$ cells (Navas et al., 2006; Zhou et al., 2007) (Figure 4.2B).

To validate the differential transcript expression, I isolated Lin^{-} c-Kit^{+} CD45.2^{+} cells from transplant recipients (24 weeks post-transplant, n=4 per genotype) that had been reconstituted with bone marrow from tamoxifen treated *R26*-CreER^{T2} *Srsf2*$^{+/+}$ or *R26*-CreER^{T2} *Srsf2*$^{P95H/+}$ animals. Real time quantitative polymerase chain reaction (RT-qPCR) validation of the differentially expressed genes was then performed. The results confirmed a significantly decreased expression of *Ly6d*, *Blk*, *Lck* (p=0.07) and *Bim* (p=0.07), as well as the increased expression of *Ncam1* in *Srsf2*$^{P95H/+}$ cells. However, no change in the expression of *Dach2* and *Hmga2* was detected in the independent validation samples (Figure 4.2C).

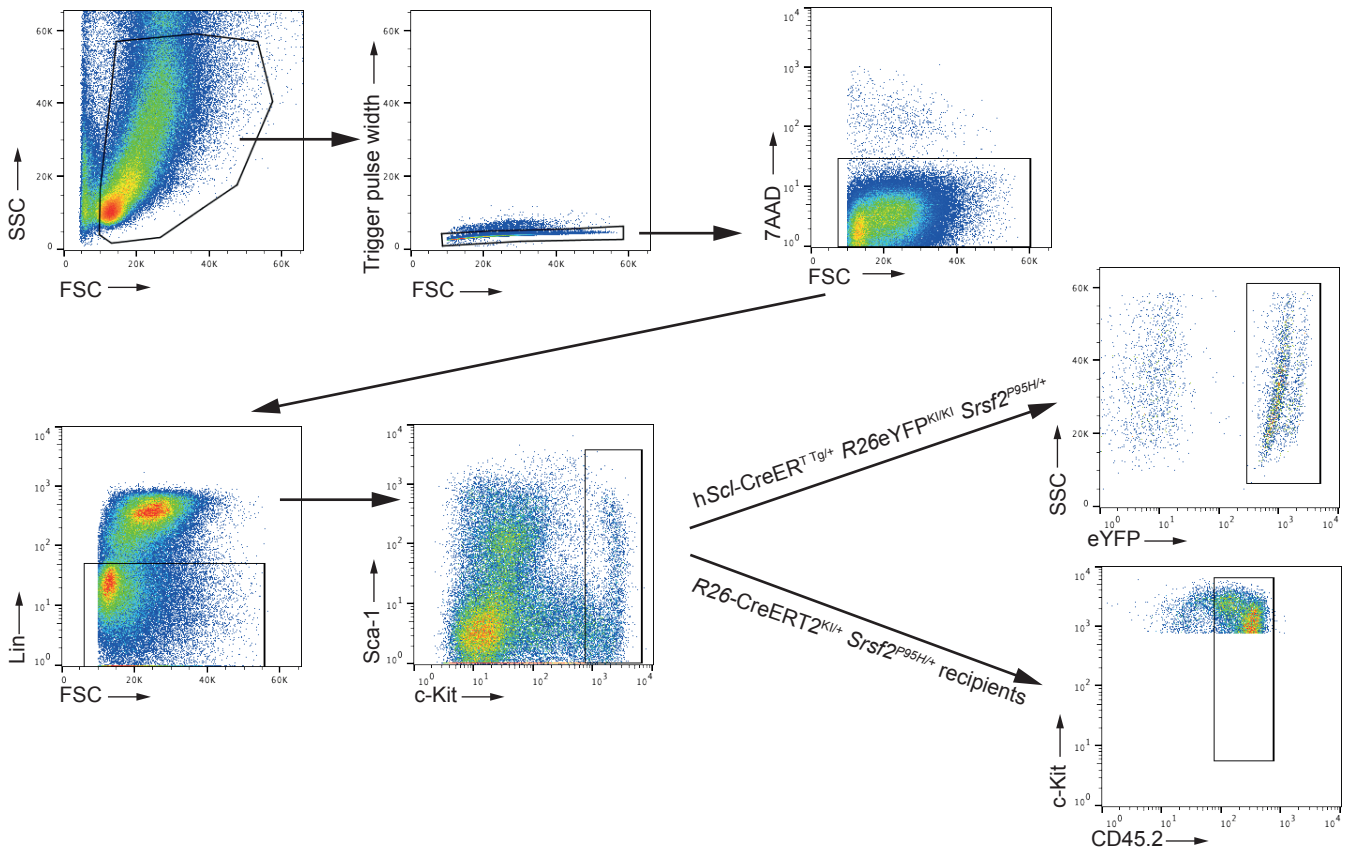
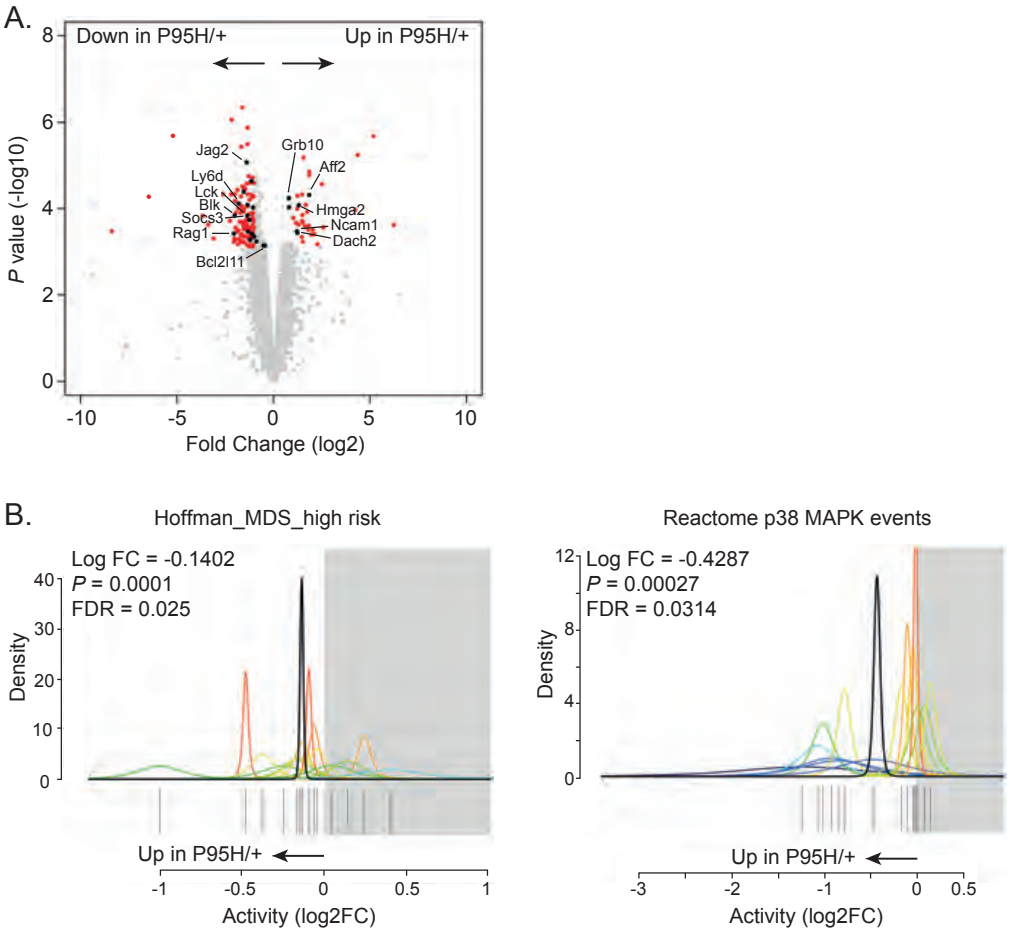


Figure 4.1: The sorting strategy of *Srsf2*$^{P95H/+}$ stem and progenitor cells for h*Scl*-CreER$^{T\ Tg+}$ R26eYFP *Srsf2*$^{P95H/+}$ and transplant recipients of *R26*-CreER^{T2} *Srsf2*$^{P95H/+}$


A.
Down in P95H/+
Up in P95H/+
P value (-log10)
Fold Change (log2)
Jag2
Ly6d
Lck
Blk
Socs3
Rag1
Bcl2l11
Grb10
Aff2
Hmga2
Ncam1
Dach2
B.
Hoffman_MDS_high risk
Log FC = -0.1402
P = 0.0001
FDR = 0.025
Density
Up in P95H/+
Activity (log2FC)
Reactome p38 MAPK events
Log FC = -0.4287
P = 0.00027
FDR = 0.0314
Density
Up in P95H/+
Activity (log2FC)

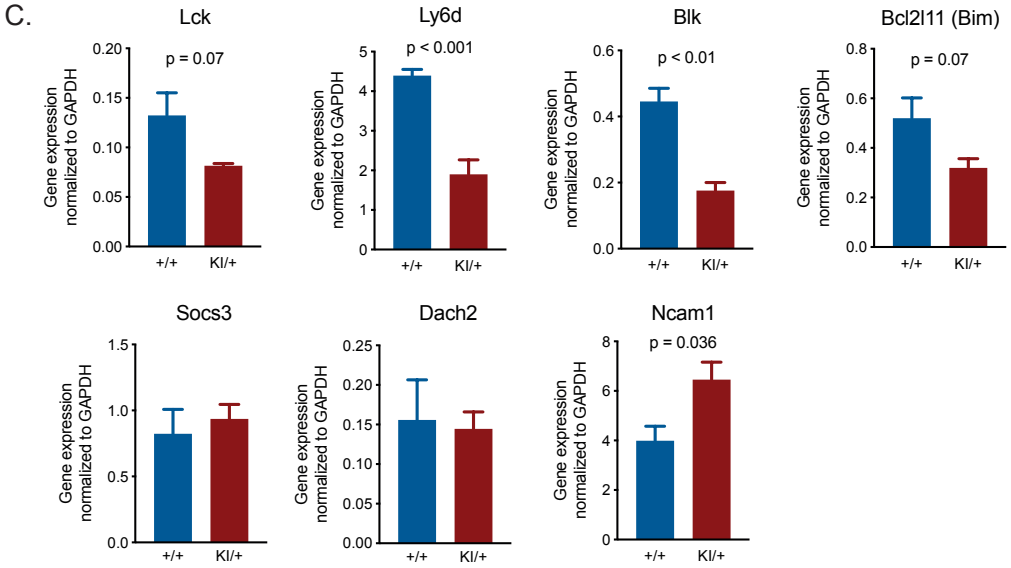


Figure 4.2: Transcriptome analysis demonstrates an up-regulation of myeloid specific transcripts and down regulation of lymphoid associated genes in *Srsf2*$^{P95H/+}$ stem and progenitor cell

A. Volcano plot showing up- and down-regulated genes in lin^{-} cKit^{+} eYFP^{+} *Srsf2*$^{P95H/+}$ cells isolated 20 weeks after tamoxifen cessation (n=3). Transcripts with significant changes are labelled in red.

B. Quantitative Set Analysis for Gene Expression Analysis (QuSage) of data from (A).

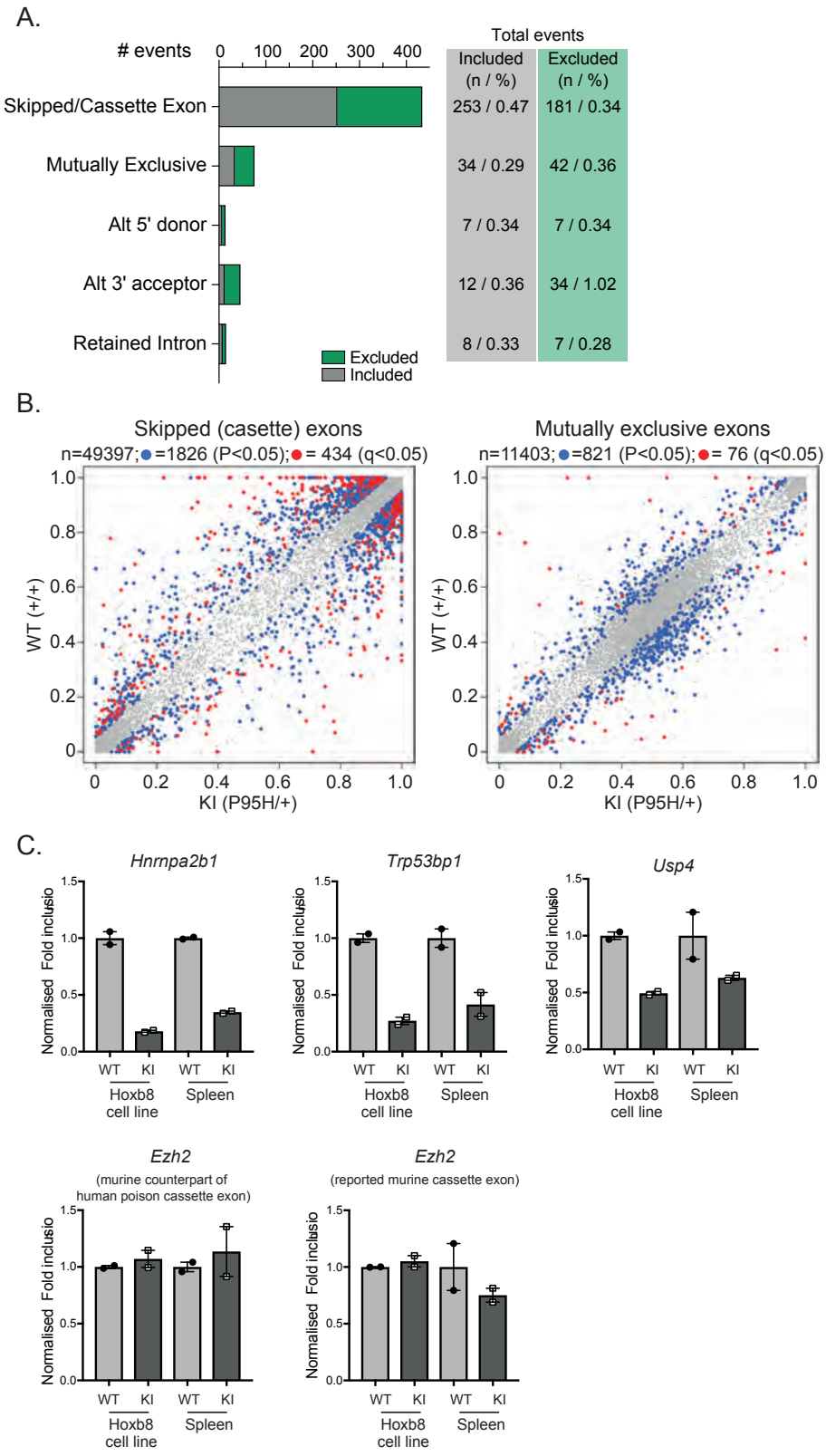
C. RT-qPCR validation of key up- and down-regulated genes identified in the *Srsf2*$^{P95H/+}$ RNA-seq using lin^{-}cKit^{+}CD45.2^{+} cells isolated 24 weeks post transplantation (n=4 per genotype). Unpaired student t test. Data presented as mean $\pm$ standard error of mean.

We then interrogated the changes in splicing in the *Srsf2*$^{P95H/+}$ cells. In total, rMATs analysis revealed 585 mis-spliced genes at an FDR of <0.05 and an inclusion cut-off of 0.05. The inclusion cut-off of 0.05 means that the mis-spliced event has to be present in at least 5% of the exon-intron boundary and splice junctions. The most frequent category of mis-splicing was skipped exons (434/585, 74.1%). Mis-spliced events that were statistically different in the *Srsf2*$^{P95H/+}$ cells composed less than 1% of all the mis-spliced events detected (1.02% in alternative 3' splice site), which is consistent with earlier reports from alternative models (Zhang et al., 2015) (Figure 4.3A-B). There was no overlap nor correlation between the differentially expressed genes and mis-spliced genes in the *Srsf2*$^{P95H/+}$ cells. I then tested several of the *Srsf2*$^{P95H/+}$ mis-spliced targets that had been previously reported. I performed isoform analysis using both splenocytes from *R26*-CreER^{T2} *Srsf2*$^{P95H/+}$ mice (20 weeks post tamoxifen cessation), as well as Hoxb8-immortalised *Srsf2*$^{P95H/+}$ cell lines (n=2 per genotype). The cell lines were generated by immortalising bone marrow cells from *R26*-CreER^{T2} *Srsf2*$^{P95H/+}$ and *R26*-CreER^{T2} *Srsf2*$^{+/+}$ mice with Hoxb8 retrovirus and are phenotypically mature myeloid cells, which I describe in Chapter 5.2.1 and Figure 5.1.

Consistent with the work of Kon et al., and using the same PCR primers as they had described, there was decreased exon inclusion in *Hnrnpa2b1*, *Trp53bp1* and *Usp4* in both *Srsf2*$^{P95H/+}$ sample types. In contrast, I did not detect any splicing changes of *Ezh2* or inclusion of the proposed 'poison exon' in *Ezh2*, as reported by Kim et al (Kim et al., 2015) (Figure 4.3C). In addition to these mis-spliced targets, the splicing of the receptor tyrosine phosphatase *Cd45* is regulated by SRSF2 during the T cell development and lack of SRSF2 protein leads to loss of exon 5 containing *Cd45* isoforms (Wang et al., 2001). This can be detected on the cell surface at the

protein level with an anti-CD45RB monoclonal antibody, which targets an epitope within exon 5. Both PCR and FACS results showed an increase in exon 5 exclusion in *Cd45*/CD45 in the *Srsf2*$^{P95H/+}$ samples (Figure 4.3D).

I also performed RNA-seq on an independent set of Hoxb8 immortalised *R26*-CreER^{T2} *Srsf2*$^{P95H/+}$ and *R26*-CreER^{T2} *Srsf2*$^{+/+}$ cell lines (7 days post tamoxifen cessation, n=3 per genotype). The culture of Hoxb8 immortalised cell lines was conducted along with the *Srsf2* CRISPR screen in Chapter 5.2.2 and Figure 5.2. There was no significant up- or down-regulation of gene expression in these cell lines (FDR<0.1, data not shown). By using rMATs splicing analysis, a total of 1,511 mis-spliced events were detected (FDR<0.05, inclusion=0.05). Skipped exons were the most frequently mis-spliced category of alternative splicing (910), followed by mutually exclusive exons (360) (Figure 4.4A-B). The *Srsf2*$^{P95H/+}$ cell lines are a much more homogeneous population than the primary cell population that was sequenced, which may explain the increase in the number of significantly mis-spliced events that could be detected.


A.
events
0
100
200
300
400
Skipped/Cassette Exon
Mutually Exclusive
Alt 5' donor
Alt 3' acceptor
Retained Intron
Excluded
Included
Total events
Included (n / %)
Excluded (n / %)
253 / 0.47
181 / 0.34
34 / 0.29
42 / 0.36
7 / 0.34
7 / 0.34
12 / 0.36
34 / 1.02
8 / 0.33
7 / 0.28
B.
Skipped (casette) exons
n=49397; =1826 (P<0.05); = 434 (q<0.05)
Mutually exclusive exons
n=11403; =821 (P<0.05); = 76 (q<0.05)
WT (+/+)
KI (P95H/+)
C.
Hnrnpa2b1
Trp53bp1
Usp4
Ezh2
(murine counterpart of human poison cassette exon)
Ezh2
(reported murine cassette exon)
Normalised Fold inclusio
WT
KI
Hoxb8 cell line
Spleen

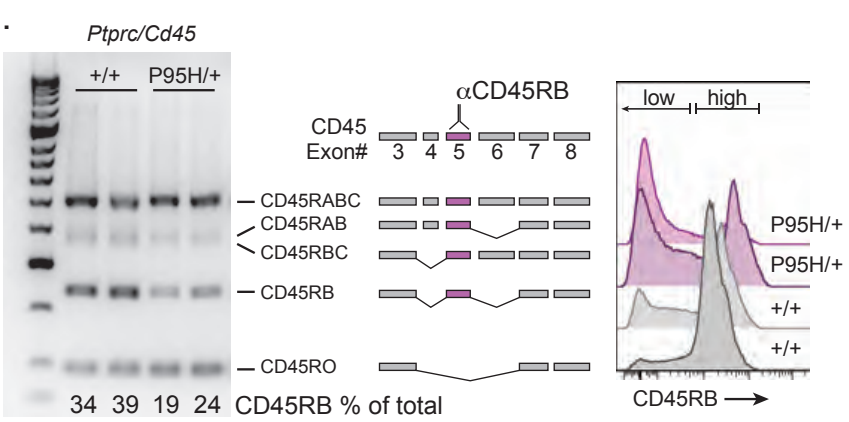


Figure 4.3: Skipped exons are the most frequently mis-spliced event in *Srsf2*$^{P95H/+}$ cells

A. The number and percentage of mis-spliced events by type in lin^{-} cKit^{+} eYFP^{+} cells 20 weeks after tamoxifen cessation (n=3 per genotype)

B. The inclusion events of cassette exons and mutually exclusive exons in *Srsf2*$^{+/+}$ and *Srsf2*$^{P95H/+}$ samples. Grey dots $\leq$ 5% difference between genotypes; blue dots = *P*<.05 and >5% difference between genotypes; red dots = q <.05 and >5% difference between genotypes.

C. Quantitative polymerase chain reaction (qPCR) analysis of alternative splicing of the indicated transcripts in splenocytes from tamoxifen treated *R26*-CreER^{T2} *Srsf2*$^{+/+}$ and *Srsf2*$^{P95H/+}$ animals or tamoxifen-treated Hoxb8 immortalised cell lines (n=2 per genotype).

D. *Ptprca/Cd45* isoform at transcript (left) and protein level of the CD45RB isoform (right) from splenocytes from tamoxifen treated *R26*-CreER^{T2} *Srsf2*$^{+/+}$ and *Srsf2*$^{P95H/+}$ animals.

Data expressed as normalised fold inclusion rate of the cassette exon compared with exclusion product (WT ratio normalized to 1).FC, fold change; FDR, false discovery rate.

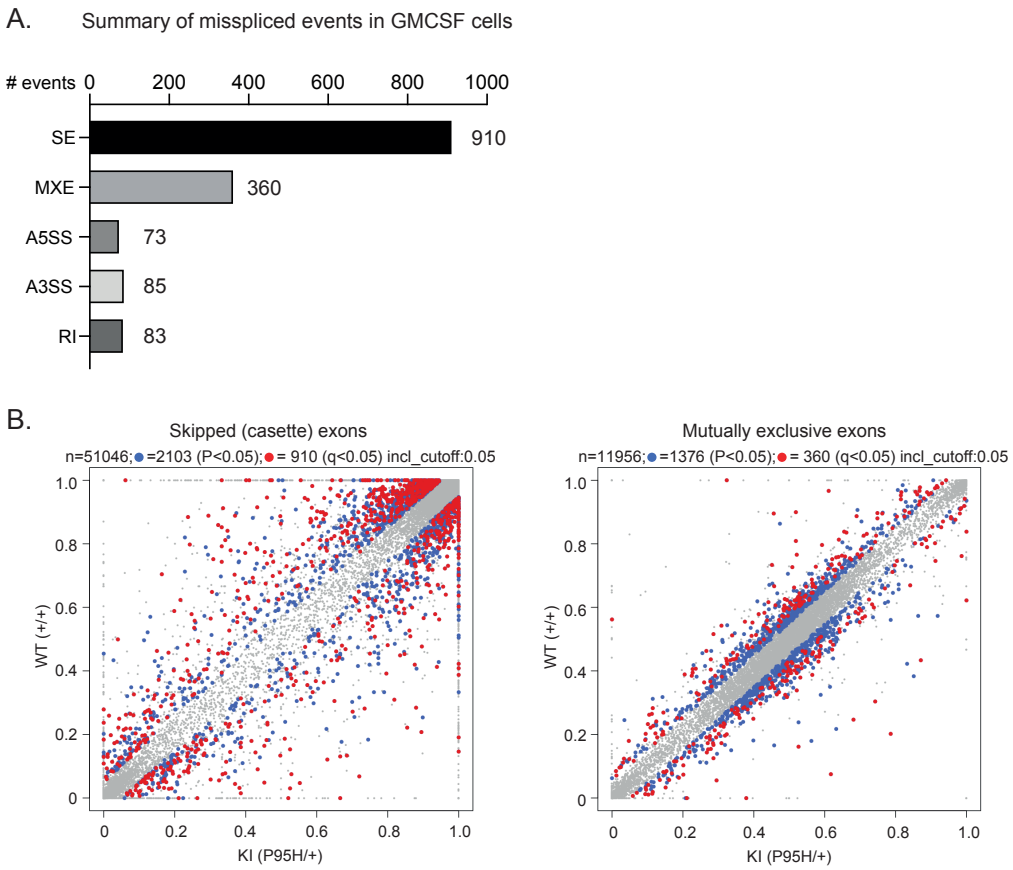


Figure 4.4: *In vitro* splicing changes caused by *Srsf2*$^{P95H/+}$ expression mirror the *in vivo* changes

A. Analysis of aberrant splicing events by type using Hoxb8-immortalised *Srsf2*$^{+/+}$ and *Srsf2*$^{P95H/+}$ cells (7 days after tamoxifen cessation; n=3 per genotype)

B. The inclusion events of cassette exons and mutually exclusive exons in *in vitro* *Srsf2*$^{+/+}$ and *Srsf2*$^{P95H/+}$ samples. Grey dots $\leq$ 5% difference between genotypes; blue dots = *P*<.05 and >5% difference between genotypes; red dots = q <.05 and >5% difference between genotypes.

SE, skipped exon; MXE, mutually exclusive exon; A5SS, alternative 5' splicing site; A3SS, alternative 3' splicing site; RI, retained intron.

4.2.2 mRNA processing and DNA repair are the top pathways represented in the mis-spliced transcripts in *SRSF2* mutant cells

Several studies have published analysis of mis-spliced events in *SRSF2* or *Srsf2* mutant cells from human AML, CMML or murine models, respectively (Hurtado et al., 2018; Kim et al., 2015; Komeno et al., 2015; Kon et al., 2018; Qiu et al., 2016; Shiozawa et al., 2018; Smeets et al., 2018; Zhang et al., 2015). The mis-spliced events identified in these studies reflect the cellular changes in each disease subtype and also revealed the biological differences between human and mouse. Nevertheless, identifying species conserved common mis-spliced events would provide crucial information on the pathways that are closely regulated by SRSF2. I compared the mis-spliced events from a total of 12 datasets, including the murine *Srsf2*$^{P95H/+}$ stem and progenitor cells and Hoxb8 immortalised *Srsf2*$^{P95H/+}$ cells that I had assessed (Table 4.1). I ranked the mis-spliced genes by the number of datasets they were detected in, and applied a cut-off for inclusion if present as mis-spliced in a minimum of 3 datasets. Using Metascape (Zhou et al., 2019), I performed pathway analysis and found that mRNA processing and DNA repair pathways were the top candidate pathways over-represented within the mis-spliced transcript sets from *Srsf2*$^{P95H/+}$ cells (Figure 4.5). Irrespective of the type of mis-spliced event, *Atf2* and *Mtmr14* were the most commonly mis-spliced transcripts (8/12 datasets), followed by *Csf3r, Dap3* and *Hnrnpa2b1* (Table 4.2).

Table 4.1: Datasets used for the comparison of *SRSF2/Srsf2* mis-spliced genes

Dataset description	Dataset URL	Database, license, and accessibility information	Reference
Human datasets			
mRNA splicing data from primary CMML (n=13; 3 with *SRSF2* mutation) and AML (n=9; 5 with *SRSF2* mutation) patients*	Dataset retrieved from table S1 in the supplement	Available in the supplement information table S1	(Kim et al., 2015)
mRNA profile of MDS-L cells (in triplicate) transfected with *Srsf2* P95H shRNA construct	https://www.ncbi.nlm.nih.gov /geo/query/acc.cgi?acc=GSE61052	Publicly available at Gene Expression Omnibus (GEO): GSE61052. Mis-spliced gene list available in the supplement information table S4	(Komeno et al., 2015)
mRNA profile of K562 CRISPR cell clones (with wild-type or mutant SRSF2, n=4 per genotype)	https://www.ncbi.nlm.nih.gov /geo/query/acc.cgi?acc=GSE71299	Publicly available at Gene Expression Omnibus (GEO): GSE71299 Mis-spliced gene list available in the supplement information table S1	(Zhang et al., 2015)
mRNA profile of primary MDS bone marrow samples (n=115)	Dataset retrieved from table S7 in the supplement	Available in the supplement information tables S7	(Qiu et al., 2016)
mRNA profile of primary CMML (n=20; 7 with *SRSF2* mutation) patients	Dataset retrieved from table SX-XIV in the supplement	Available in the supplement information tables SX-XIV	(Hurtado et al., 2018)

Murine datasets			
mRNA profiles of murine model (MP/LSK cells) expressing *Srsf2* WT and mutants*	https://www.ncbi.nlm.nih.gov /geo/query/acc.cgi?acc=GSE65349	Publicly available at Gene Expression Omnibus (GEO): GSE65349 Mis-spliced gene list available in the supplement information table S1	(Kim et al., 2015)
mRNA profile of murine model (MP/LSK cells) expressing *Srsf2* WT or *Srsf2*$^{P95H/+}$ *	http://trace.ddbj.nig.ac.jp /DRASearch/submission?acc= DRA006224	Publicly available at DNA Data Bank of Japan repository: DRA006224 Mis-splicing gene list available in supplemental table 2	(Kon et al., 2018)
mRNA profile of murine model (sorted LK+ eYFP+ cells) expressing *Srsf2* WT or *Srsf2*$^{P95H/+}$	https://www.ncbi.nlm.nih.gov /geo/query/acc.cgi?acc=GSE99852	Publicly available at Gene Expression Omnibus (GEO): GSE99852 Mis-splicing gene list available in supplemental dataset 2	(Smeets et al., 2018)
mRNA profile of Hoxb8 immortalised GM-CSF cell lines expressing *Srsf2* WT or *Srsf2*$^{P95H/+}$	N/A	Not publicly available	unpublished

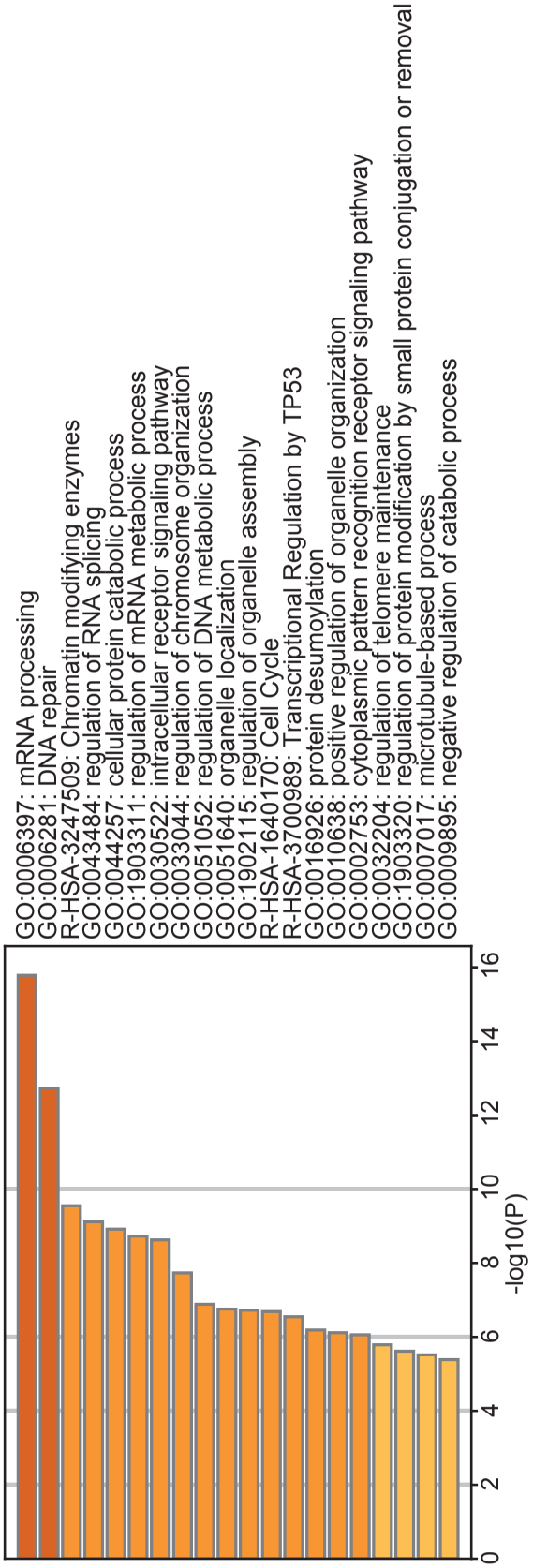


Figure 4.5: mRNA processing and DNA repair are the top mis-spliced pathways caused by *SRSF2* mutation
Pathway analysis of common mis-spliced genes among 12 datasets using Metascape.

Table 4.2: List of top mis-spliced genes ranked by the number of datasets the transcript was identified in

Gene	Description	Number of datasets
ATF2	activating transcription factor 2	8 (2H/6M)
MTMR14	myotubularin related protein 14	8 (4H/4M)
AGTPBP1	ATP/GTP binding protein 1	7 (2H/5M)
PRPSAP2	phosphoribosyl pyrophosphate synthetase-associated protein 2	7 (2H/5M)
CSF3R	colony stimulating factor 3 receptor (granulocyte)	7 (2H/5M)
DAP3	death associated protein 3	7 (1H/6M)
EGFL7	EGF-like-domain, multiple 7	7 (2H/5M)
EHMT2	euchromatic histone-lysine N-methyltransferase 2	7 (1H/6M)
HNRNPA2B1	heterogeneous nuclear ribonucleoprotein A2/B1	7 (1H/6M)
SEC24B	SEC24 family member B	7 (1H/6M)
TPM3	tropomyosin 3	7 (3H/4M)
ELP2	elongator acetyltransferase complex subunit 2	7 (2H/5M)
STRADA	STE20-related kinase adaptor alpha	7 (1H/6M)
SUN1	Sad1 and UNC84 domain containing 1	7 (2H/5M)
TIA1	TIA1 cytotoxic granule-associated RNA binding protein	7 (2H/5M)
MAPK8IP3	mitogen-activated protein kinase 8 interacting protein 3	7 (2H/5M)
PARP6	poly (ADP-ribose) polymerase family, member 6	7 (2H/5M)

H, Human datasets; M, Murine datasets.

4.2.3 Single cell sequencing of *Srsf2*$^{P95H/+}$ stem and progenitor populations reveals a unique signature in the LT-HSC

The application of single cell RNA sequencing (scRNA-seq) has dramatically reshaped the understanding of haematopoiesis in both homeostasis and perturbed settings. This powerful technology can differentiate the haematopoietic hierarchy based on transcriptome status. In combination with functional analysis, the information provided can aid the understanding of self-renewal and lineage commitment in stem and progenitor

cells. Here, I present a pilot study of the application of scRNA-seq to *Srsf2*$^{P95H/+}$ stem cells.

Purified long-term haematopoietic stem cells (LT-HSC, lin$^-$c-Kit$^+$Sca-1$^+$Flt3$^-$CD34$^-$) and short-term stem cells plus multipotent progenitors (referred to as progenitors, lin$^-$c-Kit$^+$Sca-1$^+$Flt3$^{+/-}$CD34$^+$) were sorted from *R26*-CreER^{T2} *Srsf2*$^{P95H/+}$ and *Srsf2*$^{+/+}$ mice (n=1 per genotype, 20 weeks post tamoxifen cessation, Figure 4.6). Sorted cells were processed using the Chromium Single Cell 3' Reagent Kit v2 and sequenced on the Illumina HiSeq 2500 by the Australian Genome Research Facility (AGRF). In total, 2,334 cells were sequenced across all samples. Both *Srsf2*$^{+/+}$ LT-HSC and progenitor samples have a much smaller number of cells sequenced than *Srsf2*$^{P95H/+}$ LT-HSC and progenitor samples (Table 4.3). Cells were clustered and annotated based on previously published marker genes (Cheng et al., 2019) (Figure 4.7A). This analysis was undertaken with Dr Davis McCarthy and Ms Ruqian Lyu in the Bioinformatics and Cellular Genomics Lab at SVI. Uniform Manifold Approximation and Projection (UMAP) plot analysis revealed the expected stem and progenitor populations within the pooled batch analysis. Based on the expression pattern of each dataset, the HSCs and repopulating MPPs (repMPPs) could be further subdivided into four smaller clusters (HSC_C1-C4 and repMPP_C1-C4). This is reflected on the UMAP plots of individual samples (Figure 4.7B).

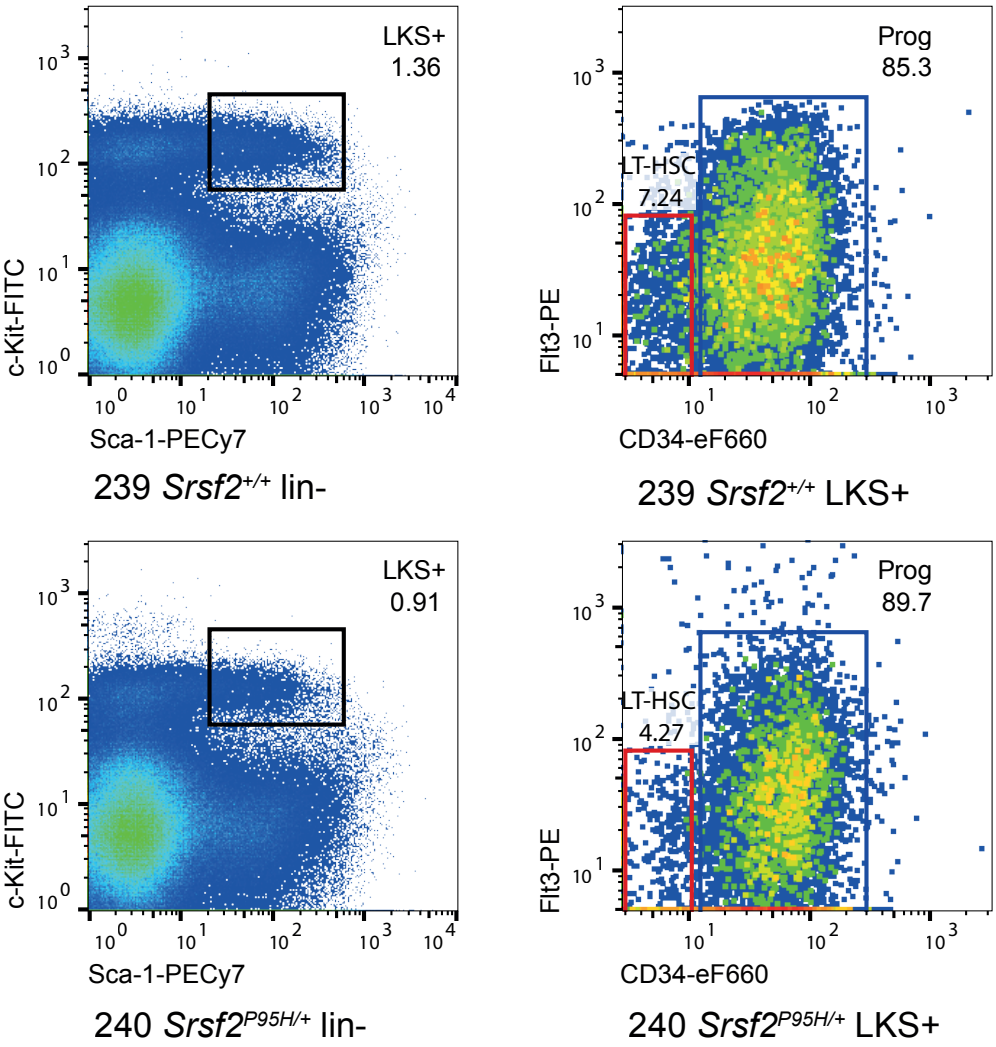


Figure 4.6: The sorting strategy of long-term haematopoietic stem cells and progenitor populations for single-cell sequencing

Table 4.3: Estimated number of cells processed for each sample

Sample	Cell number
Srsf2$^{+/+}$ LT-HSC (239)	164
Srsf2$^{+/+}$ Prog (239)	208
Srsf2$^{P95H/+}$ LT-HSC (240)	609
Srsf2$^{P95H/+}$ Prog (240)	1315

Table 4.4: The number of different cell types retrieved from each sample

Cell type	*Srsf2*$^{+/+}$ LT-HSC (239)	*Srsf2*$^{+/+}$ Prog (239)	*Srsf2*$^{P95H/+}$ LT-HSC (240)	*Srsf2*$^{P95H/+}$ Prog (240)
HSC_C1	90	2	92	7
HSC_C2	21	16	126	72
HSC_C3	9	20	22	241
HSC_C4	4	11	14	150
Erythroid	5	7	36	29
LMPP	1	44	7	261
MPP	1	23	10	139
repMPP_C1	0	11	12	153
repMPP_C2	4	35	1	0
repMPP_C3	4	8	8	30
repMPP_C4	0	6	5	47
GMP	0	1	40	10
Gr	0	0	15	5
Gr_Mo	0	3	11	16
B cell	6	0	115	5

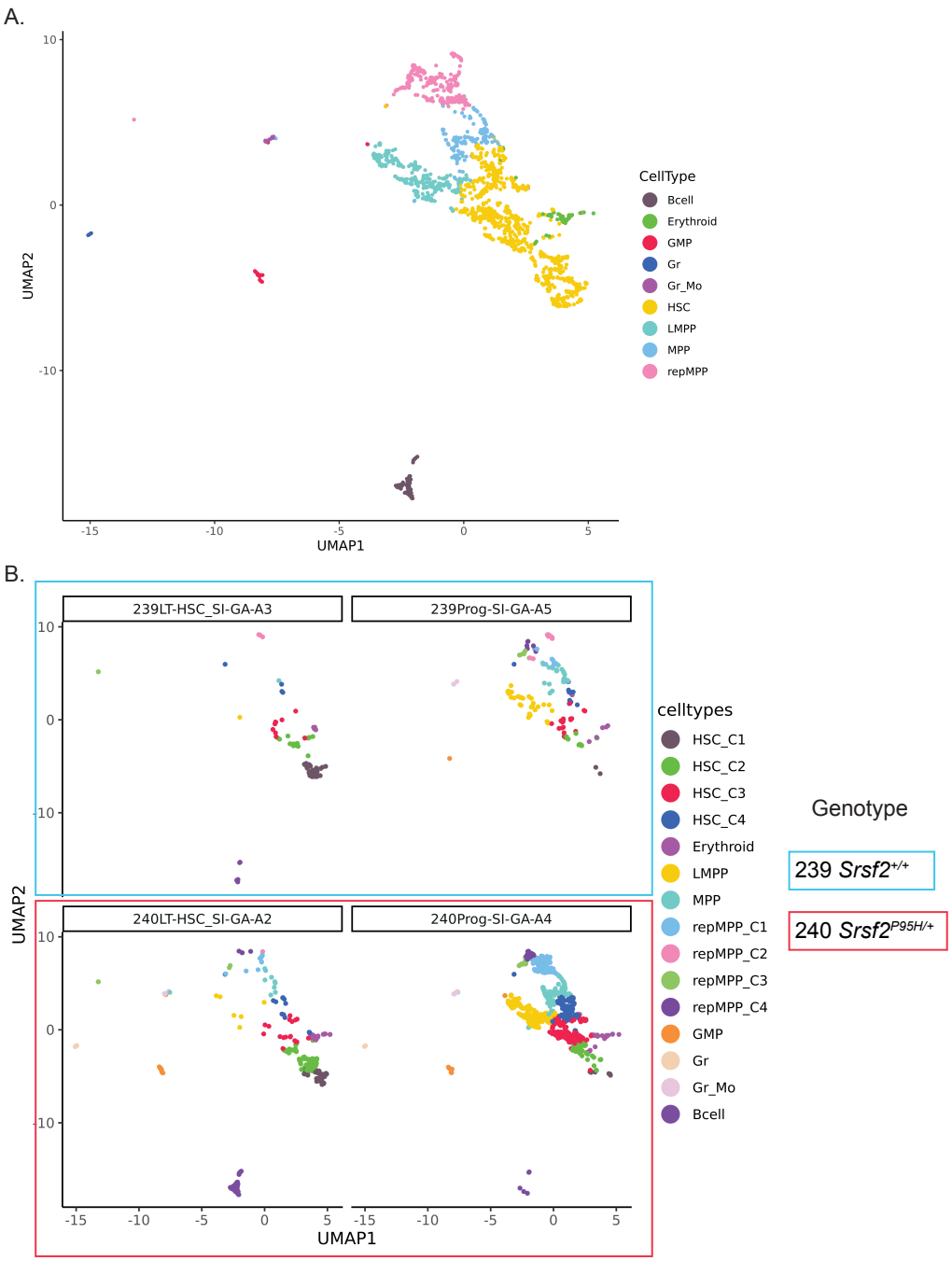


Figure 4.7: UMAP plots of all cells sequenced (A) and separate samples of *Srsf2*$^{+/+}$ (LT-HSC/Prog) and *Srsf2*$^{P95H/+}$ (LT-HSC/Prog) (B)
UMAP, Uniform Manifold Approximation and Projection; HSC, haematopoietic stem cells; LMPP, lymphoid primed multipotent progenitor; MPP, multipotent progenitor; repMPP, replicating multipotent progenitor; GMP, granulocyte-macrophage progenitor; Gr, granulocyte; Gr_Mo, granulocyte_monocyte.

Within the identified populations, there were comparable numbers of HSC_C1 clustered in both *Srsf2*$^{+/+}$ and *Srsf2*$^{P95H/+}$ LT-HSC samples, despite the lower number of cells from the *Srsf2*$^{+/+}$ sample. Interestingly, the repMPP_C2 population is almost exclusively from the *Srsf2*$^{+/+}$ sample (Figure 4.8A) (Table 4.4). On top of the stem and progenitors population, many LT-HSCs were labelled as B cells, which indicates the population bear the transcriptomic characteristics of more mature B cells. This could be due to contamination during cell sorting, but it is equally possible that these cells are primed for B lineage differentiation. Using a sling-Pseudotime analysis, the mutant HSC_C1 population displayed a faster transition into the HSC_C2 population than wild type (Figure 4.8B). This type of analysis features the temporal structure between the identified populations and infers a haematopoietic differentiation hierarchy based on the gene expression profile (Trapnell et al., 2014). HSC_C1 represents the most primitive haematopoietic stem cell population. Therefore, comparing the gene signatures of *Srsf2*$^{+/+}$ and *Srsf2*$^{P95H/+}$ HSC_C1 would reveal the effect of *Srsf2*$^{P95H/+}$ mutation on stem cell behaviour. Mutant stem cells indeed exhibit a distinct expression pattern similar to that of myeloid-biased LT-HSC (mLT-HSC) reported previously (Mann et al., 2018). This includes up-regulation of *Apoe* and *Plac8*, as well as down-regulation of *Serinc3* and *Cbx3* (Figure 4.8C). Finally, a trajectory analysis was performed to compare all the identified clusters between *Srsf2*$^{+/+}$ and *Srsf2*$^{P95H/+}$. Consistent with the myeloid bias observed in HSC_C1, *Srsf2*$^{P95H/+}$ cells had an increased differentiation trend towards myeloid progenitors when compared to *Srsf2*$^{+/+}$ (Figure 4.9). These findings from single cell transcriptome analysis are consistent with the phenotypic analysis that has been reported (Smeets et al., 2018).

A.
1.00
0.75
0.50
0.25
0.00
count
HSC_C1
HSC_C2
HSC_C3
HSC_C4
Erythroid
LMPP
MPP
repMPP_C1
repMPP_C2
repMPP_C3
repMPP_C4
GMP
Gr
Gr_Mo
Bcell
celltypes
B.
slingPseudotime_5
9
6
3
0
HSC_C1
HSC_C2
HSC_C3
HSC_C4
Erythroid
Sample group
239 Srsf2+/+
240 Srsf2P95H/+

Figure 4.8: Cluster comparison identifies a myeloid biased signature in *Srsf2*$^{P95H/+}$ haematopoietic stem cells
A. The proportion of cells from *Srsf2*$^{+/+}$ (239) and *Srsf2*$^{P95H/+}$ (240) within each cluster.
B. The slingPseudotime analysis
C. Heatmap of the top differentially expressed genes in HSC_C1 population between *Srsf2*$^{+/+}$ (239) and *Srsf2*$^{P95H/+}$ (240). Genes shaded in green belong to a myeloid-biased long-term haematopoietic stem cell signature. HSC, haematopoieitic stem cells.

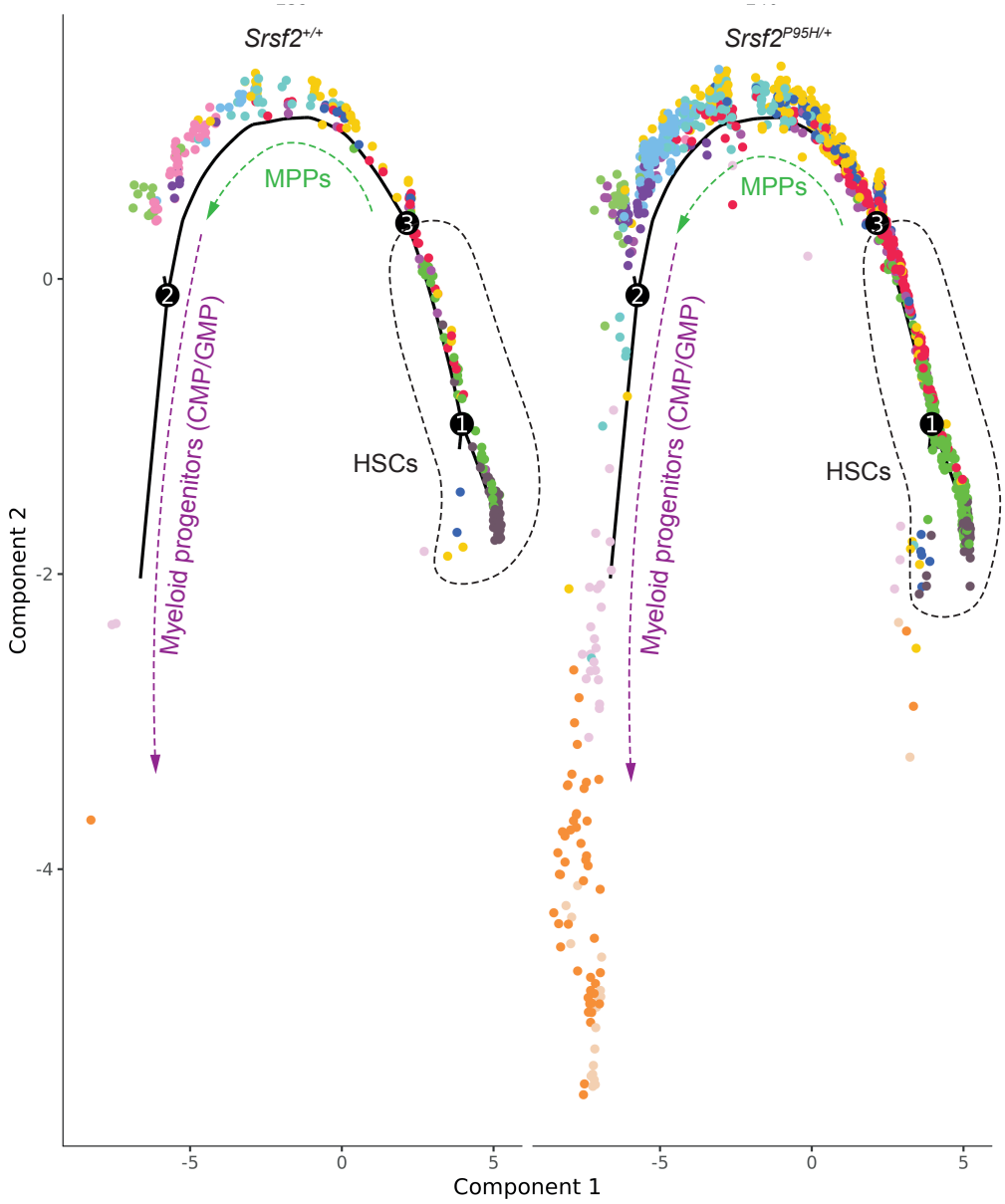


Figure 4.9: The trajectory plot of *Srsf2*$^{+/+}$ and *Srsf2*$^{P95H/+}$ samples reveals increased myeloid differentiation in the mutant cells
The arrows indicate the inferred differentiation trajectory.
HSC, haematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; GMP, granulocyte-macrophage progenitor.

4.3 Discussion

Transcriptome profiling is a great tool to gain insight into cellular behaviour. The differentially expressed genes (DEGs) reflect the response of essential pathways towards endogenous changes, such as disruptions caused by genetic mutations. In *Srsf2*$^{P95H/+}$ mutant cells, the up-regulation of myeloid genes and down-regulation of lymphoid genes corresponds to the myeloid-bias phenotype observed in the *in vivo* murine model (Smeets et al., 2018). As an RNA splicing factor, SRSF2 aids in the recognition of 3' splicing sites. SRSF2^{P95H} mutation specifically leads to a shift in the RNA binding affinity/preference, thereby potentially inducing mis-splicing (Kim et al., 2015; Zhang et al., 2015). Splicing changes in key regulatory genes were proposed to underlie the dysregulation of lineage-specific transcripts and cause the observed phenotype. Much interest has therefore focused on investigating the mis-splicing events in both human MDS samples and murine models.

The mis-splicing landscape in human SRSF2mut MDS cells has mostly revealed changes in metabolic pathways, including mRNA and cellular metabolism (Qiu et al., 2016). RNA splicing, processing and translation all fall within the mis-spliced mRNA metabolism category. Several mitochondrial related pathways were also shown to be involved (Pellagatti et al., 2018). For SRSF2mut MDS, *ATR, STAG2* and several cell cycle related genes (*CDK7, CCNL1, CARD16, SNCA, PYCARD*) were identified as mis-splicing targets in patients (Pellagatti et al., 2018; Qiu et al., 2016). Although some targets have been identified across multiple models (*Hnrnpa2b1/Trp53bp1*), the suggestion that *EZH2* is the single mis-spliced target that can account for a significant proportion of the phenotype of *SRSF2* mutation is more controversial (Kim et al., 2015; Kon et al., 2018; Smeets et al., 2018; Zhang et al., 2015). Both human studies

using K562 cells with the P95H mutation knocked into the endogenous locus conducted by Zhang et al, *Srsf2*$^{P95H/+}$ model from Kon et al., and my studies with primary and immortalised murine cells failed to demonstrate mis-splicing of *Ezh2* (Kon et al., 2018; Zhang et al., 2015). The magnitude of mis-splicing changes is only around 1% of the total splicing activity detected in the transcriptome, which is a relatively subtle overall change and it remains to be seen if one specific transcript or a small subset of mis-spliced transcripts are responsible for the pathogenesis of MDS in SRSF2 mutants. Indeed, very recently published RNA-seq analysis of >1,200 human MDS samples concluded that it was unlikely that a single transcript will be identified that accounts for phenotypes in spliceosome mutant myeloid neoplasms (Hershberger et al., 2020). It is possible that the disease is caused by the cumulative effect of many small splicing changes across essential cellular pathways in a 'death by a thousand cuts' analogy. However, confirming this would require very high-depth and high-fidelity sequencing of matched samples and multiple robust bioinformatic analysis pipelines, an undertaking that is yet to be reported.

Over the years, many datasets from SRSF2 mutant human myeloid leukaemia, or murine *Srsf2*$^{P95H/+}$ models have been generated. These datasets are from different disease backgrounds and cell types (Hurtado et al., 2018; Kim et al., 2015; Komeno et al., 2015; Kon et al., 2018; Qiu et al., 2016; Shiozawa et al., 2018; Smeets et al., 2018; Zhang et al., 2015). I hypothesized that the pathogenic network of *SRSF2* mutation might be detected as a pan-cancer signature of mutant SRSF2. Unlike introns that are more divergent across species, exonic sequences are highly conserved and targeted by SRSF2. Therefore, identifying the common mis-spliced genes that are shared between these datasets could unveil the crucial pathways regulated by SRSF2. Indeed, I found that mRNA processing and

DNA repair were the top two affected pathways. In terms of individual genes, activating transcription factor 2 (*ATF2*) was found to be mis-spliced in 8 out of the 12 datasets. Interestingly, ATF2 is a substrate of the p38 MAPK signalling pathway (Jiang et al., 1996). The p38 MAPK pathway is involved in normal haematopoiesis and is hyperactive in human MDS (Geest and Coffer, 2009; Navas et al., 2006). My pathway analysis of bulk cell RNA-seq from *Srsf2*$^{P95H/+}$ stem cells also revealed an upregulation of the p38 MAPK pathway. Understanding the structure of *ATF2* pre-mRNA could indicate why it is frequently mis-spliced by SRSF2^{P95H}. A more important question, however, is what these different and distinct isoforms are that are generated through mis-splicing and what the consequences of these could be at the functional/protein level. Clearly, not all the genes on the list of mis-spliced targets have a direct connection with haematopoiesis or leukaemia. Therefore, delineating the role of the top mis-spliced pathways, such DNA repair, is more likely to bridge the gap between the mechanism and the phenotypes observed in *Srsf2*$^{P95H/+}$ models.

While bulk RNA sequencing is frequently used to analyse the transcriptome and splicing changes of *Srsf2*$^{P95H/+}$ mutant cells, no study has examined the mutation via single cell sequencing. Here I present an initial single cell RNA seq (scRNA-seq) study of *Srsf2*$^{P95H/+}$ stem and progenitor cells. The single cell transcriptomic analysis examined the transcriptional state of the immunophenotypic homogeneous LT-HSC population and was able to subdivide it into four distinct LT-HSC subsets: HSC_C1 to C4. I discovered that the gene signature of the 20-week-old *Srsf2*$^{P95H/+}$ LT-HSC subset HSC_C1 is reminiscent of that of myeloid biased LT-HSC found in aged mice (20-24 months old) (Mann et al., 2018). The most immature *Srsf2*$^{P95H/+}$ LT-HSCs showed an accelerated differentiation toward

downstream stem and progenitor cell populations, particularly those that are myeloid biased/committed. This corresponds to the increased myeloid bias identified in the trajectory analysis. Combining all the information, the scRNA-seq study reinforced the hypothesis that *Srsf2*$^{P95H/+}$ mutation leads to transcriptome changes in the haematopoietic stem cell compartment, which drive a myeloid bias via cell-intrinsic mechanisms.

While scRNA-seq provided novel information regarding the stem cell compartment of *Srsf2*$^{P95H/+}$, there are major caveats to the pilot study. For instance, the overall number of *Srsf2*$^{+/+}$ cells sequenced is still low. In total, less number of wild-type cells were analysed due to insufficient number of reads mapped to the reference genome. This analysis only included one animal of each genotype. Independent replication is required to confirm the interpretation of the findings. An additional question remains about the best method to undertake such analysis. Here I have used 10X Chromium based library preparation to generate the libraries for sequencing. This was primarily due to the number of cells that can be assessed, ease of use and availability. This method has a 3' bias and is excellent for gene expression analysis. However, analysis of full-length transcriptomes and isoforms is difficult with the 10X methodology. Alternative methods that resolve fully length transcripts such as Smart-seq3 would allow a more detailed analysis of splicing and isoforms, albeit at the expense of total number of cells that can be processed due to the plate-based method required for Smart-seq3 and similar full-length transcriptome analysis methods (Hagemann-Jensen et al., 2020). Other technologies such as long-read and direct RNA sequencing are currently not feasible due to the requirement for large amounts of starting material. Despite the caveats, scRNA-seq technology remains a great tool to study haematopoietic stem cells and has great potential for studying splicing

mutant leukaemia.

Chapter 5

In vitro CRISPR screen identifies *Srsf2*$^{P95H/+}$ specific synthetic lethal pathways

5.1 Introduction

A synthetic lethal interaction involves two genes where mutation of either one separately is viable, but perturbation of both genes at the same time is lethal to the cell (Dhanjal et al., 2017). In cancer, the concept of synthetic lethality is very attractive as it enables the discovery of "second-hit" targets that the cells with a cancer-causing mutation depend on for survival. Thus, it provides opportunities to advance specific anti-cancer therapies, especially when the driver mutation is non-druggable. The key to successfully identifying such targets relies on a robust synthetic lethal genetic interaction (Wang et al., 2017). From RNAi (RNA interference) screens in yeast to CRISPR (Clustered regularly interspaced short palindromic repeats) screens in *in vivo* systems, larger scale and more clinically relevant models have been developed and have identified new targets in their respective cancers' settings (O'Neil et al., 2017).

CRISPR technology has been widely adopted for large-scale genetic screens. Conventional CRISPR-Cas9 (CRISPRko) gene editing leads to irreversible gene knockout by using specifically designed single-guide RNAs (sgRNAs) and Cas9 endonuclease (Shalem et al., 2014). The Cas9 endonuclease creates precise double-strand breaks (DSB) at target loci when guided by short RNAs. The DSB is mostly repaired through non homologous end-joining (NHEJ). NHEJ is error-prone, therefore creating indels that produce premature termination codons and frame shift mutations, resulting in non-sense mediated decay of the resulting transcript or non-functional proteins when translated (Cong et al., 2013; Mali et al., 2013). In addition, both CRISPR mediated transcriptional inhibition (CRISPRi) and CRISPR mediated transcriptional activation (CRISPRa) techniques have been developed by fusing either a transcriptional repressor domain (KRAB) or activator domain (VP64) to a catalytically dead

mutant of Cas9 (dCas9) (Gilbert et al., 2013; Qi et al., 2013). Guided by the sgRNA, the fusion protein then modifies the expression level of the target gene from its endogenous locus.

The sgRNA libraries are designed specifically for each CRISPR modification. For CRISPRko, Shalem et al. and Wang et al. first described a genome-wide CRISPRko library (GeCKO) that targeted over 18,000 genes in the human genome (Sanjana et al., 2014; Shalem et al., 2014; Wang et al., 2015; Wang et al., 2014). This was followed by Hart et al., who reported the Toronto KnockOut library (TKO). The same group, also introduced sub-pool libraries, targeting sets of genes with similar functions such as cell cycle proteins and kinases (Hart et al., 2015). For the mouse genome, Shalem et al., Wang et al. and Koike-Yusa et al. all described whole genome pooled libraries targeting nearly 20,000 genes (Koike-Yusa et al., 2014; Shalem et al., 2014; Wang et al., 2014). Most sgRNA libraries have become commercially available or are able to be accessed through repositories such as Addgene. Since these original genome-scale libraries, better performing libraries have been released with maximized on-target efficiencies and lower off-target effects (Doench et al., 2016).

With the advent of CRISPR technology, researchers have started to utilize genome-wide CRISPR screens to identify genetic vulnerabilities or mechanisms of drug resistance in multiple cancer cell lines. For haematological cancers, this was explored by Tzelepis et al. who looked at gene essentiality in 5 AML cell lines; MOLM-13, MV4-11, HL-60, OCI-AML2 and OCI-AML3. They performed a CRISPR screen and defined common and mutation-specific synthetic lethal genes (Tzelepis et al., 2016). Yamauchi et al. investigated the synthetic lethality of p53 WT AML using *in vitro* and *in vivo* screens (Yamauchi et al., 2018). Both studies discovered known therapeutic targets as well as new ones, such as KAT2A

and DCPS. More recently, Bajaj et al., developed an *in vivo* screen for a BCR-ABL/NUP98-HOXA9 driven model of blast-crisis chronic myeloid leukaemia (bcCML) and identified the RNA-binding protein STAUFEN2 as a key regulator of myeloid leukaemia (Bajaj et al., 2020). However, whole genome genetic interactions of *SRSF2* mutations are yet to be assessed or reported.

Here, I present an *in vitro* CRISPR knockout screen of *Srsf2*$^{P95H/+}$ cells to identify synthetic lethal interactions. I discovered that *Srsf2*$^{P95H/+}$ cells are more sensitive to perturbation of cell cycle and DNA repair pathways. Furthermore, pharmacological inhibition of CDK6, a cell cycle kinase, showed preferential suppression on *Srsf2*$^{P95H/+}$ cells.

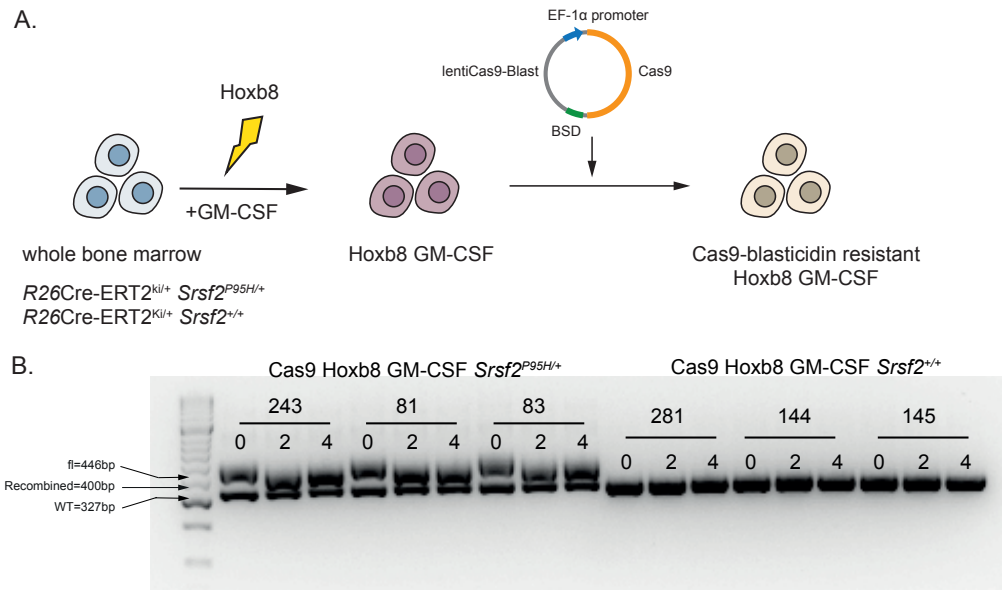
5.2 Results

5.2.1 Generation and characterization of *in vitro Srsf2* $^{P95H/+}$ cell lines

To explore the synthetic lethal genetic interactome of *Srsf2*$^{P95H/+}$ mutant cells *in vitro*, I utilized the Hoxb8 immortalized cell lines that I described in Chapter 4.2.1 (n=3 independently derived lines from 3 separate animals per genotype). To generate cells stably expressing Cas9, the cell lines were infected with Cas9-blasticidin lentivirus and then selected with 3μg/ml blasticidin. The cells were selected for a minimum of two weeks with blasticidin before characterization (Figure 5.1A).

The Cas9-blasticidin Hoxb8 cell lines (called Cas9 *Srsf2*$^{P95H/+}$ or *Srsf2*$^{+/+}$ cell lines from hereon) were treated with tamoxifen (400nM) for 4 days and then cultured without tamoxifen for an additional 14 days. The

floxed allele was fully recombined by day 4 of tamoxifen treatment, indicating activation of *Srsf2*$^{P95H/+}$ mutation (Figure 5.1B). FACS analysis of the Cas9 Hoxb8 cell lines confirmed that the cell lines are consistent with a granulocyte phenotype (Figure 5.1C). During the 14 days of culture, the Cas9 *Srsf2*$^{P95H/+}$ cell lines showed no difference in the proliferation rate compared to Cas9 *Srsf2*$^{+/+}$ cell lines (Figure 5.1D). To test Cas9 activity, the six cell lines were transfected with sgRNA targeting the *Ifih1* gene. Six days after transfection, RT-qPCR analysis showed a marked decrease in *Ifih1* expression, indicating functional Cas9 activity in the cells (Figure 5.1E).



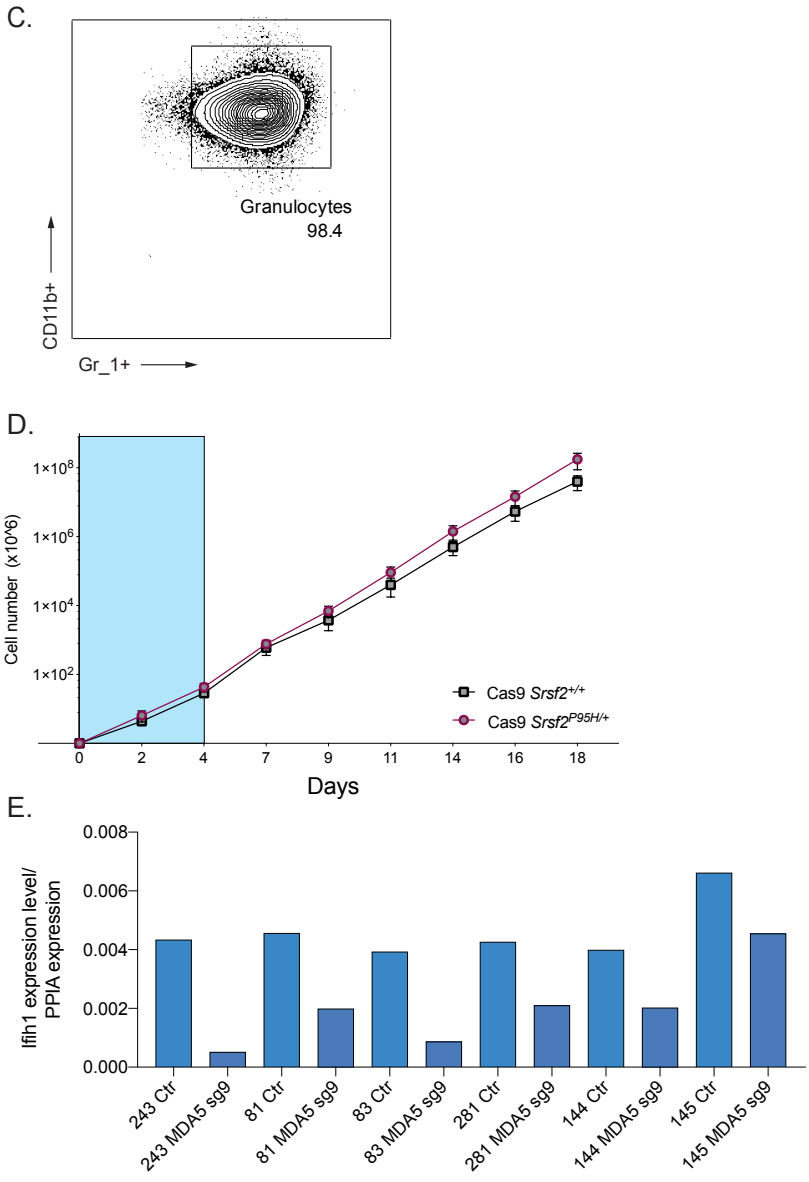


Figure 5.1: Generation and characterization of Hoxb8 immortalized cell lines and Cas9-blasticidin resistant cell lines

A. Generation of Hoxb8 immortalized GM-CSF cell lines and Cas9-blasticidin resistant cell lines.

B. Genomic DNA PCR showing full recombination of *Srsf2* P95H allele after 4 days of *in vitro* tamoxifen treatment.

C. Representative FACS plot of Cas9 Hoxb8-immortalized GM-CSF cells showing mature myeloid (granulocyte) phenotype.

D. Proliferation curve of *Srsf2*$^{+/+}$ and *Srsf2*$^{P95H/+}$ cell lines. Blue area indicates when cells were cultured with tamoxifen. n=3 per genotype.

E. RT-qPCR showing decreased expression of *Ifih1* 6 days after sgRNA transduction of Cas9 *Srsf2* cell lines.

5.2.2 Pooled CRISPRko negative selection screen in *Srsf2*$^{P95H/+}$ mutant cells

To investigate synthetic lethal genetic interactions with *Srsf2*$^{P95H/+}$, I designed a pooled CRISPRko negative selection (drop-out) screen using the Cas9 *Srsf2* cell lines (n=3 for *Srsf2*$^{P95H/+}$ and n=3 for *Srsf2*$^{+/+}$) (Figure 5.2). I made use of the mouse Brie library, which contains 78,637 sgRNA including 1,000 control sgRNA, and targets 19,674 genes with 4 unique sgRNA per gene (Doench et al., 2016). For each cell line, a total of 130 million Cas9 cells were infected with the Brie library at a multiplicity of infection (MOI) of 0.3 and cultured for 48hrs before adding puromycin (0.5μg/mL) and being selected for 4 days. I was aiming for an average copy number of 500 copies per sgRNA within the cell population. Cells were cultured in 3μg/mL blasticidin after infection and for the whole duration of the puromycin culture, to enforce Cas9 expression. The four-day puromycin treatment allows enough time for Cas9-mediated gene knockout to occur before activation of the *Srsf2*$^{P95H/+}$ mutation. This time point was designated as the day 0 timepoint. The achieved initial copy number of sgRNA from the cells was established by analysis of cells collected at this timepoint and served as the baseline sgRNA representation for the screen. All six cell lines were then treated with tamoxifen for 4 days to activate the *Srsf2*$^{P95H/+}$ allele in the mutant cells. After 4 days, the cells were washed to remove the tamoxifen and cultured for up to 14 days post-tamoxifen (Outline of screen in Figure 5.2). A full media change and dilution of the cells to a concentration of 200,000 or 100,000 cells/ml was performed every two to three days to avoid overgrowing and apoptosis due to cytokine deprivation.

Three additional timepoints, besides day 0, were selected to collect DNA from for sgRNA sequencing. Day 4 is when cells were taken off

tamoxifen, while days 11 and 18 are 7- and 14-days post tamoxifen, respectively. Choosing three different timepoints allows the identification of essential genetic interactions with *Srsf2*$^{P95H/+}$ immediately post mutation activation and after a short or more extended period of activation. For each timepoint, 40 million cells were collected to ensure a representation of a minimum of 500 copies per sgRNA, and all the DNA was amplified to avoid bias in quantitation of sgRNA. Finally, a total of 24 libraries (6 cell lines, 4 timepoints) were generated after gel-purification of PCR products and then sequenced (See Materials and Methods chapter for detailed description). In parallel to the CRISPR drop-out screen, I tamoxifen-treated non-sgRNA library infected Cas9 *Srsf2* cell lines and collected RNA at 7 days post tamoxifen treatment. These samples were used to undertake RNA-seq to establish gene expression and mis-splicing in these cell lines. This data was included in Chapter 4.2.1.

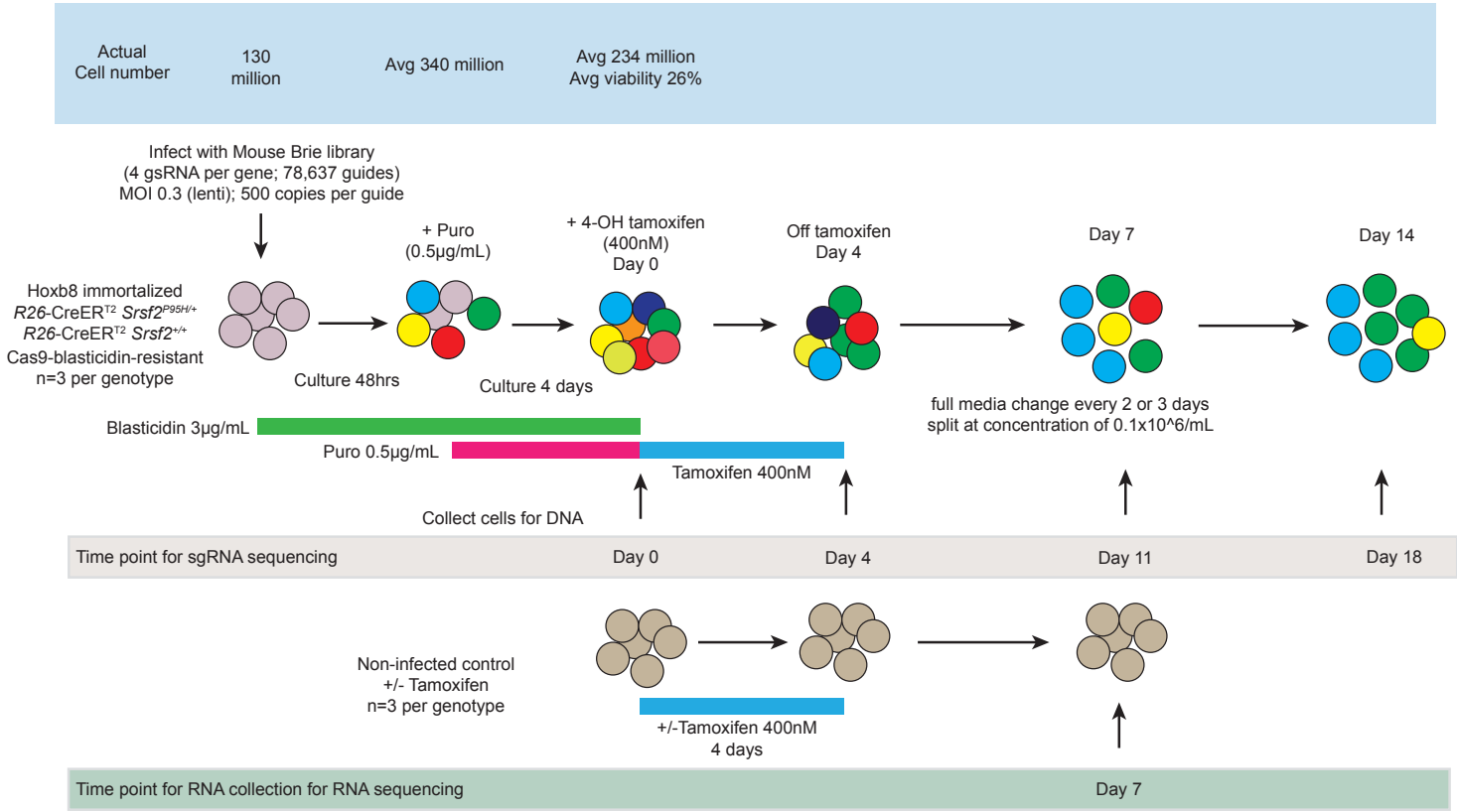


Figure 5.2: Outline of the pooled CRISPR knock-out *Srsf2* screen
Detailed schematic outline of the steps, time points and calculation of estimated cell numbers of the Srsf2 CRISPR screen. A parallel cell culture for RNA seq is included. Three biological replicates were used for both CRISPR screen and parallel cell culture for RNA seq.

