Optimising CRISPR for targeting of primary haematopoietic stem cells



Lucy Field

A thesis submitted for the degree of Doctor of Philosophy in Medical Sciences

St John's College

Trinity Term 2021

ABSTRACT

CRISPR (clustered regularly interspaced short palindromic repeats) offers an unprecedented array of genetic tools for treating and modelling haematological disorders. However, there are concerns about its safety. CRISPR-based editing can induce upregulation of TP53 associated pathways, a finding particularly important for clinical applications. The impact of CRISPR-based editing on haematopoietic stem cells (HSCs) has not, however, been interrogated in detail.

In the work described in this thesis, I initially optimised the CRISPR approach in mouse haematopoietic stem and progenitor cells (HSPCs) by targeting the fluorescent reporter tdTomato, as this enabled lineage tracking of edited HSCs. I then sorted edited HSCs for RNA-Sequencing (RNA-Seq) 24 hours post-editing. This allowed me to assess the immediate transcriptional consequences of genome editing. Using *in vivo* transplantation, I serially tracked the contribution of edited cells to all blood lineages. I carried out RNA-Seq on edited HSCs 18 weeks post-transplant which allowed me to assess whether the short-term changes in transcriptional profiles were maintained long-term.

To demonstrate the applicability of the genome editing approach I had optimised to a disease model setting, I targeted the *NF-1* gene which harbours loss of function mutations in 15% of JMML patients. While mouse models exist for this disorder, most lack a human fetal context. I first optimised targeting of the *NF-1* locus in a GM-CSF dependent cell line (TF-1). I then carried out *in vitro NF-1* targeting studies using human cord blood (CB) to establish a GM-CSF hypersensitivity phenotype. Finally, I set up *in vivo* transplants of *NF-1* edited human CB to establish whether this could propagate the myeloid biased expansion characteristic of JMML.

In this way, I applied the CRISPR genome editing approach to study both normal and malignant haematopoiesis: first to establish the molecular and functional consequences of genome editing on normal haematopoietic output and second to model the consequence of *NF-1* loss of function in a human fetal-like context.

DEDICATION

This thesis is dedicated to my Grandma, Delyse Taylor. You inspire me everyday with your determination to overcome adversity and never let your illness set you back, nor ever complain about the pain. You don't let anything stop your love for gardening and painting even when your arthritis makes it so much harder. I've never met anyone with such generosity and selflessness as you, you welcome everyone you meet as if they are your own family and so many of us who know you look to you for wisdom and support. You taught me how to lose myself in my art, something I will always cherish; you've helped me through the toughest of times and are always there to celebrate my successes. I love you.

ACKNOWLEDGEMENTS

I would like to first thank my supervisors Professor Irene Roberts and Professor Adam Mead. It really has been a privilege to work under the supervision of such inspiring, dedicated, exceptional scientists. They have guided me through what has been a very exciting and enjoyable project, I am indebted to their scientific expertise and guidance. Not only that, I am incredibly grateful for their support on a personal level. As well as always being available to offer insightful advice on any project questions I have, they have been very supportive of my personal development. They have encouraged my endeavours in public engagement and medical writing, which I really appreciate. I would also like to thank the members of my thesis committee as well as transfer and confirmation committees- Professor Tatjana Sauka-Spengler, Professor Anna Schuh, Professor Claus Nerlov and Dr Lynn Quek, for all their scientific insights.

It has also been a great privilege to work with the inspiring and talented scientists in both Irene and Adam's labs. Thank you to each and every one of you for making the MHU WIMM a wonderful place to work and for imparting your scientific wisdom. I would like to particularly thank Nguyen Tran and Patricia Ciccone for processing hundreds of orders and making sure all the lab equipment is maintained.

I would also like to thank everyone who has helped during the course of my project. Firstly to Wei Wen for dedicating a lot of time to the bioinformatics analysis in my thesis. Thank you also for the stimulating discussions, I am very grateful to you for investing keen interest in the biological questions we wanted to answer and helping me think about the results in different ways, which really enhanced the analysis process. Haseeb Rahman and Merve Aksöz I would like

to thank for kindly teaching me the Target-Seq library preparation protocol, Nick Fordham and Joana de Teixeira Carrelha I would like to thank for their help with the mice transplants. I would also like to thank Zemin Ren for is insight in the CRISPR optimisation. David Santa Cruz Hernandez I would like to thank for his help with the JMML project. I would also like to thank Nikolaos Sousos and Eleni Louka for their help and expertise with morphology slide preparation and analysis. I am also very appreciative of all the help from the WIMM and JR Hospital core facilities; I would like to thank all the staff in the BMS, FACS, Genome Engineering and Virus Screening facilities. In particular thanks to Sara Close, Sally-Ann Clark, Kevin Clark, Craig Waugh, Paul Sopp, Philip Hublitz, Zoe Holloway and Ryan Beveridge. A special thanks also to my friends Adi and Lynn for keeping me company during long sessions spent looking down the microscope!

I would like to thank my wonderful friends in Oxford as well as further afield for all their kindness and support. Thank you especially to Janina, Merve, Caitlin, Adi, Ed, Henry, Tom, Rhys and Falk for all the punting trips, May balls and Christmas dinners. To the members of St John's college boat club, thank you for helping me fall in love with rowing, for the most memorable sunrise outings and for inspiring me with your enthusiasm and dedication. Thank you to Sarah for all the early morning bike rides, being the best co-pilot, cooking me dinner after a long day in the lab and of course insisting on listening to me practice every single presentation. Thank you to my best friend Emma for the marathon Skype calls and endless hilarious stories. Thank you to Jodie and Lisa for touring so many countries with me by bike, your gritty determination and endless positivity really inspires me.

Finally, and most importantly, I would like to thank my family for all their love and support. My Dad I would like to thank for driving the 25+ worth of roundabouts it takes to get from Cambridge to Oxford more times than anyone deserves, not to mention sharing my passion for cycle touring and making the 90-mile journey from Cambridge by bike to visit! I would like to thank my Mum for the 1000s of hours you spent listening to me recite science to you to pass my exams and sharing my passion for all things biology. I can't thank you both enough for all your support and encouragement. My sister I would like to thank for always keeping me fuelled for studying by posting the most amazing home-baked treats and for agreeing to completely disassemble your bicycle, put it on a train and come cycle French mountains with me! My brother I would like to thank for always keeping me entertained with your new songs, compositions and videos. Also thank you for putting up with my persistence in convincing you of the beauty of outdoor adventures! Thank you to both of you for the 100s of Skype calls, visits and all your support.

DECLARATION

All the experiments laid out in this thesis were designed, performed and analysed by myself. In cases where colleagues have contributed their help to specific experiments, these are detailed in the relevant sections.

CONTENTS

ABSTRACT……2

DEDICATION……..4

ACKNOWLEDGEMENTS ……5

DECLARATION……7

CONTENTS……8

LIST OF FIGURES……..19

LIST OF TABLES……29

MAIN ABBREVIATIONS……32

CHAPTER 1: INTRODUCTION……34

1.1 Normal murine haematopoiesis……...35

1.1.1 Properties and identification of mature blood lineages……35

1.1.2 *In vitro* methods of enriching for murine haematopoietic stem and progenitor cells……37

1.1.3 *In vivo* methods for identifying long term murine HSCs……39

1.1.4 Maintenance of a homeostatic balance in blood production: asymmetric vs symmetric division……40

1.1.5 Maintenance of a homeostatic balance in blood production: dynamic and fluid commitment of the HSC compartment to the myeloid and lymphoid lineages……41

1.1.6 Maintenance of a homeostatic balance in blood production: heterogeneity in quiescence and functional output……..45

1.1.7 Maintenance of a homeostatic balance in blood production: heterogeneity in transcriptional signature……47

1.2 Normal human haematopoiesis……49

1.2.1 The properties of human haematopoietic stem cells (HSCs): their function and identification……..49

1.2.2 Evidence for heterogeneity within the HSC compartment…….50

1.2.3 Evidence for distinct developmental differences in haematopoiesis: changes in site, gene expression profile and lineage output from the HSC compartment between adult and fetal life……..52

1.3 Juvenile Myelomonocytic Leukaemia (JMML)……55

1.3.1 Interaction of haematopoietic cells with the microenvironment: haematopoietic development and cancer surveillance

1.3.2 Clinical and haematological features of JMML……...55

1.3.3 Molecular pathogenesis of JMML……57

1.3.4 Evidence for a fetal origin of JMML……..60

1.3.5 *In vitro* modelling of JMML……61

1.3.6 Mouse and iPSC models for JMML……..63

1.4 How the CRISPR technique is applied to study normal and malignant haematopoiesis……..67

1.4.1 A brief overview of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) as an adaptive immune system in bacteria and archaea…..67

1.4.2 The Cas9 protein: structure and modification for use in mammalian cells……71

1.4.3 Methods developed to design guide RNAs targeting mammalian genomes.....74

1.4.4 The advantages and disadvantages of plasmid, viral vector and ribonuclear protein (RNP) delivery…….....76

1.4.5 The advantages and disadvantages of lipofection, transfection and transduction……..79

1.4.6 Application of the CRISPR approach to disease correction and modelling in haematopoiesis……83

1.4.7 Using the CRISPR technique to correct monogenic haematological diseases: non-homologous end joining repair……..85

1.4.8 Using the CRISPR technique to correct monogenic haematological diseases: Cas9-induced homology directed repair……86

1.4.9 Using NHEJ to induce loss of function mutations for modelling of haematological malignancies…….87

1.4.10 Using HDR for 'correction' of haematological malignant disease……..89

1.4.11 CRISPR in the context of disease correction and modelling……..90

1.4.12 TP53 association with CRISPR-based editing……81

1.5 Hypotheses and aims……92

1.5.1 Aim 1: To optimise a genome editing approach in mouse HSPCs using in vitro methods……96

1.5.2 Aim 2: Molecular and functional characterisation of CRISPR edited mouse HSCs……..97

1.5.3 Aim 3: To model the JMML associated NF-1 loss of function in human HSPCs……...99

CHAPTER 2: METHODS……101

2.1 Murine and human samples……102

2.1.1 Murine strains……102

2.1.2 Human samples……103

2.1.3 Cell lines……104

2.2 Optimisation of CRISPR in primary mouse HSPCs using in vitro methods……104

2.2.1 Guide RNA design……104

2.2.2 Guide RNA cloning……107

2.2.3 Single and dual guide RNA vector virus production……112

2.2.4 HEK293T cell line transfection by lipofection……114

2.2.5 Ba/F3 cell line transfection optimisation……115

2.2.6 tdTomato induction in primary mouse HSPCs: Neon electroporation…117

2.2.7 tdTomato induction in primary mouse HSPCs: Amaxa nucleofection…120

2.2.8 tdTomato induction in primary mouse HSPCs: Viral transduction……121

2.2.9 tdTomato knockout in primary mouse HSPCs: plasmid and RNP nucleofection.....123

2.2.10 tdTomato knockout in primary mouse HSPCs: transduction……124

2.2.11 Flow cytometry preparation and analysis……124

2.3 Molecular and functional characterisation of CRISPR-edited mouse HSCs…..126

2.3.1 Sample selection……..126

2.3.2 tdTomato knockout time course in mouse HSPCs……..126

2.3.3 Cell cycle status analysis of tdTomato edited HSPCs……128

2.3.4 *In vivo* transplantation of tdTomato edited HSPCs……..129

2.3.5 Flow cytometry analysis of long-term lineage and stem cell engraftment of tdTomato edited HSPCs……130

2.3.6 Fluorescence activated cell sorting (FACS) of tdTomato edited HSCs for Target/RNA-Seq 24 hours post-nucleofection and 18 weeks post-transplant…….....132

2.3.7 Target-Seq library preparation……134

2.3.8 tdTomato genotyping……...136

2.3.9 Bioinformatic analysis of 24-hour and 18-week tdTomato edited HSC transcriptomic data……137

2.4 Introduction of the JMML associated *NF-1* loss of function mutation in human cell lines and primary haematopoietic cells……140

2.4.1 NF-1 guide RNA design……...140

2.4.2 TF-1 cell line selection……141

2.4.3 TF-1 cell line nucleofection optimisation……141

2.4.4 NF-1 guide RNA selection in the TF-1 cell line……142

2.4.5 TF-1 low cell number sorting protocol optimisation……143

2.4.6 Assessment of *NF-1* edited TF-1 sensitivity to GM-CSF……144

2.4.7 TF-1 cell line NF-1 Western blot……145

2.4.8 Cord blood CD34+ cell selection……146

2.4.9 Minimal cytokine optimisation for CB CD34+Lin- single cell colony growth……..147

2.4.10 *NF-1* targeting in CD34+Lin- CB and in vitro growth assessment……148

2.4.11 *NF-1* colony PCR of CD34+Lin- CB colonies……149

2.4.12 *NF-1* and *AAVS1* targeting and NSG transplantation of edited CB CD34+ cells.....150

2.4.13 Flow cytometric and morphological analysis of NSG transplant recipients.....151

2.5 Statistical analysis……153

CHAPTER 3: RESULTS……...154

3.1 Optimisation of CRISPR in primary mouse HSPCs using *in vitro* methods.....155

3.1.1 Experimental approach……...155

3.1.2 Guide RNA design……...156

3.1.3 Guide RNA cloning and virus production……160

3.1.4 Selection of tdTomato induction guide RNAs in the HEK293T cell line……163

3.1.5 Optimisation of electroporation and nucleofection conditions in the Ba/F3 cell line……165

3.1.6 Optimisation of CRISPR delivery mode and format for tdTomato induction in mouse BM HSPCs……172

3.1.7 Optimisation of CRISPR delivery mode and format for tdTomato knockout in mouse BM HSPCs……181

3.1.8 Summary……...192

3.2 Molecular and functional characterisation of CRISPR-edited mouse HSCs……197

3.2.1 Experimental approach……..197

3.2.2 RNP nucleofection editing time course……199

3.2.3 Transcriptomic analysis of the HSC compartment 24h after nucleofection.....211

3.2.4 Long-term engraftment of edited cells in the peripheral blood, bone marrow and spleen……233

3.2.5 Transcriptomic analysis of the HSC compartment 18 weeks post-transplant.....252

3.2.6 Summary……...265

3.3 Introduction of the JMML associated NF-1 loss of function mutation in human cell lines and primary haematopoietic cells……..272

3.3.1 Experimental approach……...272

3.3.2 *NF-1* guide RNA selection……..274

3.3.3 TF-1 cell line selection and nucleofection optimisation……277

3.3.4 Establishment of GM-CSF independent TF-1 clones edited at the *NF-1* locus using the optimised SF DN-100 nucleofection program……...280

3.3.5 Introduction of *in vitro* GM-CSF hypersensitivity in human CB cells through *NF-1* editing using RNP nucleofection…….291

3.3.6 NSG transplantation of *NF-1* targeted CB to evaluate the propensity for in vivo clonal expansion……299

3.3.7 Summary……...317

CHAPTER 4: CONCLUSIONS AND FUTURE DIRECTIONS……..321

4.1 Overview……..322

4.2 Optimisation of CRISPR in primary mouse HSPCs……...324

4.3 Molecular and functional characterisation of CRISPR edited mouse HSCs…..327

4.4 Introduction of the JMML associated NF-1 loss of function mutation in human cell lines and primary haematopoietic cells……331

4.5 Implications of this work……336

4.5.1 Direct comparison of three modes and methods of CRISPR delivery in primary HPSCs targeting the same non-haematopoietic genomic locus with the same guide RNA……336

4.5.2 Use of tdTomato RNP editing as a proxy for editing at another locus……337

4.5.3 Evidence for potential enrichment of TP53 mutated cells by CRISPR editing…..337

4.5.4 Evidence for changes in HSPC composition and function……..338

4.5.5 Direct comparison of short-term and long-term transcriptomic consequences of CRISPR editing……340

4.5.6 NF-1 targeting as an in vitro JMML model……341

4.5.7 NF-1 targeting as an in vivo JMML model……341

4.6 Limitations of the experimental approaches……..342

4.6.1 Refining editing techniques for different cell types……...342

4.6.2 Time-points of transcriptomic evaluation……..343

4.6.3 Impact of species differences……344

4.6.4 Examination of off-target editing……345

4.6.5 Modelling childhood human leukaemias using gene editing of normal CB and limited recipient numbers……..345

4.6.6 NSG mice as a xenograft model for myeloid malignancy……347

4.7 Future directions……...347

4.7.1 Use of tdTomato reporter to enrich for edited populations……..347

4.7.2 Assessment of the effect of TP53 mutations on the engraftment of edited HSCs….....348

4.7.3 Assessment of potential disease propagating populations…….349

4.7.4 Assessment of the importance of a fetal-like context in *NF-1* driven malignancy……..350

4.7.5 Assessment of the importance of timing of acquisition of additional mutations in NF-1 driven malignancy……..351

4.8 Conclusions……352

REFERENCES……...353

Appendix 1: List of monoclonal antibodies used in this project……369

Appendix 2: Detailed large cell number RNP nucleofection protocol for tdTomato knockdown in mouse HSPCs……370

Appendix 3: RNA library indexing using the Nextera XT (Illumina, FC-131-1096) kit……..372

Appendix 4: TF-1 home-brew SF buffer……374

Appendix 5: Lonza 4D nucleofection matrix optimisation table……..375

Appendix 6: Initial alignment mapping rate results for 24hr post-nucleofection and 18wk post-transplant RNA-Seq samples…….376

Appendix 7: Annotated list of genes contributing to the allograft rejection hallmark gene set……378

Appendix 8: HSC gene sets contributing to HSC, LT-HSC and ST-HSC/Progenitor groups for HSC score RNA-Seq analysis……399

LIST OF FIGURES

1.1: Early model of the mouse haematopoietic hierarchy

1.2: Current model of the mouse haematopoietic hierarchy

1.3: Schematic of haematopoietic stem cells (HSCs) within the bone marrow microenvironment

1.4: Genetic landscape of initiating mutations in JMML

1.5: CRISPR mediated adaptive immunity in bacteria/archaea

1.6: Structure of Cas9

1.7: The repair of CRISPR-Cas9 induced double strand breaks

1.8: Extrinsic and intrinsic mechanisms of responding to cellular stress

3.1: Ai9-Rosa26 schematic, clone chart and guide RNA sequences

3.2: tdTomato region schematic, sequence and guide RNA sequences

3.3: Example clone charts and Sanger traces for tdTomato induction and knockdown guide RNAs cloned into PX458 and EFS backbones

3.4: Guide pair 1 and 6 were optimal for tdTomato induction in HEK293T cells

3.5: Optimisation of Ba/F3 transfection conditions experimental workflow

3.6: 1400V 30ms 1p was the optimal electroporation condition for Ba/F3 cells

3.7 Microscopy and flow cytometry of the optimal electroporation condition: 1400V 30ms 1p

3.8: 2.8µg was the optimal plasmid amount for Ba/F3 cell electroporation

3.9: 2.8µg was the optimal plasmid amount for Ba/F3 cell nucleofection

3.10: Comparison of three most optimal electroporation and nucleofection conditions

3.11: Schematic of experimental approach for targeting loxP-STOP-loxP tdTomato mice with dual guide RNAs to induce tdTomato expression

3.12: Flow cytometry gating strategy for analysis tdTomato induction efficiency in mouse HSPCs

3.13: Nucleofection was superior to electroporation for tdTomato induction by plasmid CRISPR delivery in mouse BM HSPCs

3.14: RNP CRISPR induction of tdTomato expression was inefficient despite nucleofection being a highly efficient transfection approach

3.15: Viral transduction was not efficient in delivering CRISPR components to induce tdTomato despite high levels of control virus transduction

3.16: Schematic of tdTomato knockout approach and targeting guide RNAs

3.17: Example flow cytometry plots for analysing tdTomato knockdown in mouse c-Kit BM using RNP nucleofection

3.18: RNP nucleofection was efficient in targeting tdTomato knockout in mouse BM HSPCs. Guide 3 gave optimal editing without significantly reducing viability

3.19: Plasmid nucleofection gave significantly reduced tdTomato expression but significantly reduced viability

3.20: Virus transduction was highly efficient with a small control but significantly less efficient with guide RNA vectors

3.21: Virus transduction significantly reduced tdTomato expression with a delayed editing window compared to plasmid and RNP transfection

3.22: Virus transduction did not significantly reduce viability of mouse HSPCs

3.23: RNP nucleofection was superior to plasmid nucleofection and viral transduction in inducing tdTomato knockout by CRISPR editing

3.24: Experimental workflow for dissecting the short- and long-term molecular and functional consequences of CRISPR-editing on mouse HSCs

3.25: Gating strategy for analysing tdTomato expression in each HSPC compartment

3.26: No tdTomato expression loss was seen in the NT RNP group

3.27: Loss of tdTomato expression was delayed in HSCs compared to more mature cell populations

3.28: tdTomato knockdown was highly efficient in all haematopoietic compartments although delayed in HSCs

3.29: tdTomato RNP targeted HSPC compartment editing summary

3.30: The proportion of cells in the LSK compartment was significantly reduced in tdTomato edited samples compared to controls

3.31: Gating strategy for separating HSPC compartments for cell cycle analysis

3.32: The cell cycle status of primitive HSPC compartments was not significantly altered by tdTomato targeted gene editing

3.33: Wild type tdTomato sequence was less detectable by genomic PCR 24h post nucleofection

3.34: tdTomato HSC editing at 24h HSC correlated with 72 tdTomato expression-tdTomato RNP

3.35: tdTomato HSC editing at 24h HSC correlated with 72 tdTomato expression- NT RNP

3.36: Low Synthego ICE R score lead to underestimation of tdTomato expression loss

3.37: Low level off-target editing was present at the top tdTomato off-target site

3.38: Principal component analysis separates CRISPR edited samples both from fresh controls and nucleofection controls with increasing intervention

3.39: tdTomato gene expression was reduced in TOM RNP targeted groups. Differential gene expression between TOM RNP and control samples decreased with increasing CRISPR intervention

3.40: Differential gene expression displayed by TOM RNP samples showed some overlap between control conditions

3.41: Differential gene expression identified transcriptional changes suggesting implications on HSC function

3.42: TP53 and DNA damage response pathways were upregulated with increasing significance across nucleofection conditions ordered by increasing intervention

3.43: Random forest analysis separated nucleofection groups by gene expression

3.44: Innate immune response, degradation of foreign protein/RNA material, and DNA damage response separated nucleofection groups with increasing intervention

3.45: tdTomato edited samples possessed a stronger LT-HSC score compared to non-edited nucleofection controls

3.46: Gating strategy for analysing peripheral blood lineage engraftment of tdTomato edited and control groups

3.47: Low volume protocol CRISPR editing impeded the engraftment ability of targeted cells more than nucleofection itself or presence of the Cas9 protein

3.48: Low volume protocol CRISPR editing impeded the engraftment in primitive BM compartments of targeted cells more than nucleofection itself or presence of the Cas9 protein

3.49: High volume protocol partially rescued the impeded engraftment caused by the low volume protocol

3.50: High volume protocol rescued the impeded engraftment in primitive stem cell compartments caused by the low volume protocol, but with variable efficiency

3.51: Total peripheral blood engraftment was negatively impacted by editing although the severity of the effect differed between lineages

3.52: Platelet and red blood cell engraftment decreased over serial readouts exclusively in edited samples when accounting for total overall engraftment

3.53: Myeloid engraftment in the bone marrow and spleen contributed to a lower overall proportion in edited samples compared to controls

3.54: Engraftment in the haematopoietic stem/progenitor compartments was significantly reduced for TOM RNP samples compared to controls

3.55: Contribution of tdTomato negative TOM RNP and control groups to peripheral blood mature lineages and bone mature lineages/HSPC compartments for two biological repeats combined

3.56: Proportions of tdTomato negative donors contributing to lineage/HSPC compartments and HSC lineage output for two biological repeats combined

3.57: Technical and biological replicates clustered closely

3.58: Unsupervised clustering failed to separate groups by principal component analysis

3.59: tdTomato reads were significantly diminished in transplant samples. No gene sets were significantly enriched between fresh tdTomato positive and negative samples

3.60: Principal component analysis by differential gene expression separated TOM RNP targeted samples from transplant and fresh controls

3.61: Differential gene expression displayed by TOM RNP samples showed some overlap between control conditions

3.62: Differential gene expression identified transcriptional changes suggesting implications on HSC function

3.63: Gene set enrichment analysis of transplanted TOM RNP targeted samples compared to sham-edited and fresh transplant controls

3.64: Random forest analysis separated transplanted TOM RNP targeted samples from sham-edited and fresh transplanted controls

3.65: Random forest analysis revealed transcriptomic up-regulation of genes associated with innate immune and stress responses, as well as cell cycle control

3.66: Differentially expressed genes in TOM RNP targeted samples 18 weeks post-transplant showed no overlap with differential gene expression patterns 24 hours post-nucleofection

3.67: HSC signature analysis revealed the LT-HSC signature seen in 24h post nucleofection was not sustained post-transplant although the overall stem signature was strong in edited HSCs

3.68: Frequency of additional events in NF-1 initiating JMML

3.69: NF-1 exon 2 region guide RNAs

3.70: GM-CSF significantly reduced TF-1 growth. DN100 was the optimal column-optimised nucleofection program for TF-1 cells in SG buffer

3.71: SG buffer maintained higher viability but SF buffer gave higher transfection efficiency in nucleofection row-optimisation for TF-1 cells

3.72: NF-1 guides 1 and 2 were significantly more efficient than guide 3 or triple guide combination

3.73: Single cell TF-1 colony growth was supported in Terasaki non-conditioned medium

3.74: NF-1 targeting increased the clonogenicity of TF-1 cells, particularly in the absence of GM-CSF

3.75: TF-1 control and AAVS1 non-targeted clones maintained an intact NF-1 locus

3.76: NF-1 edited clones harboured 87-99% editing at the NF-1 locus

3.77: NF-1 editing caused loss of protein expression in targeted clones

3.78: NF-1 editing caused GM-CSF independent proliferation in targeted clones

3.79: Wild type CB clonogenicity was reduced in minimal cytokine (SCF + GM-CSF) growth

3.80: NF-1 edited clones produced bigger colonies than AAVS1 controls under the most stringent cytokine conditions (10ng/mL SCF + 0.1ng/mL GM-CSF)

3.81: Effect of NF-1 editing on CB clonogenicity at low GM-CSF concentrations

3.82: NF-1 CB targeting was not 100% efficient in producing edited colonies in minimal cytokine conditions

3.83: NF-1 editing produced heterozygous, compound heterozygous and homozygous mutated clones

3.84: NF-1 guide 1 produced more compound heterozygous clones than NF-1 guide 2

3.85: CB NF-1 targeting showed a single base insertion dominated at the bulk level

3.86: Gating strategy for assessing human mature lineage engraftment in NSG CB CRISPR edited recipients

3.87: Human CD45 peripheral blood lineage engraftment showed a similar pattern in NF-1 and AAVS1 targeted CB NSG recipients

3.88: Mouse CD45 chimerism reduced over the course of transplant in NF-1 and AAVS1 targeted CB NSG recipients

3.89: Peripheral blood counts of NF-1 CB NSG recipients

3.90: Differential WBC counts were similar between AAVS1 control and NF-1 targeted groups

3.91: Engraftment of NF-1- and control-edited hCD45+ cells in peripheral blood, bone marrow and spleen at 13-18 weeks post-transplant

3.92: Flow cytometry gating strategy for analysis of engraftment of human HSPC populations in the bone marrow and spleen of NF-1 and AAVS1 edited CB recipient mice

3.93: AAVS1 edited CB NSG recipient mice showed a normal human bone marrow composition

3.94: NF-1 edited CB NSG recipient mice showed an expansion of human CD45RA+ CD123- and LMPP in the BM

3.95: NF-1 edited CB NSG recipient mice showed an expansion of CD45RA+ CD123- and LMPP in the BM

3.96: AAVS1 CB recipient BM showed no expansion of an NF-1 edited population

3.97: NF-1 CB recipient bone marrow showed an expansion of the edited population compared to at the time of transplant

3.98: NF-1 CB NSG recipients did not show an enlarged spleen

3.99: NF-1 CB NSG recipients showed reduced megakaryopoiesis but no significant infiltrations of blast populations in the bone marrow

LIST OF TABLES

Table 1.1: Cell surface markers used to identify mouse and human HSCs and progenitors

Table 1.2: Summary of the advantages and disadvantages of different modes and methods of delivery of the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) machinery

Table 2.1: tdTomato induction guide RNA sequences

Table 2.2: tdTomato knockout and non-targeting guide RNA sequences and tdTomato PCR primers

Table 2.3: Cosmid predicted tdTomato guide RNA 3 off-targets and PCR primer sequences

Table 2.4: PCR reagent components for guide RNA amplification from the PX458 plasmid

Table 2.5: PCR program for guide RNA amplification from the PX458 plasmid

Table 2.6: Antibody panel used to stain primary mouse HSPCs for in vitro transfection and editing efficiency analysis

Table 2.7: Antibody panel used for tdTomato time course analysis and sorting of HSPCs for Target-Seq 24h post-nucleofection and RNA-Seq 18wk post-transplant

Table 2.8: Platelet and red blood cell lineage panel for assessing engraftment of tdTomato edited HSCs

Table 2.9: Lymphoid and myeloid lineage panel for assessing engraftment of tdTomato edited HSCs

Table 2.10: Components of lysis buffer for Target-Seq

Table 2.11: Components of Target-Seq reverse transcription master mix

Table 2.12: Target-Seq protease heat inactivation

Table 2.13: Target-Seq reverse transcription PCR program

Table 2.14: Components of Target-Seq PCR master mix

Table 2.15: Target-Seq PCR amplification program

Table 2.16: Components of tdTomato genotyping PCR

Table 2.17: tdTomato genotyping PCR program

Table 2.18: NF-1 exon 2 targeting guide RNA sequences

Table 2.19: NF-1 genotyping PCR components

Table 2.20 NF-1 genotyping PCR program

Table 2.21: Terasaki and 96-well sorting colony counting key

Table 2.22: FACS panel for sorting CD34+ Lin- CB single cells

Table 2.23: Cytokine cocktails used for CD34+ Lin- CB single cell sorting

Table 2.24: Cytokine cocktail for sorting NF-1 and AAVS1 targeted CD34+Lin- CB

Table 2.25: PCR master mix for CB colony genotyping

Table 2.26: PCR program for CB colony genotyping

Table 2.27: Kapa 2G PCR components for CB genotyping

Table 2.28: Antibody panel for peripheral blood analysis of NSG CB recipients

Table 2.29: Antibody panel used to stain the human HSPC compartment of NSG CB recipients

Table 3.1: NF-1 germline initiating mutations genomic and protein regions

Table 3.2: Summary of NSG mouse CB recipients and timeline of disease onset

Table 3.3: White cell, red blood cell and platelet counts for NSG recipients of NF-1 1 and NF-1 4 edited CB at time-point culled sick- 17 and 13 weeks post-transplant respectively

MAIN ABBREVIATIONS

AML	Acute myeloid leukaemia
BM	Bone marrow
Cas	CRISPR-associated
Cascade	CRISPR associated complex for anti-viral defense
CB	Cord blood
CDK	Cyclin dependent kinase
CLP	Common lymphoid progenitor
CML	Chronic myeloid leukaemia
CMP	Common myeloid progenitor
CRISPR	Clustered Regular Interspaced Short Palindromic Repeats
cRNA	CRISPR RNA
dCas9	Catalytically dead Cas9
DS	Double stranded
DSB	Double strand break
dsDNA	Double stranded DNA
eCas9	Engineered Cas9 to reduced sgRNA mis-match off targets
ES	Embryonic stem
GM-CSF	Granulocyte/Macrophage colony stimulating factor
GMP	Granulocyte/Macrophage progenitor
gRNA	Guide RNA
HBG	Normal haemoglobin
HBS	Sickle haemoglobin
HDR	Homology directed repair
HSC	Haematopoietic stem cell
HSPC	Haematopoietic stem and progenitor cell
indels	Insertions or deletions
iPSC	Induced pluripotent stem cell
IT-HSC	Intermediate term HSC
JMML	Juvenile Myelomonocytic leukaemia
LCR	Locus control region
LT-HSC	Long-Term HSC
MEP	Megakaryocyte/Erythrocyte progenitor
MPP	Multipotent progenitor
nCas9	Nickase Cas9
NHEJ	Non-homologous end joining
NK cell	Natural Killer cell
nt	Nucleotide
NT	Non-targeting
PACE	Phage-assisted continuous evolution
PAM	Protospacer Associated Motif
PB	Peripheral blood
pre-cRNA	Long primary transcript of a CRISPR locus

rAAV	Recombinant adeno associated virus
RNA-Seq	RNA sequencing
RNP	Ribonuclear protein
SCF	Stem cell factor
sgRNA	Single guide RNA
spCas9	Cas9 purified from *Streptococcus pyogenes*
SR1	Stem Regenin 1
SS	Single stranded
ssDNA	Single stranded DNA
ssODN	Single stranded oligonucleotide
ST-HSC	Short-Term HSC
tracrRNA	Trans-encoded small RNA
VWF	Von Willebrand factor
xCas9	Engineered Cas9 to increase PAM recognition variety

CHAPTER 1: INTRODUCTION

1: INTRODUCTION

1.1 Normal murine haematopoiesis

1.1.1 Properties and identification of mature blood lineages

Haematopoietic stem cells create all blood lineages of the haematopoietic system. Haematopoiesis is the process by which all blood lineages are produced. In adult mice, this occurs predominately in the bone marrow (BM) in a hierarchical manner; although certain cell types undergo further maturation elsewhere (Spits 2002). At the apex of the haematopoietic hierarchy are stem cells. True haematopoietic stem cells (HSCs) are identified by their capability to create the entire haematopoietic system for the lifetime of the mouse (Laurenti and Göttgens 2018).

Blood production is separated into lymphoid and myeloid arms (**Figure 1.1**) with specific properties and cell surface markers, as illustrated in **Figure 1.2**.

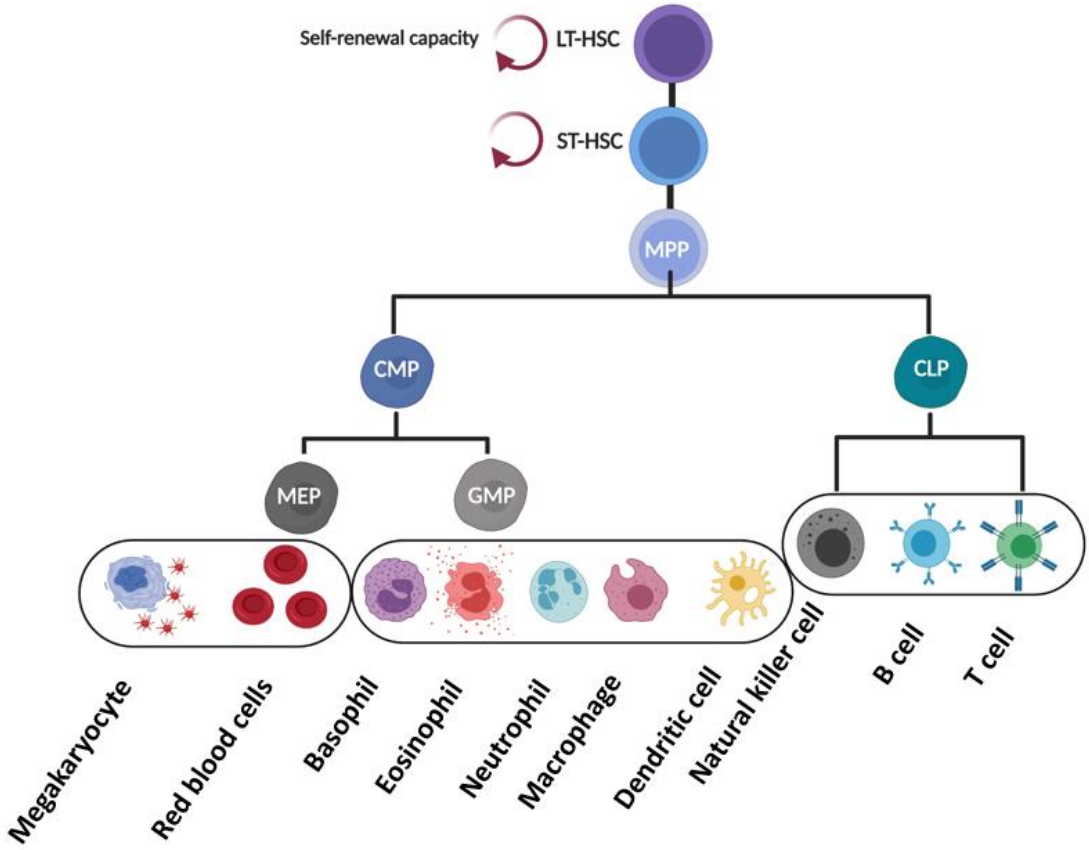


Figure 1.1: Early model of the mouse haematopoietic hierarchy (adapted from (Liu 2018) Created with BioRender.com) A single multipotent progenitor lacking self-renewal capacity generated from the HSC pool immediately branches into two progenitor arms with either lymphoid (common lymphoid progenitor, CLP) or myeloid (common myeloid progenitor, CMP)

The myeloid arm includes megakaryocytes, erythrocytes and macrophages; and the lymphoid arm B cells, T cells and NK cells (Liu 2018). From their niche in the BM, stem cells maintain a homeostatic level of all mature cells of the peripheral blood (PB). The largest output from the BM is that of red blood cells and platelets due to their relatively short lifespan and constant demand (Hoffmann-Fezer, Mysliwietz et al. 1993) (Lebois and Josefsson 2016). Differentiation to these lineage committed cells occurs through a series of branching multi and oligo-potent progenitor fates, with gradually more restricted output, coupled with the

loss of self-renewal (Lu, Czechowicz et al. 2019). Specific cell surface markers (Challen, Boles et al. 2009) (as shown in **Figure 1.2**), morphological properties (Boisset and Robin 2012), and *in vitro* culture assays (Xavier-Ferrucio and Krause 2018) (Gupta, Shah et al. 2014) (Lewis, de Leenheer et al. 2015) (Flaherty and Reynolds 2015) (Lim, Berger et al. 2016) (Vijey, Posorske et al. 2018) allow the identification and propagation of terminally differentiated cells. *In vitro* methods for assessing mature lineages allow the assessment of HSC and progenitor potential.

1.1.2 *In vitro* methods of enriching for murine haematopoietic stem and progenitor cells

Despite their crucial role in maintaining homeostatic haematopoiesis, HSCs constitute a very small proportion (0.01%) (Challen, Boles et al. 2009) of the adult mouse BM. Therefore to unravel their complex function requires careful enrichment and purification.

A notable feature of both mouse and human HSCs is their lack of mature lineage cell surface makers such as Ter 119 (marking red blood cells), CD11b and F4/80 (marking macrophages), CD19 and B220 (marking B cells), CD4 and CD8 (marking T cells) (Challen, Boles et al. 2009). However, a large number of positive enrichment markers have been identified. Cell surface markers which enrich for HSCs include but are not limited to c-Kit, a growth factor receptor; Sca-1, a Ly-6 family antigen; and the glycoprotein receptor SLAM (CD150) (Spangrude, Heimfeld et al. 1988) (Kiel, Yilmaz et al. 2005) (Yilmaz Ö, Kiel et al. 2006) (Oguro, Ding et al. 2013). Applying microarray technology identified

endothelial cell-selective adhesion molecule (ESAM) as a marker of HSCs in mouse fetal liver as well as adult bone marrow primitive progenitors (Yokota, Oritani et al. 2009). In addition, endothelial protein C receptor (EPCR), originally identified as differentially expressed in murine bone marrow HSCs (Ramalho-Santos, Yoon et al. 2002), was used to purify cells from the bone marrow and resulted in high levels of *in-vivo* reconstitution capacity, confirming it as a bone fide enrichment marker for murine stem cells. More recently, CD201 and CD27 have also been identified as markers of LT-HSCs and provide additional labels to use in combination with LSK markers or as alternatives in models of haematopoietic injury where c-Kit can be down-regulated but these markers aren't (Vazquez, Inlay et al. 2015). Despite the wide array of positive enrichment markers, those described above are still not sufficient to guarantee purification of homogenous HSCs with functional long-term re-populating capacity. Ali and colleagues developed a system in which to identify further markers of bona fide functional stem cells (Ali, Fuse et al. 2017). Having established LT-HSCs express the highest levels of the nucleostemnin gene, previously identified in neural stem cells, within the HSPC compartment, they purified high GFP LT-HSCs and found nine genes particularly up-regulated in this group compared to GFP low LT-HSCs: *Tcf15*, *Slamf1*, *Vwf*, *Mpdz*, *Obsl1*, *Sdpr*, *Sult1a1*, *Clec1a*, and *Cdkn1c*. This highlighted molecular heterogeneity even within the phenotypic LT-HSC compartment, a subset of the HSPC population. Interestingly, CD34, a surface sialomucin-like adhesion molecule which enriches for human haematopoietic stem cells (Civin, Strauss et al. 1984) is absent from the most primitive mouse HSCs, as determined by the lack of long-term re-populating capacity of CD34+ mouse bone marrow (Osawa, Hanada et al. 1996). Human HSC markers are

discussed in more detail in *1.2 Normal human haematopoiesis*. Progress in identifying cell populations enriched in HSCs and sub-types of HSCs has come about through a combination of multi-parameter fluorescence-activated cell sorting (FACS), clonogenic assays and transplantation experiments. These have allowed selection of mouse HSCs with higher purity by including additional markers that distinguish HSCs from multipotent progenitors (MPPs, 0.055% of total mouse bone marrow) (Seita and Weissman 2010) (Challen, Boles et al. 2009). Functional HSCs were identified by their ability to produce myeloid, erythroid and megakaryocytic lineages from single cells *in vitro,* whereas MPPs showed lower clonogenicity and mainly myeloid output (Yilmaz Ö, Kiel et al. 2006). Currently, the most accepted combination of surface markers to enrich for mouse HSCs is: lineage negative (Lin-, a cocktail of terminally differentiated lineage markers which are excluded) c-Kit+, Sca-1+, CD48-, CD150+, CD41-, Rho123^{lo}, CD34-/lo, Flt3R-, CD49^{lo}, CD244-, CD229- (Ng and Alexander 2017).

1.1.3 *In vivo* methods for identifying long-term murine HSCs

Functional stem cells are identified by their ability to contribute to all blood lineages in primary and secondary transplants. Although mouse HSCs can be enriched through antibody selection with the markers described above, this only identifies these cells phenotypically and they remain a heterogeneous population in terms of their lineage potential. Functional verification is required to confirm their long-term stem cell capability. A functional mouse HSC is capable of reconstituting all blood lineages in a lethally irradiated recipient mouse for at least four months post-transplantation (Challen, Boles et al. 2009). Upon purification from a primary recipient, true long-term mouse HSCs are also capable of reconstituting all blood lineages in secondary recipients (Rosendaal, Hodgson et

al. 1979, Sanjuan-Pla, Macaulay et al. 2013). These are termed long-term (LT)-HSCs. These transplantation methods conventionally use the pan-haematopoietic surface marker CD45 which has two isotypes, CD45.1 and CD45.2, which are very useful experimentally to distinguish donor from recipient in engraftment studies, e.g. to identify which lineages the transplanted HSC has contributed to (Jafri, Moore et al. 2017). However, this method cannot be used to determine donor-derived platelets and erythrocytes as they lack this surface marker. Common approaches to overcome this use erythrocyte and platelet inclusive transgene fluorescent markers, expressed by platelets and erythrocytes as well as other blood lineages (Mansson, Hultquist et al. 2007) (Suzuki, Moriguchi et al. 2009).

1.1.4 Maintenance of a homeostatic balance in blood production: asymmetric vs symmetric division

Haematopoietic stem cells maintain the balance of lineage production and allow the blood system to adapt to stress or environmental change. They do this by controlled levels of quiescence (G_0 cell cycle phase) and asymmetric versus symmetric division. Asymmetric and symmetric self-renewal cell division is a unique property of stem cells (Höfer and Rodewald 2016). Through symmetric cell division they are able to self-renew and thus maintain the quiescent stem cell pool. Asymmetric division also maintains the stem cell pool while at the same time generating a more differentiated haematopoietic cell. Similarly, symmetric division can also produce two daughter cells of a more differentiated state. In contrast to more lineage committed cells, stem cell proliferation is low with the majority found in G_o of the cell cycle (70% at any one time) (Wilson, Laurenti et al. 2008) and total numbers limited. However, in times of stress, such as tissue

damage or infection these quiescent cells can be rapidly recruited into the cell cycle and produce the required lineage (Baldridge, King et al. 2010). Through asymmetric division, HSCs are able to commit to specific lineage trajectories down the differentiation hierarchy.

1.1.5 Maintenance of a homeostatic balance in blood production: dynamic and fluid commitment of the HSC compartment to the myeloid and lymphoid lineages

Differentiation from an HSC to a mature lineage at the base of the trajectory is complex and dynamic, without one single path leading from HSC to specific mature blood lineage. The dynamic and fluid nature of this process allows the haematopoietic system to effectively respond to stress and changes in the environment.

As illustrated in **Figure 1.1**, initial studies suggested the differentiation process separating HSCs into progenitors, committed to the myeloid or lymphoid lineages, occurred in a stepwise manner through a multipotent progenitor (MPP) which then bifurcated into a common lymphoid progenitor (CLP) (Kondo, Weissman et al. 1997) or common myeloid progenitor (CMP) (Akashi, Traver et al. 2000)**.** It was shown through *in vitro* differentiation and transplant assays that CLPs give rise exclusively to B, T and NK cells and lack long-term repopulation capacity (Kondo, Weissman et al. 1997). Similarly, CMPs, which were separated further to granulocyte/macrophage progenitors (GMPs) and megakaryocyte/erythrocyte progenitors (MEPs), produced exclusively

granulocytes, macrophages, megakaryocytes and erythrocytes, with an equal lack of long-term lineage maintenance *in vivo* (Akashi, Traver et al. 2000).

More recent studies have shown that, although committed progenitors are present in the haematopoietic hierarchy, the trajectory is more complex (Schulte, Wilson et al. 2015) (Nestorowa, Hamey et al. 2016) and this can alter in response to stress (Lu, Czechowicz et al. 2019). One approach to investigating this was through single cell transplantation assays of paired 'daughter' cells (Seita, Ema et al. 2007) which showed that the HSC population can be separated into LT-HSCs, short-term (ST) HSCs and intermediate-term (IT) HSCs (Weissman and Shizuru 2008) (Yamamoto, Morita et al. 2013). More importantly, this approach demonstrated that myeloid, erythroid and megakaryocyte re-populating progenitors could be produced directly from the HSC compartment. Another notable finding was that, upon transplantation, platelets tended to be the first lineage detected as well as the first lineage lost when HSC self-renewal capacity became reduced, as indicated by a lack of contribution to PB lineages, thus a strong indicator of stem cell self-renewal activity (Yamamoto, Morita et al. 2013).

The concept of a bifurcating lineage trajectory has been further challenged with the identification of branching populations of MPPs (Pietras, Reynaud et al. 2015). Single cell RNA-sequencing (RNA-Seq) data has also complemented these findings showing that lineage commitment is a highly dynamic process (Paul, Arkin et al. 2015) (Wilson, Kent et al. 2015) (Dahlin, Hamey et al. 2018). By sorting MPPs based on CD150, CD48 and Flk2 expression, Pietras and colleagues identified, through liquid culture assays, MPPs which, while they were biased towards specific subsets of myeloid (MPP2, MPP3) and lymphoid (MPP4) cell production, still retained the propensity to produce the other cell types though

at much lower levels. MPP4 are also referred to as lymphoid-primed multipotent progenitors (LMPP). Jacobsen and colleagues showed that while they demonstrated myeloid potential, LMPPs lacked the capacity to produce cells down the megakaryocyte or erythroid lineage branches (Adolfsson, Mansson et al. 2005). *In vitro*, a reduced plating efficiency distinguished multipotent progenitors functionally from LT and ST-HSCs; they showed a clearly diminished propagation capacity. Their lack of self-renewal was confirmed *in vivo* by the lack of engraftment past four weeks post-transplantation (Pietras, Reynaud et al. 2015). The lineage biases seen *in vitro* were also replicated in the *in vivo* setting. Further, gene expression analysis by microarray separated the HSCs, MMP2/3 and MPP4 by cell cycle and myeloid/lymphoid gene signatures, although MPP4 portrayed poised myeloid gene expression which was complemented *in vivo* by a contribution to the initial myeloid reconstitution storm seen immediately post-transplantation. **Figure 1.2** summarises the current model of the mouse haematopoietic hierarchy indicating which compartments possess self-renewal capacity and the trajectories by which each lineage can be produced. Identifying cell surface markers are also summarised (Yamamoto, Morita et al. 2013) (Pietras, Reynaud et al. 2015) (Kondo, Weissman et al. 1997, Akashi, Traver et al. 2000).

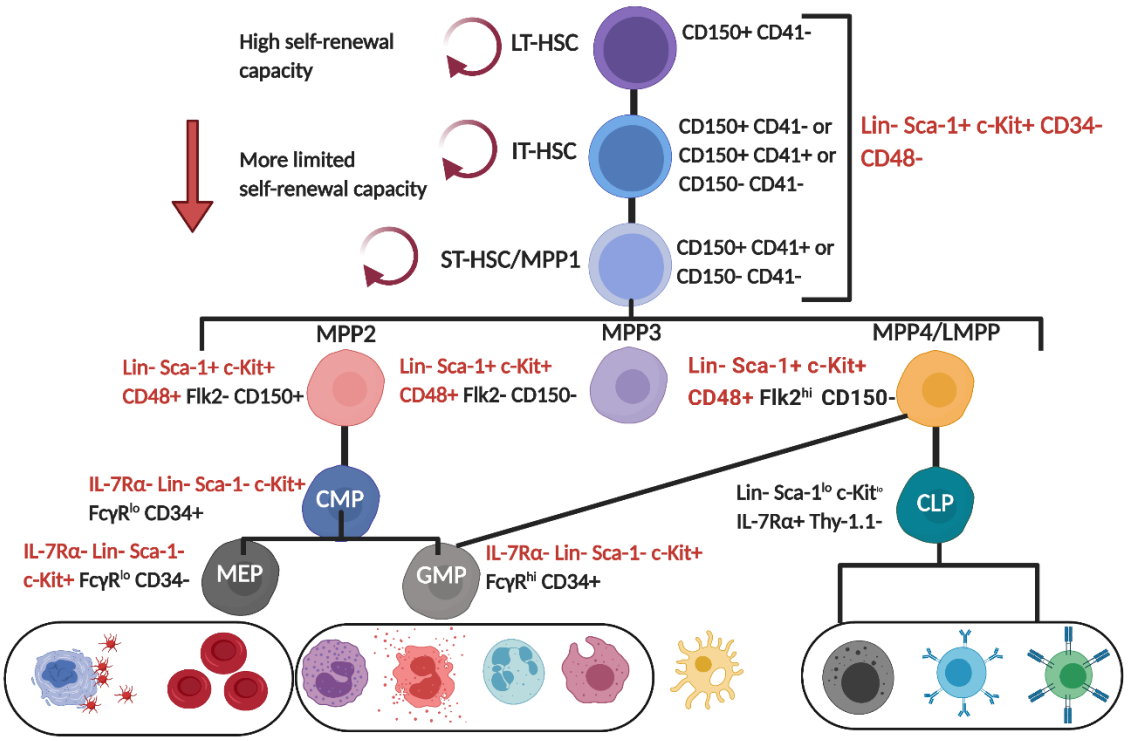


Figure 1.2: Current model of the mouse haematopoietic hierarchy (adapted from (Liu 2018) Created with BioRender.com) Cell surface markers used to enrich for each sub-population are labelled. Markers common to several populations are labelled in red, with unique markers in black. Self-renewal capacity is indicated by red arrows. Haematopoietic stem cell (HSC), Long-term HSCs (LT-HSC), intermediate-term HSC (IT-HSC), short-term HSC (ST-HSC), multipotent progenitor (MPP), lymphoid-primed multipotent progenitor (LMPP), common myeloid progenitor (CMP), common lymphoid progenitor (CLP), megakaryocyte-erythroid progenitor (MEP), granulocyte-macrophage progenitor (GMP)

In vitro methods for separating haematopoietic progenitors involve flow cytometry analysis/sorting based on the presence and/or absence of specific cell surface markers. In this way the relative abundance of particular progenitors in the BM can be deduced and individual types of HSC and their progenitors can be isolated based on their pattern of cell surface marker expression. **Table 1.1** summarises the cell surface markers used to identify the populations present in the mouse and human BM HSPC compartments for direct comparison.

Cell type	Mouse cell surface markers for enrichment	Human cell surface markers for enrichment
LT-HSC	Lin- Sca-1+ c-Kit+ CD34- CD48- CD150+ CD41-	Lin- CD34+ CD38- CD90+ CD45RA- CD49f+ Rho^{lo}
IT-HSC	Lin- Sca-1+ c-Kit+ CD34- CD48- (CD150+ /CD41- **or** CD150+/CD41+ **or** CD150-/CD41-)	**MPP:** Lin- CD34+ CD38- CD45RA- CD90- CD49f-
ST-HSC/MPP1	Lin- Sca-1+ c-Kit+ CD34- CD48- (CD150+/CD41+ **or** CD150-/CD41-)	
MPP2	Lin- Sca-1+ c-Kit+ CD48+ Flk2- CD150+	
MPP3	Lin- Sca-1+ c-Kit+ CD48+ Flk2- CD150-	
MPP4/LMPP	Lin- Sca-1+ c-Kit+ CD48+ $Flk2^{hi}$ CD150-	Lin- CD34+ CD38- CD90-/lo CD45RA+ CD10-
CLP	Lin- $Sca\text{-}1^{lo}$ $c\text{-}Kit^{lo}$ IL-7Rα+ Thy-1.1-	Lin- CD34+ CD38-/lo CD45RA+ CD90-
CMP	IL-7Rα- Lin- Sca-1- c-Kit+ $Fc\gamma R^{lo}$ CD34+	Lin- CD34+ CD38+ CD45RA-
GMP	IL-7Rα- Lin- Sca-1- c-Kit+ $Fc\gamma R^{hi}$ CD34+	Lin- CD34+ CD38+ CD45RA+
MEP	IL-7Rα- Lin- Sca-1- c-Kit+ $Fc\gamma R^{lo}$ CD34-	Lin- CD34+ CD38lo CD45RA-

Table 1.1: Cell surface markers used to identify mouse and human HSCs and progenitors

1.1.6 Maintenance of a homeostatic balance in blood production: heterogeneity in quiescence and functional output

The HSC compartment is heterogeneous in terms of quiescence, lineage output, and transcriptional profile. This has implications for better understanding haematological disease and developing targeted therapies. Indeed, many leukaemias arise due to a block in differentiation of a certain lineage (Sell 2005), as well as overproduction of a particular lineage/progenitor (Petrov, Suntsova et al. 2016). Cancer stem cells are responsible for disease propagation and can lead to therapy resistance and relapse (Li, Rogoff et al. 2015) (Flemming 2015);

so understanding the natural heterogeneity of the stem cell compartment is important for modelling leukaemias and developing targeted therapies.

HSCs with identical cell surface markers display variation in their level of quiescence and lineage contribution. Indeed, there exist HSCs with balanced contribution to myeloid and lymphoid lineages, as well as those with lymphoid, myeloid, and platelet biased outputs (Muller-Sieburg, Cho et al. 2004) (Dykstra, Kent et al. 2007) (Benz, Copley et al. 2012) (Sanjuan-Pla, Macaulay et al. 2013) (Carrelha, Meng et al. 2018) . Similarly, LT-HSCs, ST-HSCs and IT-HSCs differ in their lifespan and transplantation kinetics (Iscove 2010) (Benveniste, Frelin et al. 2010) (Ema, Morita et al. 2014).

It has been shown that myeloid-biased HSCs display a delayed contribution to PB lineages in primary transplants but maintain repopulation capacity for an extended number of serial transplants compared to lymphoid-biased HSCs (Muller-Sieburg, Cho et al. 2004). In a study using single cell transplant assays, Dykstra and colleagues showed that LT-HSCs can be grouped into three more neutral groups depending on lineage deficiency: balanced, myeloid deficient and lymphoid deficient. Similarly, it was found that these lineage imbalances were maintained through serial transplant but with myeloid deficient HSCs losing their propagation ability earlier (Dykstra, Kent et al. 2007). More recent studies have also shown that mouse HSCs can be primed towards a specific lineage. Mouse BM cells purified using the stem cell surface markers described above with the addition of von Willebrand factor (Vwf) can reconstitute all PB lineages but with a higher contribution to the platelet lineage than Vwf negative HSCs in primary and secondary transplants (Sanjuan-Pla, Macaulay et al. 2013). Further, single stem cell transplantation studies revealed the presence of a small proportion of

platelet restricted HSCs with uni-lineage output or platelet restriction in combination with retention of erythroid, myeloid and B cell potential. (Carrelha, Meng et al. 2018).

1.1.7 Maintenance of a homeostatic balance in blood production: heterogeneity in transcriptional signature

Transcriptional profiling has shown that HSCs with both different levels of quiescence and lineage output display distinct transcriptional signatures. RNA-Seq has also demonstrated the complex dynamic lineage branching that initiates close to the apex of the haematopoietic hierarchy.

By combining index sorting with single cell gene expression profiling and transplantation, it has been shown that the level of Sca-1 expression (high/low) can be used to purify more quiescent/cycling HSCs respectively, with Sca-1 high producing more robust engraftment through mainly HSCs with balanced myeloid and lymphoid lineage output (Wilson, Kent et al. 2015). Further, microarray transcriptome profiling has identified transcriptional signatures that can segregate ST and LT-HSC based on their gene expression, with significant overlap with human HSC signatures (Ivanova, Dimos et al. 2002). More recent barcoding techniques have also further elucidated stem-cell fates with the ability to map lineage restricted clones to specific transcriptional signatures, by barcodes which can track the cells (Pei, Shang et al. 2020). In addition to this, singe cell RNA-Seq has revealed subtle intrinsic differences in young and old HSCs which may explain the changes in lineage bias seen in ageing. For example it has been shown that old Vwf+ HSCs maintain a higher level of

expression of platelet and myeloid associated lineage genes compared to young HSCs, despite their equivalent surface markers; which is matched by their marked higher contribution to the platelet lineage in single cell transplantation, concomitant with a marked reduction in lymphoid reconstitution (Grover, Sanjuan-Pla et al. 2016). In addition, it has been shown that in fact myeloid-restricted re-populating progenitors (MyRPs) can bypass the multipotent to lymphoid/myeloid bifurcation in the traditional haematopoietic hierarchy and arise directly from the HSC compartment (Yamamoto, Morita et al. 2013). Interestingly, the authors subsequently showed that levels of these MyRPs increased with age (Yamamoto, Wilkinson et al. 2018).

As discussed above*, in vitro* culture and *in vivo* transplantation assays have identified heterogeneous sub-populations of HSCs which differentiate into various different multi-, bi- or uni-potent haematopoietic lineages. With the advances in single cell transcriptome profiling technologies, gene expression data has allowed further dissection of this differentiation process. The pseudo time analysis generated by these types of data has allowed trajectories to be identified, modelling the precursors that specific lineages go through to reach the terminally differentiated state (Nestorowa, Hamey et al. 2016).

More recently, in depth single cell transcriptional profiling within the LSK (Lin- Sca-1+ c-Kit+) and LK (Lin- c-Kit+) populations has revealed as many as eight branch points within the HSPC compartment specified by their enrichment in gene expression for specific lineage markers such as *Rag2* (lymphoid), *Klf1* (erythroid) and *Ly86* (monocytes) (Dahlin, Hamey et al. 2018). This supports the model that branching occurs early in the haematopoietic hierarchy at the level of the HSC/MPP compartment.

1.2 Normal human haematopoiesis

1.2.1 The properties of human haematopoietic stem cells (HSCs): their function and identification

In a similar fashion to mouse HSCs, human HSCs are defined by the capacity to produce all blood lineages, while also retaining self-renewal potential (Rieger and Schroeder 2012). Complementing research in the mouse, similar approaches have been taken to establish cell surface enrichment markers, gene expression patterns, and functional characteristics to identify human HSCs.

CD34 was the first cell surface marker to be identified as enriching for immature human haematopoietic cells (Civin, Strauss et al. 1984). Indeed Lapidot and colleagues found that only the CD34+ CD38- proportion of leukaemia initiating cells of an AML patient peripheral blood could engraft into immune deficient non-obese diabetic/severe combined immunodeficient (NOD/SCID) mice and successfully produce colony forming progenitors (Lapidot, Sirard et al. 1994). Similarly, using normal human cord blood (CB) and BM transplantation into (NOD/SCID) mice, Bhatia and colleagues demonstrated that, within the lineage depleted CD34+ compartment, only CD38- cells could achieve long-term repopulation of blood lineages (Bhatia, Wang et al. 1997). Further, analysis of the mouse BM eight weeks post-transplantation of human cells revealed the presence of human CD34+ CD38- cells, demonstrating their self-renewal, as well as proliferative capacity.

Mayani and colleagues elucidated additional markers of stem cell self-renewal using long-term serum free liquid culture assays, using cytokines promoting either myeloid or erythroid output (Mayani, Dragowska et al. 1993). They

demonstrated that low or no expression of CD45RA and CD71 marked HSPCs with higher clonogenicity, greater ability to sustain CD34 positivity long-term, and capability to contribute to myeloid and erythroid lineages. In contrast, CD45RA+CD71^{lo} and CD45RA^{lo}CD71+ had reduced capabilities of CD34+ retention and contributed predominantly to myeloid or erythroid colony output respectively. Notta and colleagues used stringently sorted populations of CB HSPCs based initially on Thy1 and subsequently Thy1 (CD90)/CD49f to establish Thy1+ as an enrichment, but not exclusive, marker for HSCs, with the Thy1- fraction of CD34+CD38-CD45RA- cells still capable of long-term re-population but at a lower efficiency (Notta, Doulatov et al. 2011). Sorting based on the integrin CD49f revealed this as an enrichment marker for HSCs, loss of which in the Thy1- fraction identified MPPs. Similar to mouse HSCs, human HSCs with long-term re-populating capacity are rare in the bone marrow, comprising about 0.01% (Brendel, Goebel et al. 2015). Taken together, these studies have established methods to enrich for human HSCs although, even with these stringent selection procedures, they remain a heterogeneous population whose true long-term re-populating ability can only be determined by long-term engraftment in xenograft recipients, equivalent to those required to confirm mouse HSCs.