5.2.3 CRISPR screen quality control and data analysis

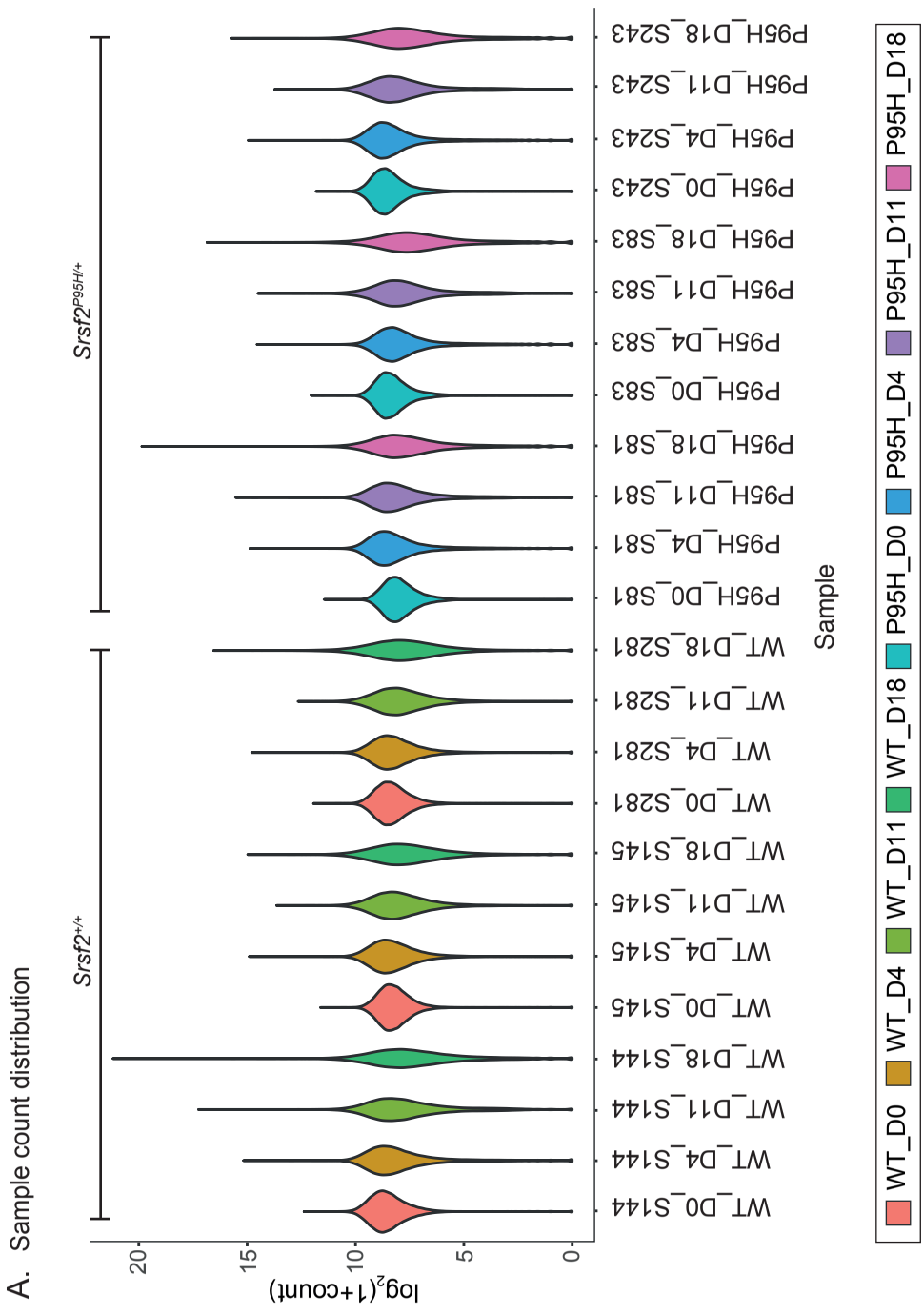
All the bioinformatic analysis of the screen was conducted by Dr. Alistair Chalk. After sequencing, FASTQ data were mapped to the reference gRNA library and analysed with the MaGeCK program (Li et al., 2014). The mapped reads ranged from 24 to 33 million per sample. On average, the mappability was 85% with an average coverage of 355 copies per sgRNA (Table 5.1). This is sufficient for analysis and to enable detection of those sgRNA that are lost following *Srsf2*$^{P95H/+}$ mutation activation, although lower than the original target of 500 copies per sgRNA. There was no evidence of misrepresentation of sgRNA due to skewed or inaccurate amplification or sequencing errors (Figure 5.3A), while the clonal dynamics of sgRNA were decreased over time as expected (Figure 5.3B). When examining the correlation between samples, the Cas9 *Srsf2*$^{+/+}$ cell lines exhibited more variance within the genotype. At day 0 and day 4, Cas9 *Srsf2*$^{+/+}$ achieved a 10% better correlation with Cas9 *Srsf2*$^{P95H/+}$ cell lines than with other cell lines of the same genotype. Over time, correlations between *Srsf2*$^{+/+}$ and *Srsf2*$^{P95H/+}$ decreased, indicating a divergent response to pooled genome knockout (Figure 5.4). The quality control results rule out the possibility of technical errors contributing to misleading results and confirm the validity of the negative enrichment CRISPR screen.

To corroborate the functional output of the screen, I then checked the overall positively and negatively selected genes in each *Srsf2*$^{+/+}$ or *Srsf2*$^{P95H/+}$ group. For the majority of the genes, 3 or 4 out of the 4 sgRNA per gene were recovered and aligned to the library sequence, which yields sufficient representation of the targeted genes for analysis and hit identification. During the MaGeCK analysis, enrichment data and statistical output from the 3/4 or 4/4 sgRNA were combined to present an

overview status of gene selection. The depletion of essential genes (CEG2) was assessed and mapped in the MaGeCK MLE visualisation plot (Hart et al., 2017) (Figure 5.5). In the visualisation plot of gene/sgRNA selection, a positive beta score means the gene is positively selected or enriched over time, whereas a negative beta score means the gene is negatively selected or lost over time (Li et al., 2015). At day 4, multiple genes involved in the essential ribosome pathway (*Rps15A, Rps7, Rpl23, Rplp2, Rps8* and *Rpl8*) were depleted. As for cell line-specific essential genes, *Hoxb8* and *Csf2ra* are among the top depleted sgRNA in both genotypes tested (highlighted in green in Figure 5.5). *Hoxb8* and *Csf2ra* (GM-CSF receptor) would be expected to be lost as the cell lines are immortalized with Hoxb8 and dependent on GM-CSF for proliferation and survival. Next, I assessed the selection of MDS related genes in the screen at day 4, day 11 and day 18. Genes that were depleted at day 4 include a selection of the spliceosome components (*Srsf2*, *Sf3b1* and *U2af1*), which is consistent with the lethality of null alleles of these essential factors. On the positive selection side, genes that are known to have a positive correlation with $Srsf2^{P95H/+}$, such as *Tet2* and *Cbl*, were enriched at the day 18 timepoint (Figure 5.6). The results from the visualisation plots confirm the depletion of expected essential and splicing-related genes, which validates the robustness of my *Srsf2* CRISPR screen.

Table 5.1: Mapped read count and coverage for 24 libraries in *Srsf2* screen

Sample_name	Mappability	Total_reads	Mapped_reads	Average Coverage
WT_D0_S144	85.72347	38583293	33074936	420.6
WT_D4_S144	85.87031	34396145	29536075	375.6
WT_D11_S144	84.44593	32585779	27517364	349.9
WT_D18_S144	87.12968	36646854	31930287	406.0
WT_D0_S145	86.08064	30683490	26412546	335.9
WT_D4_S145	86.90123	34345631	29846775	379.6
WT_D11_S145	84.85053	32027437	27175449	345.6
WT_D18_S145	86.29772	30632262	26434944	336.2
WT_D0_S281	85.63385	33078080	28326034	360.2
WT_D4_S281	86.1783	32858200	28316637	360.1
WT_D11_S281	85.12764	30345497	25832405	328.5
WT_D18_S281	86.19909	37822483	32602635	414.6
P95H_D0_S81	84.96886	26147819	22217505	282.5
P95H_D4_S81	84.81972	34788027	29507107	375.2
P95H_D11_S81	84.13214	32715519	27524267	350.0
P95H_D18_S81	86.06658	32080958	27610984	351.1
P95H_D0_S83	85.048	32903333	27983628	355.9
P95H_D4_S83	85.34779	28279733	24136126	306.9
P95H_D11_S83	84.42695	30106447	25417955	323.2
P95H_D18_S83	85.38084	28122359	24011106	305.3
P95H_D0_S243	85.01022	36751029	31242132	397.3
P95H_D4_S243	85.58722	36307614	31074678	395.2
P95H_D11_S243	85.15725	32322150	27524653	350.0
P95H_D18_S243	86.3542	28848293	24911713	316.8


A. Sample count distribution
Srsf2+/+
Srsf2P95H/+
log2(1+count)
0
5
10
15
20
WT_D0_S144
WT_D4_S144
WT_D11_S144
WT_D18_S144
WT_D0_S145
WT_D4_S145
WT_D11_S145
WT_D18_S145
WT_D0_S281
WT_D4_S281
WT_D11_S281
WT_D18_S281
P95H_D0_S81
P95H_D4_S81
P95H_D11_S81
P95H_D18_S81
P95H_D0_S83
P95H_D4_S83
P95H_D11_S83
P95H_D18_S83
P95H_D0_S243
P95H_D4_S243
P95H_D11_S243
P95H_D18_S243
Sample
WT_D0
WT_D4
WT_D11
WT_D18
P95H_D0
P95H_D4
P95H_D11
P95H_D18

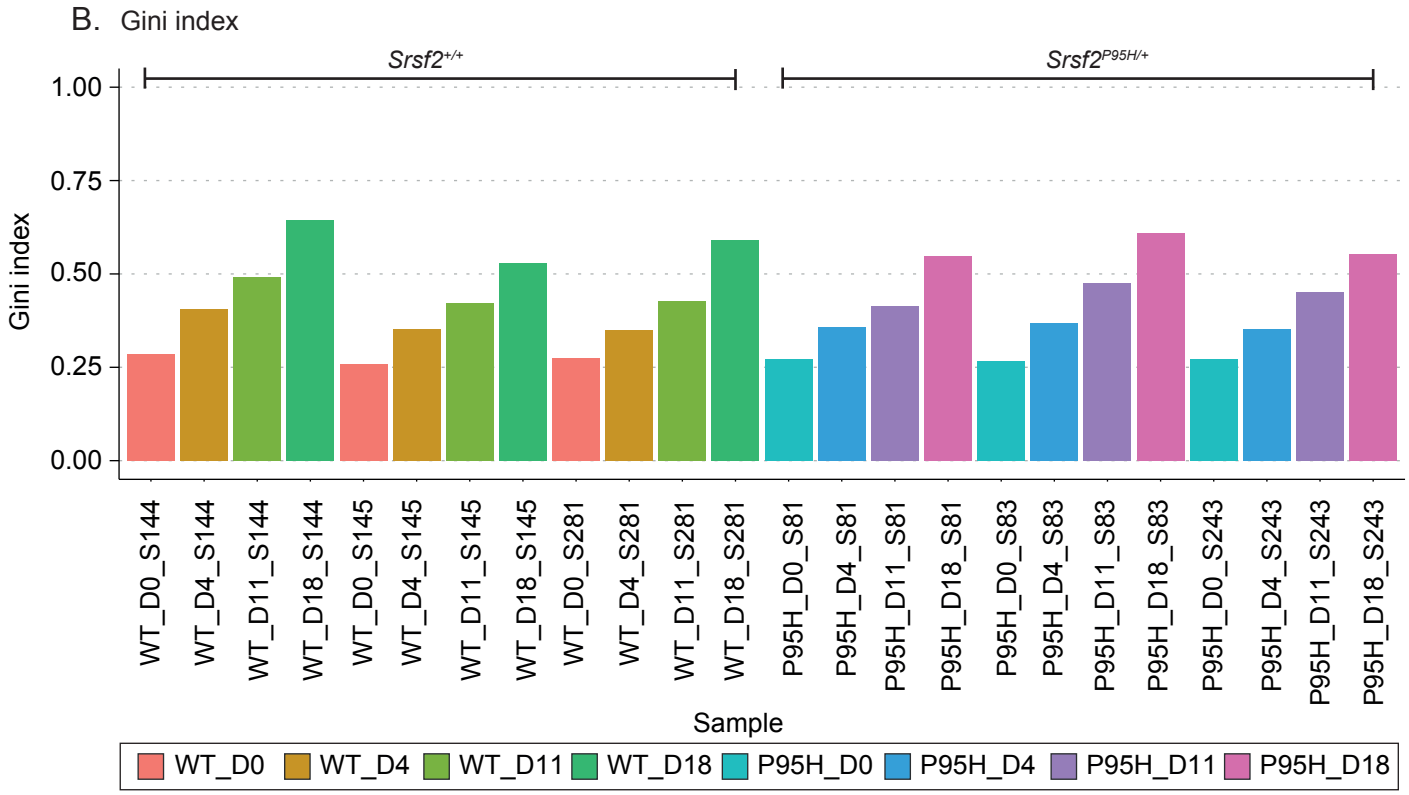


Figure 5.3: The sgRNA distribution and dynamics of the sgRNA pool in the *Srsf2* CRISPR screen
A. The sample count distribution of sgRNA of all 24 libraries showing unbiased representation of sgRNA.
B. Gini index of sgRNA in 24 libraries indicating fewer clones of the sgRNA pool as the screen progresses.

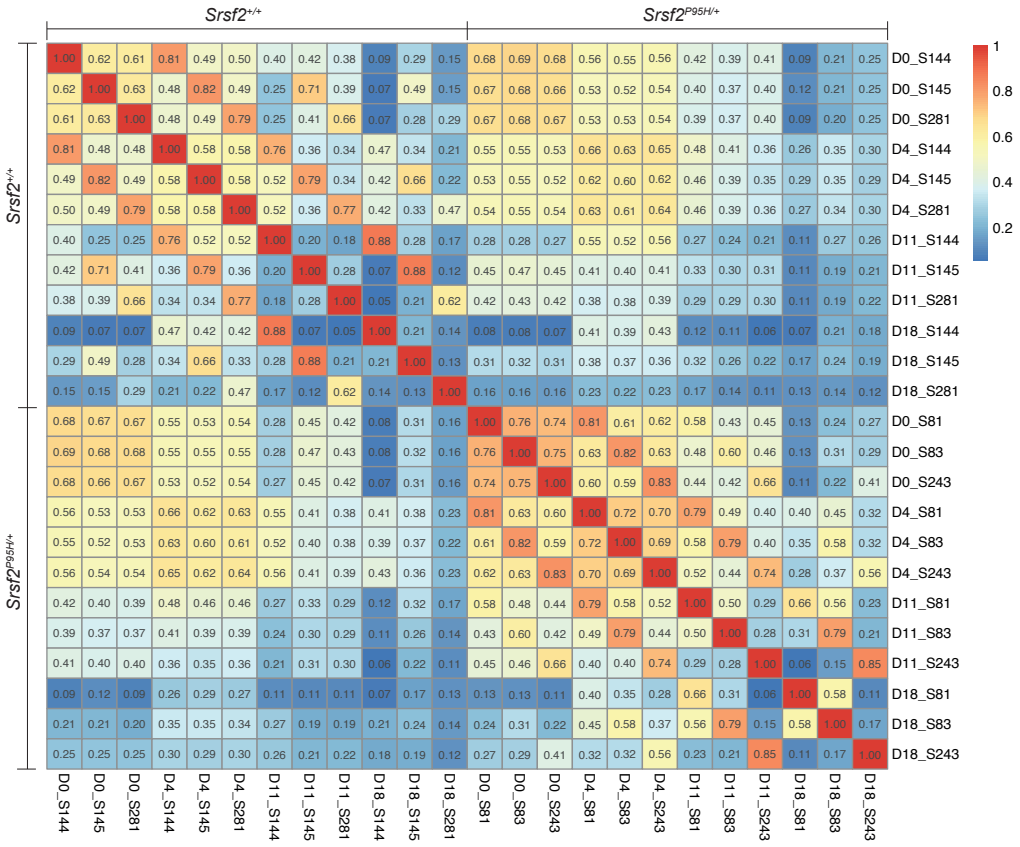


Figure 5.4: The correlation between the six Cas9 *Srsf2* cell lines across the time course of the screen

The colour code and number in each box indicates the concordance between two samples.The colour scale is indicated at the top right. A red colour and number close to 1 means two samples are tightly related. In contrast, a blue colour and number close to 0 means two samples have no correlation.

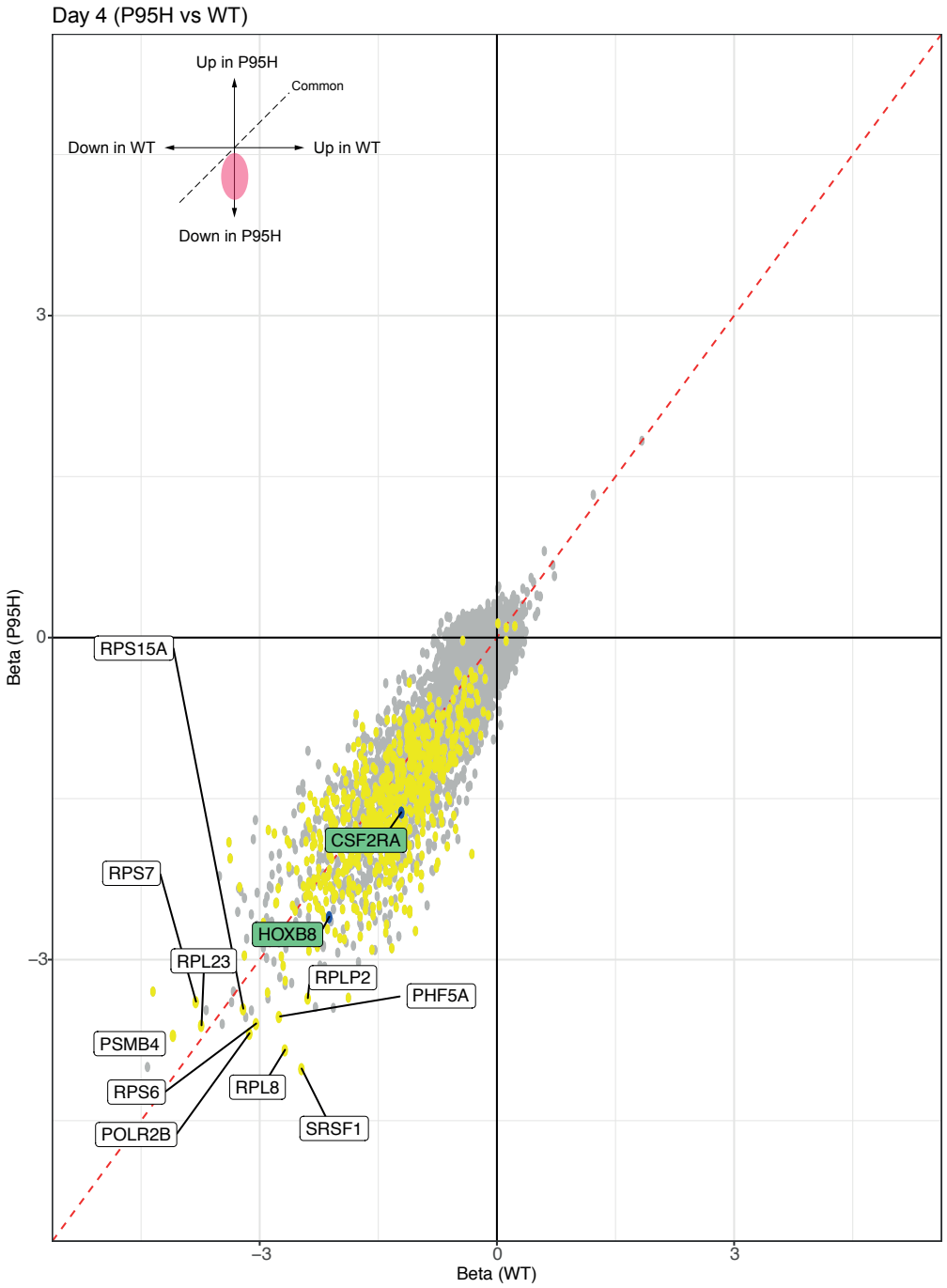
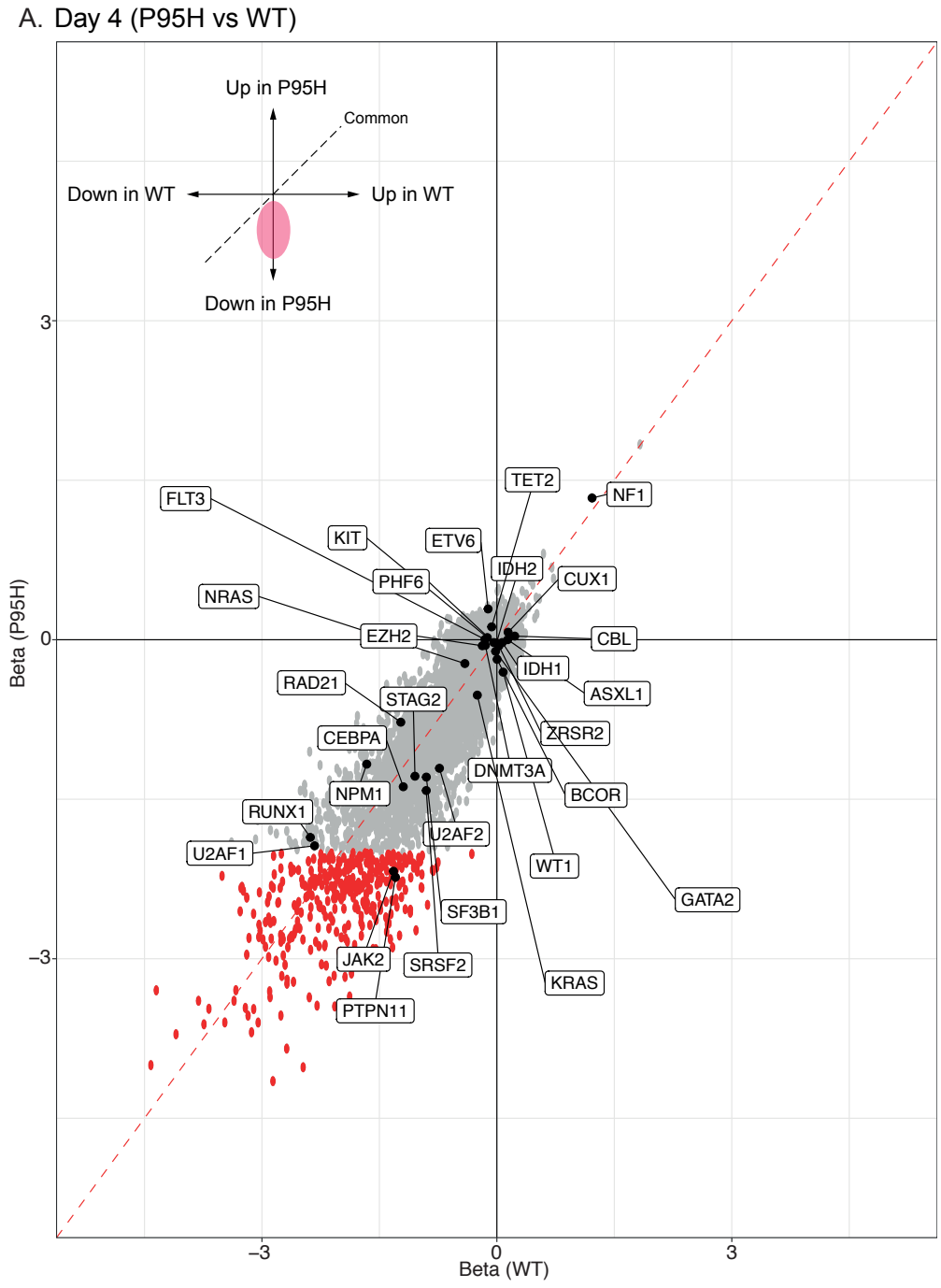
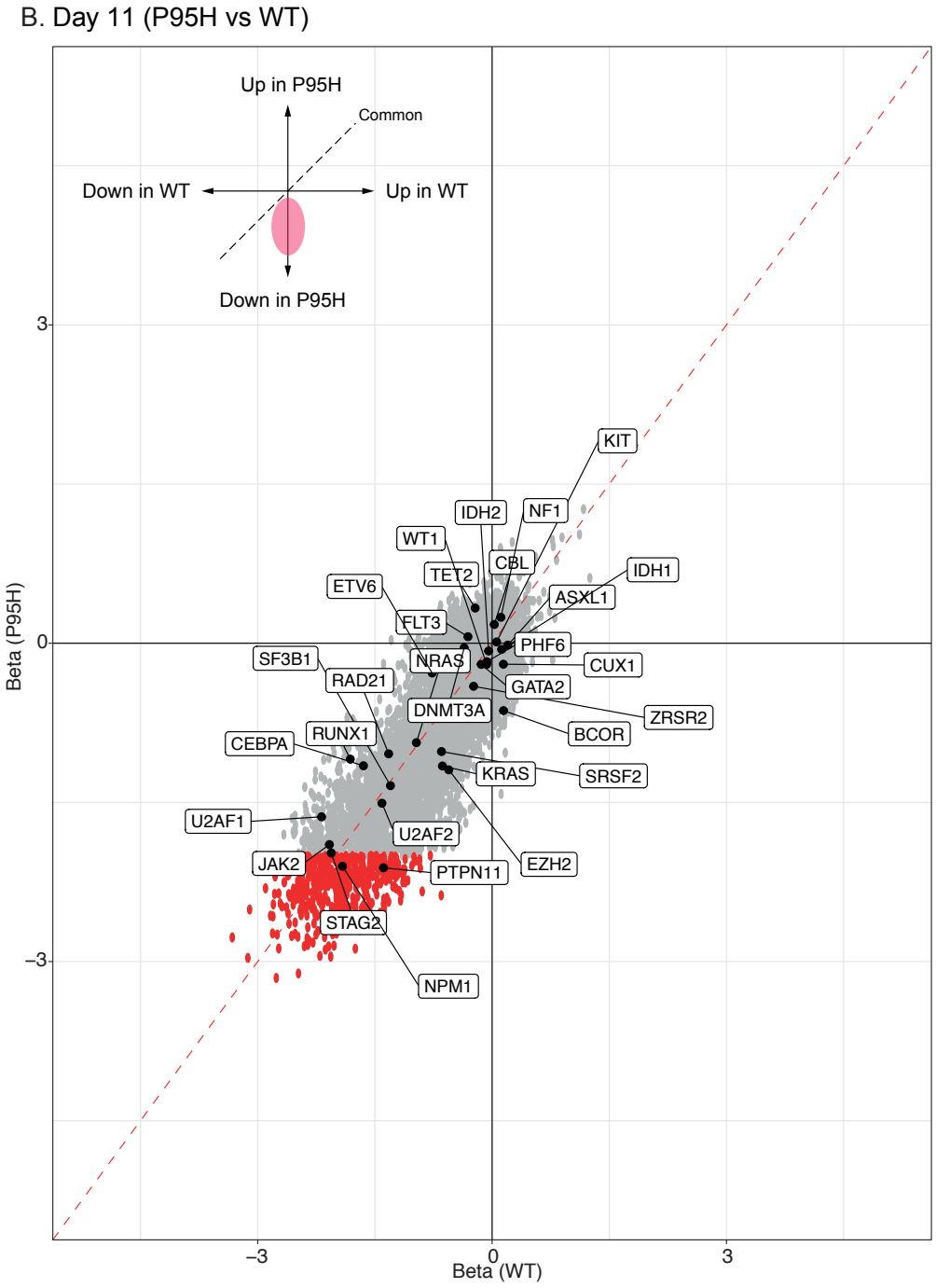


Figure 5.5: Depletion of the CEG essential genes at day 4 of the CRISPR screen

The top 10 depleted CEG2 essential genes are labeled. *Hoxb8* and *Csf2ra* are highlighted in green and marked in blue dots. Grey dots: all the sgRNA identified; Yellow dots: sgRNA of CEG2 essential genes.


A. Day 4 (P95H vs WT)
Up in P95H
Common
Down in WT
Up in WT
Down in P95H
Beta (P95H)
Beta (WT)
3
0
−3
−3
0
3
FLT3
KIT
ETV6
TET2
NF1
IDH2
CUX1
PHF6
NRAS
EZH2
CBL
IDH1
ASXL1
RAD21
STAG2
ZRSR2
CEBPA
DNMT3A
BCOR
NPM1
RUNX1
U2AF2
U2AF1
WT1
GATA2
SF3B1
JAK2
SRSF2
KRAS
PTPN11


B. Day 11 (P95H vs WT)
Up in P95H
Common
Down in WT
Up in WT
Down in P95H
Beta (P95H)
Beta (WT)
3
0
−3
−3
0
3
KIT
IDH2
NF1
WT1
CBL
IDH1
TET2
ETV6
ASXL1
FLT3
PHF6
SF3B1
NRAS
CUX1
RAD21
GATA2
DNMT3A
ZRSR2
RUNX1
BCOR
CEBPA
KRAS
SRSF2
U2AF1
U2AF2
EZH2
JAK2
PTPN11
STAG2
NPM1

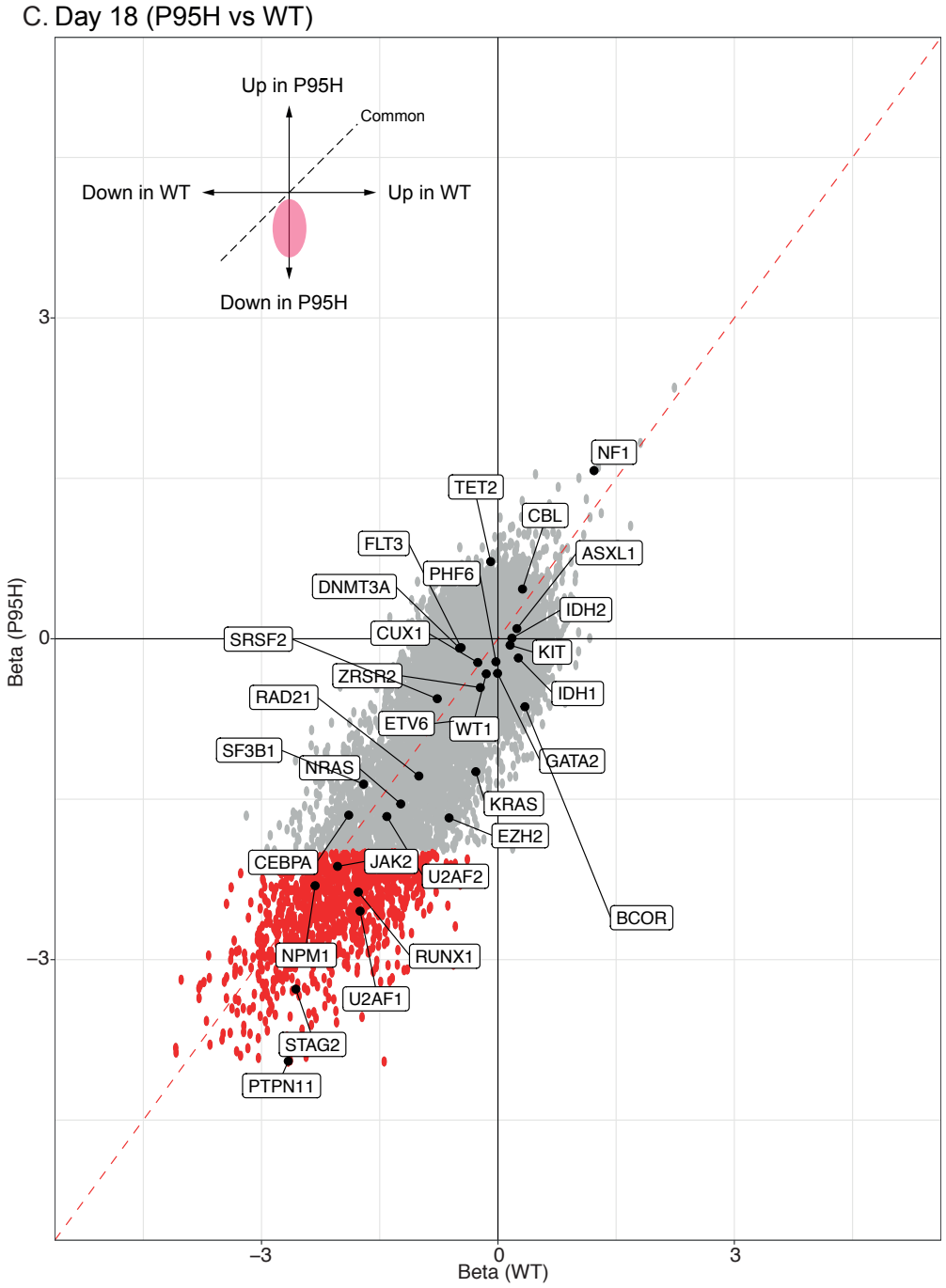


Figure 5.6: Visualisation of MDS related genes that are positively and negatively selected at day 4 (A), day 11 (B) and day 18 (C) of *Srsf2* CRISPR screen.
Both enriched and depleted genes that are associated with MDS are labeled. Grey dots: all the sgRNA identified; Red dots: sgRNA with false discovery rate (FDR) <.01.

5.2.4 Identification of RNA metabolism, Cell cycle and DNA damage response pathways as genetic vulnerabilities in *Srsf2*$^{P95H/+}$ mutant cells

To distinguish genes that are essential for *Srsf2*$^{P95H/+}$ but not *Srsf2*$^{+/+}$ cells, I first applied an FDR cut-off of <0.05 based on the FDR at the P95H_D18 timepoint for all the negatively selected genes. This removed all the genes that were not negatively selected in the *Srsf2*$^{P95H/+}$ cells 14 days after tamoxifen treatment. This included genes that were only negatively selected at any earlier timepoint or in the *Srsf2*$^{+/+}$ cells. I then applied an FDR cut-off of <0.01 for each individual timepoint as a more stringent criteria for inclusion of candidates. Overall, significantly more genes were lost in the *Srsf2*$^{P95H/+}$ cells than in the *Srsf2*$^{+/+}$ cells. A total of 762 genes were negatively selected in the *Srsf2*$^{P95H/+}$ cell lines, compared to 235 in the *Srsf2*$^{+/+}$ cell lines. Of the candidate genes, 200 were depleted in both cell lines, leaving 562 genes unique to the *Srsf2*$^{P95H/+}$ genotype and only 35 to *Srsf2*$^{+/+}$ (Figure 5.7A). Looking at the distribution within each genotype, 91, 155 and 92 genes were negatively selected in *Srsf2*$^{+/+}$ cells at day 4, day 11 and day 18 respectively, with 20 genes overlapping between timepoints. In contrast, in *Srsf2*$^{P95H/+}$ cell lines, 513, 515 and 359 genes were negatively selected at day 4, day 11 and day 18 respectively, with 180 genes overlapping between timepoints (Figure 5.7B-C).

Comparing *Srsf2*$^{P95H/+}$ with *Srsf2*$^{+/+}$ cell lines at each timepoint, 441 (day 4), 406 (day 11) and 308 (day 18) genes were exclusively depleted in mutant cells. These numbers reduced substantially when the FDR for *Srsf2*$^{+/+}$ was increased from <0.01 to <0.05, while maintaining an FDR of <0.01 for *Srsf2*$^{P95H/+}$ cells. The relaxed criteria expanded the number of candidate genes in wild type cells, while decreasing the number in the *Srsf2*$^{P95H/+}$ only group. Therefore, these filtered genes are closely linked

with *Srsf2*$^{P95H/+}$ synthetic lethality (Figure 5.8). The downward trend in the total number of negatively selected candidates across the timepoints suggested that many of the genes are synthetic lethal early after *Srsf2*$^{P95H/+}$ mutation expression, while deletion of a subset of genes led to a slow, gradual death of mutant cells and was not significant until the end of the screen. The gene lists from *Srsf2*$^{+/+}$ (FDR<0.01) and *Srsf2*$^{P95H/+}$ (FDR<0.01) are used to identify the top negatively selected pathways in Figure 5.9.

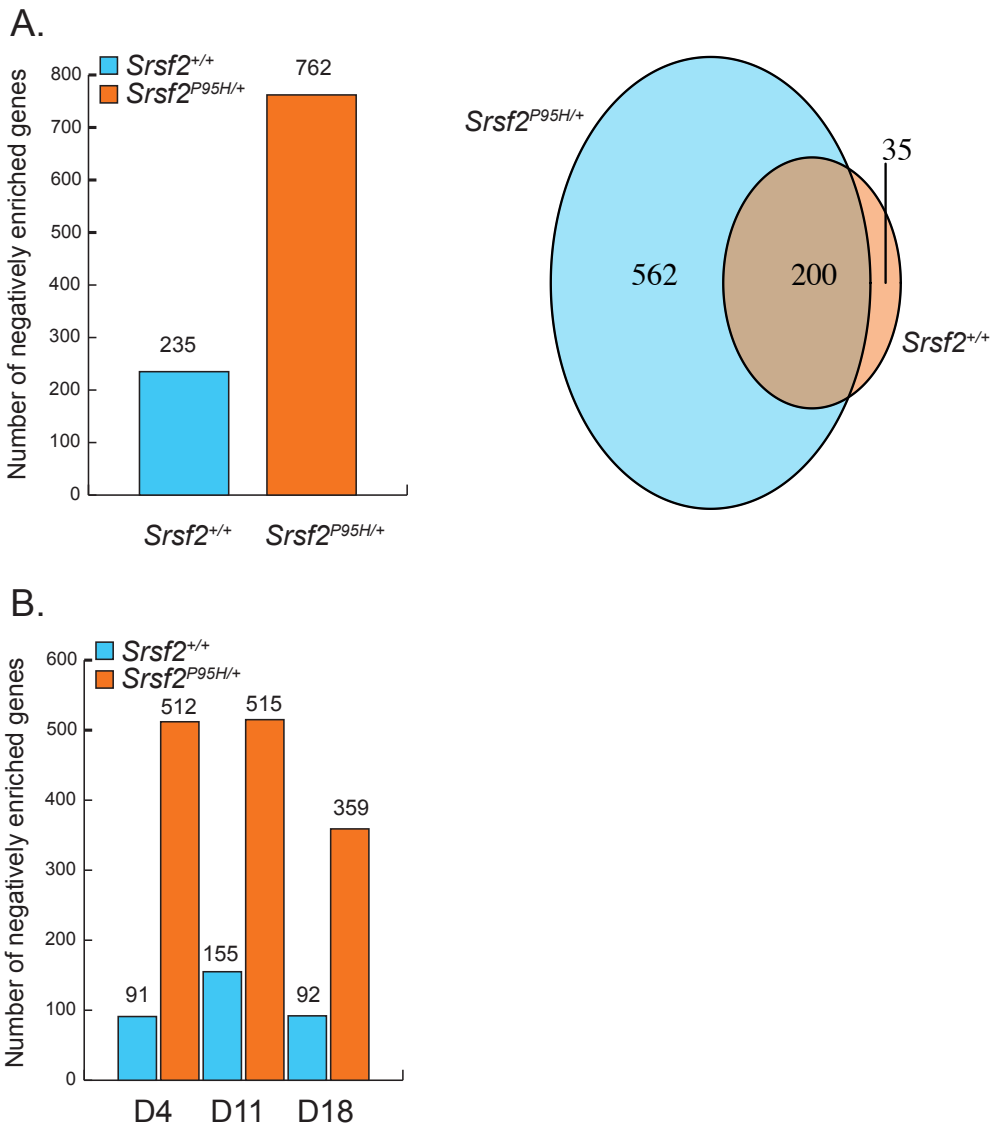
Many of the top genes that were unique to *Srsf2*$^{P95H/+}$ cells are components of pathways essential for cellular function, such as ribosome biosynthesis and DNA transcription. To provide a broader understanding of the genetic vulnerabilities of *Srsf2*$^{P95H/+}$, I decided to look at the pathways that were represented within the depleted candidates in *Srsf2*$^{P95H/+}$ cells. To do this, I created gene lists of those candidates that were negatively selected in *Srsf2*$^{+/+}$ and *Srsf2*$^{P95H/+}$ at each timepoint (FDR<0.01) and performed analysis using Metascape (Zhou et al., 2019). The top 20 most negatively selected pathways are presented in a heatmap in Figure 5.9. Overall, the top 20 pathways were significant at most timepoints in *Srsf2*$^{P95H/+}$ cells (P<0.0001), while 9 out of 20 did not reach significance at one or more timepoints in *Srsf2*$^{+/+}$ cells. The genes were also compared with a short list of significantly up- or down-regulated genes from RNA-seq of Cas9 *Srsf2* cell lines (Figure 5.9).

To achieve a more detailed comparison, I clustered the top pathways into three groups, depending on whether the pathway was significant in *Srsf2*$^{+/+}$ at all timepoints, at some timepoints or not at all. The common cluster, which contains pathways significantly selected in both genotypes at all timepoints, is the biggest (11/20) and includes pathways such as RNA metabolism, peptide biosynthesis, ribosome pathways, cell cycle and DNA repair. The partial WT cluster contains 7 out of 20 pathways,

which are only significantly selected at one or two timepoints in *Srsf2*$^{+/+}$. These include DNA replication, mRNA splicing, telomere maintenance and protein-DNA subunit organisation. Only two pathways belong to the P95H only cluster (2/20), which is specific to the *Srsf2*$^{P95H/+}$ cells. These are ribosomal small subunit biogenesis and processing of capped intronless Pre-mRNA. Within the other two clusters, *Srsf2*$^{P95H/+}$ cells showed a more significant pathway enrichment, suggesting that the mutant is more vulnerable to disruption of these pathways, even though *Srsf2*$^{+/+}$ cell viability would be impacted as well. A good example is the mRNA splicing pathway from the partial WT cluster, which is consistent with previous studies that splicing mutants have increased sensitivity to further splicing perturbation (Bonnal et al., 2012; Dvinge et al., 2016). Looking at the timing, pathways were most negatively selected at day 11 for *Srsf2*$^{+/+}$ cells and day 4 for *Srsf2*$^{P95H/+}$ cells. Finally, pathways that were negatively selected did not include genes that were differentially expressed in the cell lines. This suggests that the increased lethality of the mutant was not simply due to further disruption of already dysregulated pathways (Figure 5.9).

To identify the top negatively selected genes, the gene lists from *Srsf2*$^{+/+}$ (FDR<0.05) and *Srsf2*$^{P95H/+}$ (FDR<0.01) were applied. The number of genes that were uniquely negatively selected in *Srsf2*$^{P95H/+}$ were 303 (day 4), 231 (day 11) and 169 (day 18) (Figure 5.10A). The top 30 candidates for each timepoint were listed based on MaGeCK scoring and ranked from the lowest to the highest (Figure 5.10B). The lower the MaGeCK score, the more negatively selected the gene. *Npm1* is highlighted in the top genes at day 11, because of its association with the ribosomal small subunit biogenesis pathway, which I will discuss in Chapter 5.3. *Pkmyt1* and *Cdk6* are highlighted in the top genes at day 18, as they are two targets

selected for further pharmacological validation (Chapter 5.2.5)



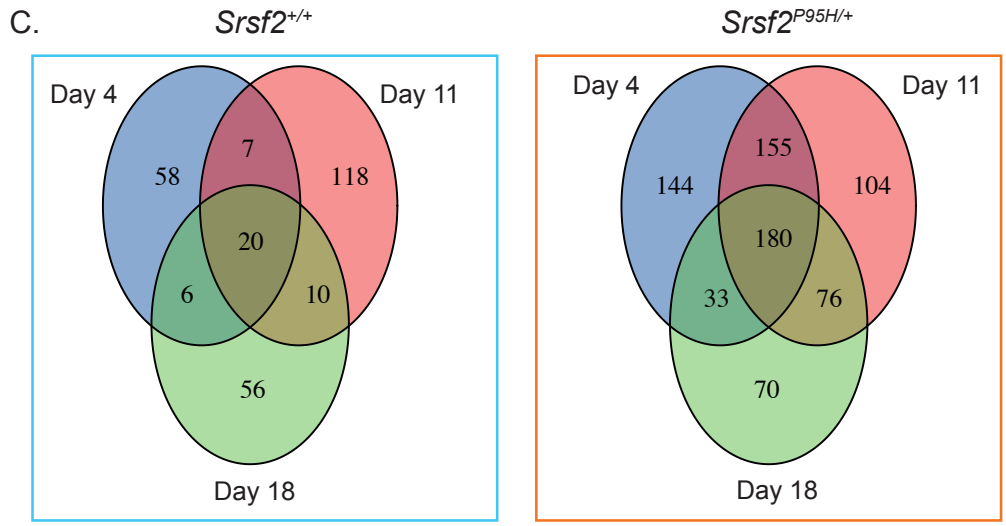


Figure 5.7: Overview of *Srsf2* negative selection CRISPR screen results
A. The total number of genes that were negatively selected in the screen. Results are presented in bar graph (on the left), as well as a Venn diagram (on the right), which shows that the majority of *Srsf2*$^{+/+}$ negatively selected genes overlap with *Srsf2*$^{P95H/+}$.
B. The number of genes that were negatively selected in *Srsf2*$^{+/+}$ and *Srsf2*$^{P95H/+}$ cell lines at each time point.
C. Venn diagrams showing the number of genes overlapping between day 4, day 11 and day 18 within *Srsf2*$^{+/+}$ and *Srsf2*$^{P95H/+}$ cells.
Genes are filtered at P95H_D18 time point at FDR <.05. Then genes are filtered at each time point individually for *Srsf2*$^{+/+}$ and *Srsf2*$^{P95H/+}$ cell lines, respectively. There are three biological replicates per genotype.

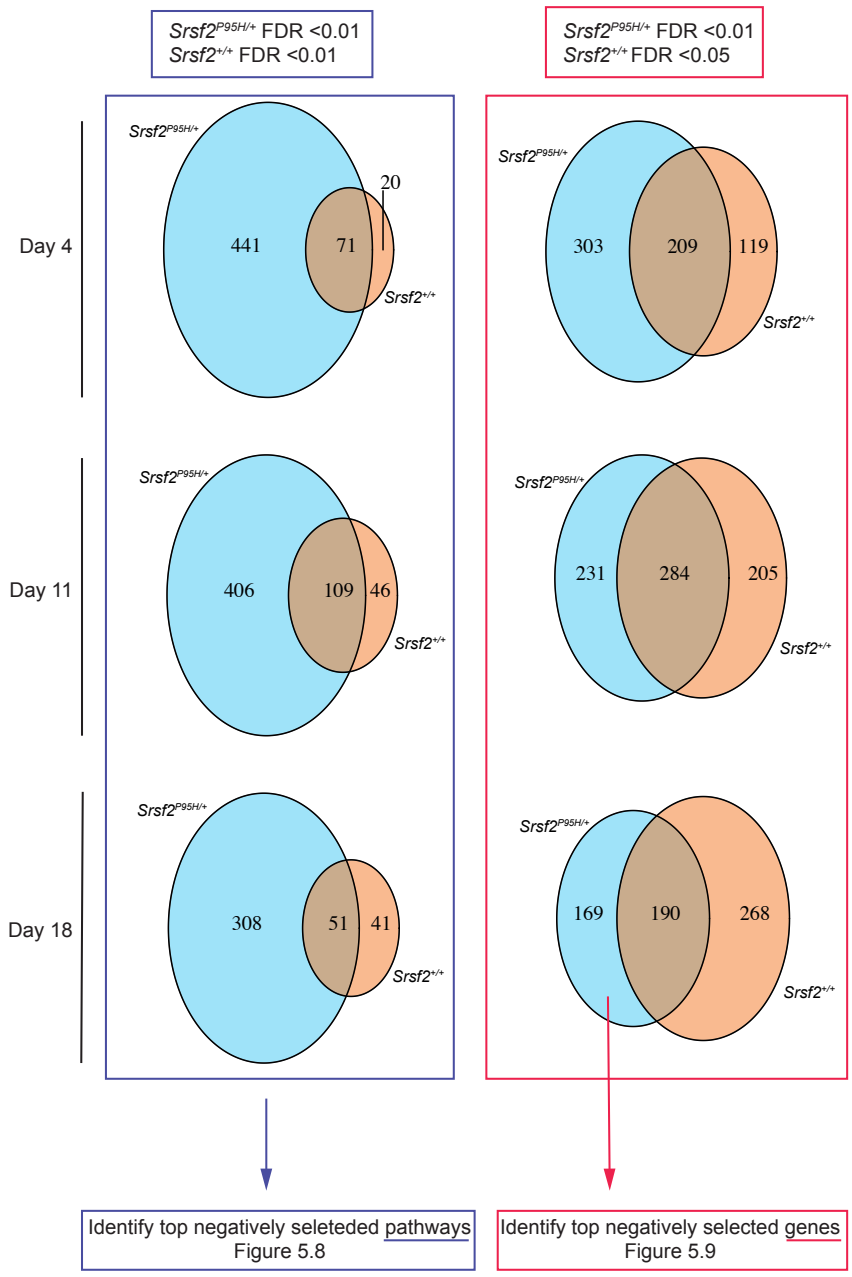


Figure 5.8: The number of negatively selected genes using different FDR cut-offs

A FDR<.01 cut-off is applied to both *Srsf2*$^{P95H/+}$ and *Srsf2*$^{+/+}$ negatively selected genes. The result is then used to identify top negatively selected pathways (Figure 5.9). To create a more stringent criteria for identifying top negatively enriched genes, a FDR<.05 is applied to *Srsf2*$^{+/+}$, while *Srsf2*$^{P95H/+}$ remains at FDR<.01. The top 30 genes are then listed from the *Srsf2*$^{P95H/+}$ only group (the light blue area) (Figure 5.10).

Genes are initially filtered at P95H_D18 time point at FDR <.05, prior to any further processing mentioned above. There are three biological replicates per genotype. FDR, false discovery rate.

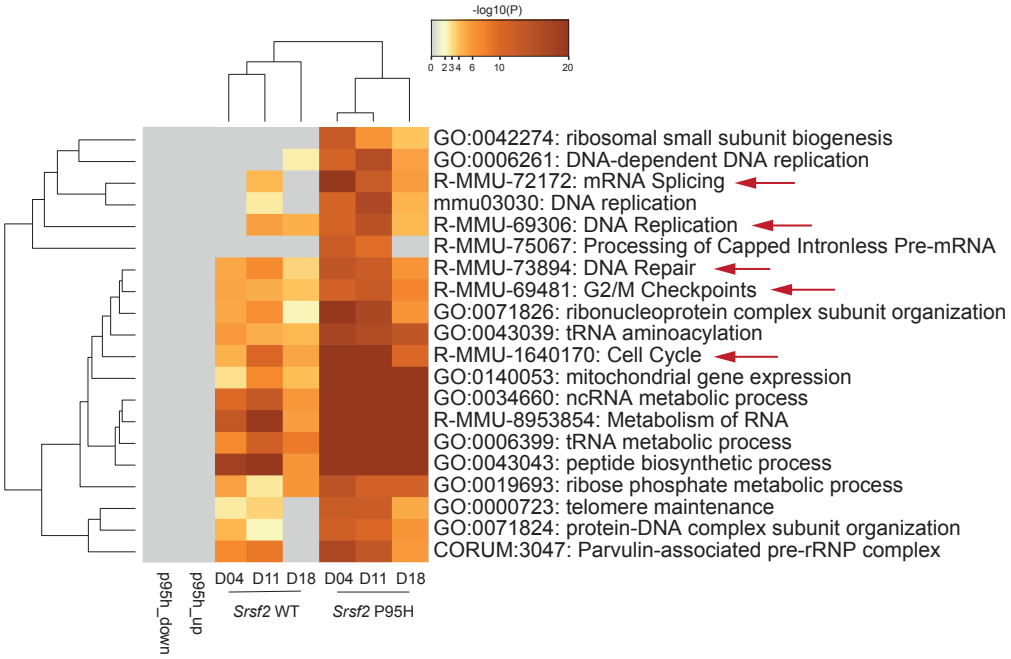


Figure 5.9: Disruption of DNA repair, DNA replication, cell cycle and RNA metabolic pathways leads to synthetic lethality in *Srsf2*$^{P95H/+}$ mutant cells

Genes are initially filtered at P95H_D18 time point at an FDR <.05, prior to applying an FDR<.01 on *Srsf2*$^{+/+}$ and *Srsf2*$^{P95H/+}$ cell lines individually at each timepoint. A list of genes that are significantly up- (p95h_up) or down-regulated in *Srsf2*$^{P95H/+}$ was also included in the comparison. Heatmap was generated using Metascape. Pathways indicated by red arrows are further discussed in the thesis. There are three biological replicates per genotype. FDR, false discovery rate.

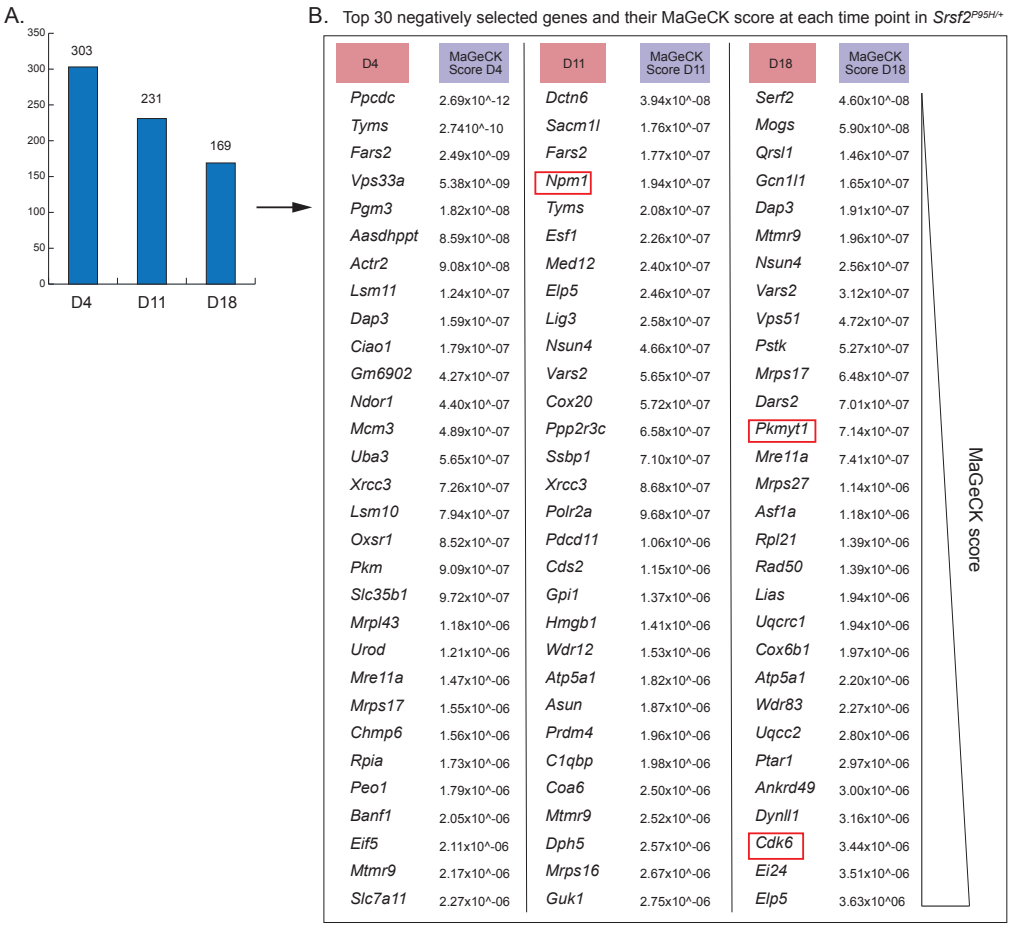


Figure 5.10: Cell cycle genes, *Pkmyt1* and *Cdk6* are among the top 30 genes that are only negatively selected in the *Srsf2*$^{P95H/+}$ cell lines at day 18

A. The number of genes that are only negatively selected in *Srsf2*$^{P95H/+}$ cell lines at day 4, day 11 and day 18.
B. The top 30 genes that are only negatively selected in the *Srsf2*$^{P95H/+}$ cell lines at day 4, day 11 and day 18.
An FDR<.05 is applied to *Srsf2*$^{+/+}$, while *Srsf2*$^{P95H/+}$ is at FDR<.01. Genes are ranked by MaGeCK score from the lowest (the most negatively enriched) to highest. Genes highlighted with red boxes are further discussed in the thesis. There are three biological replicates per genotype. FDR, false discovery rate.

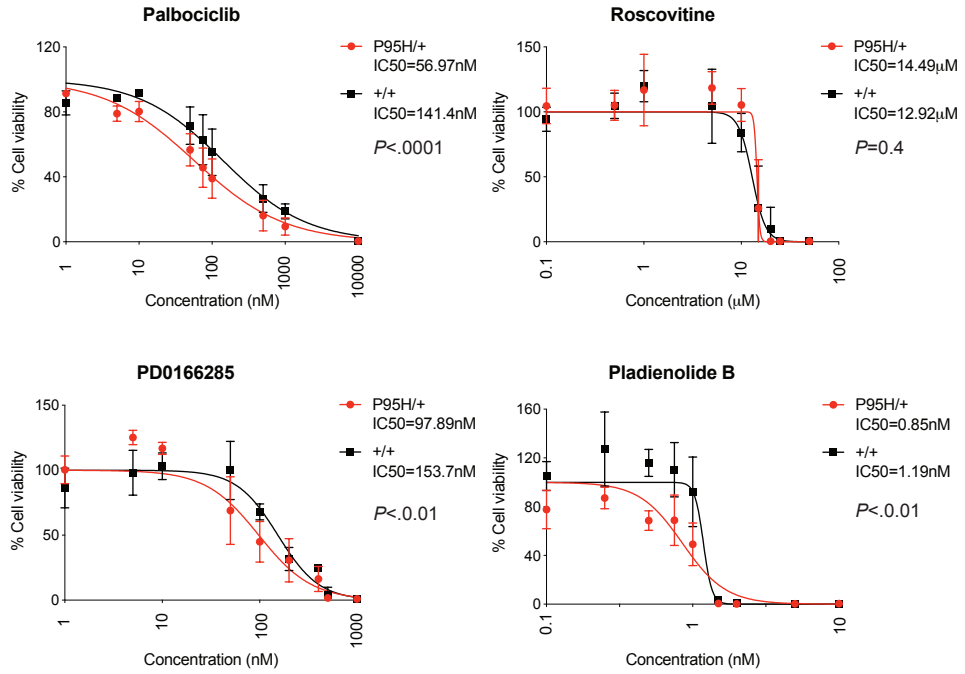
5.2.5 Pharmacological validation of *Srsf2* CRISPR Screen results

Although the majority of the negatively selected genes in *Srsf2*$^{P95H/+}$ cells belong to essential cellular pathways, I examined druggable targets within the *Srsf2*$^{P95H/+}$ day 18 list using the Drug Gene Interaction database (DGIdb) (Cotto et al., 2018). Among the 30 genes that were druggable, I selected *Cdk6* and *Pkmyt1* for further testing (Figure 5.9B). Both proteins have inhibitors available that are being used clinically for cancer treatment or have demonstrated anti-tumour activity (Kawabe, 2004; Spring et al., 2019). I analysed the efficacy of the CDK6 inhibitor, Palbociclib, and the PKMYT1 inhibitor, PD0166285, against that of Pladienolide B (PLB), a spliceosome inhibitor. Roscovitine, a CDK2/4 inhibitor, was used as a negative control for the CDK6 inhibitor as neither *Cdk2* nor *Cdk4* sgRNA had been identified as synthetic lethal with the SRSF2^{P95H} mutation.

I first performed dose response assays with the parental Hoxb8 cell lines that I used to generate the Cas9 *Srsf2* cell lines. Briefly, Hoxb8 *Srsf2* cell lines were treated with 400nM tamoxifen for 4 days to activate the mutation. Serial dilutions of Palbociclib, PLB, PD0166285, and Roscovitine were added 3 days after P95H activation. The cell viability was measured 4 days later. Palbociclib and PD0166285 both showed reduced IC50 values for the *Srsf2*$^{P95H/+}$ cells compared to *Srsf2*$^{+/+}$ cells (Figure 5.11A). To validate this in primary cells, I then sorted the Lin^{-}c-Kit^{+}Sca-1^{+} (LKS^{+}) stem cell containing population from *R26*-CreER^{T2} *Srsf2*$^{P95H/+}$ mice and activated the Srsf2^{P95H} mutation *ex vivo* with tamoxifen treatment. *Srsf2*$^{P95H/+}$ LKS^{+} cells, continued to show sensitivity towards Palbociclib while they failed to show sensitivity towards PD0166285 (Figure 5.11B). I therefore only proceeded with further testing of Palbociclib.

Conventional drug dose response assays generally assess the cellular toxicity of compounds after a 48-hour time period. This differs from the experimental design of the *Srsf2* CRISPR screen, where Cas9-mediated gene editing occurs within the 4 days before activation of the mutation and where the effects of gene disruption are measured for up to 14 days. To make the timeline more comparable between the genetic and pharmacological perturbations, I applied the IC50 value generated from the 4-day drug assay described above (Figure 5.10A) and tested the cell proliferation in the presence of Palbociclib (IC50 =57nM) for 14 days. In parallel, a culture with Pladienolide B at the IC50 of 0.85nM was performed as a positive control. The results showed that *Srsf2*$^{P95H/+}$ cells were more sensitive to Palbociclib than *Srsf2*$^{+/+}$ cells, while the spliceosome inhibitor was equally toxic to both genotypes (Figure 5.12).

A. Hoxb8-immortalized *R26*-CreER^{T2} GM-CSF cell lines



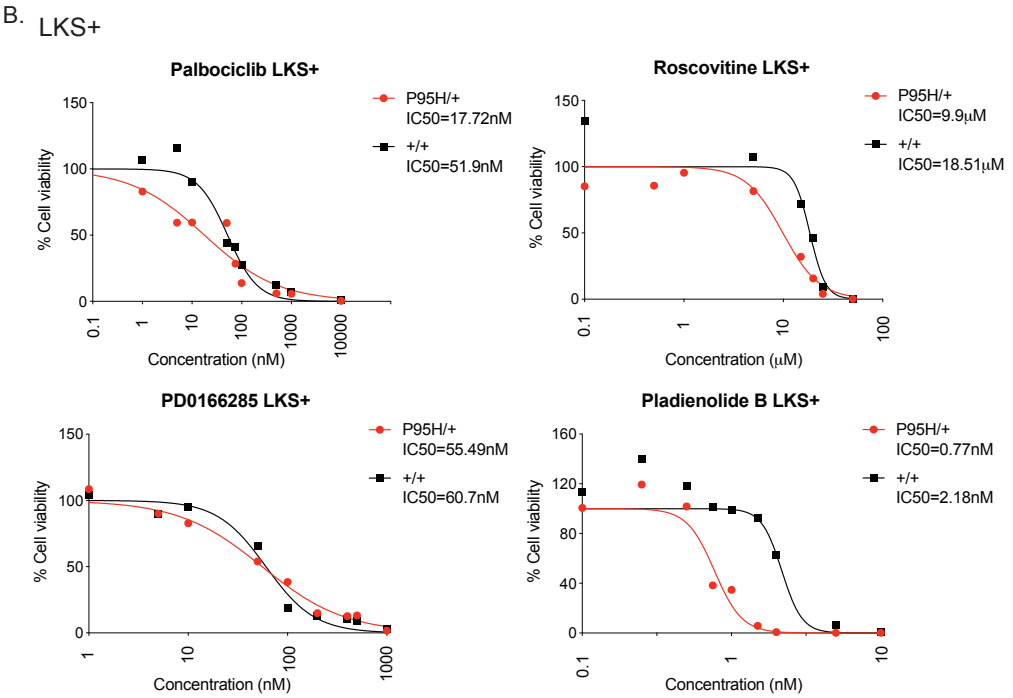


Figure 5.11: *Srsf2*$^{P95H/+}$ cells are sensitive to the CDK6 inhibitor Palbociclib, as well as to the spliceosome inhibitor Pladienolide B

A. Hoxb8-immortalized *Srsf2*$^{P95H/+}$ cell lines are sensitive to Palbociclib and PD0166285 inhibition, similar to Pladienolide B.

B. Sorted *Srsf2*$^{P95H/+}$ LKS^{+} cells are sensitive to Palbociclib and Pladienolide B, but not sensitive to PD0166285.

For Hoxb8-immortalized cell lines, n=3 per genotype for Roscovitine, PD0166285 and Pladienolide B. n=2 for Palbociclib. Extra sum-of-squares F Test. For LKS^{+} cells, n=1 per genotype. LKS^{+}, lin^{-}cKit^{+}Sca-1^{+}.

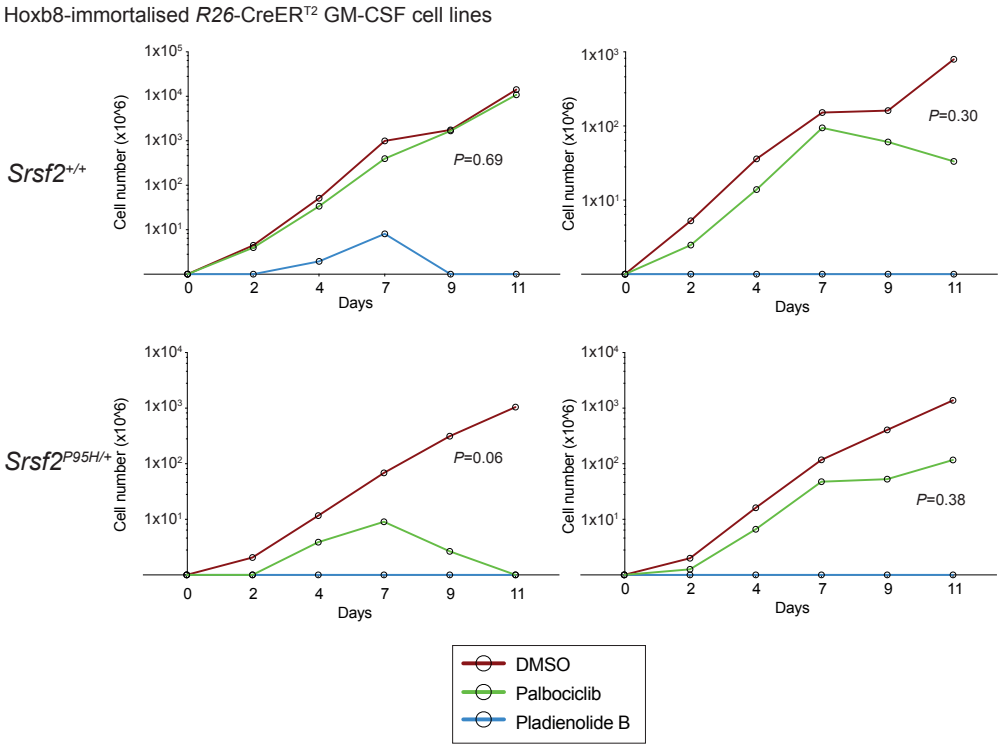


Figure 5.12: CDK6 inhibitor Palbociclib reduces the proliferation of *Srsf2*$^{P95H/+}$ cells while spliceosome inhibitor Pladienolide B is lethal for both genotypes in long-term *in vitro* culture.
Two *Srsf2*$^{+/+}$ and *Srsf2*$^{P95H/+}$ cells lines are treated with DMSO, Palbociclib or Pladienolide B *in vitro* for 11 days 3-days post mutation activation. Each graph shows the proliferation curve of one cell line treated with three compounds. No replicate for the experiment. Receiver-operating characteristic (ROC) curve.

5.3 Discussion

The experimental design is the most crucial step to a successful CRISPR screen. Multiple factors, such as the choice of cell type, the duration of the experiment, sequencing depth, quality control and downstream validation, all have a direct impact on the quality and validity of the screen results. Previous screens used commonly available leukaemia cell lines with or without specific mutations. Although this is feasible for a disease with well-established cell lines, such as AML, there were no existing MDS or CMML cell lines with *Srsf2*$^{P95H/+}$ mutation. In studies conducted by Zhang et al. and Kim et al., K562 erythroleukaemia cells were used with an *Srsf2*$^{P95H/+}$ mutation generated by CRISPR gene editing or *Srsf2*$^{P95H/+}$ mutant cell lines derived from AML or CMML patients (Kim et al., 2015; Zhang et al., 2015). In both cases, the cells contain strong oncogenic drivers, such as *BCR-ABL*, that can easily mask the true genetic interactions that lead to synthetic lethality of *Srsf2*$^{P95H/+}$ mutation. In addition, established cell lines often exhibit a complex spectrum of genetic alterations, with additional *p53* and *FLT3* mutations. Therefore, I employed Hoxb8 immortalisation to generate isogenic murine *Srsf2*$^{+/+}$ and *Srsf2*$^{P95H/+}$ cell lines with a fixed genetic background. The immortalized cells exhibit a late stage myeloid cell phenotype, which is relevant to *Srsf2*$^{P95H/+}$ mutation (Chapter 3). Unlike *Sf3b1* mutation, *Srsf2*$^{P95H/+}$ mutation is exclusively found in myeloid disease. Thus, screening *Srsf2*$^{P95H/+}$ mutation in non-myeloid cells would compromise the utility of the results.

Once efficient Cas9 function in the *Srsf2* cell lines was confirmed, it was crucial to determine the strategy for the negative selection screen based on the aims and experimental feasibility. To achieve a high statistical power, I chose to screen six independently derived cell lines, including

three *Srsf2*$^{P95H/+}$ and three *Srsf2*$^{+/+}$. Compared to technical replicates, biological replicates provide a better representation of the mean and variance of the splicing mutant biology (Blainey and Quake, 2014). Analysis from the screen showed clear biological divergence between the cell lines, especially in the *Srsf2*$^{+/+}$ cell lines (Figure 5.4). Therefore, by pooling results from three biological replicates, the identified hits reflect the biologically relevant differences between *Srsf2*$^{+/+}$ and *Srsf2*$^{P95H/+}$ cell lines.

I used a multiplicity of infection (MOI) of 0.3 to reduce the likelihood that a cell contained more than one sgRNA. On top of the MOI, I also aimed for 500 copies of each sgRNA after transduction. Achieving a high copy number of sgRNA is important for an efficient drop-out analysis by the MaGeCK algorithm. However, although the proliferation rate was assessed during the validation of Cas9 *Srsf2* cell lines and the library was titrated, the combined effects of both cell proliferation and puromycin selection make it difficult to calculate exactly how many cells to transduce at the beginning of the screen. As a result, the actual copy number achieved was an average of 355 copies of each sgRNA and was thus lower than the expected 500. This is not ideal but the copy number is still sufficiently high to perform a valid drop-out analysis (Hart et al., 2015; Shalem et al., 2014). Finally, the Hoxb8 cell lines divide rapidly (16-20 hours) and so the length of culture I have utilised is in excess of 10 population doubling, also important for statistical analysis of drop-out rates and completeness.

Studies of AML cell lines revealed that gene essentiality is largely context-dependent and reflects the genetics of each line (Wang et al., 2017). The purpose of my *Srsf2* screen was to identify which genes when deleted would lead to the selective cell death of *Srsf2*$^{P95H/+}$ mutant cells. To establish an 'essentiality profile' of the Cas9 Hoxb8 cell lines, I screened three

Cas9 *Srsf2*$^{+/+}$ cell lines alongside the mutant cell lines. Genes that were significantly depleted in these WT cells served as the best comparison to pinpoint the genetic liabilities specific to *Srsf2*$^{P95H/+}$ cells. Although it is ideal to identify synthetic lethal targets outside of the 'essential profiles', some of these 'essential genes' were also depleted in the mutant cells and often demonstrated a higher level of negative selection than that of WT cells. An example is the RNA splicing related genes. Depletion of any RNA splicing gene is detrimental to a cell, as they carry out a fundamental cellular process (Dvinge et al., 2016). However, splicing mutant cells are known to be more sensitive to further splicing disruptions, which was also observed in my screen (Obeng et al., 2019)(Figure 5.6). Identifying the differential essentialities could reveal the genetic vulnerabilities of *Srsf2*$^{P95H/+}$ cells.

In MDS, splicing factor mutations result in aberrantly spliced transcripts that cluster in cell cycle, DNA damage, RNA metabolism and RNA splicing pathways (Pellagatti et al., 2018; Qiu et al., 2016). In this screen, pathway analysis of the negatively selected genes revealed *Srsf2*$^{P95H/+}$ cells were more sensitive than *Srsf2*$^{+/+}$ cells to perturbation of these same pathways. This suggests that genetic vulnerabilities of *Srsf2*$^{P95H/+}$ lie within the pathways that are commonly mis-spliced in mutant cells. Two hypotheses can arise from this observation. One scenario is 'gain-of function', where certain mis-spliced products contribute to the oncogenic potential of *Srsf2*$^{P95H/+}$ cells. The other scenario is 'loss-of function', where mis-splicing leads to partial loss of normal cellular function and an increased dependence on the remaining intact pathways. Based on the observation that RNA splicing mutants are sensitive to further splicing perturbations, it is likely that a 'loss-of-function' underlies the increased genetic dependence on the DNA damage and Cell cycle pathways.

The timing at which synthetic lethality was most apparent also differs between the mutant and wild type cells. *Srsf2*$^{P95H/+}$ cell lines showed the most significant depletion of sgRNA targeting the RNA splicing, Ribonucleoprotein complex subunit organisation and Cell cycle pathways at day 4 post mutation activation. *Srsf2*$^{+/+}$ cell lines in contrast, did not show this depletion until day 11. The timing of cell-essential gene depletion indicates that *Srsf2*$^{P95H/+}$ cells are more sensitive to disruption of RNA splicing, Ribonucleoprotein complex subunit organisation and Cell cycle pathways than *Srsf2*$^{+/+}$ and that this leads to an early cell death. This pattern of gene depletion at different timepoints was observed previously and was thought to reflect different levels of essentiality of the genes. When examining a human CRISPR library in a colon cancer cell line, Tzelepis et al. categorised RNA splicing and ribosome pathways as essential biological processes that fall within the early depleted group (Tzelepis et al., 2016). Cell cycle and DNA replication, processes which are involved in proliferation, sit in the early to intermediate depleted group. In contrast, the mitochondrial pathway is in the late-depleted group, where genes continue to drop out throughout the screen as their loss has a lesser impact on proliferation.

From the pathway analysis, there are two pathways uniquely depleted in the *Srsf2*$^{P95H/+}$ cells. The first one is the Ribosomal Small Subunit Biogenesis pathway (GO:0042274), which is involved in the processing of ribosomal proteins that contribute to the small subunit of the ribosome. Mutations in ribosome components, generically termed ribosomopathies, can result in several haematopoietic diseases, including Diamond-Blackfan Anaemia (DBA) and MDS with 5q deletion. The decreased expression of ribosome proteins mostly leads to impaired erythropoiesis, while other lineages are preserved (Narla and Ebert, 2010).

The molecular link between RNA splicing mutant and ribosomopathy is unknown.

Within the Ribosomal Small Subunit Biogenesis pathway, *Npm1* is one of the depleted genes associated with haematological malignancies. Nucleophosmin, the protein encoded by NPM1, facilitates the nucleocytoplasmic shuttling of pre-ribosomal proteins, and is mutated in 20 to 30% of AML cases (Heath et al., 2017). Studies have shown that heterozygous deletion of *Npm1* in mice leads to the development of MDS phenotypes and myeloid malignancies, with a latency of 10-24 months. The homozygous loss of *Npm1* is embryonic lethal, due to profound defects in the forebrain and haematopoietic development (Grisendi et al., 2006; Sportoletti et al., 2008). Recently, Nachmani and colleagues reported a role for NPM1 as an RNA-binding protein. NPM1 binds predominately to small nucleolar RNAs (snoRNAs) and regulates 2'-O-methylation of ribosome RNA (rRNA) (Nachmani et al., 2019). Deletion of *Npm1* leads to centrosome abnormalities and genome instability, which might account for the synthetic lethality of *Npm1* and *Srsf2*$^{P95H/+}$. In AML patients, *NPM1* and *SRSF2* mutation are exclusive (P<0.001), which coincides with the negative selection in the CRISPR screen (Tyner et al., 2018).

The other pathway specific to the P95H mutant cells was the Processing of Capped Intronless Pre-mRNA (R-MMU-75067), where several genes encoding components of Pre-mRNA processing complexes were depleted. These include nuclear cap binding complex (NCBP) (*Ncbp1, Ncbp2*), cleavage and polyadenylation specificity factor (CPSF) (*Cpsf2, Cpsf3*) and cleavage stimulating factor (CSTF) (*Cstf1, Cstf3*). The NCBP complex is involved in the 5' capping of pre-mRNA, while the CPSF and CSTF complexes are responsible for the 3' polyadenylation (Izaurralde et al., 1994; Tian and Manley, 2017). Pre-mRNA processing is an essential

step before nuclear export of mRNA and subsequent translation by the ribosome. This process is coupled with mRNA synthesis via the carboxyl-terminal domain (CTD) of RNA polymerase II (PolII). The interaction of CTD with the capping and polyadenylation machinery also serves as the signal for transcription initiation and termination (Lee and Tarn, 2013). Interestingly, SRSF2 has been shown to interact with the PolII CTD domain and to regulate transcription elongation, in a splicing-independent manner (Lin et al., 2008). The synthetic lethality within this cluster of pre-mRNA processing genes indicates that splicing factors function in tight collaboration with other RNA modifiers. It will be interesting to see if these RNA modifications contribute to *Srsf2*$^{P95H/+}$ mediated leukaemogenesis. This could be tested through a gain of function screen with a library targeting RNA modification genes or a chemical screen with compounds known to affect Pre-mRNA processing pathway, such as JTE-607 (Ross et al., 2020).

Finally, one of my primary goals by investigating *Srsf2*$^{P95H/+}$ synthetic lethality was to probe for new therapeutic targets. *Cdk6* was identified as one of the genes selectively depleted in the *Srsf2*$^{P95H/+}$ cell lines and presents an attractive target for therapeutic inhibition. The post-screen validation showed that the CDK6 inhibitor Palbociclib was able to slow down the growth of *Srsf2*$^{P95H/+}$ cells while having minimal impact on *Srsf2*$^{+/+}$ cells. Mechanistically, Palbociclib treatment leads to a cell cycle arrest at the G0/G1 phase (Finn et al., 2009). Although the functional impact of *Srsf2* mutation on cell cycle is unclear, depletion of SRSF2 leads to a G2-phase cell cycle arrest and apoptosis (Li et al., 2005). It is possible that *Srsf2* mutation causes a mild deregulation of the cell cycle which leads to a synthetic lethality when treated with Palbociclib. Future experiments to test the CDK6 inhibitor would include administering Palbociclib in

vivo to aged Srsf2 mutant mice and assessing if there was a modification of the MDS/MPN phenotype. Palbociclib is approved for breast cancer treatment in Australia and is currently in phase I/II trials for treatment of AML (Winer and Stone, 2019) (NCT03844997). Compared to splicing modulators, such as pladienolide B and its derivative, E7107, Palbocilib could provide a new alternative for the treatment of splicing mutant cancers. Further testing in pre-clinical settings using human cell lines and *in vivo* murine models, is needed to confirm its therapeutic potential.

Chapter 6

Srsf2$^{P95H/+}$ co-operates with *Tet2* and *Cbl* mutations to promote myeloid bias and initiation of CMML

6.1 Introduction

For MDS and CMML, the heterogeneous nature of the disease is reflected in both the clinical presentation and genetic landscape. While the roles of many recurrent mutations have been well-studied as individual mutations in animal models, the co-operativity of mutations is yet to be understood. Investigating genetic co-operativity, especially between driver mutations, could determine the contribution of early clonal events in the initiation of MDS and how the different functional pathways regulate haematopoiesis. Co-mutation of *SRSF2* and *TET2* is profoundly enriched in MDS and MDS/MPN ($p=2.63x10^{-14}$; $q=1.21x10^{-12}$; cohort of n=3047) (Ochi et al., 2020). In CMML, *SRSF2* and *TET2* are the most frequently occurring mutations (50% and 60% respectively) and are the likely founder mutations in a significant proportion of these cancers. In CMML and the related chronic neutrophilic leukaemia and atypical chronic myeloid leukaemia, there is a positive association between mutation of *SRSF2* and *TET2* (p=0.002; n=218 (Meggendorfer et al., 2014)). This association is strongest in CMML (p=0.004; 52 co-mutated of 146 samples (Meggendorfer et al., 2014)). Similarly, in CMML *SRSF2* mutation associates with *CBL* mutation, which belongs to the RAS signalling pathway (Patel et al., 2017). Thus, determining how and why these mutations co-operate would provide valuable insight into the evolution of CMML.

To date, the genetic interactions with mutations of *TET2* have been assessed in experimental models with concurrent mutation of *EZH2* (chromatin remodelling) or N-RAS (Ras signalling), which I reviewed in Chapter 1 (Hasegawa et al., 2017; Kunimoto et al., 2018; Muto et al., 2013). In terms of SRSF2, Yoshimi et al., demonstrated experimentally an interaction with tumour associated point mutations of *IDH2* where this genetic interaction was noted in *de novo* AML. They did however not report any

experimental evidence for co-operativity with loss of function *TET2* mutation, even though *SRSF2* and *TET2* mutations are highly correlated in humans (Itzykson et al., 2013; Nagata and Maciejewski, 2019; Ochi et al., 2020).

To understand the interaction of *Srsf2*$^{P95H/+}$ and *Tet2* mutation, or *Srsf2*$^{P95H/+}$ and *c-Cbl* co-mutation, I have generated and characterised two compound mutant mouse models: *Srsf2*$^{P95H/+}$ *Tet2*$^{fl/fl}$ and *Srsf2*$^{P95H/+}$ *c-Cbl*$^{-/-}$. The *Srsf2*$^{P95H/+}$ *Tet2*$^{fl/fl}$ model is regulated by h*Scl*-CreER, which activates the *Srsf2*$^{P95H/+}$ mutation and deletes the *Tet2* allele at the same time within haematopoietic stem cell populations upon tamoxifen treatment. The *Srsf2*$^{P95H/+}$ *c-Cbl*$^{-/-}$ model also utilised h*Scl*-CreER however, in this instance as the *c-Cbl* mutation is a non-conditional germ-line mutation, h*Scl*-CreER activates the *Srsf2*$^{P95H/+}$ mutation upon tamoxifen treatment while the *c-Cbl* null mutation is always present. In this chapter, I present the data describing the *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ mutant model, which developed increased granulocytes and monocytes (myelomonocytic hyperplasia) with elevated immature promonocytes and monoblasts ($\sim$10 – 15% total nucleated cells in the bone marrow). These features are consistent with CMML. A similar granulocytosis was also observed in the *Srsf2*$^{P95H/+}$ *c-Cbl*$^{-/-}$ model, albeit with limited numbers analysed due to significant difficulties with breeding mice of the required genotypes.

6.2 Results

6.2.1 h*Scl*-CreER^{T} *R26*eYFP *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ compound mutant mice show modest haematological change at 20 weeks post mutation activation

To generate *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ compound mutant mice, I crossed the h*Scl*-CreER^{T} $^{Tg+}$ *R26*eYFP *Srsf2*$^{P95H/+}$ mice (Smeets et al., 2018), generated by my laboratory, and *Tet2*$^{fl/fl}$ mice which were obtained from the Jackson Lab (Moran-Crusio et al., 2011). All mice were on a C57Bl/6 background. Cohorts of h*Scl*-CreER^{T} $^{Tg+}$ *R26*eYFP *Srsf2*$^{P95H/+}$ *Tet2*$^{fl/fl}$ mice and control genotypes were given tamoxifen food for 4 weeks from 8-10 weeks of age to activate expression of the *Srsf2* mutation and/or delete *Tet2* in haematopoietic stem cells, with mutant cells also labelled with eYFP following Cre activation. The cohorts assessed included wild-type (WT, +/+), Srsf2 mutant (*Srsf2*$^{P95H/+}$), Tet2 null (*Tet2*$^{-/-}$) and double mutant mice (*Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$). At $\sim$ 20 weeks post tamoxifen treatment, a cohort of mice were analysed to investigate the impact of the mutations on haematopoiesis after a relatively short-term expression/deletion (Figure 6.1A).

In the peripheral blood, both *Srsf2* mutant and double mutant mice had significant macrocytic changes, while the double mutant mice also had elevated haemoglobin levels (Figure 6.1B). The number of red blood cells (RBCs) and leukocytes were unchanged. Similarly, there were no changes to the lineage distribution in leukocytes between genotypes at 20 weeks post gene mutation (Figure 6.1C).

Within the bone marrow, *Srsf2* mutant and double mutant mice had a hypocellular bone marrow compared to the wild type controls (Figure

6.2A). However, there were no changes in the number of the erythroid, myeloid, B and T cells in the bone marrow (Figure 6.2B-D). The stem and progenitor populations remained unchanged in absolute numbers (Figure 6.3A-B). In the spleen, the double mutant mice had a comparable cellularity to that of the wild type controls. The *Srsf2* mutant had a significantly lower splenic cellularity while the *Tet2* null mice had an increased cellularity (Figure 6.4A). No difference was detected in the myeloid, lymphoid and erythroid lineages between the double mutants and wild type controls (Figure 6.4B-C). In contrast, there was a significant increase in the early erythroid progenitor population ($Ter119^{+}CD71^{hi}$) in the *Tet2* null mice (Figure 6.4C). The thymus displayed no changes across all genotypes at this time point (data not shown).

Collectively these data demonstrated that expression of $SRSF2^{P95H}$ and deletion of TET2 leads to subtle haematological changes in the peripheral blood and bone marrow and suggested that when co-occurring these mutations are well tolerated at 20 weeks post gene mutation.

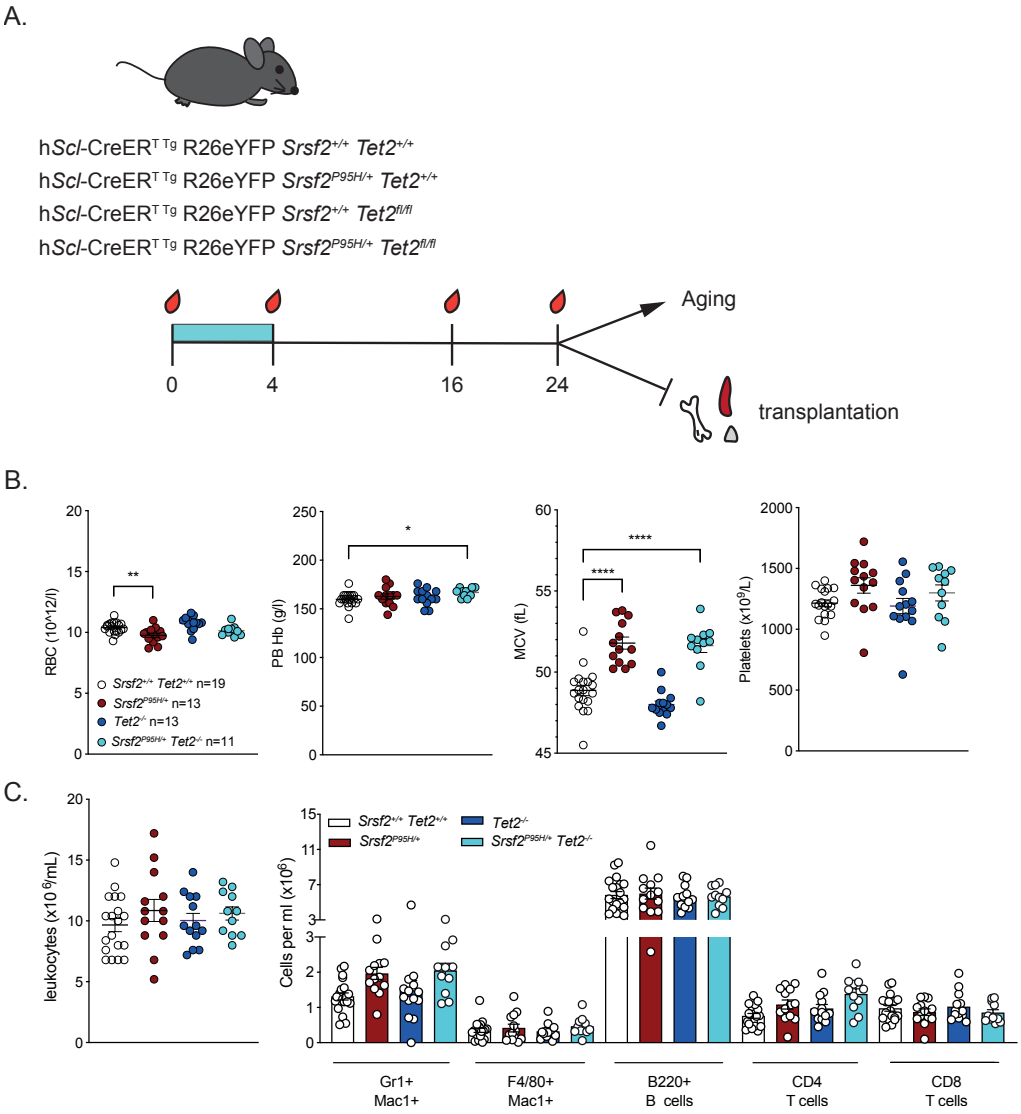
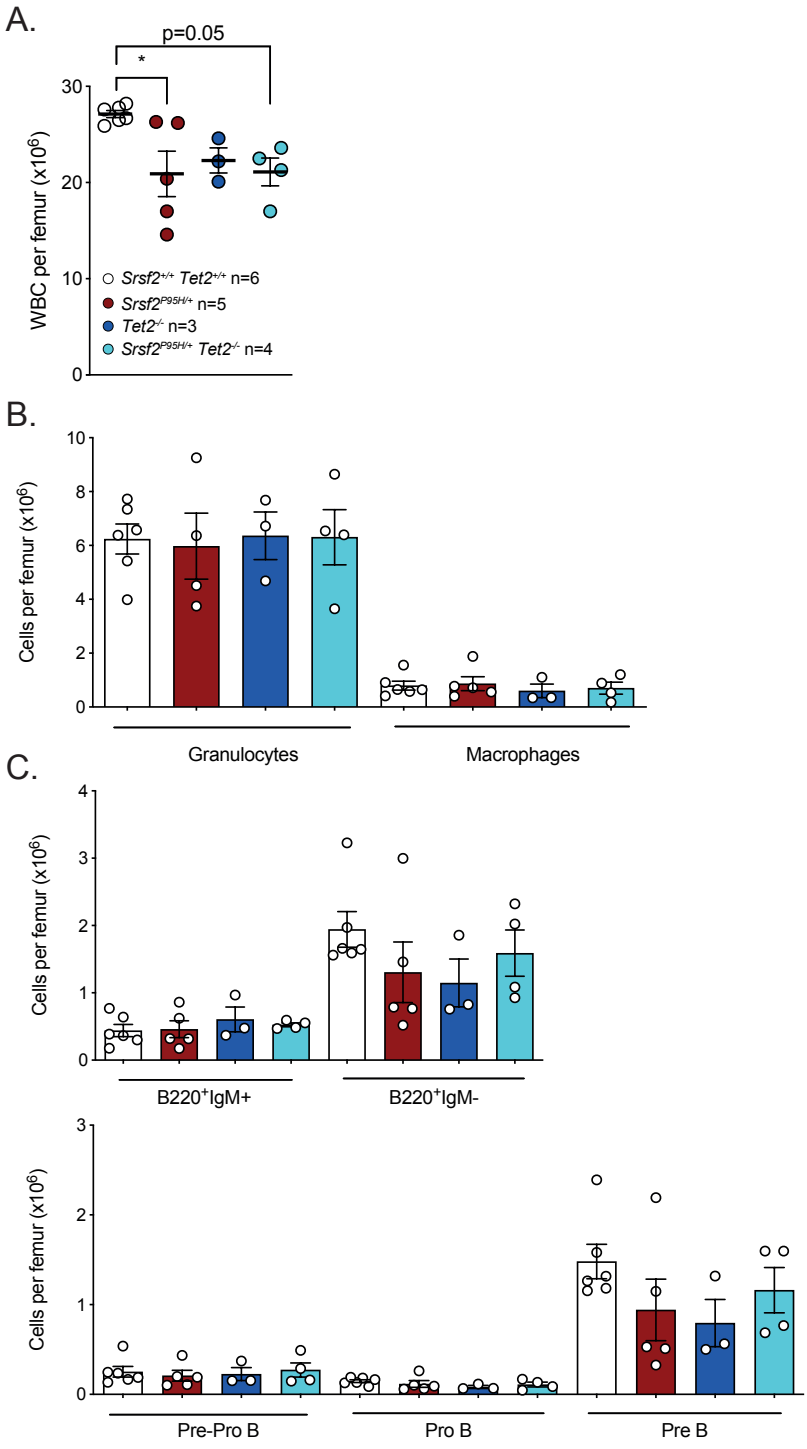


Figure 6.1: Co-expression of *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ leads to macrocytic changes rapidly after gene mutation

A. Schematic of h*Scl*-CreER *R26*eYFP *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ experiments.

B. PB indices after 20 weeks of Cre activation

C. PB leukocyte counts and lineage distribution after 20 weeks of Cre activation n >10 per genotype. Presented as mean ± standard error of mean. One-way ANOVA. *P<.05, ****P<.0001.


A.
p=0.05
*
WBC per femur (x10^6)
30
20
10
0
Srsf2+/+ Tet2+/+ n=6
Srsf2P95H/+ n=5
Tet2-/- n=3
Srsf2P95H/+ Tet2-/- n=4
B.
Cells per femur (x10^6)
10
8
6
4
2
0
Granulocytes
Macrophages
C.
Cells per femur (x10^6)
4
3
2
1
0
B220+IgM+
B220+IgM-
Cells per femur (x10^6)
3
2
1
0
Pre-Pro B
Pro B
Pre B

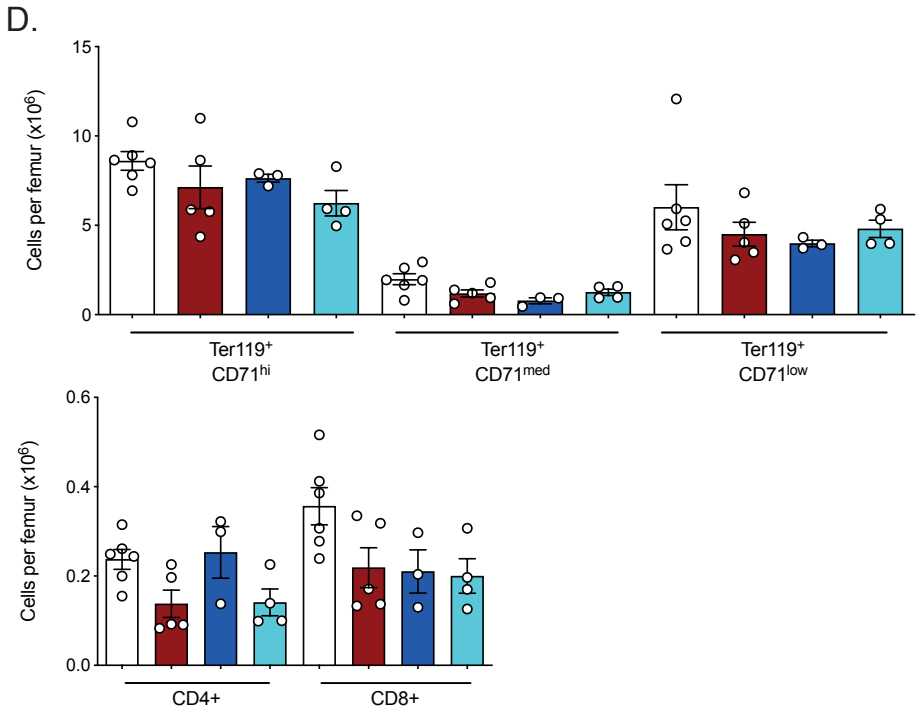


Figure 6.2: Co-expression of *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ leads to hypocellular bone marrow at early time points post gene mutation
A. Bone marrow cellularity per femur.
B. Number of granulocytes and macrophages per femur.
C. Number of mature (IgM^{+}), immature (IgM^{-}) B cells and B cell progenitor populations per femur.
D. Number of erythroid precursors and T cells per femur.
IgM, immunoglobulin M. n=3-6 per genotype. Presented as mean $\pm$ standard error of mean. One-way ANOVA. *P<.05.

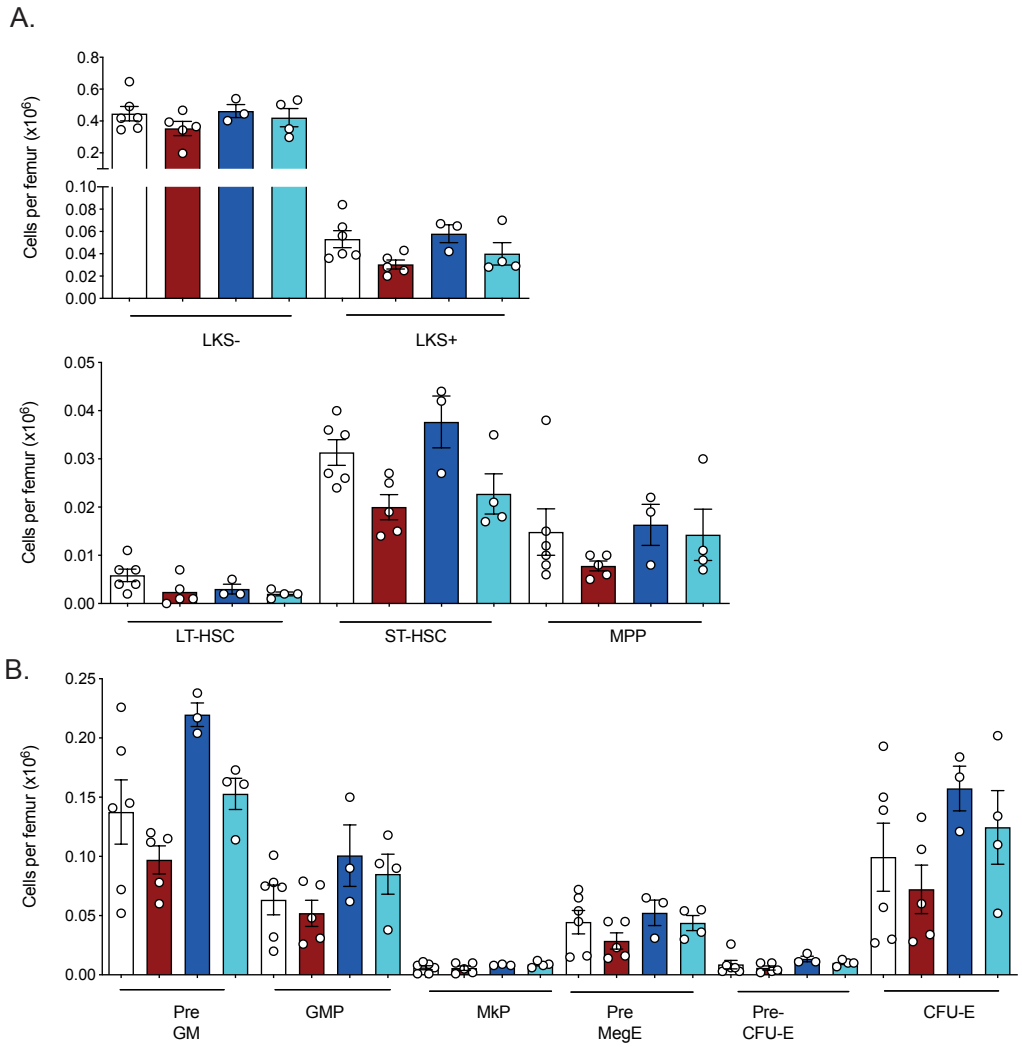


Figure 6.3: Co-expression of *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ has no effect on stem and progenitor populations after 20 weeks of mutation activation
A. Number of LKS$^-$, LKS$^+$ and phenotypic LT-HSC, ST-HSC and MPP in LKS$^+$ population using CD34/FLT3.
B. Number of myeloerythroid progenitors per femur.
LKS$^-$, lin$^-$c-Kit$^+$Sca-1$^-$; LKS$^+$, lin$^-$c-Kit$^+$Sca-1$^+$; LT-HSC, long-term haematopoietic stem cell; ST-HSC, short-term haematopoietic stem cell; MPP, multipotent progenitor; PreGM, Pre-granulocyte-macrophage; GMP, granuocyte-macrophage progenitor; MkP, megakaryocyte progenitor; PreMegE, pre-megakaryocyte erythroid progenitor; Pre-CFU-E, pre-colony-forming unit erythroid; CFU-E, colony-forming unit erythroid. n=3-6 per genotype. Presented as mean $\pm$ standard error of mean.

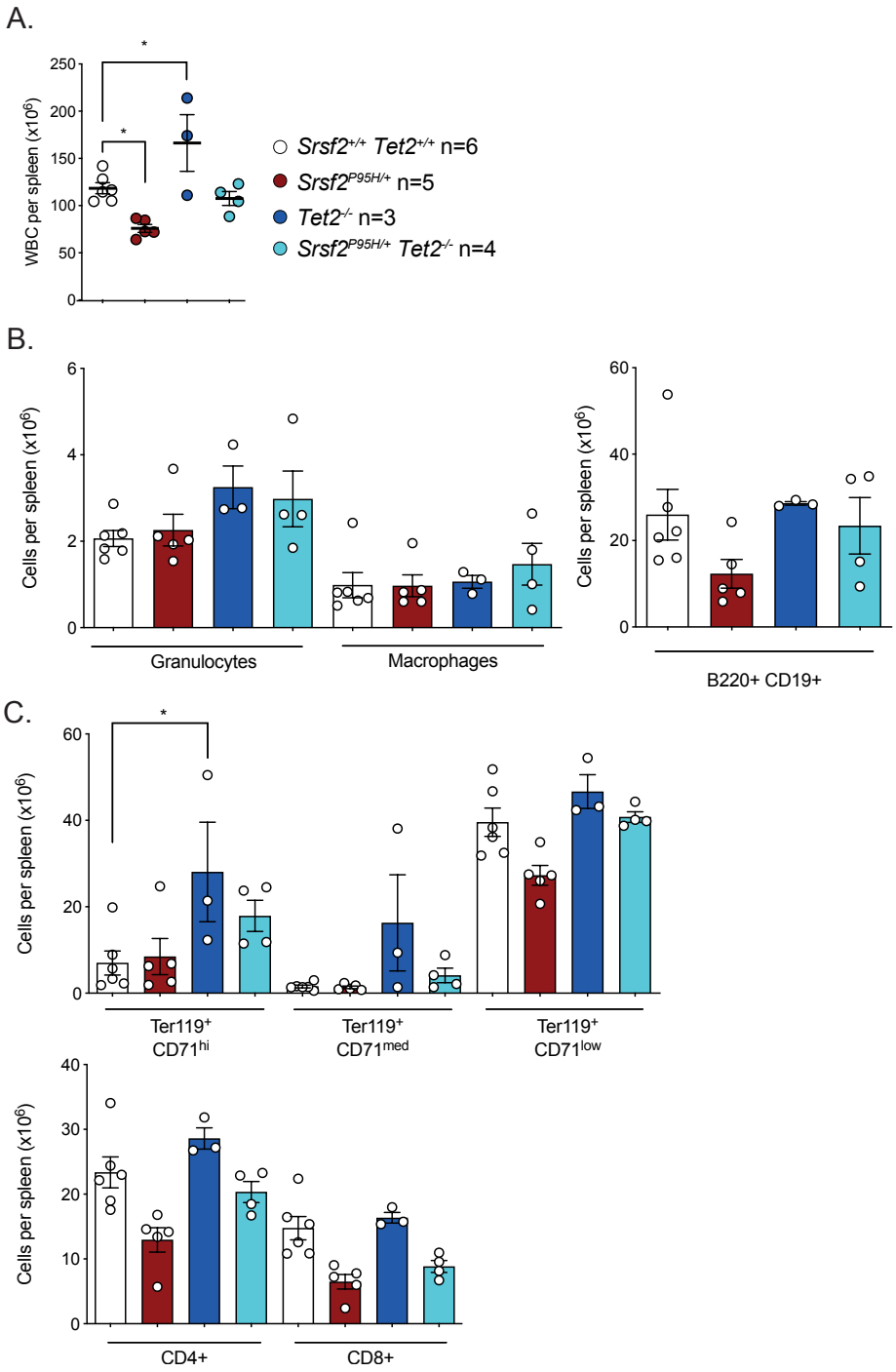


Figure 6.4: Co-expression of *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ leads to mild myeloid bias in the spleen after 20 weeks of mutation activation
A. Spleen cellularity.
B. Number of myeloid and B cells in the spleen.
C. Number of erythroid precursors and T cells in the spleen.
n=3-6 per genotype. Presented as mean ± standard error of mean. One-way ANOVA. *P<.05.

6.2.2 h*Scl*-CreER^{T} *R26*eYFP *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ mice develop more evident myeloid expansion and lymphoid suppression with time

As MDS and CMML are primarily diseases of the elderly, additional cohorts of h*Scl*-CreER^{T} *R26*eYFP *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ and control cohorts of mice were followed for up to 52 weeks post gene mutation to assess the effect on haematopoiesis with aging. In the peripheral blood, both the *Srsf2* mutant and double mutant mice showed macrocytic changes, as was observed at 20 weeks post gene mutation (Figure 6.5A). *Tet2*$^{-/-}$ mice, however, had an increase in red blood cells and haemoglobin, while also presenting with thrombocytopenia in the peripheral blood (Figure 6.5A). There were no changes to the lineage distribution of the leukocytes, except for the *Tet2*$^{-/-}$ mice that had a decreased number of B cells compared to wild type controls (Figure 6.5B). The same finding was observed for the lineage distribution within the eYFP positive population (Figure 6.5C).

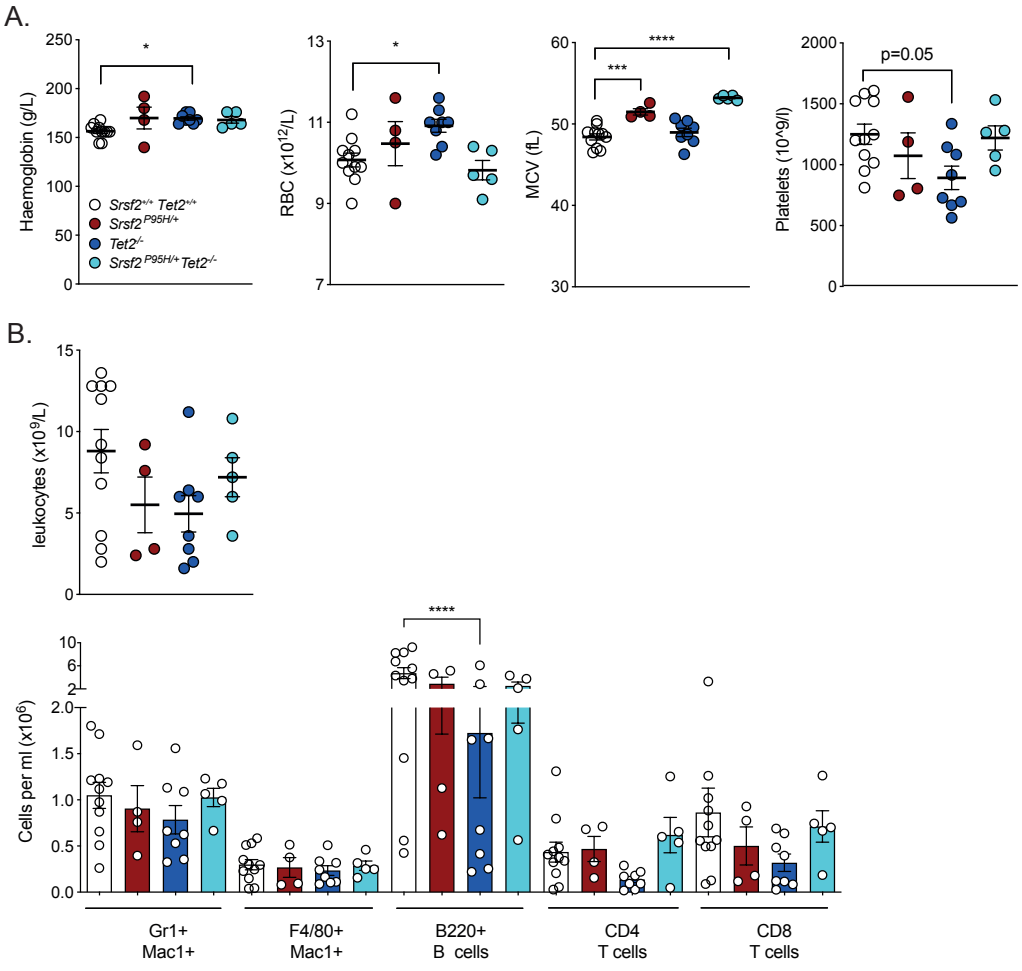
In the bone marrow, there was a significant hypocellularity in the double mutant mice (Figure 6.6A). There were no changes in the number of myeloid cells (Figure 6.6B). In contrast, there was extensive lymphoid suppression from the Pre-B cell stage to mature B cells with the effect increasing in severity from the single mutants to the double mutants (Figure6.6C). Similarly, the early stage erythroid progenitors (Ter119^{+}CD71^{hi}) and late stage erythroid progenitors (Ter119^{+}CD71^{low}) were decreased in the double mutant mice. T lymphoid cells remained unchanged (Figure 6.6D). By percentage, the double mutant mice had increased granulocytes, indicating myeloid expansion (Figure 6.6E). Most of the B lymphoid populations were decreased, except for the Pre-Pro B cells, which indicated a block in B lymphoid differentiation. A similar trend was observed in the *Tet2*$^{-/-}$ mice, except that this genotype exhibited a normal cellularity but

an elevated percentage of granulocytes (Figure 6.6E). The same extent of myeloid expansion and lymphoid suppression was present in the eYFP positive fraction for *Tet2* null mice and double mutant mice. The erythroid lineage had no change in the eYFP positive fraction (Figure 6.7).

Surprisingly, within the stem and progenitor cell populations the double mutant mice showed a trend to a reduced number of LKS^{+} (lin^{-}Sca-1^{+}c-Kit^{+}) and a significantly reduced number of LKS^{-} (lin^{-}Sca-1^{+}c-Kit^{-}) cells. Within the LKS^{+} population, double mutant mice had a lower number of short-term haematopoietic stem cells (ST-HSC) (Figure 6.8A). There were no obvious alterations in the myeloid and erythroid progenitors in the double mutant mice (Figure 6.8B). Interestingly, the percentage of phenotypic ST-HSCs was significantly increased in the double mutant mice while the percentage of LKS^{-} remained low (Figure 6.8C). The same shift in LKS^{-} and ST-HSC populations was apparent in the eYFP positive fraction of the bone marrow in the double mutant mice (Figure 6.8D). These analyses demonstrate that with aging, *Srsf2/Tet2* double mutant cells acquire a myeloid bias and expansion of the ST-HSC fraction.

Compared to the bone marrow, the double mutant mice displayed no change in the total spleen cellularity (Figure 6.9A). There was an increase in granulocytes in the *Tet2*$^{-/-}$ spleen, while *Srsf2*$^{P95H/+}$ and double mutant mice had reduced B cells. There were no changes in the erythroid and T lymphoid lineages in the spleen (Figure 6.9B). The double mutant mice presented with an increased percentage of granulocytes and a decrease in B cell percentages in the spleen (Figure 6.9C). Within the spleen, the *Srsf2*$^{P95H/+}$ mice had an increased total eYFP percentage compared to wild type. The double mutant eYFP cells produced an increased percentage of granulocytes and early erythroid progenitors (Ter119^{+}CD71^{hi}), while the percentage of B cells declined. There was no change in splenic

T cells (Figure 6.9D) and the thymus also showed no change (data not shown). Overall, these analyses demonstrate that *Srsf2/Tet2* double mutation drives an increased myeloid bias and B lymphoid suppression in bone marrow and spleen upon aging. The erythroid lineage is compromised in the bone marrow, while the T lymphoid lineage is unaffected.



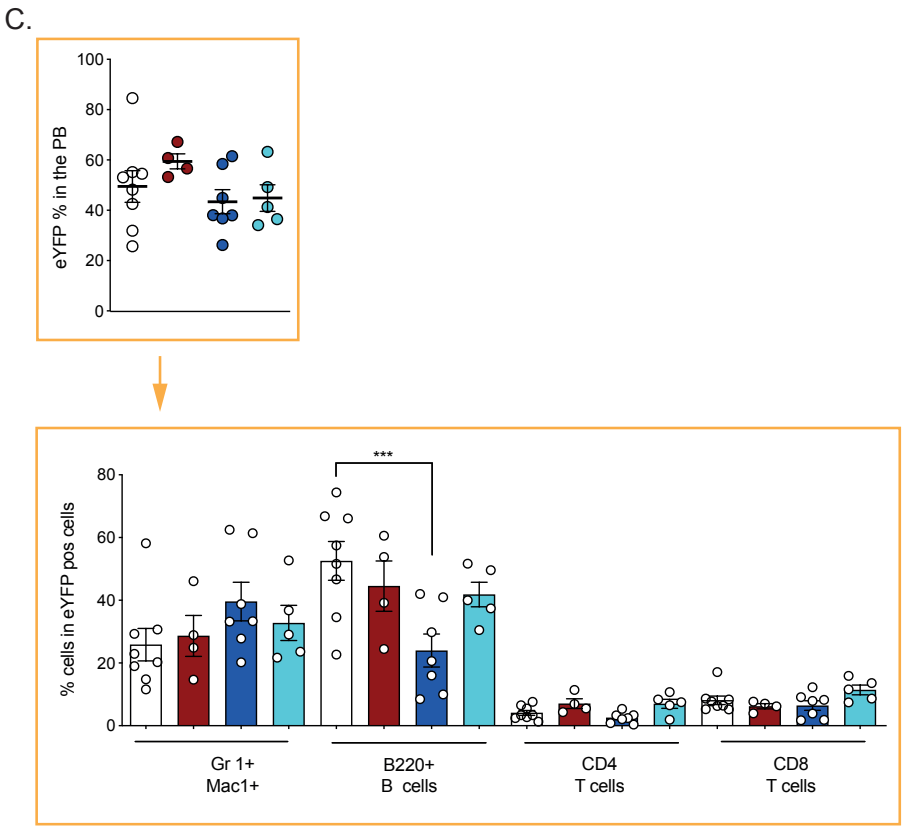


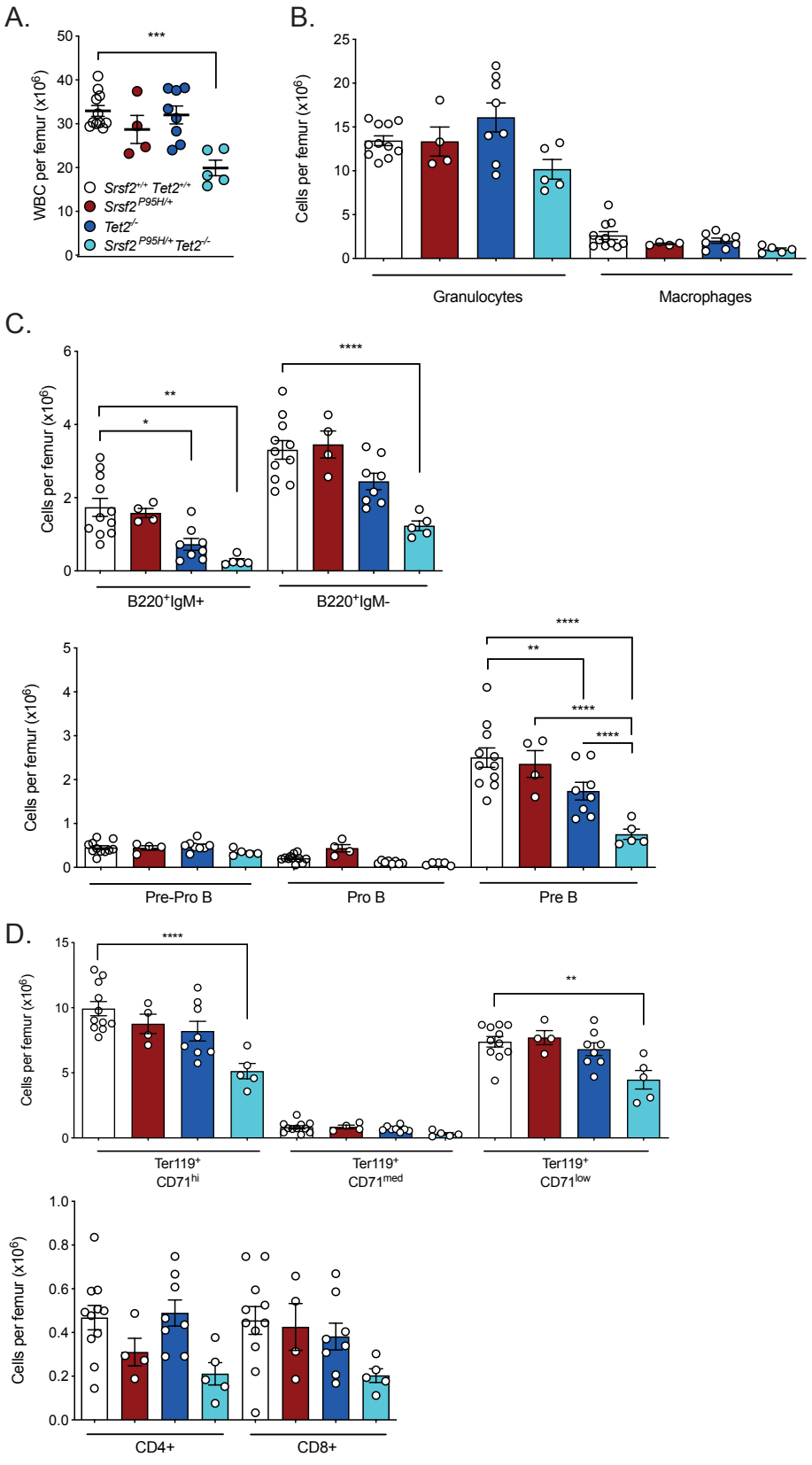
Figure 6.5: Co-expression of *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ leads to macrocytic red blood cells and lymphoid suppression in the peripheral blood 52 weeks after mutation activation

A. PB indices after 52 weeks of Cre activation.

B. PB leukocyte counts and lineage distribution after 52 weeks of Cre activation.

C. eYFP percentage in the PB and lineage distribution within the eYFP positive population.

PB, peripheral blood. n=4-11 per genotype. Presented as mean ± standard error of mean. One-way ANOVA. ***P*<.05, *****P*<.001, ******P*<.0001.


A.
WBC per femur (x10^6)
***
Srsf2+/+ Tet2+/+
Srsf2 P95H/+
Tet2-/-
Srsf2 P95H/+ Tet2-/-
B.
Cells per femur (x10^6)
Granulocytes
Macrophages
C.
Cells per femur (x10^6)
B220+IgM+
B220+IgM-
Pre-Pro B
Pro B
Pre B
D.
Cells per femur (x10^6)
Ter119+ CD71hi
Ter119+ CD71med
Ter119+ CD71low
CD4+
CD8+

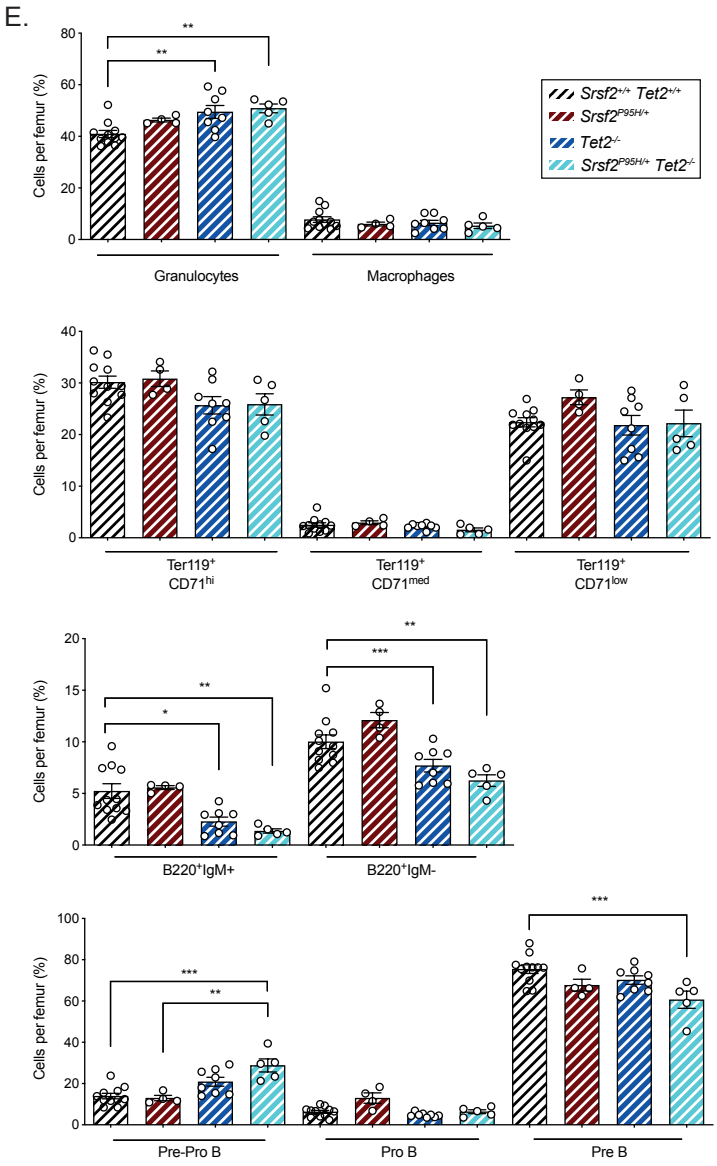


Figure 6.6: Co-expression of *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ leads to significant myeloid bias and B cell suppression in the bone marrow 52 weeks after gene mutation

A. Co-expression of *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ mutation leads to hypocellular bone marrow.

B. Number of granulocytes and macrophages in the bone marrow.

C. Co-expression of *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ mutation leads to B cell suppression from PreB to mature B cells (B220^{+} IgM^{+})

D. Co-expression of *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ mutation leads to decreased erythroid precursors while displaying no effect in T cells.

E. Percentage of myeloid, erythroid and B lymphoid lineages in the bone marrow. IgM, immunoglobin M. n=4-11 per genotype. Presented as mean $\pm$ standard error of mean. One-way ANOVA. *P<.05, **P<.01, ***P<.001,****P<.0001.

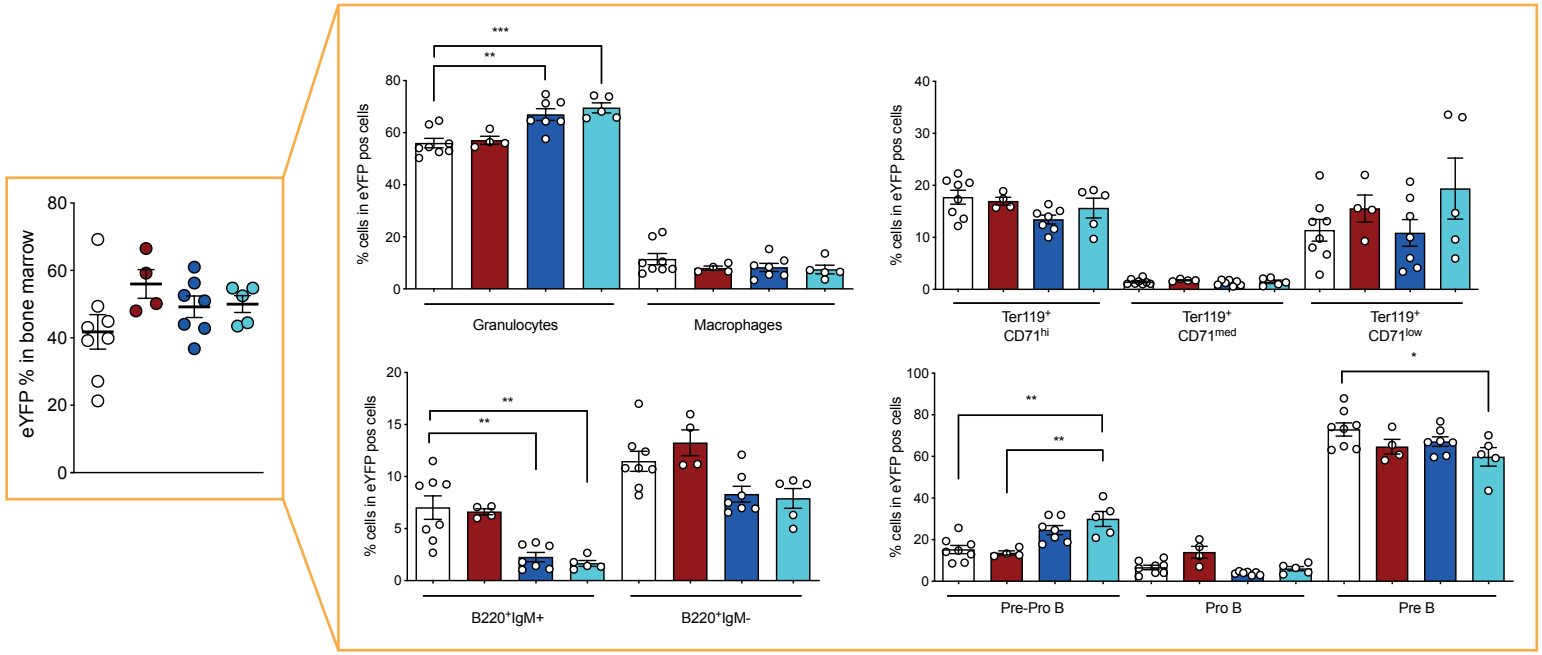
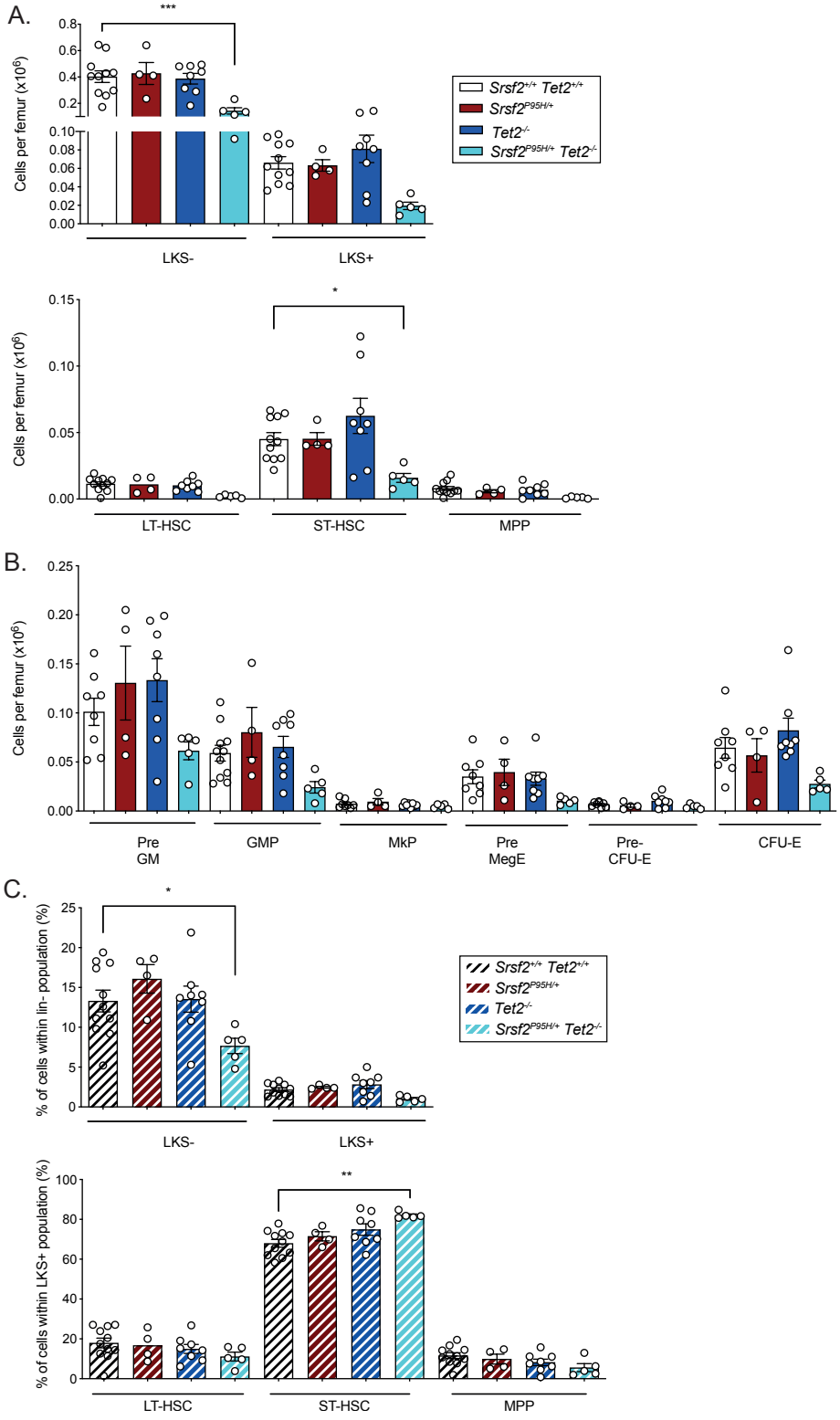


Figure 6.7: The eYFP positive *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ bone marrow cells are significantly biased towards the myeloid lineage instead of B lymphoid lineage
IgM, immunoglobin M. n=4-11 per genotype. Presented as mean ± standard error of mean. One-way ANOVA. *P<.05, **P<.01, ***P<.001.


A.
Cells per femur (x10^6)
***
LKS-
LKS+
Srsf2+/+ Tet2+/+
Srsf2P95H/+
Tet2-/-
Srsf2P95H/+ Tet2-/-
*
LT-HSC
ST-HSC
MPP
B.
Pre GM
GMP
MkP
Pre MegE
Pre-CFU-E
CFU-E
C.
% of cells within lin- population (%)
% of cells within LKS+ population (%)
**

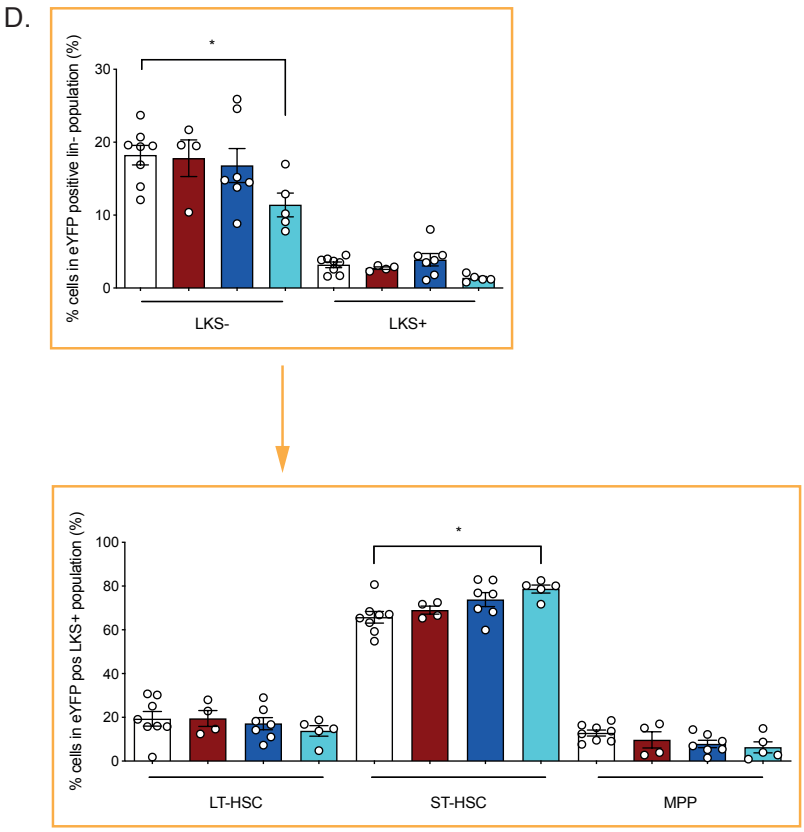
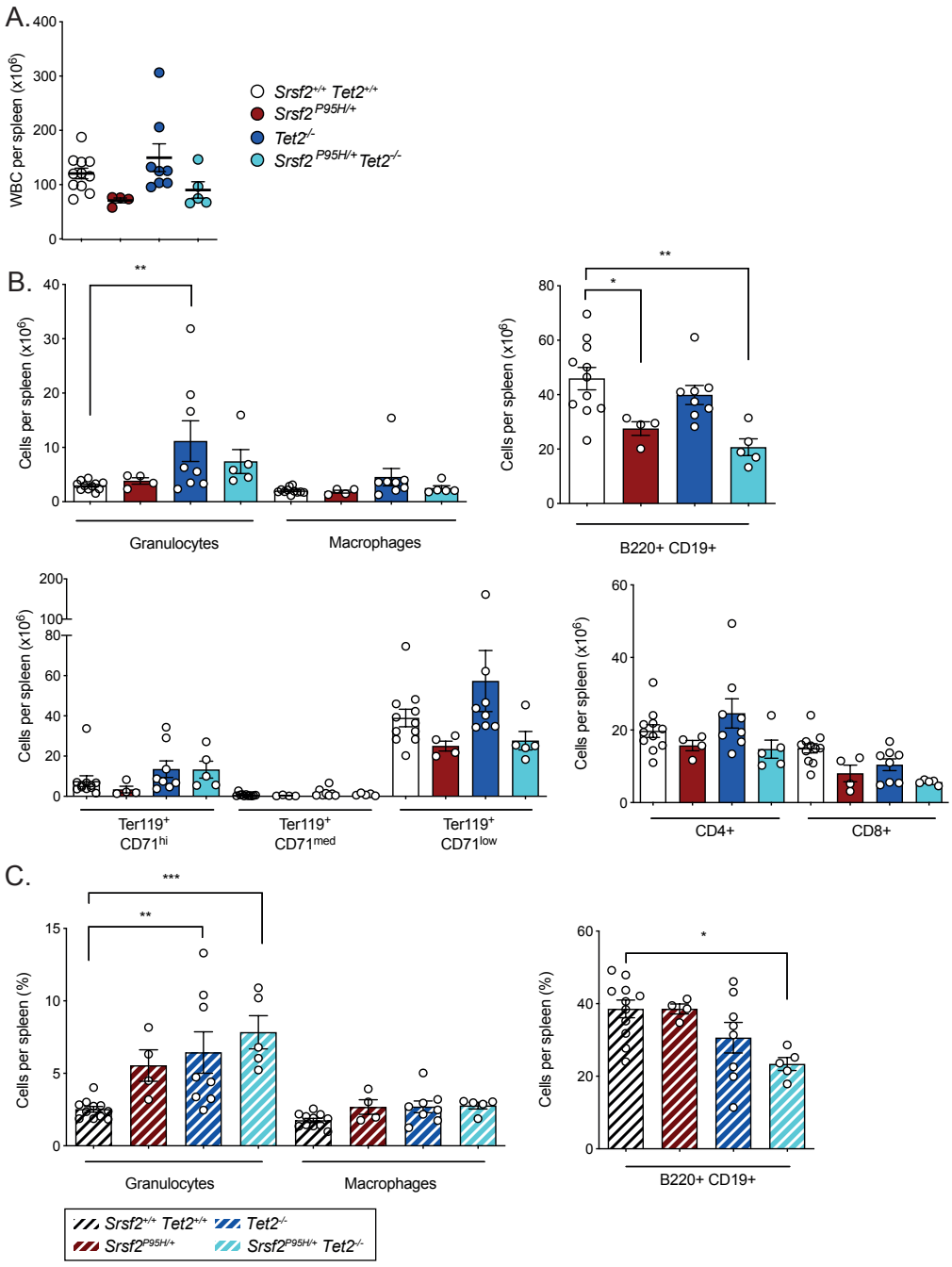


Figure 6.8: Co-expression of *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ leads to an increased proportion of phenotypic short-term haematopoietic stem cells
A. Number of LKS^{-}, LKS^{+} and phenotypic LT-HSC, ST-HSC and MPP in LKS^{+} population using CD34/FLT3.
B. Number of myeloerythroid progenitors per femur.
C. Percentage of LKS^{-} and LKS^{+} within the lineage negative population and LT-HSC, ST-HSC and MPP in the LKS^{+} population.
D. Percentage of LKS^{-} and LKS^{+} within the eYFP positive lineage negative population and LT-HSC, ST-HSC and MPP in the eYFP positive LKS^{+} population.
LKS^{-}, lin^{-}c-Kit^{+}Sca-1^{-}; LKS^{+}, lin^{-}c-Kit^{+}Sca-1^{+}; LT-HSC, long-term haematopoietic stem cell; ST-HSC, short-term haematopoietic stem cell; MPP, multipotent progenitor; PreGM, Pre-granulocyte-macrophage; GMP, granuocyte-macrophage progenitors; MkP, megakaryocyte progenitor; PreMegE, premegakaryocyte erythroid progenitor; Pre-CFU-E, pre-colony-forming unit erythroid; CFU-E, colony-forming unit erythroid.
n=4-11 per genotype. Presented as mean ± standard error of mean. One-way ANOVA. *P<.05, **P<.01, ***P<.001.


A.
WBC per spleen (x10⁶)
400
300
200
100
0
Srsf2+/+ Tet2+/+
Srsf2P95H/+
Tet2-/-
Srsf2P95H/+ Tet2-/-
B.
Cells per spleen (x10⁶)
**
Granulocytes
Macrophages
*
**
B220+ CD19+
Ter119+ CD71hi
Ter119+ CD71med
Ter119+ CD71low
CD4+
CD8+
C.
Cells per spleen (%)
***
**
*
Granulocytes
Macrophages
B220+ CD19+
Srsf2+/+ Tet2+/+
Tet2-/-
Srsf2P95H/+
Srsf2P95H/+ Tet2-/-

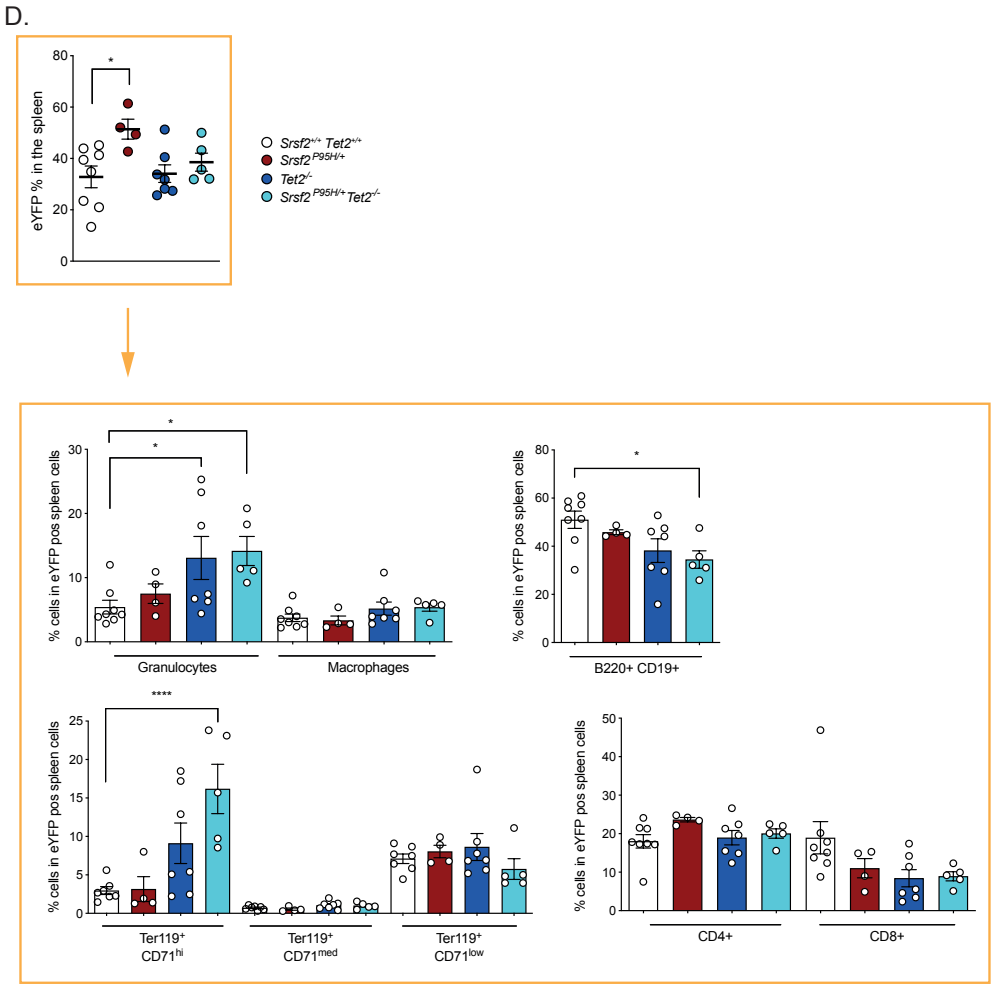


Figure 6.9: *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ mutation leads to myeloid bias and B cell suppression in the spleen
A. Spleen cellularity.
B. B cell suppression is significant in the spleen of *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ mice.
C. *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ mice have a significantly higher percentage of granulocytes and lower percentage of B cells.
D. The eYFP positive fraction of *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ spleen is biased towards myeloid lineage and early erythroid precursors instead of B cells.
n=4-11 per genotype. Presented as mean ± standard error of mean. One-way ANOVA. *P<.05, **P<.01, ***P<.001.

6.2.3 h*Scl*-CreER^{T} *R26*eYFP *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ cells stably engraft but cannot compete with wild-type bone marrow in transplantation assays

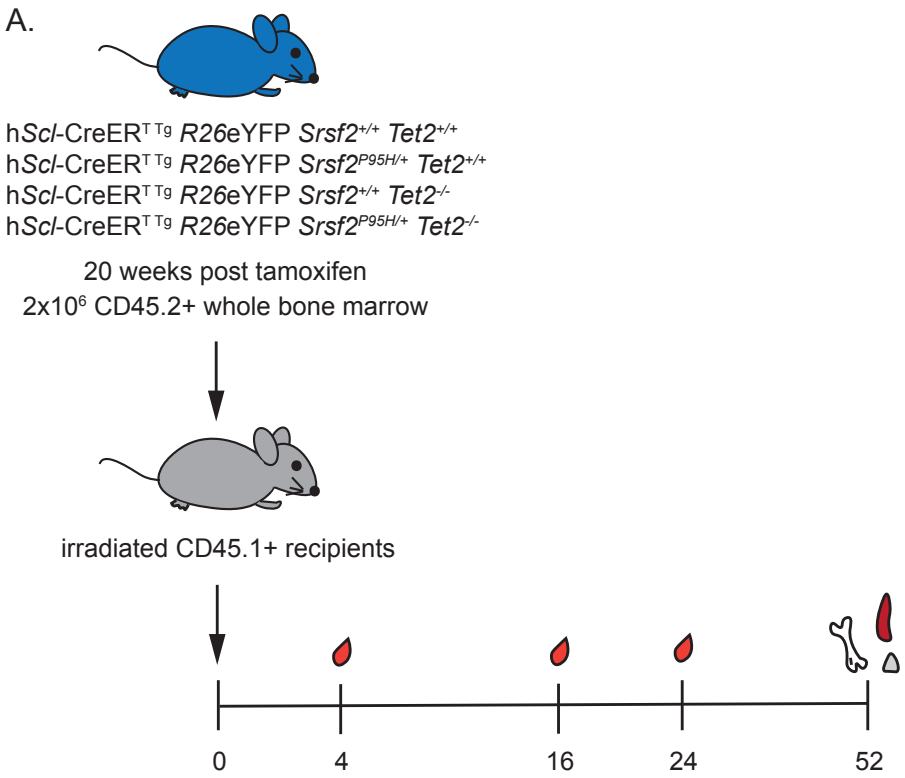
To determine the effect of compound *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ mutation on haematopoiesis following stress, bone marrow transplantation analysis was carried out. I performed both non-competitive and competitive transplants using whole bone marrow of h*Scl*-CreER^{T} *R26*eYFP *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ mice 20 weeks after activation of the mutations. I also transplanted the tamoxifen-treated and age-matched bone marrow from Cre-positive wild-type (WT, +/+), Srsf2 mutant (*Srsf2*$^{P95H/+}$) and Tet2 null (*Tet2*$^{-/-}$) mice. For non-competitive transplant, a total of 2 million whole bone marrow cells (cells from the WT and mutant mice are CD45.2^{+}) were injected intravenously into irradiated congenic recipient mice (CD45.1^{+}) (Figure 6.10A). For competitive transplant, 1 million donor whole bone marrow cells (CD45.2^{+}) along with 1 million bone marrow cells from a wild-type donor (CD45.1/CD45.2^{+}; competitor cells) were intravenously injected into irradiated congenic recipient mice (CD45.1^{+}) (Figure 6.11A). The transplant cohort was then regularly monitored for up to one-year post transplantation.

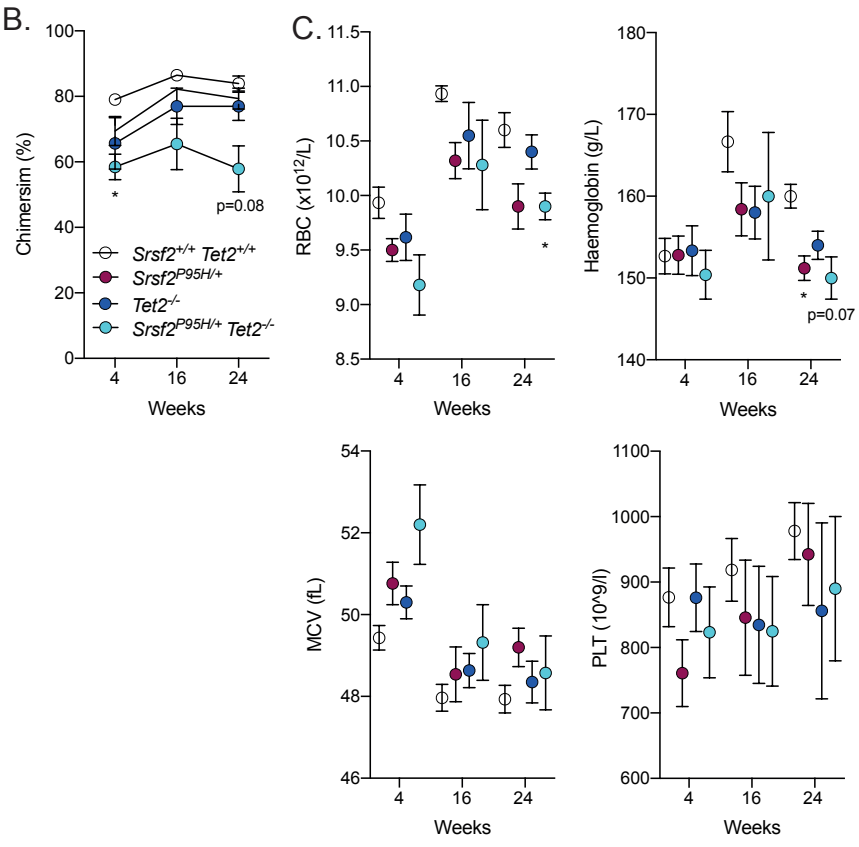
6.2.3.1 Double mutant cells undergo myeloid biased differentiation at the expense of the lymphoid lineage

Non-competitive transplantation – From four weeks after transplantation, recipient mice were monitored through peripheral blood analysis to detect chimerism (CD45.2%) and lineage distribution. In non-competitive transplants, the double mutant cells achieved a stable but lower engraftment level compared to wild-type cells and this remained low at 24 weeks

post-transplantation (Figure 6.10B). By 24 weeks post-transplant, the double mutant recipients had decreased red blood cells, haemoglobin and leukocytes, which indicates multilineage cytopenia and anaemia (Figure 6.10C). In terms of lineage distribution, double mutant cell recipients had significantly lower B cell frequencies throughout the analysis post-transplant. There was a trend of increased granulocytes, but the change was not significant. There were no changes in T cells (Figure 6.10D). Within the CD45.2^{+} donor population, there was an increased percentage of granulocytes and decline in B cells at 16- and 24 weeks post transplantation. In contrast, the T cell percentages were high at the 16 weeks timepoint but not at 24 weeks post transplantation (Figure 6.10E).

Competitive transplantation - Upon competitive transplantation, the double mutant cells displayed a significantly reduced engraftment compared to wild-type and *Tet2*$^{-/-}$ BM recipients, though the reconstitution was comparable to the *Srsf2* mutant group (Figure 6.11B). In terms of lineage distribution, the recipients of the double mutant cells had a trend of decreased B cell percentages in the total population and elevated granulocyte percentages arising from the CD45.2^{+} population, although these changes were not significant (Figure 6.11C-D). The myeloid bias was not as evident as in the non-competitive transplant, which is likely due to the lower overall engraftment of the double mutant cells.


A.
hScl-CreER$^{T\,Tg}$ R26eYFP Srsf2$^{+/+}$ Tet2$^{+/+}$
hScl-CreER$^{T\,Tg}$ R26eYFP Srsf2$^{P95H/+}$ Tet2$^{+/+}$
hScl-CreER$^{T\,Tg}$ R26eYFP Srsf2$^{+/+}$ Tet2$^{-/-}$
hScl-CreER$^{T\,Tg}$ R26eYFP Srsf2$^{P95H/+}$ Tet2$^{-/-}$
20 weeks post tamoxifen
2x10^6 CD45.2+ whole bone marrow
irradiated CD45.1+ recipients
0
4
16
24
52


B.
Chimersim (%)
p=0.08
Srsf2$^{+/+}$ Tet2$^{+/+}$
Srsf2$^{P95H/+}$
Tet2$^{-/-}$
Srsf2$^{P95H/+}$ Tet2$^{-/-}$
Weeks
C.
RBC (x10^{12}/L)
Haemoglobin (g/L)
p=0.07
MCV (fL)
PLT (10^9/l)
Weeks

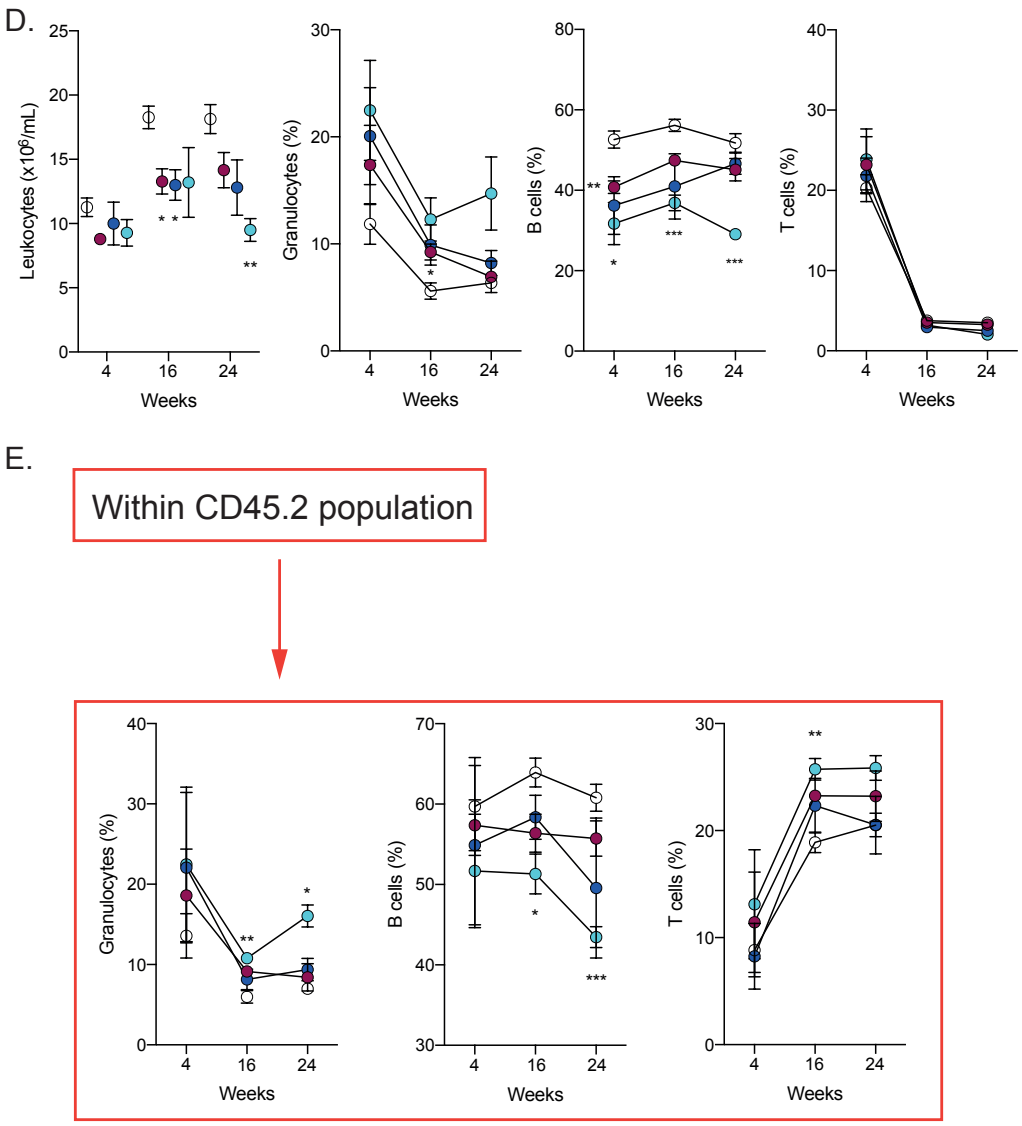
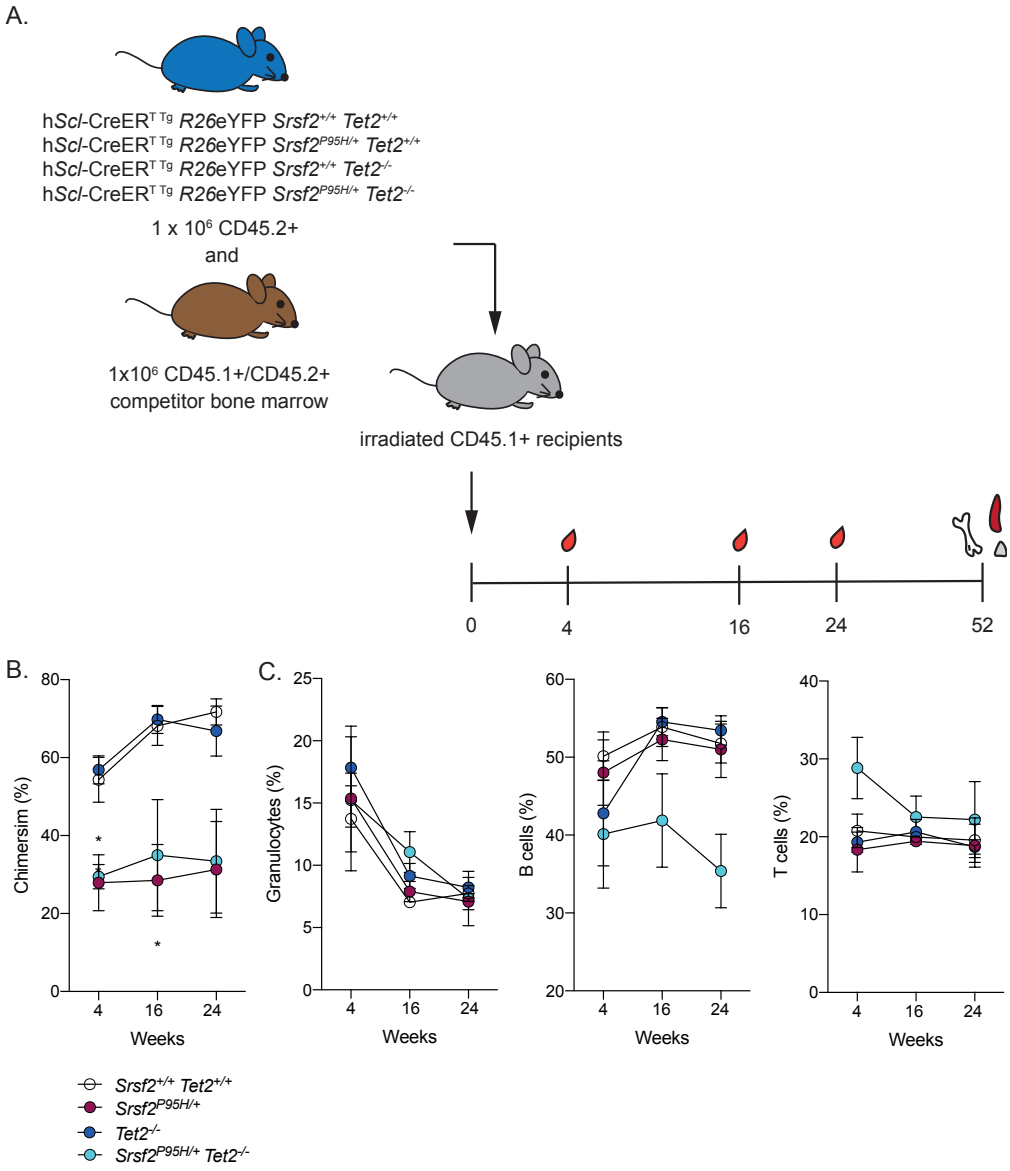


Figure 6.10: *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ cells have decreased engraftment capacity and strong myeloid bias with B cell suppression
A. Schematic illustration of non-competitive transplantation experiments.
B. *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ cells achieve a lower but stable engraftment in non-competitive transplantation.
C. PB indices of recipients up to 24 weeks post transplantation.
D. PB leukocyte counts and lineage distribution of non-competitive transplant recipients.
E. PB lineage distribution of donor (CD45.2^{+}) population in non-competitive transplant recipients.
n=3-6 per genotype. Presented as mean ± standard error of mean. One-way ANOVA performed against wild-type cells. *P<.05, **P<.01, ***P<.001.


A.
hScl-CreER$^{T\,Tg}$ R26eYFP Srsf2$^{+/+}$ Tet2$^{+/+}$
hScl-CreER$^{T\,Tg}$ R26eYFP Srsf2$^{P95H/+}$ Tet2$^{+/+}$
hScl-CreER$^{T\,Tg}$ R26eYFP Srsf2$^{+/+}$ Tet2$^{-/-}$
hScl-CreER$^{T\,Tg}$ R26eYFP Srsf2$^{P95H/+}$ Tet2$^{-/-}$
1 x 10^6 CD45.2+
and
1x10^6 CD45.1+/CD45.2+
competitor bone marrow
irradiated CD45.1+ recipients
0
4
16
24
52
B.
Chimersim (%)
Weeks
C.
Granulocytes (%)
B cells (%)
T cells (%)
Srsf2$^{+/+}$ Tet2$^{+/+}$
Srsf2$^{P95H/+}$
Tet2$^{-/-}$
Srsf2$^{P95H/+}$ Tet2$^{-/-}$

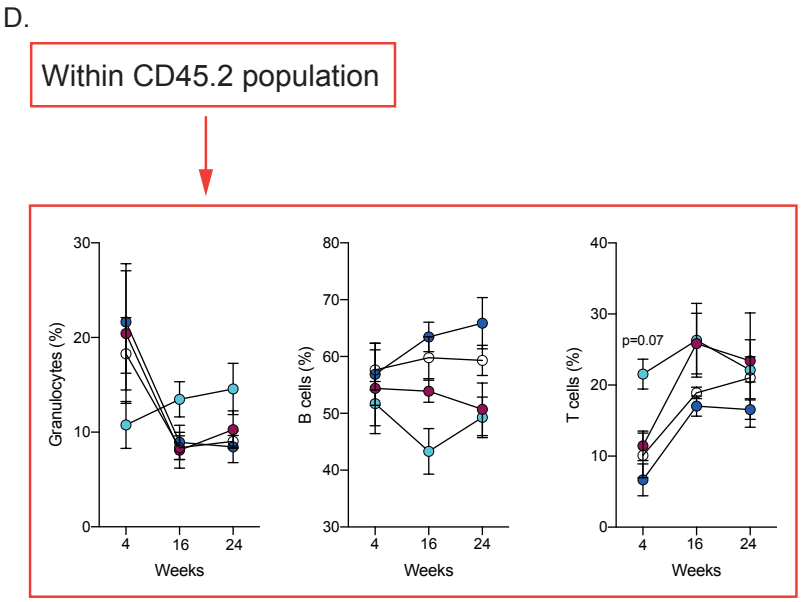


Figure 6.11: *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ cells are out-competed by wild-type cells upon competitive transplantation
A. Schematic illustration of competitive transplantation experiments.
B. *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ cells have significantly lower engraftment in competitive transplantation.
C. PB lineage distribution of competitive recipients up to 24 weeks post transplantation.
D. PB lineage distribution of donor (CD45.2^{+}) population in competitive transplant recipients.
n=3-6 per genotype. Presented as mean ± standard error of mean. One-way ANOVA performed against wild-type cells. *P<.05, **P<.01, ***P<.001.