1.2.2 Evidence for heterogeneity within the human HSC compartment

Similar to murine haematopoiesis, it has become evident that a trajectory in which a multipotent stem cell, residing at the apex of the haematopoietic tree, gradually loses self-renewal capacity through a distinct set of cell fate decisions, from bi-

potent to uni-potent potential, is not in-fact an accurate depiction of what occurs (Laurenti and Göttgens 2018). In contrast, experimental evidence generated predominantly from single cell molecular and functional assays has recently led to the understanding that the HSC pool itself is heterogeneous, in which fluid transition between myeloid and lymphoid fates can occur. Specific lineage priming can occur at very primitive stages, and the balance of self-renewal and differentiation changes both through ontogeny and in response to intrinsic/extrinsic cellular stress (Velten, Haas et al. 2017). Velten and colleagues used index sorting of single BM HSPCs, followed by transcriptomic analysis and *in vitro* functional profiling to elucidate the pathway from multipotency to lineage commitment in detail. Strikingly, they observed that cells within the Lin-CD34+CD38+ compartment progressed in discrete bifurcations towards unipotency, similar to previously adopted models. Whereas, the Lin-CD34+ CD38- compartment segregated into at least ten distinct populations by RNA-Seq, and were much more fluid in their ability to adopt different mature lineage trajectories. Interestingly, some single cells in which the index phenotype corresponded to a bi-potent progenitor, were already primed to uni-potency at the gene expression level, indicating that lineage commitment occurs at a more primitive stage than previously thought (Velten, Haas et al. 2017). Similarly, Psaila and colleagues demonstrated in the megakaryocyte erythroid progenitor (MEP), using single cell transcriptional profiling and differentiation assays, that cells with an identical immunophenotype by flow cytometry can be further segregated into three distinct populations, with different lineage biases (Psaila, Barkas et al. 2016).

1.2.3 Evidence for distinct developmental differences in haematopoiesis: changes in site, gene expression profile and lineage output from the HSC compartment between adult and fetal life

Of particular importance to better understanding childhood leukaemias, is evidence that distinct haematopoietic patterns occur across ontogeny. This has implications for identifying the cell of origin of these diseases, as well as therapy approaches. In contrast to adult blood production, which relies on the haematopoietic niche in the BM, during embryonic development haematopoietic output is segregated to different tissues in a temporal manner. The first two waves, which occur in the yolk sac and then liver, give rise to large, nucleated erythrocytes, and erythro-myeloid/lymphoid progenitors respectively. These stages are short-lived and lack self-renewing HSCs. Self-renewing HSCs are first seen in the aorta-gonad-mesonephros (AGM) region of the developing embryo, followed by the yolk sac and fetal liver (FL). The BM, placenta and umbilical cord are the final regions to be colonised, concomitant with blood vessel protrusion into the BM (Ivanovs, Rybtsov et al. 2017). Although all these tissue-associated HSCs possess CD34 positivity, other surface markers vary between the groups, including endothelial marker VE-Cadherin, which is expressed in the earlier stage HSCs.

Further, HSCs at different developmental stages possess distinct lineage programs, with different biases, and unique gene expression profiles (Popescu, Botting et al. 2019). Using an extensive flow cytometry panel and single cell *in vitro* colony assays, Notta and colleagues revealed that HSCs (CD34+CD38-) and MPPs (CD34+CD38+) with identical surface markers possessed different clonogenic and lineage potential at different stages of ontogeny (Notta, Zandi et

al. 2016). Clonogenicity decreased from FL through CB and adult BM. Interestingly, multi-lineage potential existed in both the HSC and MPP compartments for FL and CB but only in the HSC compartment for BM. Further, bi-potency was a feature almost exclusive to FL and CB whereas BM was characterised by an either multi or uni-potent output. Psaila and colleagues also noted bi-potency as a rare occurrence within the adult MEP compartment (Psaila, 2016).

Corroborating evidence from mouse developmental haematopoiesis supports the observation that the re-populating capacity and lineage output of HSCs is strictly controlled across ontogeny. Transplantation of mouse E18.5 FL, 3 and 4wk post-natal BM, and adult BM showed that HSC expansion in recipient mice was much more rapid for FL and 3wk BM compared to 4wk and adult BM, most likely due to different rates of symmetric vs asymmetric division; and that this switch was intrinsic as opposed to influenced by the environment or co-transplanted cells (Bowie, Kent et al. 2007). This switch was coupled with increased gene expression of *Ezh2*, *Atm* and *Gata-2* between post-natal weeks three and four. Further, fetal and 3wk post-natal BM derived HSCs contributed at a higher level to monocyte and granulocyte engraftment compared with 4wk post-natal and adult BM. Similarly, Benz and colleagues observed that three subtypes of mouse HSCs: alpha (lymphoid deficient), beta (balanced myeloid/lymphoid output) and gamma (myeloid deficient) changed in their proportions across ontogeny with beta HSCs dominating in embryonic FL and BM (Benz, Copley et al. 2012).

In line with a HSC self-renewal and lineage program that is intrinsically primed, which was suggested by single cell transcriptomic studies in human FL (Popescu, Botting et al. 2019), and controlled across ontogeny, several studies have

established the presence of distinct transcriptional differences between fetal and adult HSCs. Kim and colleagues compared gene expression profiles of fetal, neonatal and adult HSCs which identified Sox17 as a transcriptional regulator expressed in both fetal and neonatal but not adult HSCs (Kim, Saunders et al. 2007). Importantly, this study was followed up by an investigation in which ectopic expression of Sox17 was tested for its ability to convert adult HSCs into more primitive like cells (He, Kim et al. 2011). Indeed, it was found that adults HSCs with induction of Sox17 expression showed increased self-renewal potential, characteristic of a more fetal like HSC, as well as induction of fetal HSC gene expression programs and presence of fetal HSC surface markers (He, Kim et al. 2011). Copley and colleagues took a similar approach studying the *Lin28b*, *Let7*, *Hmga2* axis using murine cells as a model system (Copley, Babovic et al. 2013). They observed increased expression of *Lin28b* and *Hmga2*, concurrent with decreased expression of *Let7*, in FL compared to adult BM. *Lin28b*, *Let7* and *Hmga2* are part of a transcriptional network; *Lin28B* is a gene involved in self-renewal of which upregulation induces down regulation of *Let7*, a micro-RNA. *Let7* reduction leads to upregulation of its gene targets, including *Hmga2*. *Hmga2* is a transcriptional regulator which binds DNA and induces changes in chromatin structure. Indeed, transduction of adult BM with *Lin28b* or *Hmga2* overexpression constructs, followed by transplantation, led to an increase in engraftment without changing the proportion of lineage contribution compared to control, suggesting an alteration in the symmetric division capacity towards that of fetal HSCs (Copley, Babovic et al. 2013). Importantly, this transcriptional axis has been observed to be aberrantly active in certain childhood leukaemias, including Juvenile Myelomonocytic Leukaemia (JMML).

1.3 Juvenile Myelomonocytic Leukaemia (JMML)

1.3.1 Interaction of haematopoietic cells with the microenvironment: haematopoietic development and cancer surveillance

A crucial part of maintaining the appropriate levels of production of specific blood lineages as well as sustaining the appropriate levels of stem cell quiescence comes from the interaction between haematopoietic stem cells and their microenvironment (Galán-Díez, Cuesta-Domínguez et al. 2018). In adults, the bone marrow provides this microenvironment whereas in fetal development this changes as the haematopoietic system first colonies the AGM, followed by fetal liver and bone marrow. Indeed, there is evidence that the fetal microenvironment plays a crucial role in the development of infant and childhood leukaemias (Cazzola, Cazzaniga et al. 2020). Within the niche where the HSCs reside, they receive chemokine and cytokine signals as well as interacting directly with specific cells within the niche. For example, mesenchymal stromal cells (MSCs) reside in the bone marrow niche in close proximity with HSCs and contribute to HSC maintenance by producing factors such as stem cell factor (SCF). Indeed, leukemic cells have the capability to re-program MSCs to promote a leukemic microenvironment, by direct cell contact as well as production of chemokines including chemokine ligand 3 (CCL3) and thrombopoietin (TPO). **Figure 1.3** summarises the key interactions of HSCs within the BM microenvironment.

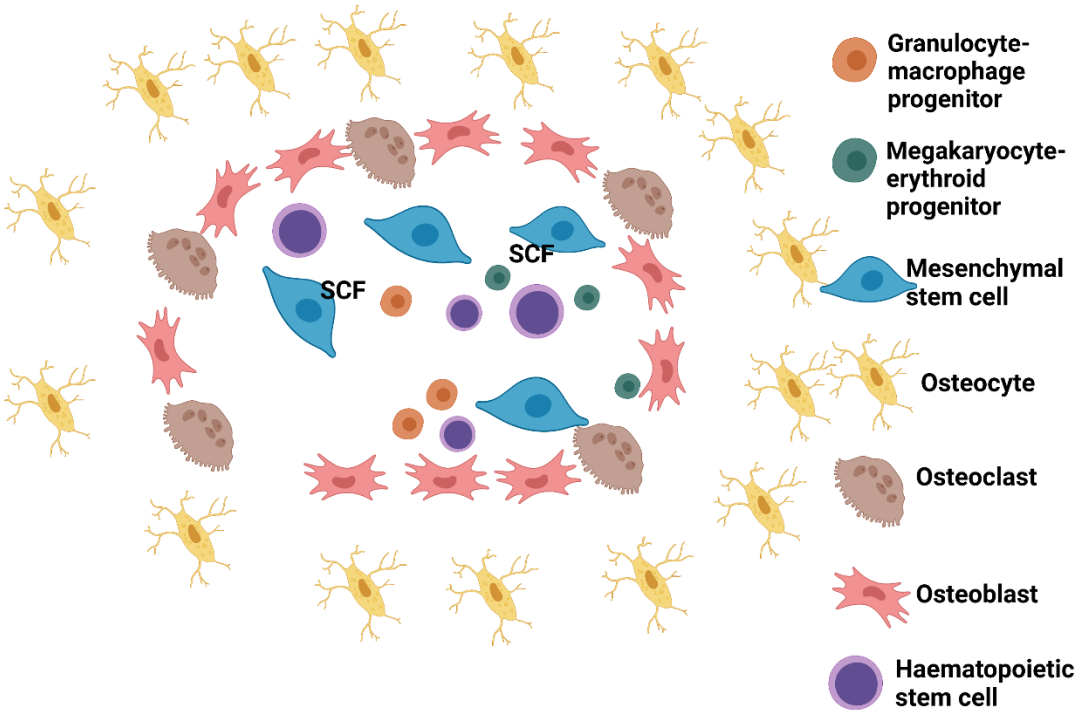


Figure 1.3: Schematic of haematopoietic stem cells (HSCs) within the bone marrow microenvironment HSCs reside in the bone marrow niche surrounded by osteocytes, osteoblasts and osteoclasts. They interact closely with mesenchymal stromal cells (MSCs) which produce important factors for stem cell maintenance including stem cell factor (SCF). Created using Biorender.com. Adaptor from (Galán-Díez, Cuesta-Domínguez et al. 2018)

1.3.2 Clinical and haematological features of JMML

JMML is a rare aggressive childhood leukaemia with poor prognosis. The only curative therapy is BM transplant but relapse can occur due to residual leukaemic stem cells. The main clinical features of JMML are the presence of leukaemic skin lesions and leukaemic infiltrates in the lung, liver and spleen. PB analyses show leucocytosis dominated by monocytosis, immature granulocytes and erythroid precursors but also the presence of small numbers of blast cells (Niemeyer and Flotho 2019). In the BM, there is also a moderate increase in blasts as well as reduced megakaryocytes, causing thrombocytopenia in

patients. Although a small number of patients spontaneously resolve (Matsuda, Shimada et al. 2007), for those who don't the prognosis is poor compared to some other childhood leukaemias. Not only this, it is also a relatively difficult leukaemia to diagnose (Liao, Qiu et al. 2019). For this reason, extensive research has, and is, being carried out to better characterise the genetic landscape (Caye, Strullu et al. 2015) (Stieglitz, Taylor-Weiner et al. 2015), cell of origin (Caye, Rouault-Pierre et al. 2020) (Louka, Povinelli et al. 2021), and properties of the leukaemic cells (Emanuel, Bates et al. 1991) (Donovan, See et al. 2002) (Chan, Leedy et al. 2005) (Schubbert, Lieuw et al. 2005) (Metzner, Horstmann et al. 2007) (Kotecha, Flores et al. 2008) (Niemeyer, Kang et al. 2010); as well as to develop mouse models which more closely recapitulate the human disease (Mulero-Navarro, Sevilla et al. 2015) (Krombholz, Aumann et al. 2016) (Yoshimi, Balasis et al. 2017) (Wunderlich, Chou et al. 2018) (Caye, Rouault-Pierre et al. 2020) (Louka, Povinelli et al. 2021)

1.3.3 Molecular pathogenesis of JMML

JMML may arise *de novo* due to a somatic initiating event or secondary to germline mutations (Caye, Strullu et al. 2015). In both cases, the clinical outcome is largely dependent on the presence and nature of the co-operating additional mutations. All JMML initiating mutations affect genes that are part of the RAS signalling pathway. Three cancer genetic susceptibility syndromes are associated with JMML (McGee and Nichols 2016): Noonan syndrome (NS) (Tartaglia and Gelb 2005) (Schubbert, Zenker et al. 2006), Neurofibromatosis syndrome (*NF-1*) syndrome (Flotho, Steinemann et al. 2007), and *CBL* syndrome

(Niemeyer, Kang et al. 2010). In the case of these three syndromes, they all harbour mutations in regulators of the RAS pathway and have a variable risk of developing JMML depending on the genes involved and the nature of the mutations.

It is clear from clinical data that not all JMML patients have the same prognosis or respond in the same way to therapies. Therefore Caye, Stieglitz, and colleagues carried out genetic and transcriptomic profiling of large cohorts of patients and correlated the mutational landscape with disease outcomes (Caye, Strullu et al. 2015) (Stieglitz, Taylor-Weiner et al. 2015). In a cohort of 118 patients, Caye and colleagues observed that all initiating mutations occurred in components of the RAS pathway (19% *NRAS*, 15% *KRAS*) or regulators of the RAS pathway (50% *PTPN11*, 9% *CBL*, 7% *NF-1*). *KRAS*, *NRAS* and *PTPN11* harboured gain of function mutations (Kratz, Niemeyer et al. 2005) (Niemeyer 2018) whereas *NF-1* harboured loss of function mutations (de Vries, Zwaan et al. 2010). Interestingly, *CBL* mutations converted the tumour suppressor to a proto-oncogene (Shiba, Kato et al. 2010). **Figure 1.4** summarises the initiating mutational landscape of JMML.

In the case of *NRAS*, all mutations were somatic, almost all being heterozygous. In the Caye cohort, *KRAS* mutations were somatic heterozygous. *PTPN11* could either be syndromic germline, all of which were heterozygous; or somatic, including a small number of patients with homozygous mutations. Germline and somatic *NF-1* mutations were evident and were almost exclusively heterozygous, although there was one patient with a germline deletion and somatic mutation on the other allele. *CBL* was exclusively germline, mostly homozygous (Caye, Strullu et al. 2015). Some, but not all, patients were seen to harbour additional

secondary mutations; these were almost all in additional RAS pathway genes or in components of the polycomb repressive complex 2 (PRC2), involved in depositing repressive chromatin marks. In some cases, loss of heterozygosity or partial chromosome deletions lead to loss of the wild type allelic counterpart to the initiating mutation. In all cases, presence of secondary mutations was predictive of poor prognosis (Caye, Strullu et al. 2015).

Corroborating evidence from tracking a cohort of 100 patients before and after therapy showed that the prevalence of additional somatic mutations at diagnosis was predictive of poor clinical outcome (Stieglitz, Taylor-Weiner et al. 2015). Additionally, secondary mutations in PRC2 complex and spliceosome genes (*EZH2*, *ASXL1*, and *ZRSR2*) were found to be common. An additional secondary mutation, not found in the Caye cohort, was *SH2B3*- a negative regulator of JAK-STAT signalling. This secondary mutation was found to be a predictor of particularly poor clinical outcome, and in one case was associated with transformation to AML (Stieglitz, Taylor-Weiner et al. 2015).

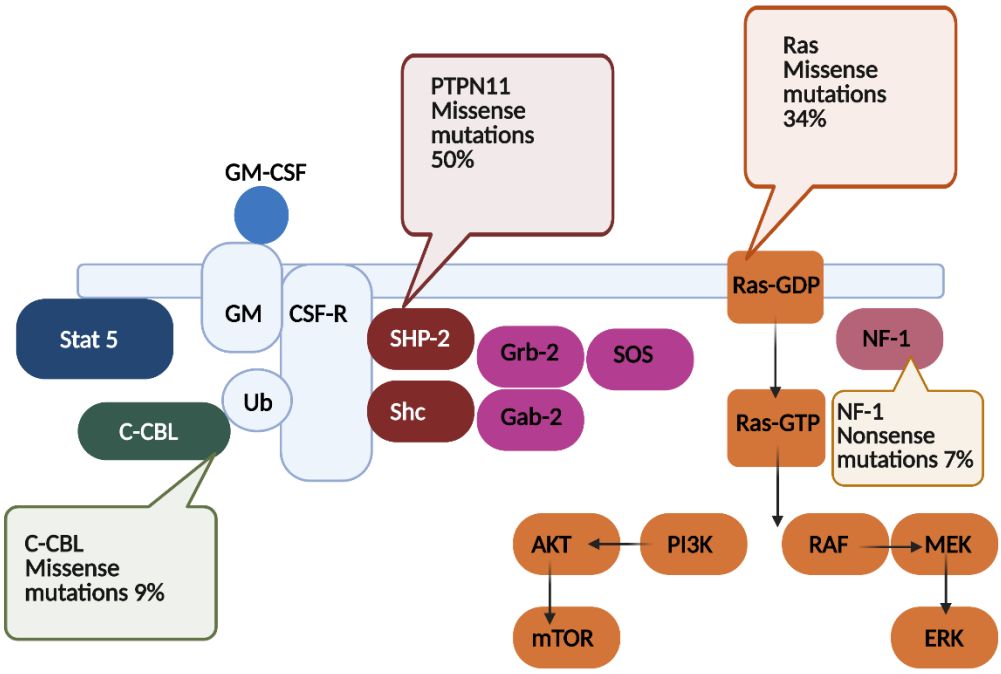


Figure 1.4: Genetic landscape of initiating mutations in JMML (Created with BioRender.com) All initiating mutations are regulators (*PTPN11*, *CBL*, *NF-1*) or components (*NRAS*, *KRAS*) of the RAS pathway. Mutations lead to over activation of the RAS signalling pathway.

1.3.4 Evidence for a fetal origin of JMML

Transcriptional profiling and clonal mutation tracking has provided considerable evidence suggestive of a fetal stem cell origin of JMML. One of the strongest pieces of evidence of a fetal origin is the persistence or re-expression of fetal haemoglobin which normally declines to <1-2% of total haemoglobin by six months of age (Loh 2011). While evidence of a stem cell origin has been provided both by mouse models which show the presence of identical genetic aberrations in multiple cell lineages not limited to the disease expanded group (Cooper, Shannon et al. 2000) and by studies with human cells (Nakamura, Ito et al. 2005) (Louka, Povinelli et al. 2021). Transcriptomic data have shown *LIN28b*, a key

player in regulating fetal haematopoiesis, including the fetal gamma globin gene *HBG*, to be one of the most upregulated genes in JMML patients (Helsmoortel, Bresolin et al. 2016). High *LIN28b* expression correlated both with fetal haemoglobin concentration and also with a poorer prognosis (Helsmoortel, Bresolin et al. 2016).

More recently, research from our lab has characterised the transcriptomic, phenotypic and clonal hierarchy of disease propagating JMML populations (Louka, Povinelli et al. 2021). Flow cytometric and clonogenic assays demonstrated the presence of an aberrant Lin-CD34+CD38-CD90+CD45RA+ population in greater than 50% of patients analysed compared to CB controls; and that these cells had exclusive myeloid output. This aberrant population was not present in patients successfully cured by BM transplantation but was dominant at relapse. Colony genotyping of index sorted Lin-CD34+ single cell derived colonies revealed the presence of initiating and secondary mutations that could be tracked back to the phenotypic HSC compartment, with the presence of both linear and branching clonal evolution. Single cell RNA-Seq revealed the upregulation of fetal specific genes, including *HMGA2*, and proliferation-associated gene sets.

1.3.5 *In vitro* modelling of JMML

Granulocyte-macrophage colony stimulating factor (GM-CSF) hypersensitivity is a property specific to JMML leukaemic cells and has been used to explore the downstream signalling consequences of patient mutations and to test potential therapeutics. In culture, a key diagnostic property of JMML patient cells is their

hypersensitivity to GM-CSF. JMML patient PB and BM cells are capable of producing large numbers of granulocyte, macrophage and erythroid colonies in the presence of low GM-CSF concentrations (Emanuel, Bates et al. 1991) (Metzner, Horstmann et al. 2007). Kotecha and colleagues also showed in patient BM cells there was a hyper-activation of phosphorylated STAT5 in response to GM-CSF. Further, in a case of JMML that had transformed to AML, BM cells displayed GM-CSF hypersensitivity whereas other non-JMML derived AMLs didn't (Kotecha, Flores et al. 2008). They also demonstrated that JAK2 but not MEK inhibitors could reduce the hyper-activation of phosphorylated STAT5 response to GM-CSF.

This study is supported by work carried out by Donovan and colleagues in a murine FL-derived GM cell line, harbouring truncated *C-myb* (Donovan, See et al. 2002). They demonstrated GM colony growth, in conjunction with increased ERK-2 phosphorylation, in the absence of GM-CSF, only when an *Nf1* -/- mutation was introduced. This proliferation could be reduced by PI3K but not MEK inhibitors (Donovan, See et al. 2002). Similar work was also carried out in murine FL by Schubbert and colleagues who demonstrated that transduction with JMML-associated *PTPN11* mutations (Chan, Leedy et al. 2005) led to increased colony growth in the absence of GM-CSF, as well as increased colony size and hypersensitivity to erythropoietin (Schubbert, Lieuw et al. 2005). This GM-CSF sensitivity was not seen if the PTPase domain was mutated or the NS N308S mutation was transduced. Niemeyer and colleagues performed analogous experiments in which they showed that mouse FL transduced with human *CBL* Y371H, and shRNA targeting the endogenous wildtype mouse *Cbl*, exhibited sensitivity to GM-CSF (Niemeyer, Kang et al. 2010).

1.3.6 Mouse and iPSC models for JMML

Both mouse haematopoietic cell-derived and JMML patient PB/spleen/BM derived xenotransplantation models have been developed to model this leukaemia in an *in vivo* context. These studies have modelled the myeloid skew and GM-CSF sensitivity as well as providing suitable contexts for therapeutic testing. However, they lack the ability to study the disease progression from the first initiating event in the human fetal HSC compartment.

Doan and colleagues used the Mx1-Cre system to induce somatic *Nf-1* exon 31 loss in mouse haematopoietic cells. This led to a reduced survival and increased white blood cell count (Le, Kong et al. 2004). The presence of higher levels of Gr-1+ and Mac-1+ cells in the BM was observed. Importantly, BM mononuclear cells showed GM-CSF sensitivity, although they did not exhibit GM-CSF independence. However, the latency of this disease was not in line with the equivalent disease progression in the human context.

Chan and colleagues used a similar system to induce the *Kras* G12D mutation in mouse haematopoietic cells (Chan, Kutok et al. 2004). These mice developed a myeloproliferative disorder, mainly characterised by a granulocyte expansion, with a median survival of 35 days. Although they didn't show evidence of GM-CSF hypersensitivity, BM from *Kras* mutated mice showed higher colony output in the absence of growth factors SCF, IL-3, and IL-6 (Chan, Kutok et al. 2004). Despite these models recapitulating the myeloid skew and growth factor sensitivity, neither model caused progression to acute leukaemia. Although this is not a common occurrence in JMML either, the lack of acute disease

progression meant the latency recapitulated a more adult-like timescale whereas JMML is a disease exclusive to early childhood.

Cutts and colleagues addressed the latency question by combining both previous models to generate a haematopoietic cell inducible *Nf-1 Kras* G12D strain (Cutts, Sjogren et al. 2009). This led to an acute myeloid leukaemia phenotype with a shorter latency than each mutation alone, increased blasts in the BM, and expansion of CD11b+ Gr-1+ granulocytes (Cutts, Sjogren et al. 2009).

An important question that arose from these earlier models was whether introducing JMML associated mutations in a human cell context, as opposed to mouse, was important in order to fully recapitulate the disease. This was partially addressed by Navarro and colleagues when they studied induced pluripotent stem cells (iPSCs) from JMML and NS patients harbouring *PTPTN11* mutations (Mulero-Navarro, Sevilla et al. 2015). In this model system, they demonstrated GM-CSF hypersensitivity, particularly marked in the NS with JMML patient iPSCs as well as the presence of larger colonies, biased towards macrophage and activated granulocyte production. Further, they observed segregation of patient vs control CD33+ myeloid-derived iPSCs by gene expression, with the upregulation of *MYC*. However, as an *in vitro* model, this approach lacked the microenvironment in which this leukaemia would naturally develop, hence only partially recapitulating the disease.

Several models have been developed to address both the microenvironment and human initiating cell context, using JMML patient cells transplanted into immunocompromised mice (Krombholz, Aumann et al. 2016) (Yoshimi, Balasis et al. 2017). Krombholz and colleagues transplanted spleen mononuclear cells

from five patients with *PTPN11* initiating mutations into Rag2–/–γc–/– mice (Krombholz, Aumann et al. 2016). Three of five mice showed human cell infiltration in the liver, lung, spleen, and BM, dominated by CD33+ myeloid engraftment. One patient failed to engraft and the other led to a non-*PTPN11* mutated T cell expansion. Mouse BM was taken from one group of primary mice and transplanted into secondary recipients, followed by tertiary transplants. All tertiary transplants showed human cell engraftment and the latency of disease was similar to the corresponding primary for both secondary and tertiary recipients. This is perhaps unsurprising as the cells originally used for transplantation were derived from patients who had already progressed to an aggressive disease. Importantly, weekly injection of human GM-CSF into recipient mice increased the proportion of CD33+ cells within the human CD45+ compartment, in both the spleen and BM (Krombholz, Aumann et al. 2016).

Yoshimi and colleagues took a similar approach using transplantation of BM and PB mononuclear cells, from JMML patients with different genetic backgrounds, into NSG-S mice to generate patient derived xenograft (PDX) models (Yoshimi, Balasis et al. 2017). Three of four patients showed robust engraftment with a myeloid skew, short latency (median survival 43 days) and enlarged spleen. For the patient with a *PTPN11* initiating and *SH2B3* secondary mutation-associated with hyper-activation of STAT5 signalling and poor prognosis- the xenograft model recapitulated the aggressive disease progression seen in the patient. Unfractionated spleen cells from the primary recipient caused aggressive myeloid-biased engraftment, enlarged spleen and lung haemorrhage in the secondary recipients, with a latency of 19 days.

Similarly, research from our lab has used PDX approaches to establish sub-populations of JMML patient BM capable of propagating disease in a xenograft setting (Louka, Povinelli et al. 2021). Indeed, the HSC population was able to engraft and cause JMML-like disease in NSG recipients, including an expansion of the human myeloid compartment, significantly enlarged spleen compared to CB controls and mice succumbing to illness at approximately 11 weeks post-transplant. In some cases, the CD45RA+ CD90+ and GMP compartments were also able to engraft and cause disease. Within the aberrant CD45RA+, CD90+ JMML patient population described earlier, CD96 was identified as a potential biomarker of long-term disease propagating cells. Indeed, it was demonstrated by transplanting Lin- CD34+ CD38-CD96+ patient BM, that it was possible to cause JMML-like disease in recipient mice with a latency of approximately 11 weeks whereas this was not the case for Lin- CD34+ CD38- CD96- BM for the patient tested. This identified CD96 as a biomarker for long-term disease propagating JMML stem cells.

In a similar manner, Caye and colleagues interrogated patient samples harbouring different mutations for their ability to propagate disease in a xenograft model from select BM fractions (Caye, Rouault-Pierre et al. 2020). They tested HSC, MPP, LMPP, CMP and GMP populations from JMML patients harbouring *NRAS*, *KRAS* and *PTPN11* mutations. In support of the results shown by Louka and colleagues, these data demonstrated multiple populations, outside of the most primitive HSC and MPP groups, capable of disease propagation *in vivo*, in particular the LMMP fraction in the case of one *PTPN11* mutated and one *KRAS* mutated patient. Interestingly, they reported a longer disease latency in primary NSG recipients of CD34+ patient BM compared to NSG-S- which express human

cytokines including GM-CSF (Wunderlich, Chou et al. 2018) (12 weeks in NSG compared to 6 weeks in NSG-S). On the other hand, in secondary transplants, the latency was 12 weeks for both types of NSG recipients (Caye, Rouault-Pierre et al. 2020).

In summary, while these models provide valuable tools for understanding the detailed phenotype of the disease and testing potential therapies in a more humanised setting, they lack the capacity to model disease progression from the first initiating event in its human fetal HSC context. While retrospective generation of clonal hierarchies from patient data and PDX models have provided key insights into the pattern of mutations within HSPC compartments and populations with disease propagation potential, they lack the ability to recapitulate how the initiating genetic event gradually leads to the observed abnormal haematopoietic outcomes. This is an important aspect of the disease to understand in more detail as therapies which directly target the initiating event in the specific population in which it occurs would be the most likely to prevent relapse.

1.4 How the CRISPR technique is applied to study normal and malignant haematopoiesis

1.4.1 A brief overview of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) as an adaptive immune system in bacteria and archaea

Clustered Regular Interspaced Short Palindromic Repeats (CRISPR) are present in most bacteria and archaea. It serves predominantly as a memory based adaptive immune system (Hille, Richter et al. 2018).

The CRISPR phenomenon was first identified in *Escherichia coli* in 1987 and has since been attributed to many other bacteria and archaea (Ishino, Shinagawa et al. 1987) (Mojica, Ferrer et al. 1995). It was noted that 26-72 nucleotide (nt) unique sequences (spacers) were present in the bacterial genome equally spaced in between repeated (21-48 nt) sequences (Mojica, Ferrer et al. 1995) (Jansen, Embden et al. 2002) (Rath, Amlinger et al. 2015).

Subsequently it was found that very often adjacent to these CRISPR regions are genes encoding CRISPR-associated (Cas) proteins of which some possess exonuclease motifs (Jansen, Embden et al. 2002). These spacer sequences showed homology with bacteriophage and plasmid sequences suggesting a role in the response to infection (Pourcel, Salvignol et al. 2005). It has since been confirmed that CRISPR prevents plasmid transformation (Marraffini and Sontheimer 2008) as well as conferring bacteriophage resistance (Barrangou, Fremaux et al. 2007). The resistance is conveyed by the presence of spacer sequences in the CRISPR 1 locus that are identical to genomic sequences in the associated bacteriophages (Barrangou, Fremaux et al. 2007).

The genomic CRISPR array is composed of a single leader sequence followed by multiple repeat, spacer, repeat, spacer segments (Bult, White et al. 1996) (**Figure 1.5**). The mechanism by which adaptive immunity is conferred by these CRISPR arrays can be largely grouped into three systems (Makarova, Haft et al. 2011).

In the first stage, the foreign genetic material is acquired and integrated into the host genome at the CRISPR site. In bacteriophage immunity, sequences suitable to act as a spacer must have a distinct and nuclease-specific protospacer

adjacent motif (PAM) at the 3’ end of the sequence. This is a three-five base pair sequence which allows the host to distinguish between self (not present) and non-self (PAM present). The second stage of the process is expression of the spacer sequences, the so-called CRISPR RNAs (crRNA).

The final stage is interference/neutralisation. This involves identification and degradation of the target DNA (Marraffini and Sontheimer 2008). The type 2 system relies only on Cas9 for cleavage which is achieved for the sense and anti-sense strand by different domains. For the systems using PAM recognition, the cutting is very precise and happens three base pairs upstream of the PAM site (Garneau, Dupuis et al. 2010). The diversity of the CRISPR system is evident in the complex array of accessory proteins as well as the fact that both RNA (Hale, Zhao et al. 2009) and DNA cleavage is used to remove the invading pathogen, depending on the host.

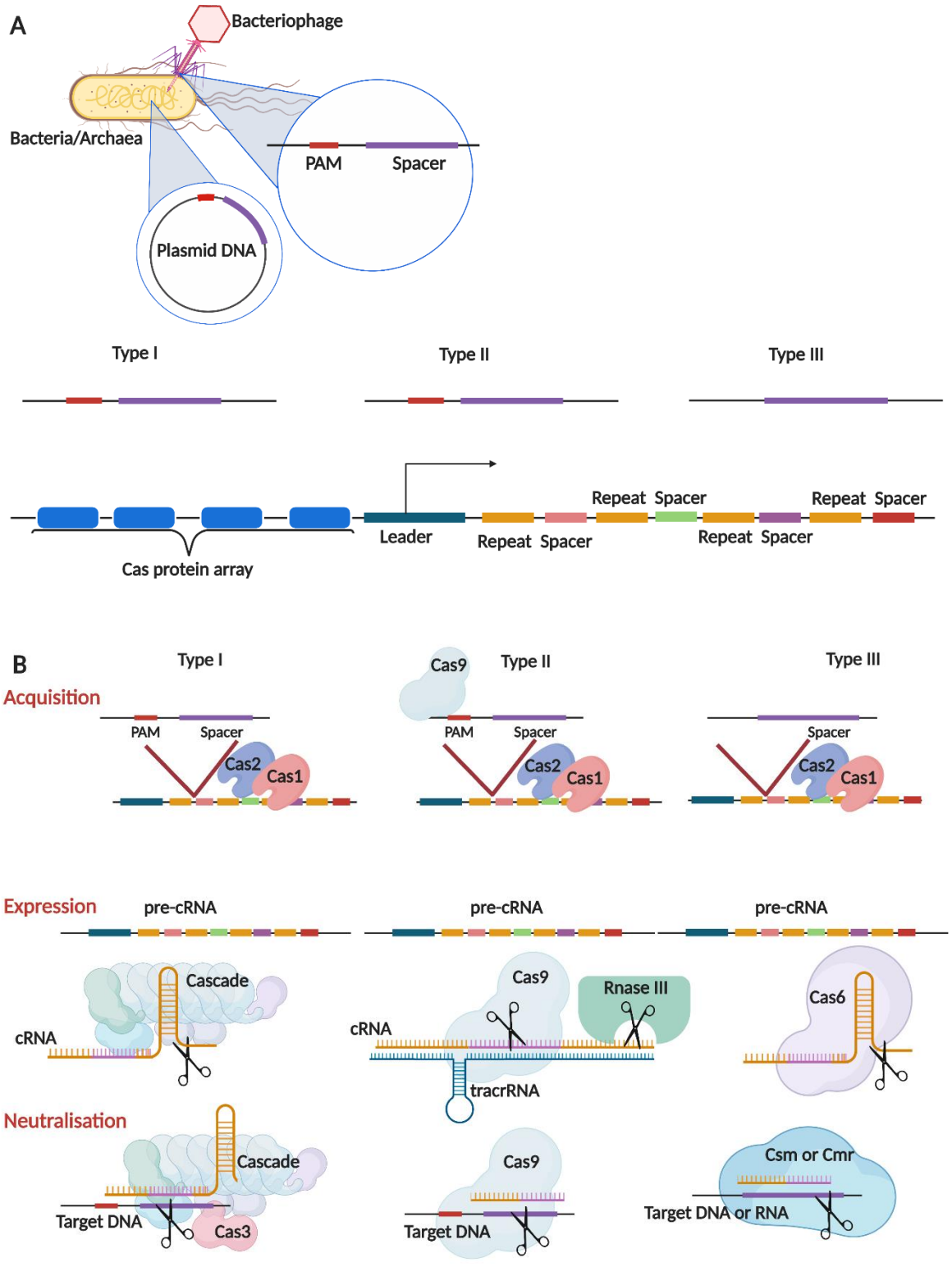


Figure 1.5: CRISPR mediated adaptive immunity in bacteria/archaea- adapted from (Makarova, Haft et al. 2011) and (Rath, Amlinger et al. 2015) Created with Biorender.com **A**) Overview of the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) locus in systems I-III. In all three systems, the locus contains an array of CRISPR associated (Cas) proteins which are

involved in processing the transcribed material as well as performing targeting cutting of foreign genetic material (DNA and RNA). Repeat and spacer regions are transcribed from the leader sequence. Spacer regions contain short segments of foreign genetic material from previous infection such that pathogens can be recognised upon re-invasion. Systems I and II use a protospacer adjacent motif (PAM) to distinguish between self (PAM absent) and non-self (PAM present). PAM containing sequences are incorporated as spacers and ensure self-cutting doesn't occur. **B**) The process by which the CRISPR system targets foreign genetic material. Cas proteins incorporate new foreign genetic material (spacers) between repeat regions when the bacteria is infected. Cas proteins first express CRISPR RNAs (crRNAs) as long pre-crRNAs containing multiple spacers which are then processed to smaller single spacer containing crRNAs which form distinct secondary structures and assist Cas protein binding at the target site. Using complementary base pairing, the crRNA and Cas protein bind the foreign genetic material and the Cas protein(s) induce double strand breaks (DSBs leading to destruction of the foreign material.

1.4.2 The Cas9 protein: structure and modification for use in mammalian cells

Cas9, the main player of the type 2 CRISPR system, is the protein which has been developed to generate an editing platform for mammalian cells with greater precision and ease of use than all its predecessors: transcription activator-like effector nucleases (TALENs), zinc finger nucleases (ZFNs) (Adli 2018) and meganucleases (Silva, Poirot et al. 2011). The Cas9 protein is composed of three main domains: the nuclease domain (NUC), a recognition domain (REC) and an unstructured C terminal domain (CTD) (**Figure 1.6**). The nuclease domain is composed of two separate functional units: RuvC and HNH cleave the guide RNA (gRNA) non-complementary as well as the complementary DNA strands, respectively (Gasiunas, Barrangou et al. 2012) (Jinek, Chylinski et al. 2012). The PAM recognition site lies within the CTD.

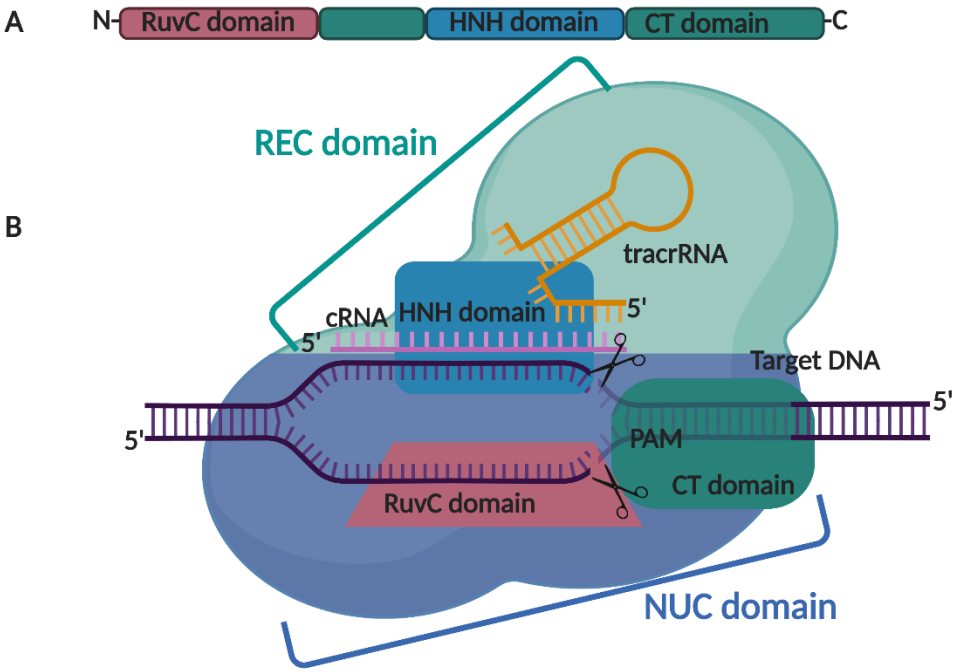


Figure 1.6: Structure of Cas9

Adapted from (Gasiunas, Barrangou et al. 2012) and (Ishino, Krupovic et al. 2018). Created with Biorender.com **A**) Linear translated sequence of structural domains. The Cas9 protein uses the RuvC and HNH domains for target cleavage and the C-terminal (CT) domain for protospacer adjacent motif (PAM) recognition. **B**) Schematic of Cas9 cleavage. The Cas9 protein consists of a recognition domain (REC) which interacts with the trans-encoded small RNA (tracrRNA). The guide RNA is formed of two components: the tracrRNA which acts as a handle to aid Cas9 binding and the CRISPR RNA (crRNA) which contains the spacer complementary to the target. Together they form a secondary structure interacting with both Cas9 and the target. The nuclease (NUC) domain is separate into three structures: the HNH domain which interacts with the crRNA and cuts the crRNA complementary strand of the target, the RuvC domain which cuts the crRNA non-complementary strand and the C-terminal (CT) domain which interacts with the PAM motif but is not involved in cleavage.

Both, the guide RNA sequence and Cas9 protein native to bacteria have been modified to enhance their use in genome editing approaches. The Cas9 protein subject to these modifications is most commonly that purified from *Streptococcus pyogenes* (spCas9). This Cas9 has been modified to both vary and reduce the PAM requirement, reduce off-target cutting, and increase specificity. spCas9 is restricted in its genomic targeting by the requirement of an NGG sequence at the 3' end of the target which is not particularly frequent in mammalian genomes. Using phage-assisted continuous evolution (PACE), whereby the protospacer with a variable PAM sequence upstream of a vital protein for viral replication, Hu and colleagues generated Cas9 mutants with a much wider PAM spectrum (xCas9) including NG, NNG, GAA, GAT and CAA (Hu, Miller et al. 2018).

Another key effort of Cas9 modification has been to improve the specificity, reducing off-target events. A major risk of using this approach to correct disease associated mutations is that the gRNA introduced gives rise to an off-target cut which leads to dysregulation of an oncogene or tumour suppressor. Indeed, it has been shown for specific targets, for example *FANCF* and *VEGFA*, that unmodified Cas9 can lead to up to 20% editing at top off-targets (Kleinstiver, Pattanayak et al. 2016) .

Modifications to reduce off-target editing include reducing sgRNA length (Fu, Sander et al. 2014) and Cas9 mutagenesis specifically targeting amino acids involved in guide RNA/DNA binding. This has led to the development of engineered Cas9 proteins (eCas9) with reduced off-target cleavage (Slaymaker, Gao et al. 2016). In a similar fashion, the lab of Kleinstiever targeted residues of Cas9 interacting with the target DNA strand for mutation, generating SpCas9-HF-1. Various combinations of mutations were tested which all showed lower off-

target cleavage compared to wild type Cas9, without significantly reducing on-target activity, although this was variable between loci (Kleinstiver, Pattanayak et al. 2016). Another method of identifying Cas9 mutants using a bacterial screening approach, with the on-target guide corresponding to a toxin and the off-target to an antibiotic resistance gene, found the R691A mutation reduced off-target cleavage in the loci tested and maintained high on-target cleavage for a therapeutically relevant gene, *HBB*, in CD34+ HSPCs (Vakulskas, Dever et al. 2018).

In addition, either or both of the nuclease domains (HNH and RuvC) have undergone single point mutagenesis to develop nickase (nCas9) or dead Cas9 (dCas9). nCas9 only cuts a single strand depending on which domain has been mutated and dCas9 doesn't cut the target at all (Ran, Hsu et al. 2013). Using nCas9 for target cleavage improves specificity by requiring two guide RNAs and nCas9 to achieve a full cleavage, one in complex with RuvC null Cas9 (D10A) and the other HNH null Cas9 (H840A).

1.4.3 Methods developed to design guide RNAs targeting mammalian genomes

A crucial part of ensuring high on-target and low off-target editing at the desired genomic location comes at the guide RNA design stage. To this end, algorithms have been designed to predict suitable guide RNAs with highest on-target to off-target ratio. Separate tools have been produced to predict on-target and off-target activity. This is due to the fact that on-target activity prediction requires experimental validation of guide RNA properties whereas off-target prediction is

more definitive, relying on the number of mismatches and the position of the mismatch within the guide sequence (Thomas, Parry-Smith et al. 2019). Off-target activity algorithms scan the genome for all PAM sites with adjacent protospacers with up to four mismatches. These off-targets are then scored depending on the number and position of these mismatches; those with minimal mismatches and within the seed region closest to the PAM have the highest off-target score. On-target activity has been validated by testing groups of guides in a U6- promoter based *in vitro* setting and T7 promoter based *in vivo* setting and comparing several empirically determined scores with several prediction tools (Haeussler, Schönig et al. 2016).

While this evaluation demonstrated the prediction tools value in greatly reducing the number of guides required to be empirically tested, it also revealed large discrepancies depending on the algorithm used and context of the guide. For example, high-on target guides used with the U6 promoter differed greatly from those used with the T7 promoter for the same target, due to specific transcription requirements for the two systems. The on-target and off-target scores are then combined to give a value that comprises both; the choice of guide is therefore a balance of low off-targets whilst maintaining high on-target efficiency. What these algorithms don't take into account is the position of the off-target therefore results are outputted with the location of the off-target. If the top off-targets are within intronic regions it can be deduced that these would have less biological effect therefore can be considered as more suitable than a guide with the same overall score but with off-targets in exonic regions. These computational tools are valuable in consolidating a group of guides with high predicted efficiency and specificity; however, these algorithms don't circumvent the requirement to

empirically test a set of guides for a particular locus. The predictions output by the algorithms don't always completely match the experimental result.

1.4.4 The advantages and disadvantages of plasmid, viral vector and ribonuclear protein (RNP) delivery

The CRISPR machinery can be delivered into mammalian cells in three main modes. A bacterial plasmid, a viral vector or as a ribonucleoprotein (RNP) (Lino, Harper et al. 2018). Plasmids and RNPs can be delivered by electroporation, nucleofection or lipofection, whereas lenti and adeno-associated viral particles are subject to transduction. **Table 1.2** summarises the key advantages and disadvantages of each mode and method of delivery of the CRISPR machinery.

Delivery mode	Advantages	Disadvantages
Plasmid	Fluorescent readout of efficiency. Sustained expression of CRISPR machinery.	Potential safety risk of integration. Large constructs can be toxic and inefficient.
RNPs	No risk of integration. Transient expression reduces potential for off-target cutting. Small construct likely to improve efficiency and reduce toxicity.	No fluorescent readout of efficiency unless fused to large fluorescent reporter protein which could reduce efficiency and increase toxicity.
Viral vector	Sustained expression of CRISPR machinery.	Potential safety risk of integration.
Delivery method		
Lipofection	Reagents optimised for low toxicity	Very low efficiency in HSPCs
Electroporation	Higher efficiency than lipofection. Transient electrical impulse.	Electrical impulse can be toxic particularly to primary cells.
Nucleofection	Direct delivery of CRISPR machinery to nucleus	Electrical impulse can be toxic particularly to primary cells.
Transduction	Cell surface pseudotype and promoter can be optimised for target cell.	Large constructs can have low transduction efficiency.

Table 1.2: Summary of the advantages and disadvantages of different modes and methods of delivery of the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) machinery Ribonuclear proteins (RNPs) and plasmids can be delivered by lipofection, electroporation or nucleofection. Viral vectors are delivered by transduction.

The advantages of plasmid and viral delivery are direct and simple readouts of transfection/transduction efficiency (Heckl, Kowalczyk et al. 2014) as they deliver sustained expression of the CRISPR machinery coupled with a fluorescent marker (Shalem, Sanjana et al. 2014). Due to the size of the Cas9 protein and any potential reporters/donors, the vector constructs are large and associated with cell toxicity (Kim, Kim et al. 2014). Packaging constraints of viral particles limits the number of guides that can be multiplexed. Further, although strategies

have been adopted to reduce the potential off-target integration of lentiviruses, this remains a possibility and therefore a risk (Lino, Harper et al. 2018). Indeed, recombinant adeno-associated viral vectors have been developed with an almost entirely episomal expression of their machinery in the host cell but still a small fraction becomes part of the target cell genome (Chen and Gonçalves 2016). Any choice of a delivery method varies depending on the cell type and application. Generally, it centres on the efficiency of editing achieved and safety in the case of therapeutic use.

Approaches to overcome the sustained expression of Cas9 include the introduction of a drug inducible element within the plasmid. This has been demonstrated successfully using doxycycline in mouse embryonic stem cells targeting a number of tumour suppressors (Dow, Fisher et al. 2015). In contrast, RNPs are small entities in which a synthetic and chemically protected single guide RNA (sgRNA) is pre-complexed with the Cas9 protein and delivered directly to the cell, an approach which has been simplified from the naturally occurring system. Although the Cas9 DNA cleavage method, the hallmark of the type 2 system, is the simplest of the naturally occurring CRISPR systems, it still requires two RNA entities. Using synthetic processes, sgRNAs in which the trans-encoded small RNA (tracrRNA) and crRNA components are fused have been developed and shown to cleave targets as expected *in vitro* (Jinek, Chylinski et al. 2012). Further modifications to these sgRNAs has led to improved stability allowing robust editing (Hendel, Bak et al. 2015).

The RNP approach is associated with lower toxicity due to the size as well as stability of the complex; it is degraded rapidly once inside the cell which doesn't prevent the potential of an innate immune response to the foreign RNA and

protein, but leaves less opportunity for off-target editing to occur (Kim, Kim et al. 2014). In human HSPCs, editing has been shown to be evident two hours after nucleofection (Wu, Zeng et al. 2019). In contrast, the same recombinant Cas9 and *CCR5* guide RNA delivered by plasmid showed the first signs of editing at six hours with maximal editing at 72 hours post-nucleofection. In RNP form, Cas9 was detectable immediately after transfection, and almost entirely degraded at 24 hours whereas plasmid expression was not detectable until six hours post-nucleofection, and still present at 72 hours. Further, RNP targeting of *CCR5* or *EMX1* induced fewer off-target effects compared to plasmid delivery without a reduction in on-target editing (Kim, Kim et al. 2014, Lin, Staahl et al. 2014). Taken together, these data demonstrated that RNP CRISPR delivery lead to more rapid on-target editing and lower off-target editing most likely due to the transient expression of the CRISPR machinery. A further improvement has been the development of Cas9 RNP fluorescent fusions which allow sorting of these transfected populations (Nasri, Mir et al. 2019).

1.4.5 The advantages and disadvantages of lipofection, transfection and transduction

As described above, there are three main methods of delivering the CRISPR machinery: lipofection, transfection and transduction. DNA plasmids and RNPs are typically delivered by lipofection, electroporation or nucleofection, and viruses by transduction. Lipofection involves coating DNA (or RNPs) in cationic lipid particles which fuse with the cell membrane allowing plasmid or RNP entry. Transfection involves the application of an electrical impulse to the cells which

generates a transient pore in the membrane allowing plasmid or RNP entry. Once inside the cell, the Cas9 and guide RNA are directed to the nucleus using host cell mechanisms addressed by the nuclear localisation sequence (NLS) tagged on Cas9. Nucleofection allows delivery of Cas9 and sgRNA to the nucleus directly without requirement for nuclear envelope breakdown. This is achieved by the use of particular buffers. Electroporation and nucleofection are both cell type specific approaches which use different buffers and voltage conditions for individual cell types (Maurisse, De Semir et al. 2010). Transduction involves the formation of a replication-deficient virus particle which is produced in HEK293TLA cells. Three separate vectors are employed: these encode the sgRNA(s) and Cas9, the viral structural proteins, and the viral replication required proteins. There are two types of viruses that have been adapted for introduction of the genome editing machinery. These are adeno-associated viruses and lentiviruses. Lentiviruses are versatile in their cell tropism as their cell surface pseudo types can be modified, and they also transduce non-dividing cells.

Efficiency and applicability of each method depends on the cell type, although a study carried out on a range of cell lines and primary cells demonstrated that nucleofection was in general a superior method to lipofection (Maurisse, De Semir et al. 2010). Lipofection methods are particularly inefficient in primary haematopoietic cells (Gul-Uludag, Xu et al. 2012) therefore studies on this cell type are restricted to transduction and transfection.

Once inside the cell, the CRISPR machinery, in its native form, induces a double strand break (DSB) at the site targeted by the gRNA. By harnessing the cell's intrinsic repair pathways, the DSB is repaired by one of two mechanisms: homology directed repair (HDR) or non-homologous end joining (NHEJ) (**Figure**

1.7). HDR requires an additional template sequence (donor) to be delivered with the Cas9 and gRNA, as either double stranded (ds) or single stranded (ss) DNA. This template contains homology arms to the sites upstream and downstream of the cut site, with the sequence to be added/modified in its centre. The intracellular repair machinery uses this sequence to direct repair of the cut site which allows introduction of precise single base pair modifications and insertion of reporter constructs. NHEJ occurs when the DSB is repaired without a template which results in insertions and deletions (indels). These arise due to rounds of the Cas9 cutting, cell repairing while retaining the PAM, Cas9 re-cutting etc until mistakes are made by the repeated DNA damage. If these occur within gene bodies causing out of frame sequence, or introduction of premature stop codons, they can cause loss of protein expression. NHEJ therefore has the potential to be used to model diseases caused by loss of function mutations.

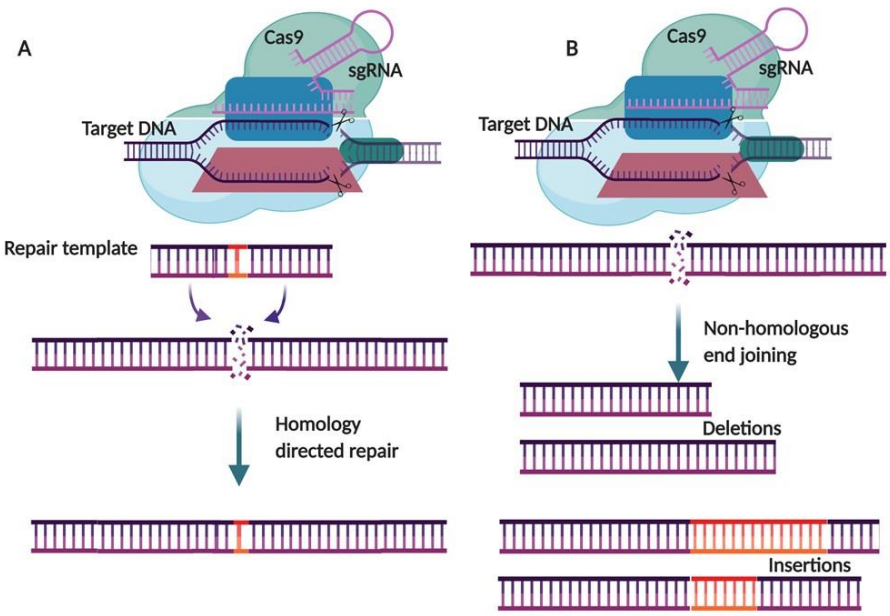


Figure 1.7: The repair of CRISPR-Cas9 induced double strand breaks A) Homology directed repair. **B**) Non-homologous end joining. Homology directed repair requires an additional template sequence which acts as a guide for the repair machinery leading to specific nucleotide changes. Non-homologous end joining is an error prone repair mechanism which leads to random insertions and deletions at the cut site.

Although in principle both approaches can be applied to any cell type, HDR is inefficient and requires DNA replication to be complete hence is restricted to late S and G2 phases of the cell cycle. In contrast NHEJ is active at all stages and therefore competes with HDR at late S and G2, and is the only pathway for CRISPR repair in G1/early S/G2 (Heyer, Ehmsen et al. 2010).This phenomenon means that cells which are quiescent, including stem cells, mainly residing in G_0/G1 of the cell cycle are limited in their ability to direct HDR; instead they are, for the most part, subject to NHEJ mediated repair. However, it has been shown in human embryonic stem cells that HDR efficiency can be increased by synchronising the cells at the G2 to M phase transition which doesn't affect subsequent proliferation (Lin, Staahl et al. 2014).

1.4.6 Application of the CRISPR approach to disease correction and modelling in haematopoiesis

Haematological disorders which have the potential to be treated using CRISPR technology, such as beta thalassaemia, require correction at the LT-HSC level for the potential of a long-term cure to be possible (Mettananda, Fisher et al. 2017). Correction at the LT-HSC level would ensure permanent correction of the erythroid lineage as opposed to a transient presence of normal red blood cells, which is the case in transfusion treatment (Shah, Sayani et al. 2019). Similarly, the HSPC compartment has been shown in some, but not all, haematological malignancies to possess the disease initiating population, as well as the leukaemic stem cells which are often responsible for relapse (Caye, Rouault-Pierre et al. 2020) (Louka, Povinelli et al. 2021). Therefore, for the CRISPR editing approach to be successful in therapeutic applications and disease modelling for haematopoiesis, the LT-HSC population within the HSPC compartment needs to be targeted. Both transduction and nucleofection approaches are being investigated for targeting LT-HSCs. Zonari and colleagues carried out transduction optimisation studies in human CB and BM CD34+ populations and established that the VSV-G pseudotype lentivirus delivery, as well as pre-culture with StemRegenin 1 (SR1) and UM171, was optimal in achieving high levels of transduction (87%) and engraftment in mice (Zonari, Desantis et al. 2017).

One important consideration in this approach is whether to purify the successfully edited population pre-transplantation so they are not competing with unedited counterparts. Bak and colleagues explored the possibility of enriching for edited cells as well as targeting multiple loci using the HDR approach in human

mobilised PB or CB. They targeted multiple loci through RNP nucleofection assisted by transduction with recombinant adeno-associated virus HDR donors, delivering fluorescent reporters. They demonstrated HDR in up to 4 genes targeted at one time but with reduced efficiency: 1% HDR for four targets compared to 10% for two. By sorting for single/double positive reporter expression and mouse transplantation, they demonstrated human cell engraftment in the myeloid and B cell compartments concomitant with high levels of the relevant reporter expression (Bak, Dever et al. 2017). These two studies demonstrated the feasibility of HDR approaches in human HSPCs at the bulk level and showed long-term engraftment in mouse xenografts. However, the efficiency of this method was a significant hurdle. Wagenblast and colleagues explored the possibility of sorting human CB LT-HSCs ahead of the editing process. Using an RNP nucleofection approach, they achieved 30% and 20% editing for NHEJ and HDR methods, respectively. Using a limiting dilution approach, they determined that transplantation of 100 LT-HSCs gave rise to a single engrafted edited LT-HSC in the majority of cases. This method gave an efficiency of 35% LT-HSC engraftment when targeting the olfactory receptor, a gene not known to be expressed in haematopoietic cells. The presence of CD45+ cells in all lineages tested suggested the edited cells maintained multi-potent lineage output (Wagenblast, Azkanaz et al. 2019).

1.4.7 Using the CRISPR technique to correct monogenic haematological diseases: non-homologous end joining repair

The most studied application of CRISPR technology for treatment of genetic haematological disorders is in the inherited haemoglobinopathies, beta thalassaemia and sickle cell disorder, which are caused by mutations in the beta globin gene (*HBB*) (Cao and Galanello 2010, Thein 2013) (Gravitz and Pincock 2014). Extensive work to validate this approach for the treatment of beta thalassaemia has been carried out by Mettananda and colleagues. Using CD34+ selected CB, enriched for HSPCs, they targeted the MCS-R2 enhancer of the alpha globin gene using a plasmid based dual guide approach in order to improve the high alpha/beta globin ratio which causes the phenotypic manifestations in some forms of beta thalassaemia (Mettananda, Fisher et al. 2017). *In vitro* they demonstrated up to 80% editing efficiency in CD34+ cells by gel band intensity and droplet digital PCR and a significantly reduced alpha/beta globin ratio. Interestingly, a range of genetic aberrations were seen at the target site not limited to deletion of the region flanked by the guide pairs but also inversions. When transplanted into NSG recipient mice, up to 34% human cell engraftment was seen, including myeloid and lymphoid lineage output at 12 weeks suggestive of HSC engraftment (Mettananda, Fisher et al. 2017).

Other groups used a CRISPR targeting approach to induce the expression of fetal haemoglobin (HbF) in severe beta thalassaemia patients, e.g. by targeting *BCL11A*, a transcriptional repressor of gamma globin. Guide RNA pairs were designed against the *BCL11A* erythroid enhancer region 200bp apart. Neon electroporation of various guide RNA pairs followed by clone production demonstrated up to 64% editing efficiency, although this was variable. Reduced

BCL11A mRNA levels, increased gamma globin mRNA for one clone, and a significant increase in HbF protein for two clones were demonstrated (Khosravi, Abbasalipour et al. 2019). Targeting the *BCL11A* enhancer is an approach that is being explored for therapeutic use in both beta thalassaemia and sickle cell disease. Wu and colleagues demonstrated efficient editing at the *BCL11A* enhancer using RNP nucleofection of human patient CD34+ HSPCs and sustained human cell engraftment in secondary xenotransplants. This was optimised by generating a Cas9 with multiple nuclear localisation sequences and the addition of glycerol to the nucleofection suspension (Wu, Zeng et al. 2019). Interestingly, they also noted that smaller indels were seen for HSPCs at G_0 compared to SG2M and that these edits were increased in the engrafted samples (Wu, Zeng et al. 2019). A CRISPR based *BCL11A* deletion approach is now in clinical trial for both beta thalassaemia and sickle cell disease patients and preliminary data look encouraging (Frangoul, Altshuler et al. 2021).