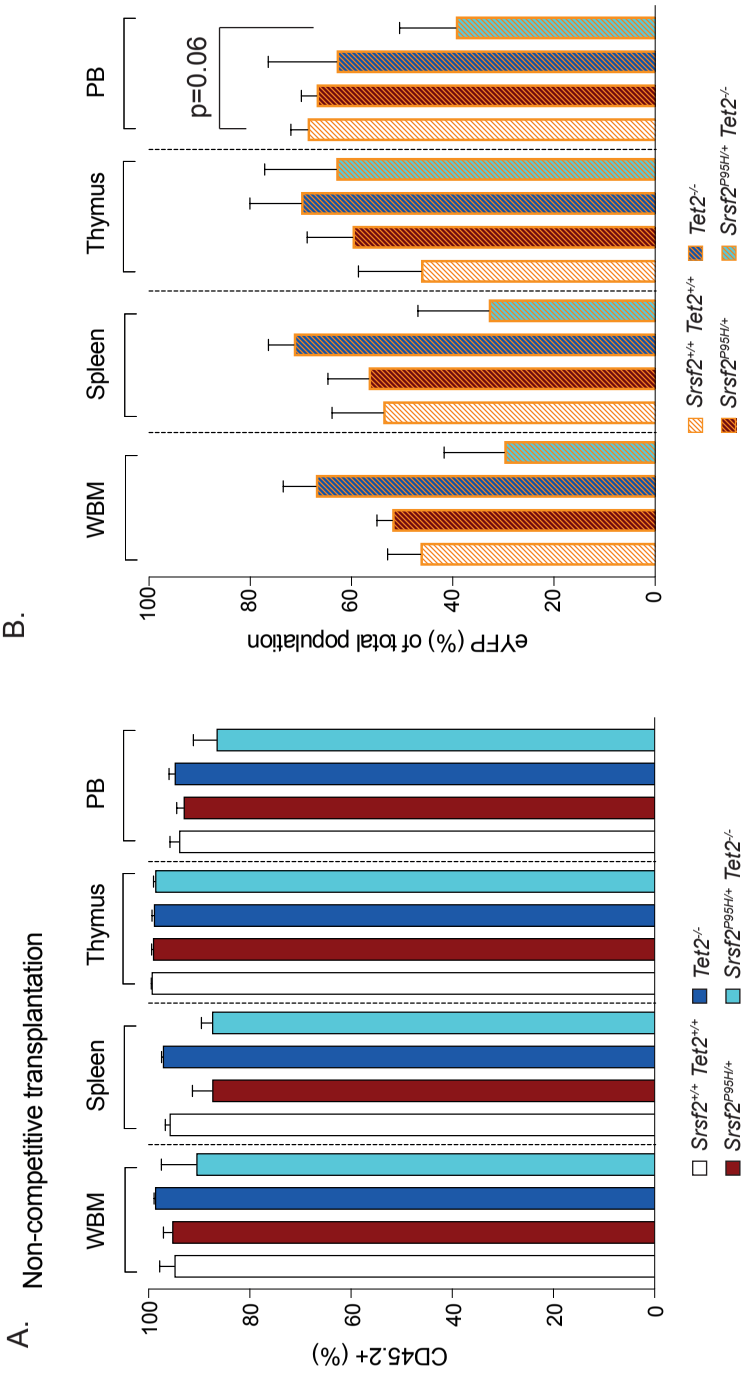
6.2.3.2 Double mutant cells still exhibit a myeloid bias but erythropoiesis is improved in aged stress haematopoiesis

Cohorts of both non-competitive and competitive transplantation recipients were aged for up to 52 weeks post-transplantation to analyse the long-term effect of double mutation following reconstitution. In the non-competitive transplant cohort, the double mutant recipients achieved stable but lower engraftment in all tissues (Figure 6.12A). Interestingly, double mutant recipients had a lower eYFP positive percentage in the peripheral blood compared to *Srsf2*$^{+/+}$ recipients. In bone marrow and spleen, the eYFP positive percentages were also lower, but the difference was not significant (Figure 6.12B). I then compared the eYFP percentage within each tissue at 52 weeks post-transplantation to the eYFP percentage of the donor tissues. This comparison allowed me to assess whether the mutant cell populations have expanded upon aging in the transplant recipients. Compared to the eYFP percentage in donor tissues, double mutant cells showed no expansion at 52 weeks post transplantation, while single *Srsf2*$^{P95H/+}$ and single *Tet2*$^{-/-}$ cell populations were increased in the recipients in bone marrow and spleen. There was no change in eYFP percentage in the thymus in any of the genotypes (Figure 6.12C). In the competitive transplant cohort, the compound mutant recipients had a worse engraftment compared to wild-type and *Tet2*$^{-/-}$ mutant recipients, but similar to *Srsf2* mutant recipients. The overall eYFP positivity was lowest in the double mutant recipients and in the *Srsf2* mutant recipients (Figure 6.13).

In the peripheral blood, the compound mutant non-competitive recipients showed no difference in red blood cell numbers, haemoglobin or mean corpuscular volume (MCV), unlike the data from 24 weeks post transplantation. The platelet and leukocytes numbers were largely

normal, although there was variation within the double mutant recipients (Figure 6.14A-B). Assessing lineage distribution, the double mutant cells displayed no difference from other genotypes. There was a large variation in leukocyte numbers in the double mutant recipients due to the small sample size (Figure 6.14B). One of the compound mutant recipients (#835, tangerine circle) developed high leukocyte counts, which was driven by increased granulocytes and monocytes in the peripheral blood. I will discuss the haematological features of this individual in Section 6.2.4.

In the bone marrow, the cellularity was normal unlike the hypocellular bone marrow that was present in the native (non-transplanted) compound mutant mice (Figure 6.15A). However, an elevated percentage of granulocytes was present in the bone marrow of the double mutants (Figure 6.15B). The number of immature B cells at the PreB stage was significantly decreased (Figure 6.15C). Interestingly, there was an accumulation of late stage erythroid precursors in the bone marrow in the total population, which may be the source of improved red blood cell numbers and haemoglobin levels in the double mutants (Figure 6.15D). Overall, these analyses demonstrate that myeloid bias and B cell suppression is a cell-intrinsic change driven by *Srsf2/Tet2* mutation. The double mutation also decreased the engraftment capacity of stem cells when transplanted competitively.


A.
Non-competitive transplantation
CD45.2+ (%)
0
20
40
60
80
100
WBM
Spleen
Thymus
PB
Srsf2+/+ Tet2+/+
Srsf2P95H/+
Tet2-/-
Srsf2P95H/+ Tet2-/-
B.
eYFP (%) of total population
p=0.06

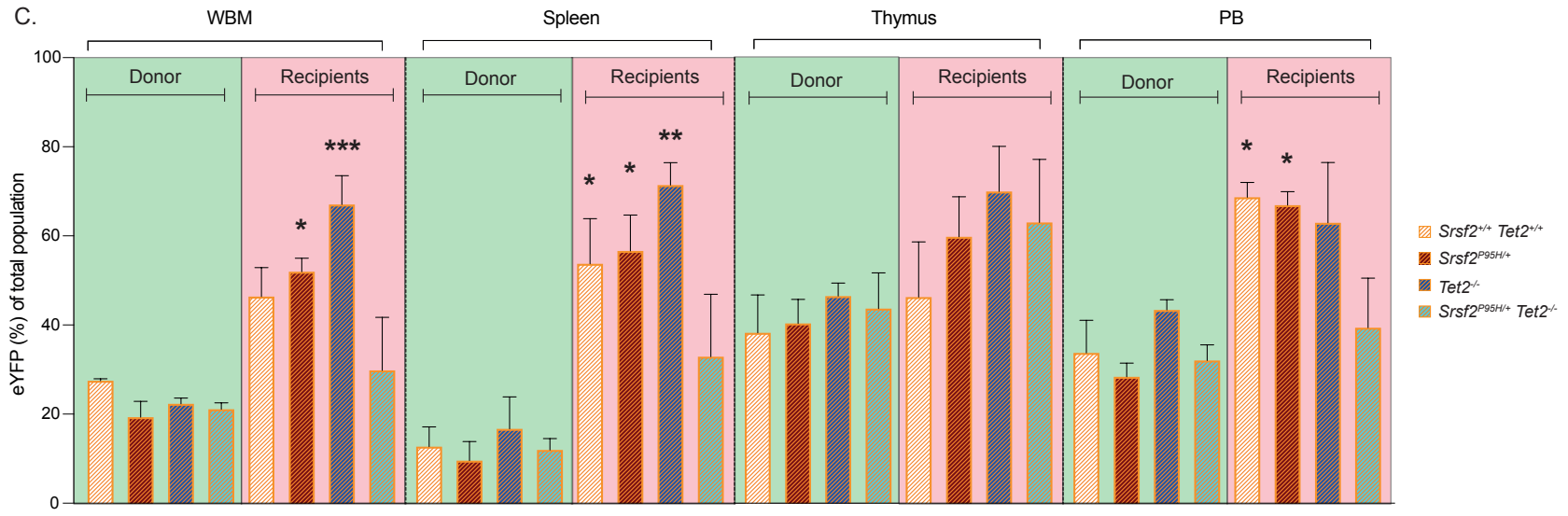


Figure 6.12: *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ cells can stably engraft haematopoietic tissues 52 weeks post transplantation
A. The percentage of donor population (CD45.2^{+}%) at 52 weeks post non-competitive transplantation.
B. The eYFP (%) in all haematopoietic tissues at 52 weeks post non-competitive transplantation.
C. The comparison of eYFP (%) in all haematopoietic tissues in donors (green shaded area) and in recipients (pink shaded area). Within each tissue, the eYFP percentage at 52 weeks post transplantation is compared to the eYFP percentage of the same genotype at 20 weeks post mutation activation.
n=3-6 per genotype. Presented as mean ± standard error of mean. One-way ANOVA performed against wild-type cells. *P<.05, **P<.01, ***P<.001.

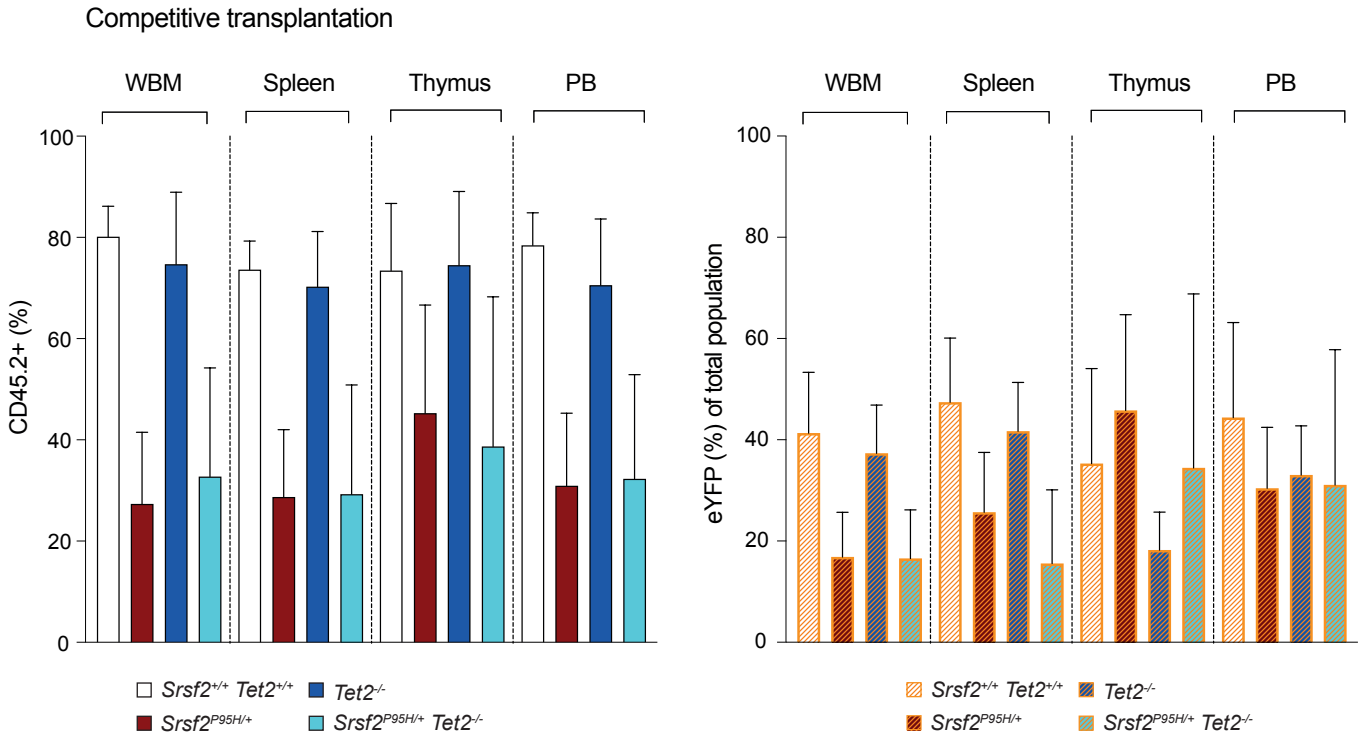


Figure 6.13: *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ cells are out-competed by wild-type cells and fail to efficiently engraft haematopoietic tissues at 52 weeks post transplantation

The percentage of donor population (CD45.2^{+}%) and eYFP (%) in all haematopoietic tissues at 52 weeks post competitive transplantation.

n=2-6 per genotype. Presented as mean ± standard error of mean.

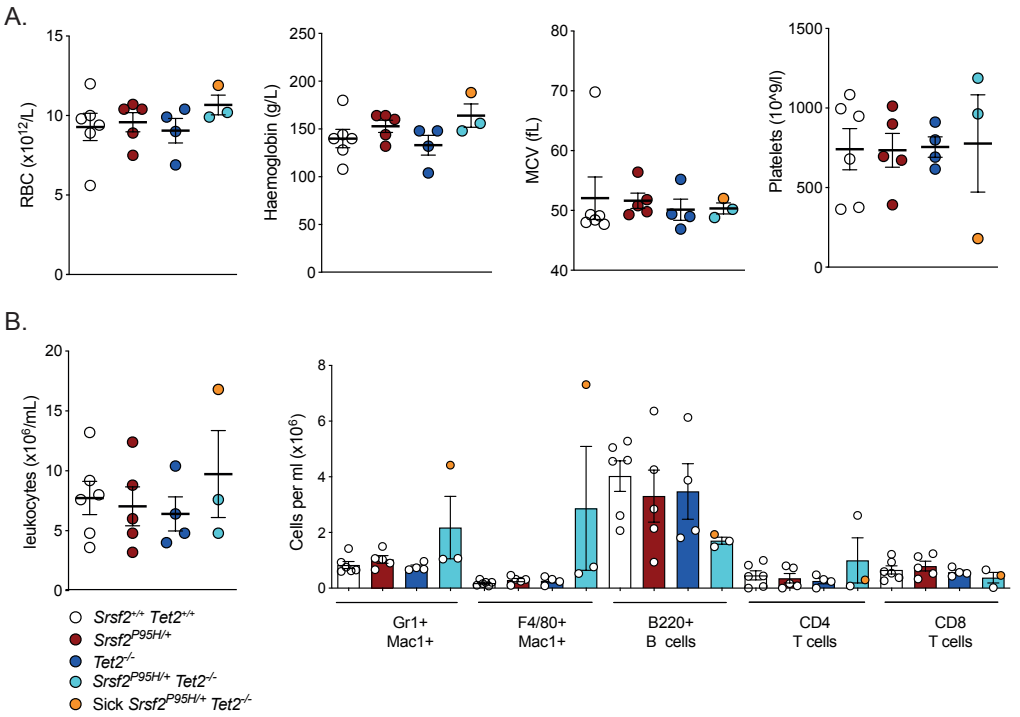
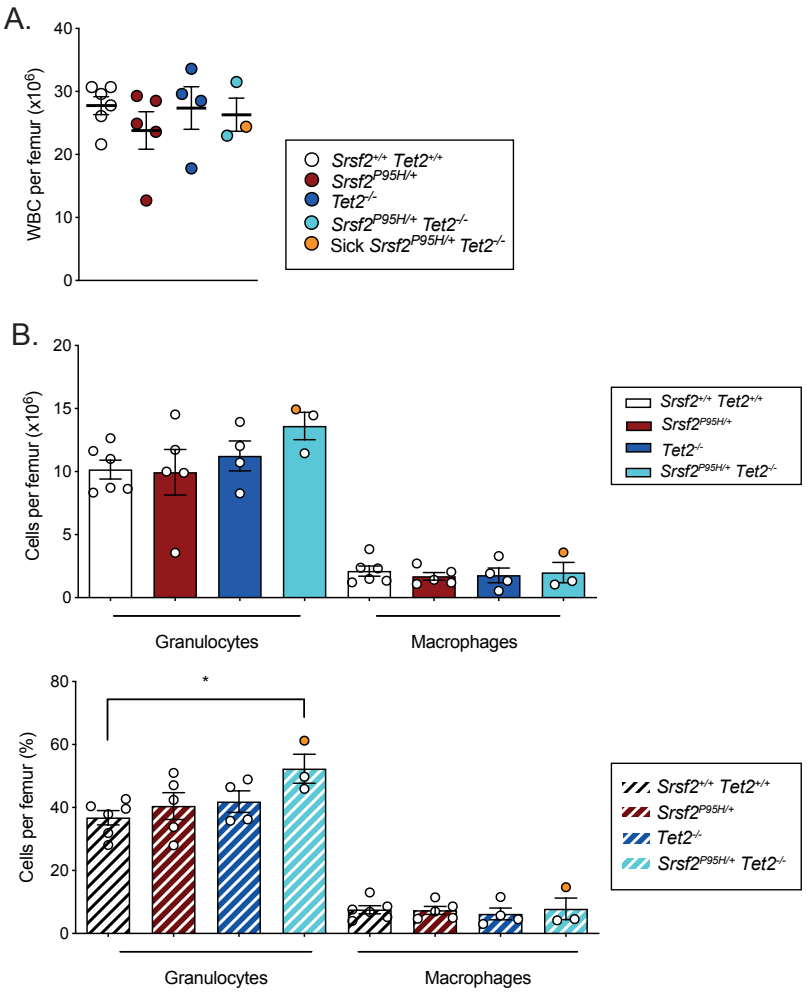


Figure 6.14: PB indices and lineage distribution of non-competitive transplant recipients at 52 weeks post transplantation
A. PB indices of the non-competitive transplant recipients.
B. PB leukocyte numbers and lineage distribution of the non-competitive transplant recipients.
The tangerine circle represents a diseased *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ transplant recipient.
See Figure 6.21 to 6.24 for more detailed analysis.
n=3-6 per genotype. Presented as mean ± standard error of mean.


A.
WBC per femur (x10⁶)
40
30
20
10
0
Srsf2+/+ Tet2+/+
Srsf2P95H/+
Tet2-/-
Srsf2P95H/+ Tet2-/-
Sick Srsf2P95H/+ Tet2-/-
B.
Cells per femur (x10⁶)
20
15
10
5
0
Srsf2+/+ Tet2+/+
Srsf2P95H/+
Tet2-/-
Srsf2P95H/+ Tet2-/-
Granulocytes
Macrophages
Cells per femur (%)
80
60
40
20
0
*
Srsf2+/+ Tet2+/+
Srsf2P95H/+
Tet2-/-
Srsf2P95H/+ Tet2-/-
Granulocytes
Macrophages

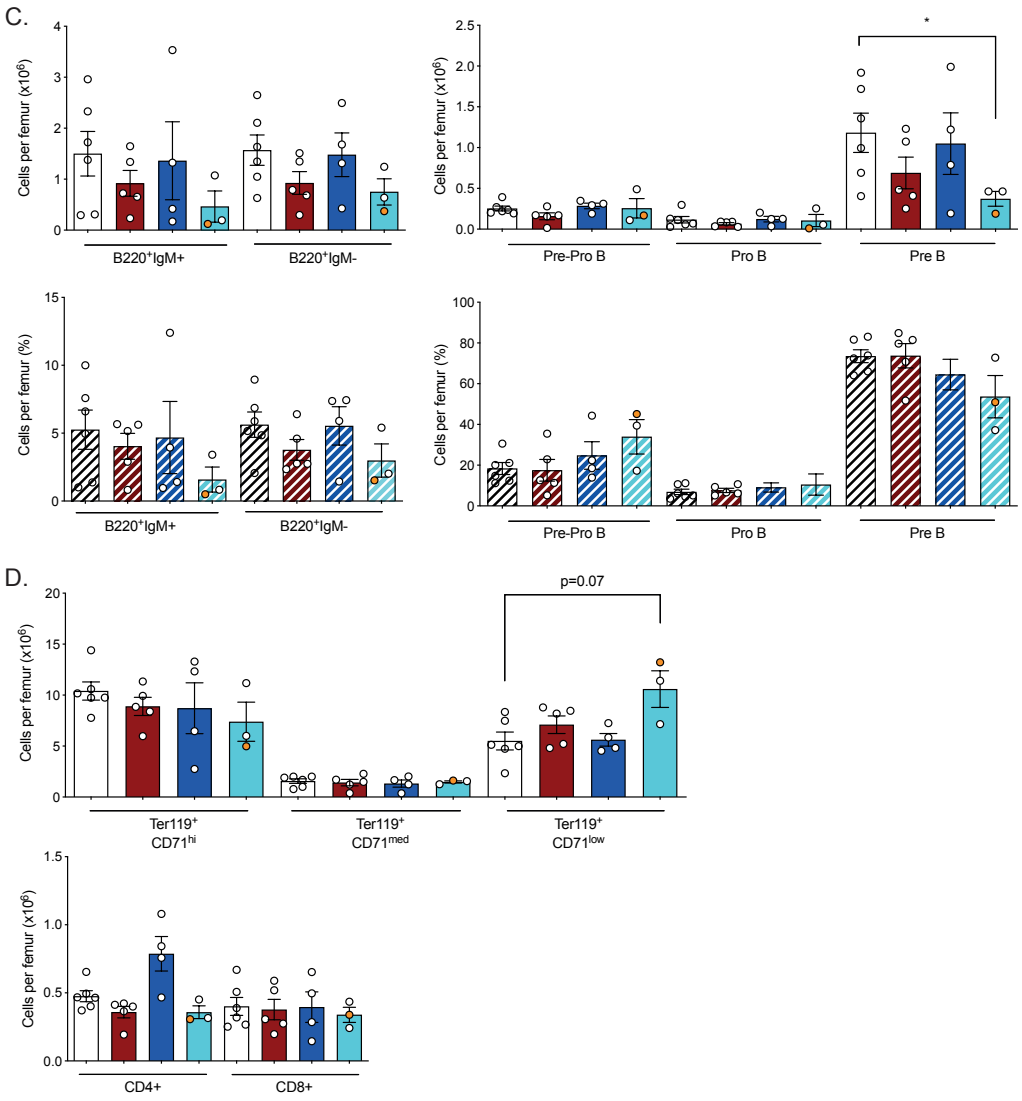


Figure 6.15: *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ non-competitive transplant recipients show alterations in myeloid, lymphoid and erythroid lineages in the bone marrow at 52 weeks post transplantation

A. Bone marrow cellularity of non-competitive transplant recipients.

B. Number and percentage of granulocytes and macrophages in the bone marrow.

C. Number and percentage of B cell precursors and mature B cells in the bone marrow.

IgM, immunoglobulin M. n=3-6 per genotype. Presented as mean ± standard error of mean. One-way ANOVA. *P<.05.

6.2.4 *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ leads to development of CMML *in vivo*

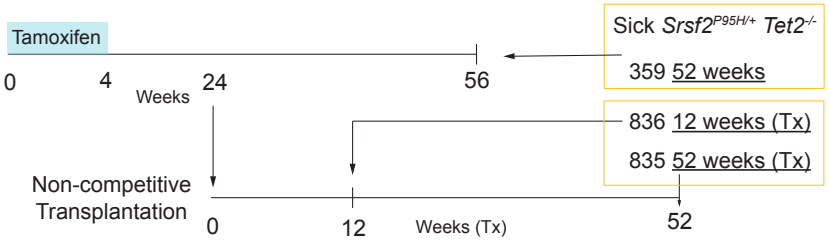
During the course of monitoring to date, three mice have developed phenotypes consistent with disease progression, one from an aging cohort of h*Scl*-CreER^{T} *R26*eYFP *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ (#359) and two from the non-competitive transplantation of bone marrow from 20 weeks post tamoxifen treated h*Scl*-CreER^{T} *R26*eYFP *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ mice (#836, 835). Of the three sick mice, #359 became moribund at 52 weeks post tamoxifen treatment, while #835 and #836 were detected at 52 weeks and 12 weeks post transplantation, respectively (Figure 6.16A). At the time of analysis, the bone marrow cellularity for the three mice was relatively normal or lower than the age-matched wild-type context-specific control (for #359 the average BM cellularity for control animals was 32.9 $\pm$ 1.2 x 10^6; for #835 and #836 the average BM cellularity for transplant controls was 27.8 $\pm$ 1.4 x 10^6). In contrast to other genotypes analysed at the same time, #359 and #835 displayed splenomegaly and increased splenic cellularity (Figure 6.16B-C). #836 also had an increased splenic cellularity, with a spleen weight of 183mg. In the peripheral blood, #835 and #836 had significantly increased leukocyte counts while #359 had a normal leukocyte level (Figure 6.16B). Analysis of peripheral blood smears demonstrated that there were dysplastic (hypersegmented) neutrophils in the blood of all three sick mice (Figure 6.16D). In the bone marrow cytospin of #359, there is an increased percentage of promonocytes and monoblasts (10-15%) along with occasional dysplastic monocytes (Figure 6.16E). A/Prof Meaghan Wall (Victorian Clinical Genetic Services) performed the analysis of bone marrow cytospins. The assessment of the blood film and bone marrow cytospin identified dysplastic features and monocytosis. This is consistent with pathological features of CMML, most closely aligning

to CMML-2 (5-19% peripheral blasts and 10-19% marrow blasts, WHO diagnostic criteria) (Patnaik and Tefferi, 2018). All three moribund mice also had a decreased percentage of CD45RB expression on splenocytes, demonstrating aberrant splicing of *Cd45* isoforms, which we previously reported in the *Srsf2*$^{P95H/+}$ model and which is a direct measure of the mis-splicing of the RNA of this protein (Figure 6.16F) (Smeets et al., 2018).

A.

8-10 weeks old

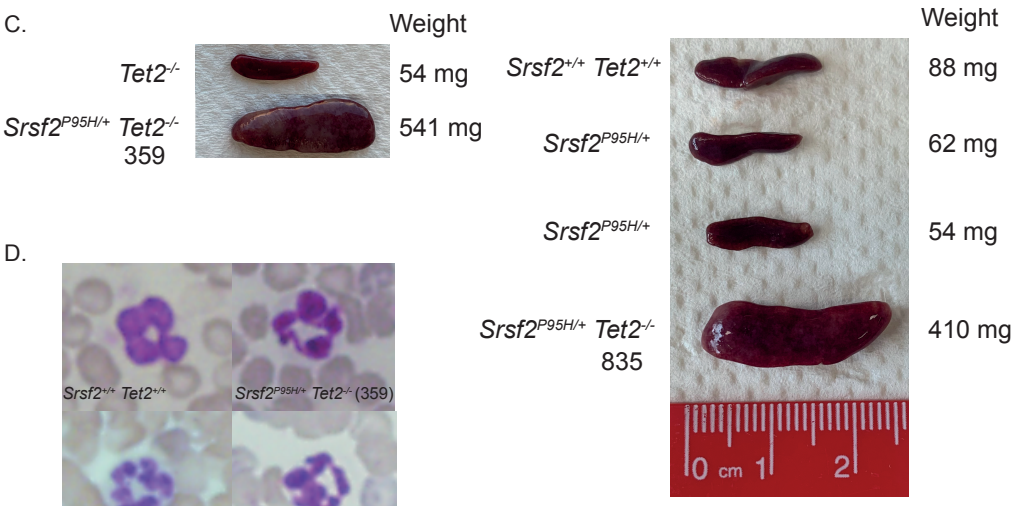
hScl-CreER$^{T\ Tg+}$ R26eYFP Srsf2$^{+/+}$ Tet2$^{+/+}$
hScl-CreER$^{T\ Tg+}$ R26eYFP Srsf2$^{P95H/+}$
hScl-CreER$^{T\ Tg+}$ R26eYFP Tet2$^{fl/fl}$
hScl-CreER$^{T\ Tg+}$ R26eYFP Srsf2$^{P95H/+}$ Tet2$^{fl/fl}$



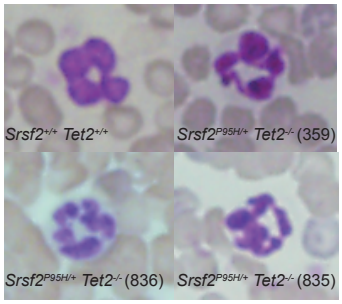
B.

Sick mouse	PB WBC (10^9/l)	Bone marrow cellularity per femur (10^6)	Spleen (10^6)
359	6.4	14.7	356.5
836	26.8	18.6	291
835	16.8	24.4	323.5

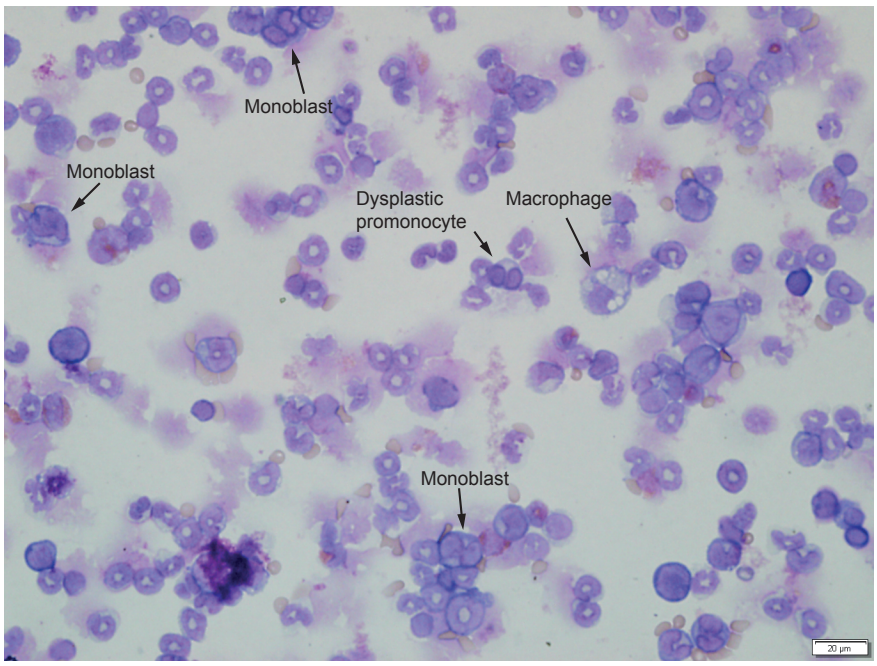
C.



D.



E.



F.

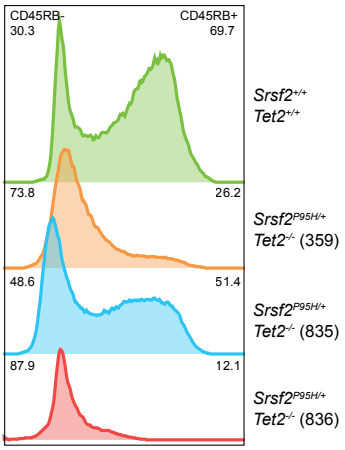


Figure 6.16: Sick *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ mice display features consistent with CMML including splenomegaly, dysplastic neutrophils and an increased number of immature monocytes

A. A timeline of sick mice post mutation initiation or post transplantation.
B. The cellularity of peripheral blood, bone marrow and spleen of moribund mice.
C. Images of the spleens of #359 and #835 and genotype controls in the cohort.
D. Representative image of dysplastic neutrophils in the peripheral blood of moribund mice. May-Grünwald-Giemsa staining. Image taken at 40x.
E. The bone marrow cytospin of #395 at the time of analysis. Cells from different maturation stages of the monocytic lineage are labelled. May-Grünwald-Giemsa. Image taken at 20x.
F. FACS plot of CD45RB expression on splenocytes of the moribund mice compared to a 52 weeks post tamoxifen wild-type control.

As the diseased animals were identified in different contexts and at different times, I will separate them and discuss them relative to their respective controls. I will also treat them as individual cases and discuss them in this context. For the purposes of presentation of these individuals I have included data for reference from the individual *Srsf2* and *Tet2* mutants and the double mutants from the most closely matched timepoint to the age the moribund mouse was identified. At the time of analysis, #359 had an increased MCV and decreased platelets compared to age-matched controls. The number of red blood cells and haemoglobin levels were slightly elevated (Figure 6.17A). Within the peripheral blood leukocytes, there was an increase in granulocytes and a dramatic decrease in the number of B cells, while the other cell populations remained normal (Figure 6.17B). The bone marrow was hypocellular compared to age/treatment matched control or single mutants with an increased percentage of granulocytes (Figure 6.18A-B). There was a profound suppression of both immature ($B220^{+}IgM^{-}$) and mature B cells ($B220^{+}IgM^{+}$). The number and percentage of Pro B and Pre B populations were decreased compared to control genotypes, while the percentage of Pre-Pro B cells were increased (Figure 6.18C). The erythroid lineages were affected as well, with a decrease in the immature erythroid population (Figure 6.18D). The myeloid bias and B cells suppression was more evident within the eYFP positive fraction of the peripheral blood in #359 where ~60% of the eYFP positive cells were granulocytes (Figure 6.19A). This was also the case in the BM (Figure 6.19B-C). In the spleen, the increased cellularity was the result of a profound expansion of myeloid cells, both granulocytes and monocytes, as well as increased immature erythroid progenitors ($Ter119^{+}$ $CD71^{hi}$). In contrast to the myeloid and erythroid expansion, there was a near absence of B cells in the spleen (Figure 6.20A-B). The eYFP positive fraction had similar phenotypes (Figure 6.20C). The thymus was relatively normal

(data not shown).

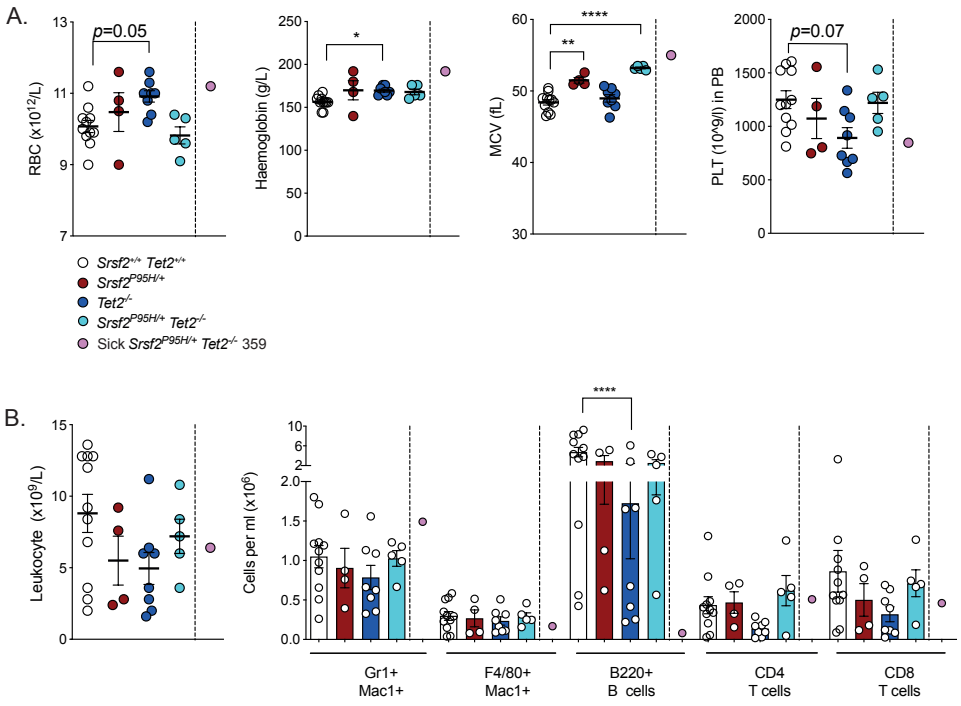
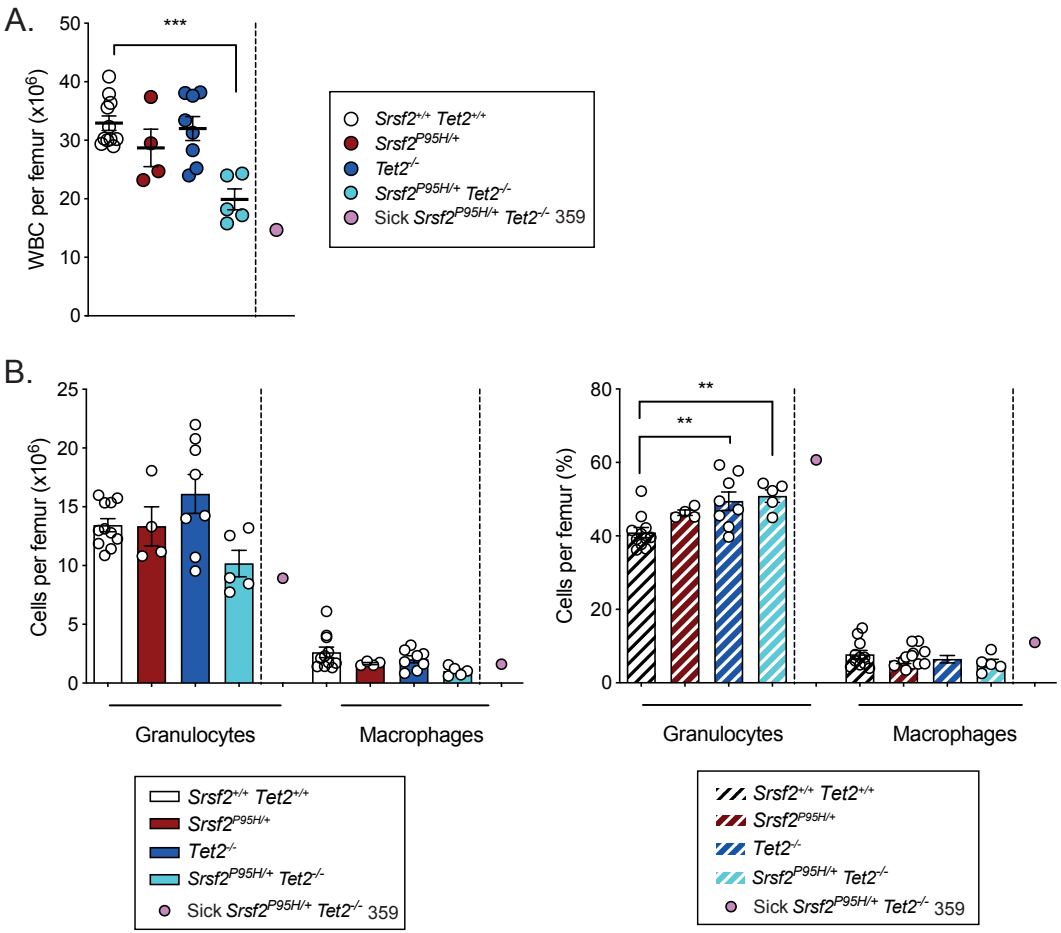
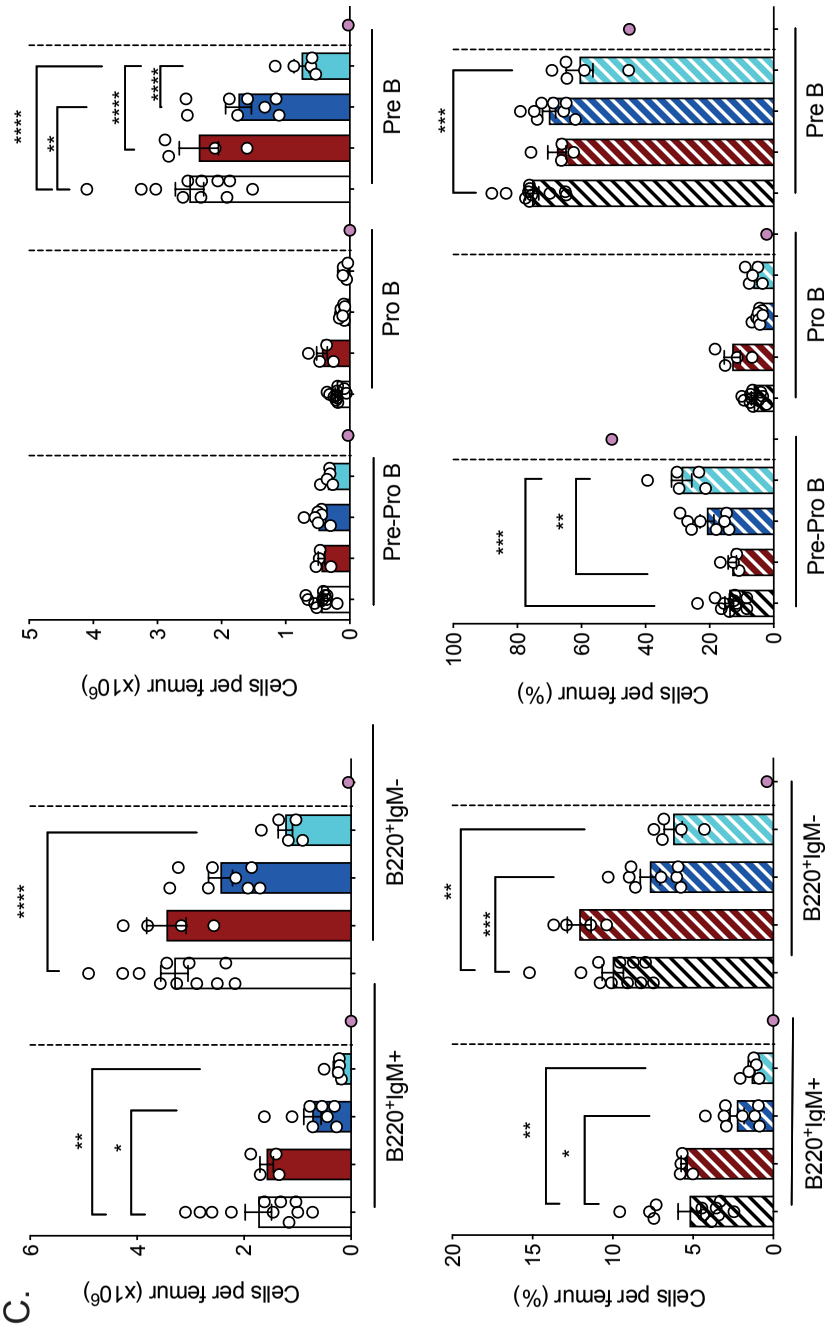


Figure 6.17: *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ mouse #359 developed myeloid bias and significant B cell suppression at 52 weeks after tamoxifen treatment
A. PB indices of #359 and other genotypes after 52 weeks of Cre activation.
B. PB leukocyte counts and lineage distribution.
n=4-11 for *Srsf2*$^{+/+}$ *Tet2*$^{+/+}$, *Srsf2*$^{P95H/+}$,*Tet2*$^{-/-}$ and *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ genotypes. Presented as mean ± standard error of mean. One-way ANOVA. **P*<.05, ***P*<.01, *****P*<.0001.


A.
WBC per femur (x10^6)
50
40
30
20
10
0
***
Srsf2+/+ Tet2+/+
Srsf2P95H/+
Tet2-/-
Srsf2P95H/+ Tet2-/-
Sick Srsf2P95H/+ Tet2-/- 359
B.
Cells per femur (x10^6)
25
20
15
10
5
0
Granulocytes
Macrophages
Srsf2+/+ Tet2+/+
Srsf2P95H/+
Tet2-/-
Srsf2P95H/+ Tet2-/-
Sick Srsf2P95H/+ Tet2-/- 359
Cells per femur (%)
80
60
40
20
0
**
**
Granulocytes
Macrophages
Srsf2+/+ Tet2+/+
Srsf2P95H/+
Tet2-/-
Srsf2P95H/+ Tet2-/-
Sick Srsf2P95H/+ Tet2-/- 359


C.
Cells per femur (x10⁶)
Cells per femur (%)
B220⁺IgM+
B220⁺IgM-
Pre-Pro B
Pro B
Pre B

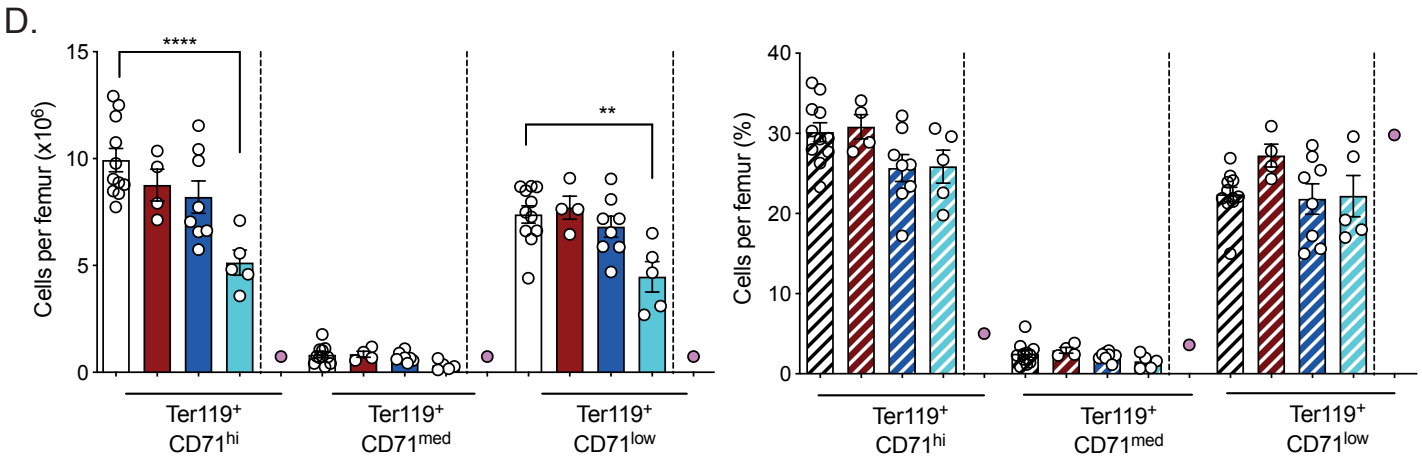


Figure 6.18: Moribund *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ mice displayed severe disruption of lymphoid and erythroid lineages, with myeloid bias in the bone marrow

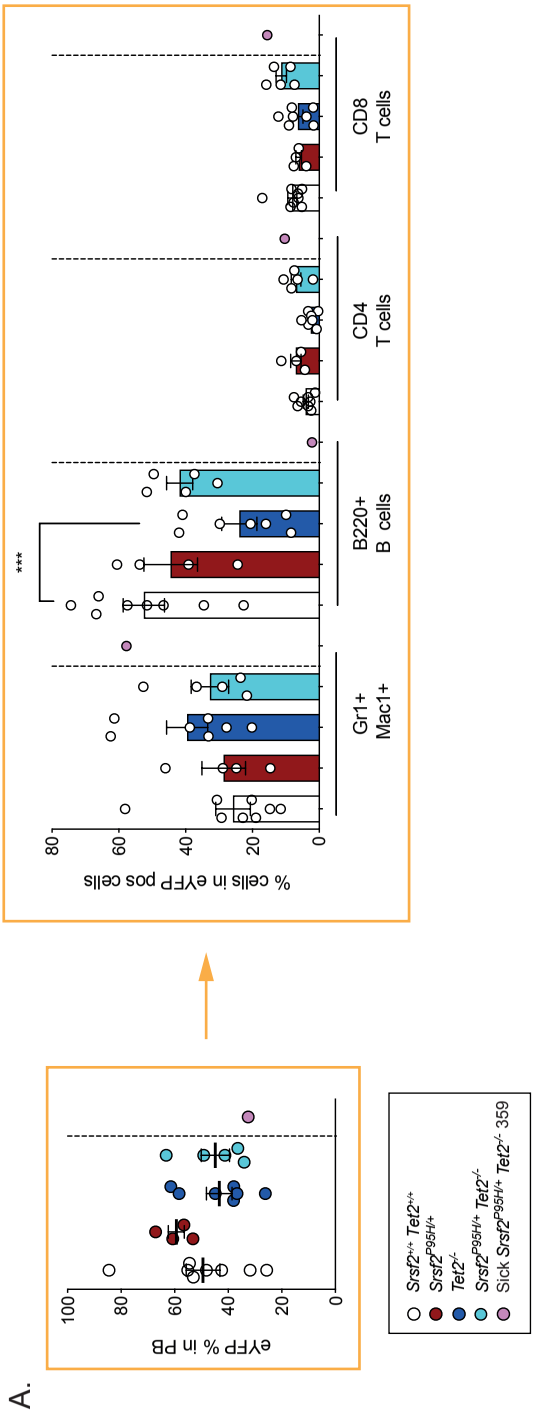
A. Bone marrow cellularity of #359, compared to other genotypes after 1 year of mutation activation.

B. The number and percentage of granulocytes in the bone marrow of #359.

C. The number and percentage of B lymphoid lineages in the bone marrow of #359.

D. The erythroid lineages in the bone marrow of #359.

IgM, immunoglobulin M. n=4-11 for *Srsf2*$^{+/+}$ *Tet2*$^{+/+}$, *Srsf2*$^{P95H/+}$,*Tet2*$^{-/-}$ and *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ genotypes. Presented as mean ± standard error of mean. One-way ANOVA. **P*<.05, ***P*<.01, ****P*<.001, *****P*<.0001.


A.
eYFP % in PB
0
20
40
60
80
100
Srsf2+/+ Tet2+/+
Srsf2P95H/+
Tet2-/-
Srsf2P95H/+ Tet2-/-
Sick Srsf2P95H/+ Tet2-/- 359
% cells in eYFP pos cells
***
Gr1+ Mac1+
B220+ B cells
CD4 T cells
CD8 T cells

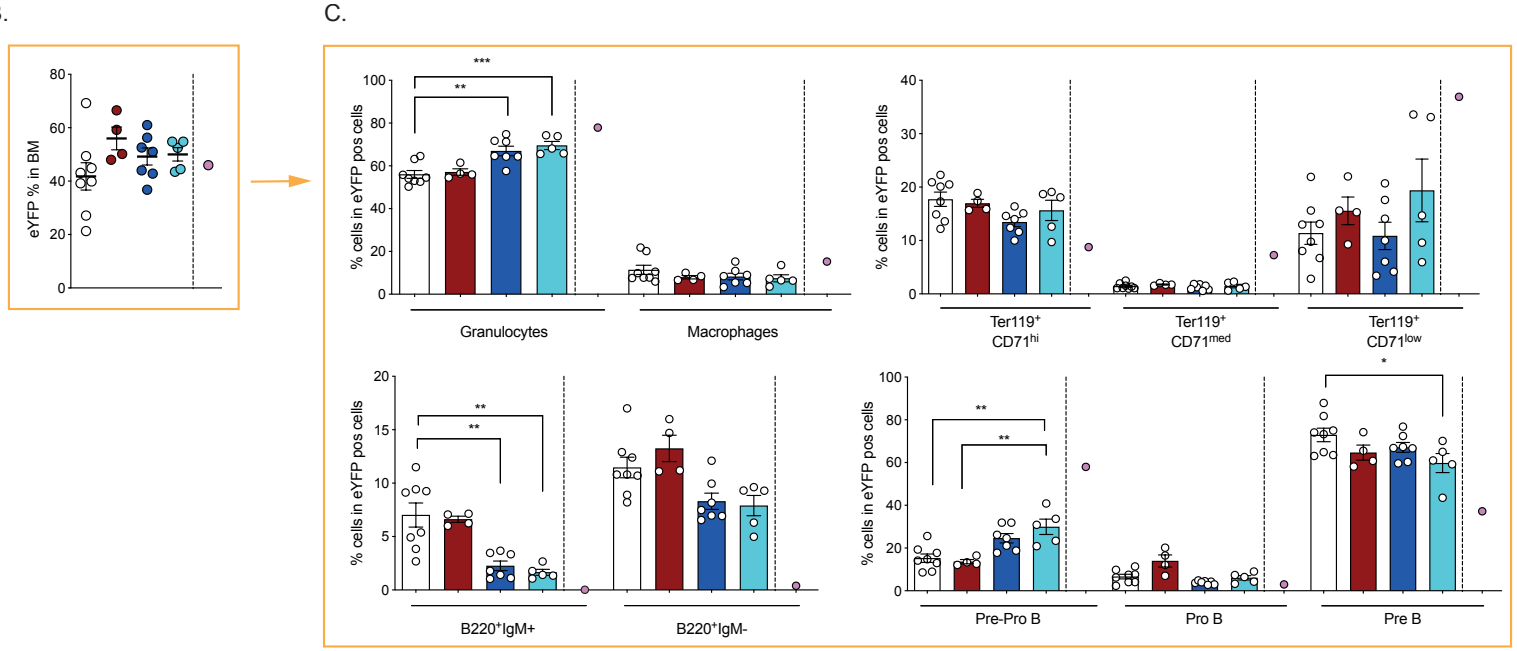


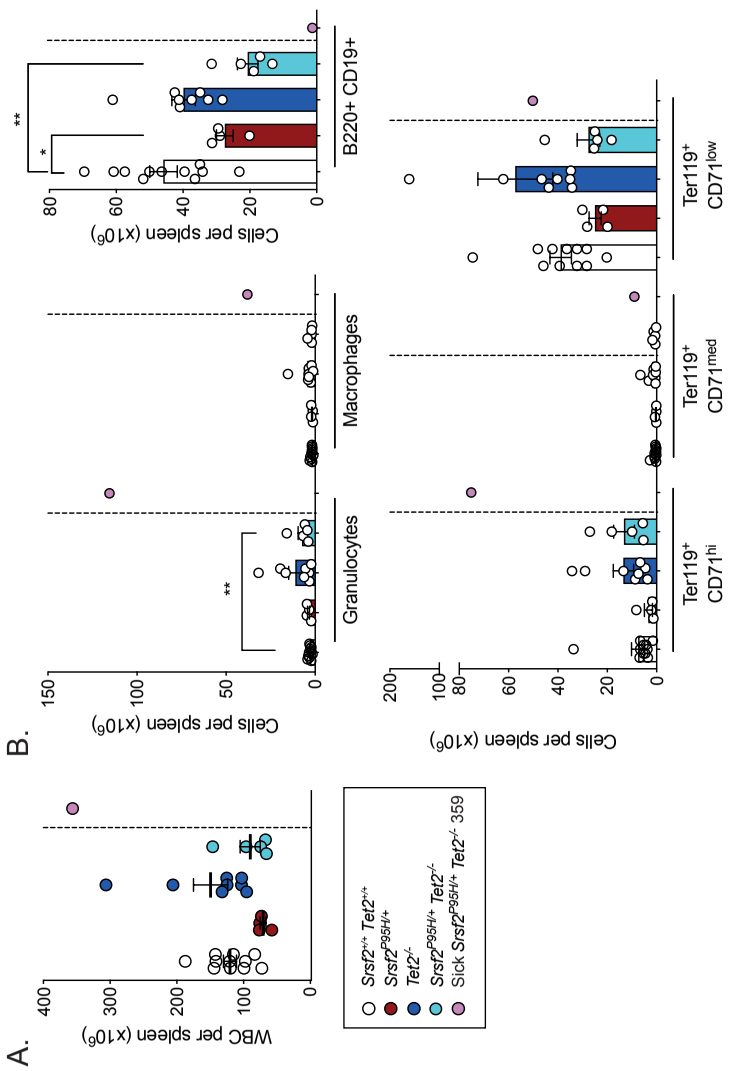
Figure 6.19: The myeloid bias and lymphoid suppression is evident in the eYFP positive fraction of peripheral blood and bone marrow in #359

A. The eYFP percentage of peripheral blood and the lineage distribution of eYFP positive peripheral blood leukocytes.

B. The eYFP percentage of bone marrow.

C. The proportion of myeloid, erythroid and lymphoid compartments in the eYFP positive fraction of bone marrow.

IgM, immunoglobulin M. n=4-11 for $Srsf2^{+/+}$ $Tet2^{+/+}$, $Srsf2^{P95H/+}$,$Tet2^{-/-}$ and $Srsf2^{P95H/+}$ $Tet2^{-/-}$ genotypes. Presented as mean ± standard error of mean. One-way ANOVA. *P<.05, **P<.01, ***P<.001.


A.
WBC per spleen (x10⁶)
0
100
200
300
400
Srsf2+/+ Tet2+/+
Srsf2P95H/+
Tet2-/-
Srsf2P95H/+ Tet2-/-
Sick Srsf2P95H/+ Tet2-/- 359
B.
Cells per spleen (x10⁶)
0
50
100
150
**
Granulocytes
Macrophages
Cells per spleen (x10⁶)
0
20
40
60
80
*
**
B220+ CD19+
Cells per spleen (x10⁶)
0
20
40
60
80
100
200
Ter119+ CD71hi
Ter119+ CD71med
Ter119+ CD71low

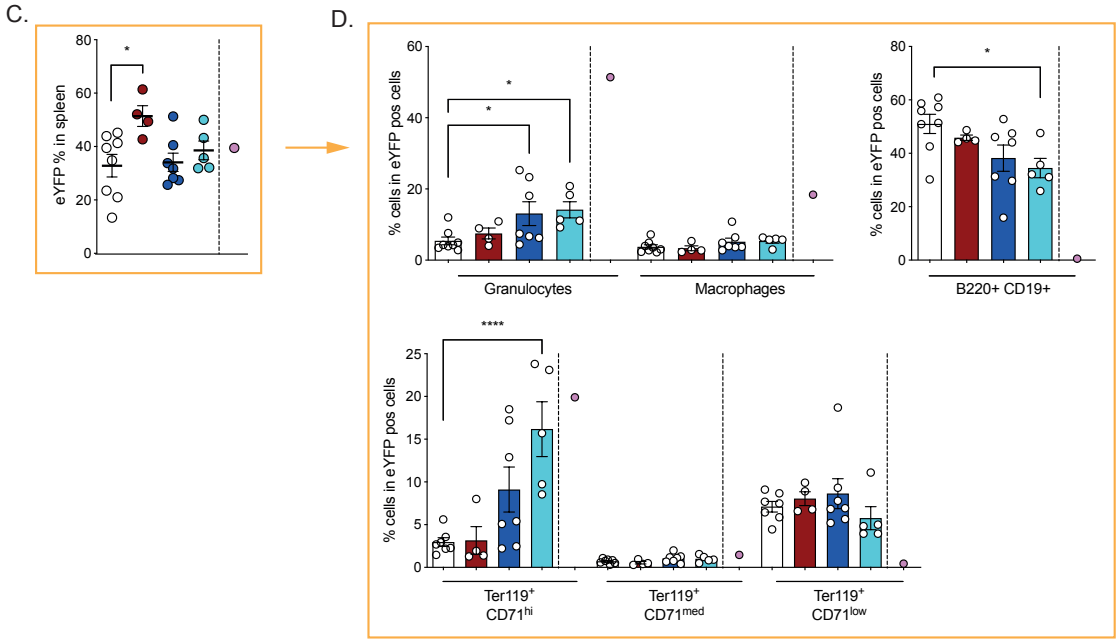


Figure 6.20: Sick mouse #359 has splenomegaly with significant myeloid expansion, erythroid compensation and near absence of B cells

A. Spleen cellularity of #359 compared to other genotypes after 52 weeks of Cre activation.

B. The number of myeloid lineage, B cell and erythroid lineage cells in the spleen.

C. The eYFP postitive percentage in the spleen.

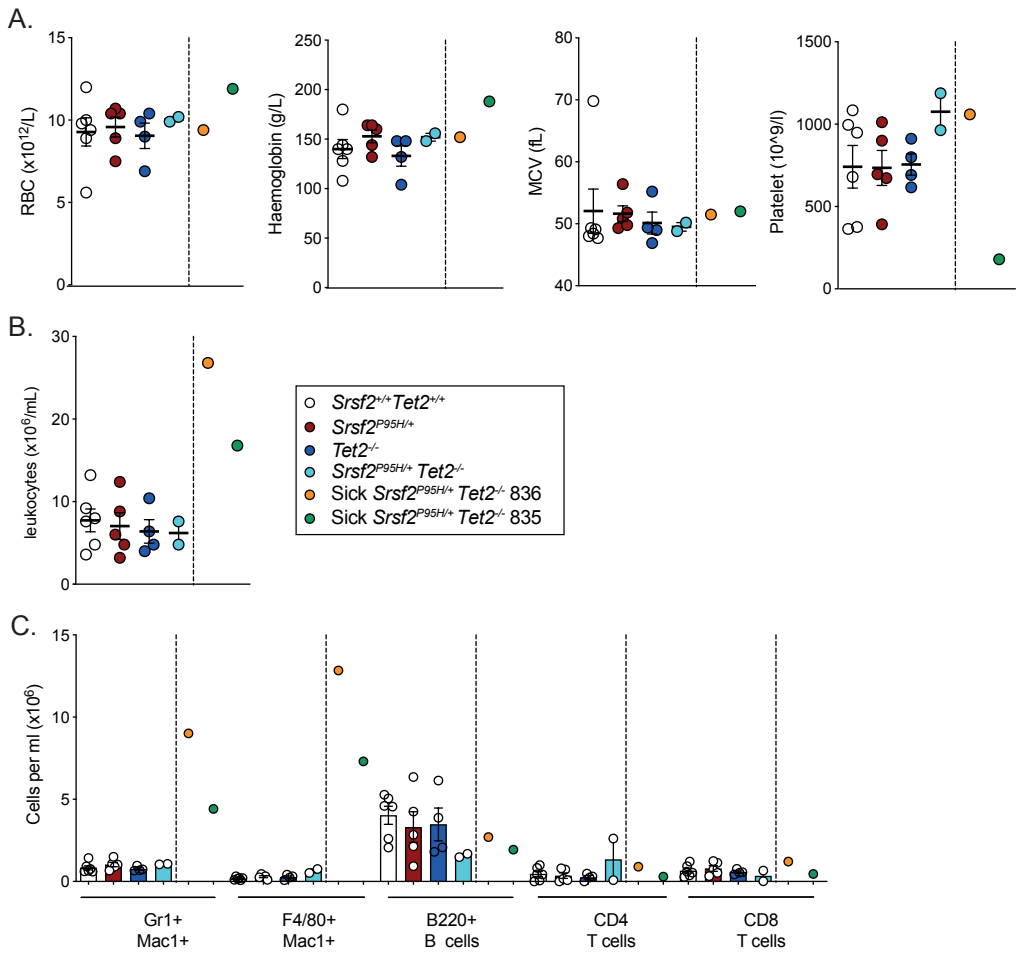
D. The percentage of myeloid lineage, B cell and erythroid lineage cells within the eYFP positive fraction in the spleen.

n=4-11 for *Srsf2*$^{+/+}$ *Tet2*$^{+/+}$, *Srsf2*$^{P95H/+}$,*Tet2*$^{-/-}$ and *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ genotypes. Presented as mean ± standard error of mean. One-way ANOVA. *P<.05, **P<.01, ****P<.0001.

The two sick mice identified from non-competitive transplant experiments (#835 and #836) displayed a more proliferative phenotype in the peripheral blood than #359 which was identified from cohorts with native haematopoiesis. The red blood cell parameters, including red blood cell numbers, haemoglobin and MCV were largely comparable to the controls (Figure 6.21A). The platelet count for #836 (12 weeks post-transplant) remained normal while #835 (52 weeks post-transplant) displayed thrombocytopenia (Figure 6.21A). The leukocyte counts were dramatically increased, with granulocytes and monocytes being highly elevated in both #835 and #836 (Figure 6.21B-E). In the bone marrow, both #835 and #836 had myeloid bias, and #836 also exhibited monocytosis and B cell suppression (Figure 6.22A-B). In contrast to #836, recipient #835 had a more drastic increase in granulocytes and monocytosis with a loss of mature B cells in the spleen (Figure 6.22C-D). The same phenotype was seen within the eYFP positive population for both moribund mice (Figure 6.23). I have included FACS plots of granulocytes, monocytes and B cells in the bone marrow and spleen of the three sick mice within the total cell population and within the eYFP positive fraction in Figure 6.24 and Figure 6.25. For granulocytes, I labelled both immature granulocytes ($Gr1^{med}CD11b^{+}$) and mature granulocytes ($Gr1^{hi}CD11b^{+}$) in the bone marrow and spleen plots. For monocytes, I labelled both immature monocytes ($CD11b^{med}F4/80^{+}$) and mature monocytes ($CD11b^{hi}F4/80^{+}$) in the bone marrow, and the whole monocyte ($CD11b^{+}F4/80^{+}$) population in the spleen. The B cells are labelled as $CD19^{+}B220^{+}$ in the spleen.

The collective features of these moribund mice are consistent with a diagnosis of chronic myelomonocytic leukaemia (CMML). This diagnosis is supported by the morphological findings in the bone marrow in #359 and marked increase of PB monocytes in #835 and #836. None of these

phenotypes are present in the respective single mutant haematopoiesis. This observation validated my hypothesis that *Srsf2*$^{P95H/+}$*Tet2*$^{-/-}$ mutation can lead to CMML *in vivo*. Given the small number of cases that have been identified at present, the transition to a more overt CMML phenotype most likely involves the acquisition of additional mutations. This would require further analysis of the available samples and assessment of a greater number of animals in future.



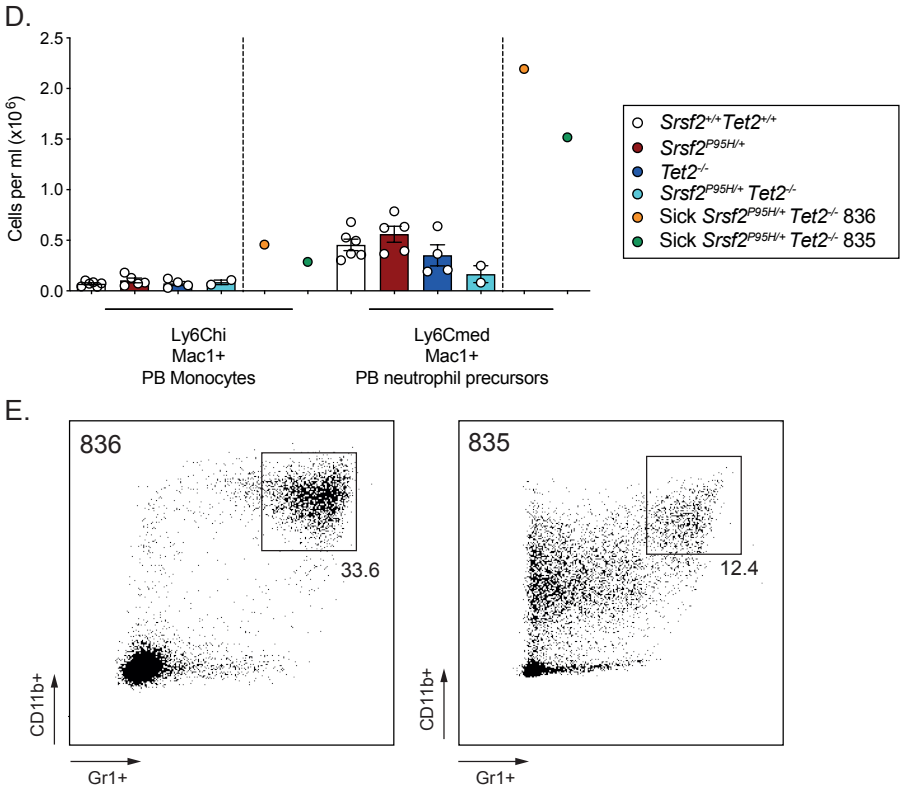
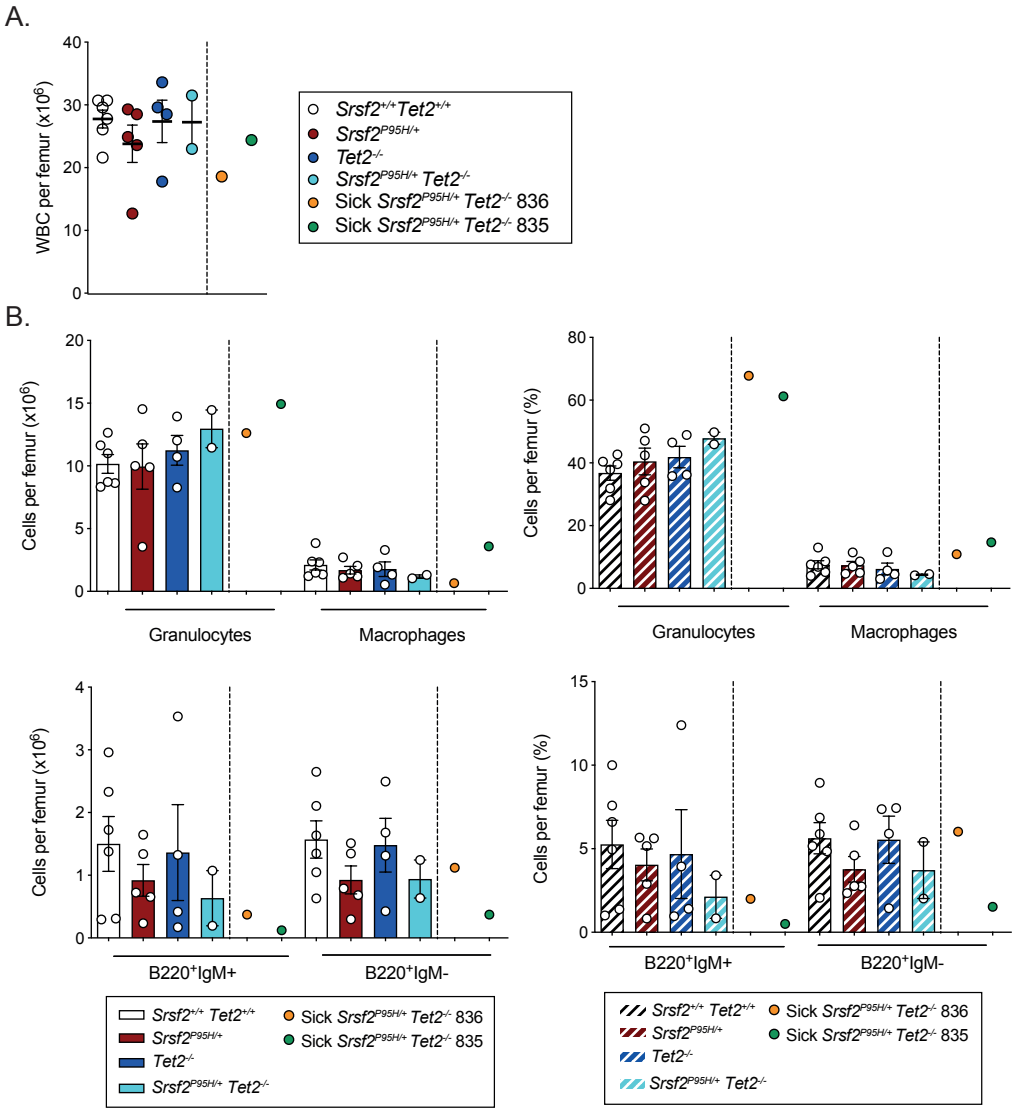


Figure 6.21: *Srsf2*$^{P95H/+}$*Tet2*$^{-/-}$ non-competitive transplant recipients #835 and #836 developed leukocytosis, with granulocytosis and monocytosis in the peripheral blood

A. The PB indices for two sick *Srsf2*$^{P95H/+}$*Tet2*$^{-/-}$ recipients compared to other genotypes from 52 weeks post transplantation.
B. PB leukocyte counts.
C. The PB lineage distribution.
D. The number of monocytes (Ly6C^{hi}Mac1^{+}) and neutrophil precursors (Ly6C^{med}Mac1^{+})in the PB.
E. The FACS plots of granulocytes in the peripheral blood.
n=2-6 for *Srsf2*$^{+/+}$ *Tet2*$^{+/+}$, *Srsf2*$^{P95H/+}$,*Tet2*$^{-/-}$ and *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ genotypes. Presented as mean $\pm$ standard error of mean.


A.
WBC per femur (x10^6)
Srsf2^+/+ Tet2^+/+
Srsf2^P95H/+
Tet2^-/-
Srsf2^P95H/+ Tet2^-/-
Sick Srsf2^P95H/+ Tet2^-/- 836
Sick Srsf2^P95H/+ Tet2^-/- 835
B.
Cells per femur (x10^6)
Cells per femur (%)
Granulocytes
Macrophages
B220+IgM+
B220+IgM-

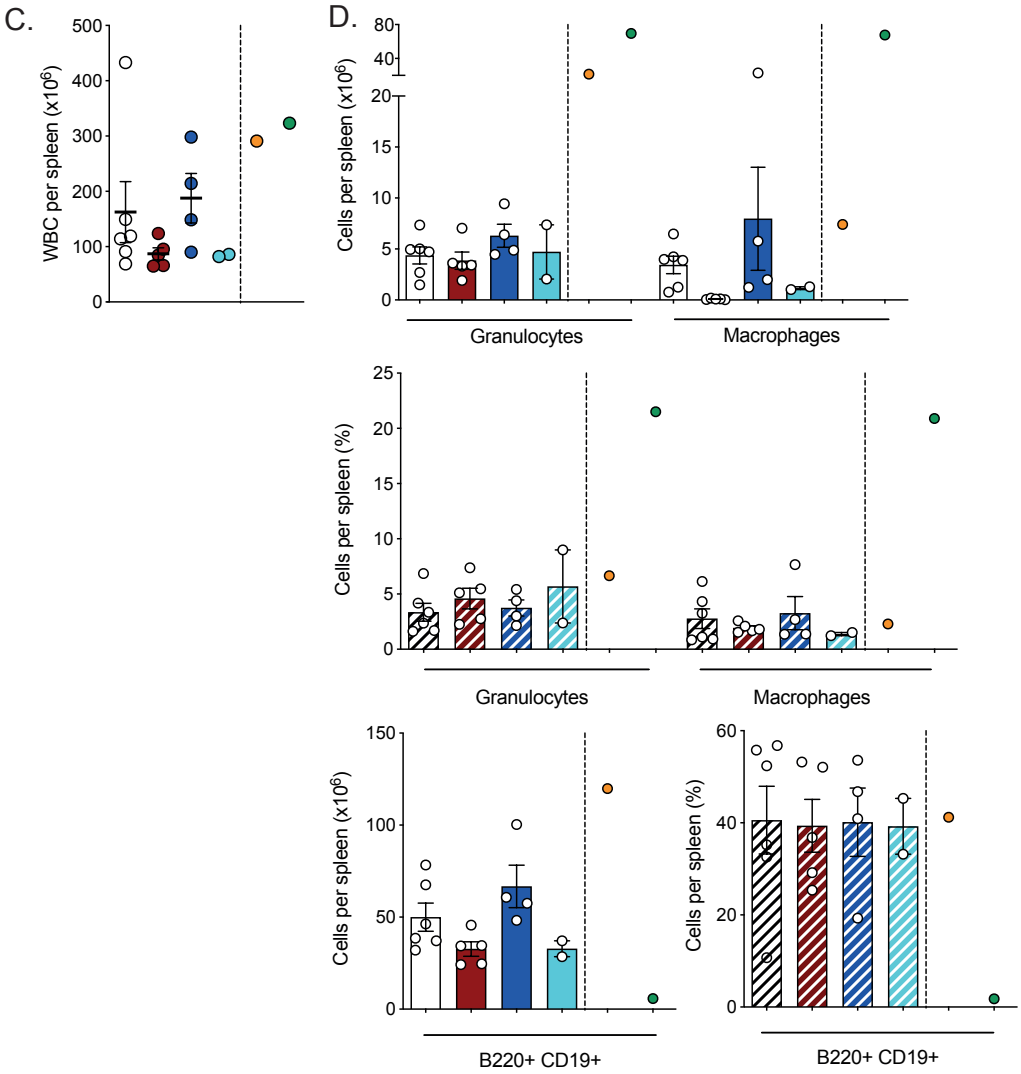


Figure 6.22: *Srsf2*$^{P95H/+}$*Tet2*$^{-/-}$ leads to hypercellularity and myeloid bias in the bone marrow and spleen of sick non-competitive transplant recipients

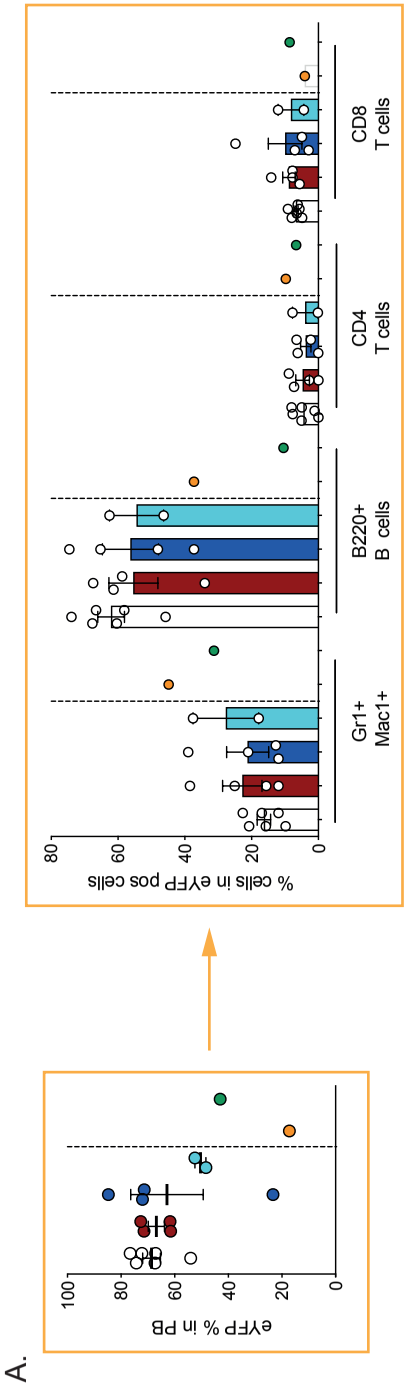
A. Bone marrow cellularity of #835 and #836, compared to non-competitive transplant recipients at 52 weeks post transplantation.

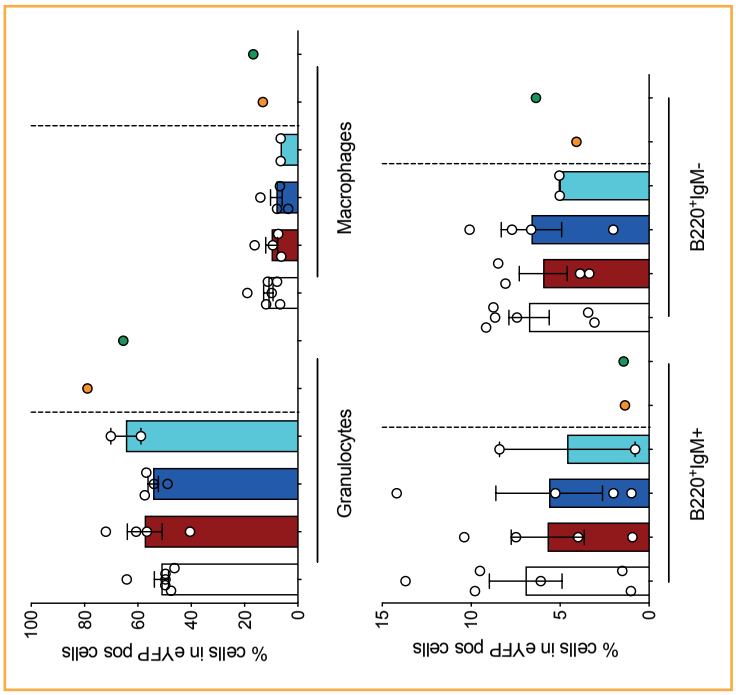
B. The number and percentage of myeloid and B lymphoid cells in the bone marrow.