1.4.8 Using the CRISPR technique to correct monogenic haematological diseases: Cas9-induced homology directed repair

In parallel with NHEJ based approaches, HDR editing approaches are also being investigated for treatment of beta thalassaemia. Dever and colleagues applied a method aimed at targeting the *HBB* locus itself in human mobilised PB. Using a 2′-O-methyl-3′-phosphorothioate modified gRNA within the *HBB* exonic region and a rAAV6 carrying 400-500bp homology arms flanking a GFP reporter, they achieved 29% HDR in HSPCs, which also showed long-term engraftment, although at much lower levels than non-edited controls. Successful viral

integration was demonstrated by PCR against the homology arm from the lentivirus (Dever, Bak et al. 2016). They tested pre-purifying HDR positive cells which increased the total edited population engrafted compared to bulk transplantation of equivalent cell numbers. They demonstrated *in vitro* a similar approach targeting the sickle haemoglobin (*HBS*) Glu6Val mutation with 40% transduction, 50% HDR and a 60% reduction in HBS mRNA when differentiated into erythrocytes (Dever, Bak et al. 2016). Other groups have targeted human induced pluripotent stem cells (iPSCs) derived from fibroblasts of a compound heterozygous beta thalassaemia patient (Xie, Ye et al. 2014) although this approach is less likely to be useful in a clinical setting.

1.4.9 Using NHEJ to induce loss of function mutations for modelling of haematological malignancies

Inactivating mutations in HSPC populations are common drivers of haematological malignancies and NHEJ lends itself to modelling these mutations. Ebert's group used a lentiviral mediated NHEJ approach in mouse BM LSK cells, targeting multiple genes where loss of function is commonly seen in myeloid malignancies. They demonstrated efficient editing (50-70%) in the HEL cell line for *NF-1*, *EZH2*, *RUNX1*, *TET2*, *DNMT3A* and *SMC3* using a fluorescent reporter based system (Heckl, Kowalczyk et al. 2014). In order to induce NHEJ events at multiple loci in the same single LSK cell, they used a compound vector approach. From one vector, Cas9, GFP and a gRNA were expressed. Other gRNAs were expressed from separate RFP vectors without Cas9 and produced at a higher titre. Transplantation of these transduced cells led to an expansion of

GFP/RFP positive cells in the PB and a myeloid skew although not all of the mice developed leukaemia. One mouse which did develop leukaemia showed the presence of an expanded clone with mutations in 4 genes, *Dnmt3a*, *Ezh2*, *Smc3* and *Nf-1*, which also caused disease in secondary recipients. Interestingly, in a sample containing control sgRNA, they didn't see an altered contribution to myeloid, B or T cell lineages in edited compared to non-edited cells suggesting that HSPC lineage output was not affected by the editing process in their system (Heckl, Kowalczyk et al. 2014).

One advantage of the CRISPR editing system is the ability to introduce mutations in a stepwise fashion within a specific cell population. Kotini and colleagues used this approach to elucidate the progression of myelodysplastic syndrome (MDS; pre-leukaemia) to acute myeloid leukaemia (AML). They derived iPSCs from patients with low risk MDS, high risk MDS or MDS/AML in which the mutations at each stage had been identified. They demonstrated, using lentiviral transduction (bi-allelic) or nucleofection (wild type allele inactivation only) with a *GATA-1* targeting RNA, of a pre-leukaemic clone already harbouring a *GATA-1* mutation, that the additional mutation caused progression to low risk MDS, characterised by an expansion of the CD34+ CD90+ compartment. Similarly, they modelled the progression from a normal to pre-leukaemic stage by nucleofection of a normal iPSC clone with an *ASXL-1* gRNA which on its own showed no phenotypic differences. However, when subsequently targeted with an AAV virus inducing a mono-allelic chromosome 7 deletion via loxP site integration, this modelled the progression to high-risk MDS (Kotini, Chang et al. 2017). Such approaches provide the opportunity to elucidate key disease drivers and candidates for therapy resistance (Tothova, Krill-Burger et al. 2017).

As well as facilitating further understanding of important known oncogenic events, CRISPR allows the screening of potentially important novel genes at a genome wide level. Yamauchi and colleagues used a lentiviral approach in mouse AML cell lines to identify gene targets essential for AML survival. AML cell lines, with a simple genetic background, were transduced with a retrovirus containing either the *CALM/AF10* or *MLL/AF9* fusion genes followed by lentiviral transduction of Cas9. These were propagated in three serial transplantations. A genetic screen was then carried out using lentiviral transduction targeting 20,611 protein coding genes followed by depletion analysis. This identified the DCPS protein as a candidate required for AML propagation which was further validated using shRNA knockdown which reduced AML cell line survival, as well as the DCPS inhibitor RG3039 which improved survival in patient derived AML xenograft recipients (Yamauchi, Masuda et al. 2018). In this way, potential novel therapeutic targets can be narrowed down with greater efficiency than previous approaches.

1.4.10 Using HDR for 'correction' of haematological malignant disease

CRISPR is being explored as a therapeutic approach for targeting oncogenic drivers in various haematological malignancies. HDR offers the opportunity to correct loss of function mutations and restore protein expression. In the chronic myeloid leukaemia (CML) cell line KBM5, which harbours a homozygous *ASXL-1* mutation, Valletta and colleagues used plasmid nucleofection of sgRNAs targeting the exon 12 mutation in combination with a single stranded oligonucleotide (ssODN) containing the corrected sequence. Heterozygous and

homozygous corrected clones were generated which showed gain of *ASXL-1* expression as well as increased H3K27 methylation. Edited clones demonstrated reduced proliferation compared to unedited KBM5, as well as an increased level of morphologically mature myeloid cells. In the transplant setting, recipients of edited clones showed increased survival compared to unedited (Valletta, Dolatshad et al. 2015). *ASXL-1* is associated with a poor prognosis in myeloid malignancies, therefore restoring its function presents a potentially promising therapeutic strategy. Other groups have targeted the *BCR-ABL* fusion gene which causes chronic myeloid leukaemia (CML) and although successful HDR at the *BCR-ABL* locus was obtained, long-term disease correction in xenotransplant models has not yet been achieved (Luo, Gao et al. 2019).

1.4.11 CRISPR in the context of disease modelling and correction

A critical consideration when using CRISPR for therapeutic purposes and disease modelling is the impact of the genetic modification on the target cell. To be confident that altered stem cell properties seen when tumour suppressors are targeted by CRISPR can be attributed to the loss of that target gene, it must be evident that the DSB induced doesn't cause any of these known phenotypes when a tumour suppressor is not targeted. Otherwise, disease phenotypes observed cannot be confidently attributed to the specific tumour suppressor or oncogene targeted to model a certain disease associated mutation.

Equally, in terms of therapeutic genome editing, it is crucial that the phenotype of the modified HSPCs is a result of the gene targeted, as opposed to the DSB process itself. Edited stem cells for therapeutic purposes must behave as normal

LT-HSCs and be able to contribute to long-term production of the corrected lineage as well as all other blood lineages. It is crucial that these stem cells are not aberrantly affected in any way. Indeed, previous gene therapy approaches have had tragic consequences. In a study for treatment of severe combined immunodeficiency (SCID)-X1 disease, retroviral transfer of a normal version of the gammac receptor in patient CD34+s was used (Cavazzana-Calvo, Hacein-Bey et al. 2000). Despite showing initial promise, a patient receiving this treatment developed a T-cell leukaemia, most likely due to interrogation of the retrovirus close to a tumour suppressor or proto-oncogene, resulting in aberrant up-/down-regulation (Check 2002). In another study using an adenoviral approach to treat ornithine transcarbamylase (OTC) deficiency, cells infected with inactivated adenovirus carrying a normal copy of the OTC gene led to severe inflammation caused by an over-reactive immune response which proved fatal (Lehrman 1999). Indeed, it has also been shown in CRISPR mouse models in which Cas9 is constitutively expressed, that there was a sustained immune response to the presence of the Cas9 protein (Wang, Mou et al. 2015). CRISPR-based gene therapy approaches therefore do not come without associated safety concerns. Due to the fact that DNA damage induced by double strand breaks can lead to the induction of apoptosis (Kaina 2003), an important consideration is whether cells which are able to circumvent this DNA-damage induced apoptosis are more likely to survive CRISPR editing, and whether this could lead to enrichment of cancer propagating populations.

1.4.12 TP53 association with CRISPR-based editing

An emerging concern around the CRISPR approach is the role of TP53 in inhibiting the process and therefore selection of *TP53* mutated cells by the genome editing machinery.

TP53, often referred to as the guardian of the genome, plays an importance role in controlling cell proliferation through directly regulating the transcriptional of various cell cycle, apoptosis and DNA damage genes (Aubrey, Kelly et al. 2018). Under normal conditions, TP53 protein levels are low, but they increase in response to cellular stress or damage, such as UV or radiation. The TP53 protein responds to this damage by activating DNA damage repair pathways through transcription. If damage can't be repaired, TP53 causes activation of genes leading to programmed cell death (apoptosis). In addition to TP53 pathway activation, immune cells, such as T cells, play an important role in cancer surveillance. Tumour cells often aberrantly expressed antigens on their cell surface which are recognised by T cells, in a similar way to the recognition of 'foreign' bacterial and viral antigens on the surface of infected cells. Tumour antigens on the cell surface cause T cells to use cytotoxic approaches to destroy the tumour cells (Rao, Gharib et al. 2019). Indeed, new emerging therapries for certain cancers, including leukaemia, involves *ex vivo* modifying of T cells to react against novel cell surface tumour antigens and re-introducing these into patients (Sterner and Sterner 2021). In this way, there are multiple intrinsic and extrinsic mechanisms for preventing cancer development which cells must overcome before a solid tumour or leukaemia can develop. Mutations in TP53, the most commonly mutated gene in solid and liquid cancers, allow cells to circumvent this damage surveillance and therefore continue to aberrantly proliferate (Aubrey,

Kelly et al. 2018). **Figure 1.8** summarises the role of the immune system and TP53 in damage surveillance and responses to cellular stress.

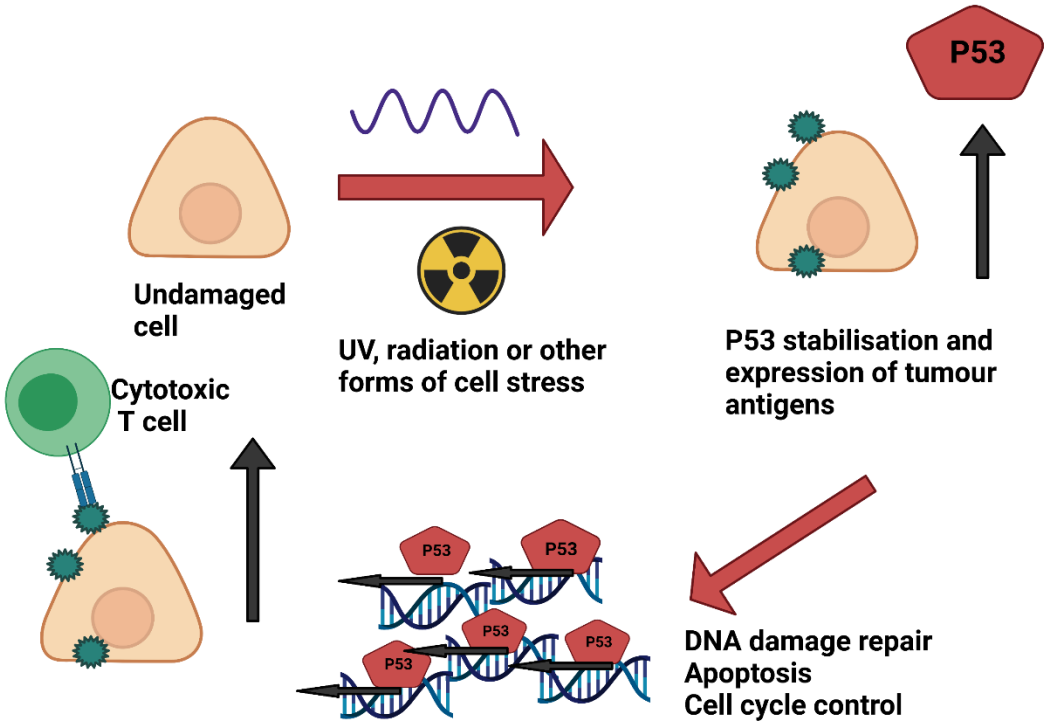


Figure 1.8: Extrinsic and intrinsic mechanisms of responding to cellular stress The immune system and intrinsic cell pathways are present to respond to cellular damage such as UV, radiation and infection. Damaged cells stabilise the TP53 protein which causes transcriptional activation of gene networks involved in DNA damage repair, apoptosis and cell cycle control. Damaged cells also express surface antigens which are recognise by T cells with specific receptors in a process called immune surveillance. Created with BioRender.com. Adapted from (Aubrey, Kelly et al. 2018).

One particular concern about CRISPR is whether it can select for mutated TP53 cells and therefore expand the pool in currently healthy patients. A CRISPR dropout screen carried out in the non-transformed cell line RPE1 found significant enrichment of guide RNAs targeting *TP53* and one of its transcriptional targets cyclin dependent kinase (*CDK*) inhibitor 1A (Haapaniemi, Botla et al. 2018). The authors also noted that some essential genes which are expected to be depleted

in this type of screen did not show significant dropout compared to non-essential genes. Interestingly, in the same cell line mutated at *TP53*, the essential genes were depleted suggesting that survival of these cells after a CRISPR induced DSB was most significantly affected by genes controlling cell cycle and apoptosis. In support of this observation, HDR mediated correction of a mutated GFP allele introduced in this cell line was more efficient in *TP53* mutated or supressed samples. These effects were abrogated by CDK4/6 inhibition or attenuation of TP53 suppression. These data point towards the potential for *TP53* mutant cells to selectively survive genome editing above their *TP53* wild type counterparts.

Similar findings were published by Ihry and colleagues who studied CRISPR editing in human ES cells and iPSCs. Using an inducible Cas9 line derived from the ES cells and lentiviral delivery, they showed significantly reduced proliferation specific to targeting guide RNAs (neuronal, not required for stem cell survival) which was mirrored in transient RNP delivery, suggesting there was a toxic response to the DSB. This was accompanied by enrichment for genes transcriptionally targeted by TP53 including *P21*, when comparing an on-target to a non-targeting guide. An additional finding which points towards a TP53 specific response was a dropout screen in both ES cells and iPSCs, as well as tumour cell lines, which showed the most prominent enrichment in non-targeting guides in the stem cells with almost no enrichment in *TP53* mutated cell lines. Introduction of a *TP53* targeting guide RNA prior to neuronal gene targeting partially rescued the proliferation of the stem cells. Similarly, transfection of a dominant negative *TP53* mutant transgene increased HDR delivery and stem cell survival. (Ihry, Worringer et al. 2018). Further, Enache and colleagues performed

extensive screening of *TP53* WT and mutant (MUT) cell lines to assess the dependency of Cas9 activity on TP53 as well as the specificity of Cas9 selection for *TP53* mutated cells, as opposed to other tumour suppressors (Enache, Rendo et al. 2020). They found a significant competitive advantage for *TP53* MUT lines that was specific to Cas9 introduction as opposed to empty vector controls. What was quite striking was that this advantage was specific to TP53 and not seen when competing other MUT tumour suppressors ARID1A or FBXW7 with their wild type counterparts.

Schiroli and colleagues interrogated the transcriptional and proliferation impact on human CB CD34+ selected HSPCs upon targeting of the *IL2RG* locus with nucleofection of ZFN mRNA, a low specificity RNP, a high specificity RNP or a high specificity RNP with a HDR template provided by AAV6 transduction. Although all conditions showed a decrease in proliferation compared to the control, this was only significant in the low specificity RNP, or high specificity RNP with AAV6, suggesting a cumulative effect of DSB acquisition. Similarly, the most significantly upregulated genes were TP53 targets and apoptotic associated pathways. Further, more robust human CD45+ engraftment of edited cells was seen when the samples were co-transfected with the editing components and a transient TP53 inhibitor. This suggests that TP53 provides a protective response to CRISPR mediated DSBs which can be circumvented by its transient suppression (Schiroli, Conti et al. 2019). It is not however known whether the TP53 associated transcriptional response induced immediately subsequent to CRISPR induced DSBs has a long-lasting effect on HSC gene expression patterns post-transplant; although by flow cytometry the lineage output of edited HSCs looked normal.

1.5 Hypotheses and aims

The principal aim of this thesis was to gain an in-depth understanding of the CRISPR genome editing tool as an approach for studying both normal and malignant haematopoiesis. In order to achieve this therefore, the CRISPR method was employed first in a setting in which no haematopoietic genetic landscapes were targeted for perturbation and second in which a specific RAS pathway tumour suppressor was edited. This broad main aim was broken down into three principal objectives. The first was to optimise a CRISPR delivery system in mouse HSPCs such that 90-100% editing efficiency by NHEJ could be achieved. The second aim was to use the optimised NHEJ directed editing approach to interrogate the molecular and functional consequences of genome editing on mouse HSCs both short-term, 24 hours post-nucleofection, and long-term, 18 weeks post-transplant. The final aim was to use the optimised NHEJ directed editing approach in human HSCs to target the *NF-1* loss of function mutation associated with 15% of childhood JMML cases (Caye, Strullu et al. 2015).

1.5.1 Aim 1: To optimise a genome editing approach in mouse HSPCs using *in vitro* methods

In order to develop a readily tractable genome editing approach, gRNAs were designed targeting the fluorescent reporter tdTomato. What was paramount for the success of this approach was to achieve a 90-100% editing efficiency. Therefore, given the wide range of CRISPR delivery modes and methods (Lino, Harper et al. 2018), and variable editing and engraftment efficiency in HSPCs

depending on the approach (Zonari, Desantis et al. 2017) (Bak, Dever et al. 2017) (Wagenblast, Azkanaz et al. 2019), experiments were designed to systematically assess the efficiency of each mode and method using two editing strategies. It was hypothesised that the more sustained expression of the Cas9 and guide RNA would be advantageous in the context of plasmid and viral vector based editing whereas the RNP approach might not allow a sufficient time window for editing. On the other hand, it was predicted that RNPs might have lower toxicity particularly in the context of primary cells. The first editing strategy used a dual guide approach to target induction of tdTomato expression. The second used a single guide targeting a repeat region to induce tdTomato knockout. Plasmid and RNP transfection were tested using both electroporation and nucleofection. Viral vectors were tested using transduction. In this way, each method was optimised and tested for their relative transfection/transduction efficiency, editing efficiency and effect on cell viability. In this way, an optimal approach was consolidated which balanced cell viability with editing efficiency.

1.5.2 Aim 2: Molecular and functional characterisation of CRISPR edited mouse HSCs

In order to critically assess the short-term and long-term effects of CRISPR editing on mouse HSCs, RNA-Seq of tdTomato targeted HSCs was carried out both at 24 hours post-nucleofection and 18 weeks post-transplantation. Given the evidence that TP53 plays an important role in the short-term CRISPR-induced DSB response (Schiroli, Conti et al. 2019) it was hypothesised that this might have implications on long-term stem cell function if there was sustained

perturbation of the transcriptional signature. By carrying out short-term and long-term RNA-Seq on matched tdTomato-edited, sham-edited and fresh samples, it was possible to separate transcriptional signatures which were transient versus those which were sustained long-term. Although previous studies have assessed the lineage output of edited HSCs (Wagenblast, Azkanaz et al. 2019) (Mettananda, Fisher et al. 2017), they have lacked the ability to interrogate all terminally differentiated lineages. By targeting the tdTomato reporter transgene, not only was it possible to track the contribution of edited cells to all terminally differentiated lineages including red blood cells and platelets, it was also possible to remove any possibility of a phenotype arising as a result of targeting an endogenous gene implicated in normal haematopoiesis. In this way, all changes in lineage output and gene expression could be attributed to the genome targeting approach itself. It was hypothesised that within the HSPC population targeted, there might be certain populations more affected by apoptosis induction caused by the presence of DSBs, for example LT-HSCs which are cycling less and therefore performing less DNA replication, might be less efficient at repairing DSBs induced by CRISPR, hence be less likely to survive the process. This would lead to a lower level of LT-HSC engraftment *in vivo* and lower peripheral blood lineage output. In this way, specific sub-populations were predicted to be enriched in edited samples, and this was expected to be reflected in an RNA-Seq signature that over-represented ST-HSC/MPP in edited samples compared to controls. Another hypothesis was that genome editing might alter the transcriptional profile of HSCs in such a way that produced a signature that didn't correlate with any known sub-population gene expression profiles, thus generating a new HSC signature of editing. One final prediction was that HSC

function might be altered. It was hypothesised that in that case that LT-HSC were preserved through genome editing, their functional output might be modified. For example, edited HSC might have a higher propensity to produce certain lineages above others. A real benefit of targeting the tdTomato fluorescent reporter and competitively transplanting with tdTomato positive counterparts was that the proportion of edited cells compared to non-edited competitors contributing to each mature lineage could be tracked over time.

1.5.3 Aim 3: To model JMML associated *NF-1* loss of function in human HSPCs

In order to develop a method to model JMML progression from its initiating mutation in a human cell context, a CB *NF-1* targeting approach was developed. The rationale behind this design was to address a gap in current transplant models in which the disease is modelled both in a fetal-like human context and from the point of acquisition of the first initiating mutation. While valuable models currently exist for this disease (Krombholz, Aumann et al. 2016) (Yoshimi, Balasis et al. 2017) (Le, Kong et al. 2004) (Cutts, Sjogren et al. 2009), they don't combine both the human fetal-like context with disease initiation. By targeting *NF-1* in human CB it was possible to assess *in vitro* phenotypic consequences of loss of function of this tumour suppressor as well as monitor disease progression in transplant recipients. The key questions I wanted this novel system to provide the opportunity to address were three-fold. Firstly, the importance of the human context of the *NF-1* initiating event; crucially was it possible to reduce the latency of disease onset to one that more closely resembled a childhood disease in the mouse model setting by transplanting edited human HSPCs as opposed to activating this mutation in the mouse HSPC context? Secondly, the importance

of the fetal-like context for acquisition of the initiating mutation; importantly does a JMML-like disease only develop when *NF-1* is inactivated in a fetal-like context and not in an adult HSPC? Finally, is a single RAS pathway hit in human fetal-like HSCs sufficient to cause JMML-like disease development in a xenograft setting or are secondary mutations required; and if these secondary mutations are required, does a single *NF-1* hit provide the landscape to promote the acquisition of additional mutations?

CHAPTER 2: METHODS

2. METHODS

2.1 Murine and human samples

2.1.1 Murine strains

Murine strains were maintained by the John Radcliffe Hospital Biomedical Services (BMS) team under Home Office compliant conditions in individually ventilated cages according to the licence of Adam Mead (PPL- P2FF90EE8 2019 onwards) and Sten Eirik Jacobsen (PPL-30-3103 2016-2019).

i) Ai9 (RCL-tdT)

For tdTomato induction experiments the Ai9 (RCL-tdT) mouse strain on the C57BL/6J CD45.1 genetic background was used (JacksonLaboratory 2021) (Madisen, Zwingman et al. 2010). The tdTomato pseudo dimer is formed of two repeat regions joined by a short adapter sequence. It was derived using directed evolution of dsRed to generate a brighter version (Shaner, Campbell et al. 2004). In this mouse strain, it is controlled by the CAG promoter which lies downstream of a loxP-STOP-loxP cassette. This construct was inserted at the Gt(ROSA)26Sor locus such that expression is global. For the purpose of this thesis, this strain is referred to as Ai9-Rosa26.

ii) Rosa26-tdTomato

For tdTomato knockout experiments, an adaption of the Ai9-Rosa26 strain was used. This strain was generated by Ruggiero Norfo in the lab of Adam Mead by breeding of Ai9-Rosa26 with the CMV-Cre mouse strain (JacksonLaboratory 2021). Cre-positive, high tdTomato expressing mice were then selected to breed

out the CMV-Cre component. This yielded constitutive tdTomato mice in which the loxp-STOP-loxP component was permanently removed.

iii) C57BL/6

For sham-edited, mock-edited and fresh control transplant samples, the C57BL/6 CD45.2 in-bred strain with no germline alterations was used (JacksonLaboratory 2021).

iv) NSG

For transplant of human cord blood samples, the immune-deficient mouse strain NSG (JacksonLaboratory 2021), harbouring severe combined immune deficiency (SCID) and loss of the IL-2 receptor gamma chain on the NOD/ShiLtJ genetic background, was used. This immunodeficiency allows the support of human bone marrow and blood without rejection (Theocharides, Rongvaux et al. 2016).

2.1.2 Human samples

i) Cord blood

Cord blood was obtained from the NHS BT bank (CB samples 925, G180105121455, 326, 700M) with the appropriate informed consent (REC number: 21/LO/0195) and kindly processed to mononuclear cells (MNCs) by colleagues of Peng Hua, a post-doc in the lab of Irene Roberts, or purchased as processed MNCs (multiple donors) from Stem Cell Technologies (**Cat no. 70007.2**: Lot nos. 2007417009, 2007416006, 2007402007, 2007416007, 2008405003, 2009425031, 2008404009, 20047417012, 2009425031, 2009410007, 2007429005, 2009410007).

2.1.3 Cell lines

The Ba/F3 cell line was obtained from cultures previously expanded by lab members and stored in liquid nitrogen. The TF-1 cell line was obtained from cultures previously expanded by Alba Meira, a DPhil student in the lab of Adam Mead. The original cell clone was a kind gift from Tim Somervaille at the Cancer Research UK Institute (CRUK) in Manchester.

2.2 Optimisation of CRISPR in primary mouse HSPCs using *in vitro* methods

2.2.1 Guide RNA design

i) Inducible tdTomato

Guide RNAs were designed against the loxP sites of the loxP-STOP-loxP cassette prior to my joining the lab by Philip Hublitz, head of the WIMM Genome Engineering facility. As this sequence was in a genomic region that is not part of any species genome databases, the gRNAs were designed in Snapgene using known PAM requirements of Cas9 (NGG for the most commonly used *S.pyogenes* strain, used in this project). The UCSC BLAT and BLAST tools were used to ensure no complementarity to the mouse genome. For the vector map containing guide sequences refer to *Results section 3.1.2* ***Figure 3.1*****. Table 2.1** summarises the guide sequences.

Guide RNA	Sequence (5'-3')
1- 5' loxP 1	AAGAAGTTCCTATTCTCTA
2- 5' loxP 2	AGAGGATCTGCGACTCTAG
3- 5' loxP 3	TGCAGATCTGCGACTCTAG
4- 3' loxP 1	AGCTAGATCGAATTCGGC
5- 3' loxP 2	ATCAAGCTGATCCGGAACCC
6- 3' loxP 3	AAGCTAGATCGAATTCGGC

Table 2.1: tdTomato induction guide RNA sequences

ii) tdTomato knockout

Guide RNAs were designed against the tdTomato sequence using the tool developed by Benchling (Benchling 2018). This tool allows the design of gRNAs against input coding sequences or selected genes therefore the input tool was used as the tdTomato sequence is not a standard genomic region. The PAM sequence was set to *Sp*Cas9 5'-NGG-3' to extract gRNAs with the required 3-bp recognition sequence. GRNA parameters were set to: single guide, 20nt guide length, 5'-NGG-3' PAM. Guides with the top three on-target scores were selected for testing. Primers were designed using the Primer 3 web tool (Primer3 2021) such that the amplicon size was 800-900bp with the gRNA targeting region in the centre. **Table 2.2** summarises the gRNAs and primer sequences. Refer to *Results Section 3.1.2* ***Figure 3.2*** for sequence region and guide positions. Guide RNA sequences of non-targeting control sequences are also detailed. Potential off-target sites for the optimal gRNA chosen for downstream applications were assessed using Cosmid (Cradick, Qiu et al. 2014). Cosmid is an online tool for predicting on- and off-target scores of gRNAs. Off-target scores are based on the number of mismatches, as well as the position of the mismatches within the gRNA as Cas9 tolerates certain mismatches more easily than others, making

mismatches at these positions more likely to lead to off-target events (Cradick, Qiu et al. 2014). Parameters were set to Mus Musculus GRCm38 (mm10), PAM 5'-NGG-3', maximum no. of indels and mismatches (to yield the most comprehensive resulting list).

Guide RNA/Primer	Sequence (5'-3')
1- tdTomato guide RNA 1	GAGTTCGAGATCGAGGGCGA
2- tdTomato guide RNA 2	GGCCACGAGTTCGAGATCGA
3- tdTomato guide RNA 3	GCCAAGCTGAAGGTGACCAA
4- tdTomato forward primer	CGTGCTGGTTATTGTGCTGT
5- tdTomato reverse primer	CCGTCCTTCAGCTTCAGGG
6- Non-targeting guide RNA 1 (loxP STOP cassette)	AAGTAAAACCTCTACAAATG
7- Non-targeting guide RNA 2 (Drosophila Gal-4)	AACGACTAGTTAGGCGTGTA

Table 2.2: tdTomato knockout and non-targeting guide RNA sequences and tdTomato PCR primers

Primer design against the resulting off-target sequences was carried out in parallel using the COSMID tool. Parameters were set to: min separation un-cleaved to cleaved- 200, min cleavage product size difference- 0, min amplicon length- 600, max amplicon length- 800, optimal amplicon length- 700. **Table 2.3** summarises the top three predicted off-targets and primers designed against these targets.

Genomic region/primer	Sequence (5'-3')
Off target 1- Chr7:65070254-65070276-	CACAAGCTGAAGGTGACCAAGGG
Off target 2- Chr13:69599588-69599610	GAAGAGCTGAAGGTGACCAATGG
Off target 3- Chr12:91294119-91294141	TCAAAGCAGAAGGTGACCAAAGG
Off target 1- forward primer	CGCCTGACATGTGTGATCAACAGC
Off target 1- reverse primer	AGCTGGCTCCAGCCAGTCAT
Off target 2- forward primer	TGAACTCAAGACCTCTGCAAGAGC
Off target 2- reverse primer	AGCCTATGCTCCAACCTAGCTAAG
Off target 3- forward primer	TTTCTTAGACAACACCCAGGGGGC
Off target 3- reverse primer	AGCCTGTGGGAGCCATTCTTATTC

Table 2.3: Cosmid predicted tdTomato guide RNA 3 off-targets and PCR primer sequences

2.2.2 Guide RNA cloning

i) Single guide insertion into the PX458 CBh-Cas9 plasmid

Cloning of tdTomato induction guide RNAs was performed prior to my joining the lab by Philip Hublitz. Cloning of the tdTomato knockout guide RNAs was performed under the guidance of Philip Hublitz. The PX458 CBh-Cas9 plasmid (Ran, Hsu et al. 2013) (Addgene 2021) contains the SpCas9 sequence followed by eGFP with the T2A sequence intervening such that the fusion is cleaved. It also codes for a nuclear localisation sequence, U6 promoter and scaffold suitable for gRNA insertion, as well as Ampicillin resistance for bacteria selection.

The reverse complement of each gRNA sequence was determined in order to prepare single stranded oligonucleotides (oligos) for cloning. At the 5' end of the forward cloning oligo, the CACCG sequence was added and at the 5' end of the reverse cloning oligo the AAAC sequence was added to allow for cloning using BbsI (New England Biolabs, NEB). Single stranded oligos were purchased from ThermoFisher Scienctific at 0.025µmol in the desalted format. Desalted oligos were dissolved to 100µM in nuclease free water (multiple vendors). 10µL of each forward and reverse oligo was added to 10µL NEB buffer 2.1 and 70uL nuclease free water followed by incubation at 99ºC for five minutes. Annealed oligos were slowly cooled by switching off the heat block and leaving the tubes for five minutes, followed by lying on top of the block for five minutes. Double stranded (ds) oligos were precipitated by the addition of 30µL 3M NaOAC pH 5.6 and 300µL 100% ethanol followed by centrifugation at 21,000g for 30 minutes at 2ºC. The supernatant was discarded and the pellet washed once with 70% ethanol followed by centrifugation at 21,000g for 10 minutes at room temperature and

removal of supernatant. Pellets were air dried and ethanol removed with a tissue followed by resuspension in nuclease free water. 1µg of dsoligo was incubated with T4 polynucleotide kinase (PNK) (NEB) for one hour at 37ºC. Ligation was performed at room temperature for one hour in a 10µL final reaction volume using 20ng PX458 backbone (pre-digested with BbsI and gel purified) and 25ng dsoligo. 5µL of the ligation reaction was used to transform One Shot Stbl3 competent *E.Coli* (ThermoFisher Scientific C737303) by incubating on ice for 20 minutes followed by a two minute heat shock (42ºC), then returning to ice for five minutes. 2mL of Lysogeny broth (LB) growth medium was added followed by incubation for one hour at 37ºC with mild shaking (800rpm). Samples were centrifuged at 1,000g for five minutes followed by plating onto agar plates containing ampicillin and incubated overnight at 37ºC. Colonies were picked and grown up in 1.4mL LB broth with 0.1% ampicillin at 37ºC overnight. DNA was isolated from clones and sent for sequencing according to the method detailed below *2.2.2 ii)*.

ii) DNA purification from guide RNA plasmid and viral vector clones

Transformed clones were centrifuged at 16,000g for one minute at room temperature. The pellet was resuspended in 200µL Buffer P1 (Qiagen 19051) followed by vortexing five times, five minutes each time. 200µL Buffer P2 (Qiagen 19052) was then added followed by careful inversion and incubation for three minutes at room temperature. 200µL of chilled Buffer P3 (Qiagen 19053) was then added followed by five times inversion. Samples were then centrifuged at 16,000g for 10 minutes at 4ºC. The supernatant was transferred to a fresh tube

followed by the addition of 500µL of isopropanol and vigorous mixing. Samples were then centrifuged at 16,000g for 10 minutes at 4ºC and the supernatant discarded. 500µL of 70% room temperature ethanol was then added, left on the pellet for five minutes and then discarded. The sample tube was then drained on a Kimwipe, centrifuged at 16,000g for five minutes followed by removal of residual ethanol. DNA was eluted in 50µL nuclease free water and vortexed at 2000rpm for five minutes. Samples were sent for in-house Sequencing at 3ng/µL using one of the following sequencing primers: 5'GCCTGGTATCTTTATAGTCC3' (PX458 single guide insertion), 5'CCTTCATATTTGCATATACG3' (EFS-Cas9 and SFFV-Cas9 guide 1) or 5' TAAAATTGTGGATGAATACT3' (EFS-Cas9 and SFFV-Cas9 after cloning of guides 1 and 2).

iii) Single and dual guide insertion into the EFS-Cas9 and SSFV-Cas9 viral vectors

For virus transduction, tdTomato gRNAs were cloned as single guides (tdTomato knockout) or dual guides (tdTomato induction) into the EFS-Cas9 (Addgene 57818) (Addgene 2021) (Heckl, Kowalczyk et al. 2014) and SFFV-Cas9 (Addgene 57827) (Addgene 2021) (Heckl, Kowalczyk et al. 2014) backbones. These vectors contain the sequence for *Sp*Cas9 under either the EFS or SFFV promoter, followed by the eGFP sequence with intervening P2A which allows ribosomal skipping hence separate translation of the two proteins. This vector also contains an ampicillin resistance sequence for selection as well as U6 promoter and scaffold sequence for gRNA transcription.

The EFS-Cas9 and SFFV-Cas9 vectors (10µg each) were first digested with BsmBI (3 units) in 1X NEB buffer 3.1, diluted with nuclease free water to a total volume of 50µL. Digestion was carried out for one hour at 55ºC followed by heat inactivation for 20 minutes at 80ºC. Digested plasmids were gel purified and diluted to 20ng/µL. Dsoligo preparation, ligation and transformation into One Shot Stbl3 competent *E.Coli* was performed as described in *2.2.2 i)*. DNA from single guide insertion clones was purified as described in *2.2.2 ii*. Samples were sent for in-house Sanger sequencing at 3ng/µL using the U6 promoter primer: 5'CCTTCATATTTGCATATACG3'. To clone the second gRNA, the sequence was initially amplified from the PX458 backbone in which it was previously cloned into using the following primers: forward primer 5' GGTCACGAATTCGAGGGCCTATTTCCCATGATTCCTT 3', reverse primer 5'ACATTGGAATTCAAAAAAGCACCGACTCGGTGCCACT 3'. The PCR reaction was carried out using the reagent combination detailed in **Table 2.4** following the PCR program detailed in **Table 2.5**.

Component	**Volume per sample (µL)**
5X Q5 reaction buffer (NEB)	10
10mM dNTPs (ThermoFisher)	1
10µM Forward primer	2.5
10µM Reverse primer	2.5
Template DNA	1 ng
Q5 HF DNA Pol (NEB)	0.5
Nuclease free water	Up to 50
Total	50

Table 2.4: PCR reagent components for guide RNA amplification from the PX458 plasmid

Step	Temperature (°C)	Time
Denaturation	98	30 seconds
30 cycles	98 72 72	5 seconds 10 seconds 20 seconds
Extension	72	2 minutes
Hold	4	Hold

Table 2.5: PCR program for guide RNA amplification from the PX458 plasmid

1µL of DpnI (NEB) was then added directly to the PCR reaction and incubated for one hour at 37ºC followed by heat inactivation at 80ºC for 20 minutes. The PCR product was then purified using the Monarch PCR DNA Cleanup Kit (NEB T1030) following manufacturer's instructions. Direction ligation into the EFS-Cas9 and SFFV-Cas9 backbones containing single guide RNAs was unsuccessful so the second guide was then ligated into the pBSSK backbone (Addgene 2021) as an intermediate step. pBSSK was linearized using EcoRV (NEB) by incubating with CutSmart buffer (NEB) for one hour at 37ºC. The PCR product was ligated into the pBSSK backbone using 20ng backbone and either 7.6ng or 12.6ng PCR product as the optimal ratio was not known. Ligation and One Shot Stbl3 *E.coli* transformation was performed as described in *2.2.2 i)*. pBSSK ligated with the second guide RNA was then digested with EcoRI HF (NEB) using one unit of enzyme in 1X CutSmart buffer (NEB) to a total volume of 50µL in nuclease free water. Digestion was performed at 37ºC for one hour followed by heat inactivation for 20 minutes at 65ºC. The EFS-Cas9 and SFFV-Cas9 backbones containing gRNA 1 were digested with EcoRI HF (NEB) using three units of enzyme in 1X Cutsmart buffer (NEB) to a total volume of 50µL with nuclease free water. Digestion was carried out at 37ºC for one hour followed by

heat inactivation at 65ºC for 20 minutes. The backbones were then dephosphorylated using 2.5 units of calf intestinal phosphatase (NEB) by incubating at 37ºC for 30 minutes. Both the backbones and gRNA PCR inserts were gel purified. Ligation and transformation was carried out as described in *2.2.2 i)* using 25ng of backbone and either 1.5 or 4ng of insert as the optimal ratio was not known. DNA from clones was prepared as described in *2.2.2 ii)*. Clones were sent for in-house Sanger sequencing at 3ng/µL to confirm dual guide insertion using the following primer: 5' TAAAATTGTGGATGAATACT 3'.

2.2.3 Single and dual guide RNA vector virus production

Using the EFS-Cas9 and SFFV-Cas9 backbones containing the single or dual tdTomato knockout and induction guides resepectively, highly concentrated virus particles were kindly synthesised by Ryan Beveridge, head of the WIMM Virus Screening facility. In addition to the EFS-Cas9 and SFFV-Cas9 vectors, the psPAX2 (Didier Trono Lab) (Addgene 2021) and pMD2.G (Addgene 2021) vectors, containing the gag and pol proteins and VSVG envelope proteins respectively, were also used to prepare virus. These separate vectors allowed replication and virus particle production in the infected cell while preventing successful infection of new cells.

Initially HEK293T cells were grown in DMEM (Thermofisher 11320033) supplemented with 2mM glutamine (Thermofisher 25030149), 1% Penicillin/Streptomycin (Thermofisher 15140122), 1X NEAA (Thermofisher 11140050) and 10% FCS (Sigma Aldrich F9665) in a T175 flask to 80-90% confluence in a 37ºC, 5% CO_2 tissue culture incubator, washed with PBS

(ThermoFisher 10010023) and removed from the flask using 5mL trypsin by incubation for five minutes in the tissue culture incubator. 3-4x10^7 cells were pelleted followed by removal of the supernatant and re-suspension in 20mL DMEM with supplements as described above. Cells were counted and 1.2x10^7 cells were seeded in a 15cm plate and incubated overnight. The next day, a cocktail of 13.3µg of EFS-Cas9 or SFFV-Cas9, 10µg of psPAX2 and 6.7µg of pMD2.G was prepared in 1.5mL Opti-MEM (Thermofisher 31985062) alongside an additional 1.5mL of Opti-MEM containing 30µL of PEI-Pro (Polyplus transfection 115-010). The plasmid containing and PEI-Pro containing Opti-MEM tubes were then mixed by pipetting and incubated at room temperature for 15 minutes. Cells were transfected by dropping the Opti-MEM, plasmid and PEI-Pro mix onto the plates. Five to six hours post-transfection, the culture medium was removed and replaced with 10mL of fresh warm DMEM with supplements as described above. 48 hours after transfection, the supernatant was removed from the plates and stored at 4ºC. The plates were then replenished with 10mL of fresh warm DMEM with supplements and returned to the tissue culture incubator. 72 hours after transfection, the second supernatant was removed and added to the first. The combined supernatants were centrifuged at 2,000g for five minutes at 4ºC. The supernatant was then filtered using a 0.45µM cellulose acetate syringe filter (StarLab E4780-1453). 5mL of 20% sucrose (Sigma Aldrich RDD023) was then added to an ultracentrifuge tube followed by slowing adding the viral supernatant. Viral supernatants were centrifuged for two hours at 90,000g, max acceleration at 4ºC. The supernatant was removed and the pellet was resuspended in PBS (ThermoFisher 10010023) by pipetting for one to two

minutes, covering with parafilm and transferring to ice for 120 minutes with periodic agitation every 20 minutes. 50µL aliquots were then stored at -80ºC.

2.2.4 HEK293T cell line transfection by lipofection

HEK293T cells were maintained in DMEM with 10% FCS (Sigma Aldrich F9665), 1% Penicillin/Streptomycin (Thermofisher 15140122) and 2mM glutamine (Thermofisher 25030149) in a tissue culture incubator at 37ºC, 5% CO_2. Cells were split 1:10 every two to three days to maintain appropriate confluence. 12-15 hours prior to transfection, cells were seeded at $2.5x10^5$ cells per well in a 12-well tissue culture plate (Corning 10253041) in DMEM with supplements as described above. The following morning, 2µg of the Ai9 (Addgene 2021) (Madisen, Zwingman et al. 2010) vector and either 0.5 or 1µg of each 5' and 3' loxP guide RNA containing PX458 vectors (three vectors in total) were mixed with 50µL DMEM (**without supplements**). In a separate tube, 50µL DMEM (**without supplements**) was combined with 10µL lipofectamine$_{2000}$ (Life Technologies, 11668019), carefully flipped to mix and incubated for five minutes. The DNA and lipofectamine containing tubes were then mixed and incubated at room temperature for 20 minutes to allow liposome formation. Prior to lipofection, medium was removed from the HEK293T cells and replaced with fresh DMEM containing supplements (as above). 100µL transfection solution (DMEM, DNA, lipofectamine$_{2000}$) was then added to cells, followed by gentle swirling and incubation overnight in the tissue culture incubator. The medium was then removed and replaced every other day up until analysis. Transfection and editing efficiencies were measured by GFP and tdTomato expression respectively three

days post lipofection by flow cytometry. Viability was measured using DAPI (ThermoFisher 62248). Preparation for flow cytometry and analysis was carried out as described in *2.2.11*.

2.2.5 Ba/F3 cell line transfection optimisation

i) Electroporation voltage and plasmid amount optimisation

The murine pro-B Ba/F3 cell line cell was maintained in T25 flasks at a concentration between $4x10^4$ and $7x10^5$ cells/mL in RPMI (ThermoFisher 1640) supplemented with 10% FCS (Sigma Aldrich F9665), 1% Penicillin/Streptomycin (Thermofisher, 15140122) and 10ng/mL mouse IL-3 (BioLegend, 575506). Cells were passaged every two to three days as appropriate.

Prior to transfection, 48-well tissue culture plates (Corning 3548) were prepared with 500µL per sample of RPMI 1640 with all supplements as above except antibiotics and left to warm for at least 30 minutes in a tissue culture incubator set at 37ºC, 5% CO_2. Ba/F3 cells were diluted 1:2 with trypan blue (ThermoFisher 15250061) and counted on a haemocytometer. 200,000-250,000 cells were aliquoted per sample and centrifuged at 500g for five minutes at room temperature followed by removal of the supernatant. Cells were washed with an equal volume of PBS (ThermoFisher 10010023) and centrifuged at 500g for five minutes at room temperature followed by resuspension in 10µL resuspension buffer R, (Neon Transfection System 10µL kit: ThermoFisher MPK1025). Cells were then mixed with 0.4µg of room temperature pmaxGFP plasmid (Lonza SG Cell Line Transfection kit: Lonza V4XC-3032). The Neon transfection system (ThermoFisher MPK5000) was set up with 3mL electrolytic buffer E (Neon

Transfection System 10µL kit: ThermoFisher MPK1025) and set with the respective voltage, time and pulse conditions (voltage range tested: 950-2000V milliseconds tested: 20-30, pulses tested: 1-2). The cells were then transferred to the Neon pipette (ThermoFisher MPP100) to perform electroporation. The cells were left in the tip for 10 minutes followed by transfer to the pre-warmed culture medium. Immediate viability was counted using a 1:2 trypan blue (ThermoFisher 15250061) dilution. 24 hours post-electroporation cells were centrifuged for five minutes at 100g at room temperature and resuspended in RPMI with supplements including antibiotics. GFP positivity was analysed 24 hours post-transfection by microscopy and 48 hours post-transfection by flow cytometry. Flow cytometry preparation and analysis was carried out as detailed in *2.2.11*. Once the optimal voltages were chosen, the process was repeated with the same voltage condition but varying the plasmid amount (between 0-10µg).

ii) Nucleofection plasmid amount optimisation

At the time of carrying out this project, a Ba/F3 buffer and nucleofection program had been pre-optimised and the protocol was available commercially (Lonza 2021). Therefore optimisation was carried out to determine the desired plasmid amount. Using data from the electroporation optimisation analysis, the plasmid range was narrowed to 0-6µg.

Ba/F3 cells were maintained as described in *2.2.5 i)*. At least 30 minutes prior to nucleofection, 24-well culture plates were prepared with 500µL of RPMI with supplements as described above and placed in a tissue culture incubator at 37ºC, 5% CO_2. At the same time, the SG cell line solution and supplement (Lonza

V4XC-3032) were brought to room temperature. Ba/F3 cells were counted 1:2 in trypan blue (ThermoFisher 15250061) and aliquoted at 200,000-250,000 cells per sample followed by bringing to 1.5mL with PBS (ThermoFisher 10010023). Cells were centrifuged at 300g for seven minutes at room temperature. Cells were then resuspended in 20µL SG solution (with supplement added) per sample and mixed with the specific plasmid amount (0-6µg) followed by transfer to separate wells of a 16-well cuvette strip (Lonza V4XC-3032). The 4D nucleofector (Lonza AAF-1002X) was set to SG solution, program CM-147 and the samples run, followed by transfer to the pre-incubated medium. Transfection efficiency was measured by flow cytometry 24 hours post-nucleofection. Flow cytometry preparation and analysis was carried out as detailed in *2.2.11.*

2.2.6 tdTomato induction in primary mouse HSPCs: Neon electroporation

Neon electroporation was compared to Amaxa nucleofection and viral transduction for both the tdTomato induction and knockout editing approaches. In the case of Neon electroporation and Amaxa nucleofection, both plasmid and RNP delivery was compared. Previous optimisation carried out in the Ba/F3 cell line was used to refine the range of voltage conditions in order to reduce the requirement for primary mouse material. Total plasmid amounts were previously optimised in the Ba/F3 cell line and therefore not re-optimised for primary cells, as this is directly related to cell number which was similar for both methods.

i) Isolation of c-Kit enriched mouse BM

One day prior to electroporation, the tibiae and femurs were obtained from the required number of Ai9-Rosa26 adult (6-12 weeks) mice. Bones were crushed

with a pestle and mortar in MACS buffer (PBS (ThermoFisher 10010023), 2% FCS (Sigma Aldrich F9665), 1mM EDTA (Sigma Aldrich, E7889)) and put through a 70µM cell strainer (Greiner Bio-One 542070). Cells were then centrifuged for five minutes at room temperature followed by resuspension in MACS buffer to 300µL/100 million cells. 7.5µL of c-Kit beads (Miltenyi 130-091-224) per 100 million cells were then added and the cells incubated for 20 minutes on ice with careful inversion every two minutes. Cells were washed with 3mL MACS buffer and centrifuged for five minutes at 500g. LS MACS column(s) (Miltenyi 130-042-401) were then washed with 3mL MACS buffer followed by resuspension of the cells in 3mL MACS buffer and addition to the column through a 100µM falcon filter (Greiner Bio-One). The column was then washed once with MACS buffer followed by removal of the column from the magnet and flushing of the cells through with 10mL MACS buffer. Cells were then centrifuged at 500g for five minutes at room temperature followed by resuspension in 1mL of StemSpan SFEM (Stem cell Technologies, 09650) per million cells with the following cytokines: 50ng/mL SCF (Prepotech 250-03), 50ng/mL TPO (Prepotech 300-18), 10ng/mL IL-3 (BioLegend 575506), 10ng/mL IL-6 (RND systems, 206-IL), 50ng/mL Flt3 Ligand (Miltenyi 130-096-479). Cells were incubated at 37ºC, 5% CO_2 in 12-well plates (Corning 3513) for 24 hours prior to transfection.

ii) Plasmid and RNP mouse HPSC Neon electroporation

24 hours post c-Kit selection, cells were removed from the 12-well plate using MACS buffer and aliquoted at 200,000-250,000 cells per sample. Samples were pelleted for five minutes at room temperature and resuspended in 10µL buffer T

(Neon Transfection System 10µL kit, ThermoFisher MPK1025). PX458 plasmids containing 5' loxP guide RNA 1 and 3' loxP guide RNA 6 (0.5µg each) were combined at a maximum volume of 1µL (10% of total electroporation volume) and mixed with the cell sample. loxP guide RNA 1 and loxP guide RNA 6 (Synthetic TruGuide modified sgRNA, Invitrogen) were prepared as RNPs by aliquoting 15pmol of each guide RNA into **separate** PCR tubes followed by heating at 95°C for two minutes, returning to ice for two minutes and centrifuging at 16,400g for one minute. Guides were then separately combined with 0.5µg spCas9 WT V2 (Life Technologies A27865) and incubated for 10 minutes at room temperature. The two pre-complexed 5' and 3' loxp RNPs were then combined with the cell sample. Samples were then transferred to the Neon pipette and the selected program run (voltage range: 1400-1700V, millisecond range: 10-20 pulse range: 1-3). Cells were then placed in a 24-well plate containing 500µL pre-warmed StemSpan with appropriate cytokines as described above. Viability was assessed immediately, 24 hours and 48 hours post-transfection by counting samples diluted 1:2 in trypan blue (ThermoFisher 15250061). Transfection (by GFP positivity) and editing (by tdTomato positivity) was measured 48 hours post-transfection by flow cytometry using the antibody panel detailed in **Table 2.6**. Flow cytometry preparation and analysis was carried out as detailed in *2.2.11.* Empirical evidence (data not shown) determined c-Kit labelling by flow cytometry to be unreliable post c-Kit enrichment. It was predicted that this was due to internalisation of the receptor after c-Kit bead labelling or saturation by c-Kit bead antibodies. Therefore Lin- Sca-1+ (LS) were gated within the total c-Kit gate before gating for CD48- CD150+ HSCs. Gating strategies used are shown in the appropriate results sections.

Antibody	Fluorochrome
CD48	APC
Lineage: F4/80, CD3, Mac-1, NK1.1, B220, Ter119, Gr1	PECY5
Viability	7-AAD
c-Kit	APC-ef780
CD150	PECY7
Sca-1	PB

Table 2.6: Antibody panel used to stain primary mouse HSPCs for *in vitro* transfection and editing efficiency analysis

2.2.7 tdTomato induction in primary mouse HSPCs: Amaxa nucleofection

At the time of these experiments, pre-optimised buffer and voltage conditions were commercially available for nucleofection of mouse c-Kit+ cells and therefore these were used to compare to electroporation.

24 hours prior to nucleofection, femurs and tibiae were obtained from the appropriate number of Ai9-Rosa26 adult mice. c-Kit enriched cells were prepared as described in *2.2.6 i)*. 200,000 (+10%) cells were centrifuged per sample at 500g for five minutes at room temperature. Plasmid and guide RNAs were prepared as described in *2.2.6 ii)*. Cells were resuspended in 20µL primary cell kit P3 buffer (Lonza V4XP-3032), combined with the plasmid/RNP combinations and transferred to separate wells of the 16-well strip cuvette (Lonza V4XP-3032). The primary cell P3 DK100 program was then run on the 4D Amaxa nuclefection unit followed by transfer to a 24-well plate containing 500µL pre-warmed StemSpan with appropriate cytokines as described above. Viability was assessed immediately, 24 hours and 48 hours post-transfection by counting samples diluted 1:2 in trypan blue (ThermoFisher 15250061). Transfection (by

GFP positivity) and editing (by tdTomato positivity) was measured 48 hours post-transfection by flow cytometry using the antibody panel detailed in **Table 2.6**. Flow cytometry preparation and analysis was carried out as detailed in *2.2.11.* Empirical evidence (data not shown) determined c-Kit labelling by flow cytometry to be unreliable post c-Kit enrichment. It was predicted that this was due to internalisation of the receptor after c-Kit bead labelling or saturation by c-Kit bead antibodies. Therefore Lin- Sca-1+ (LS) were gated within the total c-Kit gate before gating for CD48- CD150+ HSCs. Gating strategies used are shown in the appropriate results sections.

2.2.8 tdTomato induction in primary mouse HSPCs: Viral transduction

24 hours prior to transduction, femurs and tibiae were obtained from the appropriate number of Ai9-Rosa26 adult mice. c-Kit enriched cells were prepared as described in *2.2.6 i)*. On the same day, retronectin coated plates were prepared by diluting rectronectin (Takara T202) in PBS (ThermoFisher 10010023) to 25µg/mL and placing 100µL in separate wells of a non-TC treated 96 well plate (Corning 351172). Plates were left overnight at 4°C.

24 hours post c-Kit enrichment, EFS-Cas9 and SFFV-Cas9 containing 5' loxP guide RNA 1 and 3' loxP guide RNA 5 as well as the small control vector PEGZ2 linker B322 were prepared to a multiplicity of infection (MOI) of 30-70 and diluted to 100µL in StemSpan without cytokines (**NB:** although 3' loxP guide RNA 6 was optimal in HEK293T experiments, it was not possible to clone this sequence as a dual guide system so the next most optimal guide RNA 5 was used). Below is an example of MOI calculations:

MOI of 50 means 50 viral particles per cell.

For 200,000 cells therefore requires 10 million particles.

For a virus of titre $3.32x10^8$ particles/mL: $(10x10^6)/ (3.32x10^8)$ = 0.03mLx1000= 30µL.

Retronectin solution was then removed from the coated plates and washed with 200µL FACS buffer (PBS (ThermoFisher 10010023), 5% FCS (Sigma Aldrich F9665)). After removal of the FACS buffer 50µL of virus suspension was added to the retronectin plates and centrifuged at 1,500g for one hour at 32°C followed by removal of the medium. 50,000 cells per sample were washed with MACS buffer and centrifuged at 500g for five minutes at room temperature followed by resuspension in 50µL of StemSpan with 2x cytokines (as described above) and added to the remaining 50µL of pre-diluted virus. Samples were centrifuged for one hour at 500g at 32°C followed by transfer to a tissue culture incubator. At the end of the day, another 100µL of StemSpan with 1x cytokines was added to the plate. Transduction (by GFP positivity) and editing (by tdTomato positivity) was measured three days post-transduction by flow cytometry using the antibody panel detailed in **Table 2.6**. Flow cytometry preparation and analysis was carried out as detailed below in to *2.2.11.* Empirical evidence (data not shown) determined c-Kit labelling by flow cytometry to be unreliable post c-Kit enrichment. It was predicted that this was due to internalisation of the receptor after c-Kit bead labelling or saturation by c-Kit bead antibodies. Therefore Lin-Sca-1+ (LS) were gated within the total c-Kit gate before gating for CD48-CD150+ HSCs. Gating strategies used are shown in the appropriate results sections.

2.2.9 tdTomato knockout in primary mouse HSPCs: plasmid and RNP nucleofection

Having established in tdTomato induction experiments that nucleofection was superior to electroporation, this approach was chosen to compare plasmid and RNP delivery for tdTomato knockout. The tdTomato induction approach also revealed several steps requiring optimisation to reduce cell loss including reducing washes and increasing centrifugation times, therefore the following protocol contains some minor modifications.

24 hours prior to nucleofection, femurs and tibiae were obtained from the appropriate number of Rosa26-tdTomato adult mice. c-Kit enriched cells were prepared as described in *2.2.6 i)*. 200,000-250,000 cells per sample were pelleted at 350g for 10 minutes at room temperature. 1.2µg of PX458 tdTomato knockout plasmid was aliquoted per sample and Cas9 (Life Technologies) was diluted to 1.5µg/µL in NEB Diluent B (NEB B8002S). 1.6µg of tdTomato knockout guide RNA (1-3) (Synthego, synthetically modified) was aliquoted per sample into PCR strip tubes and combined with 3µg Cas9. PCR tubes were incubated at 21°C for 10 minutes followed by resuspension of the cells in 20µL Primary Cell buffer P3 per sample. Cell samples were then combined with either RNP or plasmid and transferred to separate wells of a 16-well cuvette and the Primary Cell P3 DK100 program was run on the 4D nucleofector. Cells were then transferred to a 24-well plate containing pre-warmed StemSpan medium with cytokines as described above. Nucleofection (by GFP positivity) and editing (by tdTomato negativity) was measured 48 hours post-transfection by flow cytometry using the antibody panel detailed in **Table 2.6**. Flow cytometry preparation and analysis was carried out as detailed in *2.2.11.* Empirical evidence (data not

shown) determined c-Kit labelling by flow cytometry to be unreliable post c-Kit enrichment. It was predicted that this was due to internalisation of the receptor after c-Kit bead labelling or saturation by c-Kit bead antibodies. Therefore Lin-Sca-1+ (LS) were gated within the total c-Kit gate before gating for CD48-CD150+ HSCs. Gating strategies used are shown in the appropriate results sections.

2.2.10 tdTomato knockout in primary mouse HSPCs: transduction

Transduction of mouse HSPCs (Rosa26-tdTomato strain) was carried out using tdTomato guide RNA 3 containing EFS-Cas9, SFFV-Cas9 and control PEGZ2 linker B322 vectors as described in *2.2.8* with the modification of increased cell numbers (100,000-200,000 per sample) due to the higher titre of single guide RNA containing vectors allowing an MOI of 30-50 to be achieved in the required volume for a higher cell number. Transduction (by GFP positivity) and editing (by tdTomato negativity) was measured 72 hours and seven days post-transduction by flow cytometry using the antibody panel detailed in **Table 2.6**. Flow cytometry preparation and analysis was carried out as detailed in *2.2.11.*

2.2.11 Flow cytometry preparation and analysis

Flow cytometry preparation and analysis for assessment of lipofection, transfection, transduction and editing efficiency in the HEK293T and Ba/F3 cell lines as well as primary mouse HSPCs is described below. Protocol steps specific to certain methods are noted.

Optimal antibody dilutions were determined by titration for those without pre-determined dilutions or at the company recommendation for those with flow cytometry specific dilution suggestions. Antibody master mixes, as well as master mixes containing all antibodies except one (fluorescence minus one, FMOs) were prepared at 2X the final dilution requirement in FACS buffer (PBS (ThermoFisher 10010023), 5% FCS (Sigma Aldrich F9665)). Single stains were prepared using cells for viability stains and fluorescent reporters, and beads for cell surface markers. For HEK293T and Ba/F3 cell line flow cytometry analysis, DAPI viability dye was the only marker added. Antibody panels used for mouse HSPC staining are detailed in **Tables 2.6 and 2.7**.

Cells were removed from the plate by trypsinisation (HEK293T), FACS buffer (Ba/F3) or MACS buffer (mouse c-Kit+) and centrifuged at 1,000g for five minutes at 4°C followed by resuspension in 25µL Fc block (BD, 553142) per sample. Aliquots from appropriate controls/experimental samples were separated for fluorescence minus one controls (FMOs) and resuspended in 12.5µL Fc block. All samples were incubated for at least 15 minutes at 4°C followed by addition of the appropriate FMO or master mix cocktail and an additional 15 minute incubation at 4°C. Cells were then washed with 1mL FACS buffer and centrifuged at 1000g for four minutes at 4°C followed by resuspension in 200µL FACS buffer. Cells were kept on ice until acquiring by flow cytometry and viability dye was added immediately before acquiring samples. HEK293T samples were analysed on an Attune NxT Flow Cytometer (ThermoFisher). Ba/F3 and mouse HSPC samples were analysed on either an LSR Fortessa (BD Biosciences), LSR II (BD Biosciences) or LSR Fortessa X50 (BD Biosciences) flow cytometer. Flow cytometry analysis was carried out using FlowJo version 10.6.

2.3 Molecular and functional characterisation of CRISPR-edited mouse HSCs

2.3.1 Sample selection

To allow for interrogation of each stage of the nucleofection process in terms of its unique effect on engraftment and transcriptional profile of HSCs, control groups were included to separate each step. For the pilot transplant experiment, these included a 'cultured only' sample which remained in StemSpan with the appropriate cytokines from c-Kit enrichment up until the point of transplant, a 'water only' sample which underwent the nucleofection procedure without the presence of Cas9 or a guide RNA and a 'Cas9 only' control. For subsequent experiments, it was decided to replace the Cas9 only control with a non-targeting control against the STOP cassette sequence (5' AAGTAAAACCTCTACAAATG 3') of Ai9-Rosa26 (absent in the Rosa26-tdTomato strain) or the Drosophila Gal-4 gene (5' AACGACTAGTTAGGCGTGTA 3') (absent in the mouse genome) to account for presence of the full CRISPR machinery without DSB induction. tdTomato guide RNA and non-targeting guide RNAs are detailed in **Table 2.2**. An additional fresh control was also included for 24-hour RNA-Seq and transplant experiments to account for differences caused by tissue removal and cell culture manipulation.