C. Spleen cellularity.

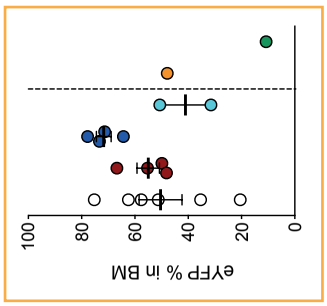
D. The number and percentage of myeloid and B cells in the spleen.

n=2-6 for *Srsf2*$^{+/+}$ *Tet2*$^{+/+}$, *Srsf2*$^{P95H/+}$,*Tet2*$^{-/-}$ and *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ genotypes. Presented as mean ± standard error of mean.


A.
eYFP % in PB
0
20
40
60
80
100
% cells in eYFP pos cells
0
20
40
60
80
Gr1+
Mac1+
B220+
B cells
CD4
T cells
CD8
T cells


% cells in eYFP pos cells
Granulocytes
Macrophages
% cells in eYFP pos cells
B220+IgM+
B220+IgM-

B.


eYFP % in BM

C.

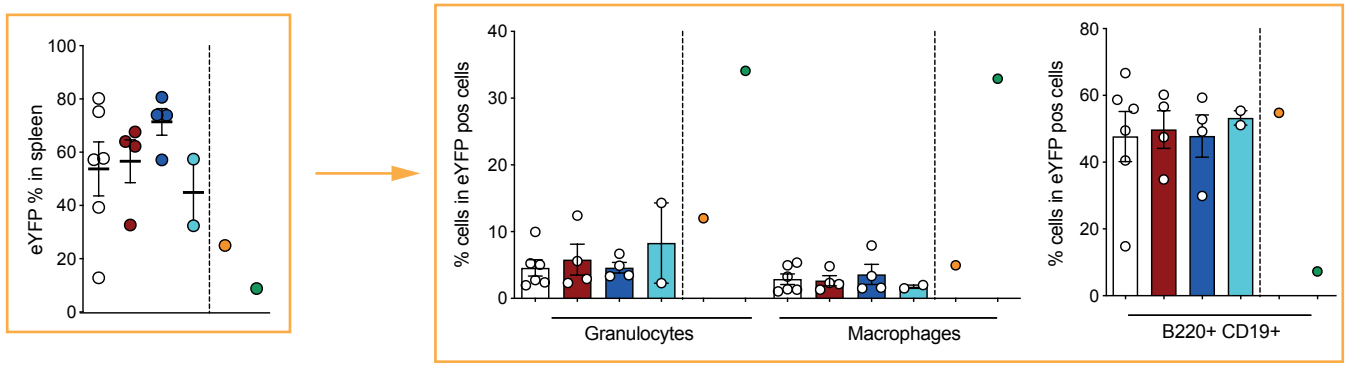
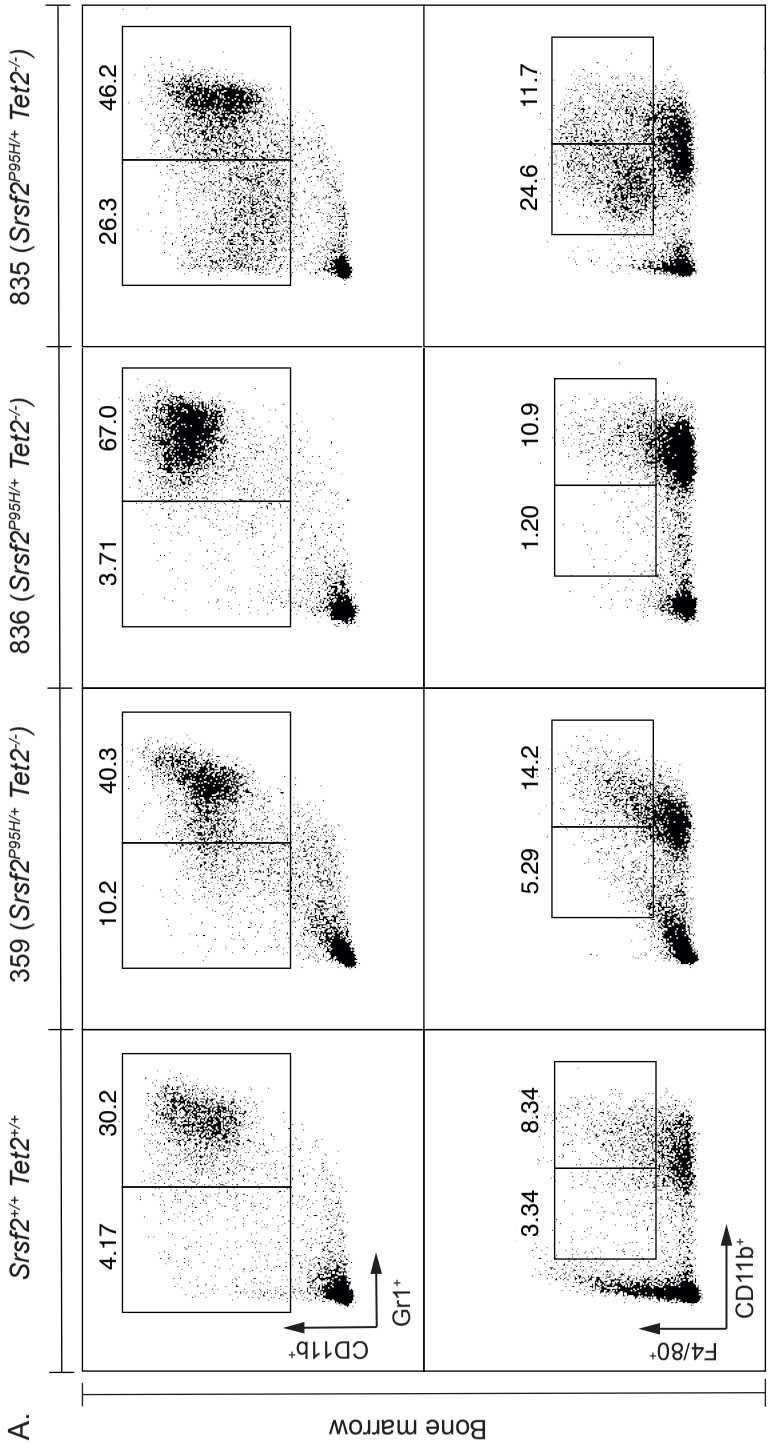
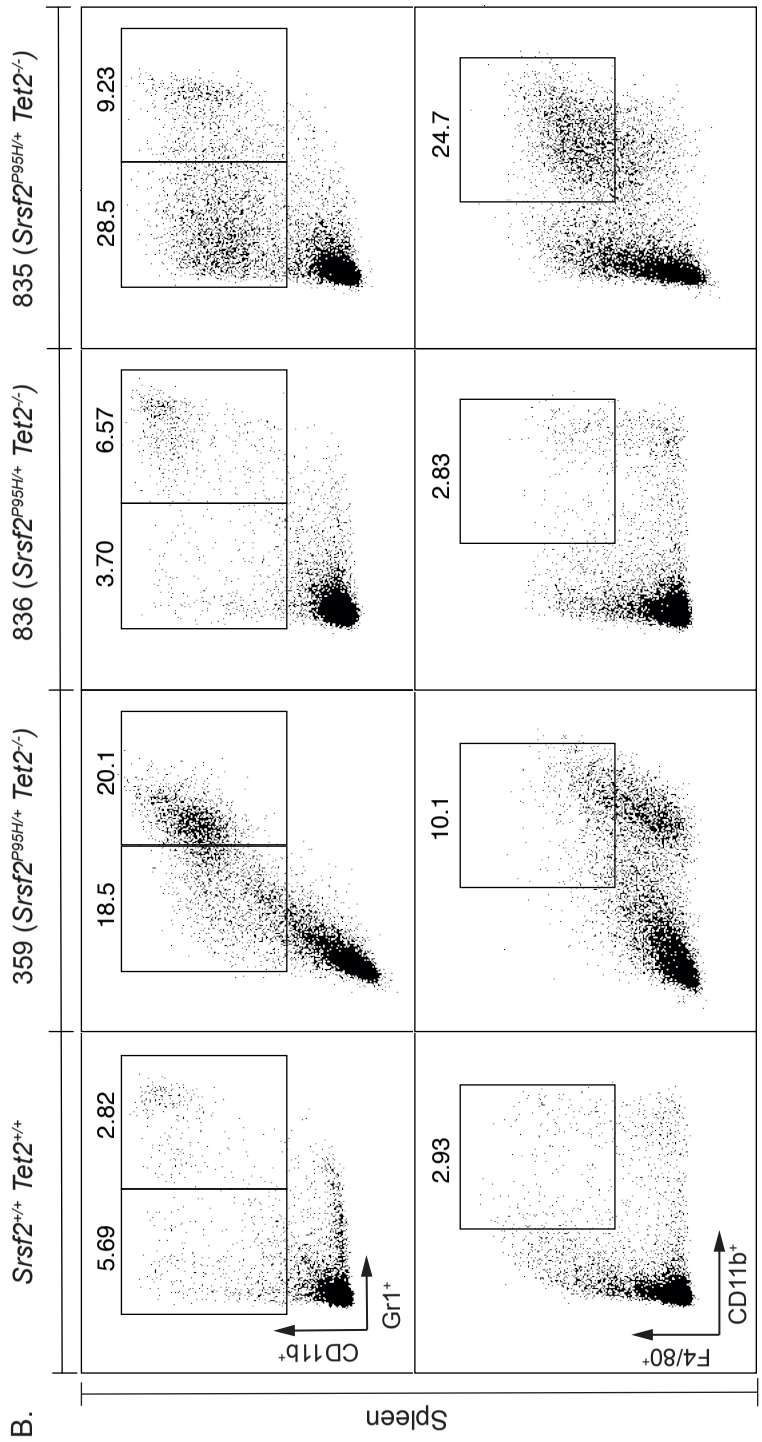


Figure 6.23: Cells expressing *Srsf2*$^{P95H/+}$*Tet2*$^{-/-}$ mutation drive the myeloid bias and B cell suppression in the non-competitive transplant recipients #835 and #836

A. The eYFP percentage in peripheral blood and lineage distribution of leukocytes within the eYFP positive fraction.
B. The eYFP percentage in bone marrow and percentage of myeloid and B lymphoid lineage within the eYFP positive fraction.
C. The eYFP percentage in bone marrow and percentage of myeloid and B cells within the eYFP positive fraction.
n=2-6 for *Srsf2*$^{+/+}$ *Tet2*$^{+/+}$, *Srsf2*$^{P95H/+}$,*Tet2*$^{-/-}$ and *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ genotypes. Presented as mean ± standard error of mean.


A.
Bone marrow
Srsf2+/+ Tet2+/+
359 (Srsf2P95H/+ Tet2-/-)
836 (Srsf2P95H/+ Tet2-/-)
835 (Srsf2P95H/+ Tet2-/-)
4.17
30.2
10.2
40.3
3.71
67.0
26.3
46.2
3.34
8.34
5.29
14.2
1.20
10.9
24.6
11.7
CD11b+
Gr1+
F4/80+
CD11b+


B.
Spleen
Srsf2+/+ Tet2+/+
359 (Srsf2P95H/+ Tet2-/-)
836 (Srsf2P95H/+ Tet2-/-)
835 (Srsf2P95H/+ Tet2-/-)
5.69
2.82
18.5
20.1
3.70
6.57
28.5
9.23
2.93
10.1
2.83
24.7
CD11b+
Gr1+
F4/80+
CD11b+

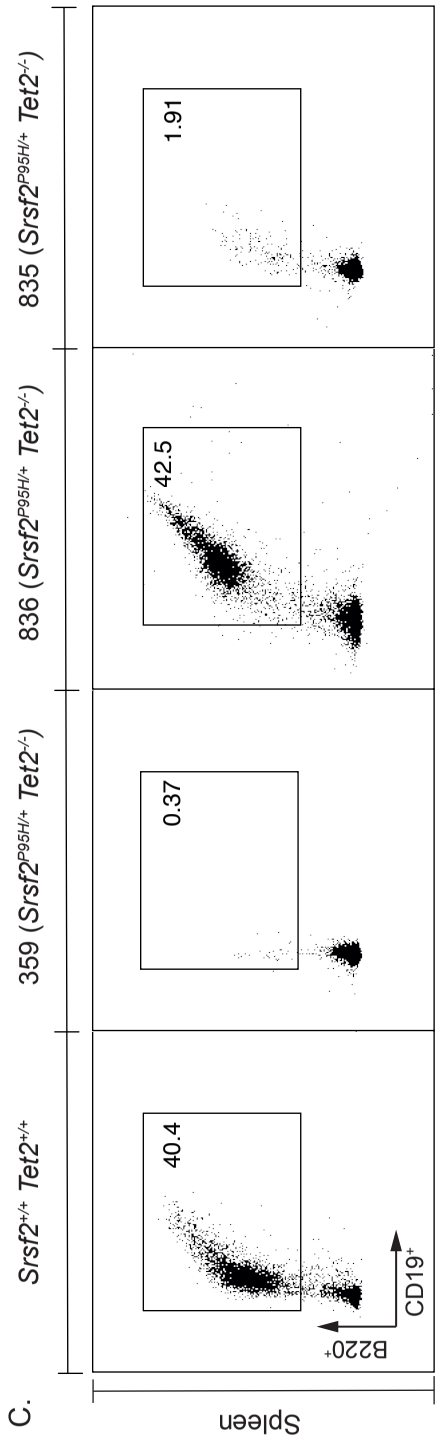
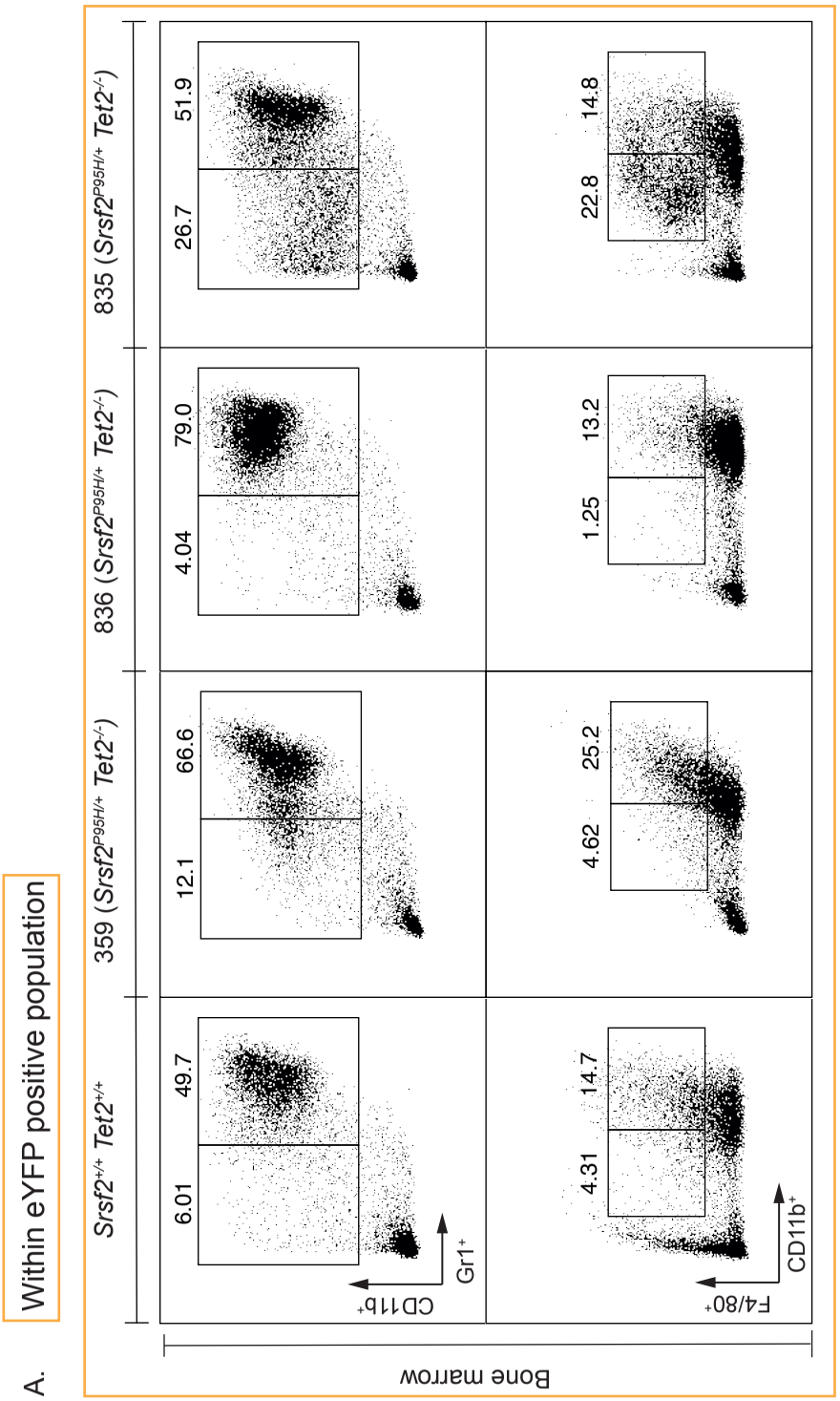
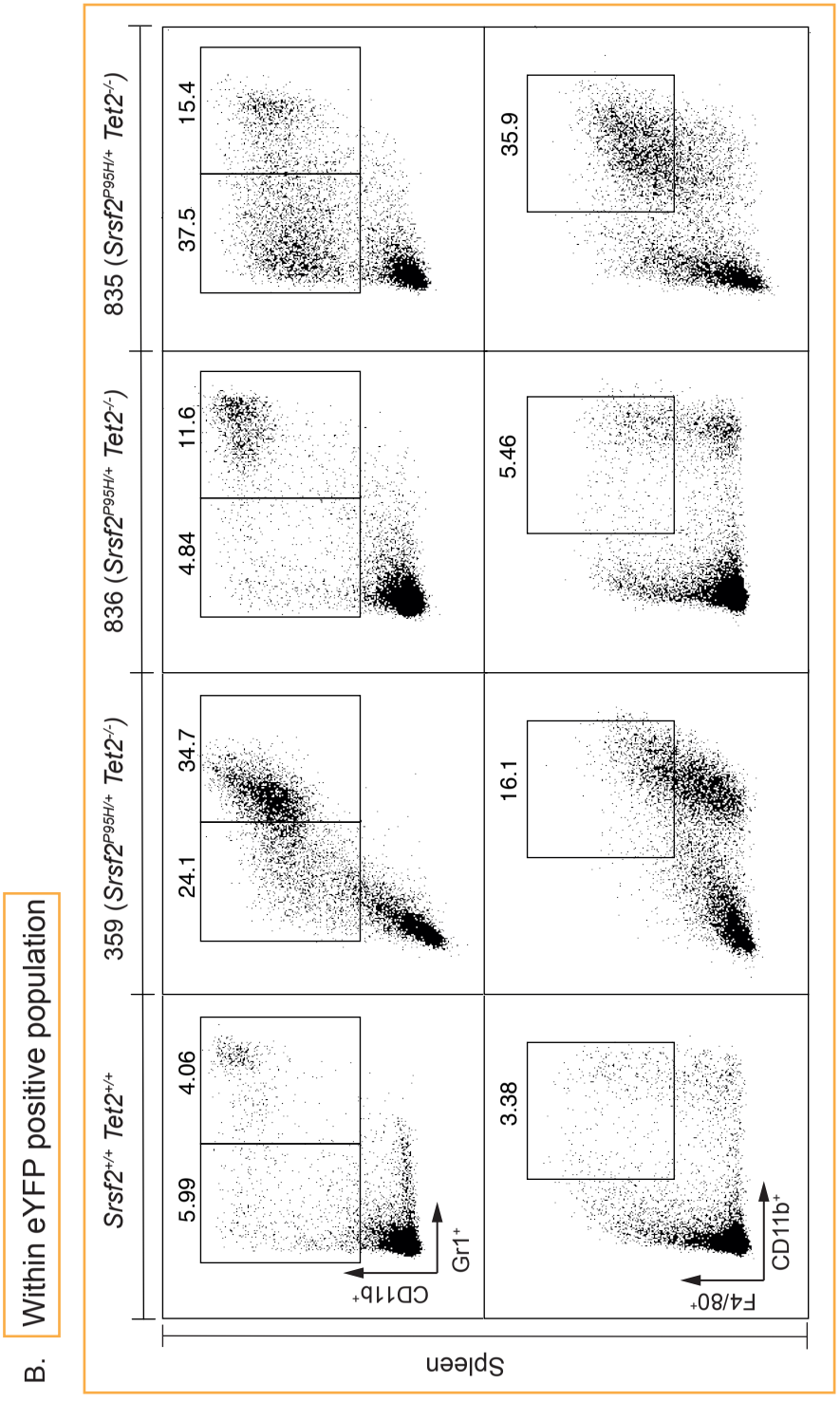


Figure 6.24: FACS plots from the moribund *Srsf2*$^{P95H/+}$*Tet2*$^{-/-}$ mice showing myeloid bias, monocytosis and B cell suppression in the bone marrow (A) and spleen (B-C)


A.
Within eYFP positive population
Srsf2+/+ Tet2+/+
359 (Srsf2P95H/+ Tet2-/-)
836 (Srsf2P95H/+ Tet2-/-)
835 (Srsf2P95H/+ Tet2-/-)
Bone marrow
CD11b+
Gr1+
F4/80+
CD11b+
6.01
49.7
12.1
66.6
4.04
79.0
26.7
51.9
4.31
14.7
4.62
25.2
1.25
13.2
22.8
14.8


B.
Within eYFP positive population
Srsf2+/+ Tet2+/+
359 (Srsf2P95H/+ Tet2-/-)
836 (Srsf2P95H/+ Tet2-/-)
835 (Srsf2P95H/+ Tet2-/-)
Spleen
CD11b+
Gr1+
F4/80+
CD11b+
5.99
4.06
24.1
34.7
4.84
11.6
37.5
15.4
3.38
16.1
5.46
35.9

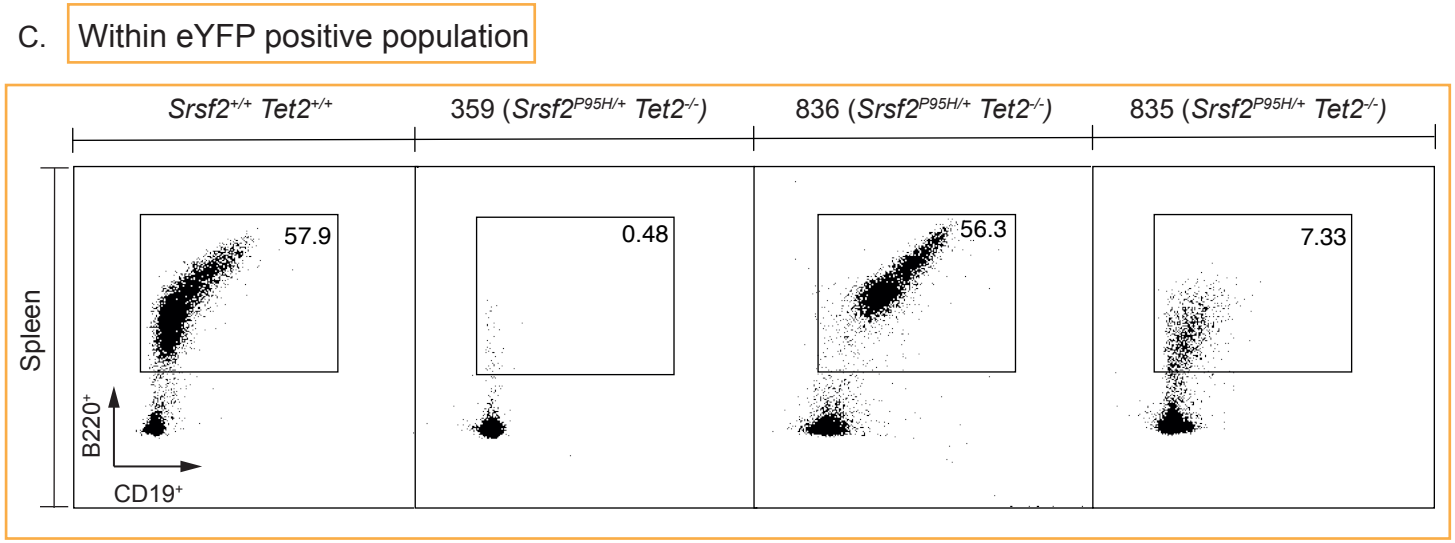


Figure 6.25: FACS plots of the eYFP positive population from the sick *Srsf2*$^{P95H/+}$*Tet2*$^{-/-}$ mice showing myeloid bias, monocytosis and B cell suppression in the bone marrow (A) and spleen (B-C)

6.2.5 *Srsf2*$^{P95H/+}$ *Cbl*$^{-/-}$ mutation leads to acceleration of myeloid proliferation

I had intended to utilise a *c-Cbl* ring finger mutation (*c-Cbl*$^{A/+}$) combined with the null allele (*c-Cbl*$^{-/-}$) as this genetic combination more closely recapitulates the mutations seen in patients. However, due to sustained and unexpected breeding difficulties, I have not been able to breed and obtain any h*Scl*-CreER^{T} *R26*eYFP *Srsf2*$^{P95H/+}$ *Cbl*$^{A/-}$ mice over three years of breeding. I chose not to assess foetal liver cells or use these for repopulation as MDS/MPN is a disease of adults and I wanted to use models of native haematopoiesis preferentially. As a result, I decided to assess if co-operativity would be apparent between *Srsf2*$^{P95H/+}$ and *Cbl*$^{-/-}$ mutation, even though the *Cbl*$^{-/-}$ murine model develops a mild myeloproliferative phenotype (Rathinam et al., 2010). I have analysed a small cohort of hScl-CreER *R26*eYFP *Srsf2*$^{P95H/+}$ *Cbl*$^{-/-}$ (n=2) along with *Srsf2*$^{P95H/+}$ *Cbl*$^{+/-}$ (n=3) that were fed tamoxifen food when the mice were 8-10 weeks old. Four weeks later, mice were taken off tamoxifen food and monitored regularly. The Cre-positive *Cbl*$^{+/-}$ and *Cbl*$^{-/-}$ genotypes were included in the cohort and treated with tamoxifen as single mutant control. This is to rule out any potential side-effects from Cre recombinase, despite the *Cbl* mutation being a germ-line mutation. During the course of monitoring, one *Srsf2*$^{P95H/+}$ *Cbl*$^{+/-}$ (referred to as 'compound het') and one *Srsf2*$^{P95H/+}$ *Cbl*$^{-/-}$ mutant (referred to as 'compound null') mouse were analysed at 20- and 10-weeks post mutation activation respectively due to skin wounds. These two mice are marked by ☒ in the figures. The rest of the cohorts were analysed at 52 weeks post mutation activation (outlined in Figure 6.26A).

Due to the low number of mice analysed for most genotypes, the data presented here are a preliminary analysis of the *Srsf2*$^{P95H/+}$ *Cbl* compound

mutation. I have also not included statistical analysis and comparison of the data due to the low number of individuals in each cohort and have provided a limited interpretation of the trends in the data rather than a more definitive conclusion. In the peripheral blood, there was a trend for a lower number of red blood cells and increased MCV in the compound null mutants. The platelet number trended to be increased, while the haemoglobin remained stable (Figure 6.26B). The leukocyte count and lineage distribution were largely unchanged (Figure 6.26C-D). In the bone marrow, there was no alteration in total cellularity (Figure 6.27A). The myeloid, erythroid and T lymphoid lineages were normal (Figure 6.27B-C). In contrast, the immature B cells ($B220^{+}IgM^{-}$) and Pre B cells were reduced in the compound het and compound null mutants (Figure 6.27D). The stem and progenitor populations did not show any obvious differences (Figure 6.27E). In the spleen, there was a decline of cellularity in the compound het and compound null mutants compared to single *Cbl* mutant controls (Figure 6.28A). The discrepancies were probably due to a reduction in B cells in the compound mutants, while myeloid, erythroid and T cells were not affected (Figure 6.28B).

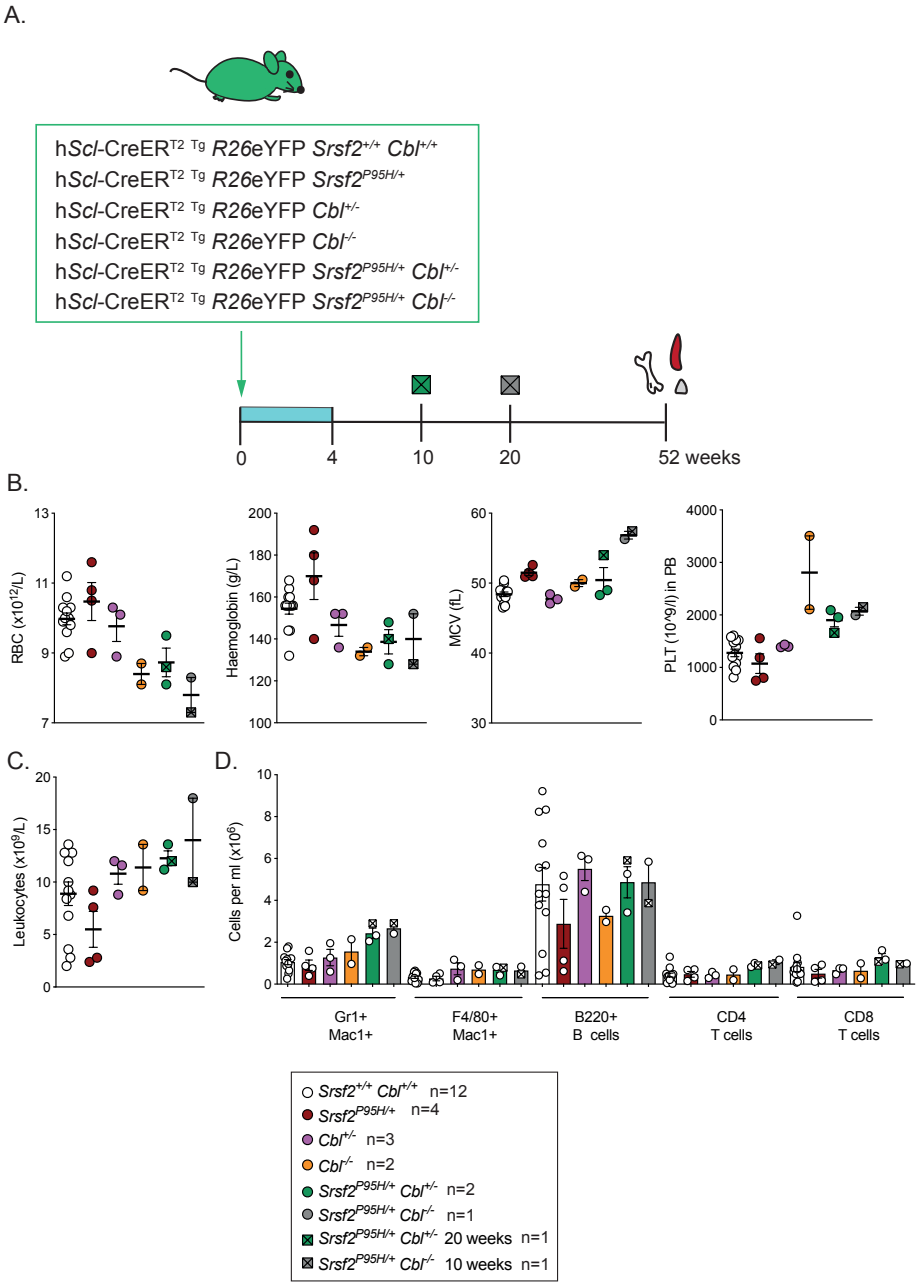
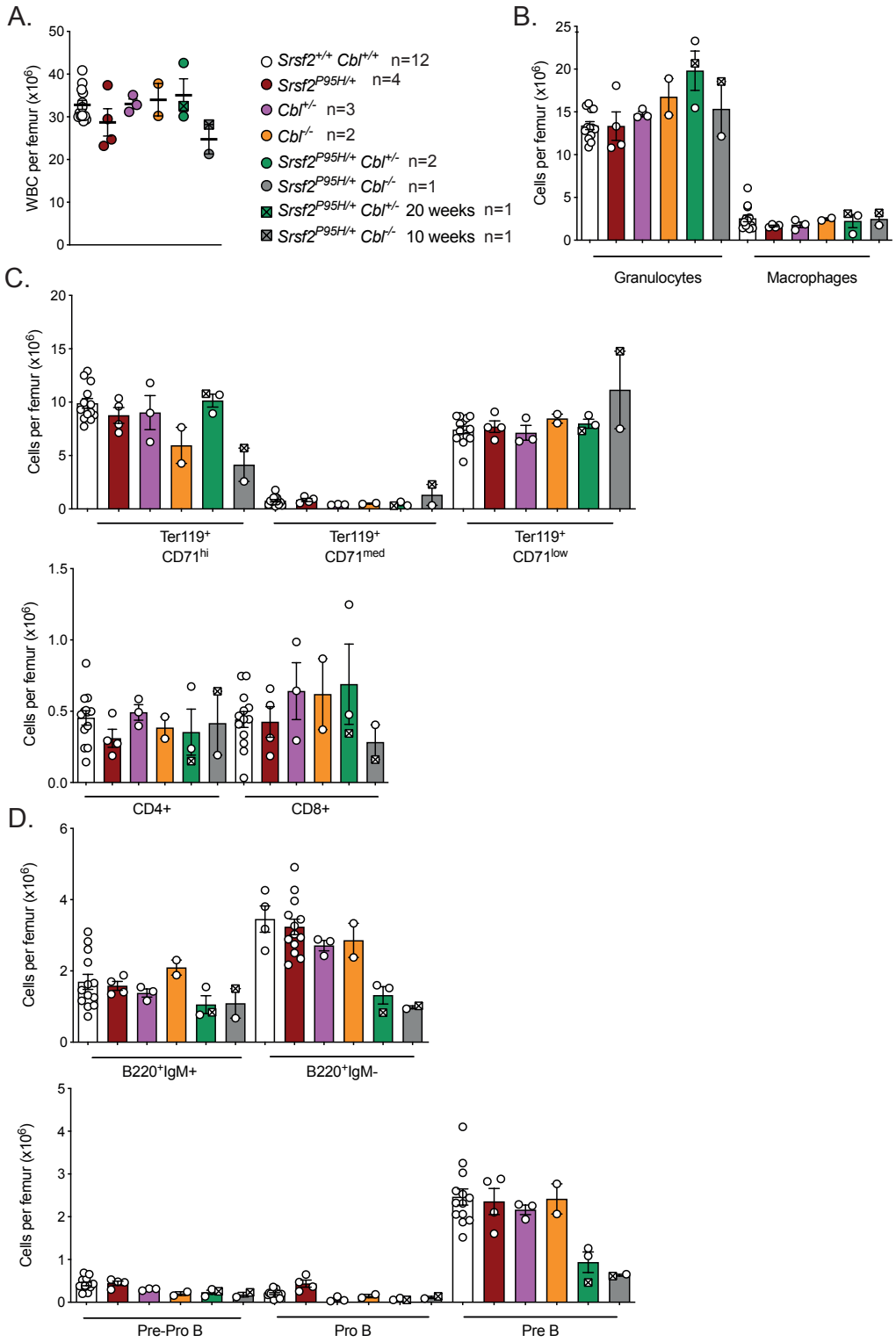


Figure 6.26: Peripheral blood analysis of Srsf2$^{P95H/+}$ *Cbl*$^{-/-}$ compound mutant mice
A. The schematic of *Srsf2 Cbl* compound mutant cohort analysis.
B. The PB indices.
C. The PB leukocyte counts.
D. The lineage distribution within leukocytes.
Presented as mean ± standard error of mean.


A.
WBC per femur (x10⁶)
Srsf2+/+ Cbl+/+ n=12
Srsf2P95H/+ n=4
Cbl+/- n=3
Cbl-/- n=2
Srsf2P95H/+ Cbl+/- n=2
Srsf2P95H/+ Cbl-/- n=1
Srsf2P95H/+ Cbl+/- 20 weeks n=1
Srsf2P95H/+ Cbl-/- 10 weeks n=1
B.
Cells per femur (x10⁶)
Granulocytes
Macrophages
C.
Cells per femur (x10⁶)
Ter119+ CD71hi
Ter119+ CD71med
Ter119+ CD71low
CD4+
CD8+
D.
Cells per femur (x10⁶)
B220+IgM+
B220+IgM-
Pre-Pro B
Pro B
Pre B

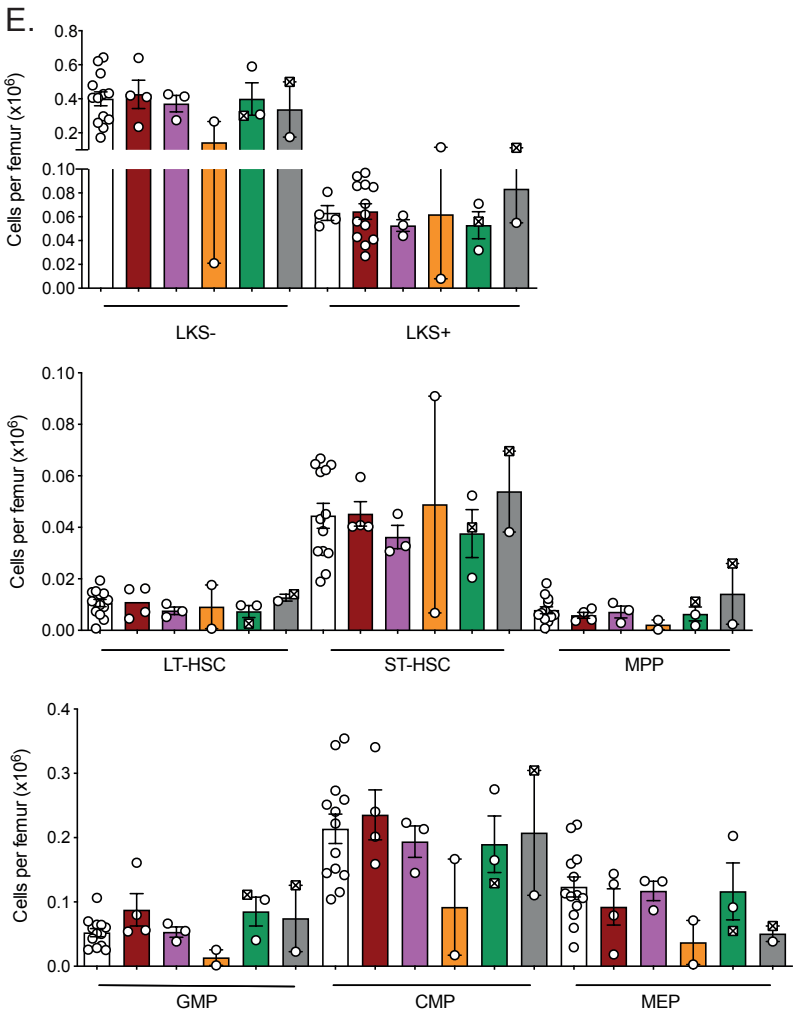


Figure 6.27: Bone marrow analysis of *Srsf2*$^{P95H/+}$ *Cbl*$^{-/-}$ compound mutant mice

A. The bone marrow cellularity.

B. The bone marrow granulocyte and macrophage counts.

C. The bone marrow erythrocyte and T cell lineage counts.

D. The bone marrow B lineage counts.

E. Number of LKS^{-}, LKS^{+}, phenotypic LT-HSC, ST-HSC, MPP in LKS^{+} population and myeloerythroid progenitors in the LKS^{+} population. IgM, immunoglobulin M; LKS^{+}, lin^{-}c-Kit^{+}Sca-1^{-}; LKS^{+}, lin^{-}c-Kit^{+-}Sca-1^{+}; LT-HSC, long-term haematopoietic stem cell; ST-HSC, short-term haematopoietic stem cell; MPP, multipotent progenitor; GMP, granuocyte-macrophage progenitors; CMP, common myeloid progenitors; MEP, myelo-erythroid progenitors. Presented as mean $\pm$ standard error of mean.

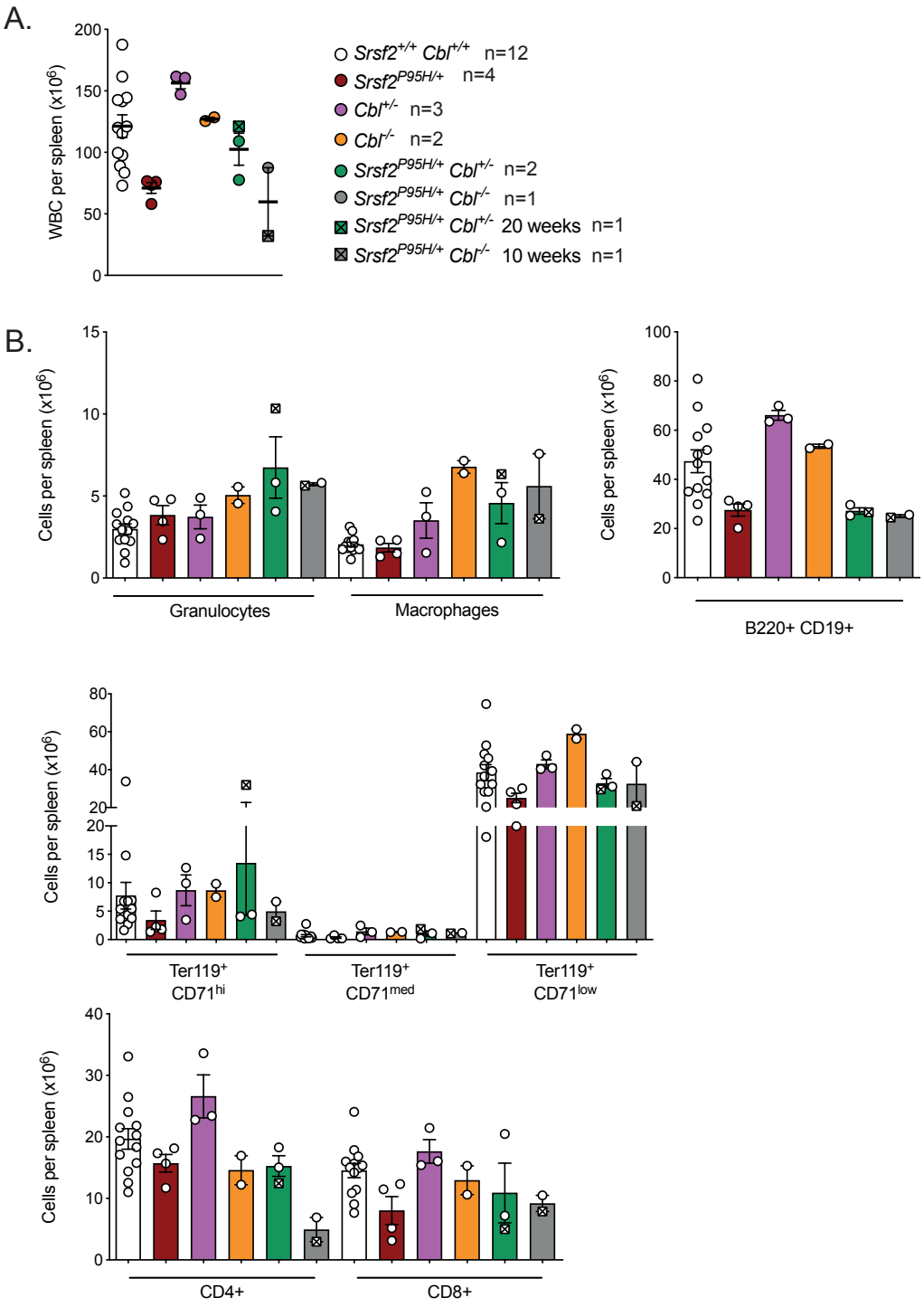


Figure 6.28: Spleen analysis of Srsf2$^{P95H/+}$ *Cbl*$^{-/-}$ compound mutant mice
A. The spleen cellularity.
B. The number of myeloid, lymphoid and erythroid lineage in the spleen.
Presented as mean ± standard error of mean.

I also analysed the compound het and compound null mutant that were collected at 20- and 10-weeks post mutation activation respectively. Compared to a wild-type or *Srsf2*$^{P95H/+}$ mutant at 20-weeks post mutation activation, FACS analysis indicated severe myeloid hyperplasia ($Gr1^+CD11b^+$) in both the bone marrow and peripheral blood of the compound het and compound null mutant mice. The myeloid biased phenotype was present in both overall and within the eYFP positive fraction (Figure 6.29). There was an elevation of the macrophage population ($CD11b^+F4/80^+$) in the bone marrow and peripheral blood (Figure 6.30).

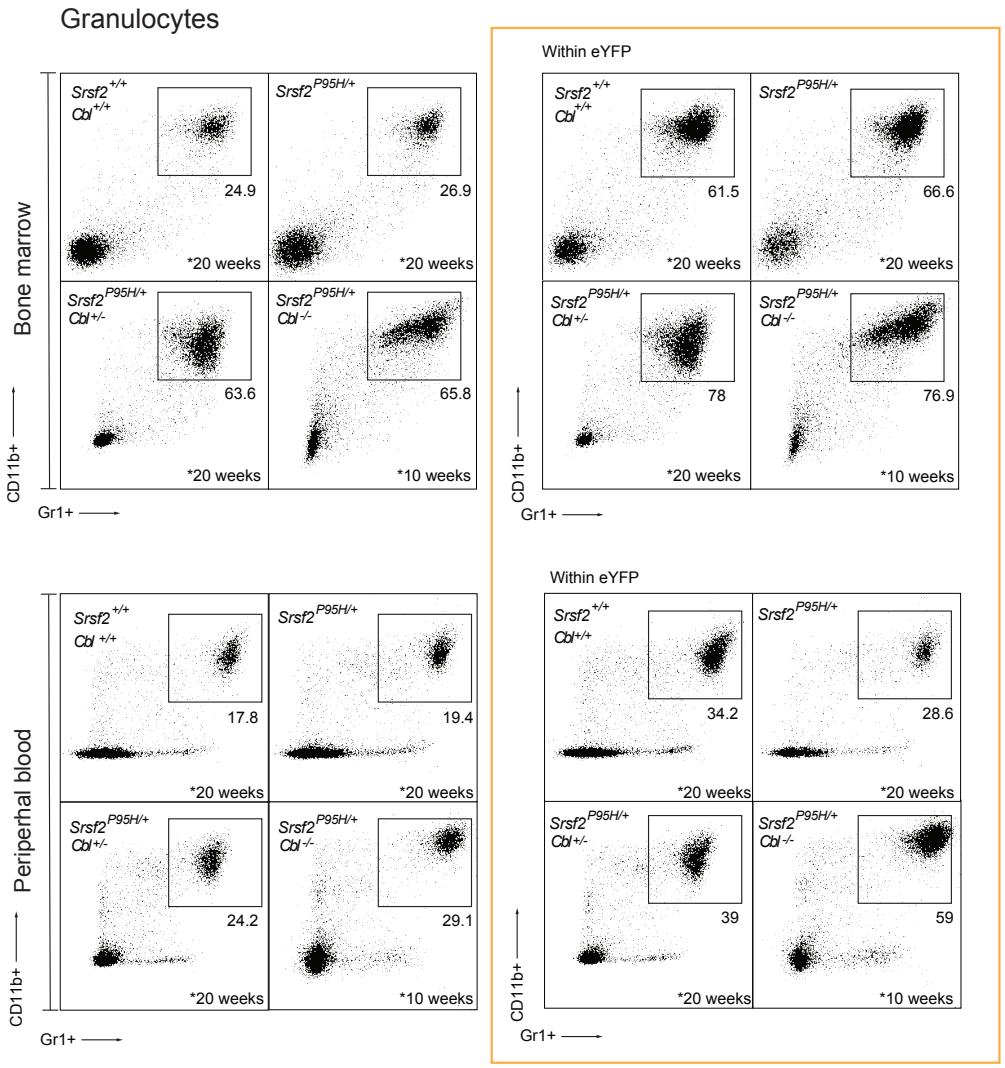


Figure 6.29: Myeloid-hyperplasia in both the bone marrow and peripheral blood of sick Srsf2$^{P95H/+}$ *Cbl*$^{-/-}$ mice, including within the eYFP positive fraction
Time labels indicate weeks post mutation activation.

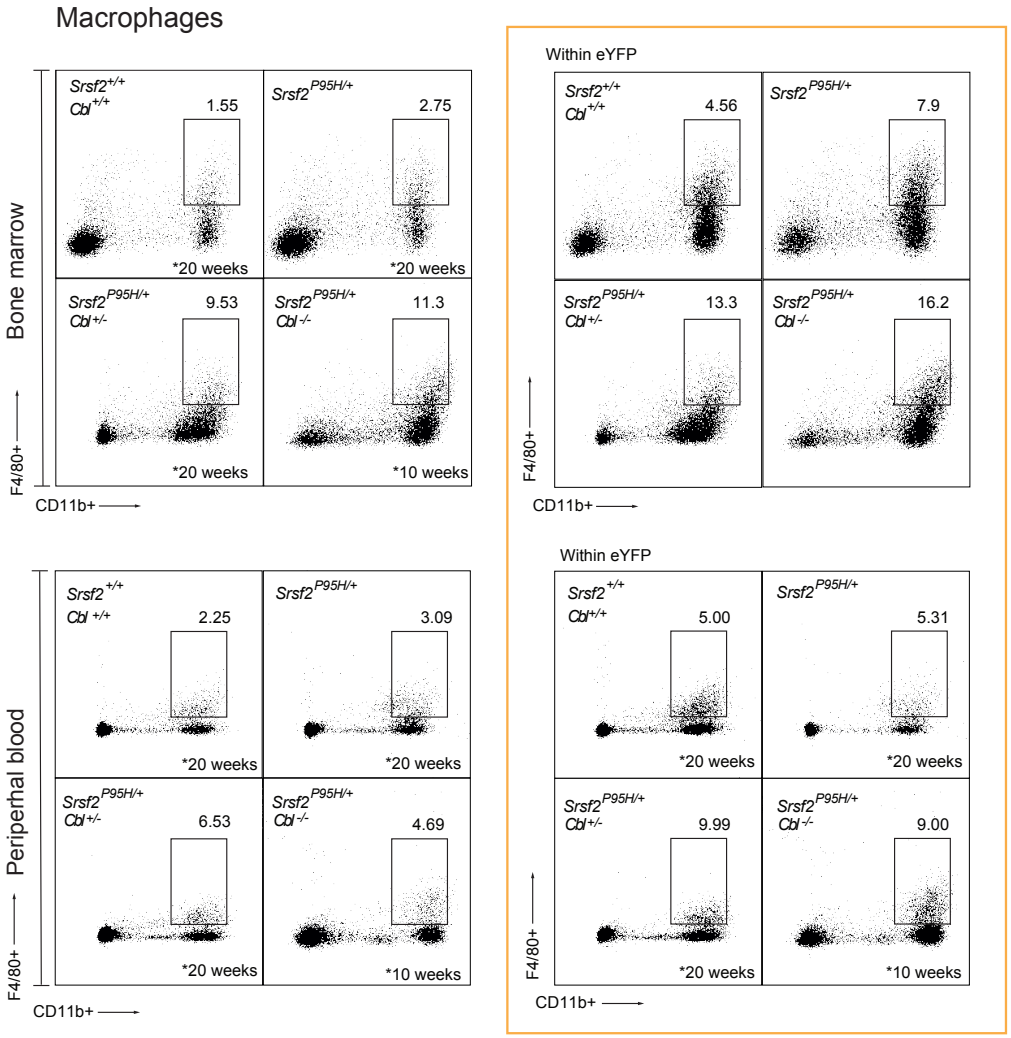


Figure 6.30: FACS plots showing a mild increase in the macrophage population in both bone marrow and peripheral blood of sick mice, including within the eYFP positive fraction. Time labels indicate weeks post tamoxifen cessation
Time labels indicate weeks post mutation activation.

6.3 Discussion

In this chapter, I have described the generation and characterisation of a *Srsf2/Tet2* compound mutant murine model. This genetic combination is profoundly enriched in MDS/MPN and CMML. Ageing of the compound mutants resulted in a more substantial myeloid bias and B cell suppression compared to either single *Srsf2* or *Tet2* mutation alone. Similar to the primary mice, recipients of *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ mutant bone marrow also developed severe myeloid bias and B cell suppression, suggesting that the lineage bias was a cell-autonomous behaviour. Compared to wild-type recipients, compound mutant recipients achieved stable long-term engraftment in a non-competitive setting. In contrast, the compound mutant bone marrow showed a disadvantage when competing against fresh wild-type bone marrow and resulted in significantly lower engraftment.

During the course of the experiments to date, I have observed development of CMML in one non-transplanted *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ mouse and in two recipients of *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ mutant bone marrow from non-competitive transplant experiments. While #836 succumbed to the disease at 12 weeks post transplantation, #359 and #835 had a latency of 52 weeks post mutation activation (#359) or 52 weeks post transplantation (#835), respectively. This long latency is similar to single *Srsf2*$^{P95H/+}$ mutants, which developed fatal MDS after at least 52 weeks post mutation activation (Smeets et al., 2018). There was no acceleration of disease initiation, despite the additional *Tet2*$^{-/-}$ mutation being activated at the same time as *Srsf2*$^{P95H/+}$. This is unexpected as combining two co-operating mutations leads to acceleration of disease in many experimental settings of cancer formation. In the case of *Srsf2/Tet2* mutations, the long latency, however, is not inconsistent with the analysis of CHIP and clonal haematopoiesis (Genovese et al., 2014; Jaiswal et al., 2014). The role of *Srsf2*$^{P95H/+}$ and

Tet2$^{-/-}$ as driver mutations is to lay the foundation for disease initiation, which did result in CMML in my experiments. For the disease to progress additional mutations have to be progressively acquired, and in the case of MDS/MPN the average number of mutations in patient samples is 3 (0-12) and in CMML it is 14 ± 5 (Haferlach et al., 2014; Merlevede et al., 2016). I speculate that additional mutations besides *Srsf2/Tet2* mutations are required to complete the transformation, such as *ASXL1* and RAS pathway (*CBL, NRAS, KRAS*) mutations. Exome sequencing of the sick compound mutant samples would be valuable to understand the additional mutations that have occurred in these animals to allow disease evolution and outline how these samples have evolved to enable disease development.

At the time of analysis, all three sick mice had splenomegaly with hypo- to normal cellularity of the bone marrow. Myeloid expansion was significant within the bone marrow and spleen while B cell suppression was only evident in the primary and transplant aged sick mice. While some phenotypes observed in the ageing and transplantation model reflect aspects of how each individual mutation affects haematopoiesis, the compound *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ mutation yielded several unique features including hypocellular bone marrow and monocytosis. In the reported single mutant models, none of the *Tet2* models described a decreased bone marrow cellularity (An et al., 2015; Moran-Crusio et al., 2011; Pan et al., 2017; Quivoron et al., 2011), while our *Srsf2* model showed a trend of lower bone marrow counts in the *Srsf2*$^{P95H/+}$ mice (Smeets et al., 2018). Looking at the populations in the bone marrow, a drop in the progenitor, lymphoid and erythroid compartments are likely to be the source of hypocellularity. It was surprising to find a decreased myeloid progenitor population especially when there was a significant increase in

mature granulocytes. It is unknown at present whether this was caused by excess apoptosis of the progenitor populations or an increased differentiation of progenitors without compensation from stem cells. Looking at the proportions of the stem cell containing population ($lin^{-}c\text{-}Kit^{+}Sca\text{-}1^{+}$), the increase of ST-HSC may suggest that there was an HSC response, with a greater expansion of the committed cycling population being generated (Morrison et al., 1995). However, this response was yet to be apparent at the committed progenitor level at the time of analysis. This might require additional mutations that enable the greater expansion of the committed cells.

Of all the phenotypes the monocytosis and splenomegaly is the most significant finding of the compound mutation model. Both *SRSF2* and *TET2* mutations are known to be associated with monocytosis in patients (Itzykson et al., 2013; Li et al., 2011; Patel et al., 2017). The compound mutant mice displayed monocytosis, as well as increased immature promonocytes and monoblasts, suggesting a synergistic effect of *Srsf2/Tet2* mutation on the early expansion of the monocytic blast population. Single cell sequencing of $Tet2^{-/-}$ uncommitted haematopoietic stem cells found a transcriptional program skewed towards myelomonocytic progenitors (Izzo et al., 2020). The skewing is likely due to the difference in CpG enrichment in the myelomonocytic transcription factor motifs. Compared to known erythroid transcription factor motifs, monocytic transcription factors, such as *Irf8* and *Spi1*, showed a lower CpG content in their motifs. Therefore, these transcription factors are more sensitive to hypermethylation changes induced by $Tet2^{-/-}$ (Izzo et al., 2020). Unlike the relationship demonstrated between enhancer methylation and transcriptional priming, there is currently no known direct link between changes in RNA splicing and monocytic skewing. Profiling of

the methylation patterns through DNA bisulfite sequencing on selected transcription factor motifs in the *Srsf2*$^{P95H/+}$ single mutant and *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ compound mutant could reveal any methylation changes associated with mis-splicing.

In addition to the phenotypes observed from native haematopoiesis, the transplantation study revealed that the compound mutant stem cells had a lower engraftment capacity than the *Tet2*$^{-/-}$ cells when in competition with fresh wild-type cells. This is the opposite of my initial hypothesis which proposed that *Tet2*$^{-/-}$ would enhance the fitness of *Srsf2*$^{P95H/+}$ stem cells. Instead, the compound mutant cells had a comparable engraftment rate to single *Srsf2*$^{P95H/+}$ when transplanted competitively. The poor competitiveness of the double mutant stem cells could suggest a decreased fitness, or an increased cycling activity. Looking at the lineage repopulation at 52 weeks post transplantation, there was a much lower contribution from double mutant cells in the non-competitive transplant recipients, marked by the eYFP percentage. This suggests that both the self-renewal (size of the stem cell pool) and differentiation (lineage repopulation) capacity in the double mutant stem cells was compromised. Looking at the overall effect of these phenotypes, the fitness of compound mutant stem cells is similar to single *Srsf2*$^{P95H/+}$. This suggests that splicing changes induced by *Srsf2*$^{P95H/+}$ mutation buffer the stem cell expansion caused by *Tet2*$^{-/-}$. This neutralizing phenomenon could be due to the mis-splicing of pathways that *Tet2*$^{-/-}$ mutation utilises in modifying the stem cell behaviour. Re-visiting such pathways affected by *Tet2*$^{-/-}$ and identifying any mis-splicing events associated with *Srsf2*$^{P95H/+}$ could test the hypothesis. It is also possible that the two mutations affect different stem cell populations (LT-HSC vs. ST-HSC) and jeopardize the normal re-populating process. It would be appropriate to perform transcriptome

analysis on the different stem cell populations which could unveil the consequences of bearing *Srsf2/Tet2* mutations.

Using the *Srsf2/Tet2* compound mutant model, I presented evidence of functional co-operativity of *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ in native haematopoiesis and disease initiation. However, I am yet to define the mechanistic basis of how these two mutations co-operate to initiate and enable evolution to CMML. A detailed analysis of the disease that arises in the compound mutants and comparison to the genetic evolution of CMML would be warranted. Furthermore, analysis of the transcriptomic and epigenetic changes induced by the single and co-operative mutations may provide important insight into the mechanism and cellular changes that lead to the initiation of CMML.

Lastly, I have provided preliminary analysis of the co-operativity of *Srsf2*$^{P95H/+}$ and *Cbl* mutation *in vivo*. Due to breeding issues, I only managed to generate cohorts of *Srsf2*$^{P95H/+}$ *Cbl*$^{-/-}$ mutant instead of *Srsf2*$^{P95H/+}$ *Cbl*$^{A/-}$, which was the target genotype. I analysed and presented the haematological changes of compound mutant mice after one year of mutation activation. I also included data from two individual mice analysed at 10 and 20 weeks post mutation activation. Interestingly, there was dramatic myeloid skewing in the *Srsf2*$^{P95H/+}$ *Cbl*$^{-/-}$, the homozygous compound mutant, after only 10 weeks of tamoxifen activation. The heterozygous compound mutant, *Srsf2*$^{P95H/+}$ *Cbl*$^{+/-}$, also developed myeloid bias in the bone marrow and peripheral blood after 20 weeks of mutation activation. The timing of myeloid bias might suggest a potential gene-dosage effect on the phenotype. More analysis and greater numbers of each genotype would be required to precisely determine the co-operativity of *Srsf2* and *Cbl* mutation.

Chapter 7

Final discussion and conclusion

RNA splicing is an essential process that governs cellular proliferation, differentiation and death in both normal homeostasis and in cancer. Disruption of the core splicing machinery or its key regulatory components promotes tumorigenesis and creates cancer-specific alternative splicing events (Dvinge et al., 2019; Venables et al., 2009). These events are found in tumours across different tissue types, such as haematological, breast, colorectal, ovarian, skin (melanoma) and lung cancers (Brooks et al., 2014; Kahles et al., 2018; Koh et al., 2015; Seiler et al., 2018). Alternative splicing is an important cellular function to diversify the proteome from a fixed genome (Wahl et al., 2009). However, cancer cells take advantage of the process and create aberrantly spliced genes that can lead to increased survival and drug resistance (Berman et al., 2016; Poulikakos et al., 2011; Smeby et al., 2019). In the past decade, the discovery of splicing factor mutations in myelodysplastic syndromes (MDS) and related myeloid neoplasms has further highlighted RNA splicing as a central mechanism of tumorigenesis in haematological malignancies (Haferlach et al., 2014; Yoshida et al., 2011).

Splicing factors, such as serine-arginine (SR) proteins, influence the choice of exon inclusion during RNA splicing. SRSF2 is one of the splicing factors that is frequently mutated in MDS/MPN and CMML (Haferlach et al., 2014; Itzykson et al., 2013). Interestingly, the vast majority of mutations in *SRSF2* target the P95 residue. SRSF2 mutation results in altered recognition of the RNA binding motif and can lead to exon skipping events (Kim et al., 2015; Zhang et al., 2015). To fully explore the role of $SRSF2^{P95H}$ mutation in haematopoiesis, several groups have generated genetically engineered murine models of $Srsf2^{P95H/+}$, which I reviewed in Chapter 1. These models identified phenotypic and transcriptomic

evidence of the disruption caused by *Srsf2*$^{P95H/+}$ mutation, such as initiation of MDS and increased aberrant spliced transcripts. Yet, it remained unclear whether the haematopoietic stem cell was the only cell of origin in SRSF2mut MDS. In Chapter 3, I characterised a conditional knock-in model of *Srsf2*$^{P95H/+}$, *LysM*-Cre *R26*eYFP *Srsf2*$^{P95H/+}$, which activated the *Srsf2* mutation in the myeloid progenitor population and left stem cells mutation-free. I discovered that *Srsf2*$^{P95H/+}$ mutation cannot initiate MDS from within the myeloid progenitor population, even though it produced a mild myeloid biased phenotype. Therefore, splicing changes and/or effects of *Srsf2*$^{P95H/+}$ within the stem cell containing populations were essential for initiating tumorigenesis.

The murine models have confirmed that *Srsf2*$^{P95H/+}$ can act as the founder mutation in MDS , which is consistent with the clonal analysis of human MDS. However, large-scale sequencing studies of MDS, MDS/MPN and related myeloid cancers have demonstrated that *SRSF2* mutations are not the sole cause of transformation in these cancers (Haferlach et al., 2014; Yoshida et al., 2011). One of the most frequent accomplices of *Srsf2*$^{P95H/+}$ is *TET2* mutation (Merlevede et al., 2016). TET2 participates in the DNA methylation pathway, which is well-studied regarding its role in the regulation of normal and malignant haematopoiesis (Celik et al., 2016). *SRSF2* (50%) and *TET2* (60%) are the most frequent mutations reported in CMML, and they display a strong positive association with each other. In Chapter 6, I generated and characterised a *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ compound mutant mouse model to explore and understand the genetic co-operativity of these mutations. The resulting murine model demonstrated that when both *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ were combined, it resulted in a monocytic phenotype and could initiate CMML in both native and transplantation settings. Unexpectedly, the addition of *Tet2* mutation to

the *Srsf2*$^{P95H/+}$ mutation failed to accelerate the disease onset, which indicates that additional mutations must arise to enable disease progression. The changes in the *Srsf2*$^{P95H/+}$ *Tet2*$^{-/-}$ mice are distinct from either single mutant and delineate a different trajectory of disease evolution, as the compound mutant produced phenotypes most consistent with CMML, instead of MDS as was evidenced in the single *Srsf2*$^{P95H/+}$ mutants (Kim et al., 2015; Kon et al., 2018; Merlevede et al., 2016; Smeets et al., 2018). Due to the long disease latency and delays in breeding large cohorts of mice, the number of individuals identified with CMML during my PhD are low to date, however additional animals are currently ageing. Further cohorts are warranted to continue the analysis and to define the mechanistic basis of the cooperativity between these mutations.

In my analysed experiment cohorts, only a few compound mutants displayed the CMML phenotype at the time of analysis. Nevertheless, the compound mutants do present with a fully penetrant myeloid bias, which is similar to the aged human population with Clonal haematopoiesis of indeterminate potential (CHIP) (Jaiswal et al., 2014; Pang et al., 2011). Such individuals harbor some of the same pathogenic mutations seen leukaemia patients, including *SRSF2* and *TET2* mutations, although they are clinically disease-free (Genovese et al., 2014). To find out what triggers the development of overt disease in the subset of compound mutants, the first step would be to perform exome sequencing of the bone marrow cells to identify the additional mutations that the sick individuals may have acquired, compared to 'healthy' mutants and age-matched wild-type controls. These additional mutations are likely to be the enablers of disease progression. The second step would be to assess whether the CMML disease that I have documented in this model is transplantable. Serial transplantation is performed to test the capacity of haematopoietic stem

cells to sustain haematopoiesis (Purton and Scadden, 2007). In leukaemia, this method is used to demonstrate overt transformation of the cells and also to provide an additional stress to the cells to accelerate the initiation of acute leukaemia (Kennedy and Barabé, 2008; Somervaille and Cleary, 2006). Preliminary transplantation studies of bone marrow from the sick non-competitive transplant recipient #835 and #836 revealed no development of acute leukaemia despite strong myelomonocytic bias in the donor mice. The diseased bone marrow achieved the minimum desirable but variable engraftment rate (20-60%) as the transplantation recipients aged. The proportion of granulocytes detected in the peripheral blood showed no difference between wild-type and diseased mutants. The engraftment rate from the diseased compound mutant was slightly lower than the secondary transplantation of *Srsf2*$^{P95H/+}$ cells (40-80%) (Smeets et al., 2018), but despite this low engraftment, the compound mutant donor population could be detected for up to 20 weeks post transplantation. This indicates that long-term haematopoietic stem cells (LT-HSC) were engrafting and maintaining output (data not shown). As the CMML disease observed in my experimental cohorts differs from case to case, it is unknown whether CMML cells from other diseased individuals harbour similar engraftment and differentiation capacity. Transplantation studies of the remaining diseased individuals are currently ongoing and will be important to understand the frequency of disease initiating cells and also the nature of the transformation that may occur.

While the *in vivo* studies of splicing mutations recapitulated the events that occur during the complex evolution of human leukaemogenesis, the *in vitro* study of *Srsf2*$^{P95H/+}$ has led to more insight into the molecular changes resulting from this mutation. In Chapter 4, I established cell lines through Hoxb8 immortalisation to generate three conditional

isogenic *Srsf2*$^{P95H/+}$ and three *Srsf2*$^{+/+}$ cell lines (Wang et al., 2006). These cell lines are dependent on GM-CSF and represent a relatively mature myeloid cell type. Using the cell lines, I could reproduce the mis-splicing changes seen in the *Srsf2*$^{P95H/+}$ stem and progenitor cell population, which confirms the validity of the *in vitro* model. Equipped with CRISPR-Cas9 technology, I was able to perform a genetic screen using a genome-wide knock-out library to investigate the genetic vulnerabilities of *Srsf2*$^{P95H/+}$ mutation (Hart et al., 2017; Tzelepis et al., 2016).

So far, little is known about the synthetic lethality of splicing mutants. The targeting strategy for most splicing inhibitors relies on further inhibition of spliceosome activity, which is also poorly tolerated by wild-type cells (Bonnal et al., 2012; Hong et al., 2014; Lee and Abdel-Wahab, 2016). Thus, pinpointing and targeting the pathways that splicing mutant cells are sensitive to when disturbed would provide an alternative strategy for treating splicing mutant cancers. From my synthetic lethal screen (Chapter 5), the DNA damage and cell cycle pathways were identified as candidate pathways of interest. During validation, *Srsf2*$^{P95H/+}$ cells showed increased sensitivity to Palbociclib, a CDK6 inhibitor. Interestingly, the top negatively enriched pathways also contained transcripts from the most frequently mis-spliced targets when I compared multiple SRSF2mut datasets from both human and murine samples (Hurtado et al., 2018; Kim et al., 2015; Komeno et al., 2015; Kon et al., 2018; Qiu et al., 2016; Shiozawa et al., 2018; Smeets et al., 2018; Zhang et al., 2015). The overlapping of pathways would suggest that *Srsf2*$^{P95H/+}$ mutant cells rely on the aberrant transcripts for survival. Assessing isoform changes and mis-splicing could help identify any key transcripts that are involved in these processes. Complementary to the loss-of-function screen, a gain-of function screen could be established using the same cell lines and a

CRISPR activator library (Chavez et al., 2015; Gilbert et al., 2013; Konermann et al., 2015). Instead of synthetic lethality, the screen could identify genes that promote the transformation of *Srsf2*$^{P95H/+}$ mutant cells. From a therapeutic point of view, a chemical screen could also be utilised to identify more compounds that target *Srsf2*$^{P95H/+}$ mutant cells.

In summary, from murine models to CRISPR screening, I have applied multiple orthogonal experimental methods to investigate the effect of *Srsf2*$^{P95H/+}$ mutation on haematopoiesis and to understand its genetic co-operativity. Studies of positive co-operativity shed light on how co-occurring *Srsf2*$^{P95H/+}$ and *Tet2*$^{-/-}$ mutations can lead to CMML *in vivo*, while negative co-operativity indicated that pathways other than RNA splicing could hold the key to targeting splicing mutants. Furthermore, I was able to perform single-cell RNA sequencing analysis and reveal temporal and lineage discrepancies of the *Srsf2*$^{P95H/+}$ stem cell differentiation compared to wild-type stem cells. In the future, more studies are warranted to determine the mechanisms of *Srsf2*$^{P95H/+}$ induced tumorigenesis and how we can treat splicing mutant cancers efficiently.

Appendix A

Publications associated with this thesis

Smeets, M. F., Tan, S. Y., **Xu, J. J.**, Anande, G., Unnikrishnan, A., Chalk, A. M., Taylor, S. R., Pimanda, J. E., Wall, M., Purton, L. E., and Walkley, C. R. (2018). Srsf2 P95H initiates myeloid bias and myelodysplastic/myeloproliferative syndrome from hemopoietic stem cells. Blood 132, 608-621.

Xu, J. J., Smeets, M. F., Tan, S. Y., Wall, M., Purton, L. E., and Walkley, C. R. (2019). Modeling human RNA spliceosome mutations in the mouse: not all mice were created equal. Experimental Hematology 70, 10-23.

From www.bloodjournal.org by guest on August 9, 2018. For personal use only.

blood®
Regular Article

MYELOID NEOPLASIA

Srsf2^{P95H} initiates myeloid bias and myelodysplastic/myeloproliferative syndrome from hemopoietic stem cells

Monique F. Smeets,[1,2] Shuh Ying Tan,[1-3] Jane Jialu Xu,[1,2] Govardhan Anande,[4] Ashwin Unnikrishnan,[4] Alistair M. Chalk,[1,2] Scott R. Taylor,[1] John E. Pimanda,[4-6] Meaghan Wall,[1,2,7] Louise E. Purton,[1,2] and Carl R. Walkley[1,2]

[1]St. Vincent's Institute, Fitzroy, VIC, Australia; [2]Department of Medicine, St. Vincent's Hospital, University of Melbourne, Fitzroy, VIC, Australia; [3]Department of Hematology, St. Vincent's Hospital, Fitzroy, VIC, Australia; [4]Adult Cancer Program, Prince of Wales Clinical School, Lowy Cancer Research Centre, University of New South Wales, Sydney, NSW, Australia; [5]Department of Hematology, [6]Department of Pathology, Prince of Wales Hospital, Sydney, NSW, Australia; and [7]Victorian Cancer Cytogenetics Service, St. Vincent's Hospital, Fitzroy, VIC, Australia

KEY POINTS

- ***Srsf2*$^{P95H/+}$ mutation within hemopoietic stem cells is required to initiate myeloid-biased hemopoiesis.**
- **Mutation of *Srsf2* is sufficient to initiate the development of MDS/MPN in vivo in the setting of native hemopoiesis.**

Mutations in *SRSF2* occur in myelodysplastic syndromes (MDS) and MDS/myeloproliferative neoplasms (MPN). *SRSF2* mutations cluster at proline 95, with the most frequent mutation being a histidine (P95H) substitution. They undergo positive selection, arise early in the course of disease, and have been identified in age-related clonal hemopoiesis. It is not clear how mutation of *SRSF2* modifies hemopoiesis or contributes to the development of myeloid bias or MDS/MPN. Two prior mouse models of *Srsf2*^{P95H} mutation have been reported; however, these models do not recapitulate many of the clinical features of *SRSF2*-mutant disease and relied on bone marrow (BM) transplantation stress to elicit the reported phenotypes. We describe a new conditional murine *Srsf2*^{P95H} mutation model, where the P95H mutation is expressed physiologically and heterozygously from its endogenous locus after Cre activation. Using multiple Cre lines, we demonstrate that during native hemopoiesis (ie, no BM transplantation), the *Srsf2*^{P95H} mutation needs to occur within the hemopoietic stem-cell–containing populations to promote myelomonocytic bias and expansion with corresponding transcriptional and RNA splicing changes. With age, nontransplanted *Srsf2*^{P95H} animals developed a progressive, transplantable disease characterized by myeloid bias, morphological dysplasia, and monocytosis, hallmarks of MDS/MPN in humans. Analysis of cooccurring mutations within the BM demonstrated the acquisition of additional mutations that are recurrent in humans with *SRSF2* mutations. The tractable *Srsf2*$^{P95H/+}$ knock-in model we have generated is highly relevant to human disease and will serve to elucidate the effect of *SRSF2* mutations on initiation and maintenance of MDS/MPN. (*Blood*. 2018;132(6):608-621)

Introduction

Myelodysplastic syndrome (MDS) represents a spectrum of hematological cancers with a high rate of progression to acute leukemia, where recurrent somatic mutations in the RNA splicing machinery have been identified.[1-3] In MDS, >80% of patients have a mutation in the spliceosome, and in phenotypically overlapping malignancies such as MDS/myeloproliferative neoplasms (MPN) and chronic myelomonocytic leukemia (CMML), spliceosome mutations are also common.[4,5] Spliceosomal mutations arise early in disease, including as initiating events.[6] Mutations generally occur in 1 splicing gene and are heterozygous, and different spliceosomal mutations associate with distinct clinical phenotypes. How mutations in the spliceosome alter normal hemopoiesis and contribute to disease pathogenesis remains unclear.

Mutations in *SRSF2* occur in MDS (~12% to 15%), CMML (28% to 47%), and mastocytosis,[3,4,7] clustering at proline 95 (supplemental Figure 1A, available on the *Blood* Web site). Features of *SRSF2* mutation–associated disease include myeloid bias and an increased risk of transformation to acute leukemia.[8-10] *SRSF2* mutations undergo positive selection and occur in age-related clonal hemopoiesis.[11-13] P95 mutations alter messenger RNA splicing by changing the RNA-binding affinity of SRSF2, leading to misregulation of exon inclusion.[14] This leads to modest overall changes in global messenger RNA splicing, with a subset of misspliced transcripts proposed to be relevant for MDS, including *FYN*, *EZH2*, and *Csf3r*.[14-16]

The first murine *Srsf2*$^{P95H/+}$ conditional mutation model reported a mild dysplastic phenotype characterized by leukopenia and anemia, only observed after bone marrow (BM) transplantation. De novo excised animals did not develop dysplasia/disease.[15,17] These phenotypes showed variable penetrance, given the lack of leukopenia/anemia in an ensuing study with the same allele.[15,17] In a second *Srsf2*^{P95H} model, animals also did not develop disease after 90 weeks of monitoring.[16] Neither model reported myeloid-biased

© 2018 by The American Society of Hematology

From www.bloodjournal.org by guest on August 9, 2018. For personal use only.

hemopoiesis, a feature of *SRSF2* mutation in humans, or MDS in nontransplanted models. Divergent consequences on phenotypic hemopoietic stem cells (HSCs)/progenitors were concluded. As a result, how SRSF2 mutation alters normal hemopoiesis and promotes malignancy remains uncertain.

Here we describe a new *Srsf2*^{P95H} model, where the P95H mutation is conditionally expressed physiologically and heterozygously from its endogenous locus. This model recapitulates the core characteristics of human disease associated with *SRSF2* mutation and demonstrates that the primary effect of SRSF2(P95H) on normal hemopoiesis is to promote myeloid bias from HSCs.

Methods

Experiments were approved by the Animal Ethics Committee (#033/12 and 001/16; St. Vincent's Hospital, Melbourne, Australia).

Animal models

Srsf2$^{P95H/+}$ mice (C57BL/6NTac-*Srsf2*$^{tm2874(P95H)Arte}$) were generated by Taconic Biosciences (Cologne, Germany). A full description of the targeting construct and genotyping procedures is provided in the supplemental Methods. *hScl*-CreER^{T} [18-20], *Rosa26*-eYFP (Jackson Laboratory strain #006148[18-20]), *LysM*-Cre (strain #004781[21-23]), and *Rosa26*-CreER^{T2} (strain #008463[24-26]) mice have been described. B6.SJL-Ptprc$^{aPep3b/BoyJArc}$ mice were purchased from the Animal Resources Centre (Perth, Australia). Heterozygous CD45.1/CD45.2 C57BL/6 mice were bred at St. Vincent's Hospital. Food contained 400 mg/kg of tamoxifen citrate (Sigma) in standard mouse chow (Specialty Feeds, Perth, Australia).[25,26] All mouse lines were on a C57BL/6 background.

Flow cytometric analysis

Peripheral blood (PB) was analyzed on a Sysmex KX-21N (Roche Diagnostics). Cells were prepared and analyzed as described[25,27] on a BD LSRIIFortessa or sorted with either BD FACSAria or BDInflux (BD Biosciences). Results were analyzed with FlowJo version 10.0 (Treestar).

BM transplantation

All transplantations were performed in lethally irradiated (2 × 5 Gy; 3 hours apart) recipients (B6.SJL-Ptprca$^{Pep3b/BoyJArc}$ [Animal Resources Centre]; 3-6 recipients per transplantation cohort) through IV injections. Recipients received antibiotics (Baytril) in the drinking water for 3 weeks after irradiation. Experiments were performed in duplicate or as indicated.