2.3.2 tdTomato knockout time course in mouse HSPCs

In order to assess the trajectory of tdTomato loss in specific HSPC compartments, c-Kit enriched mouse HSPCs were nucleofected with the optimal tdTomato gRNA (guide 3) and analysed by flow cytometry at 24, 48 and 72 hours

post-nucleofection. RNP nucleofection was carried as described in *2.3.9*. Specific modifications detailed below were made to allow use of the large cell number version of the protocol in order to obtain enough material to obtain cells from the same well at each time-point so replicates were matched. Cell number per sample was increased to 1,000,000. Cas9 was increased to 12µg, gRNA to 6.4µg per sample and nucleofections were carried out in 100µL Primary cell P3 Buffer in large single cuvettes (Lonza, V4XP-3024). The nucleofection program (DK 100) remained the same. Samples were plated in 1mL pre-warmed StemSpan with cytokines in 12-well tissue culture plates to maintain the appropriate cell density ($1x10^6$/mL). An expanded version of the large cell number RNP nucleofection protocol, including details of critical steps, is available in *Appendix 2*. At specific time intervals, 750µL of cells were removed and stained with the appropriate antibodies. Wells were replaced with 750µL fresh medium and incubated at 37°C 5% CO_2 until the next time-point. The antibody panel was partially modified from the initial *in vitro* optimisation in order to allow viability to be measured on a separate fluorophore to lineage, as well allow it to match the panel used for sorting, in which it was found that tdTomato was incompatible with PECY5/7-AAD. **Table 2.7** summarises the antibody panel used. Cells were prepared for flow as described in *2.2.11*.

Antibody	Fluorophore
CD48	APC
Lineage: F4/80, CD3, Mac-1, NK1.1, B220, Ter119, Gr1	FITC
c-Kit	APC-ef780
CD150	PECY7
Sca-1	BV605
Viability	DAPI

Table 2.7: Antibody panel used for tdTomato time course analysis and sorting of HSPCs for Target-Seq 24h post-nucleofection and RNA-Seq 18wk post-transplant

2.3.3 Cell cycle status analysis of tdTomato edited HSPCs

These experiments were carried out with the help of Zemin Ren, a postdoctoral researcher in the lab of Adam Mead. Mouse HSPC (Rosa26-tdTomato strain) nucleofection was carried out according to the large cell number protocol described in *2.3.2* and described in detail in *Appendix 2*.

Cells were stained using the stem cell panel detailed in **Table 2.7** following the flow cytometry preparation method described in *2.2.11*. Following the final FACS buffer wash, cells were resuspended in 250µL of fixation/permeabilisation solution (BD 554714) and incubated for 20 minutes in the dark at 4°C followed by centrifugation for five minutes at 1500g. The supernatant was discarded and the pellet washed with 750µL 1x permeabilisation/wash buffer (BD 554714) followed by centrifugation at 1500g for five minutes. The pellet was then resuspended in 500µL of Ki-67 AF700 (Biolegend 652420) diluted 1:100 in 1x permeabilisation/wash buffer and incubated for 30 minutes in the dark at 4°C followed by washing with 500µL of 1x permeabilisation/wash buffer and centrifugation for five minutes at 1500g. The pellet was then resuspended in 500µL of 1x permeabilisation/wash buffer with 5µL 7-AAD (Thermofisher A1310)

and incubated for 10 minutes in the dark at 4°C. Samples were then centrifuged for five minutes at 1500g and resuspended in 300µL FACS buffer for analysis. Flow cytometry analysis was carried out on an LSR II or LSR Fortessa analyser and analysis on FlowJo version 10.6.

2.3.4 In vivo transplantation of tdTomato edited HSPCs

RNP nucleofection was carried out on c-Kit enriched mouse BM as described in *2.2.9* for the pilot transplantation experiment and as described in *2.3.2* (detailed version available in *Appendix 2*) for subsequent experiments. The Rosa26-tdTomato strain (adults: 6-12 weeks) was used for tdTomato RNP nucleofections and all control nucleofections were carried out using the C57BL/6 wild type strain (adults 8-12 weeks) such that all transplanted samples could be read out by tdTomato negativity. All transplant recipients and competitors were of the Rosa26-tdTomato strain. All recipients were in the range of adult 8-12 weeks with the exception of the large cell volume optimisation experiment (*Results section* ***Figures 3.49, and 3.50***) which was set up just prior to the Covid-19 lab closure. For this experiment, the recipient range was increased to 7-18 weeks due to limited recipient numbers and urgency to set up the experiment before the lab closure. Transplants were carried out according to the Mead project licence P2FF90EE8 protocol 4.

Cells were prepared for transplant 24 hours post-nucleofection. In brief, competitors (Rosa26-tdTomato) and fresh controls (C57BL/6 wild type) were culled and total BM was prepared as described in *2.2.6 i)* but **without c-Kit enrichment** for competitors such that donor samples were c-Kit enriched

whereas competitors were total BM cells. Competitors were prepared in 100µL transplant buffer (PBS (ThermoFisher 10010023) + 1% FCS (Sigma Aldrich F9665)) at 250,000 cells per sample. 140,000 donors per sample were prepared in 150µL transplant buffer and mixed with competitors followed by filtering through a sterile FACS tube (Corning 352235). Samples were kept on ice until the point of transplantation. On the day of transplantation, mice received a lethal dose of irradiation (total 9Gy) split four hours apart. Injections were kindly carried out by Nick Fordham, a DPhil student in the labs of Adam Mead and Irene Roberts and Joana Carrelha, a postdoctoral researcher in the lab of Sten Eirik Jacobsen. Adult Rosa26-tdTomato mice were injected with 200µL by tail vein after gently warming at 30-40°C for 20 minutes. Mice were then treated with Baytril at 1.2mL/250mL water changed once per week for six weeks after transplantation and observed closely for two weeks after irradiation for signs of ill health.

2.3.5 Flow cytometry analysis of long-term lineage and stem cell engraftment of tdTomato edited HSPCs

Tail vein bleeding was kindly carried out by BMS staff in line with the Mead project licence P2FF90EE8 protocol 4 every 4 weeks up to 18 weeks post-transplantation. In brief, blood was collected in EDTA coated tubes and centrifuged for 10 minutes at 90g followed by removal of 6-10µL of the upper phase. The upper phase was then stained with a 2x platelet and red blood cell antibody panel as detailed in **Table 2.8** and incubated for 15 minutes at 4°C. Upper phase samples were then washed with 1mL FACS buffer and centrifuged

at 3,200g for five minutes at 4°C before resuspension in 200µL FACS buffer for flow cytometry.

Antibody	Fluorophore
CD41	BV421
CD150	APC
Ter119	BV650/FITC*

Table 2.8: Platelet and red blood cell lineage panel for assessing engraftment of tdTomato edited HSCs (* for all samples, except three TOM RNP samples at the 18-week readout for the large cell volume optimisation, GFP was used. Due to limited mouse numbers, three TOM RNP recipients (large cell volume optimisation) had an additional GFP transgene at the *Vwf* promoter therefore BV650 was used).

The remaining blood was diluted to 300µL with FACS buffer followed by the addition of 2% dextran (Sigma Aldrich 9004-54-0) and incubation at 37°C for 20 minutes. The upper phase was then removed and centrifuged at 800g for five minutes at 4°C followed by red blood cell lysis in 300 µL ammonium chloride (Sigma Aldrich 254134) for two minutes. Cells were then washed with 1mL FACS buffer, centrifuged at 800g for five minutes followed by resuspension in 25µL Fc block (12.5µL for FMOs). Samples were then prepared for flow cytometry as described in *2.3.11* using the lineage antibody panel detailed in **Table 2.9** and analysed on an LSR II flow cytometer.

Antibody	Fluorophore
Mac-1	APC
CD19	PECY7
CD4 & CD8	APC-ef780
Viability	DAPI

Table 2.9: Lymphoid and myeloid lineage panel for assessing engraftment of tdTomato edited HSCs

At the terminal readout, BM and spleen was also collected under sterile conditions. BM was obtained by crushing in MACS buffer with a pestle and mortar and spleen was pushed through a 70µM filter with FACS buffer. Total BM and spleen was prepared for flow cytometry as described in *2.3.11 and* stained both with the stem cell cocktail detailed in **Table 2.7** and the lineage cocktail detailed in **Table 2.9**. The remaining cells from the spleen and bone marrow were prepared for freezing in 90% FCS (Sigma Aldrich F9665) 10% DMSO (Sigma, D2650) for subsequent sorting for RNA-Seq.

2.3.6 Fluorescence activated cell sorting (FACS) of tdTomato edited HSCs for Target/RNA-Seq 24 hours post-nucleofection and 18 weeks post-transplant

Samples were collected for Target-Seq (Rodriguez-Meira, Buck et al. 2019) 24 hours post-nucleofection due to evidence that tdTomato is still expressed at the protein level at this time. This method was developed by Alba Rodriquez-Meira, during her time as a DPhil student in the lab of Adam Mead. A Target-Seq approach allowed the identification of tdTomato editing at the genomic and transcriptional level to confirm HSCs had undergone editing as expected at this time-point. Samples collected 18 weeks post-transplant could be sorted based on tdTomato negative expression by flow cytometry and therefore did not require a Target-Seq approach. Sample library preparation was carried out in exactly the same way for 24-hour and 18-week samples, with the exception that genotyping PCR was omitted from the 18-week samples.

24-hour samples were removed from the same well as those used for the time course experiments (see *2.4.2*) such that tdTomato knockout at 72 hours could be correlated with genomic expression at 24 hours and therefore successful editing confirmed. Total BM was thawed in IMDM (ThermoFisher, 21056023) with 20% FCS (Sigma Aldrich F9665) 12 hours prior to sorting and recovered in StemSpan with cytokines as described in 2.2.6. Both 24-hour and 18-week samples were prepared for sorting under sterile conditions as described in *2.3.11* with the HSPC cocktail detailed in **Table 2.7**. 18-week samples were stained in double the antibody volume (dilutions remained the same) and resuspended in 2mL for sorting due to increased cell numbers required as they were not c-Kit enriched. 24-hour HSC samples were sorted using the following gating strategy without including c-Kit due to expression being lost in culture post bead selection: Live Lin-Sca-1+CD48-CD150+. 18-week HSC samples were sorted on the following gating strategy: Live tdTomato-Lin-Sca-1+c-Kit+CD48-CD150+. 100-220 HSCs were sorted per replicate for 24-hour samples and 50-100 HSCs were sorted per replicate for 18-week samples. Some 18-week samples were sorted as technical replicates within biological replicates to achieve target cell numbers. Cells were sorted into lysis buffer detailed in **Table 2.10**, immediately vortexed and snap frozen on dry ice followed by transfer to -80°C prior to library preparation. Sorting was carried out with the help of Paul Sopp (Head of the WIMM Flow Cytometry facility), Sally-Ann Clark, Craig Waugh and Kevin Clark (Flow Cytometry facility specialists).

Reagent	Volume	Final concentration
Triton X-100 0.4 %	1.9 µL	0.18%
RNase Inhibitor (40 U/µL)	0.1 µL	0.95 U/µL
dNTPs (10 mM)	1 µL	2.4 mM
OligodT_{30} (10 µM)	1 µL	2.4 µM
Qiagen Protease (1.07 AU/mL)	0.1 µL	2.7E-05 AU/µL
TOTAL	4.1 µL	

Table 2.10: Components of lysis buffer for Target-Seq

2.3.7 Target-Seq library preparation

Library preparation was learnt with the kind help of Haseeb Rahman, a Research Assistant in the lab of Adam Mead and Merve Aksöz, a DPhil student in the lab of Claus Nerlov. Prior to sample thawing, all reagents were prepared for reverse transcription according to **Table 2.11** with the exception of the TSO and enzyme which were added immediately prior to adding to the samples.

Reagent	**Volume per sample (µL)**	**Final concentration**
5X Buffer	2 µL	1 X
RNase Inhibitor (40 U/µL)	0.25 µL	1 U/µL
LNA-TSO (100 µM)	0.1 µL	1 µM
SMARTScribe (100 U/µL)	1 µL	10.20 U/µL
RT-PCR grade water	2.25 µL	NA
TOTAL	5.6 µL	
TOTAL volume (cumulative)	9.7 µL	

Table 2.11: Components of Target-Seq reverse transcription master mix

Samples were collected from -80°C on dry ice and the protease was heat inactivated according to the program detailed in **Table 2.12.**

Temperature	Time (min:sec)	Cycles
72°C	15:00	1
4°C	HOLD	

Table 2.12: Target-Seq protease heat inactivation

Reverse transcription was then performed according to the program detailed in **Table 2.13.**

Temperature	Time (min:sec)	Cycles
42 C	90:00	1
50 C	02:00	10
42 C	02:00	
72 C	15:00	1
4 C	HOLD	

Table 2.13: Target-Seq reverse transcription PCR program

PCR master mix was then prepared as detailed in **Table 2.14** and PCR amplification was performed using the program detailed in **Table 2.15.**

Reagent	Volume per sample (µL)	Final concentration
2X Buffer	12.5 µL	1 X
ISPCR primer (10 µM)	0.125 µL	50 nM
SeqAMP (100 U/µL)	0.5 µL	2 U/µL
tdTomato gDNA primers (F+R, 200 µM stock; HPLC purified)	0.2 µL	808 nM
RT-PCR Grade Water	1.675 µL	NA
TOTAL	15 µL	
TOTAL volume (cumulative)	24.7 µL	

Table 2.14: Components of Target-Seq PCR master mix

Temperature	Time (min:sec)	Cycles
98 C	03:00	1
98 C	00:20	22 cycles
67 C	00:15	
72 C	06:00	
72 C	05:00	1
4 C	HOLD	

Table 2.15: Target-Seq PCR amplification program

Samples were then purified using Ampure XP beads (Beckman Coulter, A63881) at a 0.6:1 beads to cDNA ratio. Finally, samples were indexed using the Nextera XT library kit (Illumina, FC-131-1096) and sent for sequencing at Novogene using the NovaSeq 6000 PE150 sequencing platform. The Nextera XT library preparation protocol including details of critical steps is described in *Appendix 3.*

2.3.8 tdTomato genotyping

24 hours post-nucleofection samples were subject to genotyping at the tdTomato locus to confirm editing. Pre-purified samples (post PCR amplification), containing the cDNA and genomic amplicon mix were used as the input material. Control DNA samples purified from the BM of the Rosa26-tdTomato strain were initially used to carry out a PCR gradient and determine the optimal annealing temperature for the tdTomato primers. Reagents were prepared according to **Table 2.16.**

Reagent	Volume per sample (µL)
DreamTaq PCR master mix	12.5
tdTomato fwd primer (10µM)	4
tdTomato rvs primer (10µM)	4
cDNA + genomic amplicon template	1
Nuclease free water	3.5
Total	25

Table 2.16: Components of tdTomato genotyping PCR

Details of the PCR program used with empirically determined optimal annealing temperature are shown in **Table 2.17**. Samples were gel purified and sent for in-house Sanger sequencing.

Step	**Temperature (°C)**	**Time**	
Activation	95	3 mins	
Denature	95	30 seconds	Cycle to step 2 34 more times
Anneal	68.7	30 seconds	
Extension	72	1 min	
Final extension	72	10 mins	
Hold	4	∞	

Table 2.17: tdTomato genotyping PCR program

2.3.9 Bioinformatic analysis of 24-hour and 18-week tdTomato edited HSC transcriptomic data

Bioinformatics analysis of tdTomato edited HSC samples was kindly performed by Wei Wen, a DPhil student in the lab of Adam Mead. The methodology and pipelines he used are described below.

i) Quality control and batch correction

Adaptors and nucleotide bases with Phred score <30 from the 3' end were trimmed from the raw sequencing reads using TrimGalore 0.5.0 (Martin 2011). Trimmed sequence reads were aligned to the GRCm38 mouse reference genome using STAR 2.6.1d in 2-pass mode (Dobin, Davis et al. 2013). The Binary Map Alignment (BAM) files for technical replicates belonging to the same sample were merged using Samtools 1.9 (Li, Handsaker et al. 2009). All BAM files were subsequently sorted by read names using Samtools. Gene expression counts for each sample were computed from the read name-sorted BAM files

using featureCounts 2.0.0 under the sub-read module (Liao, Smyth et al. 2014). Each gene count was subsequently normalised (divided) by the length of the gene and then further normalised (divided) by the total gene counts (library size) of the corresponding sample and finally multiplied by 1,000,000 to obtain gene expression values in transcript per million (TPM). Unsupervised clustering revealed samples to cluster by cell sort batch. Hence, the gene expression counts were adjusted by cell sort batch month using the recently developed ComBat-seq under the sva 3.36.0 R package (Zhang, Parmigiani et al. 2020) (not to be confused with the ComBat software that is widely used for batch-correction on microarray and single-cell RNA-seq (Johnson, Li et al. 2006). This returned batch-corrected gene expression counts for downstream analysis.

ii) Gene expression analysis by differential gene expression, gene set enrichment analysis and random forest

Differential gene expression analysis between pairs of sample groups was performed using DESeq2 1.30.1. Differentially expressed genes were demonstrated on volcano plots. Having received the differentiation gene expression data from Wei Wen, I ranked the significant genes by log2 (fold change) and adjusted p values followed by combining all TOM RNP comparisons to generate Venn diagrams of overlapping gene expression changes. Genes were ranked by false discovery rate (FDR) and subsequently used as input for gene set enrichment analysis (GSEA) using fgsea 1.16.0. Normalised enrichment scores across the different groups for significant gene sets were demonstrated on heatmaps using pheatmap 1.0.12. Ranked genes were analysed for any

significantly enriched pathways in the MSigDB Hallmark Gene Set consisting of 50 curated gene sets representing well-defined biological states and processes displaying coordinate expression (Liberzon, Birger et al. 2015). For HSC signature scores, I compiled a list of mouse HSC, LT-HSC and ST-HSC genes from the study of Ivanova and colleagues (Ivanova, Dimos et al. 2002) which were used as gene sets by Wei Wen to compute values for each group. The full list of genes contributing to each HSC set is available in *Appendix 8*. Finally, random forest was performed on selected sample groups to identify additional genes not reported from differential gene expression analysis that may distinguish the selected sample groups using randomForest 4.6-14. Batch-correction was performed on R version 4.0.1 whereas downstream analysis was performed on R version 4.0.5.

iii) Sample-specific tdTomato sequence content

A reference sequence consisting of the 1,431-bp tdTomato sequence was created and indexed. The raw sequencing reads for each sample were aligned to the tdTomato sequence using Burrows-Wheeler Alignment (BWA) 0.7.17 (Li and Durbin 2009). For each sample, the number of reads mapping to the tdTomato sequence was computed using Samtools and compared to the total number of raw sequencing reads to obtain the percentage tdTomato sequence content.

2.4 Introduction of the JMML associated *NF-1* loss of function mutation in human cell lines and primary haematopoietic cells

2.4.1 *NF-1* guide RNA design

In order to determine the spectrum of *NF-1* mutations in JMML patients, and therefore inform guide RNA design, data were corroborated from two previous in-depth studies of JMML genetic lesions and filtered for cases of *NF-1* initiating mutations only (Stieglitz, Taylor-Weiner et al. 2015) (**supplementary Tables 2 and 3**) (Caye, Strullu et al. 2015) (**supplementary Table 6**). Frequency of mutation positions and frequency/nature of co-operating mutations were also analysed.

Having deduced the level of heterogeneity in the position of *NF-1* initiating mutations within the gene body, it was decided to mimic the protein consequence (loss of function) by targeting an early exon to maximise the chances of premature stop codons early in the coding sequence. GRNAs were designed using the Synthego gRNA design tool (SynthegoGuideDesign 2021) and the three top-ranked gRNA sequences based on low-off target and high on-target activity were selected. Guides were purchased from Synthego in the synthetically modified format with 2′-O-methyl-3′-phosphorothioate incorporated at the first three bases at the 3' and 5' ends (Hendel, Bak et al. 2015). Primer sequences were designed using the Primer 3 online tool (Primer3 2021) such that amplicons were 800-900bp in length with the guide RNA target region in the centre. **Table 2.18** summarises the guide RNA and primer sequences.

Guide RNA	Sequence (5'-3')
1- *NF-1* exon 2 guide RNA 1	GTTGTGCTCAGTACTGACTT
2- *NF-1* exon 2 guide RNA 2	AGTCAGTACTGAGCACAACA
3- *NF-1* exon 2 guide RNA 3	AAGTTTTCTTTGGTTATAAG
4- NF-1 exon 2 forward primer	TCCAGAAAGAAAGCTGCAAAA
5- NF-1 exon 2 reverse primer	GGCTGGGACAACCGTAAGTA

Table 2.18: *NF-1* exon 2 targeting guide RNA sequences

2.4.2 TF-1 cell line selection

In order to confirm the TF-1 cell line as a suitable system for testing the effect of *NF-1* knockout on GM-CSF sensitivity. The cell line was maintained in RPMI 1640 (Thermofisher, 11875093) supplemented with 10% FCS (Sigma Aldrich F9665), 1% Penicillin/Streptomycin (Thermofisher 15140122), 1% Sodium-Pyruvate (Thermofisher 11360070) and 5ng/mL human GM-CSF (Peprotech, 300-03). For GM-CSF sensitivity testing, a low passage number (less than 10) aliquot was thawed, washed, and recovered overnight in RPMI with supplements as described above. Samples were then washed with PBS (ThermoFisher 10010023) and transferred to RPMI with all supplements including 5ng/mL GM-CSF or all supplements with 0ng/mL GM-CSF at a starting cell density of $1x10^5$/mL. Every two to three days, cells were counted 1:2 diluted in trypan blue (ThermoFisher 15250061) and diluted back down to $1x10^5$ cells/mL. Growth rate was calculated using the split ratio and starting concentration, to determine the cumulative cell density without splitting followed by log transformation.

2.4.3 TF-1 cell line nucleofection optimisation

Unlike the Ba/F3 cell line, an optimised protocol was not commercially available for the TF-1 cell line at the time of experimentation. Therefore a matrix

optimisation method was designed to test a range of buffer and nucleofection programs. The design and experiments were carried out with the kind guidance of David Cruz Hernandez, a DPhil student in the lab of Paresh Vyas.

TF-1 cells were maintained in RPMI 1640 with supplements as described above and split every two-three days to maintain a cell density of $5x10^4$ - $7.5x10^5$ cells/mL. At least 30 minutes prior to nucleofection, 24-well tissue culture plates (Corning, CLS3527) were prepared with RPMI with supplements and placed at 37°C 5% CO_2. 250,000 cells were aliquoted per sample and washed with PBS, followed by resuspension in 20µL of either buffer SF or SG (Lonza SG Cell Line Transfection kit, Lonza V4XC-3032). Buffer SF was a kind gift from David Cruz Hernandez, prepared in-house. Details of the buffer solution are available in *Appendix 4.* Samples were then mixed with 0.4µg pmaxGFP (Lonza SG Cell Line Transfection kit, Lonza V4XC-3032) and placed in separate wells of a 16-well cuvette (Lonza SG Cell Line Transfection kit, Lonza V4XC-3032). Voltages were applied according to the matrix optimisation table detailed in *Appendix 5.* Initial optimisation was carried out using the voltages in the centre column, followed by the appropriate horizontal column. Transfection efficiency was measured 24 hours post-nucleofection by flow cytometry. For flow cytometry preparation and analysis details refer to *2.2.11*.

2.4.4 *NF-1* guide RNA selection in the TF-1 cell line

Using the nucleofection conditions optimised in *2.4.3*, *NF-1* exon 2 gRNAs 1-3 were tested in the TF-1 cell line for their editing efficiency and effect on GM-CSF sensitivity. In brief, 250,000 cells per sample were nucleofected with 1.6µg guide

RNA (Synthego) pre-complexed with 3µg Cas9 (Life Technologies A27865) in SF buffer using program DN100. Bulk samples were transferred to RPMI with supplements containing 5ng/mL GM-CSF or 0ng/mL GM-CSF immediately post-nucleofection and cultured for 15 days followed by genotyping of the *NF-1* locus. Genotyping was carried according to the reagent components listed in **Table 2.19** and the PCR program detailed in **Table 2.20**. Gradient PCR was carried out initially on wild type TF-1 cells in order to determine the optimal primer annealing temperature.

Step	**Temperature (°C)**	**Time**	
Activation	95	3 mins	
Denature	95	30 seconds	Cycle to step 2 34 more times
Anneal	58.6	30 seconds	
Extension	72	1 min	
Final extension	72	10 mins	
Hold	4	∞	

Table 2.19: *NF-1* genotyping PCR components

Reagent	**Volume per sample (µL)**
DreamTaq Green PCR master mix	12.5
Forward primer	4
Reverse primer	4
DNA template (25ng/uL)	4
Nuclease free water	0.5

Table 2.20 *NF-1* genotyping PCR program

2.4.5 TF-1 low cell number sorting protocol optimisation

In order to determine the minimal cell requirement for colony output and optimal sorting conditions for TF-1 cells, viable cells were sorted at 1-500 cells per well into either 96-well flat bottomed (Corning, 3596), 96-well round bottomed

(Corning, 3799) or 60-well Terasaki (Greiner Bio-One, 653190) plates containing 50µL (96-well) or 20µL (Tersaki) of RPMI with supplements or conditioned medium. Conditioned medium was prepared by culturing the TF-1 cell line in RPMI with supplements followed by filtration of the medium to remove the cells, and re-addition of appropriate amounts of FCS, Penicillin/Streptomycin and GM-CSF. Sorted cells were scored for colony growth 22 days post-sorting according to the parameters detailed in **Table 2.21**. The optimal plate and culture conditions were determined as those that produced the most colonies at the lowest cell numbers sorted.

Score	**Cell number/ well coverage**
0	No cells
1	1 cell
2	2 cells
3	3-9 cells
4	10-15 cells
5	16-20 cells
6	21-30 cells
7	31-49 cells
8	50 cells- 10% coverage
9	11-49% coverage
10	50-99% coverage
11	100% coverage

Table 2.21: Terasaki and 96-well sorting colony counting key

2.4.6 Assessment of *NF-1* edited TF-1 cell sensitivity to GM-CSF

In order to test whether *NF-1* editing renders TF-1 cells GM-CSF independent, 250,000 TF-1 cells per sample were nucleofected with the two optimal *NF-1* guides or the intronic *AAVS1* control guide using the method described in *2.4.4*. 1 or 5 cells were sorted into separate wells of 60-well Terasaki plates containing 20µL unconditioned RPMI with supplements as described above with 5ng/mL or

0ng/mL GM-CSF. Colonies were scored 16-21 days post sorting according to the key detailed in **Table 2.21**. Successful colonies from the 5ng/mL group were then grown up as clonal cell lines by dilution first into 96-well plates and further re-plating up to T25 flasks. At the T25 stage, DNA was extracted and *NF-1* genotyping performed according to **Tables 2.19** and **2.20**. Successful clones from the 5ng/mL GM-CSF group were then challenged in RPMI with supplements and 0ng/mL GM-CSF for a period of 28 days according to the method described in *2.4.2*.

2.4.7 TF-1 cell line NF-1 Western blot

For all clonal cell lines used for GM-CSF sensitivity testing as described in *2.4.6*, western blotting was carried out to confirm NF-1 loss at the protein level. Western blotting was carried out with the kind guidance of Ryan Beveridge, Head of the WIMM Virus Screening facility. $5x10^6$ cells per sample were washed with PBS followed by lysis in 1.5% NP40 (Thermofisher, 85124) dissolved in BC300 buffer (20nM Tris-HCl, pH 8.0, 300mM KCl, 5mM EDTA, 20% glycerol) and 1x protease inhibitors (Calbiochem, 539134). Supernatants were stored at -20°C prior to western blotting. Protein concentration was determined using the Bradford Assay Kit (BioRad, 5000201) following manufacturer's instructions using a BSA standard curve and 20µg of protein was loaded per sample. In brief, 20µg samples were mixed with reducing agent, sample loading buffer and protease inhibitors (all final concentrations 1x) and diluted to 25µL with RIPA buffer (Sigma-Aldrich, R0278). Samples were run on a NuPAGE 3-8% Tris-Acetate gel (Thermofisher EA0375) with a 10-250kD Precision Protein plus Kaleidoscope

ladder (Bio-Rad, 1610375). After transfer, sample membranes were cut above the 30kDa marker and were stained separately with GAPDH (Bethyl, A300-641A), beta tubulin (Sigma-Aldrich, T4026) and NF-1 (Abcam, ab17963). Samples were then probed with the appropriate secondary antibodies conjugated to horseradish peroxidase (goat anti-rabbit for GAPH and NF-1: Sigma-Aldrich, A0545; goat anti-mouse for beta-tubulin: Sigma-Aldrich, A6667) followed by development using ECL prime (Cytiva, RPN2232).

2.4.8 Cord blood CD34+ cell selection

Prior to nucleofection and sorting, total CB MNCs were CD34+ selected to enrich for HSPCs. Post-thawing, cells were filtered in FCS through a 50µM filter followed by centrifugation at 300g for 10 minutes at room temperature. Cells were then washed with human FACS buffer (IMDM, 10%FCS, 10µg/mL DNAse) followed by centrifugation at 300g for 10 minutes at room temperature. Cells were then counted 1:2 in trypan blue and blocked with 100µL FCR blocking reagent (Miltenyi, 130-059-901) per $1x10^8$ cells followed by addition of 100µL human CD34 beads per $1x10^8$ cells (Miltenyi, 130-046-702) and incubation for 30 minutes at 4°C with rocking. Samples were then washed with human FACS buffer, centrifuged for 10 minutes at 300g and then applied to a MACS column equilibrated with human FACS buffer. The column was washed 3x with 500µL human FACS buffer followed by removal from the magnet and expulsion of the CD34+ cells.

2.4.9 Minimal cytokine optimisation for CB CD34+Lin- single cell colony growth

In order to determine the minimum cytokine cocktail that maintained low level colony production but significantly reduced it compared to using a full cocktail, CD34+ selected CB samples were recovered and either sorted directly or 'mock' nucleofected with water then sorted into varying cocktails of cytokines. In brief, CB MNCs were CD34 selected as described in *2.4.8* and resuspended in recovery medium (StemSpan, 25ng/mL hSCF, 50ng/mL hTPO, 50ng/mL FLT3-Ligand) at $2.5x10^4$ cells/mL followed by 24-hour incubation at 37°C 5% CO_2. 'Mock' nucleofection samples (200,000-250,000) were then washed with human FACS buffer followed by centrifugation at 300g for 10 minutes at room temperature and resuspension in 20µL Primary cell buffer P3 followed by mixing with 2µL nuclease free water. Nucleofection was carried out using the previously optimised commercially available program for this cell type E0-100 on a Lonza 4D nucleofector. Samples were then transferred to T25 flasks containing recovery medium and maintained at $2.5x10^4$ cells/mL until sort preparation. 24 hours after nucleofection, samples were stained for FACS using the panel detailed in **Table 2.22**.

Antibody	**Fluorophore**
Viability	DAPI
CD34	APC-ef780
Lineage: CD8, CD20, CD2, CD3, CD16, CD19, CD235a, CD66b, CD10, CD127	FITC

Table 2.22: FACS panel for sorting CD34+ Lin- CB single cells

Single, viable, CD34+Lin- CB cells were sorted into separate wells of 60-well Terasaki plates containing 20µL of one of three cytokine cocktails as detailed in **Table 2.23.** Colonies were scored for size 14 days post-sorting using the key detailed in **Table 2.21**.

Terasaki sort medium	Cytokine cocktail
1	EPO (25ng/mL), SCF (10ng/mL), IL-3 (20ng/mL), IL-6 (10ng/mL), G-CSF (50ng/mL), GM-CSF (50ng/mL)
2	SCF (10ng/mL), GM-CSF (10ng/mL)
3	SCF (10ng/mL), GM-CSF (1ng/mL)

Table 2.23: Cytokine cocktails used for CD34+ Lin- CB single cell sorting

2.4.10 *NF-1* targeting in CD34+Lin- CB and *in vitro* growth assessment

CB MNCs were CD34+ selected according to the protocol described in *2.4.8* and recovered overnight as described in *2.4.9*. 24 hours post-recovery, samples were nucleofected either with the control gRNA *AAVS1*, *NF-1* guide RNA 1 or 2 (1,000,000 CD34+s per condition). Samples were resuspended in 100µL Primary Cell P3 buffer post washing in human FACS buffer and centrifugation at 300g for 10 minutes at room temperature followed by mixing with pre-complexed RNPs (6.4µg gRNA, 12µg Cas9 incubated at 21°C for 10 minutes). Samples were then returned to pre-warmed recovery medium at $2.5x10^4$ cells/mL for 24 hours prior to sorting according to the antibody panel detailed in **Table 2.22** into Terasaki plates containing the cytokine cocktails detailed in **Table 2.24**. Colonies were scored for size 14 days post-sorting using the key detailed in **Table 2.21**.

Terasaki sort medium	Cytokine cocktail
1	EPO (25ng/mL), SCF (10ng/mL), IL-3 (20ng/mL), IL-6 (10ng/mL), G-CSF (50ng/mL), GM-CSF (50ng/mL)
2	SCF (10ng/mL), GM-CSF (1ng/mL)
3	SCF (10ng/mL), GM-CSF (0.1ng/mL)

Table 2.24: Cytokine cocktail for sorting *NF-1* and *AAVS1* targeted CD34+Lin- CB

2.4.11 NF-1 colony PCR of CD34+Lin- CB colonies

Large CB colonies (> 50 cells) were picked for genotyping 14-15 days post-sorting and genotyped using a colony PCR protocol previously established by Eleni Louka and Alba Rodriguez Meira, DPhil students in the lab of Adam Mead. In brief, colonies were removed from the Terasaki plate and transferred to a 96-well plate followed by washing of the well with 10µL PBS and adding to the 96-well plate (total 30µL). The plate was then centrifuged at 300g for five minutes at room temperature followed by removal of the supernatant to the 10µL mark. 10µL samples were then snap frozen on dry ice and stored at -80°C until processing. Colonies were thawed and lysed in 15µL of lysis buffer containing 10% Proteinase K (Qiagen, 19131) and 0.4% Triton (Sigma Aldrich, X-100) by incubating at 56°C for 10 minutes followed by a 20 minute heat inactivation at 75°C. PCR amplification was carried out using the Roche High Fidelity Kit (Roche, FHIFI-RO) according to the reagent components detailed in **Table 2.25** and PCR program detailed in **Table 2.26**. PCR samples were sequenced using in-house Sanger sequencing.

Component	Volume per reaction (uL)
10X FastStart High Fidelity Reaction buffer without $MgCl_2$ (Roche)	1
25mM $MgCl_2$ (Roche)	1.8
DMSO (Roche)	0.5
10mM PCR Grade Nucleotide Mix (Roche)	0.2
5U/uL FastStart High Fidelity Enzyme Blend (Roche)	0.1
PCR Certified Water (TEKnova)	0
Total MM	3.6
DNA	4
NF-1 exon 2 primers 10uM	1.2uL each fwd/rvs
Total volume	10

Table 2.25: PCR master mix for CB colony genotyping

PCR stages	Number of cycles
95°C 10 min	1
95°C 15 seconds 58.6°C 30 seconds 72°C 1 min	49
72°C 3 min	1
4°C	∞

Table 2.26: PCR program for CB colony genotyping

2.4.12 *NF-1* and *AAVS1* targeting and NSG transplantation of edited CB CD34+ cells

Total CB MNCs were CD34+ selected as described in *2.4.8*. 24 hours post-recovery, 1,000,000 cells per guide RNA group (*AAVS1* and *NF-1* guide 1) were nucleofected as described in *2.4.10*. 24 hours post-nucleofection 47,000-50,000 cells were transplanted into sub-lethally irradiated NSG mice. Transplantation was kindly carried out by Roo Bhasin, an animal technician at the BMS facility. In brief, samples were centrifuged at 300g for 10 minutes at room temperature followed by resuspension in 250µL of PBS + 1% FCS and filtering through a

sterile FACS tube (Corning 352235). An aliquot of each sample was subject to DNA extraction and amplified at the *NF-1* exon 2 locus using the Kapa 2G PCR mix (Sigma-Aldrich, KK5702) according to the reagent components detailed in **Table 2.27** and PCR program detailed in **Table 2.26.**

Component	**Vol per reaction (uL)**
Kappa 2G	7.5
NF-1 exon 2 Primers 10uM	1.2 each fwd and rvs
DNA	5.1uL

Table 2.27: Kapa 2G PCR components for CB genotyping

On the day of transplant, NSG mice were sub-lethally irradiated with two doses of 1.25g, each four hours apart, followed by transplantation two hours after the second irradiation dose, according to the Mead project licence P2FF90EE8 protocol 4. Mice were transplanted under sterile conditions with 200µL by tail vein injection 20 minutes after heating at 30-40°C and treated with Baytril at 1.2mL/250mL sterile water changed once per week for six weeks after transplantation. Mice were weighed at baseline prior to irradiation and thereafter twice a day for two weeks. Mice were monitored regularly for signs of ill health and culled if there were significant signs of weight loss and pale limbs.

2.4.13 Flow cytometric and morphological analysis of NSG transplant recipients

At four to five week intervals post-transplantation, PB of NSG recipients was analysed for human lineage engraftment by flow cytometry. Blood was processed

as described in *2.3.5* and processed for flow cytometry as described in *2.2.11*. **Table 2.28** summarises the antibody panel used.

Antibody	**Fluorophore**
hCD45	AF700
mCD45	PETx Red
CD235a	PerCP Cy5.5
CD34	APC-ef780
Myeloid: CD15, CD33, CD11b, CD14, CD66b	FITC
CD19	PE
Viability	DAPI

Table 2.28: Antibody panel for peripheral blood analysis of NSG CB recipients

At the time of culling, BM, spleen and PB was obtained from the mice and analysed by flow cytometry using the lineage panel detailed in **Table 2.28**. Spleen and BM samples were also analysed for stem and progenitor engraftment using the HSPC panel detailed in **Table 2.29**.

Antibody	**Fluorophore**
CD34	APC-ef780
CD38	PETx Red
CD90	BV421
CD45RA	BV650
CD123	PECY7
Lineage: CD8, CD20, CD2, CD3, CD16, CD19, CD235a, CD66b, CD10, CD127	FITC
hCD45	AF700
mCD45	PE
Viability	7-AAD

Table 2.29: Antibody panel used to stain the human HSPC compartment of NSG CB recipients

At 18 weeks post-transplant, two *AAVS1* controls were culled to compare to the *NF-1* recipients which were culled at 13 and 17 weeks post-transplant when they

became sick. The remaining mice in both groups were culled 26 weeks post-transplant at the point of the experiment end. DNA was also extracted from BM samples and PCR amplified using the Kapa components detailed in **Table 2.27** following the PCR program detailed in **Table 2.26** with a reduced cycle number (34) due to higher amounts of starting material. BM films and cytospins were kindly made and stained by Nikoloas Sousos using the Giemsa-May-Grünwald kit (Thermo Fisher Scientific, 3300) and analysed by microscopy with the kind help of Eleni Louka.

2.5 Statistical analysis

All graphs and statistical analyses were computed in GraphPad Prism v7.05. Kruskal-Wallis, t-tests, Mann-Whitney t-tests and one/two-way ANOVA tests were carried out to compare experimental groups depending on the data type and distribution. Details of statistical tests using for specific experiments are described in the figure legends.

CHAPTER 3: RESULTS

3: RESULTS

3.1 Optimisation of CRISPR in primary mouse HSPCs using *in vitro* methods

3.1.1 Experimental approach

The aim of this chapter was to develop an optimised method for CRISPR targeting the tdTomato locus in transgenic mouse HSCs. The rationale for targeting this fluorescent reporter in the context of mouse HSCs was in order to obtain an editing readout that was traceable and without targeting a gene that itself would alter the properties of mouse HSCs. Thus, by having a fluorescent marker that was different depending on whether the cell was edited or not, it would be possible to track the contribution of edited HSCs to all cell lineages compared to their non-edited counterparts. This would allow for tightly controlled molecular profiling and functional assessment of the genome editing process itself on the properties of HSCs, providing insights potentially important to the clinical application of this approach.

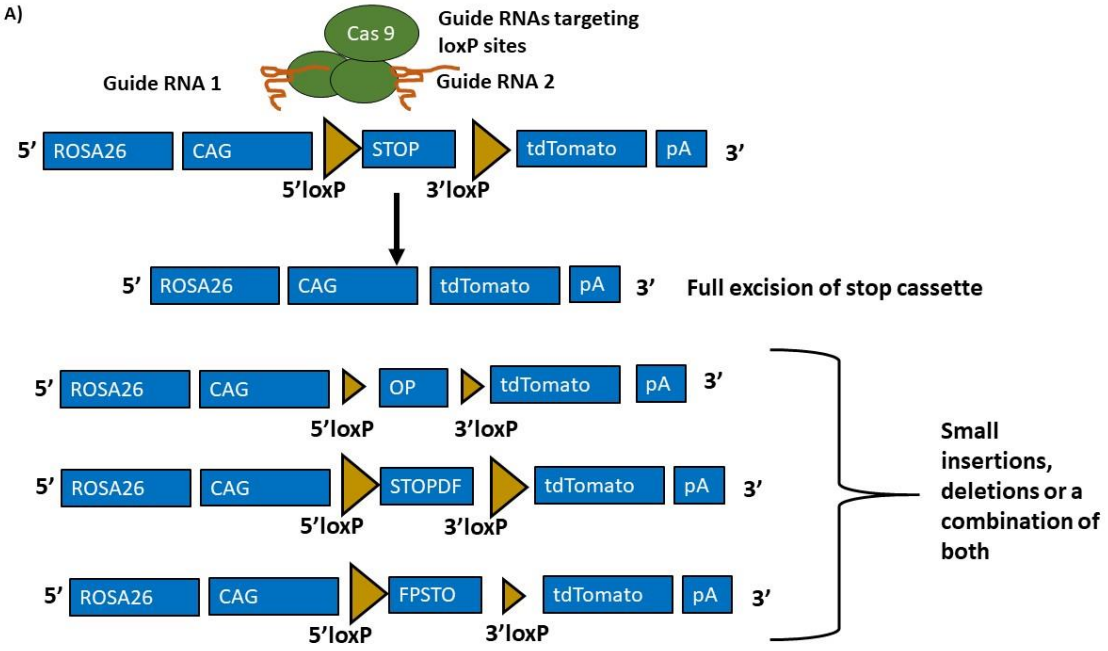
As detailed in Methods section 2.2, *in vitro* optimisation approaches were taken to ascertain the most efficient targeting approach, mode of CRISPR delivery, and format in which to use the CRISPR components in mouse HSCs. In brief, two targeting approaches were designed to either induce constitutive tdTomato expression, using dual guide RNAs targeting the loxP sites in the Ai9-Rosa26 strain; or to knock out tdTomato expression in mouse haematopoietic cells where it is constitutively expressed using a single guide RNA targeting repeat regions of the tdTomato sequence in the Rosa26-tdTomato strain. Guide RNAs were

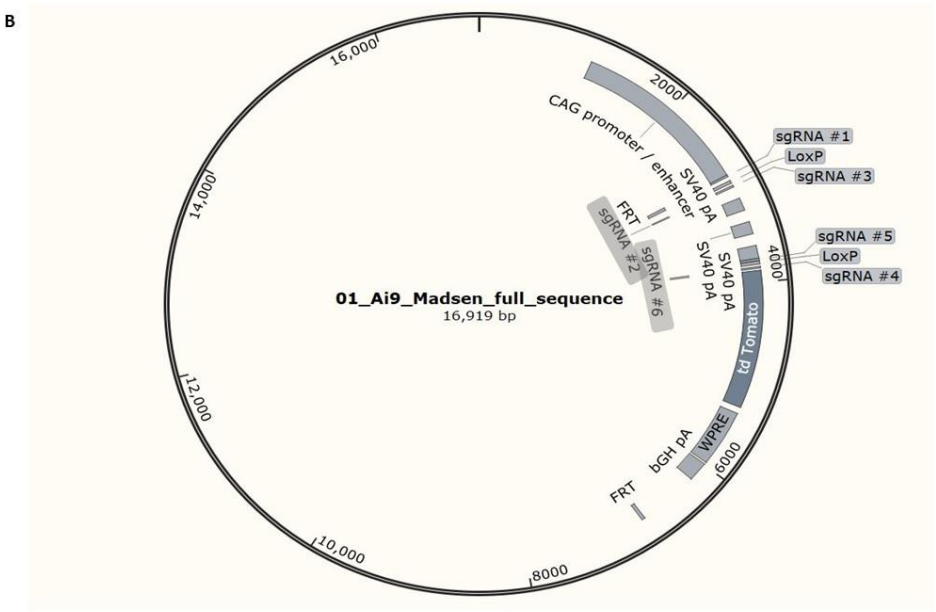
cloned into PX458 CBh-Cas9 and EFS/SFFV-Cas9 vectors for plasmid and viral delivery respectively. Highly concentrated virus particles were generated in HEK293T cells using ultra-centrifugation (by Ryan Beveridge, Virus Screening Facility). For the tdTomato induction approach, guide pairs were initially tested in HEK293T cells to choose the optimal combination. A wide range of electroporation and nucleofection conditions were initially tested in the Ba/F3 pro-B cell line, of which the most promising conditions were then further narrowed down in primary mouse BM HSPCs. For the selected targeting approach, plasmid and RNP delivery using nucleofection was compared with viral transduction of the same guides. Flow cytometry was used to analyse both the transfection/transduction efficiency, for the plasmid and viral delivery, and editing efficiency of each approach.

3.1.2 Guide RNA design

i) Induction of tdTomato reporter expression

Three guide RNAs, with 5'-NGG-3' PAM sequences, were designed targeting the loxP sites flanking the STOP cassette in the Ai9-Rosa26 strain. **Figure 3.1** summarises the guide positions within the locus. Using a dual guide RNA approach, targeting each loxP site should cause insertions/deletions (indels) at each loxP site or induce excision of the STOP cassette, usually removed by introducing Cre-recombinase, and thus induction of tdTomato expression.


A)
Cas 9
Guide RNAs targeting loxP sites
Guide RNA 1
Guide RNA 2
5'
ROSA26
CAG
STOP
tdTomato
pA
3'
5'loxP
3'loxP
Full excision of stop cassette
OP
STOPDF
FPSTO
Small insertions, deletions or a combination of both


B
01_Ai9_Madsen_full_sequence
16,919 bp
CAG promoter / enhancer
SV40 pA
sgRNA #1
LoxP
sgRNA #3
sgRNA #5
sgRNA #4
sgRNA #2
sgRNA #6
FRT
tdTomato
WPRE
bGH pA
2000
4000
6000
8000
10,000
12,000
14,000
16,000

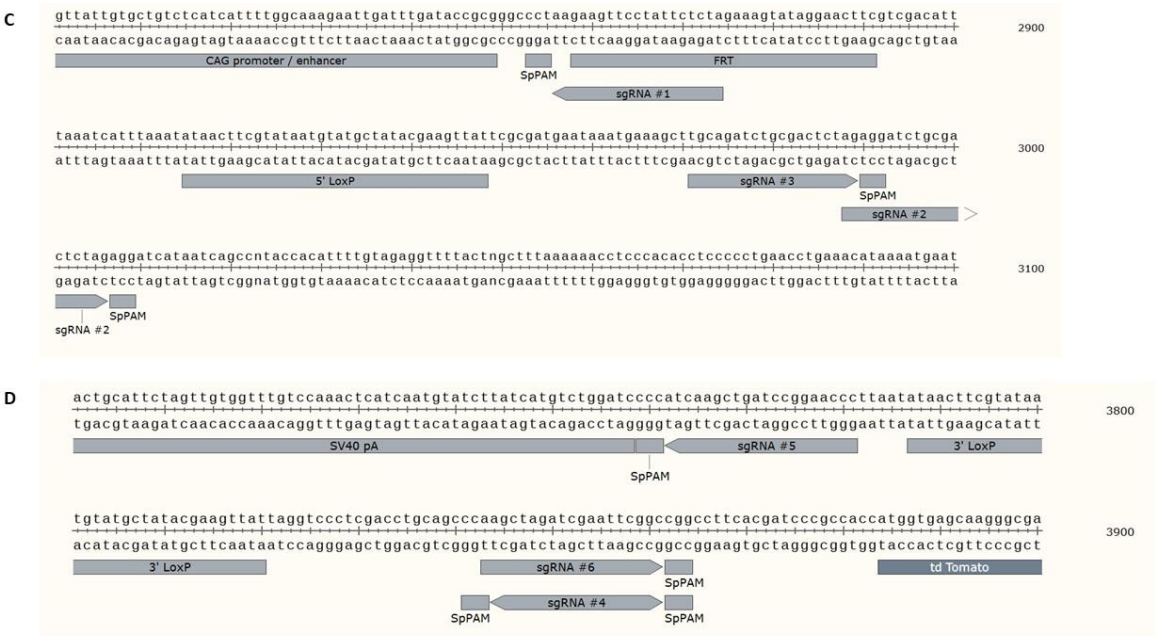
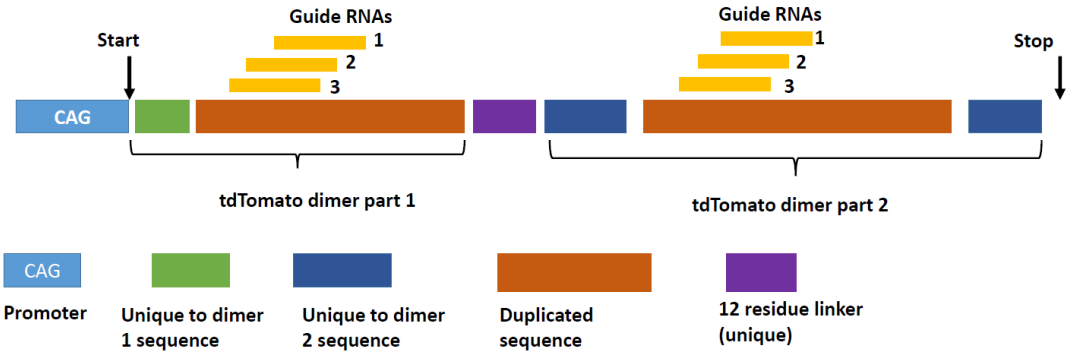


Figure 3.1: Ai9-Rosa26 schematic, clone chart and guide RNA sequences **A)** Schematic of Ai9-Rosa26 loxp-STOP-loxP-tdTomato sequence and potential outcomes of dual loxP guide RNA CRISPR targeting. **B)** Ai9-Rosa26 loxP-STOP-loxP-tdTomato region clone chart. LoxP sites, tdTomato sequences and guide RNAs are labelled in grey. **C)** 5' LoxP site sequence region. **D)** 3' LoxP site sequence region. Arrows indicate the direction of the guide 5'→3', SpPAM indicates the 5'-NGG-3' PAM motif.

ii) Knockout of tdTomato reporter expression

Three guide RNAs were designed using Benchling (Benchling 2019) against the tdTomato sequence close to the transcription start site, which is duplicated in the downstream repeat region. **Figure 3.2** summarises the tdTomato construct and the position of the guides within the target sequence region. Using a single guide approach, targeting the repeated regions should lead to indels at each target site or excision of the intervening region.

A) i)


Guide RNAs
1
2
3
Start
CAG
tdTomato dimer part 1
Guide RNAs
1
2
3
Stop
tdTomato dimer part 2

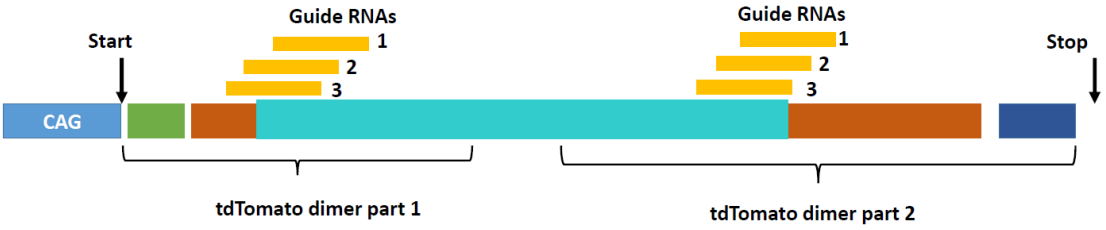
CAG
Promoter
Unique to dimer 1 sequence
Unique to dimer 2 sequence
Duplicated sequence
12 residue linker (unique)

A) ii)


Guide RNAs
1
2
3
Start
CAG
tdTomato dimer part 1
Guide RNAs
1
2
3
Stop
tdTomato dimer part 2

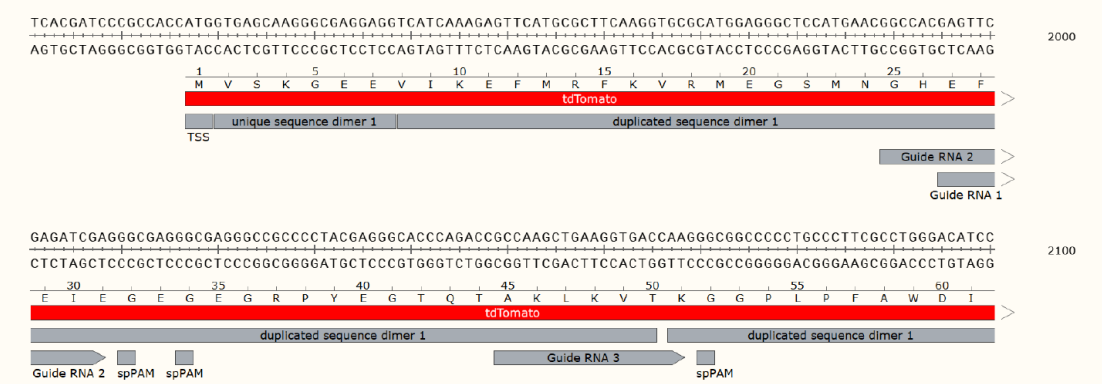

Indels

A) iii)


Guide RNAs
1
2
3
Start
CAG
tdTomato dimer part 1
Guide RNAs
1
2
3
Stop
tdTomato dimer part 2


Deleted region

B) i)



B) ii)



Figure 3.2: tdTomato region schematic, sequence and guide RNA sequences A) Schematic of **i)** tdTomato sequence region and knockout guides, **ii)** tdTomato sequence if guides induce indels at each target site, **iii)** tdTomato sequence if guides induce excision of the intervening region. **B) i)** tdTomato 5' targeted sequence region. **ii)** tdTomato 3' targeted sequence region. Transcriptional start site (TSS), duplicated regions, guide RNAs and SpPAM motifs are labelled in grey. The amino acid sequence is shown below the DNA sequence. Arrows indicate the direction of the guide 5'→3', SpPAM indicates the 5'-NGG-3' PAM motif.

3.1.3 Guide RNA cloning and virus production

i) Single and dual guide RNA insertion into the PX458 CBh-Cas9, EFS-Cas9, and SFFV-Cas9 vectors

All 6 guides for tdTomato induction and the optimal guide (determined by RNP nucleofection) for tdTomato knockout were successfully cloned into the PX458

CBh-Cas9 vector for delivery by plasmid transfection and confirmed by Sanger sequencing. The most efficient guide(s) from transfection were then cloned into EFS/SFFV-Cas9 vectors for comparison of transfection with transduction. High titre virus was kindly generated by Ryan Beveridge in HEK293T cells using ultracentrifugation. **Figure 3.3** shows the clone chart of tdTomato knockout gRNA 3 in the PX458 CBh-Cas9 backbone, and the clone chart of tdTomato induction guides 1 and 5 in the EFS-Cas9 backbone as examples.

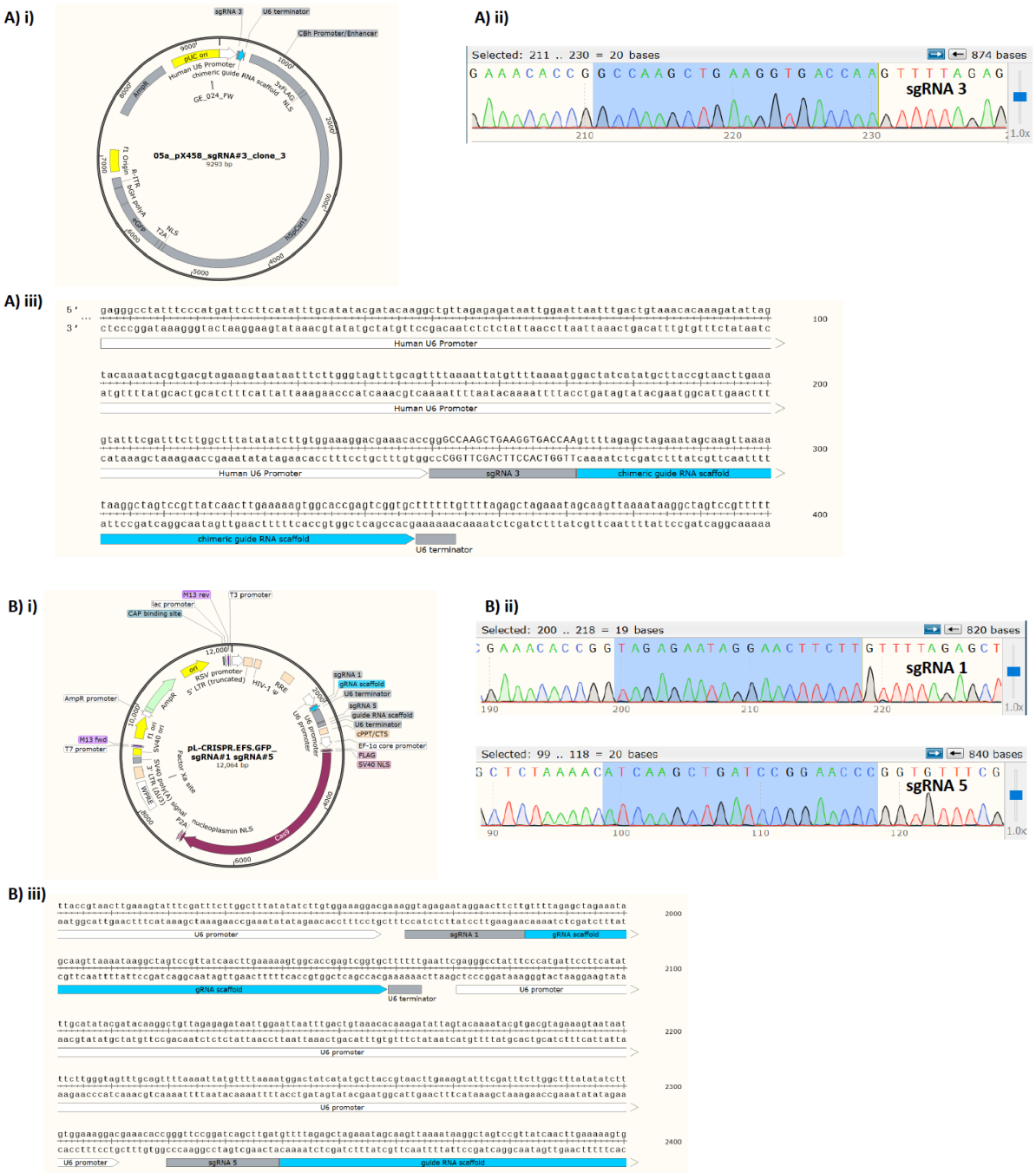
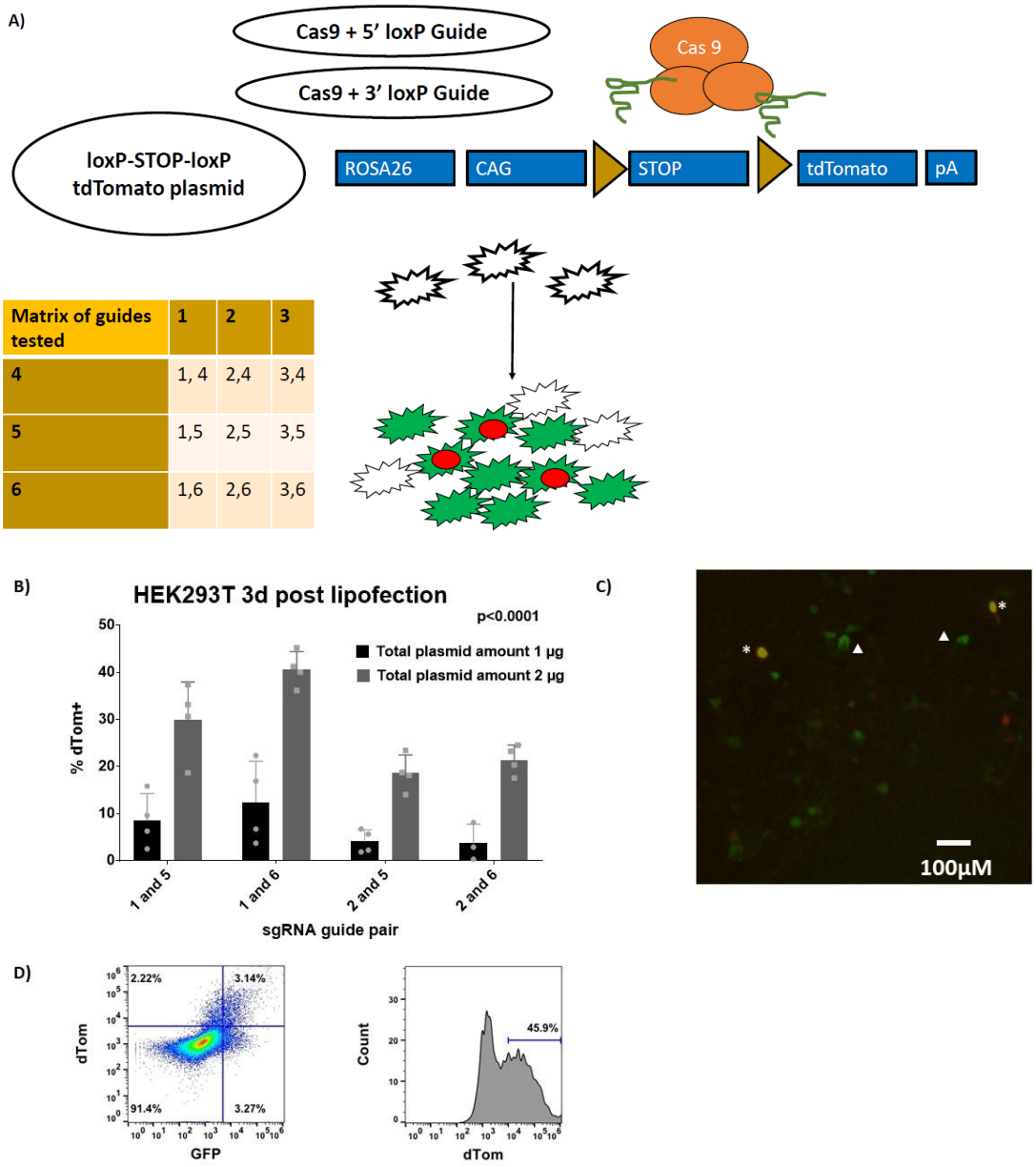


Figure 3.3: Example clone charts and Sanger traces for tdTomato induction and knockdown guide RNAs cloned into PX458 and EFS backbones A) PX458 CBh- Cas9 vector containing the U6 promoter, cloned tdTomato knockout guide RNA 3, gRNA scaffold and U6 terminator **i)** Clone chart, **ii)** Sanger trace showing tdTomato knockout guide RNA (highlighted) insertion within the U6 region **iii)** guide sequence region of tdTomato knockout guide RNA 3. **B)** EFS- Cas9 vector containing the U6 promoter, cloned tdTomato induction guide RNAs 1 and 5, gRNA scaffold and U6 terminator **i)** Clone chart, **ii)** Sanger traces showing tdTomato inductions guide RNA insertion (highlighted) in the U6 region **iii)** guide sequence region of loxP tdTomato induction guides 1 and 5.

3.1.4 Selection of tdTomato induction guide RNAs in the HEK293T cell line

i) Dual plasmid lipofection

Initial experiments were carried out using guide pairs or triplicates in order to select the optimal combination. The two best guide pairs were then used to determine the optimal total plasmid amount by transfecting guides 1 or 2 with guides 5 or 6 in the HEK293T cell line, co-transfected with the Ai9 construct. **Figure 3.4** summarises the work flow and results.



Matrix of guides tested	1	2	3
4	1, 4	2,4	3,4
5	1,5	2,5	3,5
6	1,6	2,6	3,6

Figure 3.4: Guide pair 1 and 6 were optimal for tdTomato induction in HEK293T cells A) Schematic of HEK293T lipofection with Ai9 and loxP targeting guides. **B)** tdTomato expression of HEK293T cells 3 days post lipofection. N=4 (2 biological replicates with 2 technical repeats). Data represent mean with standard deviation, significance measured using two-way ANOVA. **C)** Guide pair 1 and 6: Overlay of microscope images taken 2d post lipofection using GFP and tdTomato filters, 1cm=100uM, triangles label GFP-positive transfected cells, stars represent GFP/tdTomato double positive transfected and edited cells. **D)** Guide pair 1 and 6: Flow cytometry of GFP and tdTomato expression pre-gated on live singlets (left plot), percentage tdTomato positive pre-gated on GFP +ve (right plot).

3.1.5 Optimisation of electroporation and nucleofection conditions in the Ba/F3 cell line

In order to narrow down the set of conditions to be tested in primary mouse HSPCs, as cell numbers are more limited than with cell lines, the mouse haematopoietic pro-B Ba/F3 cell line was used. A range of voltage settings and plasmid amounts were tested for the Neon electroporation system, which were then compared to a range of plasmid amounts for Amaxa nucleofection. The Neon electroporation system was initially used to test a wide range of plasmid amounts before comparing to the Amaxa nucleofection system as a published protocol was available for application of this system in mouse HSPCs (Gundry, Brunetti et al. 2016) and therefore it was referred to as a starting point for optimisation experiments. Commercially optimised Amaxa nucleofection voltage conditions were already available for this cell line so these were not re-optimised (http://www.lonzabio.com 2017). **Figures 3.5** summarises the workflow.

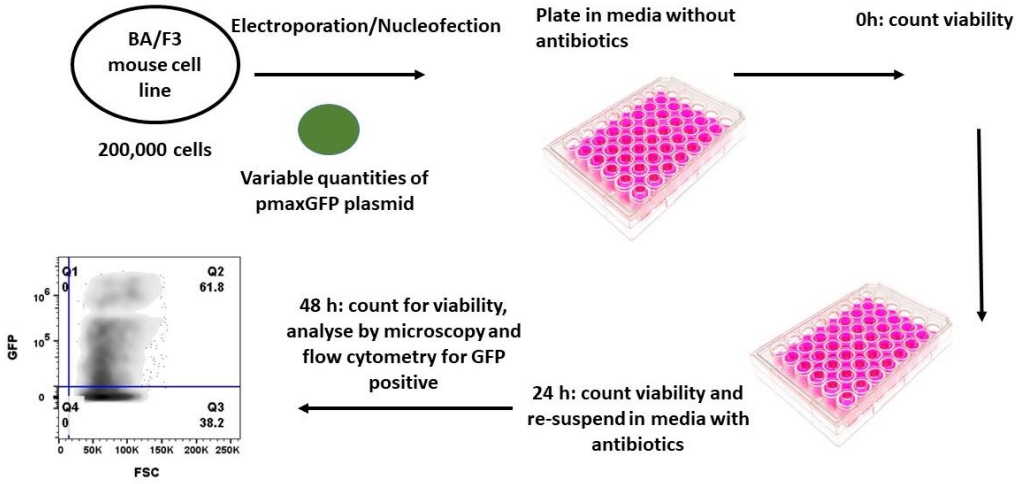


Figure 3.5: Optimisation of Ba/F3 transfection conditions experimental workflow Schematic of workflow for comparison of electroporation and nucleofection in Ba/F3 cells. For all transfections, the 4kb pmaxGFP plasmid was used to assess efficiency by microscopy and flow cytometry.

By combining data for viability (immediately, 24h and 48h post-electroporation) with transfection efficiency (GFP expression) (**Figure 3.6**) it was decided that 1400V 30ms 1p was the optimal condition that achieved high transfection without compromising viability. **Figure 3.7** shows the microscope image at 24h post-electroporation and flow cytometry plot 48h post-electroporation for the optimal condition: 1400V 30ms 1p.

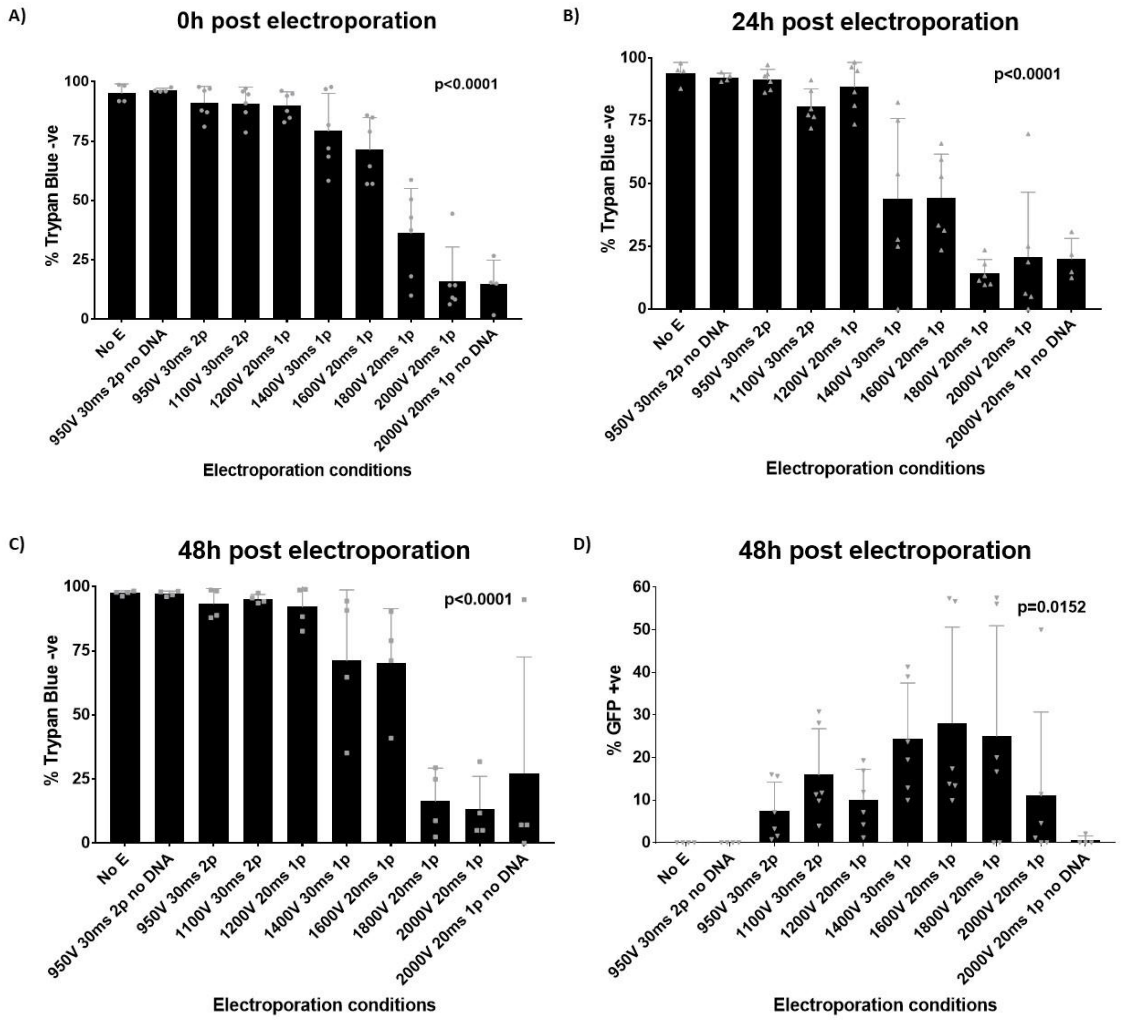


Figure 3.6: 1400V 30ms 1p was the optimal electroporation condition for Ba/F3 cells Viability **A)** 0h **B)** 24h **C)** 48h post electroporation. **D)** Percentage GFP-positive by flow cytometry 48h post electroporation, pre-gated on live singlets. Electroporation (E) condition given as voltage (V), milliseconds (ms) pulse (p). N=6 (3 biological replicates with 2 technical replicates) except for no E and no DNA controls: N=4. Data represent mean with standard deviation. Significance calculated using one-way ANOVA.

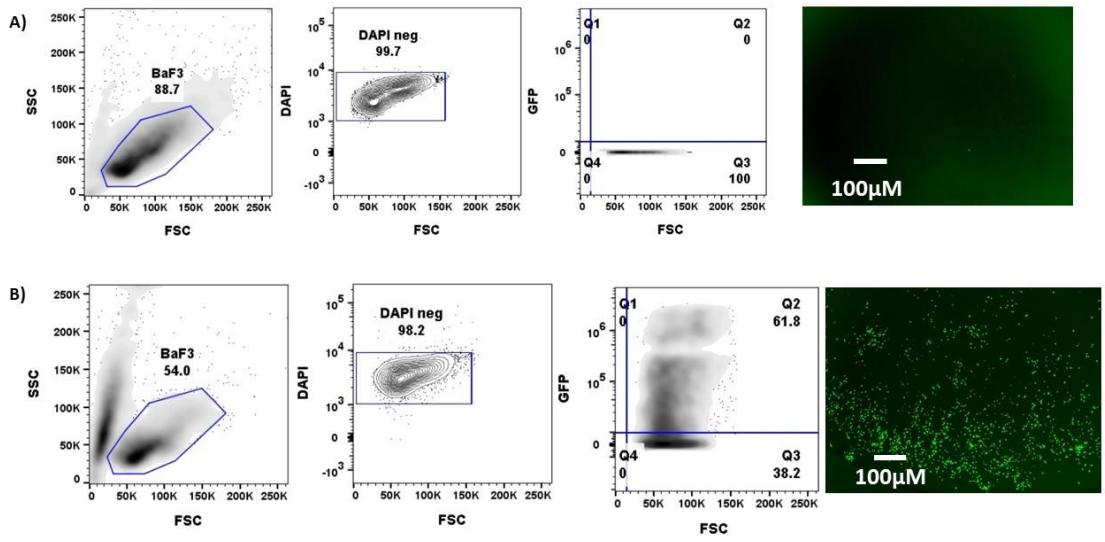


Figure 3.7 Microscopy and flow cytometry of the optimal electroporation condition: 1400V 30ms 1p Example microscopy image 24h post transfection (1cm=100µM) and flow cytometry plot 48h post transfection for **A)** No plasmid + electroporation control **B)** the optimal electroporation condition: 1400V 30ms 1p, 2.8µg pmaxGFP plasmid.

Having established the optimal electroporation condition, this program was then used to test a range of plasmid amounts and determine the balance between viability and transfection efficiency. As shown in **Figure 3.8**, at certain high plasmid amounts, transfection efficiency reduced presumably because viability became a compromising factor. 2.8µg was the plasmid amount at which optimal transfection efficiency (**Figure 3.8 D**) was achieved without compromising viability.

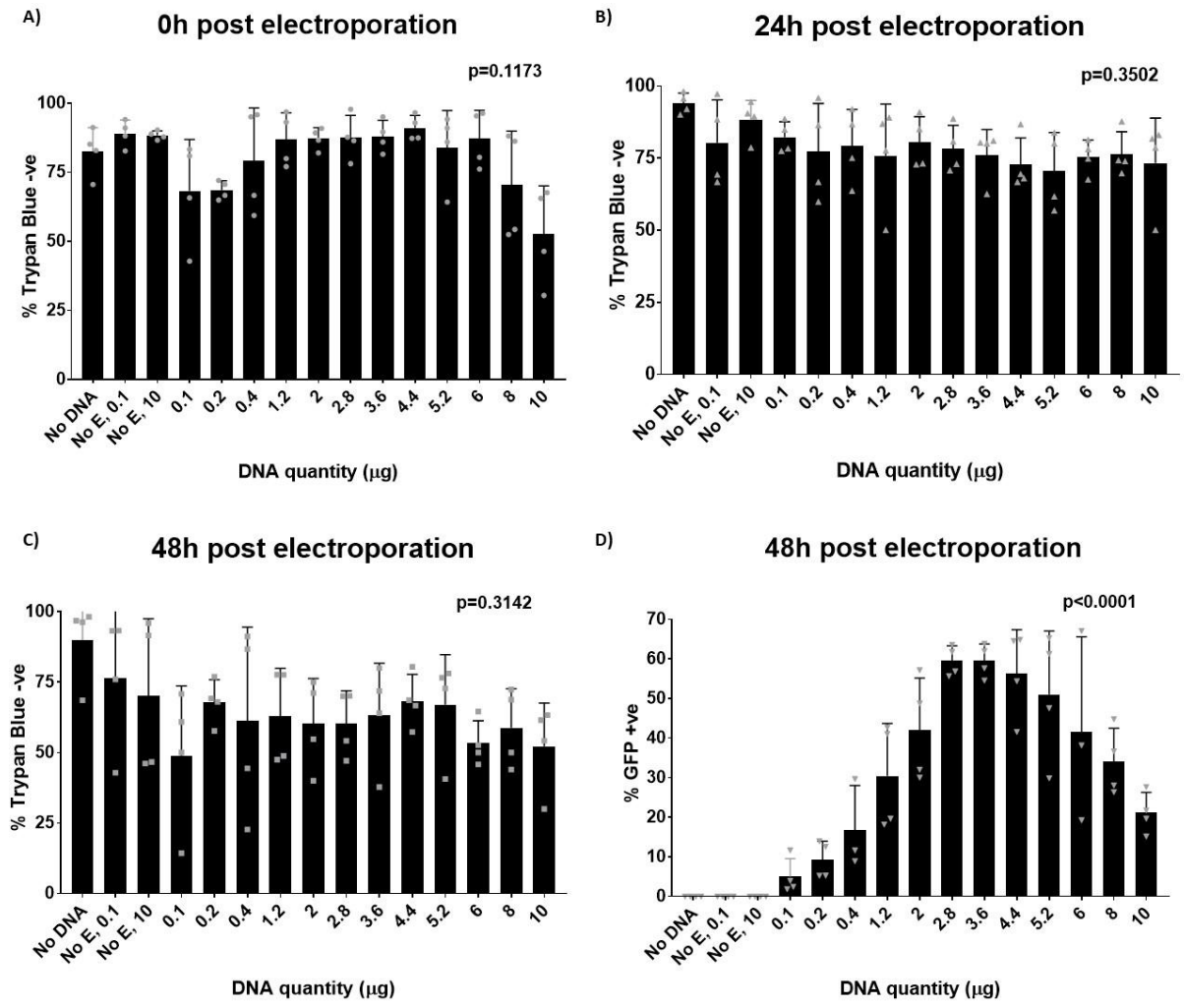


Figure 3.8: 2.8µg was the optimal plasmid amount for Ba/F3 cell electroporation Viability **A)** 0h **B)** 24h **C)** 48h post electroporation. **D)** Percentage GFP positive by flow cytometry 48h post electroporation, pre-gated on live singlets. N= 4 (3 biological replicates with 1-2 technical replicates) except for 5.2µg and 0.4µg: N=3. Data represent mean with standard deviation. Significance calculated using one-way ANOVA.

In order to directly compare electroporation with nucleofection, the range of plasmid amounts used to test electroporation was compared to the commercially optimised nucleofection program for the BaF3 cell line (http://www.lonzabio.com 2017). Due to the fact that the extreme ends of plasmid amounts (8-10µg) showed low efficiency for the electroporation system, the range was narrowed for nucleofection to 0.2-6µg. Similar to electroporation,

2.8µg was the optimal plasmid amount for transfection efficiency (**Figure 3.9**) although 1.2µg also gave a very similar efficiency.

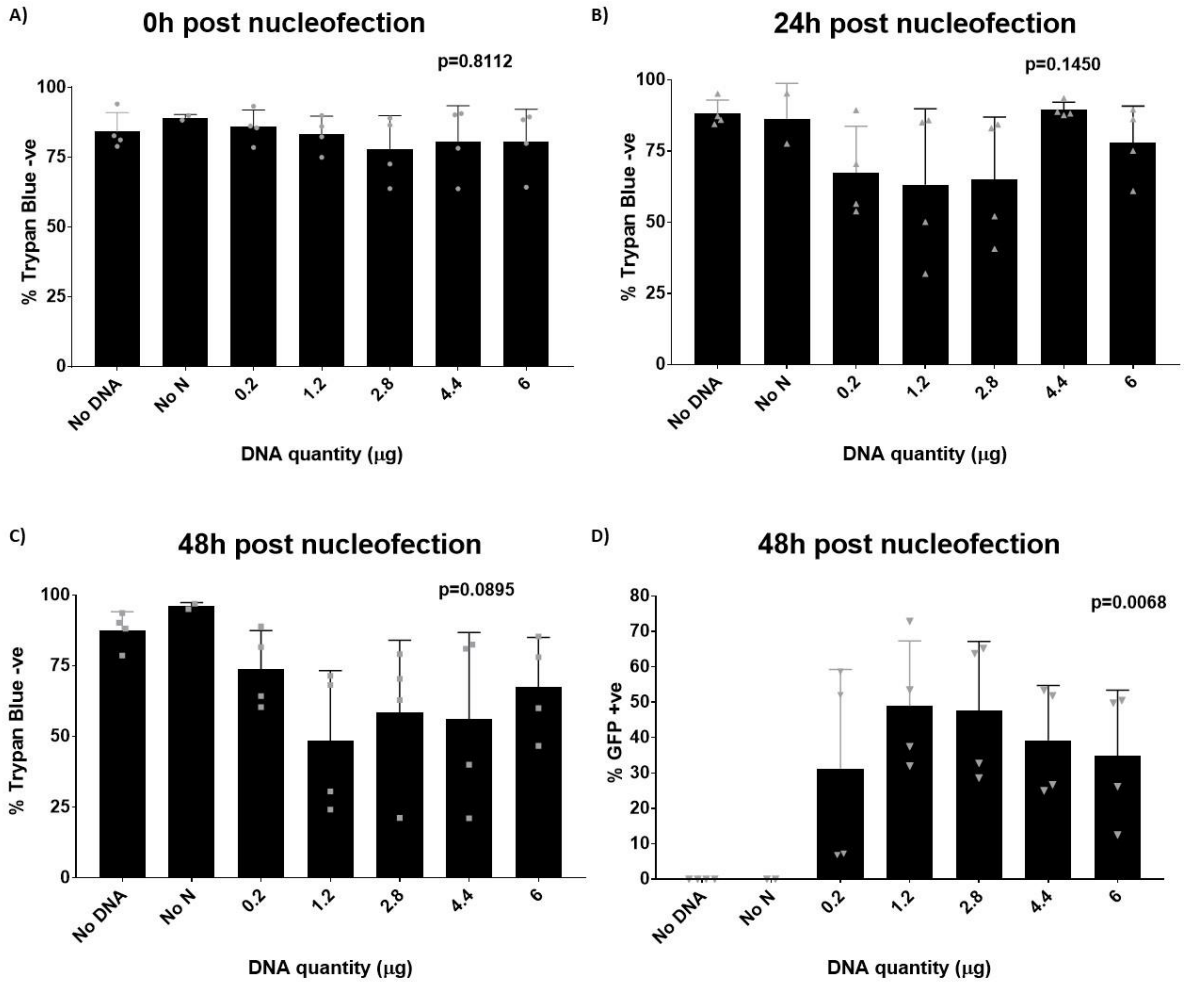


Figure 3.9: 2.8µg was the optimal plasmid amount for Ba/F3 cell nucleofection Viability **A)** 0h **B)** 24h **C)** 48h post-nucleofection (N). **D)** Percentage GFP-positive by flow cytometry 48h post-nucleofection, pre-gated on live singlets. N= 4 (2 biological replicates) except for no N: N=2. Data represent mean with standard deviation. Significance calculated using one-way ANOVA.

In order to look at electroporation and nucleofection side-by-side, the three most optimal plasmid amounts were compared on the same graph.

Interestingly, nucleofection gave higher transfection efficiency than

electroporation at lower plasmid amounts (1.2µg) although this was not statistically significant (**Figure 3.10 D**, p=0.3429).

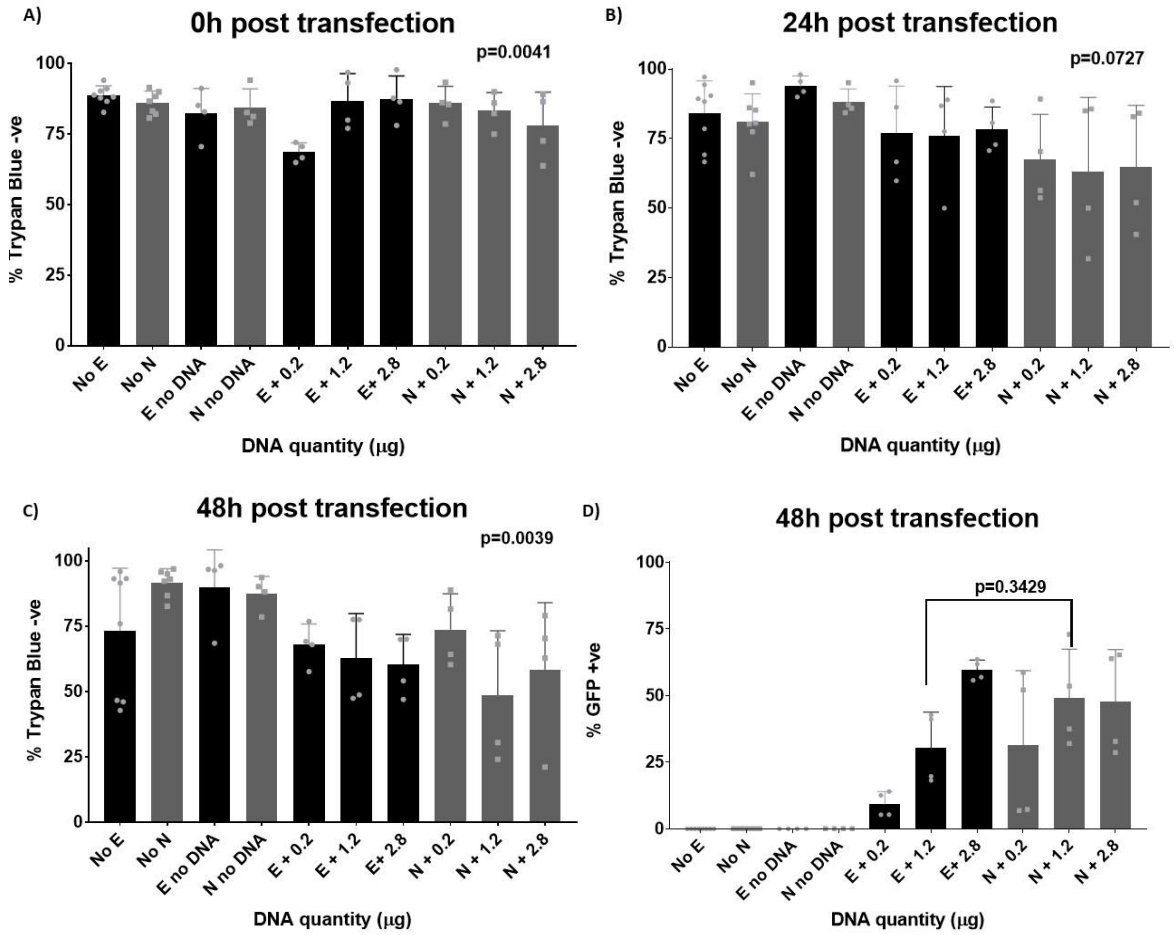


Figure 3.10: Comparison of three most optimal electroporation and nucleofection conditions For 3 most optimal E (black bars) and N (grey bars) conditions**,** viability **A)** 0h **B)** 24h **C)** 48h post transfection. **D)** Percentage GFP positive 48h post transfection. N= 4 (2 biological replicates) except for no N: N=2. Data represent mean with standard deviation. Significance calculated using one-way ANOVA. For 1.2µg comparison, significance calculated using a non-parametric un-paired Mann Whitney t-test.

Taken together, these data highlight that transfection conditions greatly affected the viability and transfection efficiency; it is therefore not surprising that significant commercial effort has focused on optimising these conditions for a wide variety

of cell lines and primary cells. It has been reported that nucleofection buffers deliver the CRISPR machinery directly to the nucleus, which may explain why higher transfection efficiencies were seen for lower plasmid amounts when compared to the optimal electroporation conditions, although this was not statistically significant (**Figure 3.10 D,** p=0.3429). From these results, it was decided to test a narrow range of both electroporation and nucleofection conditions in primary mouse HSPCs given that, although nucleofection was more efficient particularly at low plasmid amounts, this effect was subtle and optimal transfection conditions are known to vary greatly between cell types.

3.1.6 Optimisation of CRISPR delivery mode and format for tdTomato induction in mouse BM HSPCs

i) Comparison of plasmid electroporation and nucleofection

In order to be able to assess the effect of CRISPR editing at the bulk level *in vitro* and *in vivo,* it was of paramount importance to optimise an approach to achieve editing of 90-100% of BM HSPCs. This is because at the short-term 24-hour time-point, loss or gain of tdTomato expression would likely not be complete due to a delay between editing and protein turnover/translation. However, as the potential transcriptional changes induced by DNA damage may have already been induced, it was first necessary to sort the cells without gating for tdTomato expression in order to catch this short-term effect. In addition, as editing may impede the engraftment level, it was important to aim for 90-100% editing efficiency in order to generate a high enough yield of edited cells to sort 18 weeks

post-transplant. To this end, three delivery modes: electroporation, nucleofection and transduction, and three delivery formats: plasmid, RNP (by electroporation or nucleofection) and viral vector (by transduction) were tested. The following figures summarise the results for optimisation of the tdTomato induction approach.

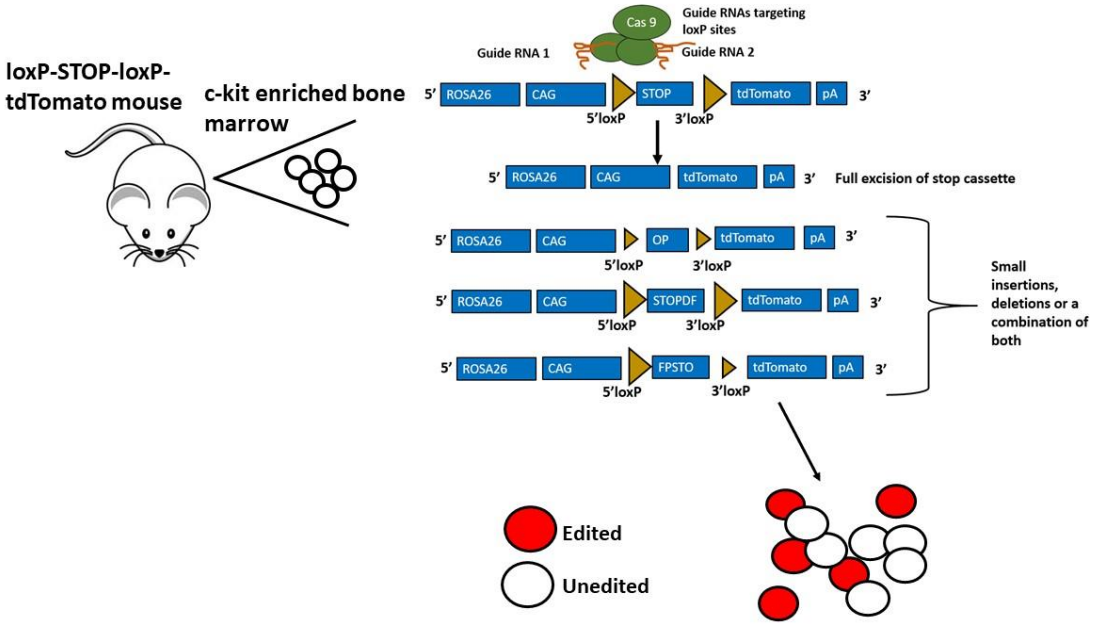


Figure 3.11: Schematic of experimental approach for targeting loxP-STOP-loxP tdTomato mice with dual guide RNAs to induce tdTomato expression

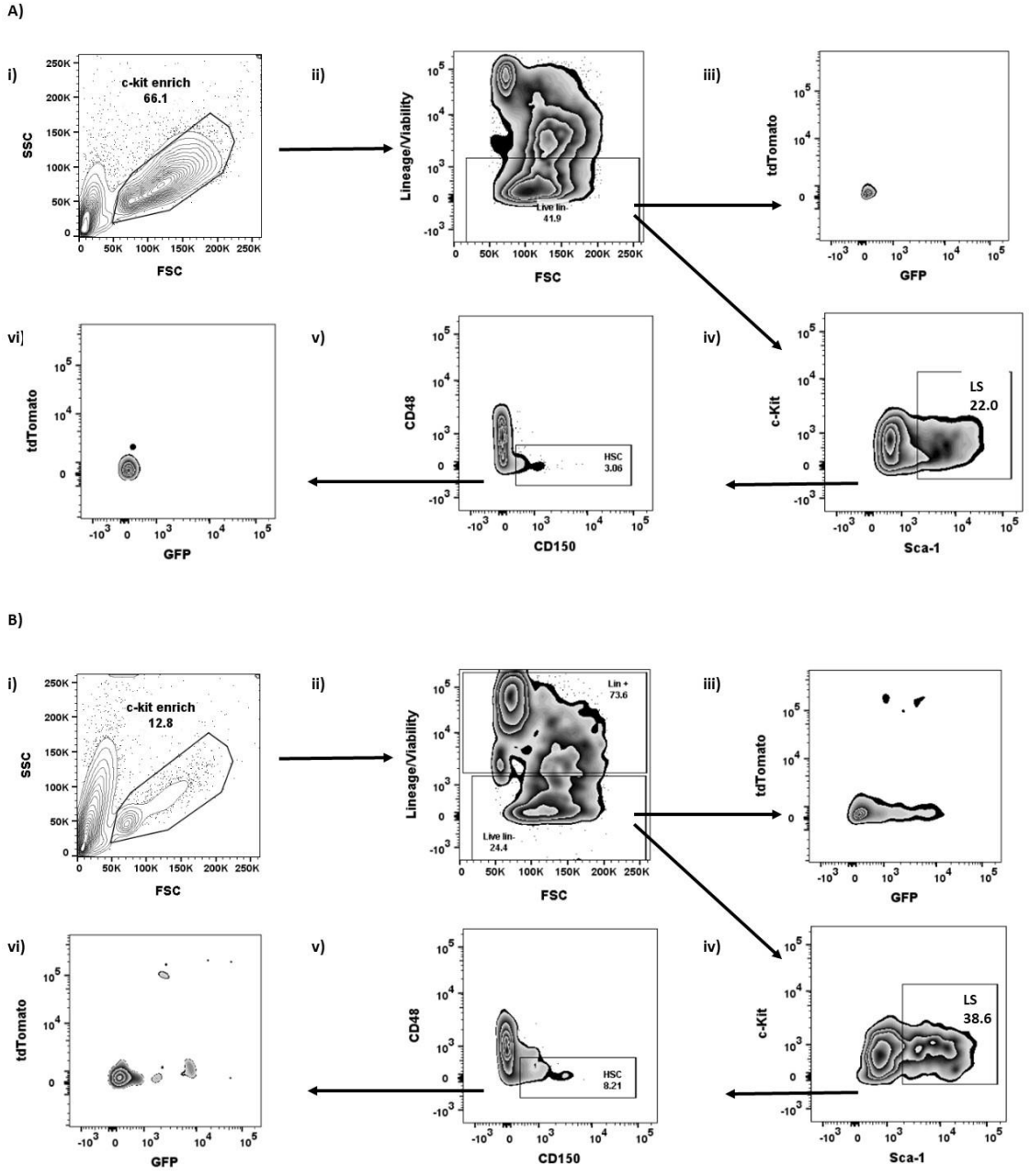


Figure 3.12: Flow cytometry gating strategy for analysis of tdTomato induction efficiency in mouse HSPCs Example flow cytometry plot for 200,000 c-Kit enriched Ai9-Rosa26 mouse BM HSPCs 48h post-nucleofection with **A)** No DNA control **B)** dual 5' and 3' loxP guide RNA plasmids (total 1µg). **i)-iv)** gating strategy, HSCs segregated as Lin-CD48-CD150+Sca-1+ **iii)** Percentage GFP- and tdTomato-positive within Lin- gate **vi)** Percentage GFP- and tdTomtato-positive within HSC gate.

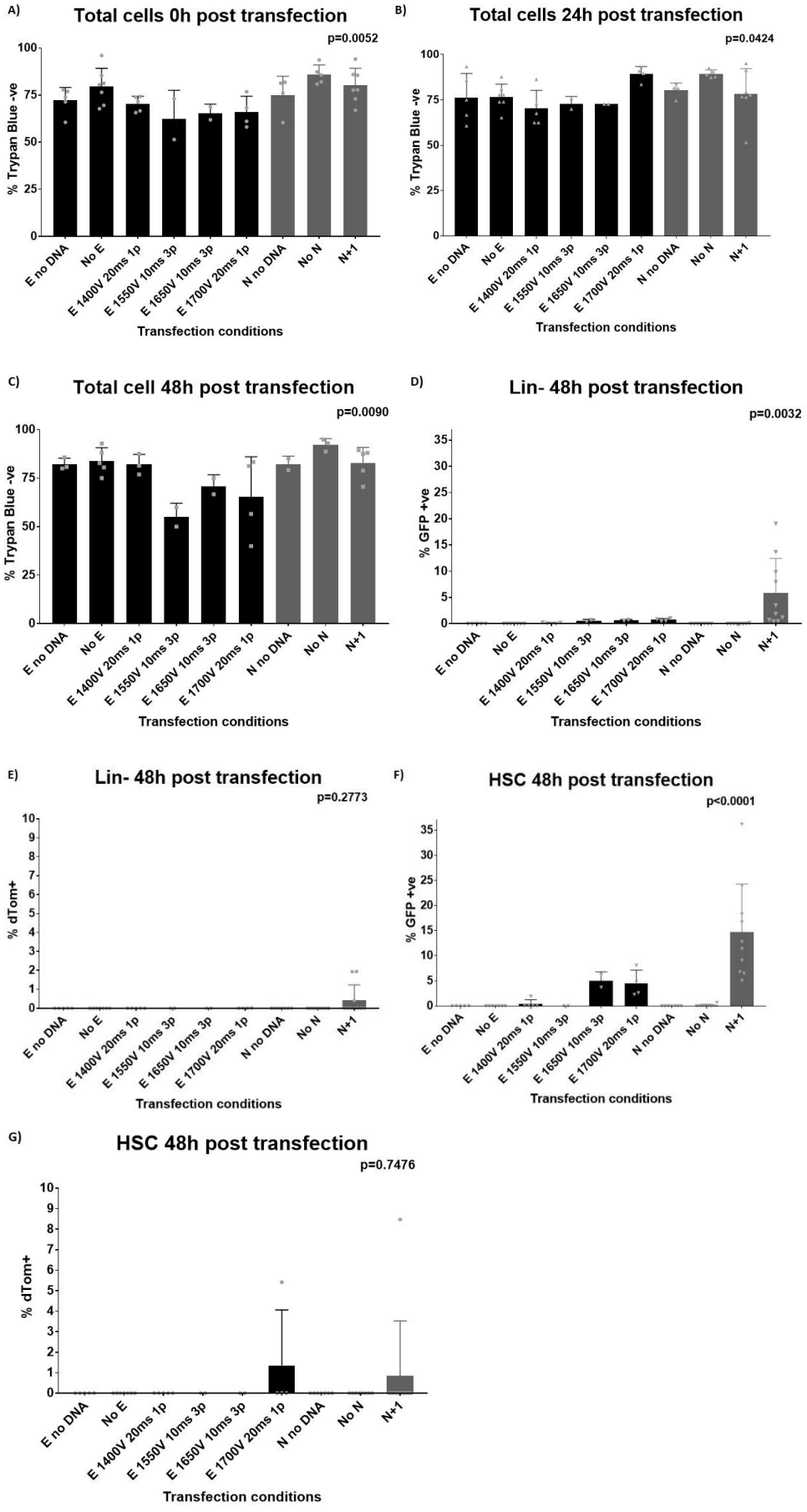

A)
Total cells 0h post transfection
p=0.0052
% Trypan Blue -ve
B)
Total cells 24h post transfection
p=0.0424
% Trypan Blue -ve
C)
Total cell 48h post transfection
p=0.0090
% Trypan Blue -ve
D)
Lin- 48h post transfection
p=0.0032
% GFP +ve
E)
Lin- 48h post transfection
p=0.2773
% dTom+
F)
HSC 48h post transfection
p<0.0001
% GFP +ve
G)
HSC 48h post transfection
p=0.7476
% dTom+
E no DNA
No E
E 1400V 20ms 1p
E 1550V 10ms 3p
E 1650V 10ms 3p
E 1700V 20ms 1p
N no DNA
No N
N+1
Transfection conditions

Figure 3.13: Nucleofection was superior to electroporation for tdTomato induction by plasmid CRISPR delivery in mouse BM HSPCs Total c-Kit+ viability **A)** 0h **B)** 24h **C)** 48h post transfection. 48h post transfection percentage **D)** GFP-positive within Lin- gate **E)** tdTomato-positive within Lin- gate **F)** GFP-positive within HSC gate **G)** tdTomato-positive within HSC gate. N= 2-10 (of 5 biological replicates, different combinations of conditions for each repeat). Data represent mean with standard deviation. Significance calculated using one-way ANOVA.

For all electroporation and nucleofection conditions tested, the GFP level (indicating transfection efficiency) was less than 10% in both the lineage negative and HSC compartments for the successful conditions (**Figure 3.13**). Although transfection was evident for both electroporation with 1700V 20ms 1p and nucleofection (program DK100), editing (tdTomato-positive) was only seen for nucleofection, again at modest levels in line with the levels of transfection (**Figure 3.13**). This level of transfection efficiency was in contrast to Ba/F3 cells in which there was ~65% efficiency for pmaxGFP plasmid transfection (**Figure 3.10**) for the optimal electroporation/nucleofection conditions. What was different between the pmaxGFP plasmid and tdTomato guide RNA induction plasmid was the size. pmaxGFP is a small (~4kb) plasmid whereas the PX458 guide RNA induction plasmid with two guides inserted is more than twice this size (~9kb). It could be that the larger plasmid size inhibited transfection although it is difficult to directly compare without data for transfection of the PX458 plasmid in Ba/F3 cells, as it could also be due to primary cells being less amenable to plasmid transfection than cell lines. From these data comparing plasmid electroporation and nucleofection it was concluded that although nucleofection was superior to electroporation, this format of CRISPR delivery was not sufficient to induce high levels of tdTomato induction in mouse BM HSPCs.

ii) Assessment of RNP nucleofection

Given the modest superiority of nucleofection compared to electroporation in mouse BM HSPCs, the same nucleofection conditions as described above were used to test the same guide RNAs in RNP format as opposed to plasmid. It was predicted that the smaller size and short-lived stability of this construct compared to the longer-lived 9kp plasmid may be a more efficient format of delivering the CRISPR machinery. The 4kb pmaxGFP plasmid was included as a positive control for transfection efficiency.

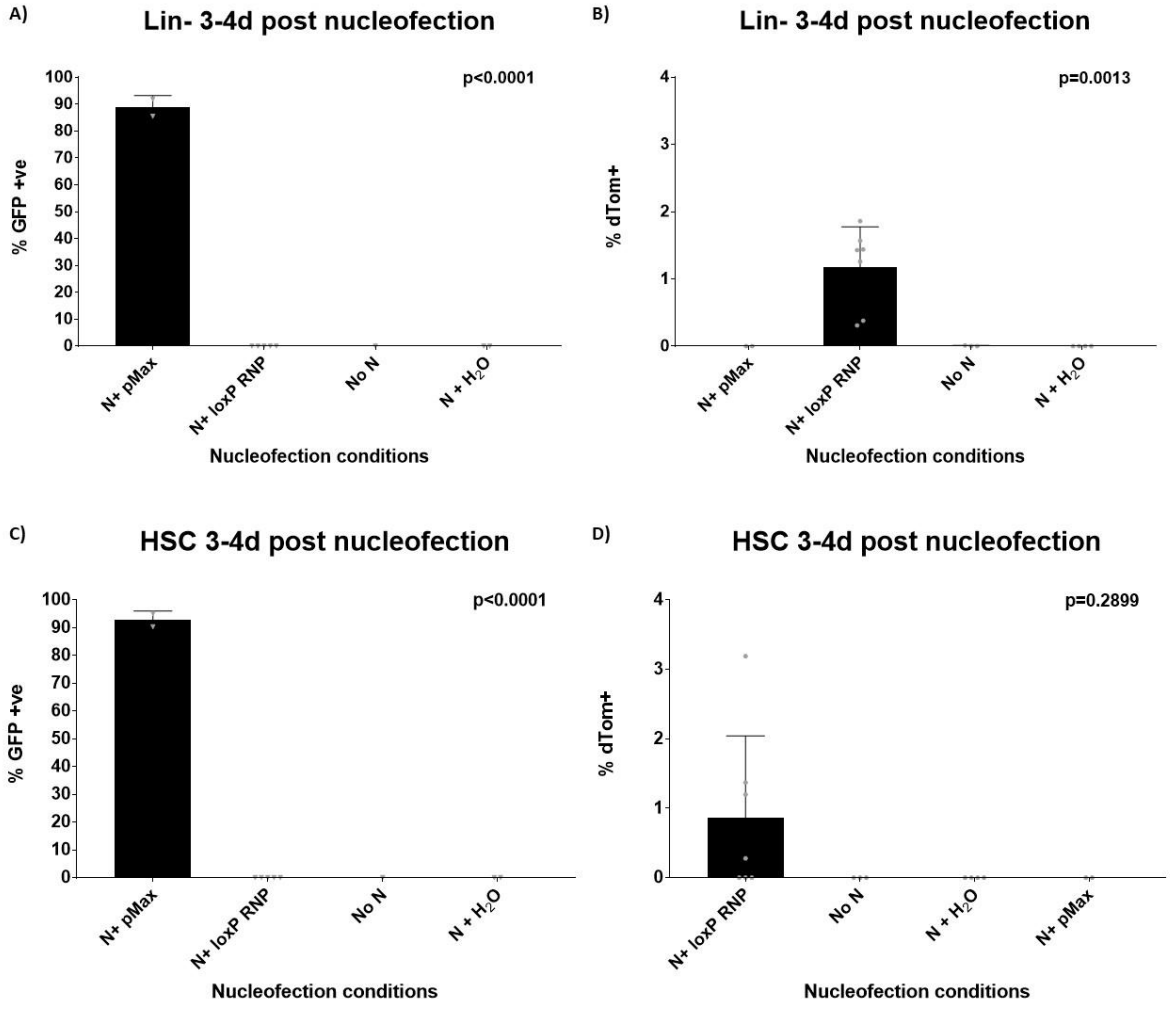


Figure 3.14: RNP CRISPR induction of tdTomato expression was inefficient despite nucleofection being a highly efficient transfection approach Percentage **A)** GFP-positive within Lin- gate **B)** tdTomato-positive within Lin- gate **C)** GFP-positive within HSC gate **D)** tdTomato-positive within HSC gate. N= 2-7 (of 3 biological replicates, different combinations of conditions for each repeat). Data represent mean with standard deviation. Significance calculated using one-way ANOVA.

Nucleofection with the 4kb GFP plasmid resulted in 75-80% transfection efficiency. Despite this, tdTomato induction with loxP targeting RNPs was ~2% suggesting low efficiency of in frame editing (**Figure 3.14**)

iii) Assessment of viral transduction

LoxP guide RNAs, with the same sequence as in the plasmid and RNP approach were introduced into mouse BM HSPCs using virus transduction with dual guide RNA vectors encoding the Cas9-GFP fusion construct under the EFS or SFFV promoters. Due to the large size of the vector (12kb), at the top end of the packaging limitations of the lentivirus, it was not possible to generate as high a titre of this construct compared to the small control virus (9.9kb). Therefore, in order to achieve a multiplicity of infection (MOI) within the functional range of transduction, a smaller number of cells (50,000 c-Kit-enriched) were used compared to nucleofection. For this reason, the total c-Kit compartment was analysed and not split into Lin-/HSC as this would have yielded too low cell numbers from which to draw accurate conclusions. **Figure 3.15** summarises the transduction and editing efficiency.

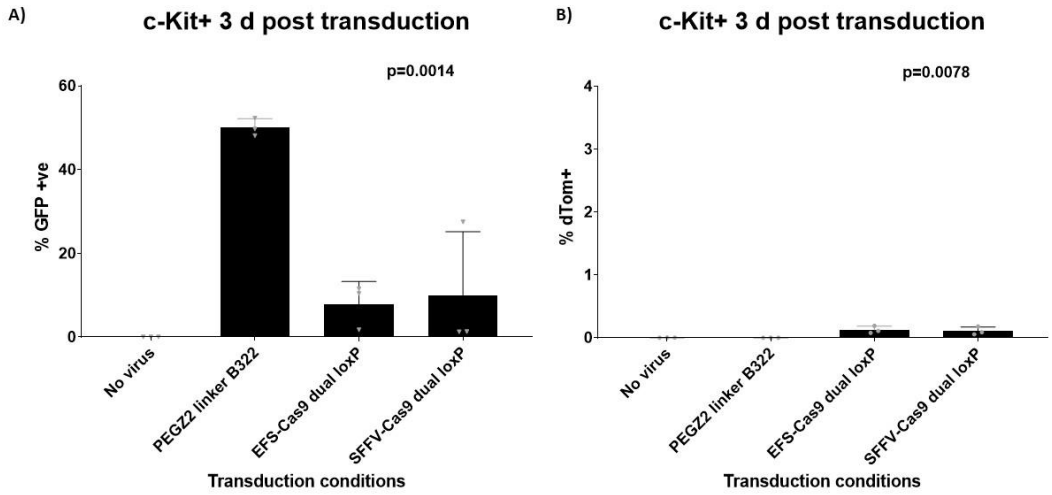


Figure 3.15: Viral transduction was not efficient in delivering CRISPR components to induce tdTomato expression despite high levels of control virus transduction Total c-Kit+ **A)** Percentage GFP-positive **B)** Percentage tdTomato-positive. N= 3. Data represent mean with standard deviation. Significance calculated using a Kruskal-Wallis test.

For the small barcoding virus vector (9.9kb), transduction efficiency was 45% on average. Whereas for the much larger (12kb) Cas9-GFP dual guide viral vectors transduction efficiency was 10% on average. Of the successfully transduced cells, there was on average 10% showing tdTomato expression, hence successful editing.

Taken together, **Figures 3.13-3.15** demonstrate that both transduction and transfection, in either plasmid or RNP format, were inefficient methods for inducing tdTomato expression by excision of the STOP cassette, normally excised by Cre-recombinase. This was despite, under control conditions, both nucleofection and transduction showing high efficiency, 75-80% and 45% respectively, in mouse BM HSPCs. This could be due to inefficient excision caused by indels at both loxP targeted sites, instead of complete excision of the cassette. Another possibility is that excision of the cassette was successful but

in most cases didn't bring the tdTomato sequence in frame leading to a lack of protein transcription/translation. Therefore, it was hypothesised that tdTomato knockout might be a more efficient editing approach for two reasons. First, a knockout approach would not require retention of an in frame coding sequence, in fact insertions and deletions leading to out of frame sequence would cause the desired effect. Given that the NHEJ approach is much more likely to produce insertions/deletions leading to out of frame sequence, it was predicted that loss of tdTomato expression would be more efficient. Secondly, the fact that the tdTomato coding sequence is made up of two repeated regions, a single guide designed against these repeated regions would allow the potential of cutting in two places, increasing the chances of indels and thus loss of tdTomato expression. A single guide approach would also reduce the size of the vectors and number of RNPs required to transfect which could have the potential to increase transfection/transduction efficiency.

3.1.7 Optimisation of CRISPR delivery mode and format for tdTomato knockout in mouse BM HSPCs

Given the electroporation vs. nucleofection results of the tdTomato induction approach, it was decided to compare plasmid and RNP guide delivery by nucleofection with viral transduction, for tdTomato knockdown. Due to the fact that this approach was amenable to single guide-based editing, it was not necessary to initially test a wide range of combinations in a cell line; instead tdTomato targeting guides were tested directly on mouse BM HSPCs. Guide optimisation was carried out using the RNP delivery format followed by cloning

of the optimal guide into the PX458 plasmid vector and both EFS-Cas9 and SFFV-Cas9 viral vectors. In order to determine an estimate for the overall viability after each delivery mode, the percentage of cells in the first flow gate (FSC/SSC), segregating viable cells, was measured. **Figure 3.2** summarises the guide design approach. The following figures summarise the results for the methods tested to induce tdTomato knockout.

i) Assessment of RNP nucleofection and guide selection

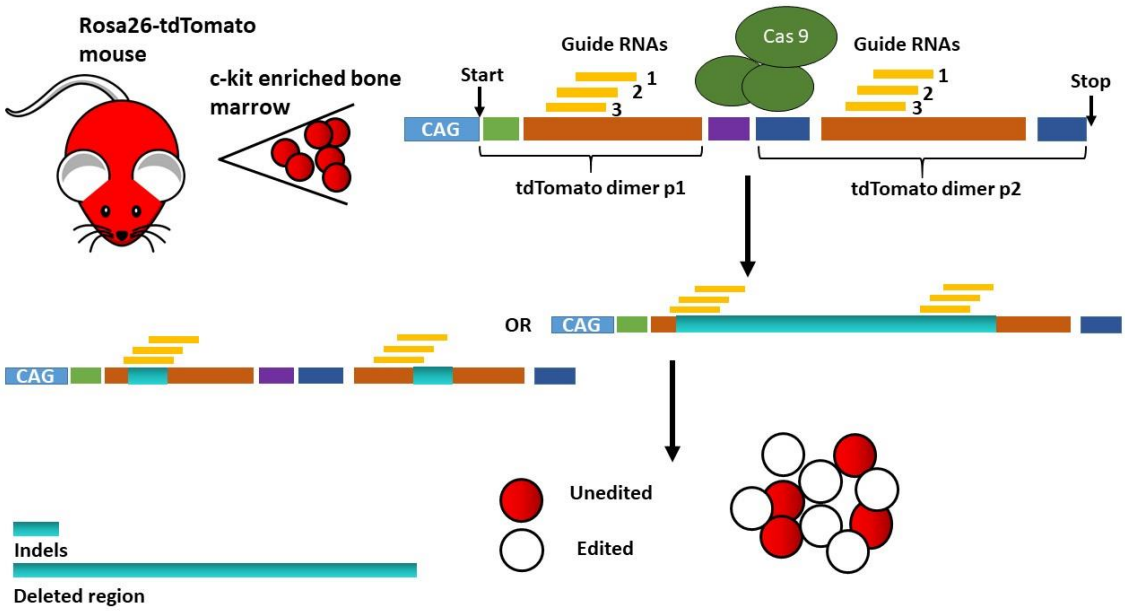


Figure 3.16: Schematic of tdTomato knockout approach and targeting guide RNAs

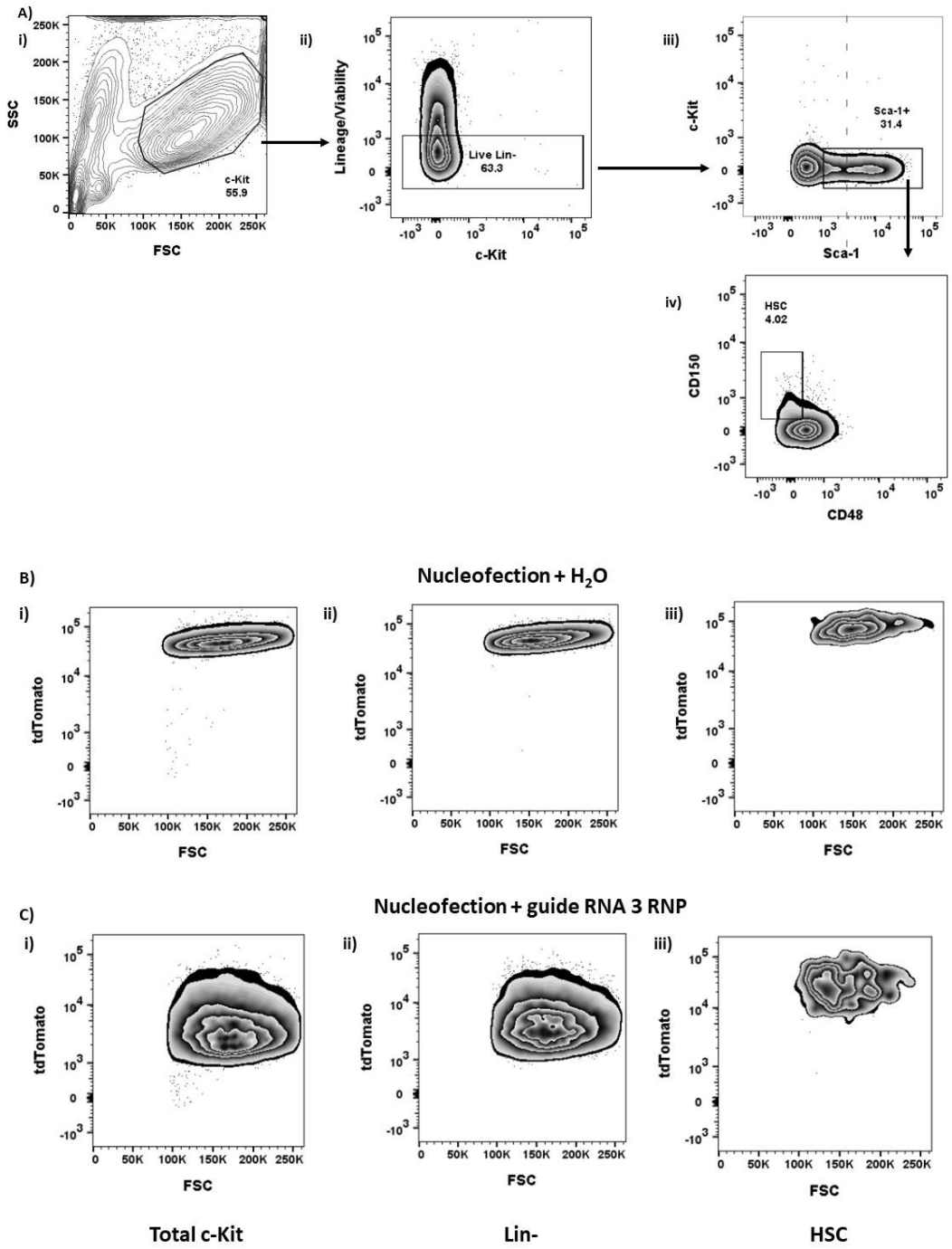


Figure 3.17: Example flow cytometry plots for analysing tdTomato knockdown in mouse c-Kit BM using RNP nucleofection Example flow cytometry plots for **A)** Nucleofection + H_2O **i)-iv)** gating strategy, HSCs segregated as Lin- Sca-1+ CD48- CD150+. tdTomato expression in **i)** c-Kit **ii)** Lin- **iii)** HSC compartments for **B)** nucleofection + H_2O **C)** nucleofection + guide RNA 3 RNP.

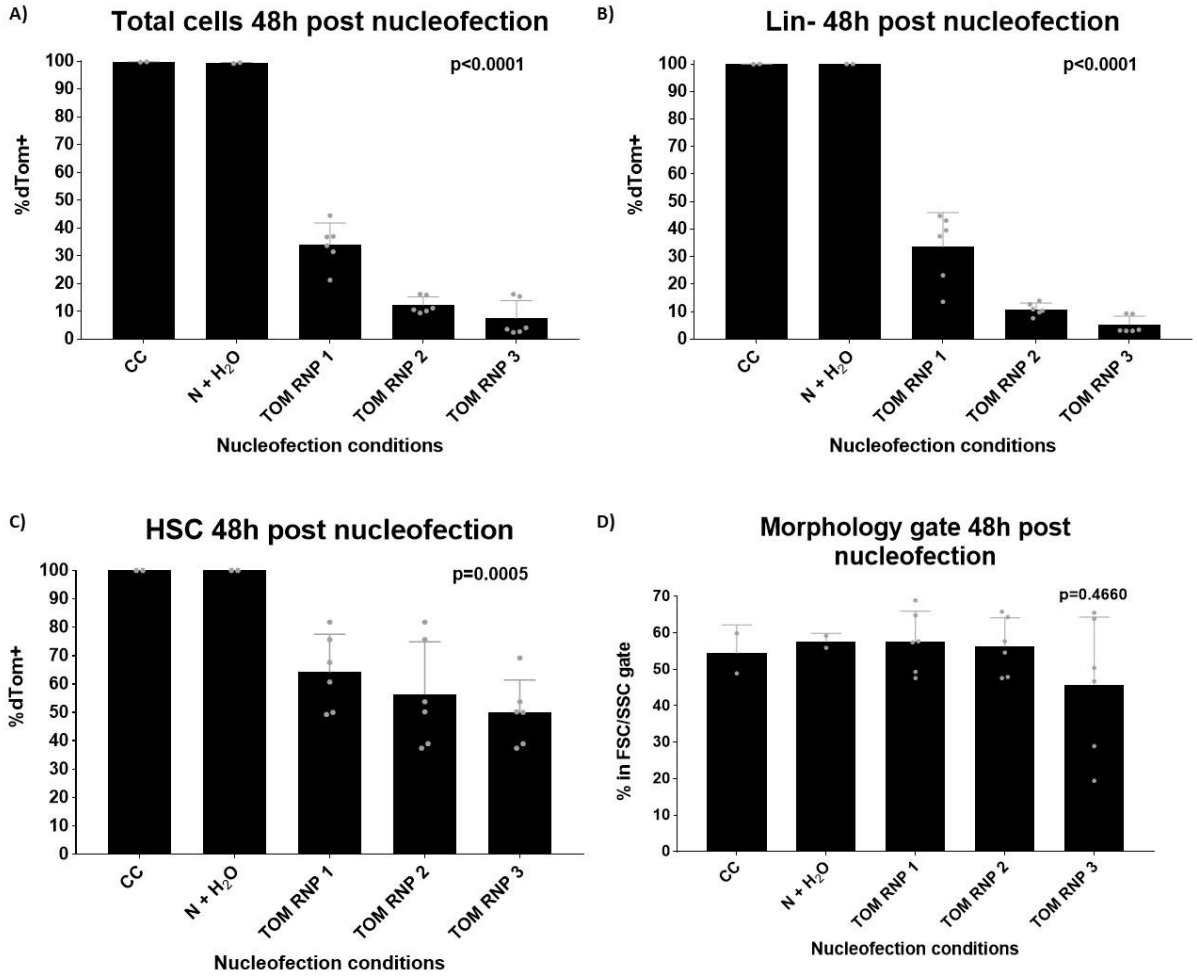


Figure 3.18: RNP nucleofection was efficient in targeting tdTomato knockout in mouse BM HSPCs. Guide 3 gave optimal editing without significantly reducing viability Percentage tdTomato positivity in **A)** Total cells **B)** Lin- and **C)** HSC compartments. **D)** Percentage cells in initial FSC/SSC gate as a marker of overall cell viability. N= 2-6 (3 biological replicates). Data represent mean with standard deviation. Significance calculated using one-way ANOVA.