RNA sequencing

RNA was isolated from lin^{-}cKit^{+}eYFP^{+} BM cells from *hScl*-CreER^{T}R26eYFP*Srsf2*$^{+/+}$ and *hScl*-CreER^{T}R26eYFP*Srsf2*$^{P95H/+}$ animals treated 20 weeks prior with tamoxifen. RNA was ribosome depleted, subjected to library preparation (Kapa Stranded RNA-Seq Library Kit; Kapa Biosystems), and sequenced on the Illumina platform by Novogene (Hong Kong). Differential expression was assessed using voom with sample weighting.[28] Quantitative Set Analysis for Gene Expression was used to determine gene set enrichment[29] against the Molecular Signatures Database collection.[20,30] Data sets were deposited in the Gene Expression Omnibus (GSE99852). RNA splicing was assessed using rMATS.[14,16,31] Results are provided in supplemental Data Set 2.

Exome analysis

Exome sequencing was performed on genomic DNA isolated from whole BM and matched ear tissue from 3 independent animals (2 *R26*-CreER^{T2} *Srsf2*$^{P95H/+}$, 1 *hScl*-CreER^{T} *R26*eYFP *Srsf2*$^{P95H/+}$) from 50 to 60 weeks post–tamoxifen treatment (time of euthanasia when moribund). Exome capture was performed with the Agilent SureSelect Mouse Exon system and sequenced on the Illumina platform by BGI Ltd (Hong Kong). Raw and analyzed data are available in the Gene Expression Omnibus (GSE99850).

Statistics

The results were analyzed using the unpaired Student *t* test or analysis of variance (with multiple comparison correction) unless otherwise stated. $P < .05$ was considered significant. All data are presented as mean ± standard error of the mean.

Results

To model SRSF2^{P95H} mutation, we generated a conditional *Srsf2*^{P95H} knock-in allele (P95H). Murine *Srsf2* has 2 closely adjacent genes at each end of the locus, so a whole-locus duplication approach was used. The proline 95 (CCG) to histidine (CAT) mutation was introduced into a duplicated *Srsf2* exon 1, generating a conditional point mutant allele (Figure 1B). In the presence of Cre, the wild-type (WT) *Srsf2* allele is deleted and the mutant *Srsf2*^{P95H} allele is transposed into the endogenous locus. All animals were heterozygous for the knock-in allele. The *Srsf2*$^{P95H/+}$ animals had a mildly elevated mean cell volume at baseline, before Cre activation (supplemental Figure 1D). There was low-level expression of the mutant *Srsf2*^{P95H} allele in the absence of Cre, with a slight increase in B cells and CD4 T cells in the PB but no change in myeloid populations (supplemental Figure 1E-J). When the *Srsf2*^{P95H} allele was activated in vivo, it was expressed heterozygously, genomically, and transcriptionally (Figure 1A; supplemental Figure 1C).

Srsf2^{P95H} causes myeloid bias

We used *R26*-CreER^{T2}*Srsf2*$^{+/+}$ and *R26*-CreER^{T2}*Srsf2*$^{P95H/+}$ animals to elicit widespread expression of P95H. Unless stated, experiments evaluated native hemopoiesis without prior BM transplantation. Tamoxifen-containing food was administered to activate P95H expression (Figure 1B). In the PB, P95H animals developed macrocytosis and reduced red blood cells (RBCs; Figure 1C), with leukocyte counts stable for the first 12 weeks but increasing by ~20 weeks after P95H expression (Figure 1C).

At 22 weeks after tamoxifen, there were increased absolute numbers of myeloid cells and CD4 T cells but no change in B lymphocytes in the PB, indicative of myeloid bias/myeloproliferation (Figure 1D). The expression of CD45RB, a characterized splicing target of Srsf2,[32] was measured on total splenocytes. In *Srsf2*$^{+/+}$, ~80% of splenocytes expressed CD45RB. *Srsf2*$^{P95H/+}$ splenocytes had reduced expression of CD45RB, demonstrating altered protein isoform expression in vivo (Figure 1E). In the BM, there was a trend to reduced overall cellularity (Figure 1F). The granulocyte and monocyte percentages significantly increased in *Srsf2*$^{P95H/+}$ (Figure 1G), while B lymphopoiesis was compromised from the pro–B-cell stage onward (Figure 1H). BM erythropoiesis was reduced, with compensatory splenic erythropoiesis (Figure 1I; supplemental Figure 2F). There was either no significant difference or

From www.bloodjournal.org by guest on August 9, 2018. For personal use only.

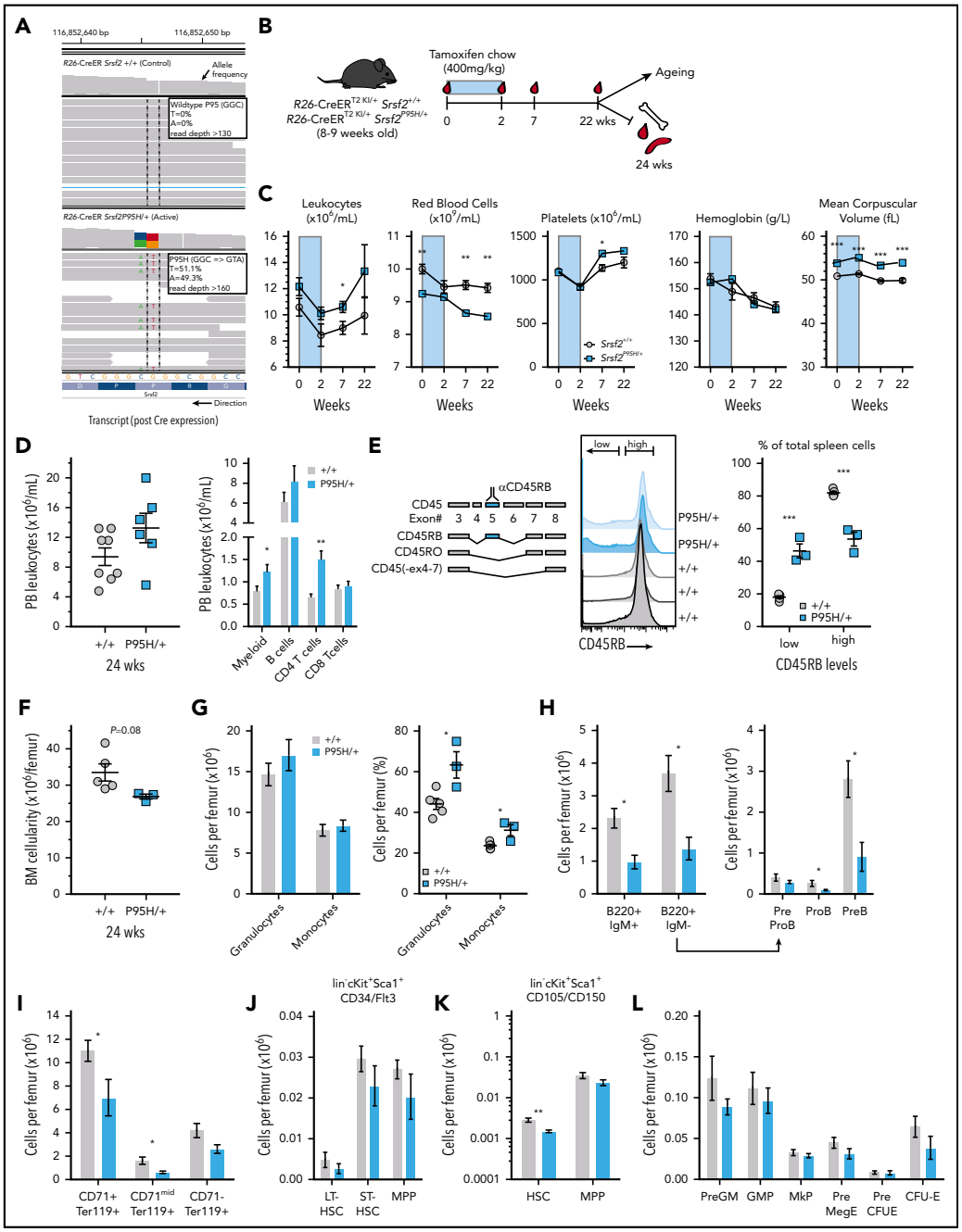


Figure 1. Widespread expression of *Srsf2*$^{P95H/+}$ promotes leukocytosis and myeloid bias. (A) *Srsf2*^{P95H} transcript expression in whole BM from a control and P95H animal. (B) Experimental outline and timeframe. (C) PB indices over 22 weeks after Cre activation. (D) PB leukocyte counts and lineage distribution at 24 weeks after Cre activation (n ≥ 6 per genotype per time point). (E) Histograms and expression of CD45RB isoform on splenocytes (n = 5$^{+/+}$; n = 3 P95H^{+}). (F) BM cellularity per femur. (G) Number and percentage of granulocytes and monocytes per femur. (H) Number of B cells and B-cell progenitor populations per femur. (I) Erythroid populations per femur. Number of phenotypic long-term (LT) and short-term (ST) stem-cell and multipotent progenitor (MPP) populations per femur based on lin^{-}cKit^{+}Sca1^{+}CD34/Flt3 (J) or lin^{-}cKit^{+}Sca1^{+}CD105/CD150 (K) phenotype. (L) Numbers of myeloerythroid progenitors per femur. (E-I) n = 5$^{+/+}$; n = 3 P95H$^{/+}$. Presented as mean ± standard error of the mean. Student *t* test. *$P < .05$, **$P < .01$, ***$P < .001$. CFU-E, colony-forming unit erythroid; GMP, granulocyte macrophage progenitor; IgM, immunoglobulin M; MkP, megakaryocyte progenitor; preMegE, pre-megakaryocyte erythroid progenitor.

From www.bloodjournal.org by guest on August 9, 2018. For personal use only.

decreased numbers of phenotypic LT-HSCs in the P95H BM, using multiple phenotypic methods (Figure 1J-L). The ST-HSCs and MPPs were reduced, with the myeloid progenitor populations least affected (Figure 1J-L; supplemental Figure 2B). Myeloid and erythroid cells increased, and B/T cells were reduced in the spleen, but overall weight/cellularity were not different (supplemental Figure 2C-G). All subsets were reduced in the thymus (supplemental Figure 2H). Therefore, *Srsf2*$^{P95H/+}$ causes myeloid-biased hemopoiesis.

Srsf2^{P95H} mutation is required within the HSCs

To determine if the cell population where P95H was activated would alter the phenotype, *hScl*-CreER^{T}*R26*-eYFP$^{ki/ki}$ was used.[18] Unlike *R26*-CreER, this is specific to the HSC-containing populations and not expressed within the BM stroma/nonhemopoietic tissues. The *hScl*-CreER^{T}–enabled establishment of mosaic hemopoiesis without prior BM transplantation, with eYFP^{+} cells marking the P95H mutants and the eYFP^{-} cells retaining WT *Srsf2*. Thus, it allowed an assessment of P95H cells compared with WT cells from the same BM microenvironment under native conditions.

The PB demonstrated a phenotype comparable to *R26*-CreER^{T2}, with the P95H animals developing leukocytosis, macrocytosis, and reduced RBC numbers (Figure 2A-B). At 20 weeks postactivation, a myelomonocytic expansion and increase in CD4 T cells without a reduction in B lymphocytes were present in the PB (Figure 2C). The PB eYFP^{+} percentage was equivalent between the genotypes (Figure 2D). Expression of CD45RB was reduced (Figure 2E). BM cellularity was similar, but the myeloid cell percentage was significantly increased in the P95H (Figure 2F-G; supplemental Figure 6A). Erythroid and B cells were reduced in the P95H BM. The LT-HSCs and ST-HSCs/MPPs showed either a reduction or trend to a reduction in the P95H BM (Figure 2J-K), with the myeloid progenitors relatively spared (Figure 2L; supplemental Figure 6A). Splenic and thymic hemopoiesis showed trends comparable to the *R26*-CreER, with a more subtle overall effect reflecting the contribution of WT cells (data not shown).

To determine if restricting P95H activation to the myeloid lineage would promote myeloid bias/myeloproliferation, *LysM*-Cre$^{ki/+}$*R26*-eYFP$^{ki/ki}$ cohorts were established (*LysM*-Cre is constitutively expressed). PB monitoring showed a mildly elevated MCV and leukocyte and platelet counts, but no evidence of morphological dysplasia (supplemental Figure 3A-D). Analysis of 20-week-old mice demonstrated that the PB and BM were largely normal (supplemental Figure 3E-K). There was a subtle increase in the percentage of granulocytes in the spleen, but no other phenotypes were apparent (supplemental Figure 3L-M). Therefore, expression of *Srsf2*$^{P95H/+}$ within the HSC population is required to efficiently initiate myeloid-biased hemopoiesis.

Srsf2^{P95H} HSC transplantation potential is determined by their competitor

To assess HSC activity, competitive transplantation of whole BM from *Rosa26*-CreER^{T2}*Srsf2*$^{P95H/+}$ and controls was performed (Figure 3A). In 2 independent experiments, the same outcome occurred: initial chimerism was low and heavily myeloid skewed; then, chimerism was rapidly reduced to low/negligible from the *Srsf2*$^{P95H/+}$ cells (Figure 3B-E). The numbers of competitive repopulating units[33] were reduced by 95% to 99% in the P95H BM (Figure 3B-C). The significantly impaired competitive repopulation of P95H HSCs is consistent with both prior models of *Srsf2* mutations in mice,[15,16] yet at odds with positive selection of *SRSF2* mutations in humans.[11-13]

Transplantation of whole BM from the *hScl*-CreER model was also performed, although unlike *R26*-CreER, the eYFP− population within the same donor BM of each genotype was used as the competitor (Figure 3F). The eYFP levels, indicative of *Srsf2*$^{P95H/+}$ expression, were comparable between genotypes. In this setting, LT chimerism from the eYFP^{+} *Srsf2*$^{P95H/+}$-expressing HSCs was achieved (Figure 3F-H). The P95H recipients had an early myeloid bias and progressive macrocytic anemia, consistent with the nontransplantation models (Figure 3G-J). The contribution to the donor-derived CD45.2 fraction from eYFP^{+} *Srsf2*$^{P95H/+}$ increased over time at the expense of the eYFP^{-} cells (Figure 3I). Therefore, *Srsf2*$^{P95H/+}$ altered HSC transplantation potential in a context-dependent manner. Importantly, it demonstrated that when transplanted with an age-/microenvironment-matched competitor, P95H cells can outcompete WT cells.

Myelomonocytic bias is intrinsic to *Srsf2*$^{P95H/+}$ HSCs

To confirm the phenotype was intrinsic to P95H-expressing hemopoietic cells, whole BM was noncompetitively transplanted before activating expression of P95H from *R26*-CreER and *hScl*-CreER. Once hemopoiesis was reestablished, P95H expression was activated within the donor hemopoietic cells. *R26*-CreER recipients had high chimerism from both genotypes (Figure 4A-C). After tamoxifen, a relative leukopenia developed in the PB, with reduced B cells and increased myeloid cells in the *Srsf2*$^{P95H/+}$ recipients (Figure 4D-E). Immature neutrophils and monocytes accumulated in the PB, as assessed using Ly6C/CD11b (Figure 4F).[34] Within the BM and spleen, the P95H recipients recapitulated the myeloid bias observed in nontransplantation settings (Figure 4G-H). A similar phenotype arose in the *hScl*-CreER recipients (Figure 4I-O), where eYFP^{+} cells produced excess myeloid cells with a reduced contribution to the B lineage in the PB (Figure 4J-N; supplemental Figure 6B). The BM was highly similar to that observed in the nontransplanted animals, with a profound myeloid bias and suppression of other lineages (Figure 4O). In this setting, there was no increase in either proportion or number of phenotypic LT-HSCs but a preservation of ST-HSCs and myeloid progenitors (supplemental Figure 6B). Therefore, *Srsf2*$^{P95H/+}$ mutation promotes hemopoietic cell–intrinsic myeloid bias.

Srsf2^{P95H} initiates MDS/MPN in vivo

Nontransplanted animals from the *hScl*-CreER (Figure 5) and *R26*-CreER (supplemental Figure 4) models were aged after P95H expression. The time to illness was ~8 to 12 months after P95H expression (Figure 5A). Both models demonstrated a similar phenotype. The animals became moribund and required euthanasia because of worsening macrocytosis/RBC counts, with a subset frankly anemic, and development of dermatitis/skin irritation on the snout and eyes (Figure 5C; supplemental Figure 4). In the PB, neutrophil and monocyte frequencies increased, with dysplastic neutrophils displaying hypersegmentation and/or abnormal segmentation (Figure 5B,D; supplemental Figure 4B-C). The increased PB monocytes and immature neutrophils were corroborated using Ly6C/CD11b staining (Figure 5E). Dysplastic cell frequency ranged from >10% to 50%

From www.bloodjournal.org by guest on August 9, 2018. For personal use only.

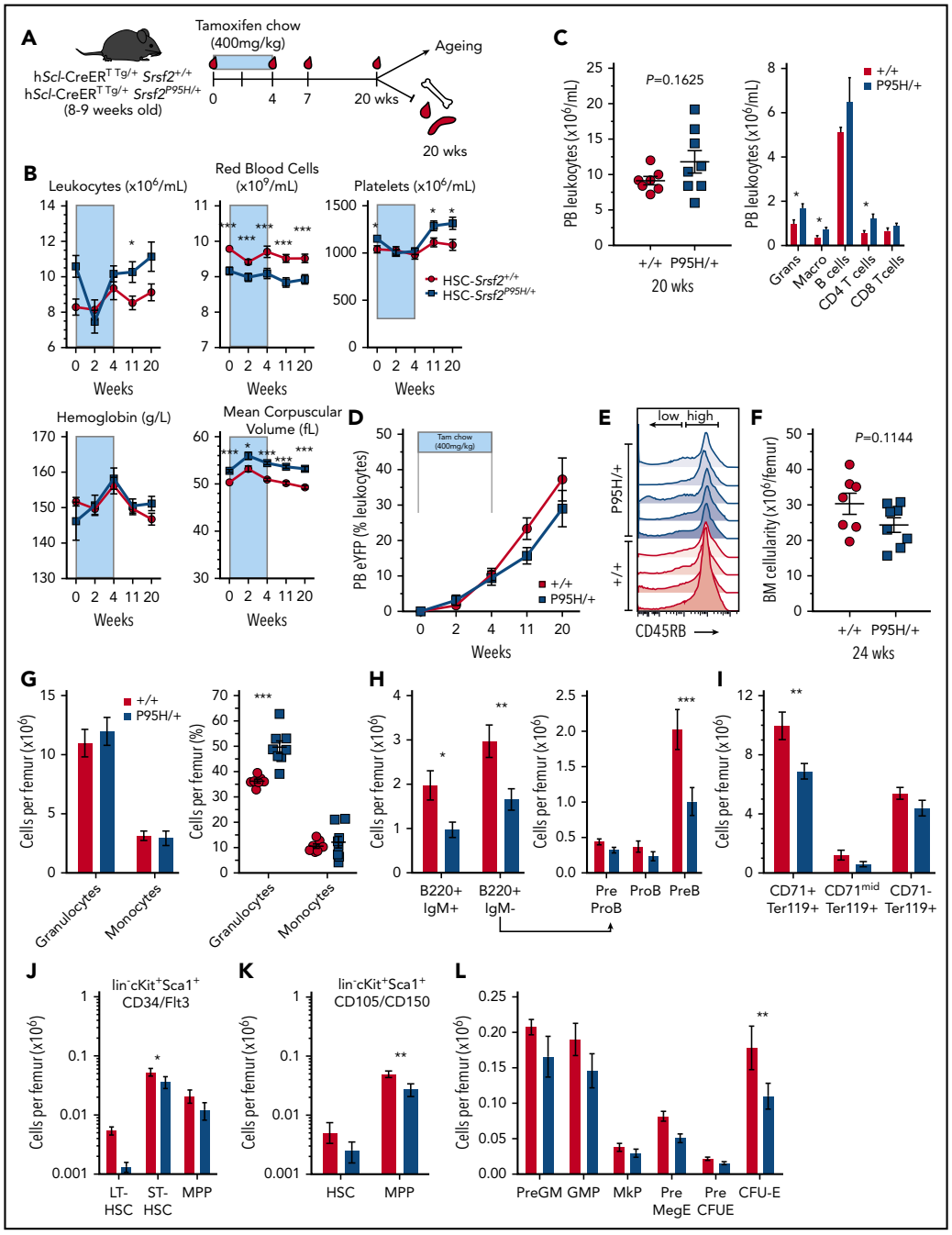


Figure 2. HSC-specific expression of *Srsf2*$^{P95H/+}$ causes myeloid bias. (A) Experimental outline and timeline. (B) PB indices over 20 weeks after Cre activation. (C) PB leukocyte counts and lineage distribution at 20 weeks after Cre activation (n = 7-23 per genotype per time point). (D) PB eYFP levels over time (n = 7-23 per genotype per time point). (E) Histograms of CD45RB isoform expression on splenocytes of the indicated genotypes (n = 4$^{+/+}$; n = 5 P95H$^{/+}$). (F) BM cellularity per femur. (G) Number and frequency of granulocytes and monocytes per femur. (H) Number of B cells and B-cell progenitor populations per femur. (I) Erythroid populations per femur. Number of phenotypic LT-HSC, ST-HSC, and MPP populations per femur using the (J) lin$^-$cKit$^+$Sca1$^+$CD34/Flt3 or lin$^-$cKit$^+$Sca1$^+$CD105/CD150 (K) phenotype. (L) Numbers of myeloerythroid progenitors per femur. (F-K) n = 7$^{+/+}$; n = 8 P95H$^{/+}$. Presented as mean ± standard error of the mean. Student *t* test. *$P < .05$, **$P < .01$, ***$P < .001$. CFU-E, colony-forming unit erythroid; GMP, granulocyte macrophage progenitor; IgM, immunoglobulin M; MkP, megakaryocyte progenitor; preMegE, premegakaryocyte erythroid progenitor.

From www.bloodjournal.org by guest on August 9, 2018. For personal use only.

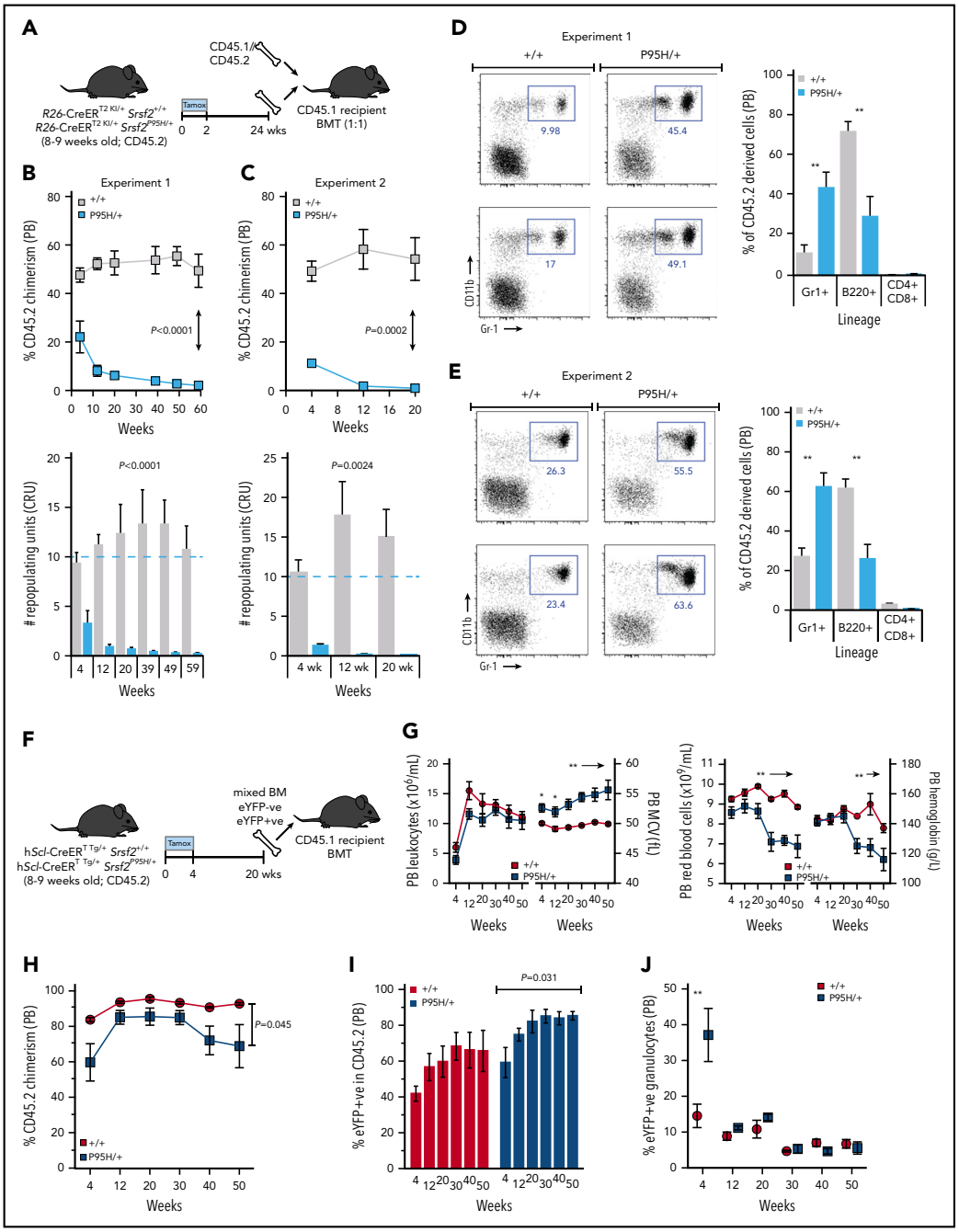


Figure 3. *Srsf2*$^{P95H/+}$ HSCs have a competitive advantage when transplanted with age- and microenvironment-matched support BM. (A) Experimental outline. (B-C) PB chimerism and repopulating units from 2 independent competitive transplantation experiments with BM from *R26*-CreER *Srsf2*$^{+/+}$ (control) and *R26*-CreER *Srsf2*$^{P95H/+}$ animals at 24 weeks after Cre activation (n = 3-5 recipients per transplantation cohort per experiment). (D-E) Fluorescence-activated cell sorting plots for PB CD11b/Gr-1 populations and lineage distribution at 4 weeks posttransplantation from *R26*-CreER model. (F) Experimental outline for *hScl*-CreER model transplantation. (G) PB leukocytes, mean cell volume (MCV), red blood cell counts, and hematocrit over 50 weeks posttransplantation. (H) PB chimerism as measured by CD45.2 staining. (I) eYFP levels within the CD45.2$^+$ fraction for each genotype. (J) Percentage of eYFP$^+$ granulocytes in the PB over time. Presented as mean ± standard error of the mean. Student *t* test or analysis of variance with repeated measures. **P* < .05, ***P* < .01. BMT, BM transplantation; CRU, competitive repopulating unit.

From www.bloodjournal.org by guest on August 9, 2018. For personal use only.

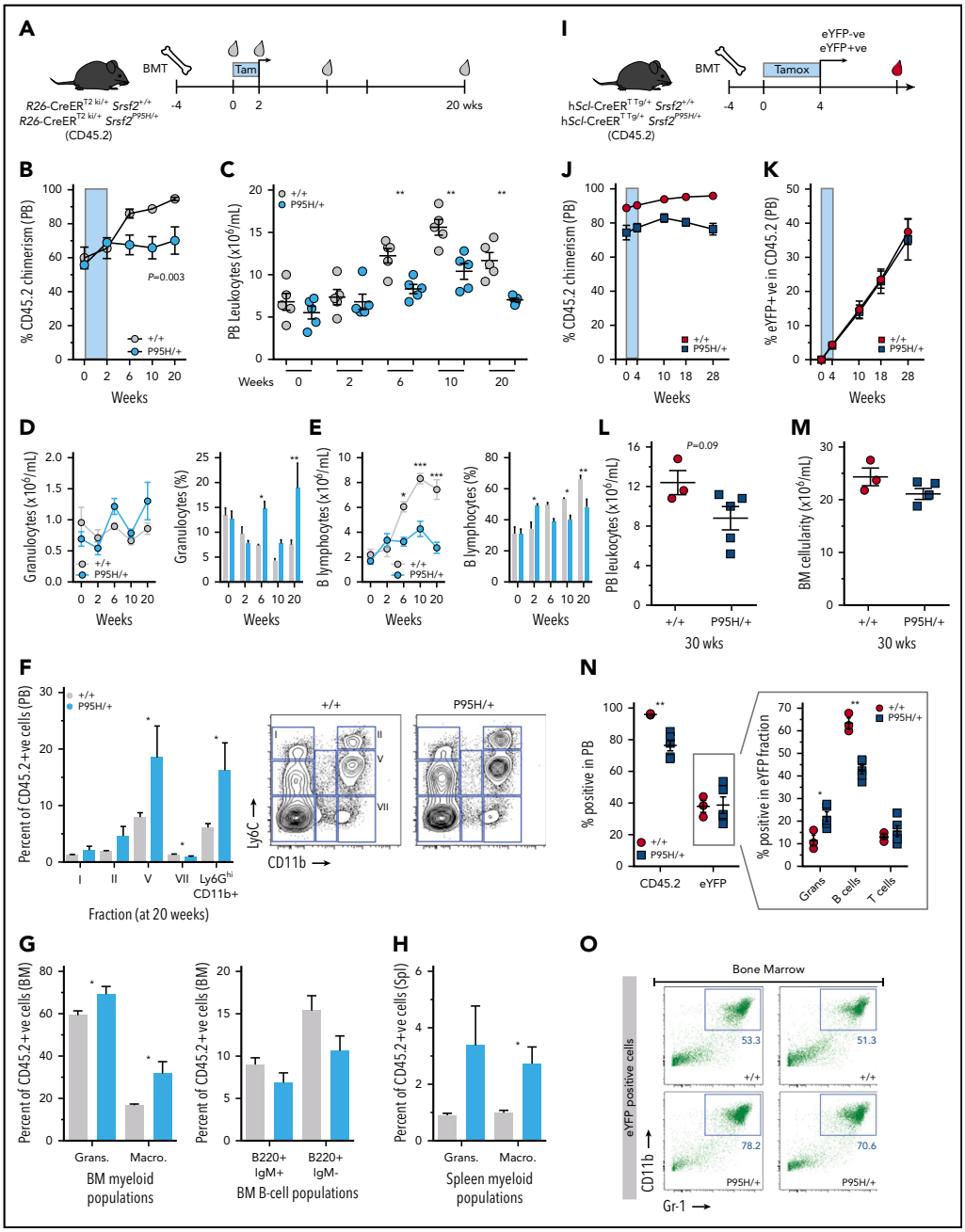


Figure 4. Intrinsic myeloid bias and impaired repopulation of *Srsf2*$^{P95H/+}$ HSCs. (A) Experimental outline and timeframe for transplantation using *R26*-CreER model. (B) PB chimerism. (C) PB leukocyte counts in the recipients of *R26*-CreER *Srsf2*$^{+/+}$ (control) and *R26*-CreER *Srsf2*$^{P95H/+}$ BM over time. Number and percentage of donor-derived granulocytes (D) or B lymphocytes (E) in the PB over time. (F) Quantitation and representative fluorescence-activated cell sorting (FACS) plots of Ly6c/CD11b populations in the CD45.2 donor-derived PB populations: fraction I = Ly6C^{hi}/Mac1^{-}; fraction II = monocytes; fraction V = neutrophil precursors; fraction VII = macrophages. (G) Percentage of CD45.2^{+} granulocytes, macrophages, and B-cell populations in the BM, respectively. (H) Percentage of CD45.2^{+} granulocytes and macrophages in the spleen. (I) Experimental outline and timeframe for transplantation using *hScl*-CreER model. PB chimerism (J) and eYFP percentage (K) within the CD45.2 population over time. PB leukocyte counts (L) and BM cellularity (M) at 30 weeks posttransplantation. (N) Analysis of lineage distribution within the eYFP/CD45.2 population at 28 weeks after Cre activation. (O) FACS plots of the eYFP^{+} CD11b/Gr-1 populations in the BM of recipients of the indicated genotype. Presented as mean ± standard error of the mean; n = 5 recipients per genotype per transplantation model at initiation of transplantation. Student *t* test or analysis of variance with repeated measures. *$P < .05$, **$P < .01$, ***$P < .001$. BMT, BM transplantation.

From www.bloodjournal.org by guest on August 9, 2018. For personal use only.

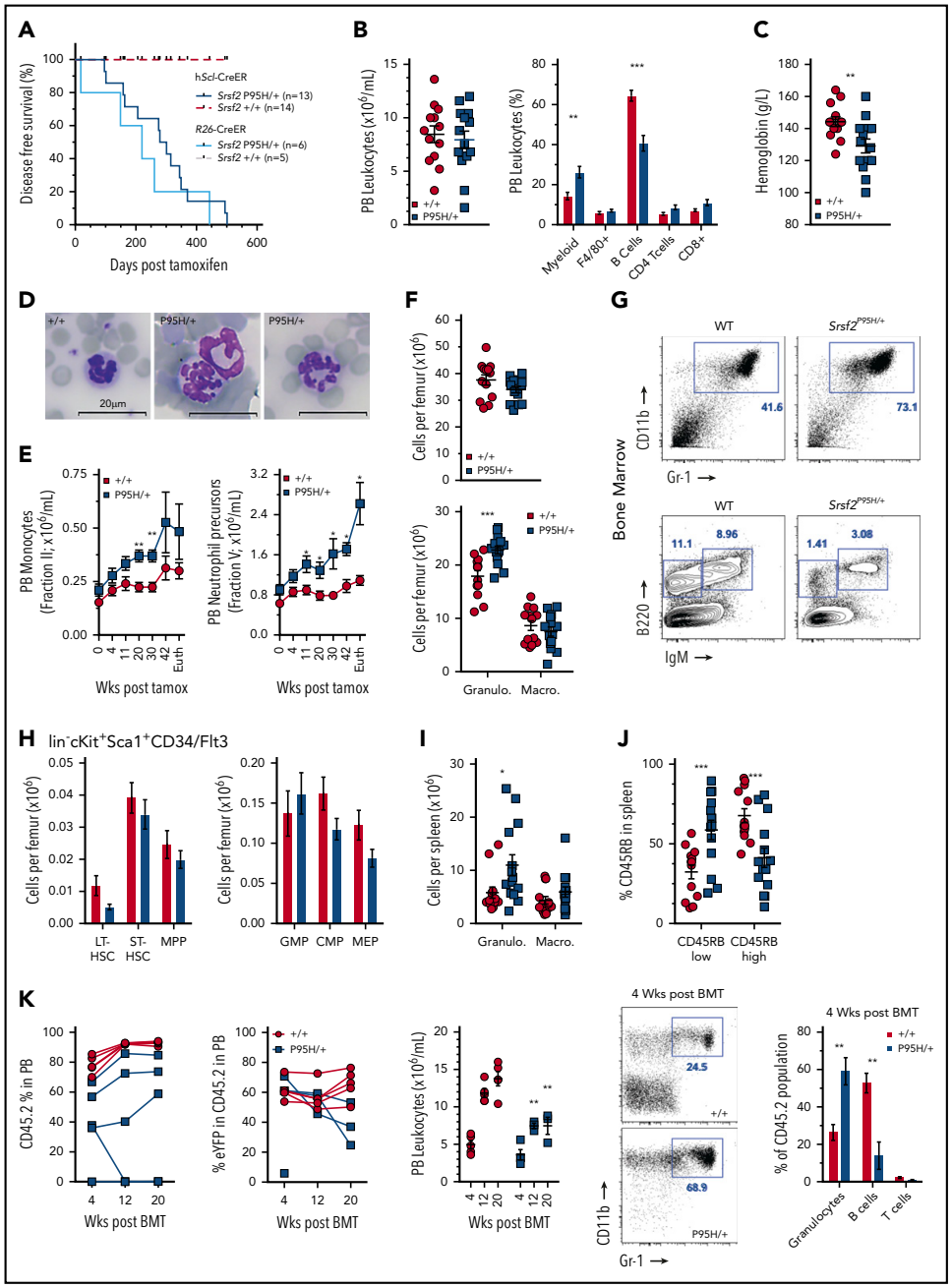


Figure 5. HSC-restricted activation of *Srsf2*$^{P95H/+}$ causes MDS/MPN. (A) Kaplan-Meier plot of disease-free survival for the *hScl*-CreER (blue) and *R26*-CreER (gray) models. (B) PB leukocyte counts and lineage distribution. (C) PB hemoglobin levels. (D) PB neutrophils from a control (age-/treatment-matched *hScl*-CreER *R26*eYFP *Srsf2*$^{+/+}$ littermate) and 3 independent *hScl*-CreER *R26*eYFP *Srsf2*$^{P95H/+}$ animals; May-Grunwald Giemsa stain. (E) Number of PB monocytes and neutrophil precursors over time after tamoxifen (n = 6-13 per genotype per time point; based on Ly6c/CD11b stain). (F) BM cellularity and number of granulocytes and monocytes per femur. (G) BM myeloid (Mac-1/Gr-1) and B lymphoid (B220/immunoglobulin M [IgM]) populations in BM. (H) Number of phenotypic LT-HSCs, ST-HSCs, and MPPs per femur and myeloerythroid progenitors per femur. (I) Number of granulocytes and monocytes per spleen. (J) CD45RB isoform expression on splenocytes. (K) PB CD45 chimerism from secondary transplantation of whole BM into lethally irradiated recipients (black circle = littermate control; blue squares = moribund *Srsf2*$^{P95H/+}$); each line represents reconstitution of an individual recipient (n = 4); PB eYFP percentage in the CD45.2$^+$ cells; PB leukocyte counts, representative PB FACS from recipients, and quantitation of lineage output from the CD45.2 population (4 weeks post–BM transplantation [BMT]). Presented as mean ± standard error of the mean. Student *t* test, with multiple comparisons correction (for panel E). **P* < .05, ***P* < .01, ****P* < .001.

From www.bloodjournal.org by guest on August 9, 2018. For personal use only.

in the PB and BM as assessed morphologically. Platelets were abundant, with occasional large/giant forms present. In the BM, a highly myeloid-biased differentiation was apparent, with a further suppression of B lymphopoiesis (Figure 5F-G). Morphological analysis of BM cytopspins confirmed the diagnosis of MDS/MPN,[35] with reduced erythropoiesis including mild to moderate dyserythropoietic changes, accompanied by granulocytic hyperplasia, monocytosis, and dysplastic changes (supplemental Figure 4D). The LT-HSCs trended toward decreased in absolute number, and the granulocyte macrophage progenitor myeloid progenitors were preserved both in number and proportion (% $P = .003$; Figure 5H; supplemental Figures 4G and 6C). Granulocytes were increased in the spleen (Figure 5I), and CD45RB expression was further compromised (Figure 5I-J; supplemental Figure 4H). Noncompetitive secondary transplantations of whole BM from the moribund *hScl*-CreER (Figure 5K) and *R26*-CreER (supplemental Figure 4I) animals recapitulated the MDS/MPN, but neither group of recipients developed leukemia. Engraftment was only achieved in a subset of recipients, suggestive of decreased HSC frequency and/or transplantability in P95H, consistent with the phenotypic data. *Srsf2*$^{P95H/+}$ expression in HSCs is sufficient to initiate MDS/MPN in vivo.

p53 loss does not promote leukemic conversion

A recurrent route of clonal evolution in *SRSF2* mutation–driven human disease is isochromosome 17q, resulting in haploinsufficiency of *TP53*,[36] and p53 deletion promoted leukemic progression in the NUP98-HOXD13 transgenic MDS mouse model.[37,38] The *R26*-CreER and *hScl*-CreER lines were crossed to a p53 conditional allele (*p53*$^{fl/fl}$). Both Cre models developed a similar phenotype to that described in Figures 1 and 2, except for a preservation of B cells, which is a p53-deficient phenotype in C57BL/6 mice (supplemental Figure 5A-D). Macrocytosis and reduced RBCs accompanied increased granulocytes, B lymphoid cells, and CD4 T cells in the PB (supplemental Figure 5D). Several animals in each cohort developed lymphoma, a p53-null phenotype, and were excluded. Aged animals developed anemia and dermatitis, as occurred in the *p53*$^{+/+}$ animals, requiring euthanasia. These animals had an MDS/MPN phenotype comparable to *p53*$^{+/+}$ animals. Upon secondary transplantation, the P95H p53-deficient cells reconstituted effectively and maintained stable LT reconstitution (supplemental Figure 5E-H). Secondary recipients recapitulated the MDS/MPN phenotype but with no evidence of leukemic transformation.

Myeloid-biased gene expression signatures and altered splicing

The most prominent effect of P95H was myeloid bias/myeloproliferation. To further understand this, we isolated lin$^-$cKit$^+$eYFP$^+$ cells from the BM of the *hScl*-CreER model at 20 weeks after tamoxifen cessation for RNA sequencing. Myeloid bias/myeloproliferation was apparent at this time point, but it occurred before MDS/MPN development. There was reduced expression of lymphoid-related transcripts (*Blk*, *Rag1*, *Lck*, *Ly6d*) and transcripts that regulate myeloid function or lifespan (*Socs3*, *Bcl2l11*; Figure 6A).[39,40] Transcripts previously associated with myeloid expansion (*Hmga2*)[41] and/or HSC function (*Grb10*; Figure 6A)[42] were increased in P95H cells. Quantitative polymerase chain reaction on lin$^-$cKit$^+$ cells from the *R26*-CreER model confirmed altered expression of these transcripts (supplemental data set 1). Pathway analysis corroborated the myeloid bias and identified the enrichment of gene signatures associated with human MDS/MPN[43] and the activation of the p38-MAPK pathway (Figure 6B-C).[44]

Changes in RNA splicing were identified with rMATS,[16,31] with on average <0.5% change in these event types (Figure 6D-E), consistent with reports from P95H-mutant primary mouse and human cells.[14,16] Altered splicing of a range of transcripts, including those previously identified (*Csf3r*, *Arhgap4*, *Tpm3*, *Hnrnpa2b1*, and *Fyn*; supplemental data set 2),[14-16,32] could be identified. *Ezh2* was not misspliced. We validated the findings on independent cell types (splenocytes and Hox-immortalized cell line) and demonstrated reduced inclusion of exon 5 of the *Cd45* transcript in *Srsf2*$^{P95H/+}$ samples, with the concordant reduction of CD45RB (Figure 6F). *Arhgap4* was misspliced in splenocytes; however, we did not observe changes in the splicing of *Dot1l* or *Bcor* (Figure 6G). *Ezh2* was not misspliced in either additional cell type using multiple approaches (Figure 6G-H; supplemental data set 2).[16]

We examined the sensitivity of the myeloid progenitor cell lines to inhibitors of RNA splicing. Two different inhibitors were tested, but there were no substantive differences in sensitivity between WT and P95H cells (Figure 7A). This result is consistent with modest sensitization to splicing inhibition seen in cells from an *Sf3b1*$^{K700E/+}$ mouse model.[45] Therefore, P95H-mutant cells are not substantively sensitized to RNA splicing inhibitors.

Mutation acquisition in *Srsf2*^{P95H} cells mirrors that of human MDS/MPN

As the aged *Srsf2*$^{P95H/+}$ animals developed MDS/MPN, we assessed if the end-stage disease acquired additional mutations similar to those in human MDS/MPN. Exome sequencing was performed on 3 independent moribund animals where whole BM was compared with ear tissue from the same individual to discriminate somatic mutations. Mutations in *Dnmt3a* were present in all P95H BM samples, along with *Phf6* mutations in 2 of 3 samples (Figure 7B). The spectrum of mutations in the P95H BM encompassed similar genes and gene classes to those recurrently mutated in human MDS/MPN and CMML, including mutations in the *Ras* and *Tet2* pathways.[7] We inferred clonal burden using PyClone, which indicated that the additional mutations were arising in a subclone present at ~8% to 10% frequency in the BM at the time of euthanasia (supplemental data set 3). Therefore, the physiologic expression of Srsf2$^{P95H/+}$ within the HSC compartment, alongside acquisition of recurrent secondary mutations commonly observed in humans, leads to the development of MDS/MPN in mice.

Discussion

The primary in vivo consequence of *Srsf2*$^{P95H/+}$ is to promote a hemopoietic cell–intrinsic myeloid bias/myeloproliferation that is apparent in both native hemopoiesis and after multiple distinct BM transplantation models. This phenotype was not observed in prior *Srsf2*^{P95H} mouse models,[15,16] yet is consistent with clinical findings where *SRSF2*-mutant allele frequency correlates with monocytosis.[10] Monocyte and neutrophil precursors increase in the PB from ~11 weeks after P95H activation. This myeloid skewing requires the HSC transcriptional program to efficiently

From www.bloodjournal.org by guest on August 9, 2018. For personal use only.

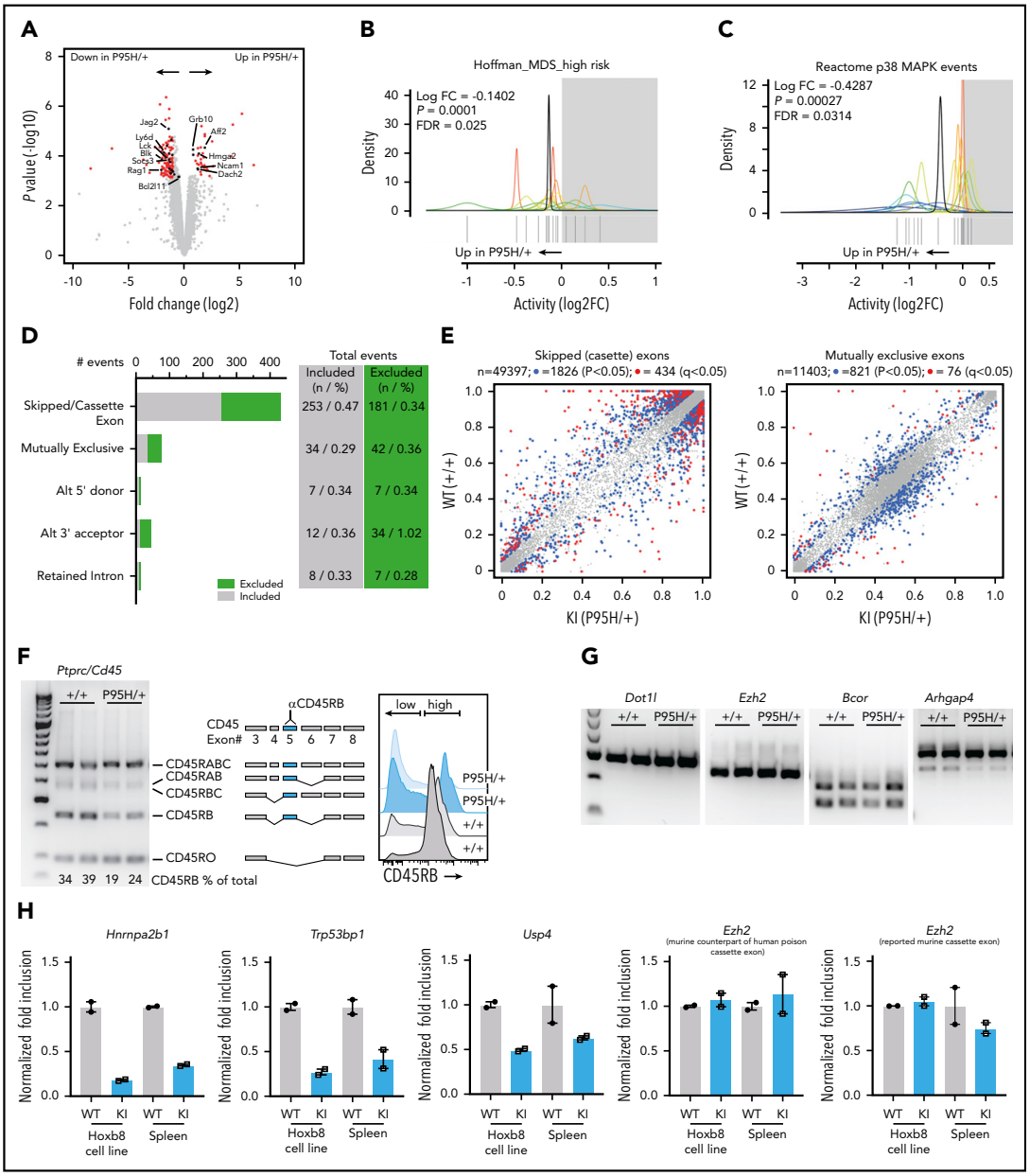


Figure 6. Altered in vivo gene expression and splicing caused by *Srsf2*$^{P95H/+}$. (A) Volcano plot of differential gene expression in BM lin$^-$cKit$^+$eYFP$^+$ cells isolated 20 weeks after tamoxifen cessation (n = 3 per genotype). (B-C) Pathway enrichment using Quantitative Set Analysis for Gene Expression analysis of data set from panel A. (D) Analysis of abnormal splicing events by type (BM lin$^-$cKit$^+$eYFP$^+$ cells; 20 weeks after tamoxifen cessation; n = 3 per genotype). (E) Plots of inclusion events of cassette exons and mutually exclusive exons in the WT and knock-in (KI) samples. Gray dots = <5% difference between genotypes; blue dots = $P < .05$ and >5% difference between genotypes; red dots = $q < 0.05$ and >5% difference between genotypes. (F) *Ptprc/Cd45* isoform at the transcript (left) and protein level of the CD45RB isoform on splenocytes from *R26*-CreER *Srsf2*$^{+/+}$ and *R26*-CreER *Srsf2*$^{P95H/+}$ animals. Polymerase chain reaction (PCR) or quantitative PCR analysis of indicated alternative splicing of the indicated transcripts in WT and KI samples from either splenocytes from *R26*-CreER tamoxifen treated animals (G) or tamoxifen-treated Hoxb8-immortalized myeloid progenitors (H) (n = 2 per genotype per sample type). Data expressed as normalized fold inclusion rate of the cassette exon compared with the exclusion product (WT ratio normalized to 1). FC, fold change; FDR, false discovery rate.

From www.bloodjournal.org by guest on August 9, 2018. For personal use only.

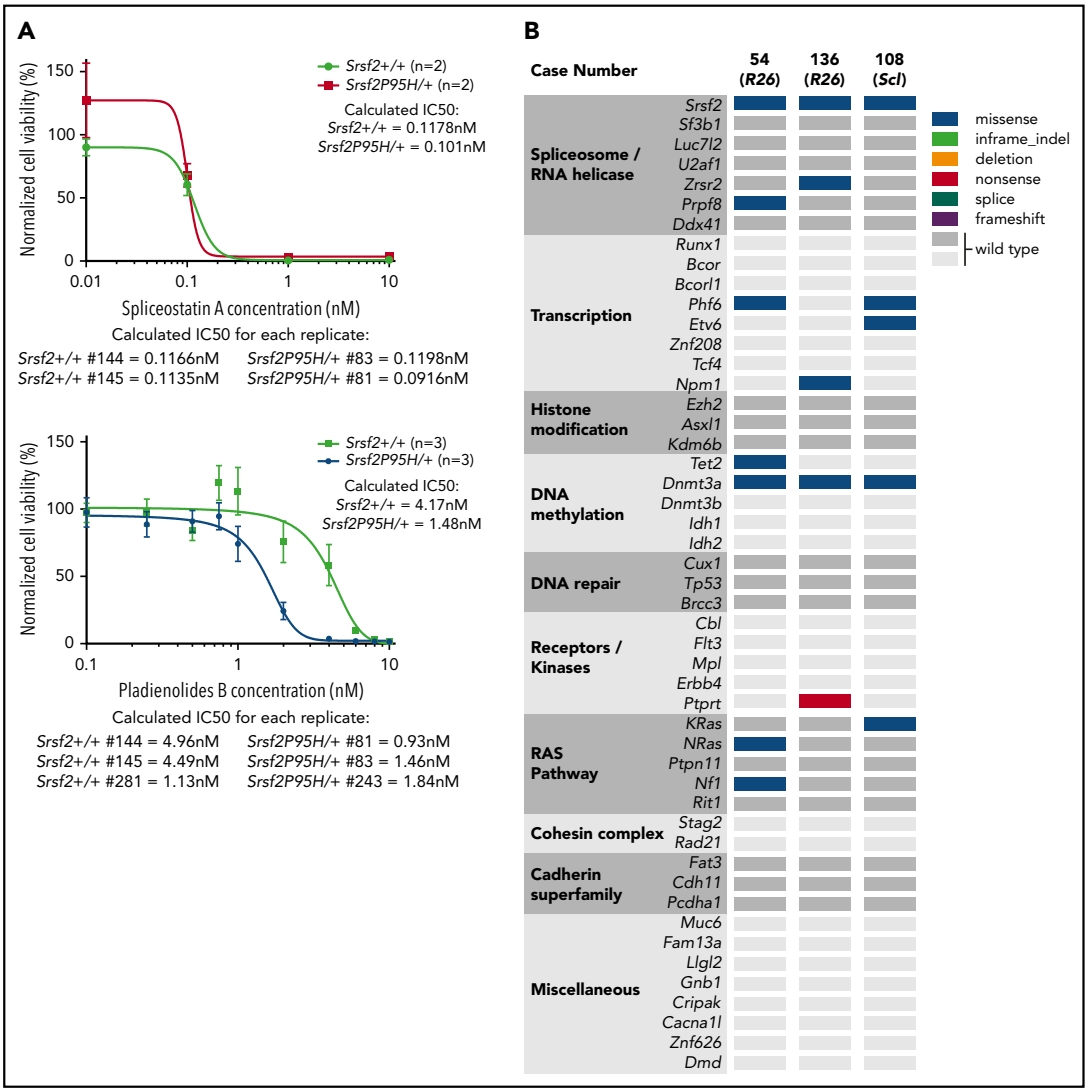


Figure 7. *Srsf2*$^{P95H/+}$ expression does not substantially sensitize to splicing inhibitors but does enable accumulation of human MDS/MPN-relevant mutations. (A) Response of Hox-immortalized cells of the indicated genotypes to the spliceosome inhibitors spliceostatin A (left) and pladienolide B (right) with calculated 50% inhibitory concentration (IC_{50}). (B) Analysis of somatic mutations accumulating in the BM of moribund *R26*-CreER *Srsf2*$^{P95H/+}$ (n = 2) and an *hScl*-CreER R26eYFP *Srsf2*$^{P95H/+}$ animal (n = 1) by exome capture analysis. Mutations are defined against the list of most frequently occurring mutations in human MDS/MPN.

promote disease development, because this phenotype was observed in the *hScl*-CreER model but not the *LysM*-Cre model. P95H corrupts normal lineage fate decisions within the HSC/MPP population, potentially because of the effects of differential splicing within genes critical for early HSC/MPP self-renewal and differentiation, although the identity of these remains elusive both from our analysis and that of others.[14-16] Only 5 transcripts were identified that were differentially expressed and had altered splicing (*Bend5*, *Dach2*, *Dpp7*, *Galns*, and *Prdm16*). Although the function of these transcripts remains to be tested, PRDM16 is involved in translocations in acute myeloid leukemia/MDS. *Srsf2*$^{P95H/+}$ cells had reduced *Prdm16* expression, and potentially relevant to the myeloid bias/monocytosis, *Prdm16*$^{-/-}$ mice had increased numbers of macrophages.[46] It remains to be demonstrated if changes in any single transcript or the cumulative burden of many changes underlie the myeloid bias and ultimately lead to MDS/MPN. The transcriptome analyses to date, including our own, have assessed relatively heterogeneous populations (eg, lin$^-$cKit$^+$ eYFP$^+$ or LKS$^+$).[14-16] The analysis of highly purified LT-HSC and distinct progenitor subsets may yield greater insight into the differentiation stage–specific effects of *Srsf2*$^{P95H/+}$, potentially discerning a greater understanding of how this mutation alters hemopoiesis.

From www.bloodjournal.org by guest on August 9, 2018. For personal use only.

The phenotypic data indicate a trend toward a reduction in absolute numbers of HSCs upon *Srsf2*$^{P95H/+}$ mutation, consistent with results from Kon et al.[16] In competitive transplantations, *Srsf2*$^{P95H/+}$ HSCs demonstrated a profound myeloid bias and >95% reduction in LT repopulation. However, and critically, this was not an absolute finding, because the competitive advantage of the *Srsf2*$^{P95H/+}$-mutant cells can be revealed by transplantation with age-/microenvironment-matched BM cells (Figure 3F-J). This later finding is significant, because the severe reduction in competitive transplantation potential and reduced phenotypic LT-HSCs were inconsistent with *SRSF2* mutations in age-related clonal hemopoiesis and the positive selection of this mutation in humans.[11-13,47] Although the mechanism behind these phenomena remains unresolved, it is tempting to speculate that BM microenvironment remodeling or paracrine factors derived from P95H cells compromise normal hemopoiesis.

With aging, nontransplanted *Srsf2*$^{P95H/+}$ animals develop MDS/MPN, characterized by both monocytosis and dysplastic neutrophils in the PB and BM and the loss of B lymphoid/erythroid populations in the BM. The *Srsf2*$^{P95H/+}$ cells acquired additional mutations in genes and pathways that cooccur in human MPN/MDS. Despite the development of fully penetrant MDS/MPN, the *Srsf2*$^{P95H/+}$ animals did not progress to acute leukemia. The loss of p53 in the *Srsf2*$^{P95H/+}$ background did not modify the myeloid bias or MDS/MPN development but increased the frequency/function of MDS repopulating cells in secondary transplantation assays.[48] The preservation of repopulation upon p53 deletion was reported in the NUP98-HOXD13 model[37,38]; however, unlike that model, we did not observe cooperativity in leukemic conversion.

The phenotypes we report are distinct from those with prior *Srsf2*$^{P95H/+}$ alleles.[15,16,49] In the first reported model, de novo mutated nontransplanted *Srsf2*$^{P95H/+}$ animals did not develop disease by 70 weeks,[15] and BM transplantation–induced stress was required to appreciate some MDS phenotypes. In transplantation-based models, *Srsf2*$^{P95H/+}$ animals did not develop myeloid bias, and there was no evidence of altered protein isoform expression in vivo (CD45RB). In a subsequent report using the same allele, leukopenia and anemia were largely absent in the *Srsf2*$^{P95H/+}$ cohorts,[17] suggesting incomplete penetrance of the phenotype. In contrast to Kim et al,[15] we did not observe expansion of phenotypic primitive progenitor cell populations, identified as LSK CD150^{+}CD48^{+} HPC-2.[50] This may relate to the effects of BM transplantation or use of *Mx*-Cre/pIpC, because the reported frequency of HPC-2 populations in the control animals (*Mx*-Cre *Srsf2*$^{+/+}$) was significantly higher than that in the reference population.[50] Using multiple phenotypic populations, we found evidence for a reduced frequency of phenotypic HSCs and impaired progenitor number/frequency. We reproduced the leukopenia, macrocytosis, and reduced RBCs reported by Kim et al, when Cre was activated post-transplantation (Figure 4A-C). However, in our analysis, the myeloid bias was preserved (supplemental Figure 6A-C). A second *Srsf2*$^{P95H/+}$ allele was recently reported, and *Vav*-Cre nontransplanted animals did not develop myeloid bias or MDS by 90 weeks of age.[16] A reduced frequency of phenotypic and functional HSCs was described.[16] Upon transplantation, erythroid-restricted dysplasia and transcriptional signatures consistent with human MDS were found. Our analysis supports the finding of a depletion of HSCs and primitive progenitors upon *Srsf2*$^{P95H/+}$ expression. However, in contrast to Kon et al,[16] our data demonstrate that *Srsf2*$^{P95H/+}$ mutation within the HSCs is an initiating lesion in MDS/MPN.

We do not have a definitive understanding of why our model develops myeloid bias and MDS/MPN from native hemopoiesis, but the prior models do not. Differences in the genomic modification used to generate the respective alleles are the most plausible source of variance. Kim et al[15] introduced a locus duplication but with a different structure/configuration to the allele we generated. Kon et al[16] used a conditionally inverted PM-bearing exon, and despite heterozygous genomic expression after Cre exposure, the mutant transcript was not heterozygously expressed (~31%). Subheterozygous expression of the mutant transcript was reported in a conditional *Sf3b1*^{K700E} mutant that used an exon-inversion approach.[45] The failure to achieve heterozygous transcript expression is a significant confounding factor in interpreting the phenotypes of these later models.

The contribution of the misspliced transcripts to the phenotypes in the P95H mutants remains an outstanding question. Altered splicing of *EZH2*, leading to reduced/absent expression, was proposed to contribute to the phenotypes in mouse and human P95H$^{/+}$ cells.[15] We did not detect altered *Ezh2* splicing in multiple cell types, yet could demonstrate altered in vivo splicing of previously defined Srsf2 targets[51] and of transcripts identified by Kon et al.[16] Additionally, heterozygous knock-in *SRSF2*$^{P95H/+}$ K562 cells had a very subtle change in *EZH2* splicing.[14] From the analysis by independent groups, missplicing of *Ezh2* does not contribute to the phenotype of *Srsf2*$^{P95H/+}$ cells. We observed a similar relatively subtle shift in overall splicing, with the majority of classes of splicing changing by <0.5%, similar to reports from K562 cells.[14] The modest impact of *Srsf2*$^{P95H/+}$ on global splicing hints at a more complex phenotype than can be solely ascribed to aberrant RNA splicing. Although not demonstrated in vivo, it was recently proposed that increased R-loop formation and DNA damage may contribute to the phenotypes associated with spliceosome mutations.[52]

The *Srsf2*^{P95H} model described here recapitulates the core traits of *SRSF2*-mutant human diseases, including myeloid bias, monocytosis, and progression to MDS/MPN with the accumulation of cooccurring mutations. Importantly, this demonstrates that Srsf2^{P95H} must be expressed within the HSC population to initiate MDS/MPN.

Acknowledgments

The authors thank S. Lane, V. Sankaran, J. Heierhorst, and D. Campbell for comments; A. Kon/S. Ogawa for providing supplemental data ahead of online publication; M. Yoshida (RIKEN) for spliceostatin A; St. Vincent's Hospital Biomedical Research Centre for care of experimental animals; St. Vincent's Institute Flow Cytometry Core Facility; Ramaciotti Centre for Genomics (University of New South Wales); and Novogene for RNA sequencing; and BGI for exome capture.

This work was supported by the Leukaemia Foundation (C.R.W., Phillip Desbrow Senior Research Fellow [SRF]; M.W., grant-in-aid; S.Y.T., PhD scholarship); Cancer Council of Victoria (C.R.W., M.W., A.M.C.; APP1126010); National Health and Medical Research Council (NHMRC; J.E.P., APP1024363/APP1102589); Cancer Institute of NSW Translational Cancer Research Network (J.E.P.); South Eastern Area Laboratory Services (J.E.P.), Anthony Rothe Memorial Trust (A.U.); Victorian Cancer Agency Research Fellowship (C.R.W.; MCRF15015); NHMRC SRF (L.E.P.; APP1003339); UNSW International PhD scholarship (G.A.); and Tom Bee

From www.bloodjournal.org by guest on August 9, 2018. For personal use only.

Stem Cell Research Fund (G.A.) and in part by the Victorian State Government OIS.

Authorship

Contribution: M.F.S., M.W., L.E.P., and C.R.W. were responsible for conceptualization; M.F.S., G.A., A.U., J.E.P., L.E.P., and C.R.W. were responsible for methodology; M.F.S., S.Y.T., J.J.X., G.A., A.U., A.M.C., S.R.T., M.W., and C.R.W. performed investigation; M.F.S., J.J.X., and C.R.W. wrote the original draft; M.F.S., S.Y.T., J.J.X., A.U., J.E.P., M.W., L.E.P., and C.R.W. reviewed and edited the manuscript; A.M.C., J.J.P., M.W., and C.R.W. were responsible for funding acquisition; G.A., A.U., and J.J.P. provided resources; and M.F.S., J.J.P., M.W., L.E.P., and C.R.W. provided supervision.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

ORCID profile: C.R.W., 0000-0002-4784-9031.

Correspondence: Carl R. Walkley, St. Vincent's Institute, 9 Princes St, Fitzroy 3065 VIC, Australia; e-mail: cwalkley@svi.edu.au.

Footnotes

Submitted 15 April 2018; accepted 9 June 2018. Prepublished online as *Blood* First Edition paper, 14 June 2018; DOI 10.1182/blood-2018-04-845602.

The online version of this article contains a data supplement.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

REFERENCES

1. Yoshida K, Sanada M, Shiraishi Y, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*. 2011; 478(7367):64-69.
2. Papaemmanuil E, Cazzola M, Boultwood J, et al; Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. *N Engl J Med*. 2011;365(15):1384-1395.
3. Meggendorfer M, Roller A, Haferlach T, et al. SRSF2 mutations in 275 cases with chronic myelomonocytic leukemia (CMML). *Blood*. 2012;120(15):3080-3088.
4. Haferlach T, Nagata Y, Grossmann V, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia*. 2014;28(2):241-247.
5. Papaemmanuil E, Gerstung M, Malcovati L, et al; Chronic Myeloid Disorders Working Group of the International Cancer Genome Consortium. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood*. 2013;122(22):3616-3627, quiz 3699.
6. Sperling AS, Gibson CJ, Ebert BL. The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. *Nat Rev Cancer*. 2017;17(1):5-19.
7. Makishima H, Yoshizato T, Yoshida K, et al. Dynamics of clonal evolution in myelodysplastic syndromes. *Nat Genet*. 2017;49(2): 204-212.
8. Wu SJ, Kuo YY, Hou HA, et al. The clinical implication of SRSF2 mutation in patients with myelodysplastic syndrome and its stability during disease evolution. *Blood*. 2012; 120(15):3106-3111.
9. Mian SA, Smith AE, Kulasekararaj AG, et al. Spliceosome mutations exhibit specific associations with epigenetic modifiers and proto-oncogenes mutated in myelodysplastic syndrome. *Haematologica*. 2013;98(7): 1058-1066.
10. Sallman DA, Komrokji R, Vaupel C, et al. Impact of TP53 mutation variant allele frequency on phenotype and outcomes in myelodysplastic syndromes. *Leukemia*. 2016; 30(3):666-673.
11. Martincorena I, Raine KM, Gerstung M, et al. Universal patterns of selection in cancer and somatic tissues [published correction appears in *Cell*. 2018;173(7):1823]. *Cell*. 2017;171(5): 1029-1041.e21.
12. Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med*. 2014; 371(26):2488-2498.
13. Genovese G, Kähler AK, Handsaker RE, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med*. 2014;371(26):2477-2487.
14. Zhang J, Lieu YK, Ali AM, et al. Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. *Proc Natl Acad Sci USA*. 2015;112(34): E4726-E4734.
15. Kim E, Ilagan JO, Liang Y, et al. SRSF2 mutations contribute to myelodysplasia by mutant-specific effects on exon recognition. *Cancer Cell*. 2015;27(5):617-630.
16. Kon A, Yamazaki S, Nannya Y, et al. Physiological *Srsf2* P95H expression causes impaired hematopoietic stem cell functions and aberrant RNA splicing in mice. *Blood*. 2018;131(6):621-635.
17. Lee SC, Dvinge H, Kim E, et al. Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins [published correction appears in *Nat Med*. 2016;22(6):692]. *Nat Med*. 2016;22(6):672-678.
18. Göthert JR, Gustin SE, Hall MA, et al. In vivo fate-tracing studies using the Scl stem cell enhancer: embryonic hematopoietic stem cells significantly contribute to adult hematopoiesis. *Blood*. 2005;105(7):2724-2732.
19. Hartner JC, Walkley CR, Lu J, Orkin SH. ADAR1 is essential for the maintenance of hematopoiesis and suppression of interferon signaling [published correction appears in *Nat Immunol*. 2009;10(5):551]. *Nat Immunol*. 2009;10(1):109-115.
20. Liddicoat BJ, Piskol R, Chalk AM, et al. RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. *Science*. 2015;349(6252):1115-1120.
21. Clausen BE, Burkhardt C, Reith W, Renkawitz R, Förster I. Conditional gene targeting in macrophages and granulocytes using LysMcre mice. *Transgenic Res*. 1999;8(4):265-277.
22. Walkley CR, Shea JM, Sims NA, Purton LE, Orkin SH. Rb regulates interactions between hematopoietic stem cells and their bone marrow microenvironment. *Cell*. 2007;129(6): 1081-1095.
23. Liddicoat BJ, Hartner JC, Piskol R, et al. Adenosine-to-inosine RNA editing by ADAR1 is essential for normal murine erythropoiesis. *Exp Hematol*. 2016;44(10):947-963.
24. Ventura A, Kirsch DG, McLaughlin ME, et al. Restoration of p53 function leads to tumour regression in vivo. *Nature*. 2007;445(7128): 661-665.
25. Smeets MF, DeLuca E, Wall M, et al. The Rothmund-Thomson syndrome helicase RECQL4 is essential for hematopoiesis. *J Clin Invest*. 2014;124(8):3551-3565.
26. Heraud-Farlow JE, Chalk AM, Linder SE, et al. Protein recoding by ADAR1-mediated RNA editing is not essential for normal development and homeostasis. *Genome Biol*. 2017; 18(1):166.
27. Singbrant S, Russell MR, Jovic T, et al. Erythropoietin couples erythropoiesis, B-lymphopoiesis, and bone homeostasis within the bone marrow microenvironment. *Blood*. 2011;117(21):5631-5642.
28. Law CW, Chen Y, Shi W, Smyth GK. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol*. 2014;15(2):R29.
29. Yaari G, Bolen CR, Thakar J, Kleinstein SH. Quantitative set analysis for gene expression: a method to quantify gene set differential expression including gene-gene correlations. *Nucleic Acids Res*. 2013;41(18):e170.
30. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA*. 2005;102(43):15545-15550.
31. Shen S, Park JW, Lu ZX, et al. rMATS: robust and flexible detection of differential alternative splicing from replicate RNA-Seq data. *Proc Natl Acad Sci USA*. 2014;111(51): E5593-E5601.
32. Wang HY, Xu X, Ding JH, Bermingham JR Jr, Fu XD. SC35 plays a role in T cell development

From www.bloodjournal.org by guest on August 9, 2018. For personal use only.

and alternative splicing of CD45. *Mol Cell*. 2001;7(2):331-342.

33. Harrison DE. Competitive repopulation: a new assay for long-term stem cell functional capacity. *Blood*. 1980;55(1):77-81.

34. Kovacic B, Hoelbl-Kovacic A, Fischhuber KM, et al. Lactotransferrin-Cre reporter mice trace neutrophils, monocytes/macrophages and distinct subtypes of dendritic cells. *Haematologica*. 2014;99(6):1006-1015.

35. Kogan SC, Ward JM, Anver MR, et al; Hematopathology subcommittee of the Mouse Models of Human Cancers Consortium. Bethesda proposals for classification of nonlymphoid hematopoietic neoplasms in mice. *Blood*. 2002;100(1):238-245.

36. Meggendorfer M, Haferlach C, Zenger M, Macijewski K, Kern W, Haferlach T. The landscape of myeloid neoplasms with isochromosome 17q discloses a specific mutation profile and is characterized by an accumulation of prognostically adverse molecular markers. *Leukemia*. 2016;30(7):1624-1627.

37. Xu H, Menendez S, Schlegelberger B, et al. Loss of p53 accelerates the complications of myelodysplastic syndrome in a NUP98-HOXD13-driven mouse model. *Blood*. 2012;120(15):3089-3097.

38. Guirguis AA, Slape CI, Failla LM, et al. PUMA promotes apoptosis of hematopoietic progenitors driving leukemic progression in a mouse model of myelodysplasia. *Cell Death Differ*. 2016;23(6):1049-1059.

39. Kotzin JJ, Spencer SP, McCright SJ, et al. The long non-coding RNA Morrbid regulates Bim and short-lived myeloid cell lifespan. *Nature*. 2016;537(7619):239-243.

40. Croker BA, Metcalf D, Robb L, et al. SOCS3 is a critical physiological negative regulator of G-CSF signaling and emergency granulopoiesis. *Immunity*. 2004;20(2):153-165.

41. Oguro H, Yuan J, Tanaka S, et al. Lethal myelofibrosis induced by Bmi1-deficient hematopoietic cells unveils a tumor suppressor function of the polycomb group genes. *J Exp Med*. 2012;209(3):445-454.

42. Yan X, Himburg HA, Pohl K, et al. Deletion of the imprinted gene Grb10 promotes hematopoietic stem cell self-renewal and regeneration. *Cell Reports*. 2016;17(6):1584-1594.

43. Hofmann WK, de Vos S, Komor M, Hoelzer D, Wachsman W, Koeffler HP. Characterization of gene expression of CD34+ cells from normal and myelodysplastic bone marrow. *Blood*. 2002;100(10):3553-3560.

44. List AF. New therapeutics for myelodysplastic syndromes. *Leuk Res*. 2012;36(12):1470-1474.

45. Obeng EA, Chappell RJ, Seiler M, et al. Physiologic expression of Sf3b1(K700E) causes impaired erythropoiesis, aberrant splicing, and sensitivity to therapeutic spliceosome modulation. *Cancer Cell*. 2016;30(3):404-417.

46. Cohen P, Levy JD, Zhang Y, et al. Ablation of PRDM16 and beige adipose causes metabolic dysfunction and a subcutaneous to visceral fat switch. *Cell*. 2014;156(1-2):304-316.

47. McKerrell T, Park N, Moreno T, et al; Understanding Society Scientific Group. Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. *Cell Reports*. 2015;10(8):1239-1245.

48. Stengel A, Kern W, Haferlach T, Meggendorfer M, Fasan A, Haferlach C. The impact of TP53 mutations and TP53 deletions on survival varies between AML, ALL, MDS and CLL: an analysis of 3307 cases. *Leukemia*. 2017;31(3):705-711.

49. Komeno Y, Huang YJ, Qiu J, et al. SRSF2 is essential for hematopoiesis, and its myelodysplastic syndrome-related mutations dysregulate alternative pre-mRNA splicing. *Mol Cell Biol*. 2015;35(17):3071-3082.

50. Oguro H, Ding L, Morrison SJ. SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. *Cell Stem Cell*. 2013;13(1):102-116.

51. Shirai CL, White BS, Tripathi M, et al. Mutant U2AF1-expressing cells are sensitive to pharmacological modulation of the spliceosome. *Nat Commun*. 2017;8:14060.

52. Chen L, Chen JY, Huang YJ, et al. The augmented R-loop is a unifying mechanism for myelodysplastic syndromes induced by high-risk splicing factor mutations. *Mol Cell*. 2018;69(3):412-425.e6.

From www.bloodjournal.org by guest on August 9, 2018. For personal use only.

2018 132: 608-621
doi:10.1182/blood-2018-04-845602 originally published online June 14, 2018

Srsf2^{P95H} initiates myeloid bias and myelodysplastic/myeloproliferative syndrome from hemopoietic stem cells

Monique F. Smeets, Shuh Ying Tan, Jane Jialu Xu, Govardhan Anande, Ashwin Unnikrishnan, Alistair M. Chalk, Scott R. Taylor, John E. Pimanda, Meaghan Wall, Louise E. Purton and Carl R. Walkley

Updated information and services can be found at:
http://www.bloodjournal.org/content/132/6/608.full.html

Articles on similar topics can be found in the following Blood collections
Hematopoiesis and Stem Cells (3513 articles)
Myeloid Neoplasia (1849 articles)

Information about reproducing this article in parts or in its entirety may be found online at:
http://www.bloodjournal.org/site/misc/rights.xhtml#repub_requests

Information about ordering reprints may be found online at:
http://www.bloodjournal.org/site/misc/rights.xhtml#reprints

Information about subscriptions and ASH membership may be found online at:
http://www.bloodjournal.org/site/subscriptions/index.xhtml

Blood (print ISSN 0006-4971, online ISSN 1528-0020), is published weekly by the American Society of Hematology, 2021 L St, NW, Suite 900, Washington DC 20036.
Copyright 2011 by The American Society of Hematology; all rights reserved.

ELSEVIER

Experimental Hematology

Experimental Hematology 2019;70:10–23

Modeling human RNA spliceosome mutations in the mouse: not all mice were created equal

Jane Jialu Xu[a,b], Monique F. Smeets[a,b], Shuh Ying Tan[a,b,c], Meaghan Wall[a,b,d], Louise E. Purton[a,b], and Carl R. Walkley[a,b,e]

[a]*St. Vincent's Institute, Fitzroy, Victoria 3065, Australia;* [b]*Department of Medicine, St. Vincent's Hospital, University of Melbourne, Fitzroy, Victoria 3065, Australia;* [c]*Department of Hematology, St. Vincent's Hospital, Fitzroy, Victoria 3065, Australia;* [d]*Victorian Cancer Cytogenetics Service, St. Vincent's Hospital, Fitzroy, Victoria 3065, Australia;* [e]*Mary MacKillop Institute for Health Research, Australian Catholic University, Melbourne, Victoria 3000, Australia*

(Received 27 September 2018; revised 26 October 2018; accepted 1 November 2018)

Myelodysplastic syndromes (MDS) and related myelodysplastic/myeloproliferative neoplasms (MDS/MPNs) are clonal stem cell disorders, primarily affecting patients over 65 years of age. Mapping of the MDS and MDS/MPN genome identified recurrent heterozygous mutations in the RNA splicing machinery, with the *SF3B1*, *SRSF2*, and *U2AF1* genes being frequently mutated. To better understand how spliceosomal mutations contribute to MDS pathogenesis in vivo, numerous groups have sought to establish conditional murine models of *SF3B1*, *SRSF2*, and *U2AF1* mutations. The high degree of conservation of hematopoiesis between mice and human and the well-established phenotyping and genetic modification approaches make murine models an effective tool with which to study how a gene mutation contributes to disease pathogenesis. The murine models of spliceosomal mutations described to date recapitulate human MDS or MDS/MPN to varying extents. Reasons for the differences in phenotypes reported between alleles of the same mutation are varied, but the nature of the genetic modification itself and subsequent analysis methods are important to consider. In this review, we summarize recently reported murine models of *SF3B1*, *SRSF2*, and *U2AF1* mutations, with a particular focus on the genetically engineered modifications underlying the models and the experimental approaches applied. **© 2018 ISEH – Society for Hematology and Stem Cells. Published by Elsevier Inc. All rights reserved.**

Myelodysplastic syndromes (MDS), myelodysplastic/myeloproliferative neoplasms (MDS/MPN), and related disorders are a heterogeneous class of blood cancers leading to ineffective hematopoiesis in the bone marrow (BM) [1,2]. Approximately 30% of MDS patients progress to acute leukemia. Median survival ranges from 97 months for low-risk categories down to 11 months for high-risk MDS [2]. The incidence of MDS in the general population is approximately four to five per 100,000 people, but this increases with age [1]. Population-based studies in both Australia and the United States indicate a significant underestimation of the true burden of MDS, with frequencies estimated at 103 per 100,000 and between 75 and 162 per 100,000, respectively, over age 65 [3–5]. A feature of MDS and MDS/MPN is the progressive establishment of clonal hematopoiesis, in which mutant cells dominate the BM at the expense of normal hematopoiesis [6]. The mechanisms behind this clonal advantage are unknown, but the identification of clonal hematopoiesis of indeterminate potential (CHIP) suggests that the clonal advantage is acquired progressively over many years following the establishment of an initiating mutation [7,8]. In general, MDS patients have limited treatment options. Current treatment strategies for MDS and MDS/MPN are mostly heterogeneous and modestly efficacious and are not associated with durable responses [9]. The dismal prognosis following hypomethylating agent failure highlights the urgent need for new treatment approaches.

Until recently, there were few described recurrent mutations or familial syndromes that could provide insight into the genetics of these cancers that could be used to develop preclinical models. Now, however, the

Offprint requests to: Dr. Carl Walkley, B.Pharm(Hons)., Ph.D., St Vincent's Institute, 9 Princes St, Fitzroy, Victoria 3065, Australia; E-mail: cwalkley@svi.edu.au

0301-472X/© 2018 ISEH – Society for Hematology and Stem Cells. Published by Elsevier Inc. All rights reserved.
https://doi.org/10.1016/j.exphem.2018.11.001

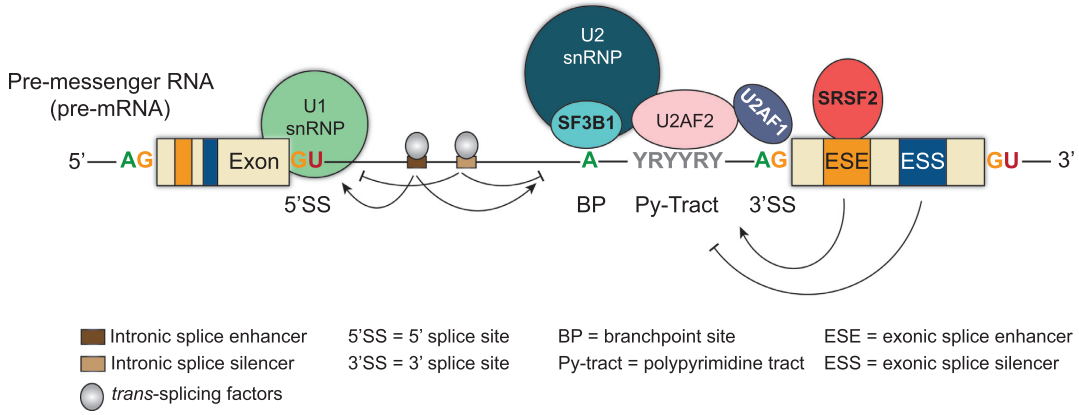


Figure 1. Schematic outline of RNA splicing and the role of those proteins mutated in MDS. Within the intron of the pre-mRNA, the U2 snRNP complex containing SF3B1 binds the branch-point site adenosine, U2AF2 binds to the polypyrimidine tract sequence, and U2AF1 binds the 3′ splice site; SRSF2 binds to the exonic splice enhancer sequence within the exonic sequence. *Cis*- and *trans*-acting factors can both positively or negative regulate splicing. Proteins in which murine models have been generated are shown in bold font.

detailed mutational architecture of human MDS, chronic myelomonocytic leukemia (CMML), and other related forms of MDS/MPN has been defined. There are recurrent mutations in key pathways, including the RNA-splicing machinery (e.g., *SF3B1*, *SRSF2*, and *U2AF1*) [10–12], transcription factors (e.g., *RUNX1*, *BCOR*, and *ETV6*), and epigenetic enzymes (e.g., *DNMT3A*, *TET2*, *EZH2*, and *ASXL1*) [6]. Additional studies have deduced clonal evolution and mutational order, highlighting distinct patterns [13]. Paralleling studies of diseased samples, analysis of healthy elderly populations has demonstrated that mutations associated with MDS and leukemia accumulate and are positively selected for in the hematopoietic cells from otherwise healthy people ≥70 years of age [7,8,14].

The identification of recurrent mutations in the RNA-splicing machinery was a breakthrough in understanding the genetics of MDS and related disorders [10–12]. RNA splicing is a highly coordinated and essential process carried out by major and minor spliceosomes to remove noncoding regions (introns) of the pre-mRNA before protein translation [15,16]. The major spliceosome consists of five small nuclear ribonucleoprotein complexes (snRNPs), U1, U2, U4, U5, and U6. The minor spliceosome includes the U5 snRNPs and other functional snRNPs mirroring the major spliceosome. Mechanistically, RNA splicing occurs with the initial recognition of the 5′ and 3′ splice sites by the U1 and U2 complexes, respectively, followed by the excision of the intron and exon ligation by U4, U5, and U6 complexes. *Trans*-acting splicing factors, such as serine-arginine-rich proteins and heterogeneous nuclear RNPs, bind to the regulatory elements located in the exons and introns to enhance or repress the splicing activity and contribute to alternative splicing (Figure 1). The regulatory elements include exonic splicing enhancers and silencers and intronic splicing enhancers and silencers [16].

In MDS, >80% of patients have a mutation in a spliceosome gene [10–12]. In phenotypically overlapping syndromes of MDS/MPN such as CMML, spliceosome mutations are also common [10–12,17]. Mutations occur most frequently in *SF3B1* and *SRSF2*, with a lower prevalence in other spliceosomal genes, including *U2AF1*, *ZRSR2*, and *U2AF2* [17,18]. Spliceosomal mutations are thought to arise early in the course of disease, including as founder/initiating events [6]. However, how mutations in the spliceosome alter normal hematopoiesis and contribute to disease pathogenesis remains unclear. A key experimental system to understand the impact of the spliceosome mutations in regulating both normal and malignant hematopoiesis is the generation of high-fidelity murine models. There is a very high conservation of the human and mouse core spliceosome proteins, including those mutated in MDS and CMML. For example, human SRSF2 and murine Srsf2 proteins are 100% sequence conserved.

The generation of conventional null alleles of RNA-splicing genes in the mouse has been useful in identifying key RNA substrates with splicing that is perturbed in their absence and in beginning to understand their normal physiological function [19–22]. Retroviral overexpression has been used to assess how the mutations present in patients may contribute to MDS development [10,22]. A caveat with both the loss-of-function alleles and retroviral overexpression of mutant cDNAs is that they fail to recapitulate the mutations as they occur in humans: heterozygous point mutations

expressed from their endogenous locus, not alleles resulting in protein deficiency. The recurrent finding that the spliceosome mutations are heterozygous indicates that retention of a wild-type (WT) copy is necessary for cells to survive, an interpretation supported by both genetic and pharmacologic evidence [23–26], and additionally that gene dosage and protein complex stoichiometry may be important in the pathogenicity of the mutations [27].

Over the last 3 years, murine models recapitulating point mutations identified in patients with MDS and CMML have been described for *Sf3b1* [28,29], *Srsf2* [30–32], and *U2af1* [33,34]. The initial *U2af1* mutation model reported was an inducible cDNA expressed from a heterologous locus, but will be discussed alongside the true knock-in alleles [33]. The phenotypes observed in these "humanized" models have ranged from very mild through to the development of MDS/MPN. The reasons for these variable phenotypes have been not been systematically outlined, but the targeting strategy itself and the analysis methods may contribute significantly to the understanding of these models. Here, we discuss in detail the murine models of these mutations developed to date, their phenotypes and how they compare to the phenotypes of humans with these mutations.

Murine models of RNA spliceosome mutations

U2AF1 *(previously known as* U2AF35*)*

U2 small nuclear RNA auxiliary factor 1 (U2AF1) binds to the AG nucleotides and polypyrimidine tract located near the 3′ splice site in the intronic sequence (Figure 1). It then recruits U2AF2 and the U2 major spliceosome complex to facilitate 3′ splicing site recognition [15,35]. U2AF1 mutations have been identified in lung, breast, colon, and various epithelial carcinomas [36]. In MDS, approximately 11% of patients have a *U2AF1* mutation, predominantly at serine 34 (S34F/S34Y) and glutamine 157 (Q157R/Q157R) [10,37]. One transgenic and two knock-in models have been developed to date.

***U2AF1*^{S34F} transgenic model (Shirai et al. [33]).** Shirai et al. [33] reported the generation of a transgenic model of *U2AF1* mutation. In this case, unlike the others that will be discussed, an inducible (tetO) human *U2AF1*^{S34F} or control *U2AF1*^{WT} cDNA were targeted to the Collagen1a1 (*Col1a1*) locus of *Rosa26*-M2rtTA ES cells [38]. Expression of the WT or S34F cDNAs was not regulated in the manner of the endogenous *U2af1* locus. To activate the expression of the *U2AF1*^{S34F} or *U2AF1*^{WT} cDNAs, animals were administered doxycycline, enabling dose-dependent induction of expression of the human *U2AF1* cDNAs in the presence of two endogenous wild-type gene copies derived from the mouse (Figure 2).

No analysis was reported of the effects of *U2AF1*^{S34F} on native hematopoiesis (i.e., in animals treated with doxycycline in the absence of prior bone marrow transplantation). The only data presented were from irradiated transplantation recipients reconstituted with whole BM of *U2AF1*^{S34F} M2rtTA or *U2AF1*^{WT} M2rtTA donors. Once reconstitution was established, the recipient animals were administered doxycycline [33]. A similar approach was chosen in a second report of this model [26]. The use of BM transplantation prior to gene induction restricts expression of the mutation to the reconstituted hematopoietic cells, of particular importance when using an inducer expressed from the widely expressed *Rosa26* locus or others such as *Mx1* that are active in non-hematopoietic cell populations [39–41]. A caveat of analysis in the post-transplantation setting is that BM transplantation is known to induce hematopoietic stress, permanently modifying the BM microenvironment [42], and to alter the clonal dynamics of hematopoietic stem cells (HSCs) [43].

Using this experimental approach, the authors showed that the doxycycline-treated *U2AF1*^{S34F} M2rtTA recipients developed a stable peripheral blood leukopenia without changes in red blood cell or platelet numbers for up to 12 months. The leukopenia was the result of reduced B cells and monocytes and was dependent on the continued expression of the mutant *U2AF1*. Within the BM, there was evidence of increased cell death, reduced B cells and monocytes, and increased neutrophils. There was an increase in the percentages of common myeloid progenitors and a subtle increase in the percentage of linage$^-$/c-Kit$^+$/Sca-1$^+$ (LKS) cells. In competitive transplantations, *U2AF1*^{S34F} cells were serially transplantable through to tertiary recipients, albeit with a reduced chimerism. With extended aging (up to 500 days after BM transplantation), the *U2AF1*^{S34F}-expressing mice were not reported to develop dysplasia, MDS, or acute myeloid leukemia (AML).

***U2af1*^{S34F} knock-in models.** Fei et al. [34] reported the generation of two conditional knock-in models of *U2af1*^{S34F} mutation targeted to the endogenous locus. The S34F mutation is encoded by exon 2 of the murine *U2af1* allele. Both models required Cre protein to induce expression of the mutant allele. Two different knock-in approaches were described, one using an inverted exon 2 (referred to in that study as *IES34F*) and the second using a minigene (referred to as *MGS34F*), which resulted in significant differences in the expression of the mutant allele (Figure 2). Both were tested as heterozygous alleles, one of which encoded endogenous wild-type *U2af1* and the second the mutant allele.

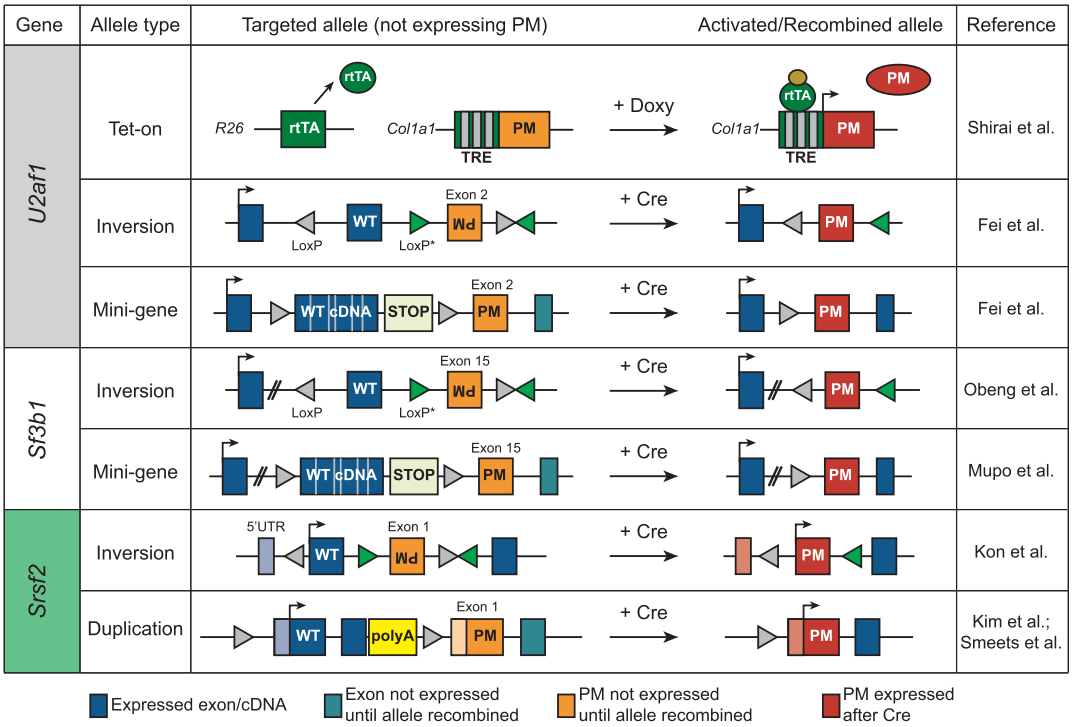


Figure 2. Simplified schematic outline of the currently described spliceosome mutant mouse models. LoxP elements are indicated as grey or green triangles indicating different types of loxP elements. Expressed exons are shown in blue, exons not expressed until after locus recombination are shown in teal, and mutation-bearing exons are shown in orange before recombination (silent) and in red when recombined and expressed.

***U2af1*^{S34F} "inverted exon" model (Fei et al. [34]).** For the inverted exon 2 model, a lox2272-flanked inverted S34F encoding exon 2 was placed in intron 2–3 and a conventional LoxP element was placed 5′ of the endogenous exon 2. Cre-induced recombination of the different LoxP sequences enabled removal of the endogenous exon 2 and replacement by an in-frame exon 2 encoding the S34F mutation. Analysis of mutant allele expression after Cre recombination in mouse embryonic fibroblasts demonstrated that the mutant mRNA was expressed at ~30% despite being genomically heterozygous. It is not clear why the *IES34F* allele failed to express the transcript heterozygously, although the authors suggested that it related to the efficiency of recombination of the locus. This allele was not further characterized [34].

***U2af1*^{S34F} "minigene" model (Fei et al. [34]).** The minigene model was examined in more detail [34]. In this model, a LoxP-flanked cDNA sequence encoding exons 2–8 followed by a transcriptional stop sequence was placed in the first intron. The S34F mutation was inserted in the endogenous exon 2. In the absence of Cre, the *U2af1* transcript from the targeted allele should be encoded by the endogenous exon 1 together with the cDNA encoding exons 2–8, with the stop cassette preventing read-through and expression from the S34F-encoding allele. This model was tested in the setting of native hematopoiesis using the broadly expressed *Mx1*-Cre and polyI:polyC (pI:pC) induction. The analysis of recombination following pI:pC administration indicated that the minigene, unlike the inverted exon model, was efficiently recombined and truly heterozygous both genomically and transcriptionally. Analysis of the *U2af1*$^{S34F/+}$ animals after pI:pC demonstrated a mild persistent macrocytic anemia and a twofold reduction in leukocytes, particularly B cells, in the peripheral blood. Within the BM, total cellularity was preserved, but there were reductions in the percentages of LKS^{+} and LKS^{+}CD48^{-}CD150^{+} phenotypic HSCs. Lineage distribution of more committed populations within the BM was not described. Morphological analysis of the BM revealed that less than 1% of cells displayed dysplastic features and no dysplastic cells were reported in the peripheral blood. No MDS or

AML was observed in aged *U2af1*$^{S34F/+}$ mice. The phenotypes described for native hematopoiesis were preserved, albeit exacerbated, by BM transplantation where expression of *U2af1*$^{S34F/+}$ was activated after transplantation, demonstrating a cell-autonomous effect of the mutant allele on hematopoiesis. Interestingly, and consistent across the splicing mutation models reported to date, there was a significantly reduced repopulating potential of U2af1(S34F)-expressing cells when BM cells from donors previously treated with pI:pC competed against WT BM cells.

SF3B1. Splicing factor 3b subunit 1 (SF3B1) is part of the SF3b protein complex that recognizes the branch point adenosine (A) base within the intron and facilitates the binding of U2 snRNP on pre-mRNA (Figure 1). *SF3B1* mutations have been identified in MDS, chronic lymphocytic leukemia and uveal melanoma. Mutations cluster between exons 14 and 16, with K700E being the most frequent mutation [10,11]. In MDS, *SF3B1* mutations occur in >80% of refractory anemia with ring sideroblasts, an MDS subtype with a favorable outcome [11,44]. Two independently generated *Sf3b1* conditional knock-in models have been reported and both have utilized the *Mx1*-Cre system to induce the expression of the mutant allele.

Sf3b1^{K700E} *knock-in models*

***Sf3b1*^{K700E} "inverted exon" model (Obeng et al. [28]).** A conditional *Sf3b1*^{K700E} model was reported by Obeng et al. [28] and was generated using an inverted exon strategy (Figure 2). In this model, an exon 15 with the A>G substitution and a WT exon 16 were introduced in an inverted orientation within introns 16–17. Upon Cre induction, the inverted exons were recombined in the correct orientation in place of the endogenous WT exons through the use of different loxP sequences and, in theory, should be physiologically expressed from the endogenous locus. Although the *Sf3b1*^{K700E} allele was successfully inverted after Cre induction (with pI:pC treatment), it did not achieve heterozygous expression when the RNA levels were assessed. The authors presented results of RNA sequencing (RNA-Seq) of lin$^-$cKit$^+$ cells from the BM of three *Sf3b1*$^{K700E/+}$ animals, where the mutant allele frequency was between 27% and 32% (see Figure S1A in Obeng et al. [28] and (see Supp Figure 1E-F in ref [23])). A similar level of subheterozygous mutant transcript expression was detected in a second recent report using this allele [24] and the phenotypes reported for these animals need to be considered with this caveat.

Monitoring of native hematopoiesis over 64 weeks of pI:pC-treated cohorts showed a progressive macrocytic anemia with decreased red blood cell counts from 20 weeks after activation of the K700E mutation. Corresponding with anemia, there was an increased level of plasma erythropoietin in *Sf3b1*$^{K700E/+}$ mice. This anemic phenotype was shown to be cell intrinsic through use of noncompetitive BM transplantation assays. There was no change in white blood cell or platelet counts. None of the *Sf3b1*$^{K700E/+}$ mice was reported to develop MDS during the period of monitoring. At 64 weeks after pI:pC, analysis of the BM showed no change in cellularity, whereas there was a significant increase in the number of long-term HSCs (LKS$^+$CD150$^+$CD48$^-$) and a decrease in the granulo-monocytic progenitors (GMPs). In the spleen, there was an accumulation of erythroblast populations (Ter119$^+$CD71^{high}) and a decrease in the more mature erythroid population (Ter119$^+$CD71^{low}), indicating a terminal erythroid maturation defect. Interestingly, this effect was only reproduced in vivo in young mice (11 weeks after pI:pC) under stress (drug-induced hemolysis) or in vitro when BM progenitor cells (24 weeks after pI:pC) were treated with cytokines to induce erythroid differentiation. Morphological examination of the spleen showed an increased number of erythroid precursors and erythroid dysplasia. However, ring sideroblasts or elevated levels of iron deposition were not observed. There was a severe impairment (<20%) in the competitive repopulation capacity of *Sf3b1*$^{K700E/+}$ cells (similar to Mupo et al. [29]), even though >95% chimerism was achieved in noncompetitive transplantation experiments.

***Sf3b1*^{K700E} "minigene" model (Mupo et al. [29]).** Mupo et al. [29] generated a conditionally activatable *Sf3b1*$^{K700E/+}$ allele using a mini-gene approach (Figure 2). They engineered a loxP-flanked construct containing a splice acceptor site (from an Engrail-2 splice acceptor sequence), exons 12–15, intron 15, exons 15–19 (exons 12–19 were codon optimized sequence), intron 19, and the naturally occurring sequences for exons 20–25 including the 3′-untranslated region (UTR), followed by a SV40 polyadenylation signal (SV40 pA). This mini-gene was inserted between exons 11 and 12 of *Sf3b1* and the endogenous exon 15 downstream of the mini-gene was replaced with a mutated, synthetic exon 15 harboring an A>G mutation encoding K700E. Upon expression of Cre protein from *Mx1*-Cre induced by administration of pI:pC, the mine-gene was deleted, allowing the expression of the *Sf3b1* allele containing the mutated exon 15 (*Sf3b1*$^{K700E/+}$).

The allele was assessed in a heterozygous setting with one mutant allele and one WT allele, mirroring the patient setting. RNA-Seq analysis confirmed the 50% expression level of the mutated allele in whole BM and lineage-negative (lin$^-$) cells from *Sf3b1*$^{K700E/+}$ mice. Four weeks pI:pC treatment native hematopoiesis was assessed and a decrease in the percentage of phenotypic HSCs (LKS$^+$CD34$^-$Flk2$^-$) was observed, but no change in the percentage of lymphoid-primed multipotent progenitors (L-MPPs), common myeloid progenitors (CMPs), GMPs, or myeloid-erythroid progenitors (MEPs). BM analysis indicated a mild myeloid bias

(Gr-1$^+$Mac-1$^+$) and impaired terminal erythroid differentiation (Ter119$^+$CD71^{low}FSC^{low}). There was no overall survival difference between *Sf3b1*$^{K700E/+}$ mice and WT littermates followed longitudinally for up to 83 weeks. Peripheral blood analysis revealed a progressive normocytic anemia with no change in white blood cell or platelet counts. None of the *Sf3b1*$^{K700E/+}$ mice was reported to develop MDS during monitoring. Morphological analysis of the BM did not find evidence of either dysplasia or ringed sideroblasts (a feature of human *SF3B1* mutant MDS), despite increased iron deposits in BM macrophages. Peripheral blood morphology was not reported. Competitive BM transplantation was performed using both young (2 months) and old (12 months) recipients to determine the potential influence of aging recipients on donor cell engraftment. In both groups, *Sf3b1*$^{K700E/+}$ cells showed poor chimerism, indicating impaired engraftment/repopulating potential. There was no difference between the young and old recipient groups with neither developing overt symptoms of MDS.

SRSF2. Serine-arginine rich factor 2 (SRSF2) binds to exonic splicing enhancers within the exonic sequence located near the 3′ splice site and mediates exon inclusion and exclusion (Figure 1). Unlike *U2AF1* or *SF3B1*, *SRSF2* mutations occur near exclusively in myeloid malignancies and at or encompassing proline 95 (P95), with the most frequent mutation reported a proline to histidine substitution (P95H) [10,12,17]. *SRSF2* mutations occur in approximately 12–15% of MDS [9] and 28–52% of CMML [45–47] cases. Mutations are associated with poor outcomes, a shortened time to leukemic conversion, and a myelo-monocytic bias [48–50]. Three independently generated models of *Srsf2*^{P95H} have now been reported and these will be discussed in greater detail than those of *U2af1* and *Sf3b1* given our own work on this gene.

Srsf2^{P95H} *knock-in models*

***Srsf2*^{P95H} "gene duplication" model (Kim et al. [31]).** Kim et al. [31] reported the first *Srsf2*^{P95H} conditional knock-in model. The allele described approximated a gene duplication strategy (Figure 2). The targeting construct inserted a loxP site 87 base pairs (bp) upstream of the 5′-UTR sequence of the endogenous exon 1. The sequence containing exon 3 (3′-UTR) and an introduced SV40 polyadenylation sequence (SV40pA) were placed immediately after the endogenous exon 2 to generate an *Srsf2* allele that would express a WT protein. The second loxP was placed after the SV40pA. Downstream of the second loxP, the sequence of exon 1 containing the *Srsf2*P95H mutation (CG>AC) and the sequence for exon 2 were introduced into introns 2–3 (Figure 3). Very little data were presented describing the phenotypic effects of this *Srsf2*$^{P95H/+}$ model on native hematopoiesis (i.e., not from transplantation recipients). The authors stated that none of the *Srsf2*$^{P95H/+}$ mice developed AML up to 70 weeks of monitoring, although it is not certain if this refers to de novo-mutated animals or transplantation recipients.

The authors principally reported analysis from noncompetitive BM transplantation assays for the remaining studies. Subsequent studies with this *Srsf2* allele from the same authors have also used this approach [23,24]. At 18 weeks after pI:pC, the *Srsf2*$^{P95H/+}$ recipients developed macrocytic anemia and leukopenia due to decreased B-cell numbers, whereas the platelet numbers remained normal. There was evidence of myeloid (hypolobated and hypogranulated neutrophils) and erythroid dysplasia in the peripheral blood/BM, but the estimated percentage of affected cells was not reported. Within the BM of the recipients, Kim et al. [31] reported that a major effect of the *Srsf2*$^{P95H/+}$ mutation was an expansion of the HPC-2 phenotype (LKS$^+$CD150$^+$CD48$^+$) within the hematopoietic stem and progenitor population. The frequencies of the HPC-2 population as reported by Kim et al. [31] in the control *Mx1*-Cre *Srsf2*$^{+/+}$ animals was 0.0231%, significantly higher than that reported in the index reference for this population of 0.0032 ± 0.0014% in WT mice [51]. The frequency of this population may have been affected by both the BM transplantation model and pI:pC use. However, the ~10-fold difference in reported frequency was not reconciled and the representative flow cytometry gating strategy presented by Kim et al. [31] is significantly different from the gating strategy in the original report of this population [51]. The authors reported an increased frequency of cells in S-phase and early apoptosis within the BM. In a subsequent report from the same authors using this *Srsf2*$^{P95H/+}$ allele and the *Mx1*-Cre model, the leukopenia and macrocytic anemia were largely absent (see Supp Fig 1E-F [23]). The inconsistency of the phenotype reported across the studies has not been addressed, but potentially indicates that the phenotype is not fully penetrant. Through analysis of RNA-Seq from both human patients and murine samples, the authors identified a group of mis-spliced candidate genes, with *EZH2* proposed as a key mis-spliced gene. *SRSF2*^{P95H} was proposed to lead to the inclusion of a premature stop codon in *EZH2*, resulting in a putative loss-of-function allele. This analysis posited a model that a single or small number of key mis-spliced genes may account for the phenotypes in the *SRSF2*^{P95H} mutant cells.

***Srsf2*^{P95H} "inverted exon" model (Kon et al. [32]).** Kon et al. [32] reported an *Srsf2*^{P95H} conditional knock-in model developed using an inverted exon strategy (termed FLEx switch inversion technique; Figure 2). For the targeted allele, the 5′-UTR and exon

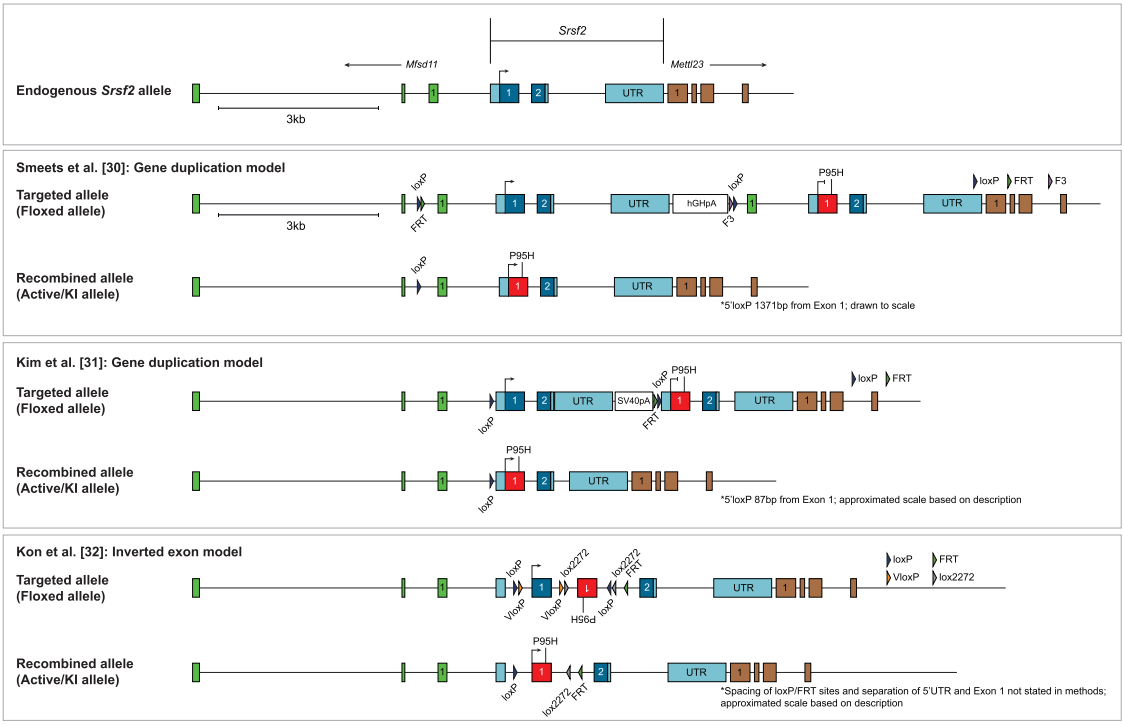


Figure 3. Detailed genomic architecture of the *Srsf2*(P95H) knock-in mutant models. **(A)** The endogenous *Srsf2* locus. Figure to scale provided. **(B)** The *Srsf2* allele described by Smeets el al. [30]. This allele is drawn to scale with the spacing of the inserted sequences as shown. **(C)** An approximation of the allele generaterd by Kim et al. [31] based on the description in the text. The 5′-loxP is 87 bp upstream of exon 1. The spacing and position of the remaining sequences is approximate. **(D)** An approximation of the allele generated by Kon et al. [32] using an inverted exon approach. The spacing and positions of the inserted sequences are approximations based on the description of the targeted allele. For **(C)** and **(D)**, the modified locus depiction is approximate and based on the descriptions available in the studies reporting these alleles.

1 were separated. Exon 1 was flanked by loxP/VloxP sites at the 5′ end and VloxP/lox2272 3′ to exon 1. An inverted exon 1 containing the P95H mutation was inserted downstream of the VloxP/lox2272 in intro 1, followed by second set of loxP/lox2272 sites. After Cre activation, the inverted exon containing P95H "flips" into the correct orientation and the endogenous exon 1 is excised at the same time (Figure 3). RNA-Seq showed no expression from the modified allele in Cre-negative heterozygous animals, indicating that the floxed allele was likely null when not recombined. This was consistent with the failure to obtain animals that had the non-recombined P95H allele homozygosed even in a Cre-negative background [32]. Why this is the case is unclear at present. The targeting strategy itself may be the most significant contributor in this particular instance. The separation of the 5′-UTR from the first coding exon, in effect exonization of the UTR, and the numerous modifications around the coding exon and the sequence insertion within the first intron may have had a significant effect on endogenous gene regulation.

The *Srsf2*^{P95H} conditional knock-in mice were crossed to *Vav1*-Cre mice to assess the effects of the heterozygous knock-in allele on hematopoiesis. Unlike the other described models, *Vav1*-Cre is a constitutive Cre and is active in the hematopoietic cells without the addition of an inducer such as pI:pC/IFNα (*Mx1*-Cre) or tamoxifen (CreER models). At approximately 15 weeks of age, the *Srsf2*$^{P95H/+}$ mice developed macrocytic anemia, which remained stable. After observation up to 90 weeks of age, none of the *Srsf2*$^{P95H/+}$ mice developed MDS or leukemia. A caveat with this model, as was described for the previous models describing inverted exon-based targeting methods, was that the mutant allele was expressed at 27–35% (mean 31%) based on RNA-Seq analysis of the *Vav1*-Cre *Srsf2*^{P95H} mice.

Upon analysis of the BM under native conditions, there was a reduction in the number of long-term repopulating

HSCs (LT-HSCs; LKS$^+$CD150$^+$CD48$^-$), MPP (LKS$^+$CD150$^-$CD48$^-$) and HPC-1 (LKS$^+$CD150$^-$CD48$^+$) populations, which are associated with increased cell cycling. There was no change in the cellularity or lineage distribution of BM and spleen, with the exception of a mild B-cell differentiation defect in the BM. There was no evidence of dysplasia reported in the BM under native settings. The authors then performed noncompetitive and competitive BM transplantation assays of the *Srsf2*$^{P95H/+}$ cells. There was engraftment in the noncompetitive transplantation setting, but the recipients of competitive BM transplantations had a significant reduction in chimerism when either whole BM or HSCs (LKS$^+$CD34$^-$) was used as the donor population. The impaired competitive transplantation potential is a common feature of all of the conditional splicing mutants to date. Extensive work demonstrated an engraftment defect of the *Srsf2*$^{P95H/+}$ HSCs. The *Srsf2*$^{P95H/+}$ recipients developed macrocytic anemia and leukopenia. In the BM, there was a significant reduction in LT-HSCs and most of the progenitor populations (MPPs, HPC-1, CMPs, CLPs). In addition, *Srsf2*$^{P95H/+}$ recipients had a significantly higher proportion of cycling and apoptotic HSCs than controls. The lineage distribution was skewed toward myeloid at the expense of B-cell differentiation and dysplastic erythroid cells and dys-megakaryopoiesis were evident in the BM.

Analysis of RNA splicing using multiple purified hematopoietic populations from both native hematopoiesis and cells isolated from transplantation recipients was described. From these analyses, the authors could identify mis-splicing of a number of genes that had been identified in human *SRSF2* mutant patient samples, including *Csf3r*, *Gnas*, *Hnrnpa2b1*, and several novel differentially spliced genes such as *Atrx* and *Mllt10*, both of which are implicated in hematological malignancies. Despite using numerous approaches, the authors could not find evidence to support *Ezh2* as a differentially spliced candidate in the mouse. It was proposed that this may be due to species differences, with human *EZH2* having two CCNG motifs, whereas the mouse has a single CCNG motif [32].

***Srsf2*^{P95H} "gene duplication" model (Smeets et al. [30]).** Recently, we reported a third conditional *Srsf2*P95H knock-in model [30] using a gene duplication method (Figure 2). The endogenous *Srsf2* locus including *Mfsd11* exon 1 and an introduced human growth hormone polyadenylation signal (hGHpA) was flanked by LoxP sites. The hGHpA was inserted downstream of the endogenous *Srsf2* 3′-UTR to prevent/reduce transcriptional read-through into the mutant allele in the absence of Cre. A duplication of the entire *Srsf2* locus and *Mfsd11* exon 1 was inserted downstream of the 3′-loxP sequence. The proline 95 (CCG) to histidine (CAT) mutation was introduced at the duplicated *Srsf2* exon 1 (Figure 3). Upon Cre-mediated excision, the endogenous *Srsf2* locus was deleted and the duplicated locus containing the P95H mutation was retained. RNA-Seq confirmed the heterozygous expression at both the genomic and transcriptional level. Analysis of Cre-negative animals and animals not treated with tamoxifen (the CreER inducer used in these studies) demonstrated a low level of transcriptional read-through in the absence of activation of the P95H bearing allele. Long-term monitoring of Cre$^-$ and non-tamoxifen-treated/recombined Cre$^+$ animals demonstrated that this "leaky" expression was not pathogenic.

Multiple Cre lines were applied to understand the cell population in which P95H was required to be able to affect normal hematopoiesis: whole body/broadly expressed (*R26*-CreER^{T2} [52]), more specific to the HSC and primitive progenitor populations (h*Scl*CreER^{T} $^{Tg/+}$ [53]), and a myeloid progenitor-targeted constitutive Cre line (*Lys*M-Cre$^{Ki/+}$ [54]). We analyzed the effect of *Srsf2*$^{P95H/+}$ on native hematopoiesis at 20 weeks after activation of the P95H mutation. In all cases, Cre$^+$ WT animals were used as controls and, for CreER models, these were tamoxifen-treated CreER$^+$ Srsf2 WT littermate controls. At 20 weeks after *Srsf2*$^{P95H/+}$ activation, *Srsf2*$^{P95H/+}$ mice developed macrocytic anemia and increased myeloid cells in the peripheral blood. The myeloid bias was more evident in the BM and was accompanied by compromised B lymphopoiesis. There was a reduction in erythropoiesis in the BM accompanied by increased splenic erythropoiesis. Within the phenotypic stem and progenitor populations, there was a reduction or trend to reduction in the number of stem cell populations using two different phenotypic methods (LKS$^+$CD34/Flt3 [55–57] or LKS$^+$CD105/CD150 [58,59]), whereas the more mature myeloid progenitor populations remained largely unchanged. The level of CD45RB, a previously characterized splicing target of *Srsf2* [21], was reduced in *Srsf2*$^{P95H/+}$ splenocytes, indicating that altered RNA splicing was reflected in the proteome. A similar phenotype was seen in both *R26*-CreER^{T2} and h*Scl*CreER^{T} $^{Tg/+}$ models, but not in the *Lys*M-Cre$^{Ki/+}$ model, indicating that the *Srsf2*$^{P95H/+}$ mutation needed to arise within the primitive populations (including HSCs) to modify native hematopoiesis. RNA-Seq analysis of lin$^-$cKit$^+$eYFP$^+$ cells (where enhanced yellow fluorescent protein [eYFP] marked cells and their daughters/progeny in which Cre was activated) isolated 20 weeks after activation of P95H demonstrated that the *Srsf2*$^{P95H/+}$ mutation induced gene expression changes consistent with myeloid bias, loss of lymphoid potential, and transcriptional signatures found in MDS. We identified similarly mis-spliced transcripts as Kon et al. [32] and also could not define mis-splicing of *Ezh2* in the murine setting.

In agreement with the previous knock-in models of spliceosome mutations, the *Srsf2*$^{P95H/+}$ cells exhibited a poor competitive engraftment potential. This appears to be a universal feature of the models described to date, with the work of Kon et al.[32] best characterizing this phenotype and demonstrating an engraftment defect in the mutant cells. This experimental result is confounding because *SRSF2* mutations are subjected to positive selection and are implicated in age-related clonal hematopoiesis [7,8,14]. However, we observed that the poor engraftment could be modified by altering the nature of the BM competitor used. Using the h*Scl*CreER$^{T\ Tg/+}$ model, we found that *Srsf2*$^{P95H/+}$ cells could competitively engraft recipient animals and expand when transplanted with the age-/microenvironment-matched competitor cells (i.e., cells taken from the same BM but in which Cre was not activated based on *R26*-eYFP reporter marking of the cells).

Upon aging, nontransplanted *Srsf2*$^{P95H/+}$ mice developed fatal MDS by ~12 months after activating the P95H mutation in both the *R26*-CreER^{T2} and h*Scl*CreER$^{T\ Tg/+}$ model. The disease in both models was highly comparable. When moribund, the mice presented with macrocytic anemia, myeloid bias (granulocytosis and monocytosis), and morphological dysplasia of myeloid and erythroid lineages in both the peripheral blood (>10–50% of cells) and BM, all characteristics of MDS/MPN. Analysis of a small cohort by exome capture demonstrated that there was a subclonal accumulation of mutations associated with human *SRSF2* mutant disease, including *Dnmt3a*, *Tet2*, *Phf6*, and *Ras* members, in the BM of the moribund animals [30,60]. Noncompetitive transplantation of the moribund BM recapitulated the MDS/MPN, but there was no evidence of progression to acute leukemia even with long-term monitoring of secondary recipients or with concurrent p53 deletion. Of the currently described *Srsf2* point mutant models, this is the only model to develop monocytosis and MDS/MPN in the setting of native hematopoiesis.

Discussion

Our understanding of the effects of spliceosome mutations comes from the detailed analysis of the mutational architecture of MDS, MDS/MPN, and related disorders in humans. In these settings, the mutation generally arises together with other mutations that combinatorially contribute to the disease manifestations and phenotypes [60]. In most cases, RNA-splicing mutations are not the sole causes of maliganancies because other driver mutations are present and familial monogenic examples of RNA splicing mutations have not been described. However, isolated mutations in RNA splicing components can sometimes cause disease directly, such as in ~20% of *SF3B1*-mutant MDS in which other driver mutations were not reported [17]. In the setting of CHIP, mutations in *SRSF2* and *SF3B1* have been identified with a range of variant allele burden, including VAFs of >10% [7,8,61]. These observations would suggest that, when present as the only driver mutation, spliceosomal mutations can confer advantages to the host cells. The development of murine models of human spliceosome mutations, both conventional knock-out models and humanized mutant models, has improved our understanding of how these mutations perturb normal hematopoiesis and ultimately promote malignancy. The establishment of tractable, autochthonous preclinical models has proven more challenging. The fidelity of the "humanized" murine models must be considered against the phenotypes associated with the same spliceosomal mutations in humans.

The independently generated "humanized" spliceosomal mutant mice described to date encompass a spectrum of phenotypes from very mild to overt MDS/MPN in the setting of native hematopoiesis (summarized in Table 1). The majority of models have also reported analysis of BM transplantation recipients. This is a useful strategy to limit the expression of the mutation to the transplantable hematopoietic cells, but is also a source of significant cellular stress. MDS, MDS/MPN, and related cancers arise in the absence of this extreme stress in the vast majority of humans. BM transplantation changes HSC clonal dynamics and contribution compared with native hematopoiesis and permanently damages and remodels the BM microenvironment itself [42,43,62–65]. The use of transplantation models also limits the capacity to understand the interplay of the mutant cells and a normal immune system, now appreciated to be an important aspect in cancer evolution and escape [66].

The models described demonstrate that the targeting strategy itself can have profound impacts on the utility of the model that is generated. The three models that have utilized an inverted exon approach have all observed ~30% variant allele frequency at the RNA level after locus recombination, demonstrating that, despite genomic heterozygosity, the expression from these modified alleles is not heterozygous. This caveat requires acknowledgement when these models are reflected against the human phenotypes associated with these mutations. The reasons that these alleles do not express heterozygously are not clear, but the results are consistent across three independent models and loci, indicating that this effect is most likely resultant of the targeting strategy itself. The inverted exon model has been successful in establishing murine models harboring the Jak2(V617F) mutation, in which the mice develop a myeloprofilerative neoplasm resembling polycythemia vera, as occurs in humans with this

Table 1. Summary of the phenotypes of the "humanized" spliceosomal mutant mice described

				Native hematopoiesis			Non-competitive transplantation			
Gene	Model	rtTA/Cre	VAF	Peripheral Blood/Spleen	BM	Dysplasia/Disease	Peripheral Blood/Spleen	BM	Dysplasia/Disease	Competitive Transplantation
U2af1	Tet-On (Shirai et al. [33])	*R26*-M2rtTA	NA[a]	Not described			Leukopenia; ↓B cells; ↓monocytes	↑CMP%;[b] ↑LKS$^+$%	No dysplasia, MDS, or AML up to 500 days post Tx	N/A
	Inversion (Fei et al. [34])	*UBC*-CreER^{T2}	~30%	MEFs only: not further characterized						
	Mini-gene (Fei et al. [34])	*Mx1*-Cre	~50%	Macrocytic anemia; leukopenia; ↓B cells	↓%LKS$^+$; ↓%LTHSCs	<1% in BM; 0% in PB; no MDS or AML	Similar to native hematopoiesis			↓↓Engraftment
Sf3b1	Inversion (Obeng et al. [28])	*Mx1*-Cre	~30%	Macrocytic anemia; ↓RBCs; ↑plasma EPO; ↑erythroblasts in spleen	↑LTHSCs; ↓GMPs	Erythroid dysplasia in spleen; no ring sideroblasts; no MDS (64 weeks post pI:pC)	Similar to native hematopoiesis			↓↓Engraftment
	Mini-gene (Mupo et al. [29])	*Mx1*-Cre	~50%	Normocytic anemia	↓HSC%; ↑granulocytes; ↓mature erythrocytes	No dysplasia/ring sideroblasts; no MDS (83 weeks post pI:pC)	N/A			↓↓Engraftment
Srsf2	Duplication (Kim et al. [31])	*Mx1*-Cre	~50%	Not described			Macrocytic anemia; leukopenia; ↓B cells	↑LKS; ↑LKS in S phase; ↑HPC2; ↓B cells	Neutrophil and erythroid dysplasia in PB; no AML (70 weeks)	↓↓Engraftment; ↑LKS; ↑HPC-2
	Inversion (Kon et al. [32])	*Vav1*-Cre	~31%	Macrocytic anemia	↓LTHSCs; ↓MPPs; ↓HPC-1	No dysplasia; no MDS or AML (90 weeks)	Macrocytic anemia; leukopenia; ↑granulocytes; ↓B cells	↓LTHSCs; ↓MPPs; ↓HPC1; ↓CMPs; ↓CLPs	Dysplastic erythroid and ↑megakaryocytes in BM; no MDS or AML	↓↓Engraftment; defect in homing
	Duplication (Smeets et al. [30])	*R26*-CreER^{T2}; h*Scl*-CreER^{T} $^{Tg/+}$; *Lys*M-Cre	~50%	Macrocytic anemia; ↑granulocytes; ↑monocytes	↓LTHSCs; ↑%granulocytes; ↓B cells; ↓erythroid cells	MDS/MPN by ~12 months; no AML; ↑ and dysplastic myeloid cells (10−50% PB); monocytosis in PB/transplantable to secondary recipients	Macrocytic anemia; leukopenia; ↑%granulocytes and monocytes; ↓B cells; ↑% monocytes in spleen	↑%granulocytes; ↑%monocytes	No AML	↓↓Engraftment in *R26*-CreER^{T2} model; myeloid bias; engraftment and ↑ mutant cells in h*Scl*/CreER^{T} $^{Tg/+}$ model

[a]The allele frequency is doxycycline dose dependent.

[b]The immunophenotypic population listed in this table is defined as following: LKS$^+$, linage$^-$/c-Kit$^+$/Sca-1$^+$; LKS$^-$, linage$^-$/c-Kit$^+$/Sca-1$^-$; LT-HSC, LKS$^+$/CD150$^+$/CD48$^-$; MPP, LKS$^+$/CD150$^-$/CD48$^-$; HPC-1, LKS$^+$/CD150$^-$/CD48$^+$; HPC-2, LKS$^+$/CD150$^+$/CD48$^+$; HSC (Mupo et al. [29]), LKS$^+$/CD34$^-$/flk2$^-$; pre-MegE, LKS$^-$/CD41$^-$/FcγR$^-$/CD150$^+$/CD105$^-$; pre-CFUE, LKS$^-$/CD41$^-$/FcγR$^-$/CD15^{-+}/CD105$^+$. %=percentage of this population.

mutation [67]. Although speculative, for alleles that provide a proliferative/survival advantage or are strongly positively selected, the inverted exon model may be feasible because even subheterozygous expression may be sufficient to afford the cell with the advantage of the mutant allele.

Another consideration is how foreign elements such as loxP sites are placed within the endogenous gene structure. Although literature describing the systematic assessment of locus modifications is limited, some principals have emerged with the evolution of the gene targeting methods and from the more recent high-throughput targeting vector generation methods [68–72]. These include the placing of loxP elements relative to the promoter and intron/exon junctions (>200 bp is generally advised); avoiding disruption of evolutionarily conserved regulatory elements and CpG islands that may be important for normal gene regulation; being cognizant of the intron and exon sizes when introducing new elements with caution when there are short introns/exons because this can affect splicing of the gene; and avoidance of modification to neighboring genes if possible. These generalized observations may be relevant to the alleles described for *Srsf2*, in which there are only two coding exons and the mutation occurs in the first exon. There is a CpG island that encompasses the 5′ promoter region and 5′-UTR/exon 1, intron and exon 2 of *Srsf2* (UCSC genome browser mm10; chr11:116,852,080-116,853,903). The *Srsf2* allele described by Kon et al. [32], for example, used an inverted exon approach and was found to be transcriptionally inactive/silent when non-recombined. This allele involved significant modifications to the gene structure, including separation of the 5′-UTR and exon 1 coding sequence and insertion of multiple recombination elements in the introns [32]. The model developed by Kim et al. [31] placed the 5′-loxP element 87 bp from the start of the exon 1-containing sequence and modified the intron spacing significantly in both the pre- and post-recombined configuration [31]. All of these modifications can potentially affect the expression of the locus.

In humans, it is thought that spliceosome mutations arise within the HSC compartment [6,73]. It is well known that the cell of origin can affect the phenotypes from experimental leukemia models [74,75], so the use of different Cre drivers that target distinct and/or overlapping cell populations may be important in the presentation of the phenotypes in these models. Two of the *Srsf2* and one of the *Sf3b1* alleles have been assessed using the constitutively active *Vav*-Cre [24,32,76]. *Vav*-Cre is active in the hematopoietic cells, including HSCs, with evidence for activity also in endothelial cells and some activity in the reproductive tissues [41,77,78]. Multiple groups have used the inducible *Mx1*-Cre, a broadly active Cre that is very efficient at recombining loxP-flanked alleles in hematopoietic cells, including HSCs and other tissues in the adult animal such as the cells composing the BM microenvironment [41,79,80]. *Mx1*-Cre requires administration of pI:pC to induce an innate immune-mediated interferon response, a known stimulus of cycling of HSCs [81,82]. Whereas this can be used to assess native hematopoiesis, the spliceosome mutant models described to date using *Mx1*-Cre have predominantly reported phenotypic/functional analysis from recipients of BM transplantations, which were treated after recovery with pI:pC. This approach is useful to restrict Cre activity in non-hematopoietic tissues because only the transplanted/engrafted BM will express Cre [39,83]. Such a strategy is effective for understanding the hematopoietic intrinsic effect of the mutation, but it introduces the stress of BM transplantation and changes the hematopoietic dynamics and BM microenvironment as a result [42,43,62–65]. Coupled with the administration of pI:pC 4–8 weeks after transplantation, this elicits substantive stress on the entire hematopoietic system. An alternative is to use CreER-based models in which tamoxifen is required to be administered to activate Cre activity. We utilized both the *Rosa26*-CreER ^{T2} and *hScl*-CreER^{T} lines to allow widespread and relatively HSC-restricted activation, respectively, of *Srsf2*^{P95H} [30]. The CreER systems and the administration of tamoxifen are not benign, with evidence that toxicity of the CreER itself and tamoxifen administered at high doses subcutaneously can have direct effects on hematopoiesis [84,85]. These effects can be reduced by oral administration of tamoxifen, shortening the duration of exposure to tamoxifen, or by the use of heterozygous CreER alleles. The contribution of the specific Cre drivers to the phenotypes reported are likely secondary to the impacts of the locus specific genetic modifications themselves. However, a direct comparison of the same splicing mutation with different Cre strains would resolve this (e.g., *Rosa26*-CreER vs. *Mx1*-Cre). In all situations, the parallel analysis of littermate Cre^{+} WT and Cre^{-} gene-modified animals is an essential control for these studies whether using *Mx1*-Cre or CreER systems.

The models of spliceosomal mutation described to date have demonstrated a range of phenotypes. Definitive reasons for this remain uncertain, with some interpretations citing a lack of conservation of splicing between species. Although the spliceosomal proteins themselves are highly evolutionarily conserved, this is not an insignificant consideration [16]. By virtue of the higher level of conservation of the exons between mouse and humans (89–94%), the exonic splice-enhancer sequences are more highly conserved [86,87]. Despite this, there is still uncertainty regarding the identity of disease relevant mis-spliced candidates in human

samples and murine models. For example, in humans, it was reported that missplicing of *EZH2* [31] and, more recently, *CASP8* [24,88] occurred in *SRSF2* mutant cells. For *Ezh2* at least, this observation was not confirmed in multiple murine *Srsf2*$^{P95H/+}$ mutant models in which other mis-spliced candidates identified in human samples could be confirmed [30,32]. The phenotypic recapitulation and conserved disease evolution and progression that occured in the *Srsf2*$^{P95H/+}$ model that we described indicates that the mouse is able to reproduce the core aspects of the human phenotypes associated with the *SRSF2* mutation. However, the identification of the specific mis-spliced genes that cause disease development has remained elusive. There are several possibilities that could account for this. First, transcriptomic analysis has yet to be performed on sufficiently purified disease initiating cells in both human and mouse and the heterogeneity of the samples assessed to date masks identification of the relevant mis-spliced transcipts. Compounding this is the fact that the computational analysis of splicing has yet to reach a consensus and the different computational methods yield distinct results. An alternative possibility is that subtle changes across many genes collectively contribute to the phenotype, rather than a small number of dramatically mis-spliced candidates. In such a possibility, it will require highly purified samples and sufficient sequencing depth to confidently identify these. Another possibility is that misspling occurs within common cellular pathways in both species, but that the underlying individual genes are more variable and not necessarily conserved. An example of species conserved transcriptional and cellular consequence despite an apparent divergence of the individual transcripts can be seen in the A-to-I RNA-editing field and in the phenoytpes of *ADAR1* mutation in human and mouse [89–91]. Unlike the exonic sequences bound by SRSF2, intronic sequences are more divergent between human and mouse, with 66.1% conservation of branchpoint sequences across mammals [92]. The coordinated recognition of the 5′ and 3′ splice sites, together with branch-point and polypyrimidine tract selection, are critical to splicing fidelity. The utilization of highly conserved branch points is likely to be conserved to a greater extent across species than those with weaker recognition motifs, less conservation, or alternative branch-point spacing. These differences in the intronic sequences and motifs between species may contribute to the phenotypic differences between mouse and humans, in particular for SF3B1, which binds the branch point within intronic sequence of the pre-mRNA.

Murine models of human disease-associated mutations allow us to understand how these mutations individually contribute to disease initiation and maintenance. To provide this understanding, they must recapitulate as closely as possible the nature of the mutation, the expression of the mutant allele, and the cell of origin. A critical assessment of the genetic modification strategies applied demonstrates that there are significant differences in the approaches that have been utilized. How the different targeting strategies and experimental approaches for individual gene mutations contribute to the phenotypes observed should be taken into consideration, along with the degree to which the model presents phenotypes that are consistent with the human disease carrying the same mutation.

Acknowledgments

This work was supported by grants from the Leukaemia Foundation (Grant-in-Aid to MW; PhD Scholarship to ST); the Cancer Council of Victoria (APP1126010 to CW and MW); the Victorian Cancer Agency (Research Fellowship MCRF15015 to CW); and the Victorian State Government Operational Infrastructure Support Scheme (OIS) (to SVI).

Author contribution statement

JX, MS, and CW wrote the original draft; JX, MS, ST, MW, LP, CW wrote, reviewed, and edited the paper.

Conflict of interest disclosure

The authors declare no competing financial interests.

References

1. Tefferi A, Vardiman JW. Myelodysplastic syndromes. N Engl J Med. 2009;361:1872–1885.
2. Tan SY, Smeets MF, Chalk AC, et al. Insights into myelodysplastic syndromes from current preclinical models. World J Hematol. 2016;5:1–22.
3. McQuilten ZK, Wood EM, Polizzotto MN, et al. Underestimation of myelodysplastic syndrome incidence by cancer registries: results from a population-based data linkage study. Cancer. 2014;120:1686–1694.
4. Cogle CR. Incidence and burden of the myelodysplastic syndromes. Curr Hematol Malig Rep.. 2015;10:272–281.
5. Cogle CR, Craig BM, Rollison DE, List AF. Incidence of the myelodysplastic syndromes using a novel claims-based algorithm: high number of uncaptured cases by cancer registries. Blood. 2011;117:7121–7125.
6. Sperling AS, Gibson CJ, Ebert BL. The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. Nat Rev Cancer. 2017;17:5–19.
7. Jaiswal S, Fontanillas P, Flannick J, et al. Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med. 2014;371:2488–2498.
8. Genovese G, Kähler AK, Handsaker RE, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med. 2014;371:2477–2487.
9. Gangat N, Patnaik MM, Tefferi A. Myelodysplastic syndromes: contemporary review and how we treat. Am J Hematol. 2016;91:76–89.
10. Yoshida K, Sanada M, Shiraishi Y, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature. 2011;478:64–69.
11. Papaemmanuil E, Cazzola M, Boultwood J, et al. Chronic Myeloid Disorders Working Group of the International

Cancer Genome Consortium. Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. N Engl J Med. 2011;365:1384–1395.
12. Meggendorfer M, Roller A, Haferlach T, et al. SRSF2 mutations in 275 cases with chronic myelomonocytic leukemia (CMML). Blood. 2012;120:3080–3088.
13. Itzykson R, Kosmider O, Renneville A, et al. Clonal architecture of chronic myelomonocytic leukemias. Blood. 2013;121:2186–2198.
14. Martincorena I, Raine KM, Gerstung M, et al. Universal patterns of selection in cancer and somatic tissues. Cell. 2017;171:1029–1041.
15. Wahl MC, Will CL, Lührmann R. The spliceosome: design principles of a dynamic RNP machine. Cell. 2009;136:701–718.
16. Lee Y, Rio DC. Mechanisms and regulation of alternative pre-mRNA splicing. Annu Rev Biochem. 2015;84:291–323.
17. Haferlach T, Nagata Y, Grossmann V, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia. 2014;28:241–247.
18. Papaemmanuil E, Gerstung M, Malcovati L, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood. 2013;122:3616–3627. quiz 3699.
19. Colla S, Ong DS, Ogoti Y, et al. Telomere dysfunction drives aberrant hematopoietic differentiation and myelodysplastic syndrome. Cancer Cell. 2015;27:644–657.
20. Visconte V, Tabarroki A, Zhang L, et al. Splicing factor 3b subunit 1 (Sf3b1) haploinsufficient mice display features of low risk myelodysplastic syndromes with ring sideroblasts. J Hematol Oncol. 2014;7:89.
21. Wang HY, Xu X, Ding JH, Bermingham Jr JR, Fu XD. SC35 plays a role in T cell development and alternative splicing of CD45. Mol Cell. 2001;7:331–342.
22. Komeno Y, Huang YJ, Qiu J, et al. SRSF2 is essential for hematopoiesis, and its myelodysplastic syndrome-related mutations dysregulate alternative pre-mRNA splicing. Mol Cell Biol. 2015;35:3071–3082.
23. Lee SC, Dvinge H, Kim E, et al. Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins. Nat Med. 2016;22:672–678.
24. Lee SC, North K, Kim E, et al. Synthetic lethal and convergent biological effects of cancer-associated spliceosomal gene mutations. Cancer Cell. 2018;34:225–241. e8.
25. Seiler M, Yoshimi A, Darman R, et al. H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers. Nat Med. 2018;24:497–504.
26. Shirai CL, White BS, Tripathi M, et al. Mutant U2AF1-expressing cells are sensitive to pharmacological modulation of the spliceosome. Nat Commun. 2017;8:14060.
27. Fei DL, Motowski H, Chatrikhi R, et al. Wild-type U2AF1 antagonizes the splicing program characteristic of U2AF1-mutant tumors and is required for cell survival. PLoS Genet. 2016;12:e1006384.
28. Obeng EA, Chappell RJ, Seiler M, et al. Physiologic expression of Sf3b1(K700E) causes impaired erythropoiesis, aberrant splicing, and sensitivity to therapeutic spliceosome modulation. Cancer Cell. 2016;30:404–417.
29. Mupo A, Seiler M, Sathiaseelan V, et al. Hemopoietic-specific Sf3b1-K700E knock-in mice display the splicing defect seen in human MDS but develop anemia without ring sideroblasts. Leukemia. 2017;31:720–727.
30. Smeets MF, Tan SY, Xu JJ, et al. Srsf2(P95H) initiates myeloid bias and myelodysplastic/myeloproliferative syndrome from hemopoietic stem cells. Blood. 2018;132:608–621.
31. Kim E, Ilagan JO, Liang Y, et al. SRSF2 mutations contribute to myelodysplasia by mutant-specific effects on exon recognition. Cancer Cell. 2015;27:617–630.
32. Kon A, Yamazaki S, Nannya Y, et al. Physiological Srsf2 P95H expression causes impaired hematopoietic stem cell functions and aberrant RNA splicing in mice. Blood. 2018;131:621–635.
33. Shirai CL, Ley JN, White BS, et al. Mutant U2AF1 expression alters hematopoiesis and pre-mRNA splicing in vivo. Cancer Cell. 2015;27:631–643.
34. Fei DL, Zhen T, Durham B, et al. Impaired hematopoiesis and leukemia development in mice with a conditional knock-in allele of a mutant splicing factor gene U2af1. Proc Natl Acad Sci U S A. 2018;115:E10437–E10446.
35. Wu S, Romfo CM, Nilsen TW, Green MR. Functional recognition of the 3′ splice site AG by the splicing factor U2AF35. Nature. 1999;402:832–835.
36. Kandoth C, McLellan MD, Vandin F, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013;502:333–339.
37. Graubert TA, Shen D, Ding L, et al. Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. Nat Genet. 2011;44:53–57.
38. Hochedlinger K, Yamada Y, Beard C, Jaenisch R. Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. Cell. 2005;121:465–477.
39. Walkley CR, Orkin SH. Rb is dispensable for self-renewal and multilineage differentiation of adult hematopoietic stem cells. Proc Natl Acad Sci U S A. 2006;103:9057–9062.
40. Purton LE, Scadden DT. Limiting factors in murine hematopoietic stem cell assays. Cell Stem Cell. 2007;1:263–270.
41. Joseph C, Quach JM, Walkley CR, Lane SW, Lo Celso C, Purton LE. Deciphering hematopoietic stem cells in their niches: a critical appraisal of genetic models, lineage tracing, and imaging strategies. Cell Stem Cell. 2013;13:520–533.
42. Quach JM, Askmyr M, Jovic T, et al. Myelosuppressive therapies significantly increase pro-inflammatory cytokines and directly cause bone loss. J Bone Miner Res. 2015;30:886–897.
43. Sun J, Ramos A, Chapman B, et al. Clonal dynamics of native haematopoiesis. Nature. 2014;514:322–327.
44. Patnaik MM, Tefferi A. Refractory anemia with ring sideroblasts (RARS) and RARS with thrombocytosis (RARS-T): 2017 update on diagnosis, risk-stratification, and management. Am J Hematol. 2017;92:297–310.
45. Mason CC, Khorashad JS, Tantravahi SK, et al. Age-related mutations and chronic myelomonocytic leukemia. Leukemia. 2016;30:906–913.
46. Patnaik MM, Lasho TL, Finke CM, et al. Spliceosome mutations involving SRSF2, SF3B1, and U2AF35 in chronic myelomonocytic leukemia: prevalence, clinical correlates, and prognostic relevance. Am J Hematol. 2013;88:201–206.
47. Patnaik MM, Parikh SA, Hanson CA, Tefferi A. Chronic myelomonocytic leukaemia: a concise clinical and pathophysiological review. Br J Haematol. 2014;165:273–286.
48. Thol F, Kade S, Schlarmann C, et al. Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. Blood. 2012;119:3578–3584.
49. Wu SJ, Kuo YY, Hou HA, et al. The clinical implication of SRSF2 mutation in patients with myelodysplastic syndrome and its stability during disease evolution. Blood. 2012;120:3106–3111.
50. Mian SA, Smith AE, Kulasekararaj AG, et al. Spliceosome mutations exhibit specific associations with epigenetic modifiers and proto-oncogenes mutated in myelodysplastic syndrome. Haematologica. 2013;98:1058–1066.
51. Oguro H, Ding L, Morrison SJ. SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. Cell Stem Cell. 2013;13:102–116.

52. Ventura A, Kirsch DG, McLaughlin ME, et al. Restoration of p53 function leads to tumour regression in vivo. Nature. 2007;445:661–665.
53. Göthert JR, Gustin SE, Hall MA, et al. In vivo fate-tracing studies using the Scl stem cell enhancer: embryonic hematopoietic stem cells significantly contribute to adult hematopoiesis. Blood. 2005;105:2724–2732.
54. Clausen BE, Burkhardt C, Reith W, Renkawitz R, Förster I. Conditional gene targeting in macrophages and granulocytes using LysMcre mice. Transgenic Res. 1999;8:265–277.
55. Osawa M, Hanada K, Hamada H, Nakauchi H. Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. Science. 1996;273:242–245.
56. Adolfsson J, Månsson R, Buza-Vidas N, et al. Identification of $Flt3^+$ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. Cell. 2005;121:295–306.
57. Yang L, Bryder D, Adolfsson J, et al. Identification of Lin-$Sca1^+kit^+CD34^+Flt3$- short-term hematopoietic stem cells capable of rapidly reconstituting and rescuing myeloablated recipients. Blood. 2005;105:2717–2723.
58. Pronk CJ, Rossi DJ, Månsson R, et al. Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. Cell Stem Cell. 2007;1:428–442.
59. Kiel MJ, Yilmaz OH, Iwashita T, Yilmaz OH, Terhorst C, Morrison SJ. SLAM family receptors distinguish hematopoietic stem and progrnitor cells and reveal endothelial niches for stem cells. Cell. 2005;121:1109–1121.
60. Makishima H, Yoshizato T, Yoshida K, et al. Dynamics of clonal evolution in myelodysplastic syndromes. Nat Genet. 2017;49:204–212.
61. McKerrell T, Park N, Moreno T, et al. Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. Cell Rep. 2015;10:1239–1245.
62. Rodriguez-Fraticelli AE, Wolock SL, Weinreb CS, et al. Clonal analysis of lineage fate in native haematopoiesis. Nature. 2018;553:212–216.
63. Busch K, Klapproth K, Barile M, et al. Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. Nature. 2015;518:542–546.
64. Henninger J, Santoso B, Hans S, et al. Clonal fate mapping quantifies the number of haematopoietic stem cells that arise during development. Nat Cell Biol. 2017;19:17–27.
65. Askmyr M, Quach J, Purton LE. Effects of the bone marrow microenvironment on hematopoietic malignancy. Bone. 2011;48:115–120.
66. Barreyro L, Chlon TM, Starczynowski DT. Chronic immune response dysregulation in MDS pathogenesis. Blood. 2018;132:1553–1560.
67. Mullally A, Lane SW, Ball B, et al. Physiological Jak2V617F expression causes a lethal myeloproliferative neoplasm with differential effects on hematopoietic stem and progenitor cells. Cancer Cell. 2010;17:584–596.
68. Economides AN, Frendewey D, Yang P, et al. Conditionals by inversion provide a universal method for the generation of conditional alleles. Proc Natl Acad Sci U S A. 2013;110:E3179–E3188.
69. Wassef M, Luscan A, Battistella A, et al. Versatile and precise gene-targeting strategies for functional studies in mammalian cell lines. Methods. 2017;121–122:45–54.
70. Lewandoski M. Conditional control of gene expression in the mouse. Nat Rev Genet. 2001;2:743–755.
71. Hall B, Limaye A, Kulkarni AB. Overview: generation of gene knockout mice. Curr Protoc Cell Biol. 2009. Chapter 19:Unit 19.12.1–19.12.17.
72. Friedel RH, Wurst W, Wefers B, Kühn R. Generating conditional knockout mice. Methods Mol Biol. 2011;693:205–231.
73. Woll PS, Kjällquist U, Chowdhury O, et al. Myelodysplastic syndromes are propagated by rare and distinct human cancer stem cells in vivo. Cancer Cell. 2014;25:794–808.
74. Huntly BJ, Shigematsu H, Deguchi K, et al. MOZ-TIF2, but not BCR-ABL, confers properties of leukemic stem cells to committed murine hematopoietic progenitors. Cancer Cell. 2004;6:587–596.
75. Krivtsov AV, Twomey D, Feng Z, et al. Transformation from committed progenitor to leukaemia stem cell initiated by MLL-AF9. Nature. 2006;442:818–822.
76. de Boer J, Williams A, Skavdis G, et al. Transgenic mice with hematopoietic and lymphoid specific expression of Cre. Eur J Immunol. 2003;33:314–325.
77. Georgiades P, Ogilvy S, Duval H, et al. VavCre transgenic mice: a tool for mutagenesis in hematopoietic and endothelial lineages. Genesis. 2002;34:251–256.
78. Croker BA, Metcalf D, Robb L, et al. SOCS3 is a critical physiological negative regulator of G-CSF signaling and emergency granulopoiesis. Immunity. 2004;20:153–165.
79. Kühn R, Schwenk F, Aguet M, Rajewsky K. Inducible gene targeting in mice. Science. 1995;269:1427–1429.
80. Velasco-Hernandez T, Säwén P, Bryder D, Cammenga J. Potential pitfalls of the Mx1-Cre system: implications for experimental modeling of normal and malignant hematopoiesis. Stem Cell Reports. 2016;7:11–18.
81. Essers MA, Offner S, Blanco-Bose WE, et al. IFNalpha activates dormant haematopoietic stem cells in vivo. Nature. 2009;458:904–908.
82. Baldridge MT, King KY, Boles NC, Weksberg DC, Goodell MA. Quiescent haematopoietic stem cells are activated by IFN-gamma in response to chronic infection. Nature. 2010;465:793–797.
83. Park D, Spencer JA, Koh BI, et al. Endogenous bone marrow MSCs are dynamic, fate-restricted participants in bone maintenance and regeneration. Cell Stem Cell. 2012;10:259–272.
84. Higashi AY, Ikawa T, Muramatsu M, et al. Direct hematological toxicity and illegitimate chromosomal recombination caused by the systemic activation of CreERT2. J Immunol. 2009;182:5633–5640.
85. Sánchez-Aguilera A, Arranz L, Martín-Pérez D, et al. Estrogen signaling selectively induces apoptosis of hematopoietic progenitors and myeloid neoplasms without harming steady-state hematopoiesis. Cell Stem Cell. 2014;15:791–804.
86. Cáceres EF, Hurst LD. The evolution, impact and properties of exonic splice enhancers. Genome Biol. 2013;14:R143.
87. Goren A, Ram O, Amit M, et al. Comparative analysis identifies exonic splicing regulatory sequences—The complex definition of enhancers and silencers. Mol Cell. 2006;22:769–781.
88. Shiozawa Y, Malcovati L, Gallì A, et al. Aberrant splicing and defective mRNA production induced by somatic spliceosome mutations in myelodysplasia. Nat Commun. 2018;9:3649.
89. Liddicoat BJ, Piskol R, Chalk AM, et al. RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. Science. 2015;349:1115–1120.
90. Heraud-Farlow JE, Walkley CR. The role of RNA editing by ADAR1 in prevention of innate immune sensing of self-RNA. J Mol Med (Berl). 2016;94:1095–1102.
91. Rice GI, Kasher PR, Forte GM, et al. Mutations in ADAR1 cause Aicardi-Goutieres syndrome associated with a type I interferon signature. Nat Genet. 2012;44:1243–1248.
92. Mercer TR, Clark MB, Andersen SB, et al. Genome-wide discovery of human splicing branchpoints. Genome Res. 2015;25:290–303.

Bibiography

Abdel-Wahab, O., Mullally, A., Hedvat, C., Garcia-Manero, G., Patel, J., Wadleigh, M., Malinge, S., Yao, J., Kilpivaara, O., Bhat, R., et al. (2009). Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. Blood 114, 144-147.

Abdel-Wahab, O., Tefferi, A., and Levine, R. L. (2012). Role of TET2 and ASXL1 mutations in the pathogenesis of myeloproliferative neoplasms. Hematol Oncol Clin North Am 26, 1053-1064.

Abelson, S., Collord, G., Ng, S. W. K., Weissbrod, O., Mendelson Cohen, N., Niemeyer, E., Barda, N., Zuzarte, P. C., Heisler, L., Sundaravadanam, Y., et al. (2018). Prediction of acute myeloid leukaemia risk in healthy individuals. Nature 559, 400-404.

Abram, C. L., Roberge, G. L., Hu, Y., and Lowell, C. A. (2014). Comparative analysis of the efficiency and specificity of myeloid-Cre deleting strains using ROSA-EYFP reporter mice. Journal of Immunological Methods 408, 89-100.

Adiconis, X., Borges-Rivera, D., Satija, R., DeLuca, D. S., Busby, M. A., Berlin, A. M., Sivachenko, A., Thompson, D. A., Wysoker, A., Fennell, T., et al. (2013). Comparative analysis of RNA sequencing methods for degraded or low-input samples. Nat Methods 10, 623-629.

Adolfsson, J., Månsson, R., Buza-Vidas, N., Hultquist, A., Liuba, K.,

Jensen, C. T., Bryder, D., Yang, L., Borge, O.-J., Thoren, L. A. M., et al. (2005). Identification of Flt3+ Lympho-Myeloid Stem Cells Lacking Erythro-Megakaryocytic Potential: A Revised Road Map for Adult Blood Lineage Commitment. Cell 121, 295-306.

An, J., González-Avalos, E., Chawla, A., Jeong, M., López-Moyado, I. F., Li, W., Goodell, M. A., Chavez, L., Ko, M., and Rao, A. (2015). Acute loss of TET function results in aggressive myeloid cancer in mice. 6, 10071.

Anczuków, O., Rosenberg, A. Z., Akerman, M., Das, S., Zhan, L., Karni, R., Muthuswamy, S. K., and Krainer, A. R. (2012). The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. Nature Structural & Molecular Biology 19, 220-228.

Arber, D. A., and Hasserjian, R. P. (2015). Reclassifying myelodysplastic syndromes: what's where in the new WHO and why. Hematology 2015, 294-298.

Arber, D. A., Orazi, A., Hasserjian, R., Thiele, J., Borowitz, M. J., Le Beau, M. M., Bloomfield, C. D., Cazzola, M., and Vardiman, J. W. (2016). The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 127, 2391-2405.

Armstrong, R. N., Steeples, V., Singh, S., Sanchi, A., Boultwood, J., and Pellagatti, A. (2018). Splicing factor mutations in the myelodysplastic syndromes: target genes and therapeutic approaches. Adv Biol Regul 67, 13-29.

Asmar, F., Punj, V., Christensen, J., Pedersen, M. T., Pedersen, A., Nielsen, A. B., Hother, C., Ralfkiaer, U., Brown, P., Ralfkiaer, E., et al. (2013).

Genome-wide profiling identifies a DNA methylation signature that associates with TET2 mutations in diffuse large B-cell lymphoma. Haematologica 98, 1912-1920.

Assi, R., Kantarjian, H. M., Garcia-Manero, G., Cortes, J. E., Pemmaraju, N., Wang, X., Nogueras-Gonzalez, G., Jabbour, E., Bose, P., Kadia, T., et al. (2018). A phase II trial of ruxolitinib in combination with azacytidine in myelodysplastic syndrome/myeloproliferative neoplasms. Am J Hematol 93, 277-285.

Australian Institute of, H., and Welfare (2013). Cancer in Australia: Actual incidence data from 1991 to 2009 and mortality data from 1991 to 2010 with projections to 2012. Asia Pac J Clin Oncol 9, 199-213.

Awada, H., Nagata, Y., Goyal, A., Asad, M. F., Patel, B., Hirsch, C. M., Kuzmanovic, T., Guan, Y., Przychodzen, B. P., Aly, M., et al. (2019). Invariant phenotype and molecular association of biallelic TET2 mutant myeloid neoplasia. Blood Advances 3, 339-349.

Bajaj, J., Hamilton, M., Shima, Y., Chambers, K., Spinler, K., Van Nostrand, E. L., Yee, B. A., Blue, S. M., Chen, M., Rizzeri, D., et al. (2020). An in vivo genome-wide CRISPR screen identifies the RNA-binding protein Staufen2 as a key regulator of myeloid leukemia. Nature Cancer 1, 410-422.

Barlow, J. L., Drynan, L. F., Hewett, D. R., Holmes, L. R., Lorenzo-Abalde, S., Lane, A. L., Jolin, H. E., Pannell, R., Middleton, A. J., Wong, S. H., et al. (2010). A p53-dependent mechanism underlies macrocytic anemia in a mouse model of human 5q- syndrome. Nat Med 16, 59-66.

Bechara, Elias G., Sebestyén, E., Bernardis, I., Eyras, E., and Valcárcel, J. (2013). RBM5, 6, and 10 Differentially Regulate NUMB Alternative Splicing to Control Cancer Cell Proliferation. Molecular Cell 52, 720-733.

Beerman, I., Bhattacharya, D., Zandi, S., Sigvardsson, M., Weissman, I. L., Bryder, D., and Rossi, D. J. (2010). Functionally distinct hematopoietic stem cells modulate hematopoietic lineage potential during aging by a mechanism of clonal expansion. Proceedings of the National Academy of Sciences 107, 5465-5470.

Bejar, R. (2018). What biologic factors predict for transformation to AML? Best Pract Res Clin Haematol 31, 341-345.

Bejar, R., and Steensma, D. P. (2014). Recent developments in myelodysplastic syndromes. Blood 124, 2793-2803.

Bejar, R., Stevenson, K., Abdel-Wahab, O., Galili, N., Nilsson, B., Garcia-Manero, G., Kantarjian, H., Raza, A., Levine, R. L., Neuberg, D., and Ebert, B. L. (2011). Clinical Effect of Point Mutations in Myelodysplastic Syndromes. New England Journal of Medicine 364, 2496-2506.

Bejar, R., Stevenson, K. E., Caughey, B. A., Abdel-Wahab, O., Steensma, D. P., Galili, N., Raza, A., Kantarjian, H., Levine, R. L., Neuberg, D., et al. (2012). Validation of a Prognostic Model and the Impact of Mutations in Patients With Lower-Risk Myelodysplastic Syndromes. Journal of Clinical Oncology 30, 3376-3382.

Bennett, J. M., Catovsky, D., Daniel, M. T., Flandrin, G., Galton, D. A., Gralnick, H., Sultan, C., and Cox, C. (1994). The chronic myeloid leukaemias: guidelines for distinguishing chronic granulocytic, atypical chronic myeloid, and chronic myelomonocytic leukaemia. Proposals by the French-American-British Cooperative Leukaemia Group. Br J Haematol 87, 746-754.

Berman, E., Jhanwar, S., Hedvat, C., Arcila, M. E., Wahab, O. A., Levine, R., Maloy, M., Ma, W., and Albitar, M. (2016). Resistance to imatinib

in patients with chronic myelogenous leukemia and the splice variant BCR-ABL1(35INS). Leuk Res 49, 108-112.

Biasco, L., Pellin, D., Scala, S., Dionisio, F., Basso-Ricci, L., Leonardelli, L., Scaramuzza, S., Baricordi, C., Ferrua, F., Cicalese, M. P., et al. (2016). In Vivo Tracking of Human Hematopoiesis Reveals Patterns of Clonal Dynamics during Early and Steady-State Reconstitution Phases. Cell Stem Cell 19, 107-119.

Blainey, P. C., and Quake, S. R. (2014). Dissecting genomic diversity, one cell at a time. Nat Methods 11, 19-21.

Bonnal, S., Vigevani, L., and Valcarcel, J. (2012). The spliceosome as a target of novel antitumour drugs. Nat Rev Drug Discov 11, 847-859.

Brooks, A. N., Choi, P. S., de Waal, L., Sharifnia, T., Imielinski, M., Saksena, G., Pedamallu, C. S., Sivachenko, A., Rosenberg, M., Chmielecki, J., et al. (2014). A Pan-Cancer Analysis of Transcriptome Changes Associated with Somatic Mutations in U2AF1 Reveals Commonly Altered Splicing Events. PLOS ONE 9, e87361.

Broske, A. M., Vockentanz, L., Kharazi, S., Huska, M. R., Mancini, E., Scheller, M., Kuhl, C., Enns, A., Prinz, M., Jaenisch, R., et al. (2009). DNA methylation protects hematopoietic stem cell multipotency from myeloerythroid restriction. Nat Genet 41, 1207-1215.

Brunetti, L., Gundry, M. C., and Goodell, M. A. (2017). DNMT3A in Leukemia. Cold Spring Harb Perspect Med 7.

Buenrostro, J. D., Corces, M. R., Lareau, C. A., Wu, B., Schep, A. N., Aryee, M. J., Majeti, R., Chang, H. Y., and Greenleaf, W. J. (2018). Integrated Single-Cell Analysis Maps the Continuous Regulatory Landscape

of Human Hematopoietic Differentiation. Cell 173, 1535-1548 e1516.

Busch, K., Klapproth, K., Barile, M., Flossdorf, M., Holland-Letz, T., Schlenner, S. M., Reth, M., Hofer, T., and Rodewald, H. R. (2015). Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. Nature 518, 542-546.

Cairns, R. A., Iqbal, J., Lemonnier, F., Kucuk, C., de Leval, L., Jais, J. P., Parrens, M., Martin, A., Xerri, L., Brousset, P., et al. (2012). IDH2 mutations are frequent in angioimmunoblastic T-cell lymphoma. Blood 119, 1901-1903.

Cairns, R. A., and Mak, T. W. (2013). Oncogenic isocitrate dehydrogenase mutations: mechanisms, models, and clinical opportunities. Cancer Discov 3, 730-741.