Less than 30% tdTomato expression was seen in the c-Kit and Lin- compartment 48h post-nucleofection. For guide RNA 3, this was consistently below 5%. Editing efficiency in the HSC compartment was 50% for guide RNA 3 at 48 hours. The percentage of cells in the initial FSC/SSC gate, a proxy for viability, was not significantly different between the RNP targeted and control groups. Therefore

RNP based editing of the tdTomato locus in c-Kit enriched mouse bone marrow did not reduce cell viability compared to unedited or cultured controls.

ii) Assessment of plasmid nucleofection and viral transduction

Having established guide RNA 3 as the most efficient using the RNP delivery format, this guide was cloned into both the plasmid PX458 and viral EFS/SFFV-Cas9 vectors. Transfection and transduction was carried out on mouse BM HSPCs to assess the efficiency of tdTomato knockout. In the case of plasmid transfection, the viability was significantly reduced and almost no HSCs remained in the phenotypic gate after 48 hours for these samples. Therefore only the c-Kit+ and Lin- populations were analysed for this group (**Figure 3.19**).

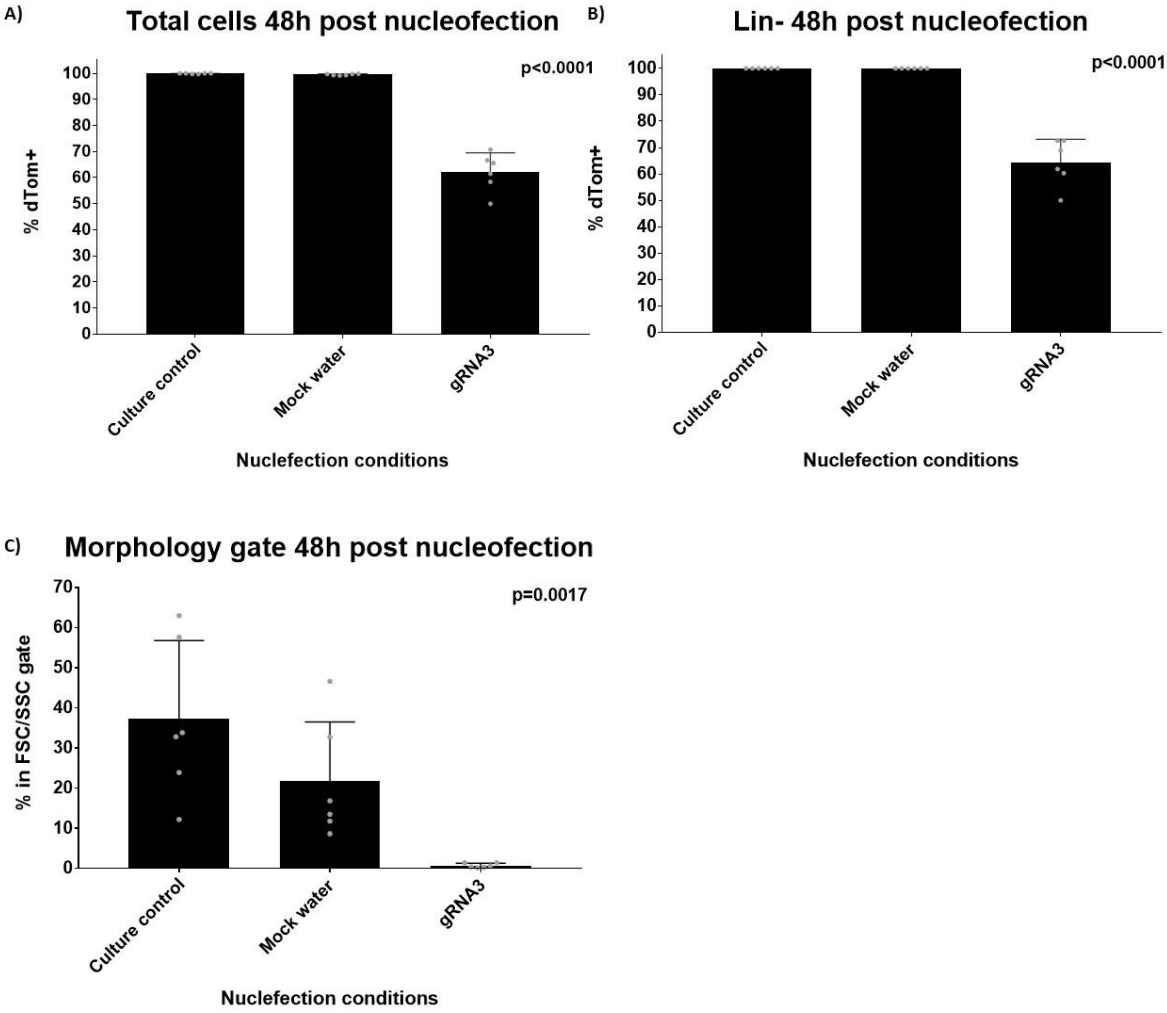


Figure 3.19: Plasmid nucleofection gave significantly reduced tdTomato expression but significantly reduced viability Percentage tdTomato positive in **A)** Total cells **B)** Lin- compartments. **C)** Percentage cells in initial FSC/SSC gate as a marker of overall cell viability. N= 6 (3 biological replicates, 2 technical replicates). Data represent mean with standard deviation. Significance calculated using one-way ANOVA.

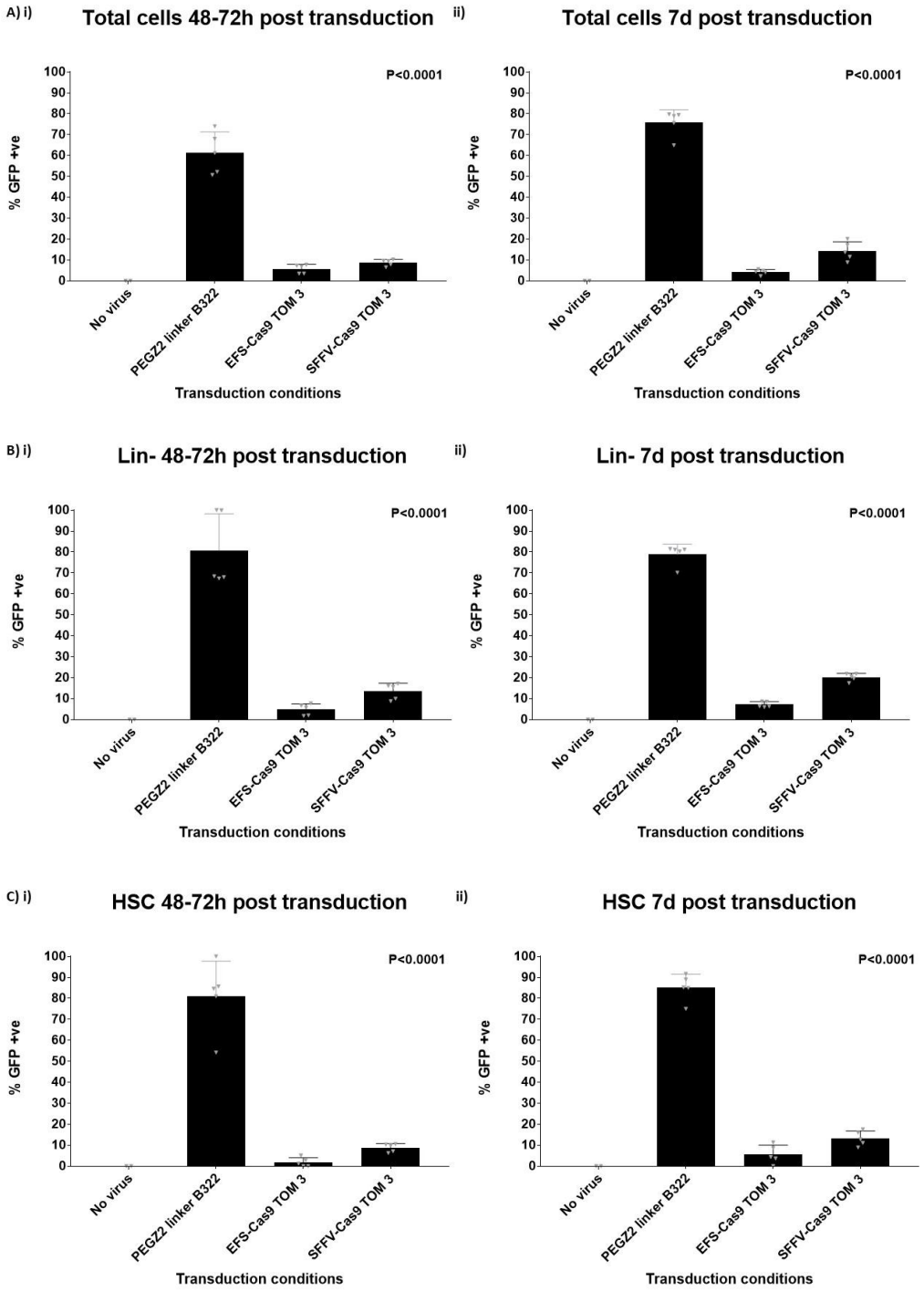


Figure 3.20: Virus transduction was highly efficient with a small control but significantly less efficient with guide RNA vectors Percentage GFP-positive in **A)** Total cells **i)** 48-72h **ii)** 7d **B)** Lin- **i)** 48-72h **ii)** 7d **C)** HSC **i)** 48-72h **ii)** 7d post-transduction. N= 2-5 (2 biological replicates, 2-3 technical replicates). Data represent mean with standard deviation. Significance calculated using one-way ANOVA.

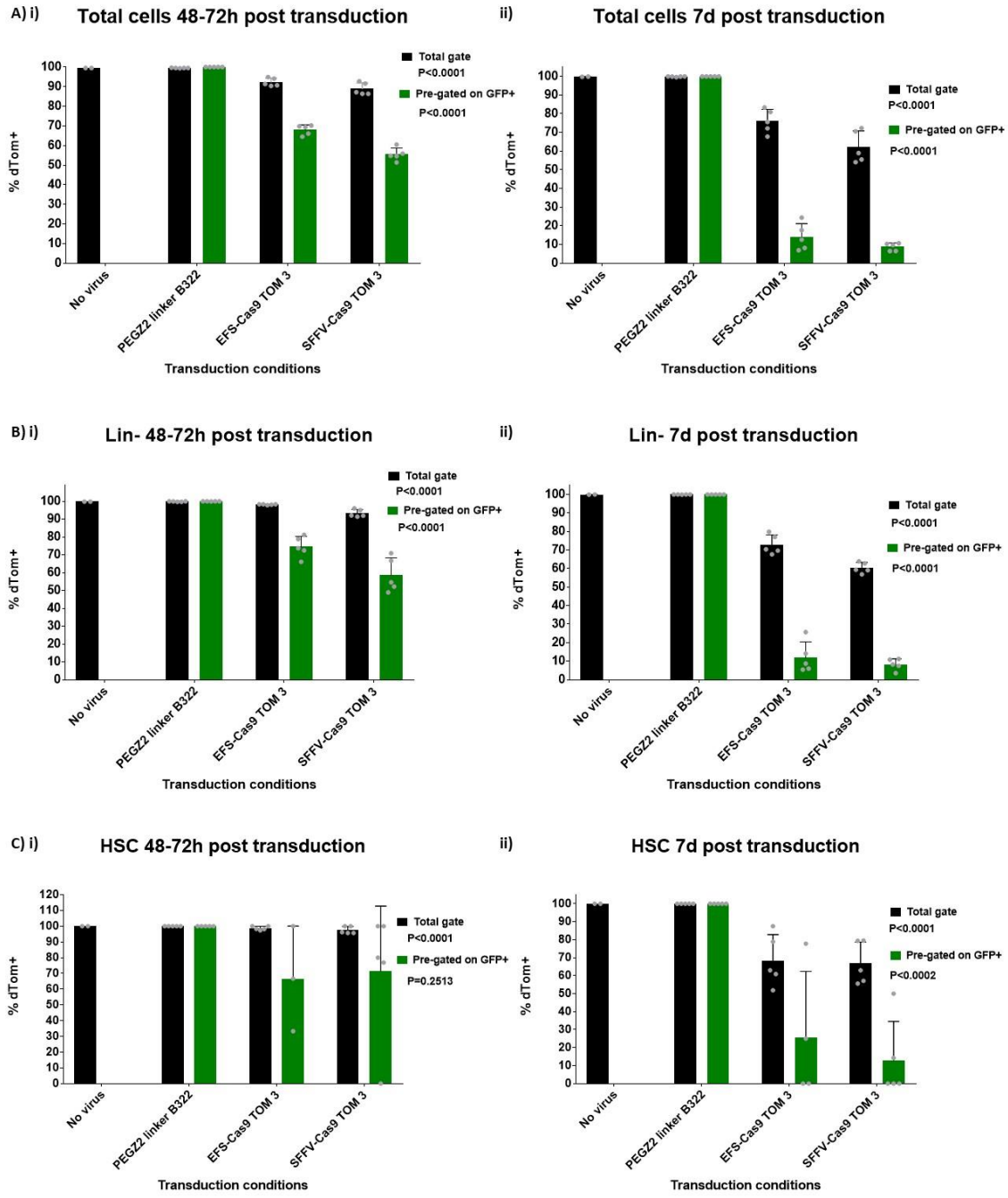


Figure 3.21: Virus transduction significantly reduced tdTomato expression with a delayed editing window compared to plasmid and RNP transfection Percentage tdTomato-positive in total gate (black) or pre-gated GFP+ population for **A)** Total cells **i)** 48-72h **ii)** 7d **B)** Lin- **i)** 48-72h **ii)** 7d **C)** HSC **i)** 48-72h **ii)** 7d post-transduction. N= 2-5 (2 biological replicates, 2-3 technical replicates). Data represent mean with standard deviation. Significance calculated using one-way ANOVA.

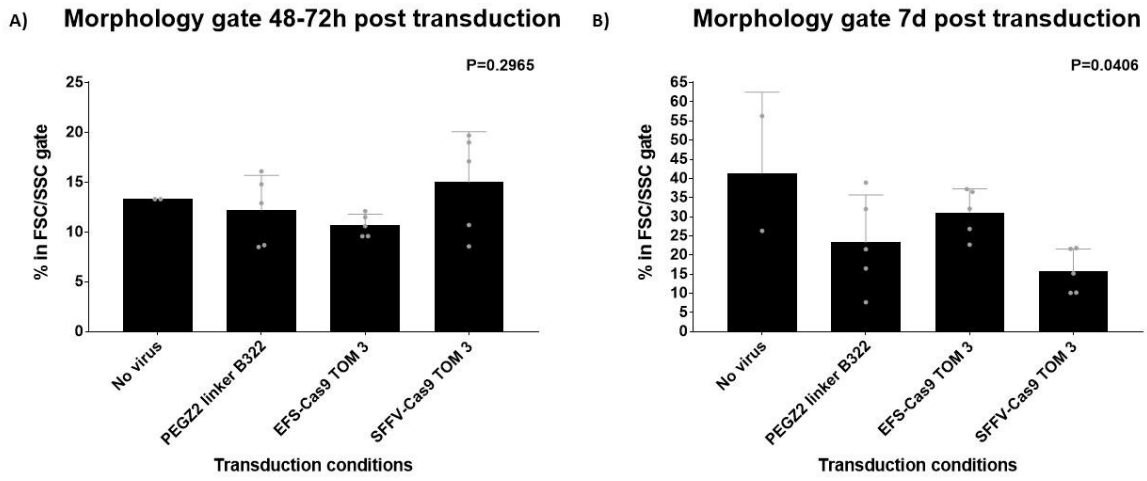


Figure 3.22: Virus transduction did not significantly reduce viability of mouse HSPCs Percentage in initial FSC/SSC gate as a marker of overall cell viability **A)** 48-72 **B)** 7d post transduction. N= 2-5 (2 biological replicates, 2-3 technical replicates). Data represent mean with standard deviation. Significance calculated using one-way ANOVA.

Taken together, **Figures 3.18-3.22** demonstrate that all three delivery modes: plasmid nucleofection, RNP nucleofection, and transduction, were efficient methods for the delivery of constructs into BM HSPCs. This is evident from the high level of GFP positivity for both plasmid nucleofection and transduction (80% with small control vectors), and by proxy the loss of tdTomato (>90% for guide RNA 3) in the RNP context. What is clear for the plasmid and viral approaches however, is that transfection and transduction of the large Cas9-guide RNA constructs was significantly less efficient than the small control constructs leading to ~50% less editing compared to RNP delivery and, in the case of plasmid transfection, significantly reduced cell viability.

This result is in contrast to several CRISPR studies in which HSPCs were successfully edited using plasmid delivery of the CRISPR machinery. For example, Tothova and colleagues used a plasmid transfection approach in

human CB CD34+s to target for CRISPR editing multiple mutations associated with myelodysplasia (Tothova, Krill-Burger et al. 2017). It is possible that introducing mutations which promote cell survival, as opposed to a relatively 'inert' locus like tdTomato, would increase the likelihood of cells to survive the CRISPR process and this is why the plasmid approach was successful in this study. Indeed, in a study optimising targeting of the CD45 locus, a pan-haematopoietic cell surface marker, in murine and human HSPCs, an RNP approach was taken instead of plasmid targeting (Gundry, Brunetti et al. 2016). In addition, in the CD34+ targeting approach to model myelodysplasia, Cas9 and guide RNAs were delivered in separate vectors which were 9kb (Addgene 44719) and ~4kb (Addgene 41824). While the Cas9 vector used is similar in size to the PX458 combined Cas9 and guide RNA vector used in this thesis, the guide RNA vector used in the CD34+ study was considerably smaller. It could be that expressing Cas9 and the guide RNA from separate vectors reduced overall toxicity in this case. Indeed, in a separate study by Mandal and colleagues, using Cas9 and guide RNAs targeting *CCR5* or *B2M* on segregated plasmids, efficient editing and engraftment in NSG mice was achieved (Mandal, Ferreira et al. 2014). Of note, far less has been reported on using plasmid CRISPR delivery approaches in mouse HSPCs. It is possible that human HSPCs are less susceptible to toxic effects of the plasmid approach than their mouse counterparts.

Interestingly, while transduction efficiency was low for the large viral vectors in the study carried out in this thesis, of those cells transduced, 80-90% were edited in the total cell and Lin- gated populations (**Figure 3.21**, **A** and **B**, pre-gated on GFP+) seven days post-transduction. This indicated that of those cells

transduced, the editing efficiency was high, therefore the transduction efficiency was the limiting factor to this approach. The editing level in the transduced HSC compartment was more variable (**Figure 3.21 C**). Figure **3.23** summarises the editing level for each of the three approaches tested; 48 hour readouts for plasmid/RNP nucleofection were compared to seven day readouts for transduction due to the delayed expression of the CRISPR machinery using this approach.

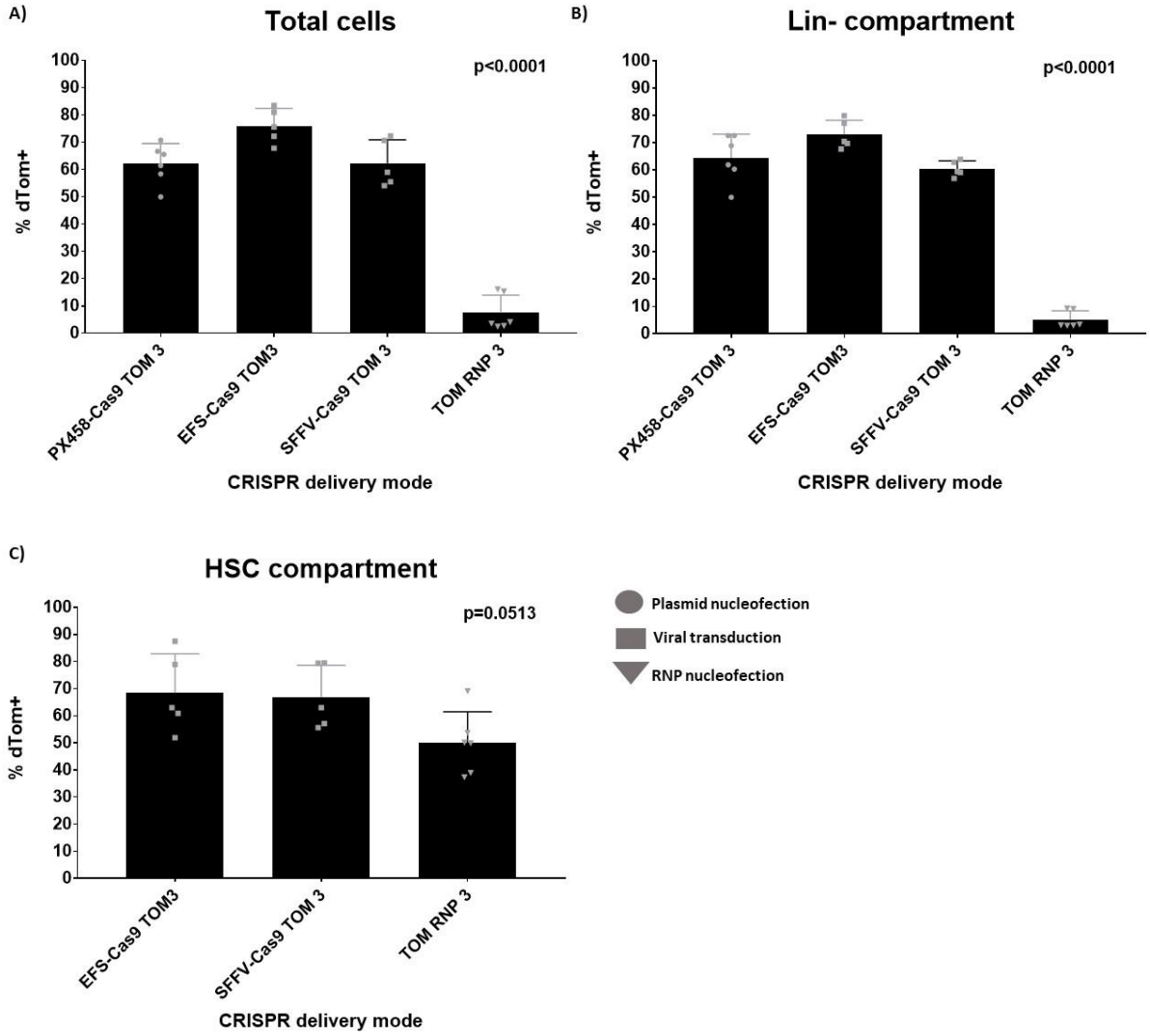


Figure 3.23: RNP nucleofection was superior to plasmid nucleofection and viral transduction in inducing tdTomato knockout by CRISPR editing
Percentage tdTomato positivity **48h (nucleofection) or 7 days (transduction)** after introduction of the CRISPR machinery in **A)** Total cells **B)** Lin- and **C)** HSC compartment. N=5-6 (of 2-3 biological replicates per condition). Data represent mean with standard deviation. Significance calculated using one-way ANOVA.

3.1.8 Summary

The overall aim of this chapter was to optimise a CRISPR editing method in mouse primary HSCs that was efficient enough to be used to dissect the molecular and functional effects of genome editing on this cell population. To this end, in the inducible Ai9-Rosa26 and constitutive Rosa26-tdTomato mouse

strains, a range of modes and formats of CRISPR delivery were tested. The most efficient approach was tdTomato knockout using nucleofection of RNPs. Using this mode and format of delivery led to >90% tdTomato knockout in the total c-Kit enriched and Lin- compartments, and 50% tdTomato knockout in the HSC compartment (**Figure 3.18 A-C)**. This was significantly more efficient than both plasmid nucleofection and viral transduction of the same guide ($p<0.0001$ for c-Kit and Lin-, $p=0.0513$ for HSC). It was predicted that the lower level of editing seen in the HSCs could be due to relative quiescence of this group compared to c-Kit enriched and Lin- populations, meaning protein turnover was slower. Hence, it could be that the time-point chosen to assess tdTomato expression in this compartment wasn't a true reflection of the editing level due to retention of the wild type translated protein for longer. On the other hand, it could be due to this population being more resistant to genome editing. This question is addressed in the second results chapter under *3.2.2 RNP nucleofection editing time course*.

Detailed assessment of all modes and formats of CRISPR delivery in both a cell line and primary cells highlighted key advantages and pitfalls of particular approaches, as well as demonstrating the importance of careful optimisation on a cell-by-cell basis. **Figures 3.6** and **3.8** clearly show that the optimal voltages and plasmid amounts were a fine balance between efficiency and viability. **Figure 3.6 D** demonstrates that increasing the voltage up to a certain point increased the GFP level but above a certain voltage, the viability as well as GFP positivity decreased ($p=0.015$). Similarly, increasing the plasmid amount had a positive linear relationship with transfection efficiency up to a point at which it plateaued before becoming a negative linear relationship (**Figure 3.8 D,** $p<0.0001$). It is at these turning points that the voltage and plasmid amount became too toxic to the

cell and therefore efficiency dropped. What is also clear from the results in Ba/F3 cells is that, at lower plasmid amounts, nucleofection was more efficient than electroporation (**Figure 3.10 D**). This can be explained by the fact that the nucleofection buffer delivers the construct directly to the nucleus rather than relying on the host cell transport system. This was also observed in the mouse BM HSPCs in which nucleofection was significantly more efficient than electroporation (**Figure 3.13 I,** $p<0.0001$).

What is quite striking from the results in mouse BM HSPCs is that in principle all three modes of delivery of genome editing constructs were feasible, despite clear discrepancies in editing efficiency. From **Figures 3.14, 3.15, 3.18 and 3.20**, it is clear that under control conditions, in which small (4kb and 9kb for plasmid and viral vectors respectively) were delivered, high transfection and transduction levels could be achieved, comparable to the editing level seen for tdTomato knockout using RNPs (90% for transfection in HSCs, **Figure 3.14 C**; 85% for transduction in HSCs, **Figure 3.20 C ii**). However, these methods were constrained by the fact that, in order to deliver the Cas9 sequence with its specific promoter together with the guide RNA sequence(s) with their specific promoter(s) and a fluorescence reporter, they were much larger than control vectors. This significantly increased the cell toxicity as well as reducing efficiency compared to RNPs, particularly in primary cells. For the CRISPR Cas-9 vectors, the editing level for tdTomato knockout was 30-40% in the c-Kit enriched (**Figure 3.23 A**), in both plasmid transfection and viral transduction, compared to 90% for RNP transfection, using the same guide RNA. In order to enrich for edited cells using these sub-optimal approaches would require sorting based on fluorescence with additional culture which risks compromising HSC function as they tend to

differentiate rapidly in culture although methods have been optimised to maintain HSCs for longer in culture (Wilkinson, Ishida et al. 2020). The advantage of the RNP approach is therefore that it doesn't require the host cell to express the Cas9 and guide RNA as it is delivered pre-complexed. This makes RNP constructs much smaller as the DNA sequences of the Cas9 and guide are not required. The level of viability compared to plasmid transfection clearly demonstrated the reduced toxicity of this approach, which could also be attributed to the fact that RNPs are more rapidly degraded once inside the cell. The disadvantage of this approach is that, without direct labelling of the Cas9 or guide RNA with a fluorescent reporter, it is not possible to detect successfully transfected cells, therefore the editing efficiency needs to be high enough that the majority of the bulk cells represent an edited population for downstream analysis.

What was also clear when comparing tdTomato induction with knockout is that the feasibility of the approach also greatly depends on the desired outcome and the proposed downstream applications. In the case of this project, 90-100% editing was required to be able to assess the molecular signature of edited cells at the bulk level as well as establish a high level of engraftment from which long-term edited HSCs could be sorted post-transplant. On the other hand, a CRISPR approach in which the edit would give the cells a growth advantage, and thus lead to expansion of the edited population, for example in leukaemia disease modelling, would likely be amenable to a much lower initial editing efficiency, similar to that seen in the tdTomato induction approach.

In conclusion, this chapter has established that although both CRISPR approaches (tdTomato induction and knockdown) can be successfully applied to primary mouse HSCs, a high enough level of editing efficiency was only seen for

the knockdown approach. Further, although in principle all modes and formats of CRISPR delivery were feasible under control conditions, the desired editing efficiency was only seen in the RNP nucleofection approach. Therefore it was decided to use the RNP nucleofection approach for molecular and functional HSC characterisation experiments.

3.2 Molecular and functional characterisation of CRISPR-edited mouse HSCs

3.2.1 Experimental approach

The aim of this chapter was to dissect the molecular and functional consequences of the genome editing process on mouse HSCs. Given the application of the CRISPR method for genome correction at the level of HSCs, it is critical that we understand how this process affects HSCs in a context where the haematopoietic gene network itself is not targeted. This is because if editing at the level of the HSCs is used to treat genetic aberrations in haematological disorders, the edited HSCs, following transplantation, will contribute to blood production in all cell lineages. For example, in the context of beta thalassaemia, the lineage affected by the mutation is the erythroid lineage even though correction must occur at the level of the HSCs in order to ensure life-long correction of the genetic defect.

To do this I chose to target the fluorescent reporter tdTomato. This allowed me to separate the genome editing process itself from any changes in HSC function that may come from the particular gene targeted; and to track the edited cells both *in vitro* and *in vivo.* As detailed in *Methods section 2.3.1 Sample selection*, in depth molecular and functional profiling of edited HSCs was carried out in comparison with sham-edited and unedited controls. Controls were designed such that each stage of the transfection and editing process could be separated and its individual effect on HSCs monitored. In order to separate the effect of culturing the cells *ex vivo*, both a fresh sample, taken directly from the mouse BM

and a 48hr culture sample, which was cultured *ex vivo* for the same length of time and in the same medium as the transfected samples were included. A sample transfected with water only, in the absence of Cas9 or a guide RNA was used to determine the specific consequences of the transfection process. A sample transfected with a non-targeting guide, complementary to other the loxP STOP cassette or the Drosophila Gal-4 gene, neither of which sequences are present in the CD45.2 wildtype mouse strain used for controls, was used to determine the consequence of the presence of Cas9 and guide RNA without DSB induction. This is due to the fact that these guides have no complementarity to any region of the genome in the mouse strain used for controls therefore should not have directed Cas9 to any genomic sequence hence no cutting should have occurred. Finally, a tdTomato targeting guide was used to determine the specific consequences of DSB induction.

In order to assess whether the efficiency of editing varied within the HSPC compartment, a time course was carried out monitoring loss of tdTomato at regular intervals for three days post-nucleofection. The composition of the HSPC compartment was also investigated at regular intervals post-transfection to determine whether the proportion of each primitive cell type was altered by any stage of the genome editing process. Cell cycle analysis was carried out to determine whether, in the short-term, genome editing altered the balance of quiescence and recruitment into the cell cycle of primitive compartments. RNA-Seq of bulk-sorted HSCs was carried out 24 hours post-transfection to assess the short-term effect on the transcriptional profile and HSC signature of edited cells. Finally, in order to functionally profile edited cells, competitive transplantation was carried out in Rosa26-tdTomato recipients and lineage

output was assessed at regular intervals for 18 weeks post-transplantation. Bone marrow and spleen was analysed at the terminal readout for lineage and stem cell engraftment. RNA-Seq on bulk-sorted BM HSCs was carried out 18 weeks post-transplant to assess the long-term transcriptional profile and HSC signature of edited cells compared to the 24-hour transcriptional profile to determine whether the alterations were retained long-term. **Figure 3.24** summarises the experimental workflow.

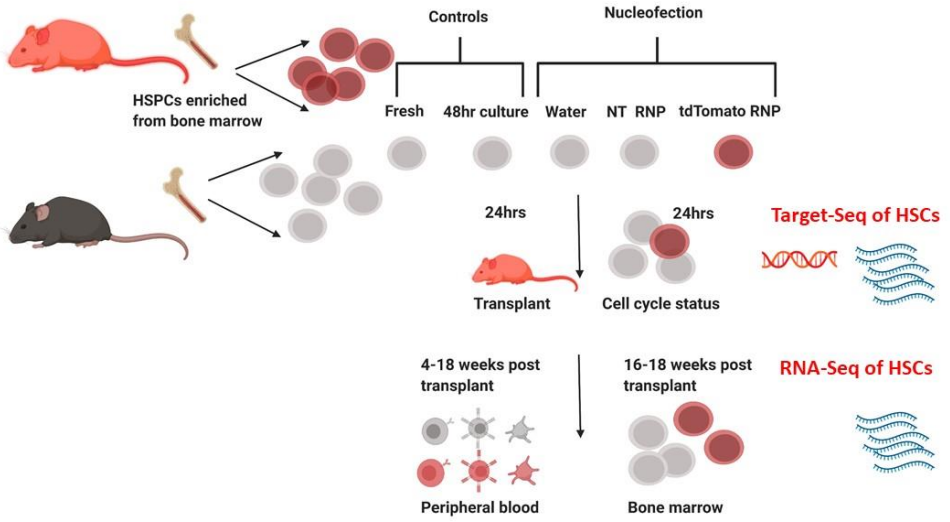


Figure 3.24: Experimental workflow for dissecting the short- and long-term molecular and functional consequences of CRISPR-editing on mouse HSCs

3.2.2 RNP nucleofection editing time course

i) Dynamics of tdTomato knockdown

Given the differences in the level of editing for the most efficient guide (3), in the HSC compartment compared to c-Kit and Lin- at 48 hours, (*Results section 3.1*

Figure 3.23, pg.192), I hypothesised that the quiescence status of most HSCs would mean that their protein turnover, and hence the loss of tdTomato expression, could be delayed in these cells. In addition, I hypothesised that quiescent HSCs might be less amenable to CRISPR-mediated editing. Therefore, as a first step, I performed a time course to assess the dynamics of editing within the mouse BM HSPC compartment using the optimal mode and format of delivery: RNP nucleofection. **Figure 3.25** shows the gating strategy and **Figures 3.26-3.29** summarise the results.

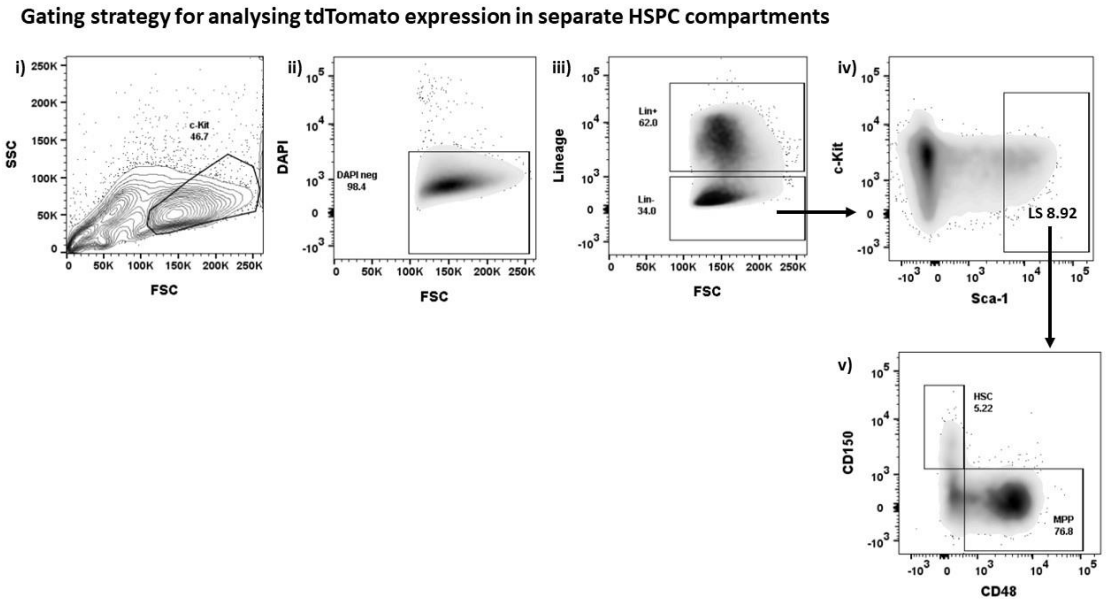


Figure 3.25: Gating strategy for analysing tdTomato expression in each HSPC compartment i-v) HSPC compartments gating strategy using a tdTomato RNP targeted sample. LS defined as Lin- Sca-1+, MPP as Lin- Sca-1+ CD48+ CD150-, HSC defined as Lin- Sca-1+ CD48- CD150+.

As shown in **Figures 3.26-3.29**, the trajectory of tdTomato loss was dynamic, differing between each compartment, compared to no expression loss in non-targeting RNP (NT RNP) controls. Loss in the total viable cells, Lin- and Lin+ happened at the same rate and was almost complete at 48 hours, whereas the

LS, MPP, and HSC compartments were delayed in their downregulation of tdTomato expression. This was most marked in the HSC compartment.

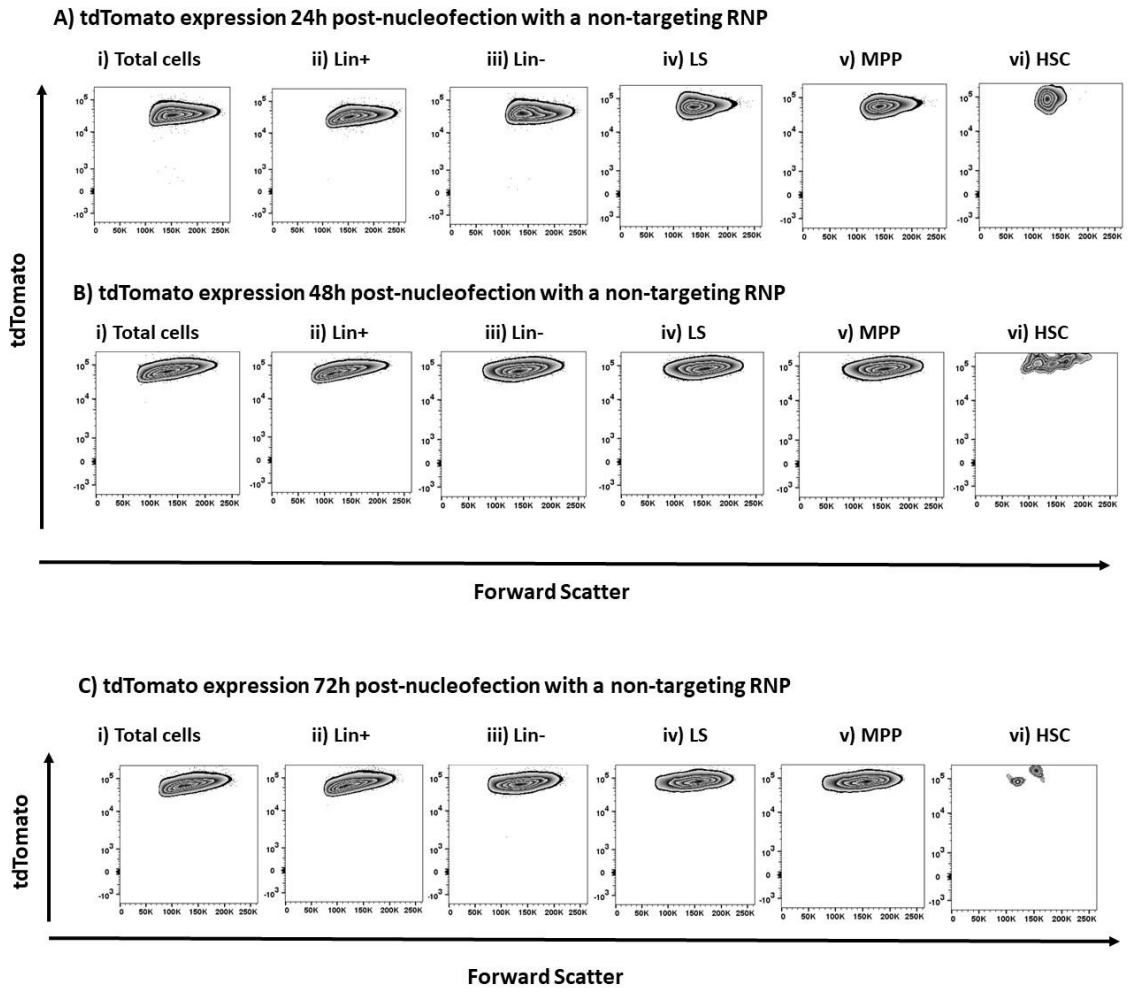


Figure 3.26: No tdTomato expression loss was seen in the NT RNP group Representative flow cytometry plots for tdTomato expression **A)** 24h post-nucleofection **B)** 48h post-nucleofection **C)** 72h post-nucleofection with non-targeting RNP in **i)** total live cells **ii)** Lin+ **iii)** Lin- **iv)** LS **v)** MPP **vi)** HSC compartments.

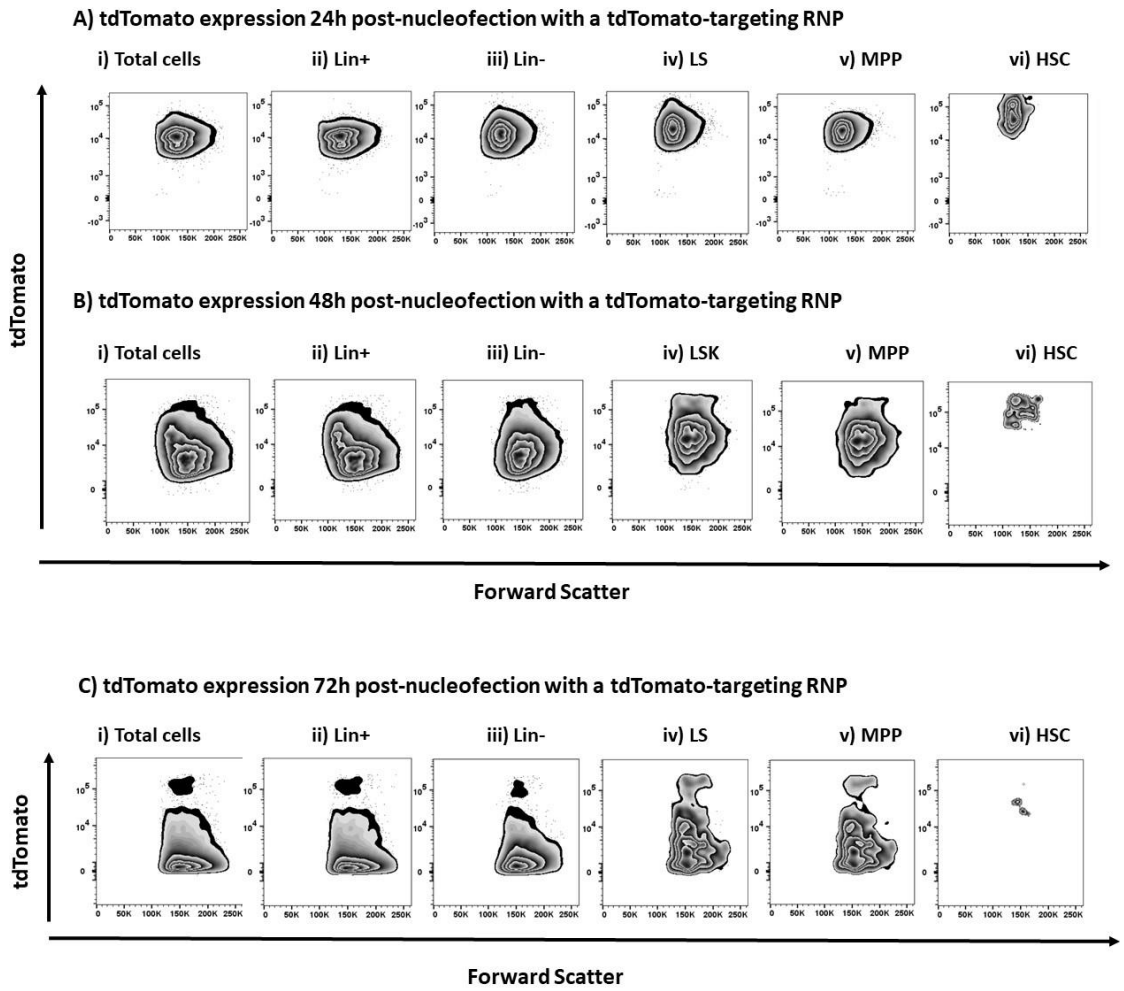


Figure 3.27: Loss of tdTomato expression was delayed in HSCs compared to more mature cell populations Representative flow cytometry plots for tdTomato expression **A)** 24h post-nucleofection **B)** 48h post-nucleofection **C)** 72h post-nucleofection with tdTomato targeting RNP in **i)** total live cells **ii)** Lin+ **iii)** Lin- **iv)** LS **v)** MPP **vi)** HSC compartments.

The time course of tdTomato loss was consistent between replicates and within compartments (**Figure 3.28**). Loss of tdTomato expression after editing had a slower trajectory in MPPs and HSCs (**Figure 3.29**), most likely as these populations reside in a more quiescent state hence protein turnover was slower. It is likely therefore that for some edited HSCs 72 hours was too early to have seen complete protein loss.

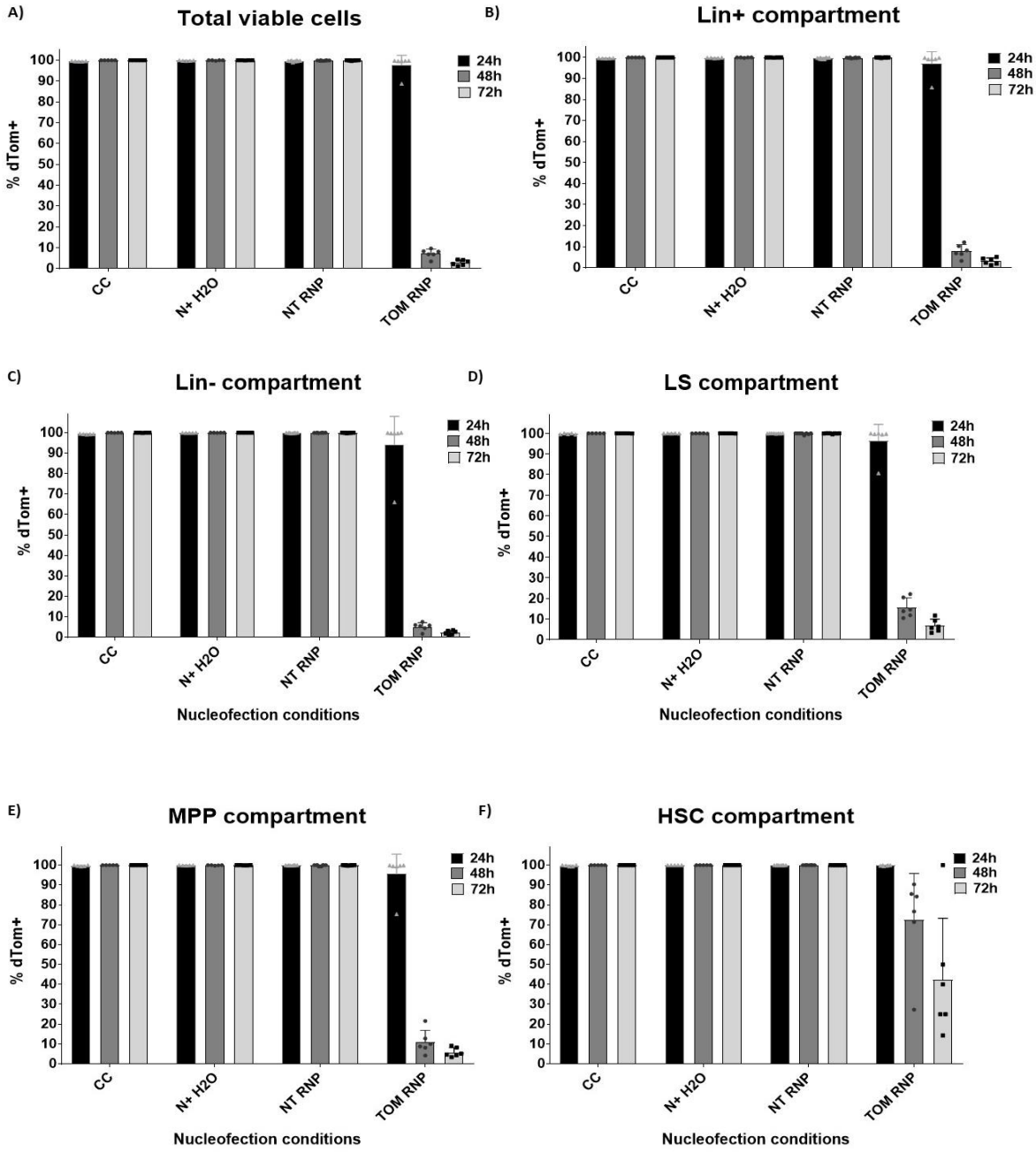


Figure 3.28: tdTomato knockdown was highly efficient in all haematopoietic compartments although delayed in HSCs Percentage tdTomato positive cells 24-72h post-nucleofection in **A)** total viable cells **B)** Lin+ **C)** Lin- **D)** LS **E)** MPP **F)** HSC compartments. N= 5-8 (3 biological replicates with 2-3 technical replicates). Data represent mean with standard deviation.

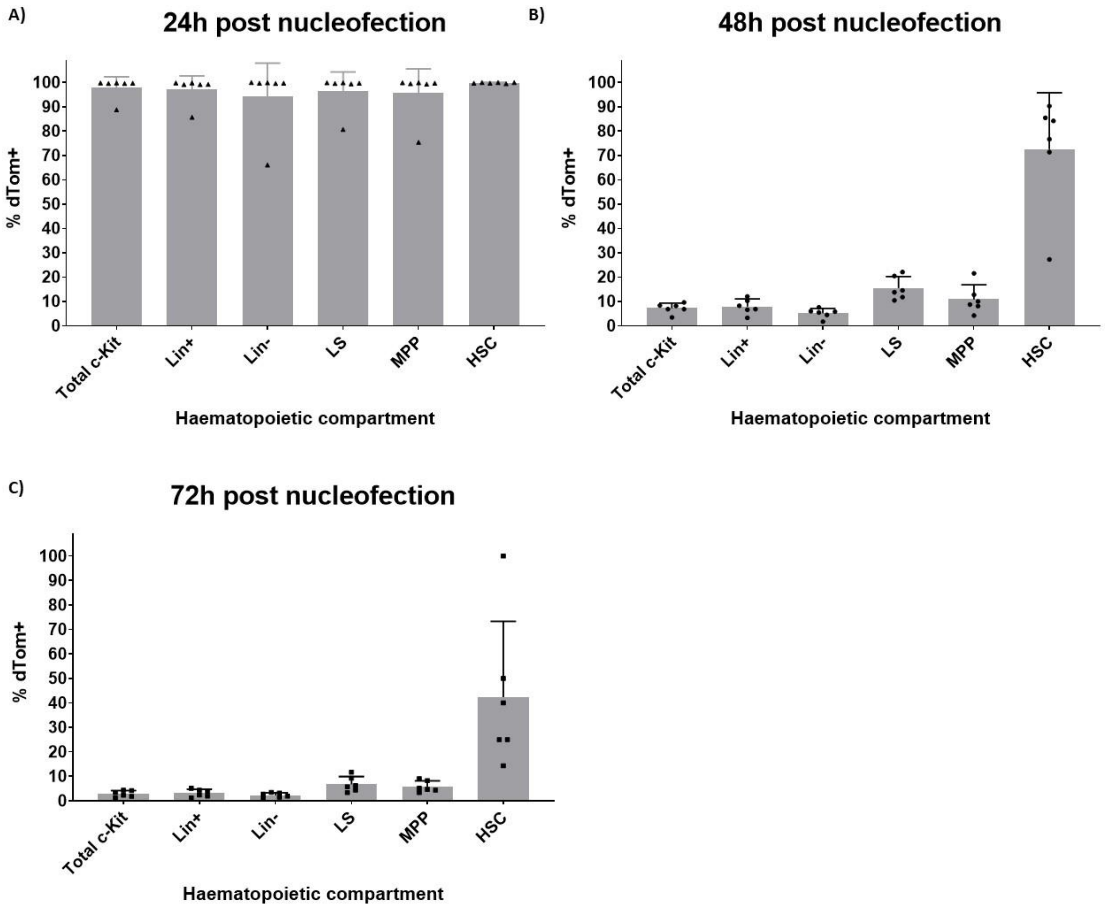


Figure 3.29: tdTomato RNP targeted HSPC compartment editing summary
tdTomato editing summary for TOM RNP samples **A)** 24h **B)** 48h **C)** 72h post-nucleofection.

ii) Cellular composition over the 72h time course

In order to determine whether editing affected the proportion of cells falling within the primitive HSPC gates, the percentage of each compartment contributing to the previous population from which it was gated (parental flow cytometry gate) was calculated for each time-point measured. **Figure 3.30** summarises the results. For the most part, the composition of c-Kit enriched BM, determined by the proportion of parental cells gated for Lin+, Lin-, LS, MPP, and HSC, was unchanged between edited, sham-edited, mock nucleofected, and culture control samples. The proportion of total cells falling in the Lin- and Lin+ gates over the

72 hour time course remained consistent and similar across the groups. Within the LS fraction, the proportion of MPPs remained consistent post-editing and similar across the groups. The proportion of HSCs within the LS fraction diminished over time across all groups at equal levels. This indicates that any DNA damage induced apoptosis did not occur more readily within specific compartments, suggesting more primitive cell types were not more sensitive to CRISPR targeting than others. However, the proportion of LS cells within the Lin- fraction consistently increased across the control groups but decreased in the edited group (**Figure 3.30 C**).

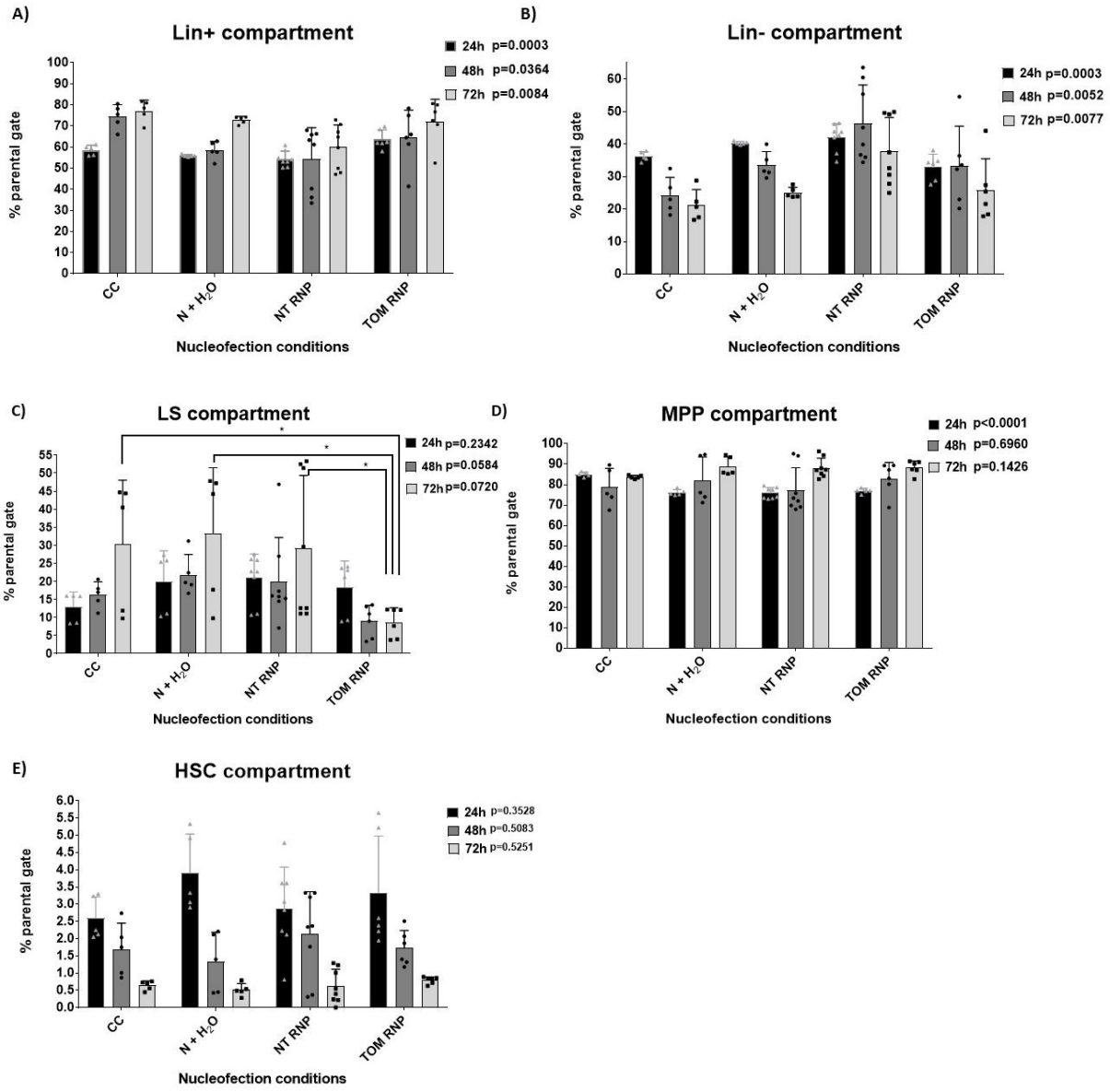


Figure 3.30: The proportion of cells in the LS compartment was significantly reduced in tdTomato edited samples compared to controls
Percentage of parental flow cytometry gate contributing to **A)** Lin+ **B)** Lin- **C)** LS **D)** MPP **E)** HSC compartments at 24h, 48h and 72h post-nucleofection. N= 5-8 (3 biological replicates with 2-3 technical replicates). Data represent mean with standard deviation. Significance at each time-point calculated using one-way ANOVA. Significance of LSK compartment at 72h calculated using a paired two-tail t test.

The difference in proportion of LS cells remaining at 72 hours was statistically significant. This difference could have resulted for several reasons; one explanation is that the HSC and MPP populations (contributing to the majority of

the LS gate) were both equally selectively reduced by DSB induced apoptosis as a result of higher sensitivity to CRISPR editing. This would explain why the proportion of MPPs and HSCs in the LS gate remained similar to controls while the LS gate was reduced. Another possibility is that edited HSCs had altered cell cycle status, in which more resided in G_0 and were therefore producing less daughter cells leading to a reduction in the HSC and MPP pool. Finally, tdTomato-targeted MPPs and HSCs may have changed their pattern of division in favour of asymmetric/symmetric division generating more lineage committed populations as opposed to self-propagation divisions which would reduce the total HSC/MPP pool. This posed the question as to whether edited HSCs would behave differently compared to controls in an *in vivo* setting if relative levels of LS populations were reduced.

iii) Cell cycle status of edited cells 24 hours after nucleofection

Due to the reduced proportion of LS cells within the tdTomato edited group, it was predicted that one possibility was that the cell cycle status between the tdTomato edited and control groups might differ, as well as the cell cycle status between different HSPC compartments within the same group. To assess this, cells were pre-stained 24 hours post-nucleofection with markers segregating primitive compartments followed by fixation and permeabilisation to stain with DAPI and Ki-67 (*Methods section 2.3.3)*. **Figures 3.31** and **3.32** summarise the results.

In each compartment analysed, there were no significant differences in the proportion of cells at each stage (G_0, G_1, SG_2M) of the cell cycle (**Figure 3.32**)

suggesting that in the short-term, 24 hours post-nucleofection, CRISPR genome editing did not significantly alter the proliferation of targeted cells. These data also suggest that editing did not cause selective apoptosis of cells at a particular stage of the cell cycle, as the proportion of cells at a particular cell cycle stage more susceptible to genome editing induced apoptosis would be reduced in the tdTomato edited samples compared to controls. Apoptosis staining, for example by Annexin V, however, would be required to confirm this observation.

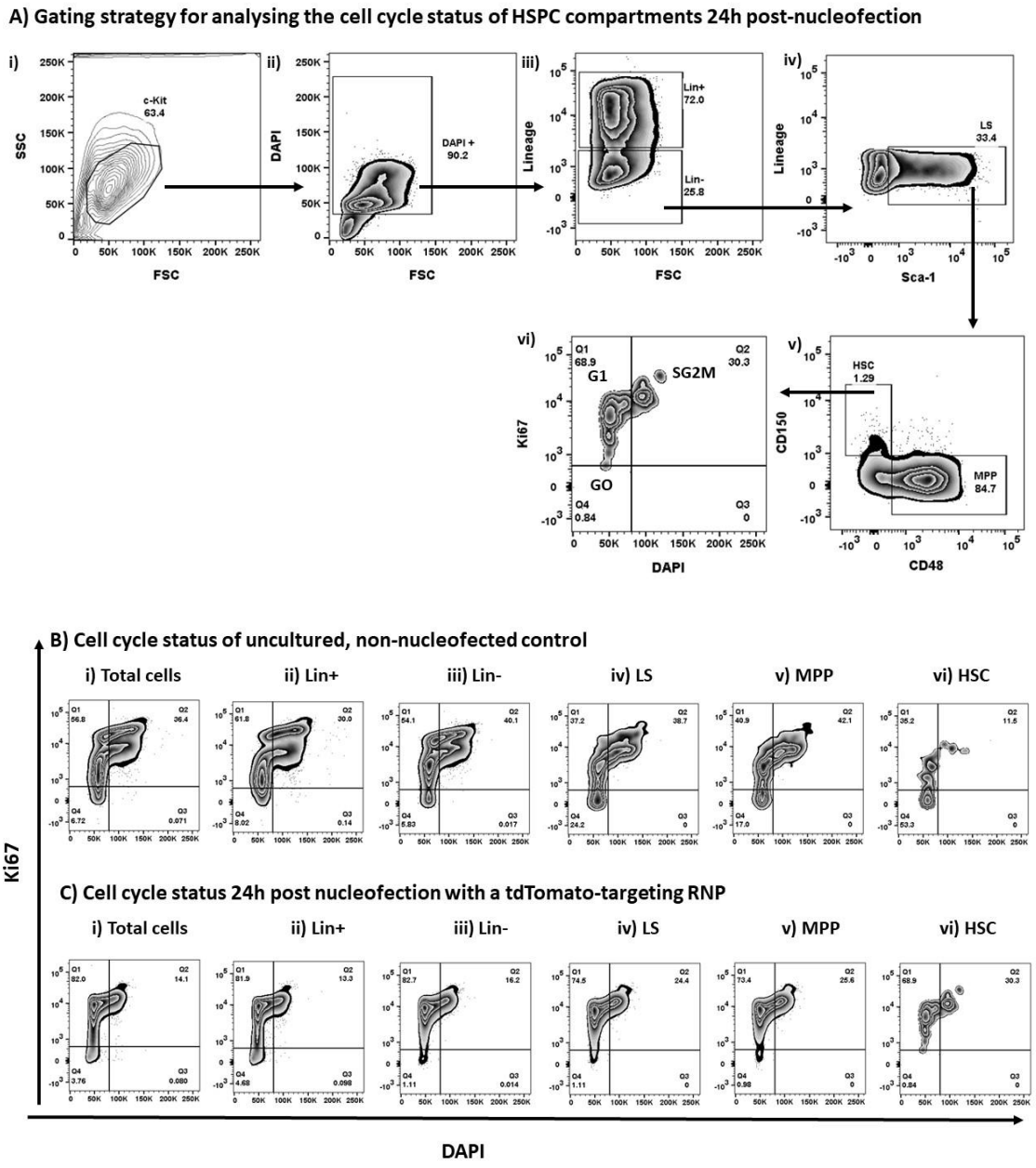


Figure 3.31: Gating strategy for separating HSPC compartments for cell cycle analysis A) i)-v) Gating strategy for segregating Lin+, Lin-, LS, MPP, HSC for cell cycle analysis **vi)** using DAPI and Ki67. Representative cell cycle plots for **B)** Uncultured control sample **C)** TOM RNP pre-gated on **i)** Total viable cells **ii)** Lin+ **iii)** Lin- **iv)** LS **v)** MPP **vi)** HSC.

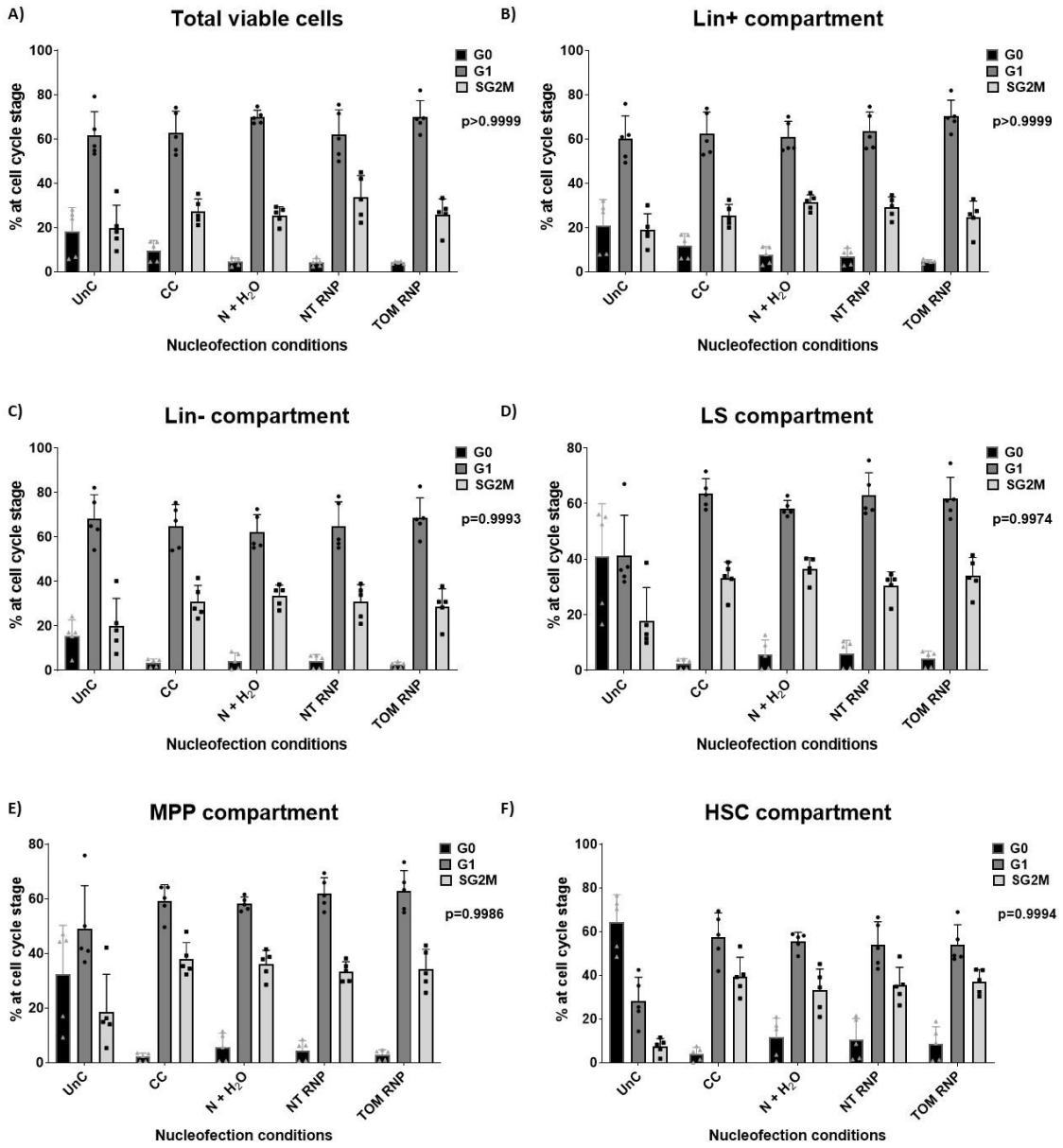


Figure 3.32: The cell cycle status of primitive HSPC compartments was not significantly altered by tdTomato targeted gene editing Percentage of cells in G_0, G_1, SG_2M 24h post-nucleofection in **A)** Total viable **B)** Lin+ **C)** Lin- **D)** LS **E)** MPP **F)** HSC compartments. N=5 (3 biological replicates, 1-2 technical replicates). Data represent mean with standard deviation. Significance calculated using column factor of two-way ANOVA. **NB.** Uncultured (UnC) samples were not included in the statistical analysis but used as a control for the expected G_0 status of HSCs.

3.2.3 Transcriptomic analysis of the HSC compartment 24h after nucleofection

i) Targeted genotyping of the tdTomato locus

As detailed in *Methods Section 2.3*, viable HSCs were stringently sorted in mini bulk (100-220 cells per replicate) 24 hours after nucleofection and subjected to Target-Seq to amplify the tdTomato targeted region and assess the transcriptome in the same sample. Genomic DNA amplified at the tdTomato region was subject to Sanger sequencing (*Methods section 2.3.8*) to confirm the editing level. This was due to the evidence that tdTomato expression was not lost at 24 hours therefore could not be used in the sorting scheme to segregate edited cells. 250,000 c-Kit enriched mouse BM cells were kept in the well from which cells were taken for Target-Seq sorting and assessed 48 and 72 hours post-nucleofection. tdTomato locus Sanger sequencing results at 24 hours were correlated with tdTomato expression in the total viable cells at 72 hours to select samples with a high enough editing level to proceed for transcriptomic analysis. A cut off was set at 65% editing at 24 hours by Sanger in HSCs and >95% tdTomato loss in the total viable cells at 72 hours in order to ensure the majority of transcriptomic signal in the bulk samples originated from edited cells. Cosmid was used to predict potential off-target sites of the tdTomato gRNA. Representative samples were amplified at the top three off-targets and subject to Sanger sequencing to determine the guide specificity (Cradick, Qiu et al. 2014). **Figures 3.33-3.37** summarise the on-target/off-target Sanger sequencing results and correlation with 72-hour flow cytometry results.

tdTomato editing in all TOM RNP samples was clear from the agarose gel (**Figure 3.33**) in which the band at the expected site was distinctly fainter than respective control bands. Equal concentrations of genomic DNA were used and therefore this was not a result of differing amounts of starting material. This result indicates an editing course resulting in loss of primer binding sites.

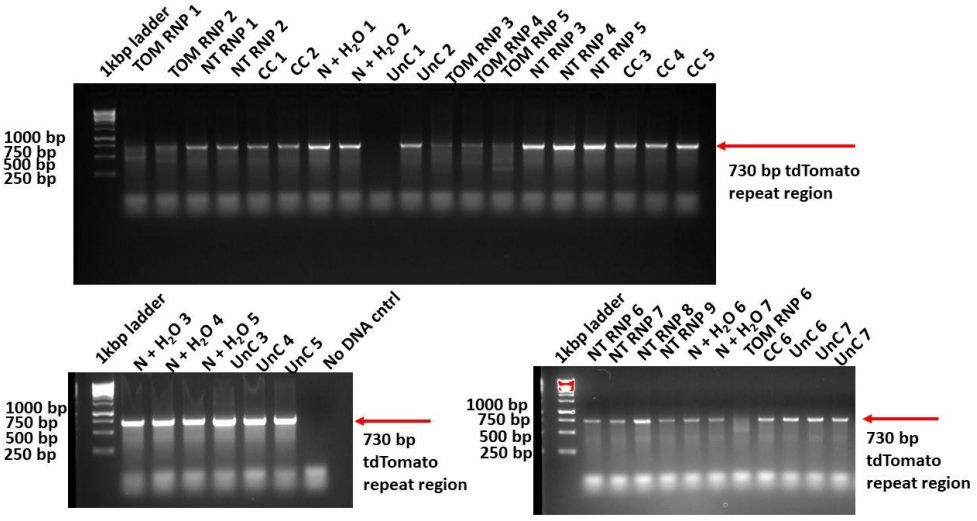


Figure 3.33: Wild type tdTomato sequence was less detectable by genomic PCR 24h post-nucleofection Agarose gel images of the tdTomato amplified region processed for Sanger sequencing.

For the faint band that remained, ICE analysis demonstrated high levels of editing (65-87%) for cells in which the primer site remained intact (**Figure 3.34**).

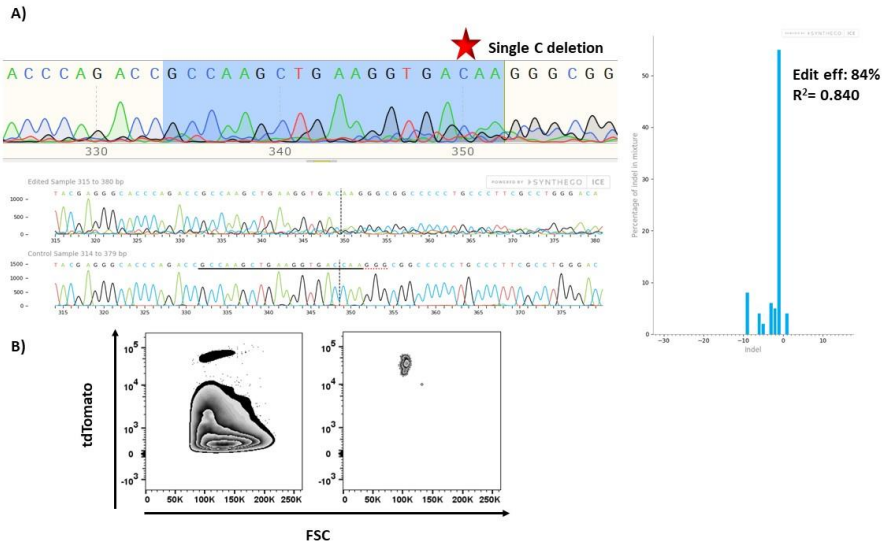


Figure 3.34: tdTomato HSC editing at 24h correlated with 72h tdTomato expression-tdTomato RNP A) Sanger trace with tdTomato knockout guide RNA 3 sequence highlighted and Synthego alignment to control sequence (left panel), percentage indels computed using Synthego ICE tool (right panel) **B)** tdTomato expression 72 post-nucleofection in total viable cells (left panel) and HSC compartment (right panel).

There was no evidence of editing at the tdTomato locus in the NT RNP control (**Figure 3.35**).

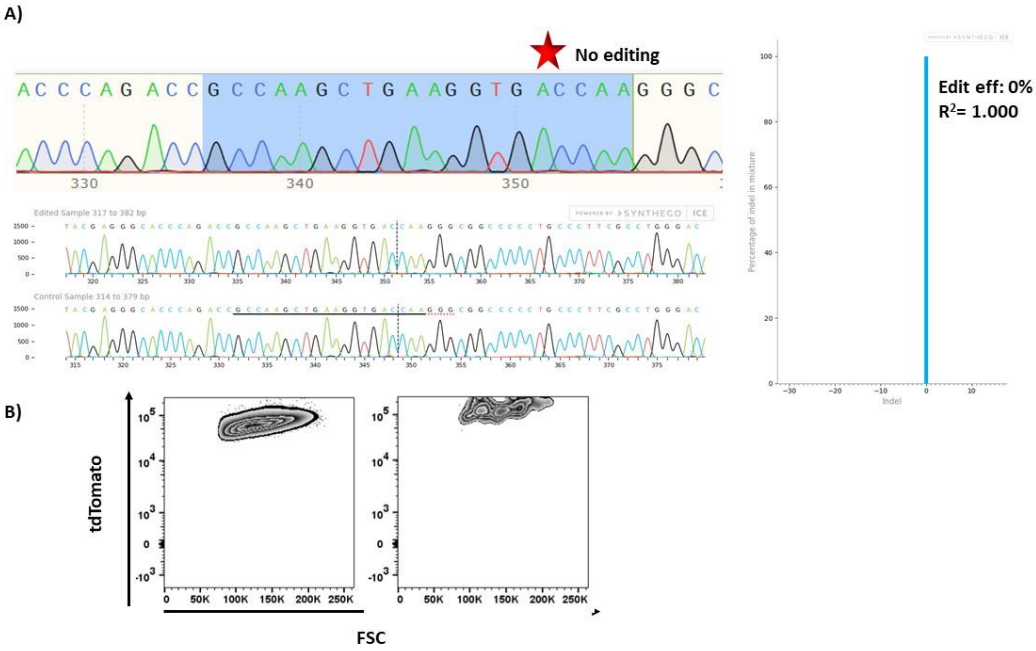


Figure 3.35: tdTomato HSC editing at 24h HSC correlated with 72 tdTomato expression-NT RNP A) Sanger trace with tdTomato knockout guide RNA sequence highlighted and Synthego alignment to control sequence (left panel), percentage indels computed using Synthego ICE tool (right panel) **B)** tdTomato expression 72 post-nucleofection in total viable cells (left panel) and HSC compartment (right panel).

For the two samples which ICE predicted to have low editing levels but showed high levels of tdTomato expression loss at 72h (>95% in total c-Kit), the Sanger trace demonstrated multiple sequences present. This resulted in a lack of successful alignment with the control sequence and a low R score (**Figure 3.36**), meaning the indel prediction was not highly accurate. What is clear from the agarose gel images and Sanger traces is that the resulting indels caused by editing were heterogeneous making predicting the specific editing landscape at the bulk level challenging. However, for the samples in which a dominant edit led to more precise ICE predictions, these correlated strongly with tdTomato expression at 72 hours by flow. tdTomato expression was subsequently

interrogated at the transcriptome level to ensure only samples with high levels of tdTomato transcript loss were processed for analysis.

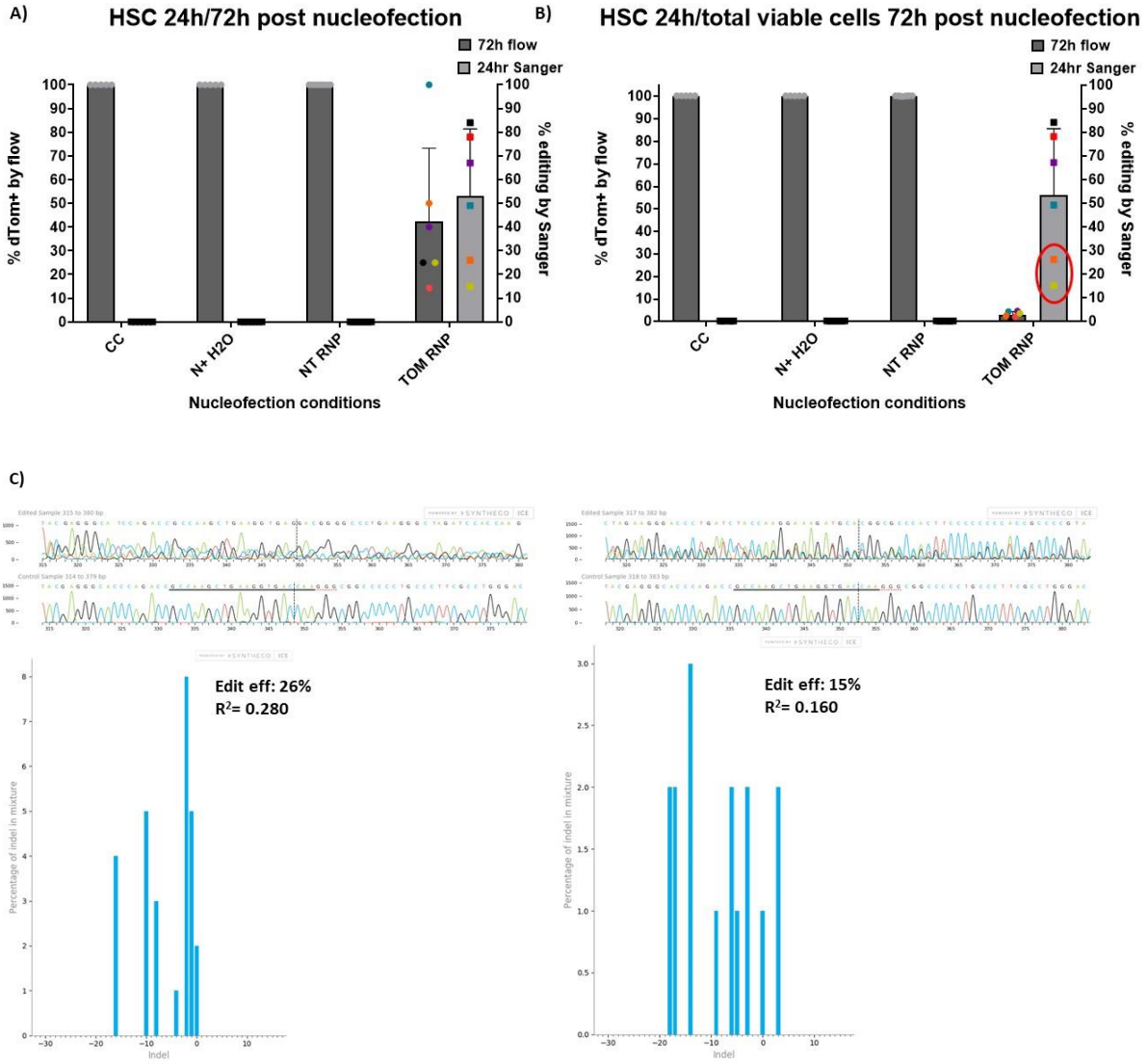


Figure 3.36: Low Synthego ICE R score lead to underestimation of tdTomato expression loss Matched tdTomato expression at 72h (left y axis) and HSC tdTomato editing level at 24h (right axis) for **A)** total viable cells and **B)** HSC compartment. N= 5-8 (3 biological replicates, 1-3 technical replicates). Samples in which Synthego ICE score doesn't correlate with 72 expression are circled in red, **C)** Sanger trace for tdTomato amplified region and ICE analysis for non-correlating samples with low R score.

Further, interrogation of the top off-targets indicated the tdTomato gRNA chosen to be highly specific with only 19% editing at the top Cosmid predicted site, and 0% editing at the second two top off-targets (**Figure 3.37**).

A)

Search result	Query type	Mismatch	Ends with RG	Chr Position	Strand	Cut site	Score
hit: CACAAGCTGAAGGTGACCAAGGG	No indel	2	Yes	Chr7:65070254-65070276	+	65070270	0.25
hit: GAAGAGCTGAAGGTGACCAATGG	No indel	3	Yes	Chr13:69599588-69599610	+	69599604	0.45
hit: TCAAAGCAGAAGGTGACCAAAGG	No indel	3	Yes	Chr12:91294119-91294141	-	91294125	0.54

Hit 1: FAM 189a1, integral membrane protein, highest expression in areas of brain at early stages of development

Hit 2: Srda1 (steroid 5 alpha reductase 1), highest expression in brain

Hit 3: Cep128 (centrosomal protein 128), highest expression in testes and livers

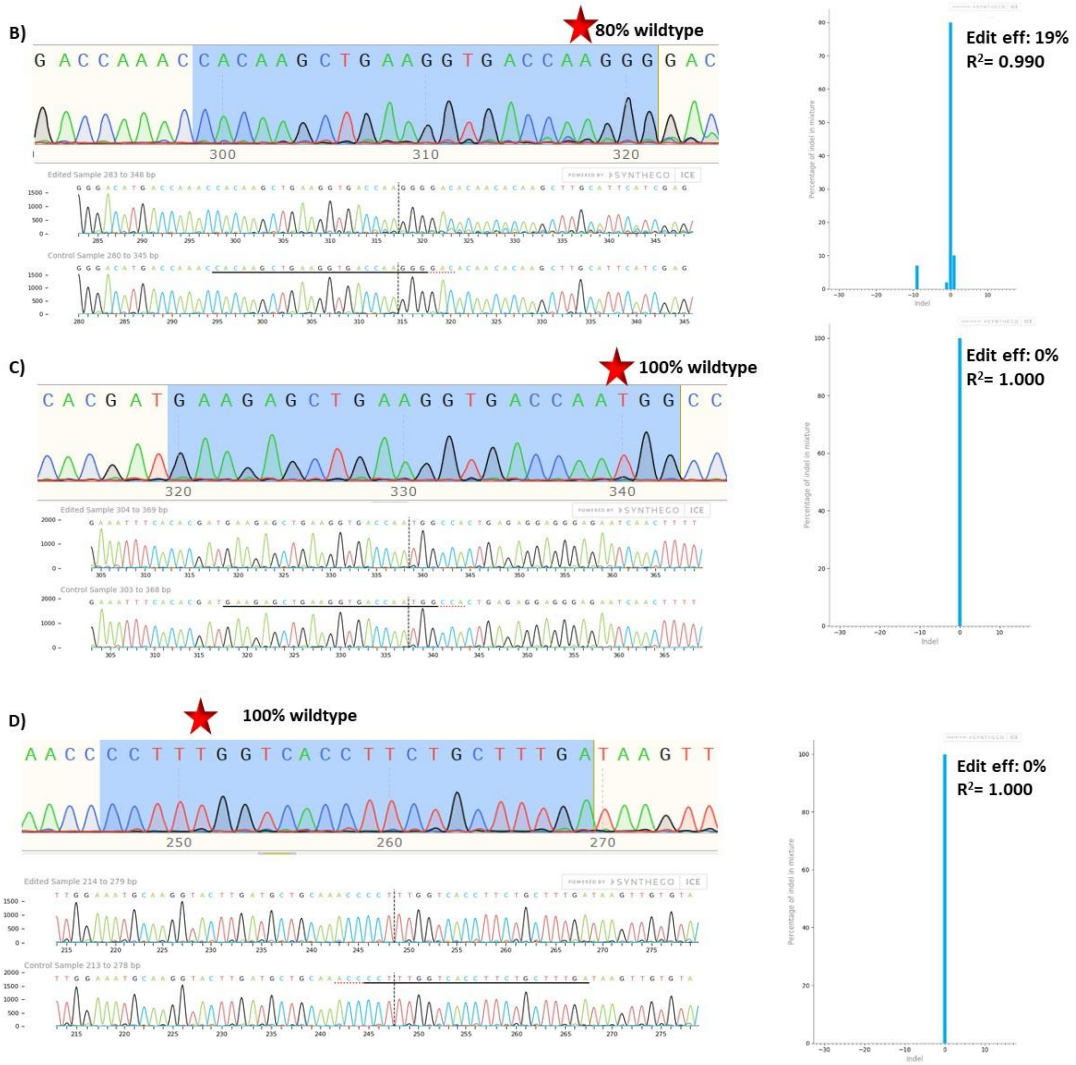


Figure 3.37: Low level off-target editing was present at the top tdTomato off-target site A) Cosmid predicted top 3 off-target genomic sites for tdTomato guide RNA 3. Sanger trace of off-target amplified region, Synthego alignment to control and percentage indels for tdTomato targeted genomic sample amplified 24h post-nucleofection at **B)** off-target 1 **C)** off-target 2 **D)** off-target 3.

ii) Whole transcriptome analysis of edited HSCs 24h post-nucleofection

As detailed in *Methods section 2.3.6*, 100-220 HSCS were stringently sorted on viability 24 hours after nucleofection and amplified using the Target-Seq (Rodriguez-Meira, Buck et al. 2019) method. Transcriptome analysis was kindly performed by Wei Wen, a DPhil student in the lab of Adam Mead. In brief, samples were aligned to the mouse genome and only those with 80% or higher reads mapped were processed for downstream analysis. Details of the alignment scores for individual samples are available in *Appendix 6*. Normalisation was then performed by taking each gene count and dividing by the length of the gene followed by dividing by the total gene counts (library size) of the corresponding sample and finally multiplied by 1,000,000 to obtain gene expression values in transcript per million (TPM). Batch correction was performed using Combat-Seq (Zhang, Parmigiani et al. 2020) as it allows for optimal correction of data which does not follow a normal distribution, as is the case for RNA-Seq. This was due to the fact that samples clustered by the date they were sorted on before batch correction, a common problem with RNA-Seq data (Zhang, Parmigiani et al. 2020). Prior to Combat-Seq, normalisation approaches were commonly used to adjust for batch effects, however these don't always completely remove effects of expression of specific genes in relation to the total counts, which is often variable between batches. Combat-Seq uses a negative binomial regression model which allows the read-count nature of RNA-Seq data to be retained. This allows differential gene expression analysis, which relies on count data for each gene to be successfully carried out with batch corrected samples. As normalisation methods transform the read counts to for example counts per million or reads per kilobase of transcript, they lose the raw count data which

removes potential inter-gene variability compared to total reads between batches. By preserving read counts, Combat-Seq is a superior method for batch correction prior to differential gene expression analysis (Zhang, Parmigiani et al. 2020). tdTomato read counts were assessed to confirm loss of expression in edited samples compared to controls. Differential gene expression analysis was carried out using DESeq2; with the FDR set at <0.05 for all pairwise comparisons. Differentially expressed genes were used to generate the PCA plot. Gene-set enrichment analysis was performed on hallmark gene sets to assess groups enriched across each stage of the editing process; with the FDR set at <0.25. Random forest analysis was performed to further segregate differentially expressed genes between the nucleofection groups. Finally, HSC gene sets for total HSC, LT-HSC and ST-HSC/Progenitors (*Appendix 8* lists all genes contributing to each set) (Ivanova, Dimos et al. 2002) were used to generate a HSC transcriptional score for each group. **Figures 3.38-3.41** summarise the results. **Figure 3.38** demonstrates Combat-Seq was an efficient method for batch correction in this context, with samples no longer clustering by sort batch in part **B**). Principal component analysis identified the fresh group as that which separated furthest from all the other groups, with increasing CRISPR intervention separating the other groups with more subtle transcriptional differences.

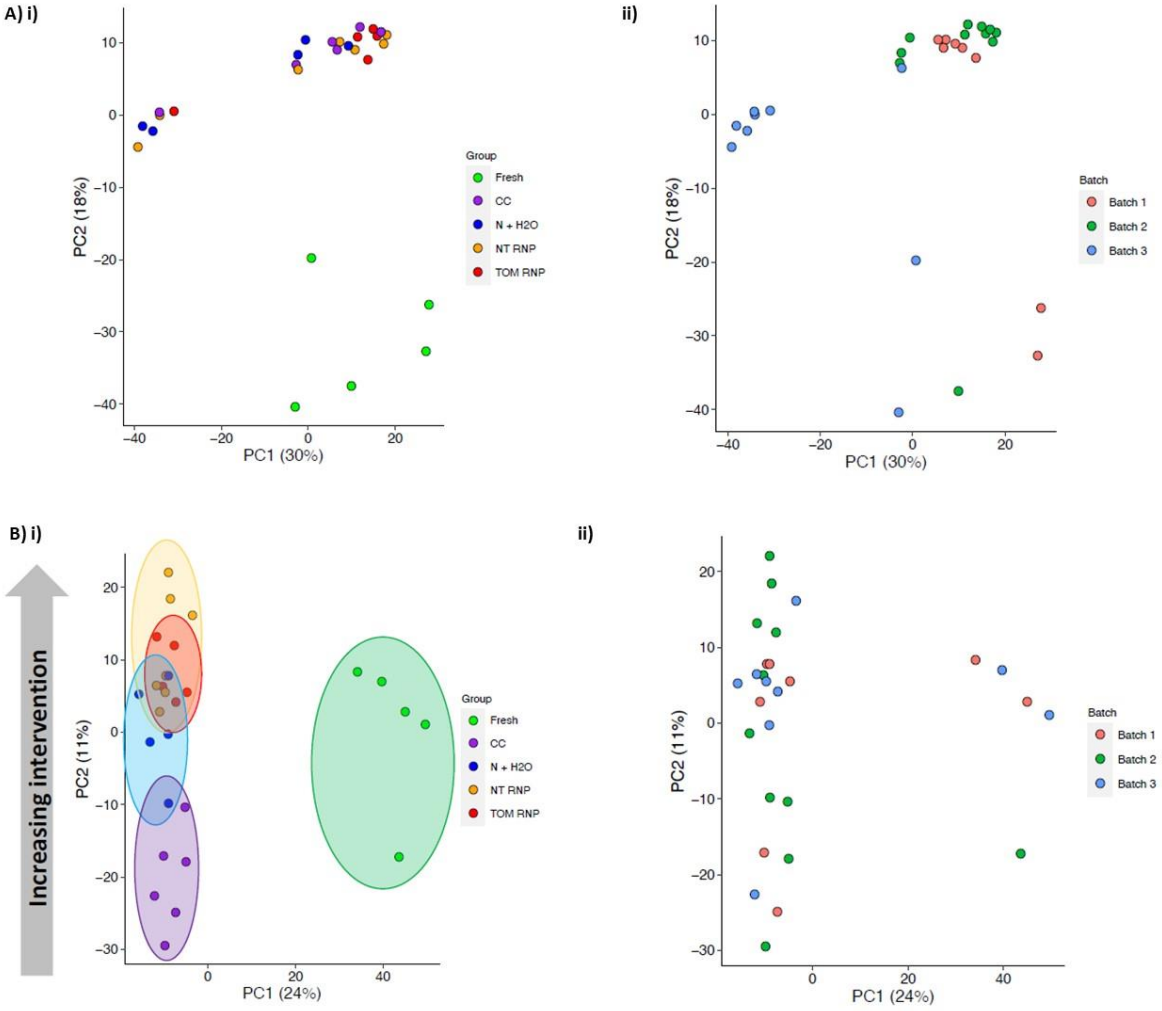


Figure 3.38: Principal component analysis separates CRISPR edited samples both from fresh controls and nucleofection controls with increasing intervention A) Supervised principal component analysis by differential gene expression showing **i)** samples labelled by condition **ii)** samples labelled by batch before batch correction with Combat-Seq. **B)** Principal component analysis showing **i)** samples labelled by condition **ii)** samples labelled by batch post-batch correction with Combat-Seq. N= 5 (fresh), 6 (culture control), 5 (water), 7 (NT RNP), 5 (TOM RNP) biological replicates per group.

Figure 3.39 A confirms that tdTomato expression was diminished at the transcript level, which correlated with both the 24-hour Sanger sequencing and 72-hour flow cytometry data (**Figure 3.36**). Part **B** demonstrates that the greatest difference in gene expression compared to TOM RNP samples was in the fresh

controls, with the least differentially expressed genes between the NT RNP and TOM RNP groups, indicating small transcriptional differences between samples targeted with a non-targeting guide compared to tdTomato targeting.

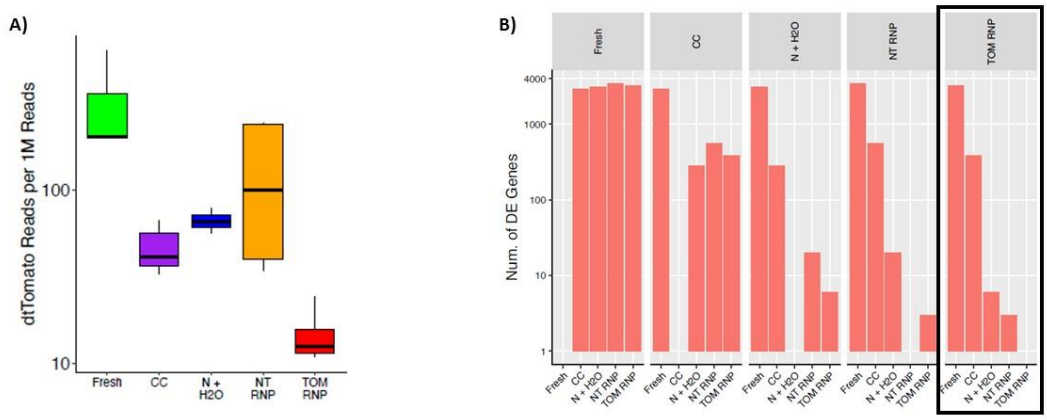


Figure 3.39: tdTomato gene expression was reduced in TOM RNP targeted groups. Differential gene expression between TOM RNP and control samples decreased with increasing CRISPR intervention A) tdTomato read count as a proportion of million reads. **B)** Log10 (no. of differentially expressed genes) for each condition (bottom x-axis) compared to every other condition as a baseline (top x-axis). Differentially expressed genes in all conditions compared to TOM RNP is outlined in black. N= 5 (fresh), 6 (culture control), 5 (water), 7 (NT RNP), 5 (TOM RNP) biological replicates per group.

Figure 3.40 demonstrates that although the level of transcriptional differences in TOM RNP samples compared to control groups varied considerably between control groups, there was a small group of genes in which differential expression overlapped in TOM RNP samples compared to at least two control groups. What is clear is that from culturing the cells *ex vivo,* through nucleofection and genome editing, the differences in gene expression reduced. A large proportion of differentially expressed genes separating the fresh group from the TOM RNP group showed no overlap (green circles) with the cultured group whereas there was a much smaller number of genes up-/down-regulated in the TOM RNP compared exclusively to the cultured group (purple circles). Finally, the number

of genes up-/down-regulated in the TOM RNP group compared to the mock nucleofection and sham-edited (NT RNP) groups was very minimal. It is possible that the transcriptional changes seen in the acute phase post-editing are transient and were not captured at this time-point and in fact an earlier time-point would have revealed greater transcriptional differences. Indeed this 24-hour time-point has also been reported by Schiroli and colleagues to demonstrate subtle changes in gene expression (Schiroli, Conti et al. 2019). Conversely, it could be that transcriptional changes are delayed and don't occur immediately in conjunction with the DSB, meaning a later time-point than 24 hours could have revealed more variation between the nucleofection groups. This will be discussed further in *Conclusions and future directions section 4.6.2*.

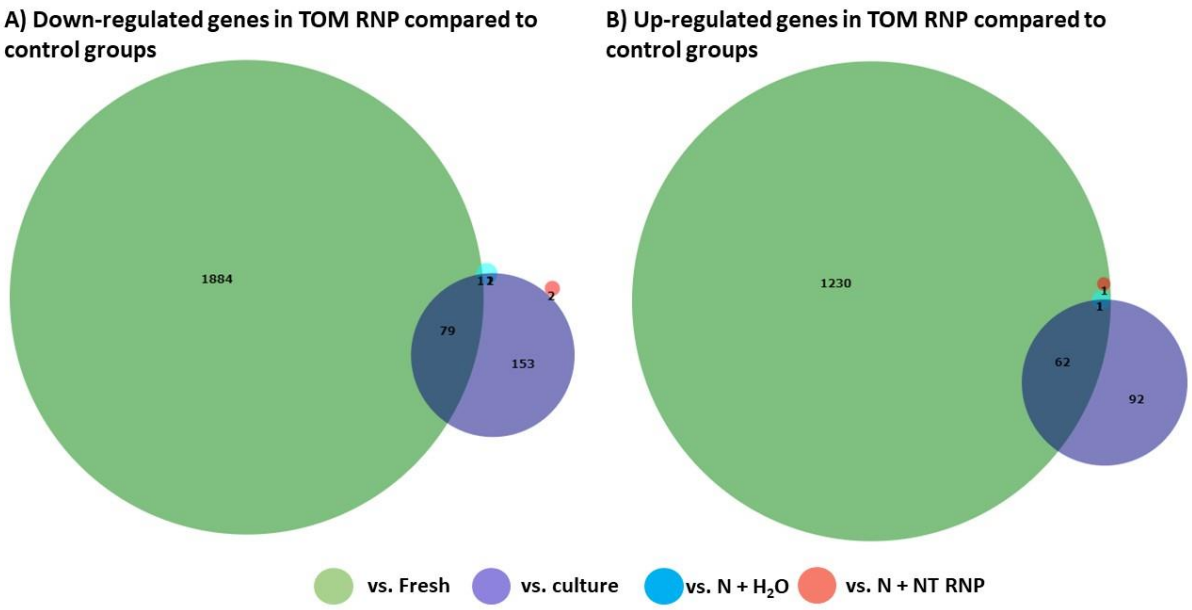


Figure 3.40: Differential gene expression displayed by TOM RNP samples showed some overlap between control conditions Venn diagrams representing overlapping differential gene expression for **A)** Genes down-regulated in TOM RNP targeted samples compared to controls **B)** Genes up-regulated in TOM RNP targeted samples compared to controls. N= 5 (fresh), 6 (culture control), 5 (water), 7 (NT RNP), 5 (TOM RNP) biological replicates per group.

Figure 3.41 illustrates that within the differentially expressed genes in TOM RNP samples, there were specific genes up- or down-regulated compared to multiple controls with potential implications for stem cell function.

A large proportion of overlapping genes up-regulated in TOM RNP samples compared to fresh and cultured controls were related to inflammatory and stress responses (*Anxa1, Anxa2, Ccl9, Gzmb, Plscr1, Rps6ka4*) and increased mitochondrial oxidative phosphorylation (*Ndufaf1, Ndufuz, Sdhb*). In addition, a number of genes involved in negative regulation of the cell cycle were upregulated (*Cdkn2b, Faf1, Noc2l, Scin*) with the exception of *Ccn3*, an important gene promoting stem cell self-renewal (Kimura, Martin et al. 2010).

Of those down-regulated genes showing overlap in TOM RNP samples compared to fresh and cultured controls, *Sult1a1* has been shown to be expressed in long-term re-populating HSCs (Gazit, Mandal et al. 2014) and *Vegfa* has been shown to be essential for HSC maintenance within the hypoxic environment of the BM niche (Gerber, Malik et al. 2002).

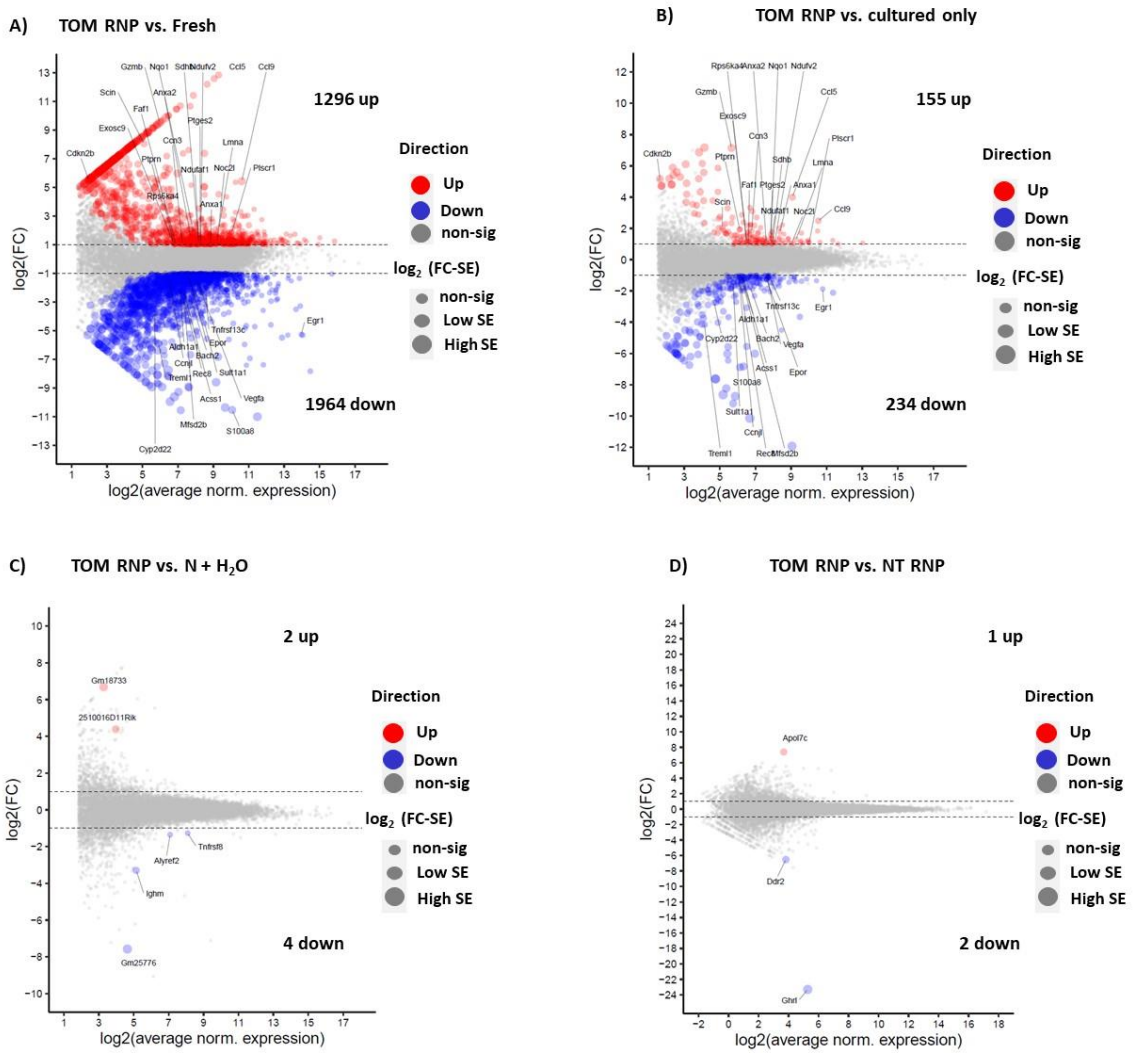


Figure 3.41: Differential gene expression identified transcriptional changes suggesting implications on HSC function Volcano plots showing up- and down- regulated genes in TOM RNP samples compared to **A)** Fresh **B)** Cultured only **C)** Nucleofection + water **D)** Nucleofection + non-targeting RNP. Plots represent log_2 average normalised expression against log_2 fold change. Up-regulated genes are labelled in red, down-regulated genes are labelled in blue. Dot sizes represent level of standard error. Smaller dots have lower standard error and vice versa. N= 5 (fresh), 6 (culture control), 5 (water), 7 (NT RNP), 5 (TOM RNP) biological replicates per group.

Due to the fact that the greatest differences in gene expression were evident when comparing fresh samples to all other samples, two plots were created to

analyse gene set enrichment scores. The first (**Figure 3.42 Ai**) represents gene set enrichment scores for all nucleofection groups compared to fresh control samples. Enriched gene sets included DNA repair. In order to then look more closely at differences within the nucleofection groups, the normalised enrichment scores were scaled for each row compared to every other nucleofection group, which removed the effect of the fresh group (**Figure 3.42 Aii**). This analysis revealed a subtle increase in the *TP53* pathway gene set in groups with the highest intervention (nucleofection + tdTomato editing). Part **B** compares the gene set enrichment level for the *TP53* pathway of each nucleofection group compared to the fresh control. Due to the fact that *TP53* was greatly enriched in fresh samples compared to all nucleofection groups (**Figure 3.42 Ai**), the scores were negative. What is clear, however, is that the TOM RNP group had the least negative score therefore, within the nucleofection sample comparison, this group had the highest *TP53* pathway enrichment.

Interestingly, when comparing the genes in the DNA repair gene set and those up-regulated in the differential gene expression analysis, there was no overlap between the two groups of genes, indicating that as a group the DNA repair gene set were subtly enriched across the nucleofection conditions with increase intervention; however, individual genes weren't enriched significantly on their own to reach the threshold value. Similarly, when looking at the overlap between genes contributing to the TP53 pathway gene set and individual differentially up-regulated genes, only *Cdkn2b* came up in both data sets. However, *Tnfrsf9* was up-regulated in the differential gene expression data, of which the receptor (*Tnfsf9*) is part of the TP53 gene set. Again this shows that with the exception of *Cdkn2b*, individual genes in the TP53 data set didn't reach the threshold of

significance alone but only as a group in gene set enrichment analysis. This is not surprising given the enrichment of the TP53 gene set, even as a group, was subtle across nucleofection with increasing intervention.

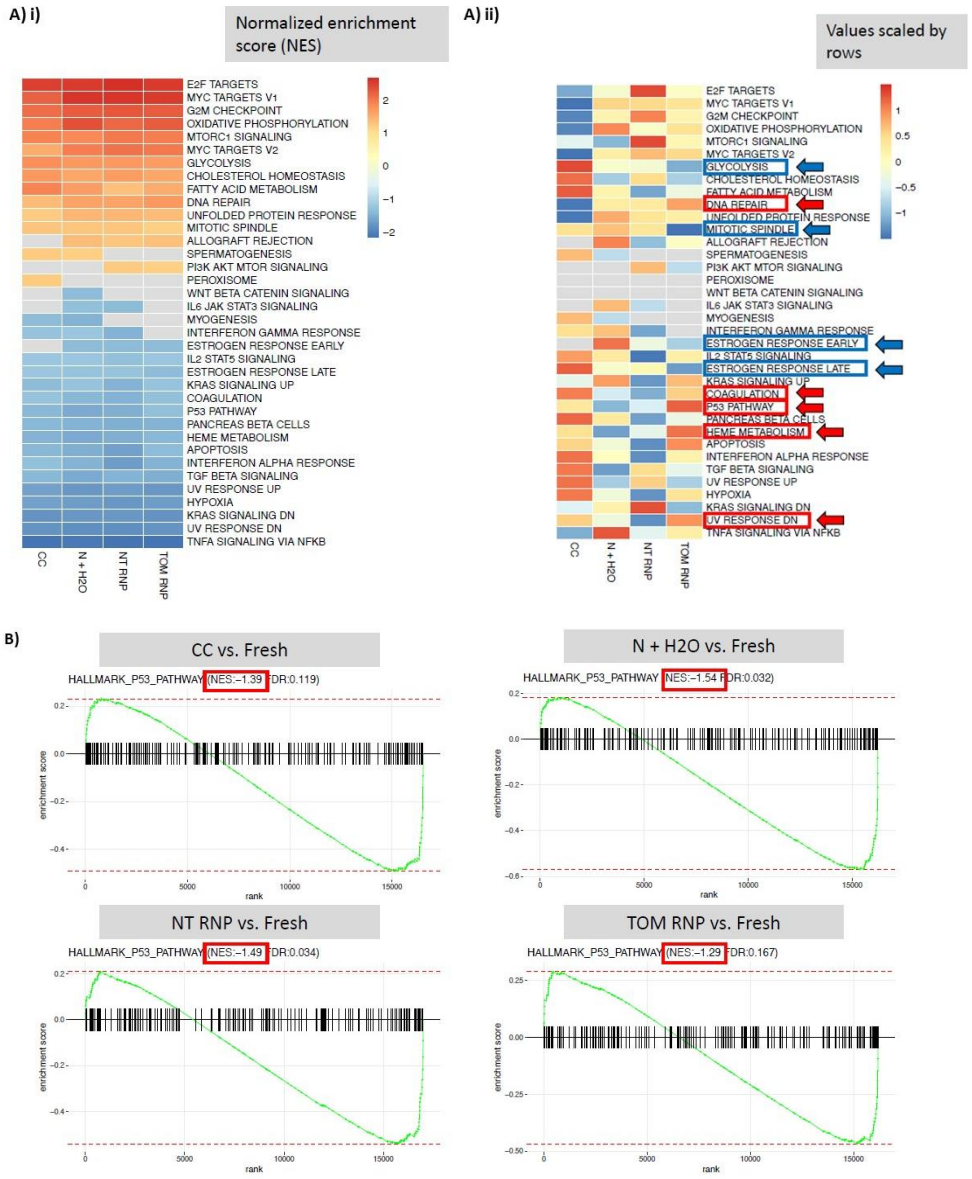


Figure 3.42: TP53 and DNA damage response pathways were upregulated with increasing significance across nucleofection conditions ordered by increasing intervention A) i) Normalised enrichment score for each group compared to fresh (no nucleofection or culturing) controls subject to gene set enrichment analysis using hallmark gene sets **ii)** Normalised enrichment scores scaled by rows ordered by nucleofection condition with increasing intervention (note scaled row values mean heat map represents differences between nucleofection groups only and not compared to fresh). **B)** GSEA plot for TP53 pathway genes for each nucleofection group compared to fresh as a baseline (note negative values due to TP53 pathway being enriched considerably in fresh samples compared to all other groups). Significance determined by FDR <0.25. N= 5 (fresh), 6 (culture control), 5 (water), 7 (NT RNP), 5 (TOM RNP) biological replicates per group.

The fact that the fresh group showed elevated *TP53* pathway enrichment was unsurprising given its role in normal haematopoiesis in preserving stem cells in a quiescent state (Asai, Liu et al. 2011). Given that this group was the only group not undergoing *in vitro* culture which promotes cycling of HSCs, it is likely that their transcriptional program would have been enriched for long-term stem cell properties, like quiescence maintenance. This was evident when analysing the stem cell signature of all groups, described later in this chapter.

Due to the limited level of separation between the nucleofection groups by differential gene expression, random forest analysis was carried out to further dissect the variation between these groups. This analysis revealed upregulation of genes associated with innate immune responses in all groups (**Figure 3.43**), with specific stress response genes most upregulated in nucleofection only (*Eef1akmt3*), specific inflammatory genes most upregulated in NT RNP samples (*Camp*) and genes involved in degradation of foreign RNA and proteins most upregulated in TOM RNP samples (*Rnf 208, Rnaset2a*) (**Figure 3.44**).

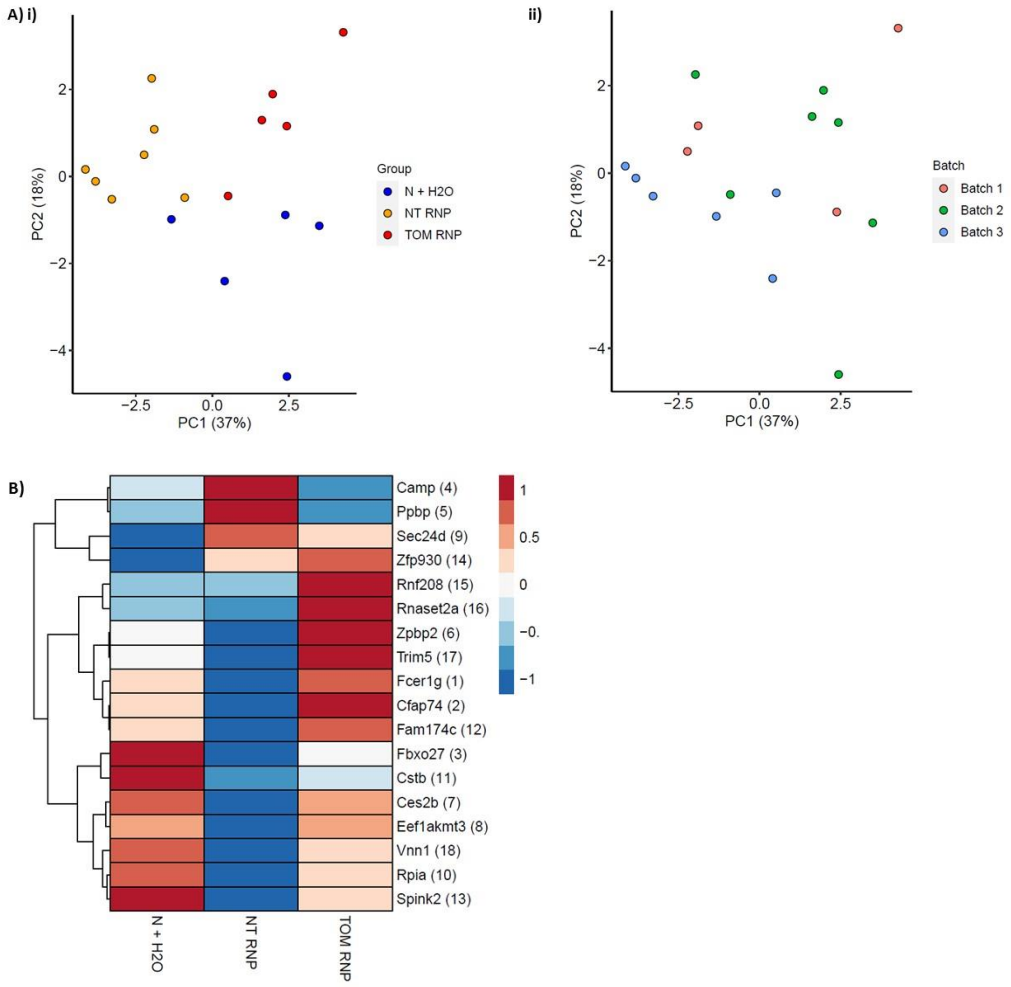


Figure 3.43: Random forest analysis separated nucleofection groups by gene expression A) Principal component analysis of nucleofection groups clustered by protein coding genes **i)** replicates labelled by nucleofection condition **ii)** replicates labelled by batch. **B)** Heat map showing normalised enrichment score for all protein coding genes clustered using random forest analysis of nucleofection groups only. N= 5 (water), 7 (NT RNP), 5 (TOM RNP) biological replicates per group.

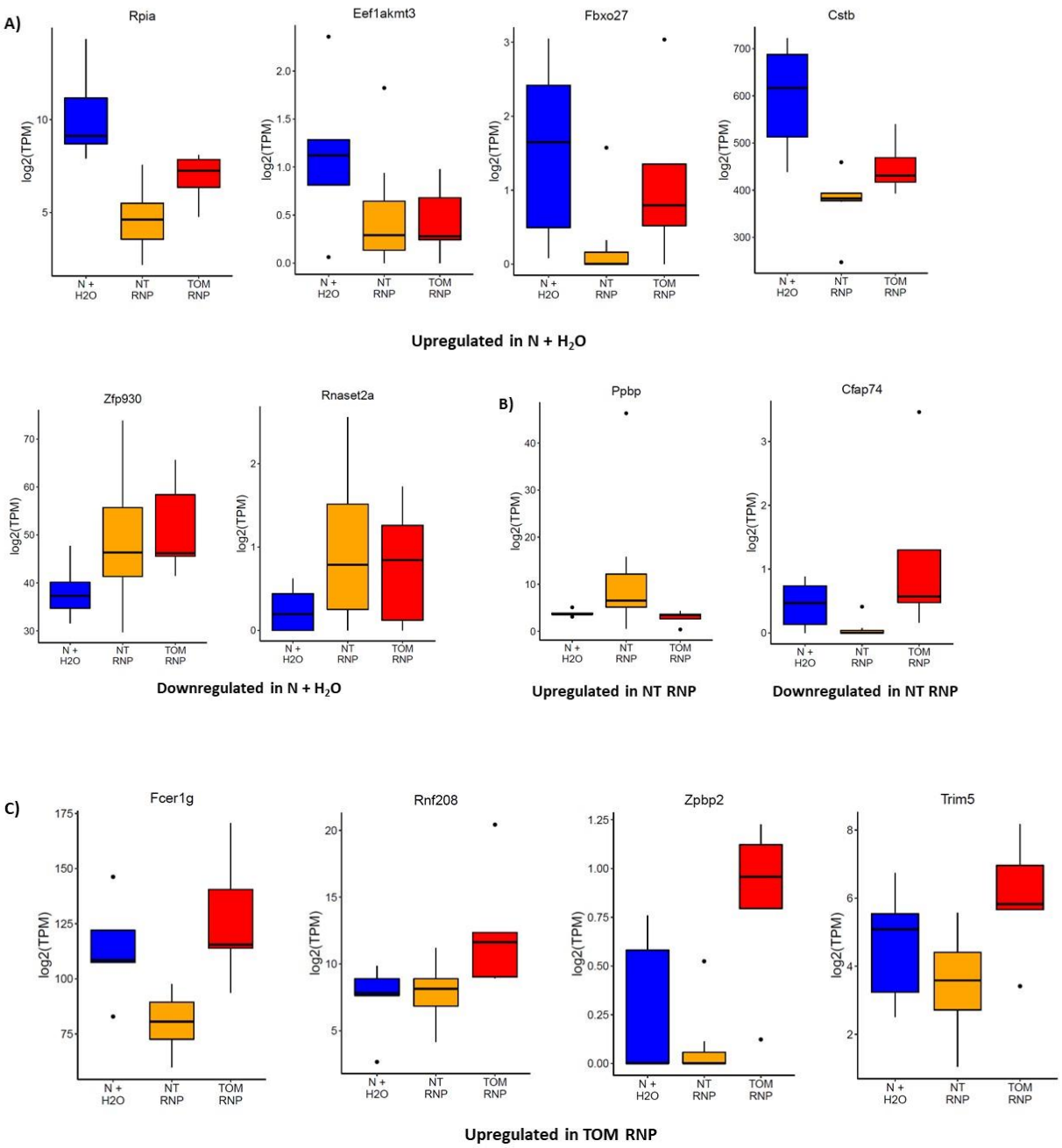


Figure 3.44: Innate immune response, degradation of foreign protein/RNA material, and DNA damage response separated nucleofection groups with increasing intervention Gene expression by log2 transcripts per kilobase million (TPM) for protein coding genes significantly up/down regulated exclusively in **A)** N + H_2O **B)** NT RNP **C)** TOM RNP nucleofection groups. N= 5 (water), 7 (NT RNP), 5 (TOM RNP) biological replicates per group.

The overall HSC and LT-HSC score was higher in edited samples compared to non-fresh controls, marked by upregulation of genes involved in the maintenance of metabolic quiescence (*Ak1*, *Hsd11b2*) (**Figure 3.45**).

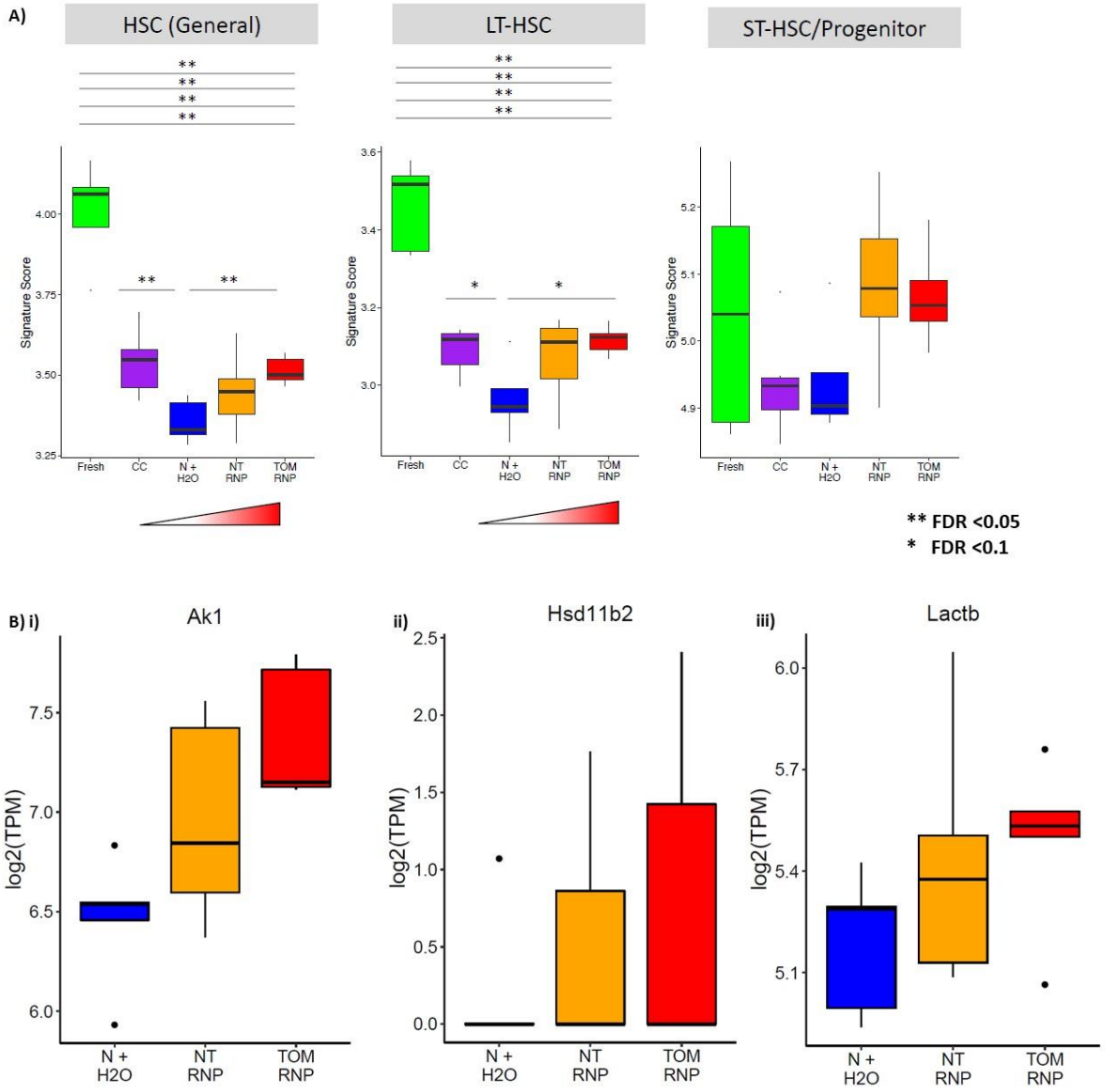


Figure 3.45: tdTomato edited samples possessed a stronger LT-HSC score compared to non-edited nucleofection controls A) Signature score for total HSC, LT-HSC and ST-HSC for tdTomato edited samples compared to nucleofected and non-nucleofected controls. **B)** Gene expression by log2 transcripts per kilobase million (TPM) for LT-HSC genes significantly upregulated in the TOM RNP group; **i)** Ak1 **ii)** Hsd11b2 **iii)** Lactb. N= 5 (water), 7 (NT RNP), 5 (TOM RNP) biological replicates per group. Significance calculated as FDR <0.1 (*) or <0.05 (**).