Carrancio, S., Markovics, J., Wong, P., Leisten, J., Castiglioni, P., Groza, M. C., Raymon, H. K., Heise, C., Daniel, T., Chopra, R., and Sung, V. (2014). An activin receptor IIA ligand trap promotes erythropoiesis resulting in a rapid induction of red blood cells and haemoglobin. Br J Haematol 165, 870-882.

Carrelha, J., Meng, Y., Kettyle, L. M., Luis, T. C., Norfo, R., Alcolea, V., Boukarabila, H., Grasso, F., Gambardella, A., Grover, A., et al. (2018). Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. Nature 554, 106-111.

Carrocci, T. J., Zoerner, D. M., Paulson, J. C., and Hoskins, A. A. (2017). SF3b1 mutations associated with myelodysplastic syndromes alter the fidelity of branchsite selection in yeast. Nucleic Acids Res 45, 4837-4852.

Celik, H., Kramer, A., and Challen, G. A. (2016). DNA methylation in normal and malignant hematopoiesis. Int J Hematol 103, 617-626.

Challen, G. A., Sun, D., Jeong, M., Luo, M., Jelinek, J., Berg, J. S., Bock, C., Vasanthakumar, A., Gu, H., Xi, Y., et al. (2011). Dnmt3a is essential for hematopoietic stem cell differentiation. Nat Genet 44, 23-31.

Chapple, R. H., Tseng, Y. J., Hu, T., Kitano, A., Takeichi, M., Hoegenauer, K. A., and Nakada, D. (2018). Lineage tracing of murine adult hematopoietic stem cells reveals active contribution to steady-state hematopoiesis. Blood Adv 2, 1220-1228.

Chavez, A., Scheiman, J., Vora, S., Pruitt, B. W., Tuttle, M., P R Iyer, E., Lin, S., Kiani, S., Guzman, C. D., Wiegand, D. J., et al. (2015). Highly efficient Cas9-mediated transcriptional programming. Nature Methods 12, 326-328.

Chen, H., and Boutros, P. C. (2011). VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics 12, 35.

Chen, J., Kao, Y.-R., Sun, D., Todorova, T. I., Reynolds, D., Narayanagari, S.-R., Montagna, C., Will, B., Verma, A., and Steidl, U. (2019). Myelodysplastic syndrome progression to acute myeloid leukemia at the stem cell level. Nature Medicine 25, 103-110.

Chen, W., Kumar, A. R., Hudson, W. A., Li, Q., Wu, B., Staggs, R. A., Lund, E. A., Sam, T. N., and Kersey, J. H. (2008). Malignant Transformation Initiated by Mll-AF9: Gene Dosage and Critical Target Cells. Cancer Cell 13, 432-440.

Cheng, Y., Luo, H., Izzo, F., Pickering, B. F., Nguyen, D., Myers, R., Schurer,

A., Gourkanti, S., Brüning, J. C., Vu, L. P., et al. (2019). m⁶A RNA Methylation Maintains Hematopoietic Stem Cell Identity and Symmetric Commitment. Cell Reports 28, 1703-1716.e1706.

Choi, J., Goh, G., Walradt, T., Hong, B. S., Bunick, C. G., Chen, K., Bjornson, R. D., Maman, Y., Wang, T., Tordoff, J., et al. (2015). Genomic landscape of cutaneous T cell lymphoma. Nat Genet 47, 1011-1019.

Christodoulou, C., Spencer, J. A., Yeh, S. A., Turcotte, R., Kokkaliaris, K. D., Panero, R., Ramos, A., Guo, G., Seyedhassantehrani, N., Esipova, T. V., et al. (2020). Live-animal imaging of native haematopoietic stem and progenitor cells. Nature 578, 278-283.

Cimmino, L., Dolgalev, I., Wang, Y., Yoshimi, A., Martin, G. H., Wang, J., Ng, V., Xia, B., Witkowski, M. T., Mitchell-Flack, M., et al. (2017). Restoration of TET2 Function Blocks Aberrant Self-Renewal and Leukemia Progression. Cell 170, 1079-1095 e1020.

Clausen, B. E., Burkhardt, C., Reith, W., Renkawitz, R., and Förster, I. (1999). Conditional gene targeting in macrophages and granulocytes using LysMcre mice. Transgenic Research 8, 265-277.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science 339, 819-823.

Cotto, K. C., Wagner, A. H., Feng, Y.-Y., Kiwala, S., Coffman, A. C., Spies, G., Wollam, A., Spies, N. C., Griffith, O. L., and Griffith, M. (2018). DGIdb 3.0: a redesign and expansion of the drug–gene interaction database. Nucleic Acids Research 46, D1068-D1073.

Couronne, L., Bastard, C., and Bernard, O. A. (2012). TET2 and DNMT3A mutations in human T-cell lymphoma. N Engl J Med 366, 95-96.

Darman, R. B., Seiler, M., Agrawal, A. A., Lim, K. H., Peng, S., Aird, D., Bailey, S. L., Bhavsar, E. B., Chan, B., Colla, S., et al. (2015). Cancer-Associated SF3B1 Hotspot Mutations Induce Cryptic 3′ Splice Site Selection through Use of a Different Branch Point. Cell Rep 13, 1033-1045.

Daubner, G. M., Cléry, A., Jayne, S., Stevenin, J., and Allain, F. H.-T. (2012). A syn–anti conformational difference allows SRSF2 to recognize guanines and cytosines equally well. The EMBO Journal 31, 162-174.

David, C. J., and Manley, J. L. (2010). Alternative pre-mRNA splicing regulation in cancer: pathways and programs unhinged. Genes & Development 24, 2343-2364.

De Boer, J., Williams, A., Skavdis, G., Harker, N., Coles, M., Tolaini, M., Norton, T., Williams, K., Roderick, K., Potocnik, A. J., and Kioussis, D. (2003). Transgenic mice with hematopoietic and lymphoid specific expression of Cre. 33, 314-325.

DeBoever, C., Ghia, E. M., Shepard, P. J., Rassenti, L., Barrett, C. L., Jepsen, K., Jamieson, C. H., Carson, D., Kipps, T. J., and Frazer, K. A. (2015). Transcriptome sequencing reveals potential mechanism of cryptic 3′ splice site selection in SF3B1-mutated cancers. PLoS Comput Biol 11, e1004105.

Delhommeau, F., Dupont, S., Della Valle, V., James, C., Trannoy, S., Masse, A., Kosmider, O., Le Couedic, J. P., Robert, F., Alberdi, A., et al. (2009). Mutation in TET2 in myeloid cancers. N Engl J Med 360, 2289-2301.

Dhanjal, J. K., Radhakrishnan, N., and Sundar, D. (2017). Identifying synthetic lethal targets using CRISPR/Cas9 system. Methods 131, 66-73.

Doench, J. G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E. W., Donovan, K. F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nature Biotechnology 34, 184-191.

Doi, A., Park, I. H., Wen, B., Murakami, P., Aryee, M. J., Irizarry, R., Herb, B., Ladd-Acosta, C., Rho, J., Loewer, S., et al. (2009). Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. Nat Genet 41, 1350-1353.

Durinck, S., Spellman, P. T., Birney, E., and Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. Nat Protoc 4, 1184-1191.

Dvinge, H., Guenthoer, J., Porter, P. L., and Bradley, R. K. (2019). RNA components of the spliceosome regulate tissue- and cancer-specific alternative splicing. Genome Res 29, 1591-1604.

Dvinge, H., Kim, E., Abdel-Wahab, O., and Bradley, R. K. (2016). RNA splicing factors as oncoproteins and tumour suppressors. Nat Rev Cancer 16, 413-430.

Elmariah, H., and DeZern, A. E. (2019). Chronic Myelomonocytic Leukemia: 2018 Update to Prognosis and Treatment. Current Hematologic Malignancy Reports 14, 154-163.

Fenaux, P., and Adès, L. (2013). How we treat lower-risk myelodysplastic syndromes. Blood 121, 4280-4286.

Fenaux, P., Platzbecker, U., Mufti, G. J., Garcia-Manero, G., Buckstein, R., Santini, V., Díez-Campelo, M., Finelli, C., Cazzola, M., Ilhan, O., et al. (2020). Luspatercept in Patients with Lower-Risk Myelodysplastic Syndromes. New England Journal of Medicine 382, 140-151.

Ferrando, A. A., and López-Otín, C. (2017). Clonal evolution in leukemia. Nature Medicine 23, 1135-1145.

Finn, R. S., Dering, J., Conklin, D., Kalous, O., Cohen, D. J., Desai, A. J., Ginther, C., Atefi, M., Chen, I., Fowst, C., et al. (2009). PD 0332991, a selective cyclin D kinase 4/6 inhibitor, preferentially inhibits proliferation of luminal estrogen receptor-positive human breast cancer cell lines in vitro. Breast Cancer Research 11, R77.

Flotho, C., Claus, R., Batz, C., Schneider, M., Sandrock, I., Ihde, S., Plass, C., Niemeyer, C. M., and Lubbert, M. (2009). The DNA methyltransferase inhibitors azacitidine, decitabine and zebularine exert differential effects on cancer gene expression in acute myeloid leukemia cells. Leukemia 23, 1019-1028.

Fu, X. D. (1995). The superfamily of arginine/serine-rich splicing factors. RNA 1, 663-680.

Fu, X. D., and Ares, M., Jr. (2014). Context-dependent control of alternative splicing by RNA-binding proteins. Nat Rev Genet 15, 689-701.

Fu, X. D., Mayeda, A., Maniatis, T., and Krainer, A. R. (1992). General splicing factors SF2 and SC35 have equivalent activities in vitro, and both affect alternative 5′ and 3′ splice site selection. Proceedings of the National Academy of Sciences of the United States of America 89, 11224-11228.

Gaidzik, V. I., Schlenk, R. F., Paschka, P., Stolzle, A., Spath, D., Kuendgen,

A., von Lilienfeld-Toal, M., Brugger, W., Derigs, H. G., Kremers, S., et al. (2013). Clinical impact of DNMT3A mutations in younger adult patients with acute myeloid leukemia: results of the AML Study Group (AMLSG). Blood 121, 4769-4777.

Gale, R. E., Lamb, K., Allen, C., El-Sharkawi, D., Stowe, C., Jenkinson, S., Tinsley, S., Dickson, G., Burnett, A. K., Hills, R. K., and Linch, D. C. (2015). Simpson's Paradox and the Impact of Different DNMT3A Mutations on Outcome in Younger Adults With Acute Myeloid Leukemia. J Clin Oncol 33, 2072-2083.

Geest, C. R., and Coffer, P. J. (2009). MAPK signaling pathways in the regulation of hematopoiesis. Journal of Leukocyte Biology 86, 237-250. Genovese, G., Kahler, A. K., Handsaker, R. E., Lindberg, J., Rose, S. A., Bakhoum, S. F., Chambert, K., Mick, E., Neale, B. M., Fromer, M., et al. (2014). Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med 371, 2477-2487.

Gilbert, Luke A., Larson, Matthew H., Morsut, L., Liu, Z., Brar, Gloria A., Torres, Sandra E., Stern-Ginossar, N., Brandman, O., Whitehead, Evan H., Doudna, Jennifer A., et al. (2013). CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. Cell 154, 442-451.

Gilliland, D. G., and Griffin, J. D. (2002). The roles of FLT3 in hematopoiesis and leukemia. Blood 100, 1532-1542.

Gothert, J. R., Gustin, S. E., Hall, M. A., Green, A. R., Gottgens, B., Izon, D. J., and Begley, C. G. (2005). In vivo fate-tracing studies using the Scl stem cell enhancer: embryonic hematopoietic stem cells significantly contribute to adult hematopoiesis. Blood 105, 2724-2732.

Graubert, T. A., Shen, D., Ding, L., Okeyo-Owuor, T., Lunn, C. L., Shao,

J., Krysiak, K., Harris, C. C., Koboldt, D. C., Larson, D. E., et al. (2011). Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. Nat Genet 44, 53-57.

Greenberg, P. L., Tuechler, H., Schanz, J., Sanz, G., Garcia-Manero, G., Sole, F., Bennett, J. M., Bowen, D., Fenaux, P., Dreyfus, F., et al. (2012). Revised international prognostic scoring system for myelodysplastic syndromes. Blood 120, 2454-2465.

Grisendi, S., Mecucci, C., Falini, B., and Pandolfi, P. P. (2006). Nucleophosmin and cancer. Nature Reviews Cancer 6, 493-505.

Grossmann, V., Haferlach, C., Weissmann, S., Roller, A., Schindela, S., Poetzinger, F., Stadler, K., Bellos, F., Kern, W., Haferlach, T., et al. (2013). The molecular profile of adult T-cell acute lymphoblastic leukemia: mutations in RUNX1 and DNMT3A are associated with poor prognosis in T-ALL. Genes Chromosomes Cancer 52, 410-422.

Grover, A., Sanjuan-Pla, A., Thongjuea, S., Carrelha, J., Giustacchini, A., Gambardella, A., Macaulay, I., Mancini, E., Luis, T. C., Mead, A., et al. (2016). Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells. Nat Commun 7, 11075.

Guillamot, M., Cimmino, L., and Aifantis, I. (2016). The Impact of DNA Methylation in Hematopoietic Malignancies. Trends Cancer 2, 70-83.

Haferlach, T., Nagata, Y., Grossmann, V., Okuno, Y., Bacher, U., Nagae, G., Schnittger, S., Sanada, M., Kon, A., Alpermann, T., et al. (2014). Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia 28, 241-247.

Hagemann-Jensen, M., Ziegenhain, C., Chen, P., Ramsköld, D., Hendriks,

G.-J., Larsson, A. J. M., Faridani, O. R., and Sandberg, R. (2020). Single-cell RNA counting at allele and isoform resolution using Smart-seq3. Nature Biotechnology 38, 708-714.

Hajkova, P., Jeffries, S. J., Lee, C., Miller, N., Jackson, S. P., and Surani, M. A. (2010). Genome-wide reprogramming in the mouse germ line entails the base excision repair pathway. Science 329, 78-82.

Hardwick, S. A., Chen, W. Y., Wong, T., Deveson, I. W., Blackburn, J., Andersen, S. B., Nielsen, L. K., Mattick, J. S., and Mercer, T. R. (2016). Spliced synthetic genes as internal controls in RNA sequencing experiments. Nat Methods 13, 792-798.

Hart, T., Chandrashekhar, M., Aregger, M., Steinhart, Z., Brown, Kevin R., MacLeod, G., Mis, M., Zimmermann, M., Fradet-Turcotte, A., Sun, S., et al. (2015). High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. Cell 163, 1515-1526.

Hart, T., Tong, A. H. Y., Chan, K., Van Leeuwen, J., Seetharaman, A., Aregger, M., Chandrashekhar, M., Hustedt, N., Seth, S., Noonan, A., et al. (2017). Evaluation and Design of Genome-Wide CRISPR/SpCas9 Knockout Screens. G3 (Bethesda) 7, 2719-2727.

Hasegawa, N., Oshima, M., Sashida, G., Matsui, H., Koide, S., Saraya, A., Wang, C., Muto, T., Takane, K., Kaneda, A., et al. (2017). Impact of combinatorial dysfunctions of Tet2 and Ezh2 on the epigenome in the pathogenesis of myelodysplastic syndrome. Leukemia 31, 861-871.

He, C., Zhou, F., Zuo, Z., Cheng, H., and Zhou, R. (2009). A Global View of Cancer-Specific Transcript Variants by Subtractive Transcriptome-Wide Analysis. 4, e4732.

Heath, E. M., Chan, S. M., Minden, M. D., Murphy, T., Shlush, L. I., and Schimmer, A. D. (2017). Biological and clinical consequences of NPM1 mutations in AML. Leukemia 31, 798-807.

Hershberger, C. E., Moyer, D. C., Adema, V., Kerr, C. M., Walter, W., Hutter, S., Meggendorfer, M., Baer, C., Kern, W., Nadarajah, N., et al. (2020). Complex landscape of alternative splicing in myeloid neoplasms. Leukemia.

Ho, Y. H., Del Toro, R., Rivera-Torres, J., Rak, J., Korn, C., Garcia-Garcia, A., Macias, D., Gonzalez-Gomez, C., Del Monte, A., Wittner, M., et al. (2019). Remodeling of Bone Marrow Hematopoietic Stem Cell Niches Promotes Myeloid Cell Expansion during Premature or Physiological Aging. Cell Stem Cell 25, 407-418 e406.

Hong, D. S., Kurzrock, R., Naing, A., Wheler, J. J., Falchook, G. S., Schiffman, J. S., Faulkner, N., Pilat, M. J., O'Brien, J., and LoRusso, P. (2014). A phase I, open-label, single-arm, dose-escalation study of E7107, a precursor messenger ribonucleic acid (pre-mRNA) splicesome inhibitor administered intravenously on days 1 and 8 every 21 days to patients with solid tumors. Invest New Drugs 32, 436-444.

Hurtado, A. M., Luengo-Gil, G., Chen-Liang, T. H., Amaral, F., Batta, K., Palomo, L., Lumbreras, E., Przychodzen, B., Caparros, E., Amigo, M. L., et al. (2018). Transcriptomic rationale for synthetic lethality-targeting ERCC1 and CDKN1A in chronic myelomonocytic leukaemia. British Journal of Haematology 182, 373-383.

Ilagan, J. O., Ramakrishnan, A., Hayes, B., Murphy, M. E., Zebari, A. S., Bradley, P., and Bradley, R. K. (2015). U2AF1 mutations alter splice site recognition in hematological malignancies. Genome Res 25, 14-26.

Itzykson, R., and Fenaux, P. (2011). [Hypomethylating agents for the treatment of myelodysplastic syndromes]. Bull Cancer 98, 927-934.

Itzykson, R., Kosmider, O., Renneville, A., Morabito, M., Preudhomme, C., Berthon, C., Adès, L., Fenaux, P., Platzbecker, U., Gagey, O., et al. (2013). Clonal architecture of chronic myelomonocytic leukemias. Blood 121, 2186-2198.

Izaurralde, E., Lewis, J., McGuigan, C., Jankowska, M., Darzynkiewicz, E., and Mattaj, I. W. (1994). A nuclear cap binding protein complex involved in pre-mRNA splicing. Cell 78, 657-668.

Izzo, F., Lee, S. C., Poran, A., Chaligne, R., Gaiti, F., Gross, B., Murali, R. R., Deochand, S. D., Ang, C., Jones, P. W., et al. (2020). DNA methylation disruption reshapes the hematopoietic differentiation landscape. Nature Genetics 52, 378-387.

Jacobsen, S. E. W., and Nerlov, C. (2019). Haematopoiesis in the era of advanced single-cell technologies. Nature Cell Biology 21, 2-8.

Jaiswal, S., Fontanillas, P., Flannick, J., Manning, A., Grauman, P. V., Mar, B. G., Lindsley, R. C., Mermel, C. H., Burtt, N., Chavez, A., et al. (2014). Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med 371, 2488-2498.

Jankowska, A. M., Szpurka, H., Tiu, R. V., Makishima, H., Afable, M., Huh, J., O'Keefe, C. L., Ganetzky, R., McDevitt, M. A., and Maciejewski, J. P. (2009). Loss of heterozygosity 4q24 and TET2 mutations associated with myelodysplastic/myeloproliferative neoplasms. Blood 113, 6403-6410.

Jeong, H. H., Kim, S. Y., Rousseaux, M. W. C., Zoghbi, H. Y., and Liu, Z. (2019). Beta-binomial modeling of CRISPR pooled screen data identifies

target genes with greater sensitivity and fewer false negatives. Genome Res 29, 999-1008.

Ji, H., Ehrlich, L. I., Seita, J., Murakami, P., Doi, A., Lindau, P., Lee, H., Aryee, M. J., Irizarry, R. A., Kim, K., et al. (2010). Comprehensive methylome map of lineage commitment from haematopoietic progenitors. Nature 467, 338-342.

Jiang, Y., Chen, C., Li, Z., Guo, W., Gegner, J. A., Lin, S., and Han, J. (1996). Characterization of the structure and function of a new mitogen-activated protein kinase (p38beta). J Biol Chem 271, 17920-17926.

Joseph, C., Julie, Carl, Steven, Cristina, and Louise (2013). Deciphering Hematopoietic Stem Cells in Their Niches: A Critical Appraisal of Genetic Models, Lineage Tracing, and Imaging Strategies. Cell Stem Cell 13, 520-533.

Kahles, A., Lehmann, K.-V., Toussaint, N. C., Hüser, M., Stark, S. G., Sachsenberg, T., Stegle, O., Kohlbacher, O., Sander, C., Rätsch, G., et al. (2018). Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. Cancer Cell 34, 211-224.e216.

Kaida, D., Motoyoshi, H., Tashiro, E., Nojima, T., Hagiwara, M., Ishigami, K., Watanabe, H., Kitahara, T., Yoshida, T., Nakajima, H., et al. (2007). Spliceostatin A targets SF3b and inhibits both splicing and nuclear retention of pre-mRNA. Nat Chem Biol 3, 576-583.

Karni, R., De Stanchina, E., Lowe, S. W., Sinha, R., Mu, D., and Krainer, A. R. (2007). The gene encoding the splicing factor SF2/ASF is a proto-oncogene. Nature Structural & Molecular Biology 14, 185-193.

Kats, L. M., Reschke, M., Taulli, R., Pozdnyakova, O., Burgess, K., Bhargava, P., Straley, K., Karnik, R., Meissner, A., Small, D., et al. (2014). Proto-oncogenic role of mutant IDH2 in leukemia initiation and maintenance. Cell Stem Cell 14, 329-341.

Kawabe, T. (2004). G2 checkpoint abrogators as anticancer drugs. Mol Cancer Ther 3, 513-519.

Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M., and Stamm, S. (2013). Function of alternative splicing. Gene 514, 1-30.

Kennedy, J. A., and Barabé, F. (2008). Investigating human leukemogenesis: from cell lines to in vivo models of human leukemia. Leukemia 22, 2029-2040. Kernytsky, A., Wang, F., Hansen, E., Schalm, S., Straley, K., Gliser, C., Yang, H., Travins, J., Murray, S., Dorsch, M., et al. (2015). IDH2 mutation-induced histone and DNA hypermethylation is progressively reversed by small-molecule inhibition. Blood 125, 296-303.

Kiel, M. J., Yilmaz, O. H., Iwashita, T., Yilmaz, O. H., Terhorst, C., and Morrison, S. J. (2005). SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. Cell 121, 1109-1121.

Kim, E., Ilagan, J. O., Liang, Y., Daubner, G. M., Lee, S. C., Ramakrishnan, A., Li, Y., Chung, Y. R., Micol, J. B., Murphy, M. E., et al. (2015). SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. Cancer Cell 27, 617-630.

Kim, H., Lee, J.-H., Lee, W.-S., Kim, I., Moon, J. H., Choi, C. W., Lee, H. S., Park, J., Choi, Y., Shin, H.-J., et al. (2019). Lenalidomide as a second-line therapy after failure of hypomethylating agents in patients with myelodysplastic syndrome. British Journal of Haematology 186,

e151-e155.

Kim, Y. J., and Abdel-Wahab, O. (2017). Therapeutic targeting of RNA splicing in myelodysplasia. Semin Hematol 54, 167-173.

Kimura, A., Kinjyo, I., Matsumura, Y., Mori, H., Mashima, R., Harada, M., Chien, K. R., Yasukawa, H., and Yoshimura, A. (2004). SOCS3 is a physiological negative regulator for granulopoiesis and granulocyte colony-stimulating factor receptor signaling. J Biol Chem 279, 6905-6910.

Kirschnek, S., Vier, J., Gautam, S., Frankenberg, T., Rangelova, S., Eitz-Ferrer, P., Grespi, F., Ottina, E., Villunger, A., Hacker, H., and Hacker, G. (2011). Molecular analysis of neutrophil spontaneous apoptosis reveals a strong role for the pro-apoptotic BH3-only protein Noxa. Cell Death Differ 18, 1805-1814.

Klampfl, T., Milosevic, J. D., Puda, A., Schonegger, A., Bagienski, K., Berg, T., Harutyunyan, A. S., Gisslinger, B., Rumi, E., Malcovati, L., et al. (2013). Complex patterns of chromosome 11 aberrations in myeloid malignancies target CBL, MLL, DDB1 and LMO2. PLoS One 8, e77819.

Klose, M., Florian, M. C., Gerbaulet, A., Geiger, H., and Glauche, I. (2019). Hematopoietic Stem Cell Dynamics Are Regulated by Progenitor Demand: Lessons from a Quantitative Modeling Approach. Stem Cells 37, 948-957.

Ko, M., An, J., Pastor, W. A., Koralov, S. B., Rajewsky, K., and Rao, A. (2015). TET proteins and 5-methylcytosine oxidation in hematological cancers. Immunol Rev 263, 6-21.

Ko, M., Bandukwala, H. S., An, J., Lamperti, E. D., Thompson, E. C., Hastie, R., Tsangaratou, A., Rajewsky, K., Koralov, S. B., and Rao, A. (2011).

Ten-Eleven-Translocation 2 (TET2) negatively regulates homeostasis and differentiation of hematopoietic stem cells in mice. Proc Natl Acad Sci U S A 108, 14566-14571.

Koh, C. M., Bezzi, M., Low, D. H., Ang, W. X., Teo, S. X., Gay, F. P., Al-Haddawi, M., Tan, S. Y., Osato, M., Sabo, A., et al. (2015). MYC regulates the core pre-mRNA splicing machinery as an essential step in lymphomagenesis. Nature 523, 96-100.

Koike-Yusa, H., Li, Y., Tan, E. P., Velasco-Herrera, M. D. C., and Yusa, K. (2014). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nature Biotechnology 32, 267-273.

Komeno, Y., Huang, Y.-J., Qiu, J., Lin, L., Xu, Y., Zhou, Y., Chen, L., Monterroza, D. D., Li, H., Dekelver, R. C., et al. (2015). SRSF2 Is Essential for Hematopoiesis, and Its Myelodysplastic Syndrome-Related Mutations Dysregulate Alternative Pre-mRNA Splicing. Molecular and Cellular Biology 35, 3071-3082.

Komrokji, R., Garcia-Manero, G., Ades, L., Prebet, T., Steensma, D. P., Jurcic, J. G., Sekeres, M. A., Berdeja, J., Savona, M. R., Beyne-Rauzy, O., et al. (2018). Sotatercept with long-term extension for the treatment of anaemia in patients with lower-risk myelodysplastic syndromes: a phase 2, dose-ranging trial. Lancet Haematol 5, e63-e72.

Kon, A., Yamazaki, S., Nannya, Y., Kataoka, K., Ota, Y., Nakagawa, M. M., Yoshida, K., Shiozawa, Y., Morita, M., Yoshizato, T., et al. (2018). Physiological Srsf2 P95H expression causes impaired hematopoietic stem cell functions and aberrant RNA splicing in mice. Blood 131, 621-635.

Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O.

O., Barcena, C., Hsu, P. D., Habib, N., Gootenberg, J. S., Nishimasu, H., et al. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature 517, 583-588.

Konieczny, J., and Arranz, L. (2018). Updates on Old and Weary Haematopoiesis. Int J Mol Sci 19.

Kotake, Y., Sagane, K., Owa, T., Mimori-Kiyosue, Y., Shimizu, H., Uesugi, M., Ishihama, Y., Iwata, M., and Mizui, Y. (2007). Splicing factor SF3b as a target of the antitumor natural product pladienolide. Nature Chemical Biology 3, 570-575.

Krainer, A. R., Conway, G. C., and Kozak, D. (1990). The essential pre-mRNA splicing factor SF2 influences 5′ splice site selection by activating proximal sites. Cell 62, 35-42.

Kunimoto, H., Meydan, C., Nazir, A., Whitfield, J., Shank, K., Rapaport, F., Maher, R., Pronier, E., Meyer, S. C., Garrett-Bakelman, F. E., et al. (2018). Cooperative Epigenetic Remodeling by TET2 Loss and NRAS Mutation Drives Myeloid Transformation and MEK Inhibitor Sensitivity. Cancer Cell 33, 44-59 e48.

Lam, K., Muselman, A., Du, R., Harada, Y., Scholl, A. G., Yan, M., Matsuura, S., Weng, S., Harada, H., and Zhang, D. E. (2014). Hmga2 is a direct target gene of RUNX1 and regulates expansion of myeloid progenitors in mice. Blood 124, 2203-2212.

Laiosa, C. V., Stadtfeld, M., Xie, H., de Andres-Aguayo, L., and Graf, T. (2006). Reprogramming of committed T cell progenitors to macrophages and dendritic cells by C/EBP alpha and PU.1 transcription factors. Immunity 25, 731-744.

Langemeijer, S. M., Kuiper, R. P., Berends, M., Knops, R., Aslanyan, M. G., Massop, M., Stevens-Linders, E., van Hoogen, P., van Kessel, A. G., Raymakers, R. A., et al. (2009). Acquired mutations in TET2 are common in myelodysplastic syndromes. Nat Genet 41, 838-842.

Laurenti, E., and Göttgens, B. (2018). From haematopoietic stem cells to complex differentiation landscapes. Nature 553, 418-426.

Law, C. W., Chen, Y., Shi, W., and Smyth, G. K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biology 15, R29.

Lee, K. M., and Tarn, W. Y. (2013). Coupling pre-mRNA processing to transcription on the RNA factory assembly line. RNA Biol 10, 380-390.

Lee, S. C., North, K., Kim, E., Jang, E., Obeng, E., Lu, S. X., Liu, B., Inoue, D., Yoshimi, A., Ki, M., et al. (2018). Synthetic Lethal and Convergent Biological Effects of Cancer-Associated Spliceosomal Gene Mutations. Cancer Cell 34, 225-241 e228.

Lee, S. C.-W., and Abdel-Wahab, O. (2016). Therapeutic targeting of splicing in cancer. Nature Medicine 22, 976-986.

Ley, T. J., Ding, L., Walter, M. J., McLellan, M. D., Lamprecht, T., Larson, D. E., Kandoth, C., Payton, J. E., Baty, J., Welch, J., et al. (2010). DNMT3A mutations in acute myeloid leukemia. N Engl J Med 363, 2424-2433.

Li, W., Köster, J., Xu, H., Chen, C.-H., Xiao, T., Liu, J. S., Brown, M., and Liu, X. S. (2015). Quality control, modeling, and visualization of CRISPR screens with MAGeCK-VISPR. Genome Biology 16, 281.

Li, W., Xu, H., Xiao, T., Cong, L., Love, M. I., Zhang, F., Irizarry, R. A.,

Liu, J. S., Brown, M., and Liu, X. S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biology 15, 554.

Li, X., Wang, J., and Manley, J. L. (2005). Loss of splicing factor ASF/SF2 induces G2 cell cycle arrest and apoptosis, but inhibits internucleosomal DNA fragmentation. Genes Dev 19, 2705-2714.

Lin, Y. W., Slape, C., Zhang, Z., and Aplan, P. D. (2005). NUP98-HOXD13 transgenic mice develop a highly penetrant, severe myelodysplastic syndrome that progresses to acute leukemia. Blood 106, 287-295.

Li, Z., Cai, X., Cai, C.-L., Wang, J., Zhang, W., Petersen, B. E., Yang, F.-C., and Xu, M. (2011). Deletion of Tet2 in mice leads to dysregulated hematopoietic stem cells and subsequent development of myeloid malignancies. Blood 118, 4509-4518.

Liang, Y., Tebaldi, T., Rejeski, K., Joshi, P., Stefani, G., Taylor, A., Song, Y., Vasic, R., Maziarz, J., Balasubramanian, K., et al. (2018). SRSF2 mutations drive oncogenesis by activating a global program of aberrant alternative splicing in hematopoietic cells. Leukemia 32, 2659-2671.

Liddicoat, B. J., Piskol, R., Chalk, A. M., Ramaswami, G., Higuchi, M., Hartner, J. C., Li, J. B., Seeburg, P. H., and Walkley, C. R. (2015). RNA editing by ADAR1 prevents MDA5 sensing of endogenous dsRNA as nonself. Science 349, 1115-1120.

Lin, S., Coutinho-Mansfield, G., Wang, D., Pandit, S., and Fu, X.-D. (2008). The splicing factor SC35 has an active role in transcriptional elongation. Nature Structural & Molecular Biology 15, 819-826.

Loghavi, S., and Khoury, J. D. (2018). Recent Updates on Chronic Myelomonocytic Leukemia. Curr Hematol Malig Rep 13, 446-454.

Lorsbach, R. B., Moore, J., Mathew, S., Raimondi, S. C., Mukatira, S. T., and Downing, J. R. (2003). TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). Leukemia 17, 637-641.

Ma, X., Does, M., Raza, A., and Mayne, S. T. (2007). Myelodysplastic syndromes: incidence and survival in the United States. Cancer 109, 1536-1542.

Makishima, H., Visconte, V., Sakaguchi, H., Jankowska, A. M., Abu Kar, S., Jerez, A., Przychodzen, B., Bupathi, M., Guinta, K., Afable, M. G., et al. (2012). Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. Blood 119, 3203-3210.

Makishima, H., Yoshizato, T., Yoshida, K., Sekeres, M. A., Radivoyevitch, T., Suzuki, H., Przychodzen, B., Nagata, Y., Meggendorfer, M., Sanada, M., et al. (2017). Dynamics of clonal evolution in myelodysplastic syndromes. Nat Genet 49, 204-212.

Malcovati, L., Hellstrom-Lindberg, E., Bowen, D., Ades, L., Cermak, J., Del Canizo, C., Della Porta, M. G., Fenaux, P., Gattermann, N., Germing, U., et al. (2013). Diagnosis and treatment of primary myelodysplastic syndromes in adults: recommendations from the European LeukemiaNet. Blood 122, 2943-2964.

Malcovati, L., Papaemmanuil, E., Bowen, D. T., Boultwood, J., Della Porta, M. G., Pascutto, C., Travaglino, E., Groves, M. J., Godfrey, A. L., Ambaglio, I., et al. (2011). Clinical significance of SF3B1 mutations in myelodysplastic syndromes and myelodysplastic/myeloproliferative

neoplasms. Blood 118, 6239-6246.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., Norville, J. E., and Church, G. M. (2013). RNA-guided human genome engineering via Cas9. Science 339, 823-826.

Mann, M., Mehta, A., De Boer, C. G., Kowalczyk, M. S., Lee, K., Haldeman, P., Rogel, N., Knecht, A. R., Farouq, D., Regev, A., and Baltimore, D. (2018). Heterogeneous Responses of Hematopoietic Stem Cells to Inflammatory Stimuli Are Altered with Age. Cell Reports 25, 2992-3005.e2995.

Marcucci, G., Maharry, K., Wu, Y. Z., Radmacher, M. D., Mrozek, K., Margeson, D., Holland, K. B., Whitman, S. P., Becker, H., Schwind, S., et al. (2010). IDH1 and IDH2 gene mutations identify novel molecular subsets within de novo cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. J Clin Oncol 28, 2348-2355.

Mardis, E. R., Ding, L., Dooling, D. J., Larson, D. E., McLellan, M. D., Chen, K., Koboldt, D. C., Fulton, R. S., Delehaunty, K. D., McGrath, S. D., et al. (2009). Recurring mutations found by sequencing an acute myeloid leukemia genome. N Engl J Med 361, 1058-1066.

Martincorena, I., Raine, K. M., Gerstung, M., Dawson, K. J., Haase, K., Van Loo, P., Davies, H., Stratton, M. R., and Campbell, P. J. (2017). Universal Patterns of Selection in Cancer and Somatic Tissues. Cell 171, 1029-1041.e1021.

Matera, A. G., and Wang, Z. (2014). A day in the life of the spliceosome. Nat Rev Mol Cell Biol 15, 108-121.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. Nature 403, 501-502.

McKerrell, T., Park, N., Moreno, T., Grove, C. S., Ponstingl, H., Stephens, J., Understanding Society Scientific, G., Crawley, C., Craig, J., Scott, M. A., et al. (2015). Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis. Cell Rep 10, 1239-1245.

McQuilten, Z. K., Wood, E. M., Polizzotto, M. N., Campbell, L. J., Wall, M., Curtis, D. J., Farrugia, H., McNeil, J. J., and Sundararajan, V. (2014). Underestimation of myelodysplastic syndrome incidence by cancer registries: Results from a population-based data linkage study. Cancer 120, 1686-1694.

Meggendorfer, M., Haferlach, T., Alpermann, T., Jeromin, S., Haferlach, C., Kern, W., and Schnittger, S. (2014). Specific molecular mutation patterns delineate chronic neutrophilic leukemia, atypical chronic myeloid leukemia, and chronic myelomonocytic leukemia. Haematologica 99, e244-246.

Meggendorfer, M., Roller, A., Haferlach, T., Eder, C., Dicker, F., Grossmann, V., Kohlmann, A., Alpermann, T., Yoshida, K., Ogawa, S., et al. (2012). SRSF2 mutations in 275 cases with chronic myelomonocytic leukemia (CMML). Blood 120, 3080-3088.

Meldi, K. M., and Figueroa, M. E. (2015). Cytosine modifications in myeloid malignancies. Pharmacol Ther 152, 42-53.

Menssen, A. J., and Walter, M. J. (2020). Genetics of progression from MDS to secondary leukemia. Blood 136, 50-60.

Merlevede, J., Droin, N., Qin, T., Meldi, K., Yoshida, K., Morabito, M., Chautard, E., Auboeuf, D., Fenaux, P., Braun, T., et al. (2016). Mutation allele burden remains unchanged in chronic myelomonocytic leukaemia responding to hypomethylating agents. Nature Communications 7, 10767.

Mian, S. A., Smith, A. E., Kulasekararaj, A. G., Kizilors, A., Mohamedali, A. M., Lea, N. C., Mitsopoulos, K., Ford, K., Nasser, E., Seidl, T., and Mufti, G. J. (2013). Spliceosome mutations exhibit specific associations with epigenetic modifiers and proto-oncogenes mutated in myelodysplastic syndrome. Haematologica 98, 1058-1066.

Montalban-Bravo, G., and Garcia-Manero, G. (2018). Myelodysplastic syndromes: 2018 update on diagnosis, risk-stratification and management. Am J Hematol 93, 129-147.

Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Maria, Vasanthakumar, A., Patel, J., Zhao, X., et al. (2011). Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation. Cancer Cell 20, 11-24.

Morlan, J. D., Qu, K., and Sinicropi, D. V. (2012). Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. PLoS One 7, e42882.

Morrison, S. J., Uchida, N., and Weissman, I. L. (1995). The biology of hematopoietic stem cells. Annu Rev Cell Dev Biol 11, 35-71.

Murphy, M. A., Schnall, R. G., Venter, D. J., Barnett, L., Bertoncello, I., Thien, C. B., Langdon, W. Y., and Bowtell, D. D. (1998). Tissue hyperplasia and enhanced T-cell signalling via ZAP-70 in c-Cbl-deficient mice. Mol Cell Biol 18, 4872-4882.

Muto, T., Sashida, G., Oshima, M., Wendt, G. R., Mochizuki-Kashio, M., Nagata, Y., Sanada, M., Miyagi, S., Saraya, A., Kamio, A., et al. (2013). Concurrent loss of Ezh2 and Tet2 cooperates in the pathogenesis of myelodysplastic disorders. J Exp Med 210, 2627-2639.

Nachmani, D., Bothmer, A. H., Grisendi, S., Mele, A., Bothmer, D., Lee, J. D., Monteleone, E., Cheng, K., Zhang, Y., Bester, A. C., et al. (2019). Germline NPM1 mutations lead to altered rRNA 2'-O-methylation and cause dyskeratosis congenita. Nature Genetics 51, 1518-1529.

Nachtkamp, K., Stark, R., Strupp, C., Kundgen, A., Giagounidis, A., Aul, C., Hildebrandt, B., Haas, R., Gattermann, N., and Germing, U. (2016). Causes of death in 2877 patients with myelodysplastic syndromes. Ann Hematol 95, 937-944.

Narla, A., and Ebert, B. L. (2010). Ribosomopathies: human disorders of ribosome dysfunction. Blood 115, 3196-3205.

Navas, T. A., Mohindru, M., Estes, M., Ma, J. Y., Sokol, L., Pahanish, P., Parmar, S., Haghnazari, E., Zhou, L., Collins, R., et al. (2006). Inhibition of overactivated p38 MAPK can restore hematopoiesis in myelodysplastic syndrome progenitors. Blood 108, 4170-4177.

Nestorowa, S., Hamey, F. K., Pijuan Sala, B., Diamanti, E., Shepherd, M., Laurenti, E., Wilson, N. K., Kent, D. G., and Gottgens, B. (2016). A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. Blood 128, e20-31.

Niemeyer, C. M. (2018). JMML genomics and decisions. Hematology 2018, 307-312.

O'Neil, N. J., Bailey, M. L., and Hieter, P. (2017). Synthetic lethality and cancer. Nature Reviews Genetics 18, 613-623.

Obeng, E. A., Chappell, R. J., Seiler, M., Chen, M. C., Campagna, D. R., Schmidt, P. J., Schneider, R. K., Lord, A. M., Wang, L., Gambe, R. G., et al.

(2016). Physiologic Expression of Sf3b1(K700E) Causes Impaired Erythropoiesis, Aberrant Splicing, and Sensitivity to Therapeutic Spliceosome Modulation. Cancer Cell 30, 404-417.

Obeng, E. A., Stewart, C., and Abdel-Wahab, O. (2019). Altered RNA Processing in Cancer Pathogenesis and Therapy. Cancer Discov 9, 1493-1510.

Ochi, Y., Kon, A., Sakata, T., Nakagawa, M. M., Nakazawa, N., Kakuta, M., Kataoka, K., Koseki, H., Nakayama, M., Morishita, D., et al. (2020). Combined Cohesin–RUNX1 Deficiency Synergistically Perturbs Chromatin Looping and Causes Myelodysplastic Syndromes. Cancer Discovery 10, 836-853.

Odejide, O., Weigert, O., Lane, A. A., Toscano, D., Lunning, M. A., Kopp, N., Kim, S., van Bodegom, D., Bolla, S., Schatz, J. H., et al. (2014). A targeted mutational landscape of angioimmunoblastic T-cell lymphoma. Blood 123, 1293-1296.

Ogawa, S. (2019). Genetics of MDS. Blood 133, 1049-1059.

Ogawara, Y., Katsumoto, T., Aikawa, Y., Shima, Y., Kagiyama, Y., Soga, T., Matsunaga, H., Seki, T., Araki, K., and Kitabayashi, I. (2015). IDH2 and NPM1 Mutations Cooperate to Activate Hoxa9/Meis1 and Hypoxia Pathways in Acute Myeloid Leukemia. Cancer Res 75, 2005-2016.

Okano, M., Bell, D. W., Haber, D. A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell 99, 247-257.

Onida, F., Barosi, G., Leone, G., Malcovati, L., Morra, E., Santini, V., Specchia, G., and Tura, S. (2013). Management recommendations for

chronic myelomonocytic leukemia: consensus statements from the SIE, SIES, GITMO groups. Haematologica 98, 1344-1352.

Osawa, M., Hanada, K., Hamada, H., and Nakauchi, H. (1996). Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell. Science 273, 242-245.

Padron, E., Dezern, A., Andrade-Campos, M., Vaddi, K., Scherle, P., Zhang, Q., Ma, Y., Balasis, M. E., Tinsley, S., Ramadan, H., et al. (2016). A Multi-Institution Phase I Trial of Ruxolitinib in Patients with Chronic Myelomonocytic Leukemia (CMML). Clin Cancer Res 22, 3746-3754.

Padron, E., Painter, J. S., Kunigal, S., Mailloux, A. W., McGraw, K., McDaniel, J. M., Kim, E., Bebbington, C., Baer, M., Yarranton, G., et al. (2013). GM-CSF-dependent pSTAT5 sensitivity is a feature with therapeutic potential in chronic myelomonocytic leukemia. Blood 121, 5068-5077.

Padron, E., and Steensma, D. P. (2015). Cutting the cord from myelodysplastic syndromes: chronic myelomonocytic leukemia-specific biology and management strategies. Curr Opin Hematol 22, 163-170.

Palomero, T., Couronne, L., Khiabanian, H., Kim, M. Y., Ambesi-Impiombato, A., Perez-Garcia, A., Carpenter, Z., Abate, F., Allegretta, M., Haydu, J. E., et al. (2014). Recurrent mutations in epigenetic regulators, RHOA and FYN kinase in peripheral T cell lymphomas. Nat Genet 46, 166-170.

Palomo, L., Meggendorfer, M., Hutter, S., Twardziok, S., Adema, V., Fuhrmann, I., Fuster-Tormo, F., Xicoy, B., Zamora, L., Acha, P., et al. (2020). Molecular landscape and clonal architecture of adult myelodysplastic/myeloproliferative neoplasms. Blood.

Pan, F., Wingo, T. S., Zhao, Z., Gao, R., Makishima, H., Qu, G., Lin, L.,

Yu, M., Ortega, J. R., Wang, J., et al. (2017). Tet2 loss leads to hypermutagenicity in haematopoietic stem/progenitor cells. Nat Commun 8, 15102.

Pang, W. W., Price, E. A., Sahoo, D., Beerman, I., Maloney, W. J., Rossi, D. J., Schrier, S. L., and Weissman, I. L. (2011). Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. Proc Natl Acad Sci U S A 108, 20012-20017.

Papaemmanuil, E., Cazzola, M., Boultwood, J., Malcovati, L., Vyas, P., Bowen, D., Pellagatti, A., Wainscoat, J. S., Hellstrom-Lindberg, E., Gambacorti-Passerini, C., et al. (2011). Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. N Engl J Med 365, 1384-1395.

Papaemmanuil, E., Gerstung, M., Malcovati, L., Tauro, S., Gundem, G., Van Loo, P., Yoon, C. J., Ellis, P., Wedge, D. C., Pellagatti, A., et al. (2013). Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood 122, 3616-3627; quiz 3699.

Parikh, S. A., and Tefferi, A. (2013). Chronic myelomonocytic leukemia: 2013 update on diagnosis, risk stratification, and management. Am J Hematol 88, 967-974.

Pastor, W. A., Aravind, L., and Rao, A. (2013). TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. Nat Rev Mol Cell Biol 14, 341-356.

Patel, B. J., Przychodzen, B., Thota, S., Radivoyevitch, T., Visconte, V., Kuzmanovic, T., Clemente, M., Hirsch, C., Morawski, A., Souaid, R., et al. (2017). Genomic determinants of chronic myelomonocytic leukemia. Leukemia 31, 2815-2823.

Pati, H., and Kundil Veetil, K. (2019). Myelodysplastic Syndrome/Myeloproliferative Neoplasm (MDS/MPN) Overlap Syndromes: Molecular Pathogenetic Mechanisms and Their Implications. Indian J Hematol Blood Transfus 35, 3-11.

Patnaik, M. M., Sallman, D. A., Mangaonkar, A. A., Heuer, R., Hirvela, J., Zblewski, D., Al-Kali, A., Kanne, D., Ismail, R., Komrokji, R. S., et al. (2019). A Phase 1 Study of Lenzilumab, a humaneered recombinant Anti-Human Granulocyte-Macrophage Colony- Stimulating Factor (anti-hGM-CSF) Antibody, for Chronic Myelomonocytic Leukemia (CMML). Blood 134, 4234-4234.

Patnaik, M. M., Sallman, D. A., Sekeres, M. A., Luger, S., Bejar, R., Hobbs, G. S., DeZern, A. E., Bolognese, J., Traynor, M., Mishra, V., et al. (2017). Preliminary Results from an Open-Label, Phase 2 Study of Tipifarnib in Chronic Myelomonocytic Leukemia (CMML). Blood 130, 2963-2963.

Patnaik, M. M., and Tefferi, A. (2018). Chronic myelomonocytic leukemia: 2018 update on diagnosis, risk stratification and management. Am J Hematol 93, 824-840.

Pei, W., Feyerabend, T. B., Rossler, J., Wang, X., Postrach, D., Busch, K., Rode, I., Klapproth, K., Dietlein, N., Quedenau, C., et al. (2017). Polylox barcoding reveals haematopoietic stem cell fates realized in vivo. Nature 548, 456-460.

Pellagatti, A., Armstrong, R. N., Steeples, V., Sharma, E., Repapi, E., Singh, S., Sanchi, A., Radujkovic, A., Horn, P., Dolatshad, H., et al. (2018). Impact of spliceosome mutations on RNA splicing in myelodysplasia: dysregulated genes/pathways and clinical associations. Blood 132, 1225-1240.

Pietras, E. M., Reynaud, D., Kang, Y. A., Carlin, D., Calero-Nieto, F. J., Leavitt, A. D., Stuart, J. M., Gottgens, B., and Passegue, E. (2015). Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions. Cell Stem Cell 17, 35-46.

Pineault, N., Buske, C., Feuring-Buske, M., Abramovich, C., Rosten, P., Hogge, D. E., Aplan, P. D., and Humphries, R. K. (2003). Induction of acute myeloid leukemia in mice by the human leukemia-specific fusion gene NUP98-HOXD13 in concert with Meis1. Blood 101, 4529-4538.

Platzbecker, U., and Germing, U. (2013). Combination of Azacitidine and Lenalidomide in Myelodysplastic Syndromes or acute Myeloid Leukemia a wise Liaison? Leukemia 27, 1813-1819.

Platzbecker, U., Germing, U., Götze, K., Kiewe, P., Wolff, T., Mayer, K., Chromik, J., Radsak, M., Donovan, E., Wilson, D. M., et al. (2016). Luspatercept Increases Hemoglobin and Reduces Transfusion Burden in Patients with Low-Intermediate Risk Myelodysplastic Syndromes (MDS): Long-Term Results from Phase 2 PACE-MDS Study. Blood 128, 3168-3168.

Poulikakos, P. I., Persaud, Y., Janakiraman, M., Kong, X., Ng, C., Moriceau, G., Shi, H., Atefi, M., Titz, B., Gabay, M. T., et al. (2011). RAF inhibitor resistance is mediated by dimerization of aberrantly spliced BRAF(V600E). Nature 480, 387-390.

Pronk, C. J., Rossi, D. J., Mansson, R., Attema, J. L., Norddahl, G. L., Chan, C. K., Sigvardsson, M., Weissman, I. L., and Bryder, D. (2007). Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. Cell Stem Cell 1, 428-442.

Purton, L. E., and Scadden, D. T. (2007). Limiting Factors in Murine Hematopoietic Stem Cell Assays. Cell Stem Cell 1, 263-270.

Qi, Lei S., Larson, Matthew H., Gilbert, Luke A., Doudna, Jennifer A., Weissman, Jonathan S., Arkin, Adam P., and Lim, Wendell A. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell 152, 1173-1183.

Qiu, J., Zhou, B., Thol, F., Zhou, Y., Chen, L., Shao, C., Deboever, C., Hou, J., Li, H., Chaturvedi, A., et al. (2016). Distinct splicing signatures affect converged pathways in myelodysplastic syndrome patients carrying mutations in different splicing regulators. RNA 22, 1535-1549.

Quesada, V., Conde, L., Villamor, N., Ordóñez, G. R., Jares, P., Bassaganyas, L., Ramsay, A. J., Beà, S., Pinyol, M., Martínez-Trillos, A., et al. (2012). Exome sequencing identifies recurrent mutations of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nature Genetics 44, 47-52.

Quivoron, C., Couronne, L., Della Valle, V., Lopez, C. K., Plo, I., Wagner-Ballon, O., Do Cruzeiro, M., Delhommeau, F., Arnulf, B., Stern, M. H., et al. (2011). TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. Cancer Cell 20, 25-38.

Raaijmakers, M. H. G. P., Mukherjee, S., Guo, S., Zhang, S., Kobayashi, T., Schoonmaker, J. A., Ebert, B. L., Al-Shahrour, F., Hasserjian, R. P., Scadden, E. O., et al. (2010). Bone progenitor dysfunction induces myelodysplasia and secondary leukaemia. Nature 464, 852-857.

Rahman, M. A., Lin, K. T., Bradley, R. K., Abdel-Wahab, O., and Krainer, A. R. (2020). Recurrent SRSF2 mutations in MDS affect both splicing and

NMD. Genes Dev 34, 413-427.

Rathinam, C., Thien, C. B. F., Flavell, R. A., and Langdon, W. Y. (2010). Myeloid Leukemia Development in c-Cbl RING Finger Mutant Mice Is Dependent on FLT3 Signaling. Cancer Cell 18, 341-352.

Roberts, D. A., and Steensma, D. P. (2015). Outlook and Management of Patients with Myelodysplastic Syndromes Failed by Hypomethylating Agents. Curr Hematol Malig Rep 10, 318-328.

Rodriguez-Fraticelli, A. E., Wolock, S. L., Weinreb, C. S., Panero, R., Patel, S. H., Jankovic, M., Sun, J., Calogero, R. A., Klein, A. M., and Camargo, F. D. (2018). Clonal analysis of lineage fate in native haematopoiesis. Nature 553, 212-216.

Roller, A., Grossmann, V., Bacher, U., Poetzinger, F., Weissmann, S., Nadarajah, N., Boeck, L., Kern, W., Haferlach, C., Schnittger, S., et al. (2013). Landmark analysis of DNMT3A mutations in hematological malignancies. Leukemia 27, 1573-1578.

Ross, N. T., Lohmann, F., Carbonneau, S., Fazal, A., Weihofen, W. A., Gleim, S., Salcius, M., Sigoillot, F., Henault, M., Carl, S. H., et al. (2020). CPSF3-dependent pre-mRNA processing as a druggable node in AML and Ewing's sarcoma. Nat Chem Biol 16, 50-59.

Rouault-Pierre, K., Mian, S. A., Goulard, M., Abarrategi, A., Di Tulio, A., Smith, A. E., Mohamedali, A., Best, S., Nloga, A. M., Kulasekararaj, A. G., et al. (2017). Preclinical modeling of myelodysplastic syndromes. Leukemia 31, 2702-2708.

Rundberg Nilsson, A., Soneji, S., Adolfsson, S., Bryder, D., and Pronk, C. J. (2016). Human and Murine Hematopoietic Stem Cell Aging Is Associated

with Functional Impairments and Intrinsic Megakaryocytic/Erythroid Bias. PLoS One 11, e0158369.

Sakata-Yanagimoto, M., Enami, T., Yoshida, K., Shiraishi, Y., Ishii, R., Miyake, Y., Muto, H., Tsuyama, N., Sato-Otsubo, A., Okuno, Y., et al. (2014). Somatic RHOA mutation in angioimmunoblastic T cell lymphoma. Nat Genet 46, 171-175.

Sanada, M., Suzuki, T., Shih, L. Y., Otsu, M., Kato, M., Yamazaki, S., Tamura, A., Honda, H., Sakata-Yanagimoto, M., Kumano, K., et al. (2009). Gain-of-function of mutated C-CBL tumour suppressor in myeloid neoplasms. Nature 460, 904-908.

Sanjana, N. E., Shalem, O., and Zhang, F. (2014). Improved vectors and genome-wide libraries for CRISPR screening. Nature Methods 11, 783-784.

Sasaki, M., Knobbe, C. B., Munger, J. C., Lind, E. F., Brenner, D., Brustle, A., Harris, I. S., Holmes, R., Wakeham, A., Haight, J., et al. (2012). IDH1(R132H) mutation increases murine haematopoietic progenitors and alters epigenetics. Nature 488, 656-659.

Sasca, D., Szybinski, J., Schuler, A., Shah, V., Heidelberger, J., Haehnel, P. S., Dolnik, A., Kriege, O., Fehr, E. M., Gebhardt, W. H., et al. (2019). NCAM1 (CD56) promotes leukemogenesis and confers drug resistance in AML. Blood 133, 2305-2319.

Savona, M. R., Malcovati, L., Komrokji, R., Tiu, R. V., Mughal, T. I., Orazi, A., Kiladjian, J. J., Padron, E., Solary, E., Tibes, R., et al. (2015). An international consortium proposal of uniform response criteria for myelodysplastic/myeloproliferative neoplasms (MDS/MPN) in adults. Blood 125, 1857-1865.

Sawai, C. M., Babovic, S., Upadhaya, S., Knapp, D., Lavin, Y., Lau, C. M., Goloborodko, A., Feng, J., Fujisaki, J., Ding, L., et al. (2016). Hematopoietic Stem Cells Are the Major Source of Multilineage Hematopoiesis in Adult Animals. Immunity 45, 597-609.

Sawen, P., Eldeeb, M., Erlandsson, E., Kristiansen, T. A., Laterza, C., Kokaia, Z., Karlsson, G., Yuan, J., Soneji, S., Mandal, P. K., et al. (2018). Murine HSCs contribute actively to native hematopoiesis but with reduced differentiation capacity upon aging. Elife 7.

Schneider, R. K., Schenone, M., Ferreira, M. V., Kramann, R., Joyce, C. E., Hartigan, C., Beier, F., Brummendorf, T. H., Germing, U., Platzbecker, U., et al. (2016). Rps14 haploinsufficiency causes a block in erythroid differentiation mediated by S100A8 and S100A9. Nat Med 22, 288-297.

Seiler, M., Peng, S., Agrawal, A. A., Palacino, J., Teng, T., Zhu, P., Smith, P. G., Buonamici, S., Yu, L., Caesar-Johnson, S. J., et al. (2018). Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types. Cell Reports 23, 282-296.e284.

Seiler, M., Yoshimi, A., Darman, R., Chan, B., Keaney, G., Thomas, M., Agrawal, A. A., Caleb, B., Csibi, A., Sean, E., et al. (2018). H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers. Nature Medicine 24, 497-504.

Selimoglu-Buet, D., Wagner-Ballon, O., Saada, V., Bardet, V., Itzykson, R., Bencheikh, L., Morabito, M., Met, E., Debord, C., Benayoun, E., et al. (2015). Characteristic repartition of monocyte subsets as a diagnostic signature of chronic myelomonocytic leukemia. Blood 125, 3618-3626.

Shalem, O., Sanjana, N. E., Hartenian, E., Shi, X., Scott, D. A., Mikkelsen, T. S., Heckl, D., Ebert, B. L., Root, D. E., Doench, J. G., and Zhang, F.

(2014). Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Science 343, 84-87.

Shapiro, I. M., Cheng, A. W., Flytzanis, N. C., Balsamo, M., Condeelis, J. S., Oktay, M. H., Burge, C. B., and Gertler, F. B. (2011). An EMT–Driven Alternative Splicing Program Occurs in Human Breast Cancer and Modulates Cellular Phenotype. PLoS Genetics 7, e1002218.

Sharma, S., Kohlstaedt, L. A., Damianov, A., Rio, D. C., and Black, D. L. (2008). Polypyrimidine tract binding protein controls the transition from exon definition to an intron defined spliceosome. Nat Struct Mol Biol 15, 183-191.

Shiozawa, Y., Malcovati, L., Gallì, A., Sato-Otsubo, A., Kataoka, K., Sato, Y., Watatani, Y., Suzuki, H., Yoshizato, T., Yoshida, K., et al. (2018). Aberrant splicing and defective mRNA production induced by somatic spliceosome mutations in myelodysplasia. Nature Communications 9, 3649.

Shirai, C. L., Ley, J. N., White, B. S., Kim, S., Tibbitts, J., Shao, J., Ndonwi, M., Wadugu, B., Duncavage, E. J., Okeyo-Owuor, T., et al. (2015). Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. Cancer Cell 27, 631-643.

Shlush, L. I., Zandi, S., Mitchell, A., Chen, W. C., Brandwein, J. M., Gupta, V., Kennedy, J. A., Schimmer, A. D., Schuh, A. C., Yee, K. W., et al. (2014). Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. Nature 506, 328-333.

Silverman, L. R., Verma, A., Odchimar-Reissig, R., Feldman, E. J., Navada, S. C., Demakos, E. P., Baer, M. R., Najfeld, V., Sparano, J. A., and Piekarz, R. (2013). A Phase II Trial Of Epigenetic Modulators Vorinostat In Combination With Azacitidine (azaC) In Patients With The Myelodysplastic

Syndrome (MDS): Initial Results Of Study 6898 Of The New York Cancer Consortium. Blood 122, 386-386.

Singbrant, S., Russell, M. R., Jovic, T., Liddicoat, B., Izon, D. J., Purton, L. E., Sims, N. A., Martin, T. J., Sankaran, V. G., and Walkley, C. R. (2011). Erythropoietin couples erythropoiesis, B-lymphopoiesis, and bone homeostasis within the bone marrow microenvironment. Blood 117, 5631-5642.

Singbrant, S., Wall, M., Moody, J., Karlsson, G., Chalk, A. M., Liddicoat, B., Russell, M. R., Walkley, C. R., and Karlsson, S. (2014). The SKI protooncogene enhances the in vivo repopulation of hematopoietic stem cells and causes myeloproliferative disease. Haematologica 99, 647-655.

Smeby, J., Sveen, A., Eilertsen, I. A., Danielsen, S. A., Hoff, A. M., Eide, P. W., Johannessen, B., Hektoen, M., Skotheim, R. I., Guren, M. G., et al. (2019). Transcriptional and functional consequences of TP53 splice mutations in colorectal cancer. Oncogenesis 8, 35.

Smeets, M. F., Tan, S. Y., Xu, J. J., Anande, G., Unnikrishnan, A., Chalk, A. M., Taylor, S. R., Pimanda, J. E., Wall, M., Purton, L. E., and Walkley, C. R. (2018). Srsf2 P95H initiates myeloid bias and myelodysplastic/myeloproliferative syndrome from hemopoietic stem cells. Blood 132, 608-621.

Smith, B. N., Savona, M., and Komrokji, R. S. (2019). Challenges in Myelodysplastic/Myeloproliferative Neoplasms (MDS/MPN). Clin Lymphoma Myeloma Leuk 19, 1-8.

Solary, E., Bernard, O. A., Tefferi, A., Fuks, F., and Vainchenker, W. (2014). The Ten-Eleven Translocation-2 (TET2) gene in hematopoiesis and hematopoietic diseases. Leukemia 28, 485-496.

Solary, E., and Itzykson, R. (2017). How I treat chronic myelomonocytic leukemia. Blood 130, 126-136.

Somervaille, T. C. P., and Cleary, M. L. (2006). Identification and characterization of leukemia stem cells in murine MLL-AF9 acute myeloid leukemia. Cancer Cell 10, 257-268.

Song, Y., Rongvaux, A., Taylor, A., Jiang, T., Tebaldi, T., Balasubramanian, K., Bagale, A., Terzi, Y. K., Gbyli, R., Wang, X., et al. (2019). A highly efficient and faithful MDS patient-derived xenotransplantation model for pre-clinical studies. Nature Communications 10.

Sperling, A. S., Gibson, C. J., and Ebert, B. L. (2017). The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. Nat Rev Cancer 17, 5-19.

Sportoletti, P., Grisendi, S., Majid, S. M., Cheng, K., Clohessy, J. G., Viale, A., Teruya-Feldstein, J., and Pandolfi, P. P. (2008). Npm1 is a haploinsufficient suppressor of myeloid and lymphoid malignancies in the mouse. Blood 111, 3859-3862.

Spring, L. M., Wander, S. A., Zangardi, M., and Bardia, A. (2019). CDK 4/6 Inhibitors in Breast Cancer: Current Controversies and Future Directions. Current Oncology Reports 21.

Srinivas, S., Watanabe, T., Lin, C.-S., William, C. M., Tanabe, Y., Jessell, T. M., and Costantini, F. (2001). Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. BMC Developmental Biology 1, 4.

Stavropoulou, V., Kaspar, S., Brault, L., Mathijs, Juge, S., Morettini, S., Tzankov, A., Iacovino, M., Lau, I. J., Thomas, et al. (2016). MLL-AF9

Expression in Hematopoietic Stem Cells Drives a Highly Invasive AML Expressing EMT-Related Genes Linked to Poor Outcome. Cancer Cell 30, 43-58.

Steensma, D. P. (2018). Myelodysplastic syndromes current treatment algorithm 2018. Blood Cancer Journal 8, 47.

Steensma, D. P., Bejar, R., Jaiswal, S., Lindsley, R. C., Sekeres, M. A., Hasserjian, R. P., and Ebert, B. L. (2015). Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. Blood 126, 9-16.

Steensma, D. P., Heptinstall, K. V., Johnson, V. M., Novotny, P. J., Sloan, J. A., Camoriano, J. K., Niblack, J., Bennett, J. M., and Mesa, R. A. (2008). Common troublesome symptoms and their impact on quality of life in patients with myelodysplastic syndromes (MDS): Results of a large internet-based survey. Leukemia Research 32, 691-698.

Stegelmann, F., Bullinger, L., Schlenk, R. F., Paschka, P., Griesshammer, M., Blersch, C., Kuhn, S., Schauer, S., Dohner, H., and Dohner, K. (2011). DNMT3A mutations in myeloproliferative neoplasms. Leukemia 25, 1217-1219.

Such, E., Germing, U., Malcovati, L., Cervera, J., Kuendgen, A., Della Porta, M. G., Nomdedeu, B., Arenillas, L., Luño, E., Xicoy, B., et al. (2013). Development and validation of a prognostic scoring system for patients with chronic myelomonocytic leukemia. Blood 121, 3005-3015.

Sun, J., Ramos, A., Chapman, B., Johnnidis, J. B., Le, L., Ho, Y.-J., Klein, A., Hofmann, O., and Camargo, F. D. (2014). Clonal dynamics of native haematopoiesis. Nature 514, 322-327.

Tadokoro, Y., Ema, H., Okano, M., Li, E., and Nakauchi, H. (2007). De novo DNA methyltransferase is essential for self-renewal, but not for differentiation, in hematopoietic stem cells. J Exp Med 204, 715-722.

Tahiliani, M., Koh, K. P., Shen, Y., Pastor, W. A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L. M., Liu, D. R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science 324, 930-935.

Taylor, J., Mi, X., North, K. D., Binder, M., Penson, A., Lasho, T. L., Knorr, K., Haddadin, M., Liu, B., Pangallo, J., et al. (2020). Single-cell genomics reveals the genetic and molecular bases for escape from mutational epistasis in myeloid neoplasms. Blood.

Tefferi, A., Lim, K. H., Abdel-Wahab, O., Lasho, T. L., Patel, J., Patnaik, M. M., Hanson, C. A., Pardanani, A., Gilliland, D. G., and Levine, R. L. (2009). Detection of mutant TET2 in myeloid malignancies other than myeloproliferative neoplasms: CMML, MDS, MDS/MPN and AML. Leukemia 23, 1343-1345.

Thien, C. B., and Langdon, W. Y. (2001). Cbl: many adaptations to regulate protein tyrosine kinases. Nat Rev Mol Cell Biol 2, 294-307.

Thota, S., and Gerds, A. T. (2018). Myelodysplastic and myeloproliferative neoplasms: updates on the overlap syndromes. Leuk Lymphoma 59, 803-812.

Tian, B., and Manley, J. L. (2017). Alternative polyadenylation of mRNA precursors. Nature Reviews Molecular Cell Biology 18, 18-30.

Traina, F., Visconte, V., Elson, P., Tabarroki, A., Jankowska, A. M., Hasrouni, E., Sugimoto, Y., Szpurka, H., Makishima, H., O'Keefe, C. L., et al.

(2014). Impact of molecular mutations on treatment response to DNMT inhibitors in myelodysplasia and related neoplasms. Leukemia 28, 78-87.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S., and Rinn, J. L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol 32, 381-386.

Trowbridge, J. J., Snow, J. W., Kim, J., and Orkin, S. H. (2009). DNA methyltransferase 1 is essential for and uniquely regulates hematopoietic stem and progenitor cells. Cell Stem Cell 5, 442-449.

Tyner, J. W., Tognon, C. E., Bottomly, D., Wilmot, B., Kurtz, S. E., Savage, S. L., Long, N., Schultz, A. R., Traer, E., Abel, M., et al. (2018). Functional genomic landscape of acute myeloid leukaemia. Nature 562, 526-531.

Tzelepis, K., Koike-Yusa, H., De Braekeleer, E., Li, Y., Metzakopian, E., Dovey, O. M., Mupo, A., Grinkevich, V., Li, M., Mazan, M., et al. (2016). A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia. Cell Rep 17, 1193-1205.

Unnikrishnan, A., Papaemmanuil, E., Beck, D., Deshpande, N. P., Verma, A., Kumari, A., Woll, P. S., Richards, L. A., Knezevic, K., Chandrakanthan, V., et al. (2017). Integrative Genomics Identifies the Molecular Basis of Resistance to Azacitidine Therapy in Myelodysplastic Syndromes. Cell Rep 20, 572-585.

Uy, N., Singh, A., Gore, S. D., and Prebet, T. (2017). Hypomethylating agents (HMA) treatment for myelodysplastic syndromes: alternatives in the frontline and relapse settings. Expert Opin Pharmacother 18, 1213-1224.

Vardiman, J. (2012). The classification of MDS: from FAB to WHO and beyond. Leuk Res 36, 1453-1458.

Velten, L., Haas, S. F., Raffel, S., Blaszkiewicz, S., Islam, S., Hennig, B. P., Hirche, C., Lutz, C., Buss, E. C., Nowak, D., et al. (2017). Human haematopoietic stem cell lineage commitment is a continuous process. Nat Cell Biol 19, 271-281.

Venables, J. P., Klinck, R., Bramard, A., Inkel, L., Dufresne-Martin, G., Koh, C., Gervais-Bird, J., Lapointe, E., Froehlich, U., Durand, M., et al. (2008). Identification of Alternative Splicing Markers for Breast Cancer. 68, 9525-9531.

Venables, J. P., Klinck, R., Koh, C., Gervais-Bird, J., Bramard, A., Inkel, L., Durand, M., Couture, S., Froehlich, U., Lapointe, E., et al. (2009). Cancer-associated regulation of alternative splicing. Nat Struct Mol Biol 16, 670-676.

Ventura, A., Kirsch, D. G., McLaughlin, M. E., Tuveson, D. A., Grimm, J., Lintault, L., Newman, J., Reczek, E. E., Weissleder, R., and Jacks, T. (2007). Restoration of p53 function leads to tumour regression in vivo. Nature 445, 661-665.

Verovskaya, E., Broekhuis, M. J., Zwart, E., Ritsema, M., van Os, R., de Haan, G., and Bystrykh, L. V. (2013). Heterogeneity of young and aged murine hematopoietic stem cells revealed by quantitative clonal analysis using cellular barcoding. Blood 122, 523-532.

Voso, M. T., Leone, G., Piciocchi, A., Fianchi, L., Santarone, S., Candoni, A., Criscuolo, M., Masciulli, A., Cerqui, E., Molteni, A., et al. (2017). Feasibility of allogeneic stem-cell transplantation after azacitidine bridge in higher-risk myelodysplastic syndromes and low blast count acute

myeloid leukemia: results of the BMT-AZA prospective study. Ann Oncol 28, 1547-1553.

Wahl, M. C., Will, C. L., and Luhrmann, R. (2009). The spliceosome: design principles of a dynamic RNP machine. Cell 136, 701-718.

Walkley, C. R., Olsen, G. H., Dworkin, S., Fabb, S. A., Swann, J., Grant, Westmoreland, S. V., Chambon, P., Scadden, D. T., and Purton, L. E. (2007). A Microenvironment-Induced Myeloproliferative Syndrome Caused by Retinoic Acid Receptor γ Deficiency. Cell 129, 1097-1110.

Walkley, C. R., Shea, J. M., Sims, N. A., Purton, L. E., and Orkin, S. H. (2007). Rb regulates interactions between hematopoietic stem cells and their bone marrow microenvironment. Cell 129, 1081-1095.

Walter, M. J., Ding, L., Shen, D., Shao, J., Grillot, M., McLellan, M., Fulton, R., Schmidt, H., Kalicki-Veizer, J., O'Laughlin, M., et al. (2011). Recurrent DNMT3A mutations in patients with myelodysplastic syndromes. Leukemia 25, 1153-1158.

Walter, M. J., Shen, D., Shao, J., Ding, L., White, B. S., Kandoth, C., Miller, C. A., Niu, B., McLellan, M. D., Dees, N. D., et al. (2013). Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. Leukemia 27, 1275-1282.

Wang, G. G., Calvo, K. R., Pasillas, M. P., Sykes, D. B., Hacker, H., and Kamps, M. P. (2006). Quantitative production of macrophages or neutrophils ex vivo using conditional Hoxb8. Nat Methods 3, 287-293.

Wang, H. Y., Xu, X., Ding, J. H., Bermingham, J. R., Jr., and Fu, X. D. (2001). SC35 plays a role in T cell development and alternative splicing of CD45. Mol Cell 7, 331-342.

Wang, L., Lawrence, M. S., Wan, Y., Stojanov, P., Sougnez, C., Stevenson, K., Werner, L., Sivachenko, A., DeLuca, D. S., Zhang, L., et al. (2011). SF3B1 and Other Novel Cancer Genes in Chronic Lymphocytic Leukemia. New England Journal of Medicine 365, 2497-2506.

Wang, T., Birsoy, K., Hughes, N. W., Krupczak, K. M., Post, Y., Wei, J. J., Lander, E. S., and Sabatini, D. M. (2015). Identification and characterization of essential genes in the human genome. Science 350, 1096-1101.

Wang, T., Wei, J. J., Sabatini, D. M., and Lander, E. S. (2014). Genetic Screens in Human Cells Using the CRISPR-Cas9 System. Science 343, 80-84.

Wang, T., Yu, H., Hughes, N. W., Liu, B., Kendirli, A., Klein, K., Chen, W. W., Lander, E. S., and Sabatini, D. M. (2017). Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras. Cell 168, 890-903.e815.

Wegrzyn, J., Lam, J. C., and Karsan, A. (2011). Mouse models of myelodysplastic syndromes. Leuk Res 35, 853-862.

Weiskopf, K., Schnorr, P. J., Pang, W. W., Chao, M. P., Chhabra, A., Seita, J., Feng, M., and Weissman, I. L. (2016). Myeloid Cell Origins, Differentiation, and Clinical Implications. Microbiol Spectr 4.

Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis: Springer International Publishing.

Winer, E. S., and Stone, R. M. (2019). Novel therapy in Acute myeloid leukemia (AML): moving toward targeted approaches. Therapeutic Advances in Hematology 10, 204062071986064.

Wolach, O., and Stone, R. M. (2015). How I treat mixed-phenotype acute leukemia. Blood 125, 2477-2485.

Woll, P. S., Kjallquist, U., Chowdhury, O., Doolittle, H., Wedge, D. C., Thongjuea, S., Erlandsson, R., Ngara, M., Anderson, K., Deng, Q., et al. (2014). Myelodysplastic syndromes are propagated by rare and distinct human cancer stem cells in vivo. Cancer Cell 25, 794-808.

Xiao, R., Sun, Y., Ding, J.-H., Lin, S., Rose, D. W., Rosenfeld, M. G., Fu, X.-D., and Li, X. (2007). Splicing regulator SC35 is essential for genomic stability and cell proliferation during mammalian organogenesis. Molecular and cellular biology 27, 5393-5402.

Xie, M., Jiang, Q., and Xie, Y. (2015). Comparison between decitabine and azacitidine for the treatment of myelodysplastic syndrome: a meta-analysis with 1,392 participants. Clin Lymphoma Myeloma Leuk 15, 22-28.

Xie, M., Lu, C., Wang, J., McLellan, M. D., Johnson, K. J., Wendl, M. C., McMichael, J. F., Schmidt, H. K., Yellapantula, V., Miller, C. A., et al. (2014). Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat Med 20, 1472-1478.

Xu, J. J., Smeets, M. F., Tan, S. Y., Wall, M., Purton, L. E., and Walkley, C. R. (2019). Modeling human RNA spliceosome mutations in the mouse: not all mice were created equal. Experimental Hematology 70, 10-23.

Xu, Q., and Lee, C. (2003). Discovery of novel splice forms and functional analysis of cancer-specific alternative splicing in human expressed sequences. Nucleic Acids Res 31, 5635-5643.

Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S. H., Ito, S., Yang, C.,

Wang, P., Xiao, M. T., et al. (2011). Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. Cancer Cell 19, 17-30.

Yaari, G., Bolen, C. R., Thakar, J., and Kleinstein, S. H. (2013). Quantitative set analysis for gene expression: a method to quantify gene set differential expression including gene-gene correlations. Nucleic Acids Res 41, e170.

Yamamoto, R., Morita, Y., Ooehara, J., Hamanaka, S., Onodera, M., Rudolph, K. L., Ema, H., and Nakauchi, H. (2013). Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. Cell 154, 1112-1126.

Yamauchi, T., Masuda, T., Canver, M. C., Seiler, M., Semba, Y., Shboul, M., Al-Raqad, M., Maeda, M., Schoonenberg, V. A. C., Cole, M. A., et al. (2018). Genome-wide CRISPR-Cas9 Screen Identifies Leukemia-Specific Dependence on a Pre-mRNA Metabolic Pathway Regulated by DCPS. Cancer Cell 33, 386-400 e385.

Yan, X., Himburg, H. A., Pohl, K., Quarmyne, M., Tran, E., Zhang, Y., Fang, T., Kan, J., Chao, N. J., Zhao, L., et al. (2016). Deletion of the Imprinted Gene Grb10 Promotes Hematopoietic Stem Cell Self-Renewal and Regeneration. Cell Rep 17, 1584-1594.

Yang, L., Bryder, D., Adolfsson, J., Nygren, J., Mansson, R., Sigvardsson, M., and Jacobsen, S. E. (2005). Identification of Lin(-)Sca1(+)kit(+)CD34(+)Flt3-short-term hematopoietic stem cells capable of rapidly reconstituting and rescuing myeloablated transplant recipients. Blood 105, 2717-2723.

Ye, M., Iwasaki, H., Laiosa, C. V., Stadtfeld, M., Xie, H., Heck, S., Clausen, B., Akashi, K., and Graf, T. (2003). Hematopoietic Stem Cells Expressing

the Myeloid Lysozyme Gene Retain Long-Term, Multilineage Repopulation Potential. Immunity 19, 689-699.

Yoshida, K., Sanada, M., Shiraishi, Y., Nowak, D., Nagata, Y., Yamamoto, R., Sato, Y., Sato-Otsubo, A., Kon, A., Nagasaki, M., et al. (2011). Frequent pathway mutations of splicing machinery in myelodysplasia. Nature 478, 64-69.

Yoshimi, A., Lin, K. T., Wiseman, D. H., Rahman, M. A., Pastore, A., Wang, B., Lee, S. C., Micol, J. B., Zhang, X. J., de Botton, S., et al. (2019). Coordinated alterations in RNA splicing and epigenetic regulation drive leukaemogenesis. Nature 574, 273-277.

Young, A. L., Challen, G. A., Birmann, B. M., and Druley, T. E. (2016). Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. Nature Communications 7, 12484.

Young, K., Borikar, S., Bell, R., Kuffler, L., Philip, V., and Trowbridge, J. J. (2016). Progressive alterations in multipotent hematopoietic progenitors underlie lymphoid cell loss in aging. J Exp Med 213, 2259-2267.

Zhang, J., Lieu, Y. K., Ali, A. M., Penson, A., Reggio, K. S., Rabadan, R., Raza, A., Mukherjee, S., and Manley, J. L. (2015). Disease-associated mutation in SRSF2 misregulates splicing by altering RNA-binding affinities. Proceedings of the National Academy of Sciences 112, E4726-E4734.

Zhang, J., and Manley, J. L. (2013). Misregulation of Pre-mRNA Alternative Splicing in Cancer. Cancer Discovery 3, 1228-1237.

Zhou, L., Opalinska, J., and Verma, A. (2007). p38 MAP kinase regulates stem cell apoptosis in human hematopoietic failure. Cell Cycle 6, 534-537.

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A. H., Tanaseichuk, O., Benner, C., and Chanda, S. K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nature Communications 10.

Zink, F., Stacey, S. N., Norddahl, G. L., Frigge, M. L., Magnusson, O. T., Jonsdottir, I., Thorgeirsson, T. E., Sigurdsson, A., Gudjonsson, S. A., Gudmundsson, J., et al. (2017). Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. Blood 130, 742-752.