Interestingly, when combining the gene set contributing to the HSC score prediction, and the up-regulated genes from the differential gene expression, only one gene (*Rps6ka4*) overlapped between the two. This demonstrates that individually the stem cell genes weren't particularly differentially expressed in edited samples but that, as a group, this set of genes were overall slightly higher. This highlights the importance of looking at genes as groups as well as individually when computing subtle differences between groups.

3.2.4 Long-term engraftment of edited cells in the peripheral blood, bone marrow and spleen

As detailed in *Methods Section 2.3*, edited, sham-edited, non-edited, cultured and fresh controls were subject to competitive transplantation, 24 hours post-nucleofection, into constitutive Rosa26-tdTomato recipients such that edited and control cells could be distinguished from competitors by their tdTomato negativity. PB engraftment was analysed 4, 8, 12, 16, and 18 weeks post-transplant, where possible. Lineage and stem cell chimerism was measured 18 weeks post-transplant in the bone marrow and spleen.

i) Transplant protocol optimisation

In order to determine the dynamics of engraftment and bone marrow chimerism, a pilot experiment was carried out using 250,000 c-Kit enriched BM starting material per nucleofection. The aim of this pilot was to estimate the starting cell numbers required to achieve chimerism in the BM sufficient for sorting edited HSCs 18 weeks post-transplant. Given the stem cell compartment comprises

0.01% of the total mouse BM (Challen, Boles et al. 2009), it was necessary to ensure that the total engraftment of edited cells in the BM was at least 15% in order to be able to sort a minimum of 50 edited HSCs per sample. Further, given that the RNA-Seq data 24 hours post-nucleofection demonstrated an upregulation of *TP53* and DNA damage pathways, it was predicted that edited samples may be at a fitness disadvantage compared to non-edited controls upon competitive transplant with wild type counterparts. Therefore it was important to determine the amount of starting material required to achieve sufficient chimerism in recipients. In addition, in order to refine the most important controls prior to the extended transplant experiment combined with RNA-Seq, all stages of the nucleofection process were tested and compared to the matched 24-hour RNA-Seq data. The only exception to this was the non-targeting control in the 24-hour RNA-Seq which was compared to a Cas9 only control in the transplant setting. It was hypothesised that given the small differences in NT RNP and TOM RNP 24-hour RNA-Seq, Cas9 only would be a better control for nucleofection without editing. **Figures 3.46-3.49** summarise the serial PB engraftment and **Figure 3.50** summarises the 18-week chimerism in the BM of tdTomato edited compared to sham-edited and control samples (**Figure 3.24** summarises the experimental workflow).

In the culture control sample detailed in the gating example shown in **Figure 3.46**, there was evidence of tdTomato negative donor cells in every lineage analysed, and clear separation between positive and negative fractions. As shown in **Figure 3.47,** in contrast to the 24-hour RNA-Seq data, the most striking sample separation came between tdTomato targeted samples compared to all other control samples. While total PB engraftment was stable across the serial time-

points for all samples, the engraftment level was ~65% for cultured only, nucleofection + water, and Cas9 only samples, whereas it was ~15% for tdTomato RNP samples (**Figure 3.47 A**). This suggests that although edited HSCs stringently sorted on viability demonstrated subtle changes in gene expression compared to sham-edited or controls, at a bulk level there was a significant fitness disadvantage in the transplant setting. Looking at each lineage separately, there was no notable difference in lineage contribution between edited compared to control samples, although in general for all samples lymphoid engraftment was higher than myeloid (**Figure 3.47**).

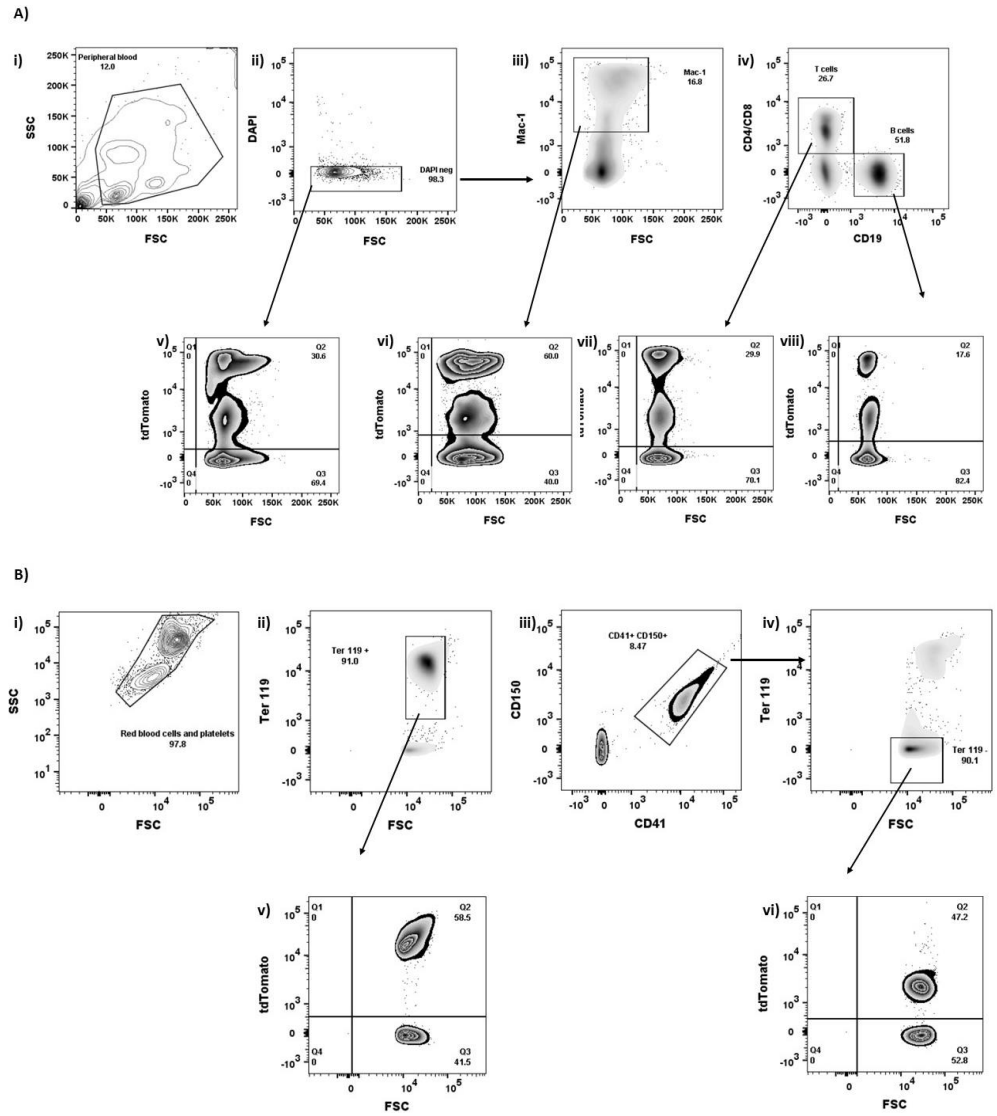


Figure 3.46: Gating strategy for analysing peripheral blood lineage engraftment of tdTomato edited and control groups A) Culture control sample 2, week 12 example peripheral blood flow cytometry gating strategy separating **i)** total peripheral blood **ii)** viable cells **iii)** myeloid cells **iv)** B cells and T cells **v)** viable cells tdTomato negative **vi)** myeloid cells tdTomato negative **vii)** B cells tdTomato negative **viii)** T cells tdTomato negative. **Note** tdTomato gating was not the same for each lineage as these were based on total positive/negative controls whose tdTomato intensity varied for different cell types. **B)** Culture control sample 2, week 12 example peripheral blood flow cytometry gating strategy separating **i)** platelets and red blood cells **ii)** Ter119+ red blood cells **iii)** CD41+ CD150+ platelets **iv)** Ter119- platelets **v)** red blood cells tdTomato negative **vi)** platelets tdTomato negative.

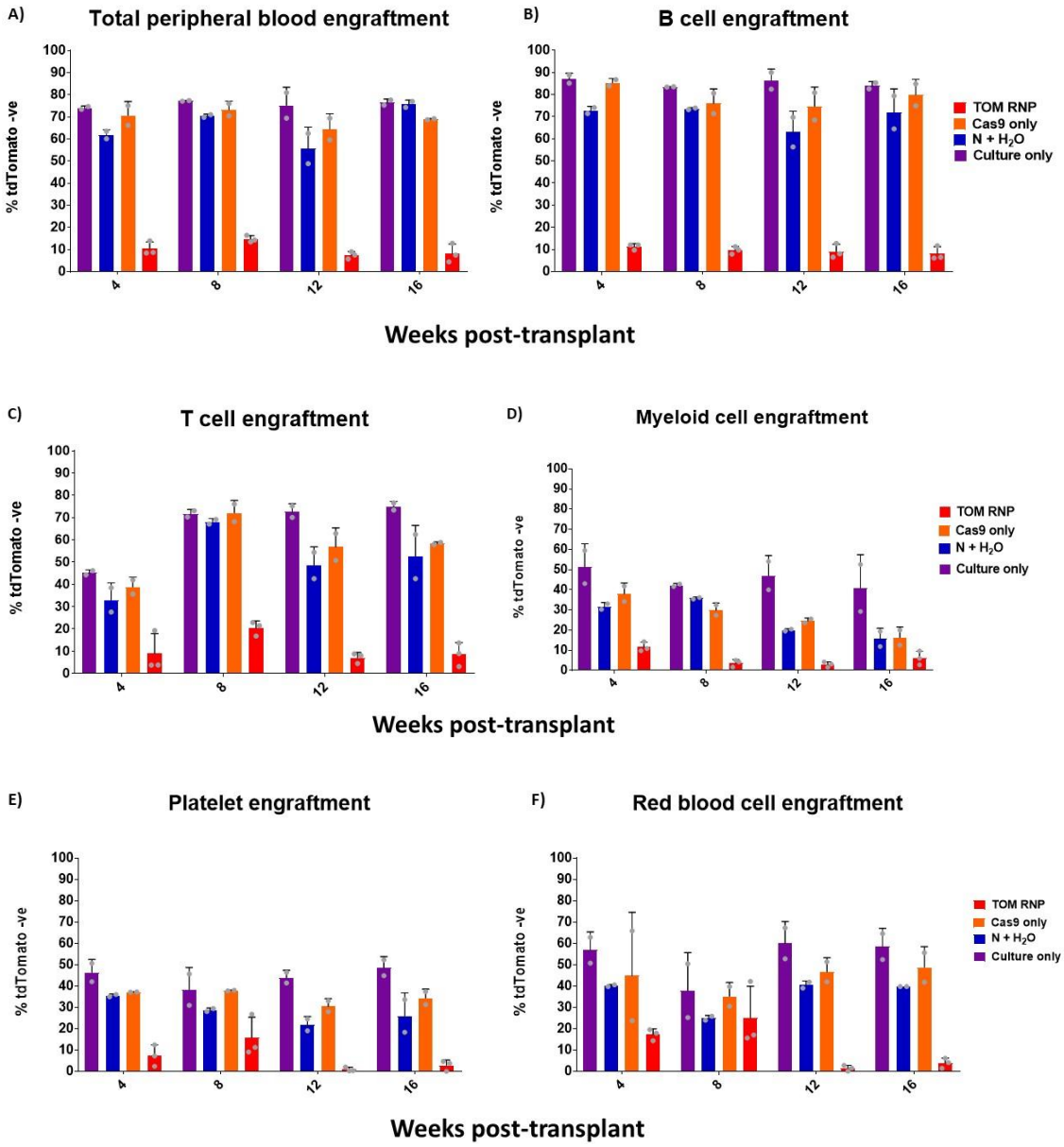


Figure 3.47: Low volume protocol CRISPR editing impeded the engraftment ability of targeted cells more than nucleofection itself or presence of the Cas9 protein Serial peripheral blood engraftment in **A)** Total peripheral blood **B)** B cells **C)** T cells **D)** Myeloid cells **E)** Platelets **F)** Red blood cells. N= 2-3 biological replicates. Data shown as individual points, error bars represent mean with standard deviation.

Unfortunately, as shown in **Figure 3.48** engraftment for edited samples in the most primitive compartments was very poor, with less than 5% engraftment in the LSK compartment for edited samples compared to 50% for nucleofection +

water and Cas9 only and 70% for culture controls. This suggests that a significant level of HSC DSB-induced apoptosis occurred post-transplantation as despite equal numbers of viable c-Kit + cells being transplanted across samples, levels of engraftment were different. The lack of any engraftment in the HSC compartment for edited cells clarified that the PB output must have come from more committed progenitors or short-lived HSCs which weren't able to maintain long-term engraftment in the BM. It is expected that if this readout had been continued beyond 16 weeks, PB engraftment would have then fallen.

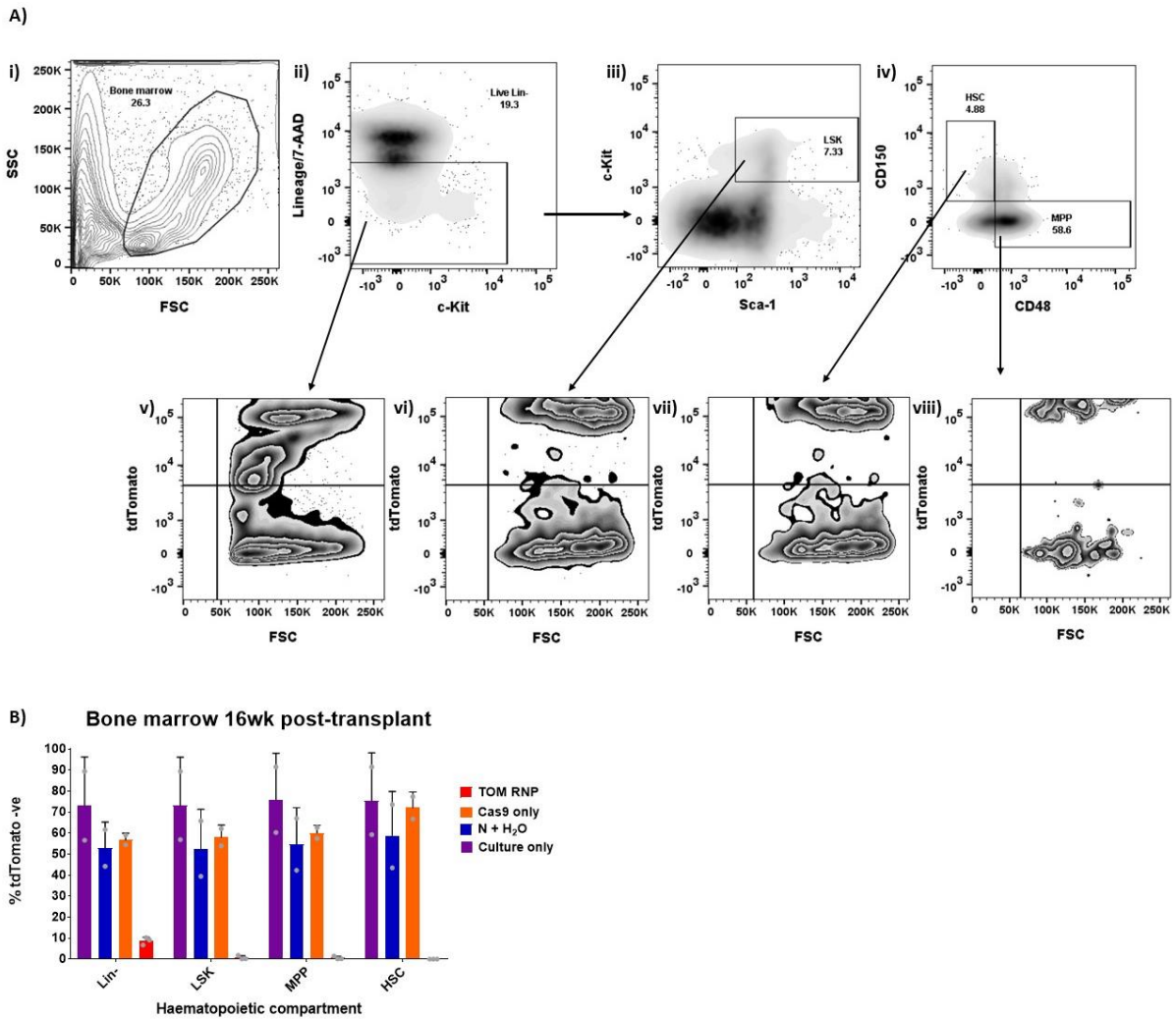


Figure 3.48: Low volume protocol CRISPR editing impeded the engraftment in primitive BM compartments of targeted cells more than nucleofection itself or presence of the Cas9 protein A) Culture control sample 2, week 16 example bone marrow flow cytometry gating strategy separating **i)** total bone marrow **ii)** viable Lin- cells **iii)** Lin- c-Kit+ Sca-1+ (LSK) cells **iv)** CD48+ CD150- (MPPs), CD48- CD150+ (HSCs) **v)** viable Lin- cells tdTomato negative **vi)** LSKs tdTomato negative **vii)** HSCs tdTomato negative **viii)** MPPs tdTomato negative. **B)** Percentage tdTomato negative in primitive BM compartments 16wk post-transplant. N=2-3 biological replicates. Data represent mean with standard deviation.

Further protocol optimisation was required to achieve successful engraftment of edited HSCs. It is known that cell number and nucleofection volume greatly influence viability post-nucleofection, particularly in primary cells. It was thus

hypothesised that modifying the protocol to use 1,000,000 cells in a 100µL volume cuvette, as opposed to 250,000 cells in a 20µL volume cuvette, could increase the viability of HSCs post-nucleofection and therefore the potential of successful long-term BM engraftment. The low cell number protocol was originally used to reduce the primary cell requirement and thus reduce the number of mice in line with the principles of 3R. In order to refine the use of primary material, it was decided from the pilot low volume protocol that, as the controls all showed similar engraftment, the best controls to carry forward to the high volume protocol would be fresh and NT RNP. These also allowed RNA-Seq data to be generated to directly match controls used for 24 hours post-nucleofection assessment. Unfortunately, due to the Covid-19 related lab closure, it was not possible to obtain material for serial readouts. However, PB and BM chimerism was assessed at 18 weeks when the lab re-opened. **Figures 3.49** and **3.50** summarise the results of this protocol adaptation.

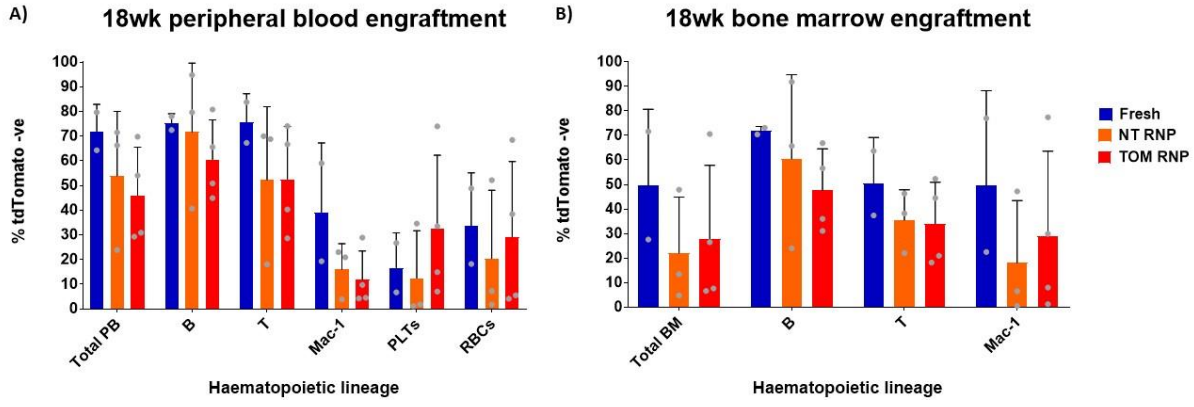


Figure 3.49: High volume protocol partially rescued the impeded engraftment caused by the low volume protocol Percentage tdTomato negative cells 18 weeks post-transplant in **A)** peripheral blood lineage compartments **B)** bone marrow lineage compartments.

Use of the large volume protocol, with transplantation of similar viable cell numbers at 24 hours post-nucleofection, rescued the engraftment level of tdTomato edited samples. **Figure 3.49 A** shows there was similar (50-70%) total PB engraftment for edited samples, compared to non-targeting and fresh controls; and this was consistent across each lineage analysed. However, there was high variability between replicates groups most likely due to the wide age range (7-18 weeks) of recipients. This experiment was required to be set up before the Covid-19 lab closure with the exception of a long closure period so only the recipient mice available at the time could be used, which limited the possibility of keeping all recipients in the 8-12 adult age range. There was a trend towards slightly lower engraftment for non-targeting and tdTomato targeting samples but this was not statistically significant probably because of the variability between the samples. **Figure 3.49 B** shows there was more disparity in the BM lineage engraftment, most marked in the myeloid compartment however, again, there was high variability between replicates.

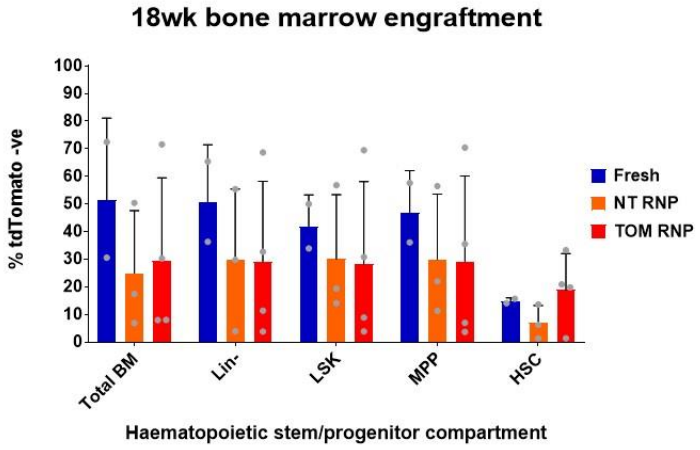


Figure 3.50: High volume protocol rescued the impeded engraftment in primitive stem cell compartments caused by the low volume protocol, but with variable efficiency Percentage tdTomato negative 18 weeks post-transplant in bone marrow HSPC compartments.

Similar levels of engraftment for all groups were seen in the primitive compartments, although overall engraftment in the HSC compartment was lower than less primitive compartments for all groups (30-50% for Lin-, LSK and MPP compared to 10-15% for HSC) (**Figure 3.50**). These data provided insights into the long-term lineage and HSC engraftment of edited HSCs, as well as establishing the large volume protocol as sufficient to allow robust HSC engraftment. However, the lack of serial data due to Covid-19 related lab closures impeded these data from shedding light on the engraftment dynamics of edited cells. It was not possible to determine whether edited cells showed delayed engraftment or long-term drop-off in particular lineages which would indicate a potential HSC defect.

ii) Serial peripheral blood engraftment

Having established the large volume protocol as efficient to achieve long-term engraftment of edited HSCs, experiments were repeated after lab re-opening to generate serial PB engraftment data. As detailed in *Methods Section 2.3*, tdTomato edited, sham-edited and fresh c-Kit enriched mouse BM samples were transplanted 24 hours post-nucleofection and subject to serial four-week PB readouts, followed by terminal 18-week BM, spleen and PB chimerism analysis. **Figures 3.51-3.54** summarise the results.

Overall, it was clear that total PB engraftment was lower for tdTomato targeted samples (**Figure 3.51 A** 60% for fresh and NT RNP compared to 20% for targeted, $p < 0.0001$). This was most likely due to reduced viability and increased DNA damage-induced apoptosis for targeted samples 24 hours post-nucleofection (at the time of transplant). The reason for this is the evidence from

the RNA-Seq data that both DNA damage and *TP53* pathways were upregulated (**Figure 3.42 A**). This reduction in viability would lead to reduced competitive ability of the HSCs, reduced numbers of successfully engrafting HSCs and an increased contribution from the competitors. This lower level of engraftment was consistent across lineages (**Figure 3.51 B-F**) with TOM RNP samples contributing less to PB output than controls. However, due to the variation in level of total PB engraftment, it was difficult to assess whether the engraftment trajectories were different between TOM RNP and control samples within different lineages.

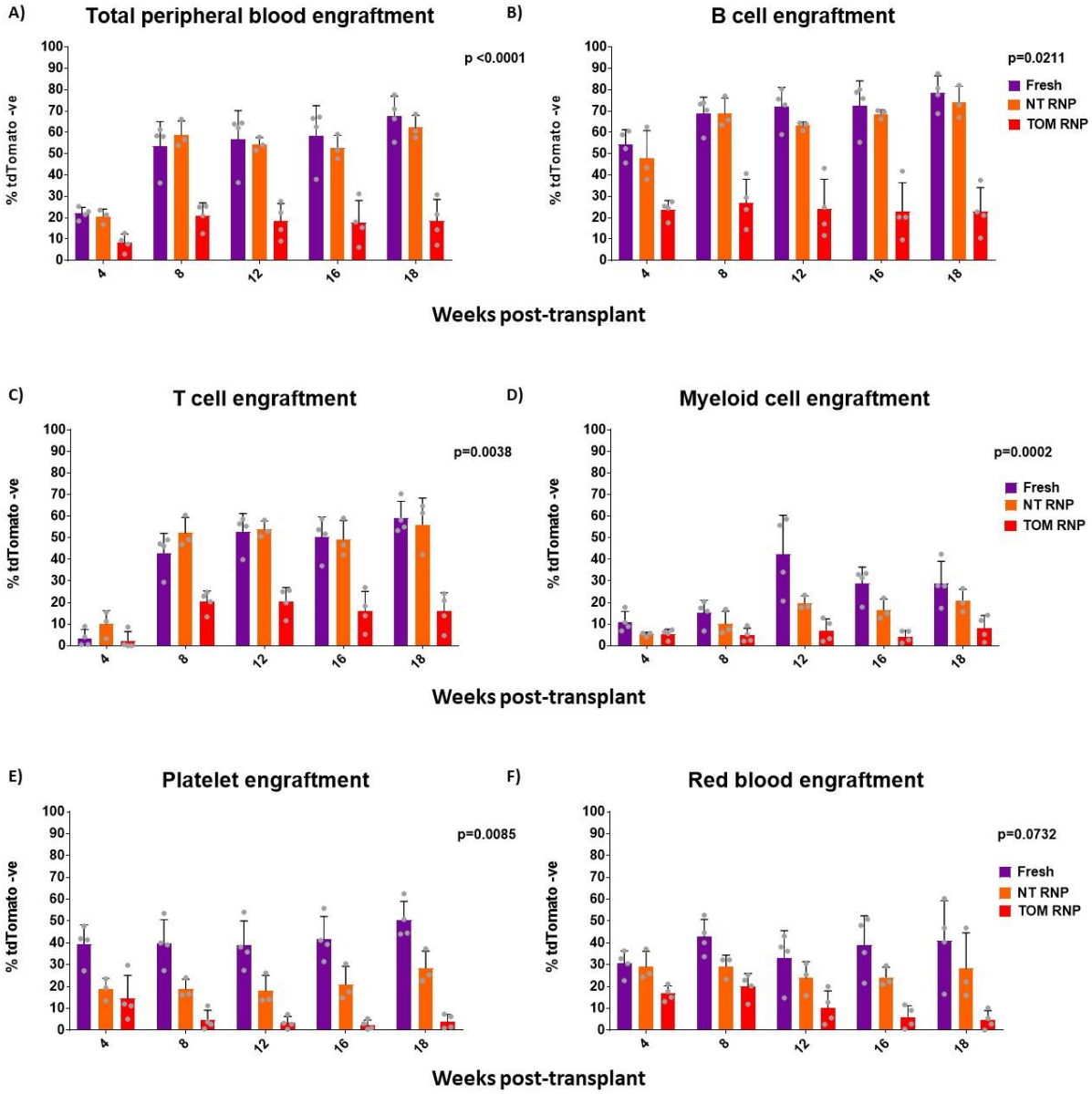


Figure 3.51: Total peripheral blood engraftment was negatively impacted by editing although the severity of the effect differed between lineages Percentage tdTomato negative cells 4, 12, 16 and 18 weeks post-transplant in **A)** Total peripheral blood **B)** B lineage **C)** T lineage **D)** Myeloid lineage **E)** Platelet lineage **F)** Red blood lineage. N= 3-4 biological repeats. Statistical significance calculated using a two-way RM ANOVA (**Note**: 4wk time-point not used in statistical analysis for T cells due to delayed engraftment of this lineage hence potential skewing of the p value).

In order to account for the variable levels of total PB engraftment between conditions, a ratio of four-week engraftment was calculated to more accurately depict the dynamics of lineage engraftment (**Figure 3.52**). To adjust for the

differing levels of total PB engraftment, each lineage was taken separately and the four-week percentage tdTomato negative was set as a baseline. Engraftment at each subsequent time-point was calculated as a ratio of the initial four-week level. What this analysis clearly revealed was the negative trajectory of platelet (**Figure 3.52 E**, $p < 0.0001$) and red blood cell (**Figure 3.52 F**, $p=0.0029$) engraftment for TOM RNP samples compared to controls.

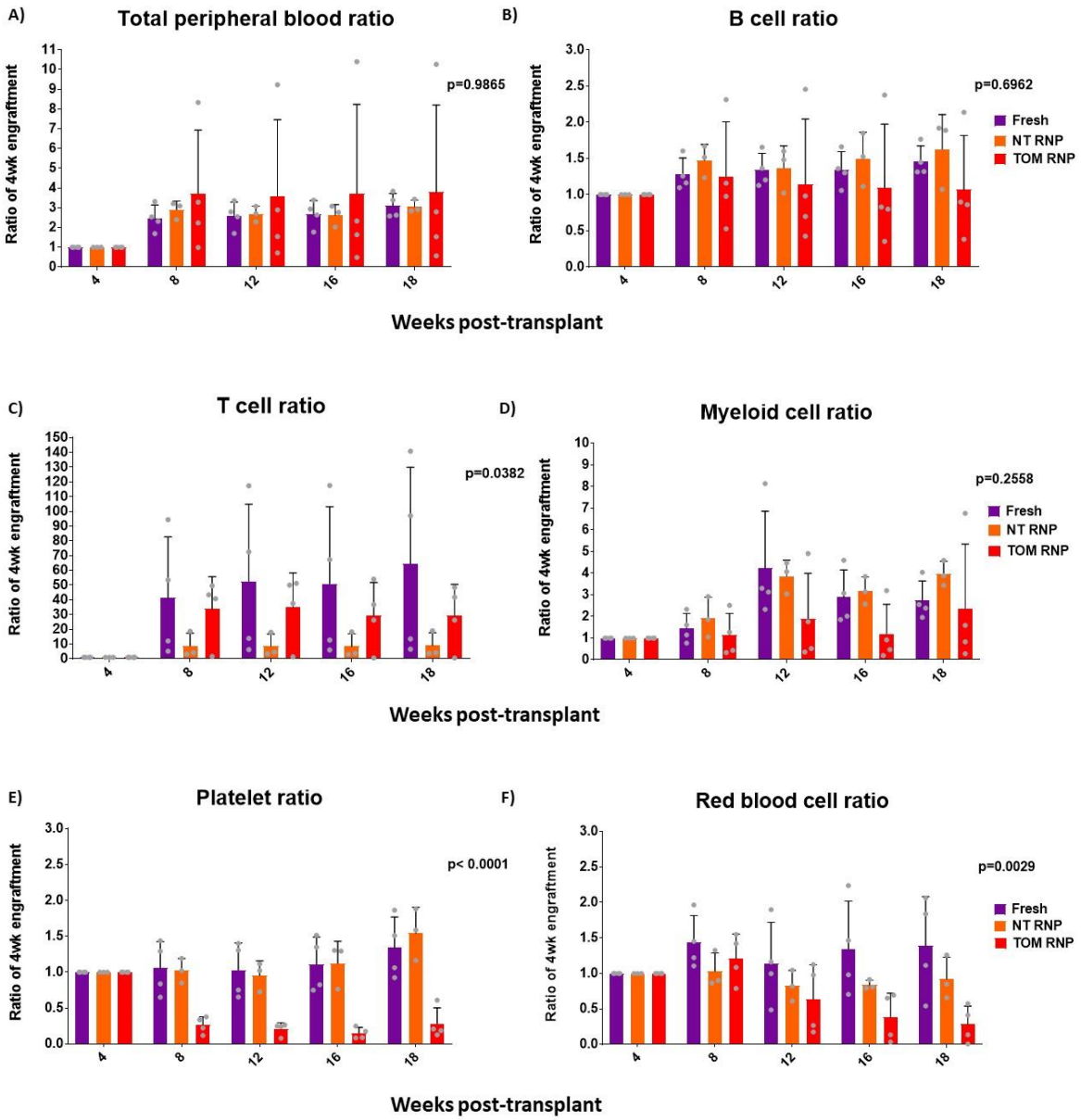


Figure 3.52: Platelet and red blood cell engraftment decreased over serial readouts exclusively in edited samples when accounting for total overall engraftment tdTomato negative engraftment as ratio of 4-week lineage engraftment in **A)** Total peripheral blood **B)** B lineage **C)** T lineage **D)** Myeloid lineage **E)** Platelet lineage **F)** Red blood lineage. N= 3-4 biological repeats. Data represent mean with standard deviation. Significance in **A)**-**F)** calculated using a two-way RM ANOVA (**Note**: 4wk time-point not used in statistical analysis for T cells due to delayed engraftment of this lineage hence potential skewing of the p value).

In the BM and spleen, the total engraftment of all lineages was significantly less for TOM RNP samples. (**Figure 3.53 A-B**, p=0.0142-0.0062). The proportion of myeloid engraftment was significantly less in the both the BM and spleen for TOM RNP samples (**Figure 3.53 C**, p= 0.0009, p= 0.0051 respectively).

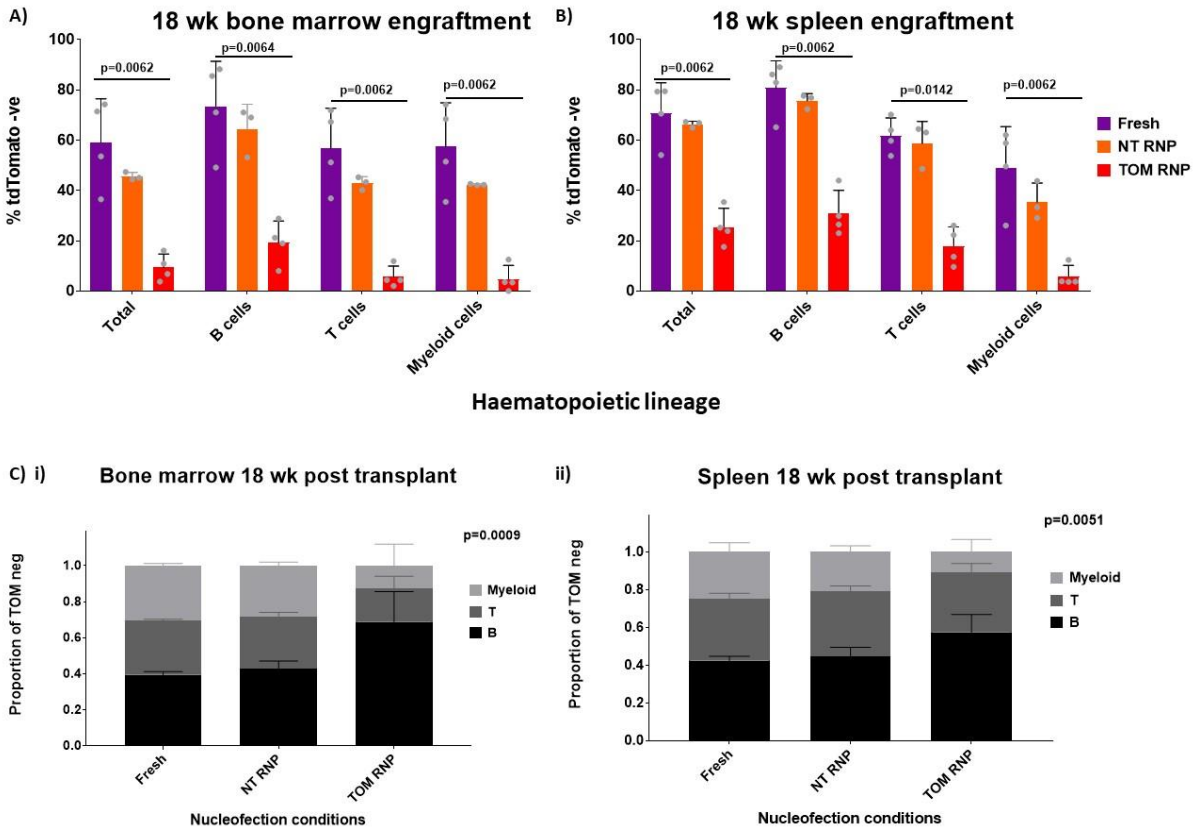


Figure 3.53: Myeloid engraftment in the bone marrow and spleen contributed to a lower overall proportion in edited samples compared to controls Percentage tdTomato negative 18wk lineage engraftment **A)** Bone marrow **B)** Spleen. **C)** Proportion of myeloid and lymphoid lineages contributing to 18wk engraftment in **i)** Bone marrow **ii)** Spleen. N=3-4 biological replicates. Data represent mean with standard deviation. Significance in **A)** and **B)** calculated using a non-parametric Kruskal-Wallis test. Significance in **C)** calculated using a two-way RM ANOVA.

Similar to the mature lineage engraftment in the PB, BM and spleen, the overall engraftment in all stem/progenitor populations analysed in the BM and spleen

was significantly lower in TOM RNP samples compared to NT RNP and Fresh controls (**Figure 3.54 A**, **B**, p=0-0159-0.0029). Interestingly, the population with most decreased tdTomato negativity in the TOM RNP group was the LSK compartment (p=0.0029). This was the same compartment that was reduced in the TOM RNP group compared to controls at 24 hours post-nucleofection (**Figure 3.30 C**). What was also noticeable, was that the increase in engraftment between the HSC and MPP compartments when each group was compared separately was particularly different. While the Fresh and NT RNP groups showed two to three times MPP tdTomato negative cells compared to their corresponding HSC engraftment; for TOM RNP samples however this was closer to a 1:1 ratio (**Figure 3.54 A**, outlined). This reduced MPP output suggests there could have been a partial stem cell defect in this group. Indeed, it has been shown that in a mouse model of impaired DNA damage response (carrying a proliferating cell nuclear antigen K164R mutation) HSC and MPP compartment proportions were disrupted (Pilzecker, Buoninfante et al. 2017). This difference was not as a striking in the spleen with MPP fractions increasing compared to HSC in a similar proportion for both TOM RNP targeted and control samples (**Figure 3.54 B**).

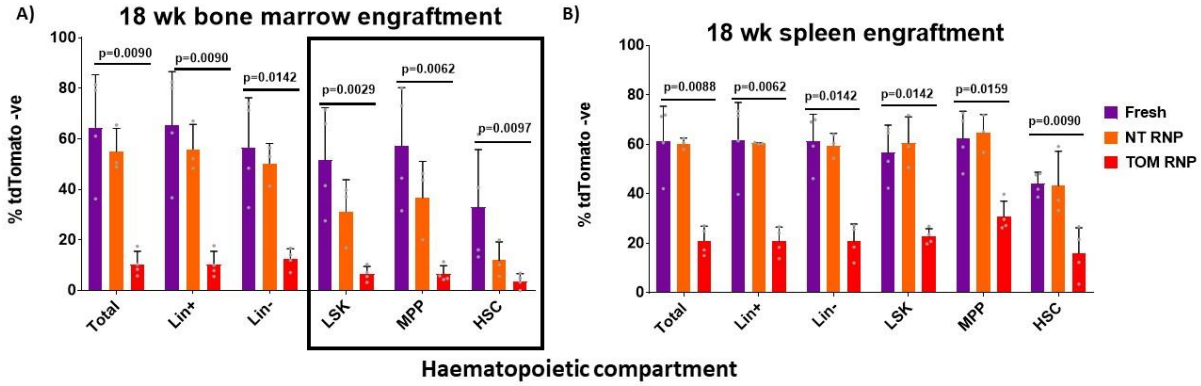


Figure 3.54: Engraftment in the haematopoietic stem/progenitor compartments was significantly reduced for TOM RNP samples compared to controls Percentage tdTomato negative engraftment in stem and progenitor compartments 18wk post-transplant in **A)** Bone marrow **B)** Spleen.

Due to the lack of serial data for the first set of biological replicates of this experiment, as a consequence of the Covid-19 related lab closure, data could only be combined for the 18-week time-point. **Figures 3.55** and **3.56** summarise the combined results for these two sets of biological replicates.

Figure 3.55 A shows that the overall PB engraftment for TOM RNP targeted samples was significantly lower than controls (p=0.0166) although this was not statistically significant for all lineages, particularly the red blood cells and platelets. However, there is an outlier in the first biological repeat, which was considerably higher for all lineages hence drawing up the average. It was not possible to analyse the engraftment course of this particular replicate as it was transplanted prior to the Covid-19 lab closure hence only data for the terminal readout was collected. It could be that this sample had a sudden increase in engraftment or consistent steady engraftment. In contrast, in the BM, engraftment of all lineages was significantly lower in TOM RNP samples (**Figure 3.55 B**). Interestingly, although the total BM engraftment was significantly lower for TOM RNP targeted samples, within the LSK, MPP and

HSC compartment this was no longer statistically significant when both biological replicates were combined (**Figure 3.55 C**). Although this was different to the significance seen in **Figure 3.54 A** it remained supportive of the observation that the edited HSCs had a lower output of mature lineages compared to the sham-edited and fresh counterparts.

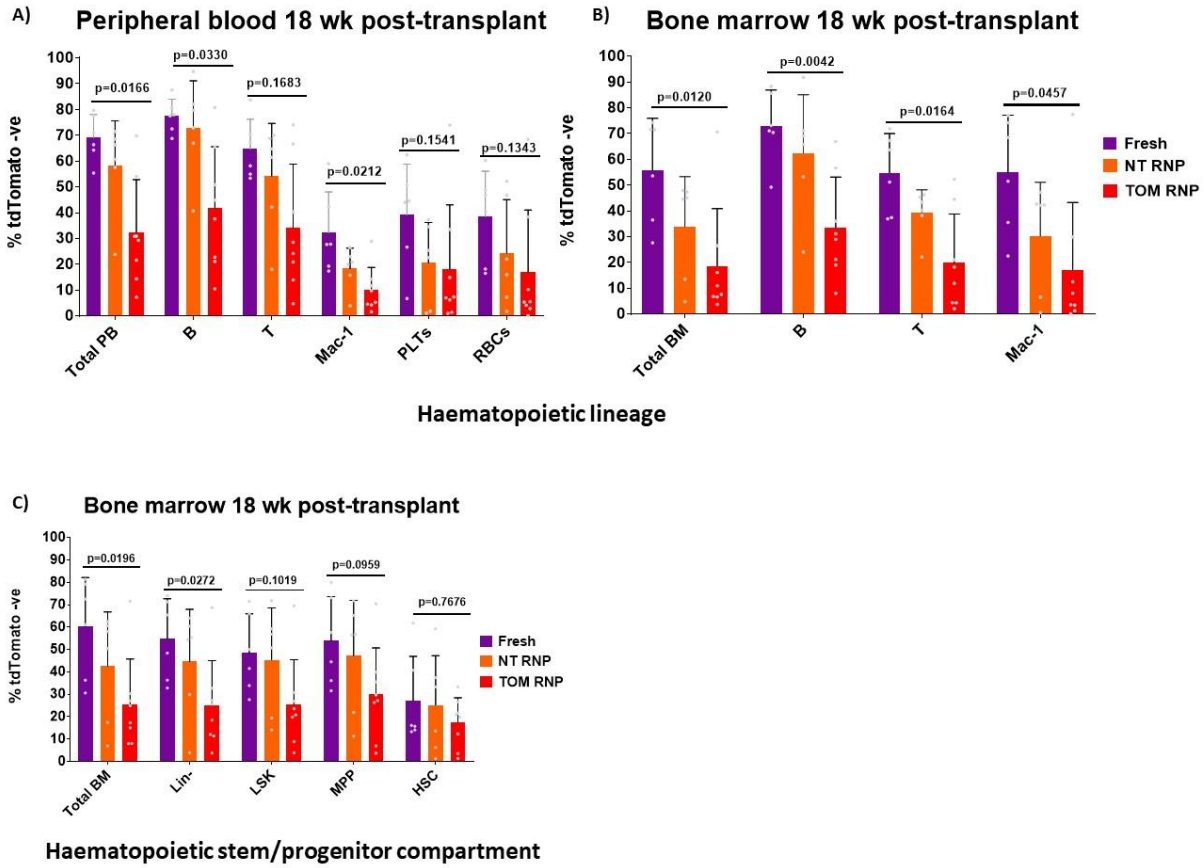


Figure 3.54: Contribution of tdTomato negative TOM RNP and control groups to peripheral blood mature lineages and bone mature lineages/HSPC compartments for two biological repeats combined
Percentage tdTomato negative 18 weeks post-transplantation in **A)** Peripheral blood lineages **B)** Bone marrow lineages **C)** Bone marrow HSPC compartments. N= 5-8 biological replicates from two independent experiments. Data represent mean with standard deviation. Statistical significance calculated using a non-parametric Kruskal-Wallis test.

Similar to the conclusion drawn from **Figure 3.55 C**, **Figure 3.56 B** demonstrates a significantly larger proportion of HSC compared to MPP contributed to the LSK compartment in TOM RNP targeted samples compared to controls. Interestingly, the myeloid lineage shared a significantly lower proportion of the tdTomato negative engraftment of the BM (**Figure 3.56 A ii**, p=0.0059) in TOM RNP samples but this was not the case for the peripheral blood (**Figure 3.56 A i**, p=0.1797).

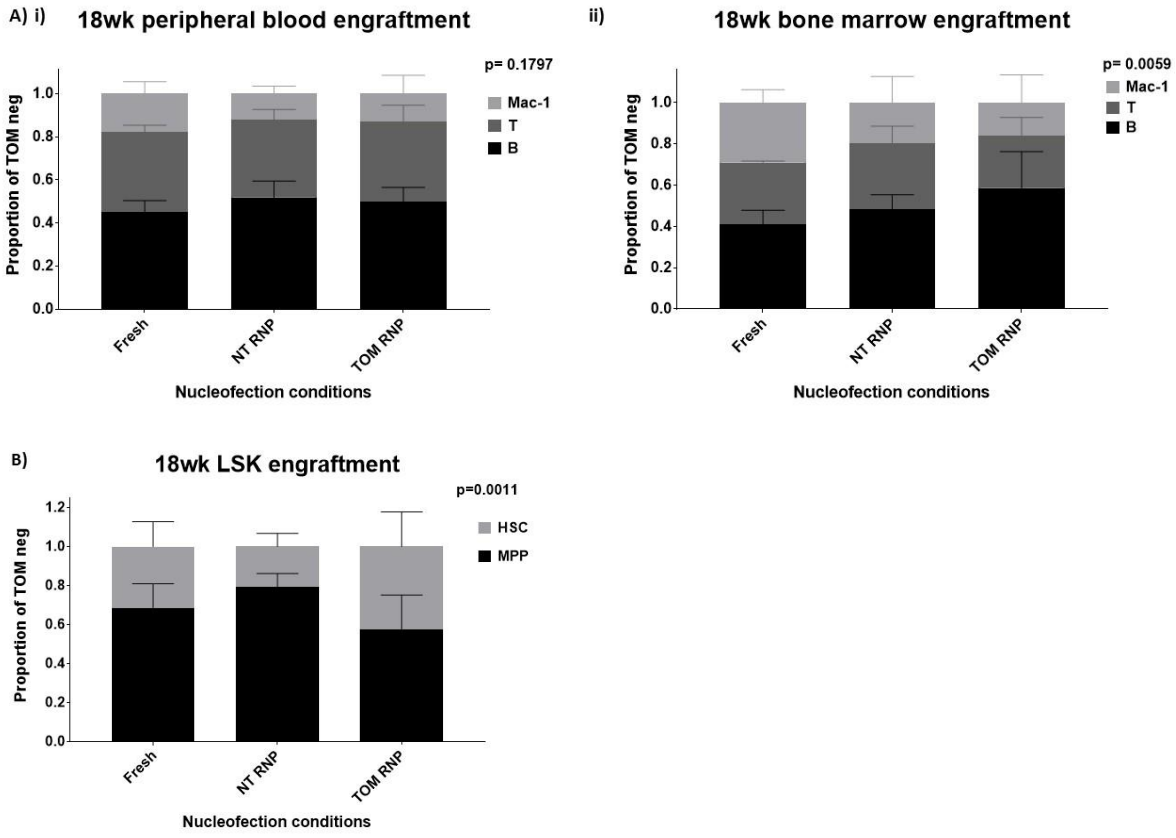


Figure 3.56: Proportions of tdTomato negative donors contributing to lineage/HSPC compartments and HSC lineage output for two biological repeats combined A) 18wk lineage proportion contributing to tdTomato negative engraftment in **i)** peripheral blood ii**)** bone marrow. **B)** Proportion of HSC/MPP contributing to tdTomato negative LSK compartment. N- 5-8 biological replicates from two independent experiments. Significance calculated using two-way ANOVA. Data represent mean with standard deviation.

3.2.5 Transcriptomic analysis of the HSC compartment 18 weeks post-transplant

In order to determine whether 24-hour transcriptional changes were maintained long-term, as well as to address potential reasons to explain the impact on HSC function of tdTomato editing seen in the engraftment data, tdTomato negative HSCs were purified from the BM of transplanted recipients 18 weeks post-transplant. Purified tdTomato negative HSCs were then subject to whole-transcriptome sequencing; *Methods section 2.3* details the experimental approach and phenotypic definition of stem cells used. Analysis was kindly performed by Wei Wen. In brief, samples were aligned to the mouse genome and only those with 80% or higher reads mapped were processed for downstream analysis. A summary of initial alignment scores is available in *Appendix 6*. Batch correction was performed using Combat-Seq (Zhang, Parmigiani et al. 2020). Differential gene expression analysis was carried out using DESeq2; with the FDR set at <0.05 for all pairwise comparisons. Gene-set enrichment analysis was performed on hallmark gene sets to assess groups enriched across each stage of the editing process; with the FDR set at <0.25. Random forest analysis was performed to further segregate differentially expressed genes between the nucleofection groups. Finally, HSC gene sets (Ivanova, Dimos et al. 2002) were used to generate a HSC transcriptional score for each group. A full list of HSC, LT-HSC and ST-HSC/Progenitor gene sets used to compute the score is available in *Appendix 8*. **Figures 3.57-3.67** summarise the results.

Figure 3.57 demonstrates the technical robustness of the approach, in which technical and biological replicates clustered tightly. **Figure 3.58** demonstrates the efficiency of the Combat-Seq approach in removing batch effects caused by

sort month. What is clear from **Figure 3.58 B) ii)** is that unsupervised clustering was unable to separate the transplanted groups from each other. This was also the case for the 24-hour post-nucleofection samples. This is not surprising as the cells were all stringently sorted based on phenotypic HSC markers therefore any transcriptional differences were expected to be subtle as the cell type sorted was the same. Therefore, supervised approaches were used to further cluster these groups.

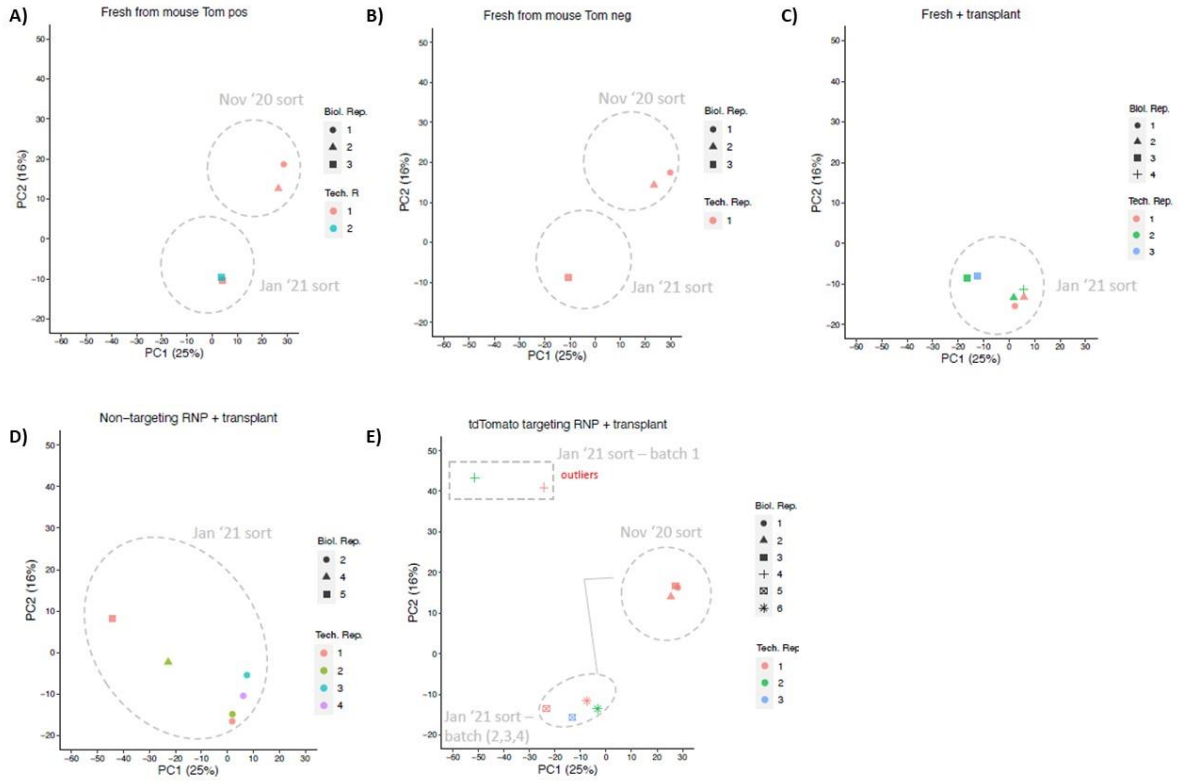


Figure 3.57: Technical and biological replicates clustered closely

Principal component analysis of each sample labelled by technical and biological replicate for **A)** Fresh tdTomato positive **B)** Fresh tdTomato negative **C)** Fresh + transplant **D)** NT RNP + transplant **E)** TOM RNP + transplant. N=6 (Fresh), 4 (Fresh + transplant), 3 (NT RNP + transplant), 6 (TOM RNP + transplant) biological replicates per group.

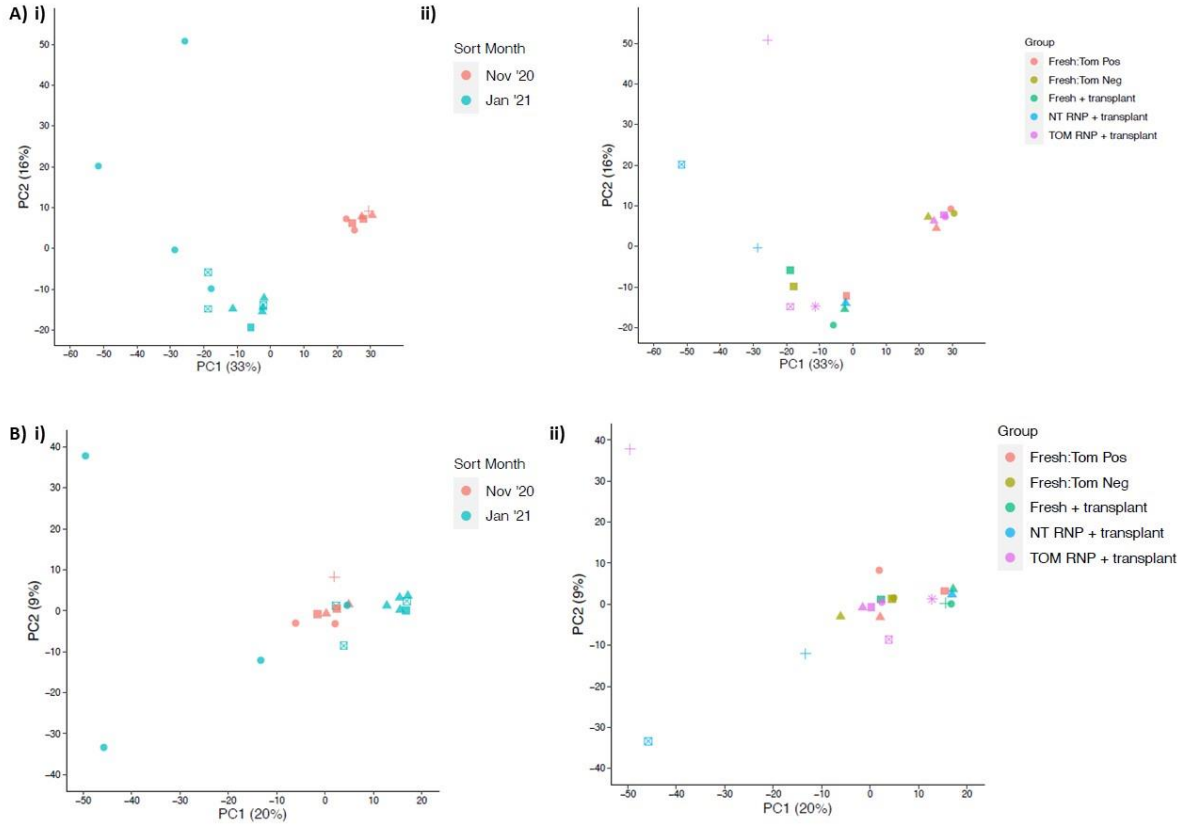


Figure 3.58: Unsupervised clustering failed to separate groups by principal component analysis Principal component analysis of fresh and transplanted samples with total gene expression counts **A)** pre Combat-Seq batch correction **B)** post Combat-Seq; batch correction labelled by **i)** batch **ii)** sample. N=6 (Fresh), 4 (Fresh + transplant), 3 (NT RNP + transplant), 6 (TOM RNP + transplant) biological replicates per group.

Prior to supervised clustering of the groups, tdTomato transcript levels and gene set enrichment between the fresh tdTomato negative and positive groups were assessed as a control step. **Figure 3.59 A** confirms that the tdTomato transcript levels were at the level of fresh tdTomato negative controls for transplanted samples, as expected from the stringent tdTomato negative sorting gates. **Figure 3.59 B** demonstrates that no enrichment of gene sets by GSEA was present between fresh tdTomato positive and negative groups. The reason for testing this was to confirm that 24 hours post-nucleofection and 18 weeks post-transplant samples could be compared without the possibility that

transcriptional differences were arising from expression of the tdTomato reporter itself. This is due to the fact the HSCs sorted at 24 hours post-nucleofection were tdTomato positive whereas those sorted at 18 weeks post-transplant were tdTomato negative. Having established no gene set enrichment between these fresh controls, they were merged for downstream analysis.

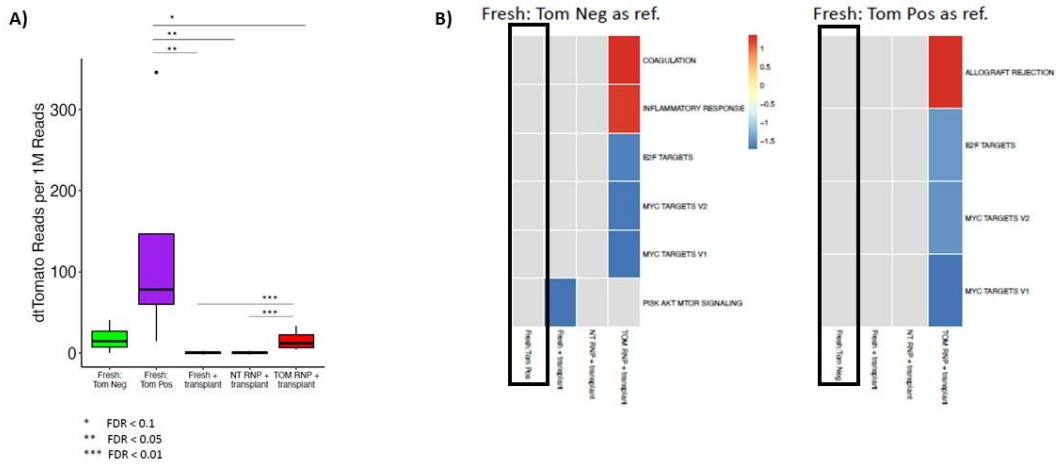


Figure 3.59: tdTomato reads were significantly diminished in transplant samples. No gene sets were significantly enriched between fresh tdTomato positive and negative samples A) tdTomato read count as a proportion of million reads. **B)** Gene set enrichment analysis of all groups setting fresh samples as a reference. Significance determined by FDR <0.25. Heat map displays normalised enrichment scores (-1.5-1). N=6 (Fresh), 4 (Fresh + transplant), 3 (NT RNP + transplant), 6 (TOM RNP + transplant) biological replicates per group.

Figure 3.60 demonstrates that transplanted TOM RNP samples clearly clustered separately from fresh and transplanted controls (part **B**) and that this variation was not related to batch (part **A**). Similar to the pattern seen at 24 hours post-nucleofection (**Figures 3.38** and **3.39**), differential gene expression was highest in TOM RNP transplanted samples compared to fresh transplanted counterparts and lowest compared to NT RNP counterparts.

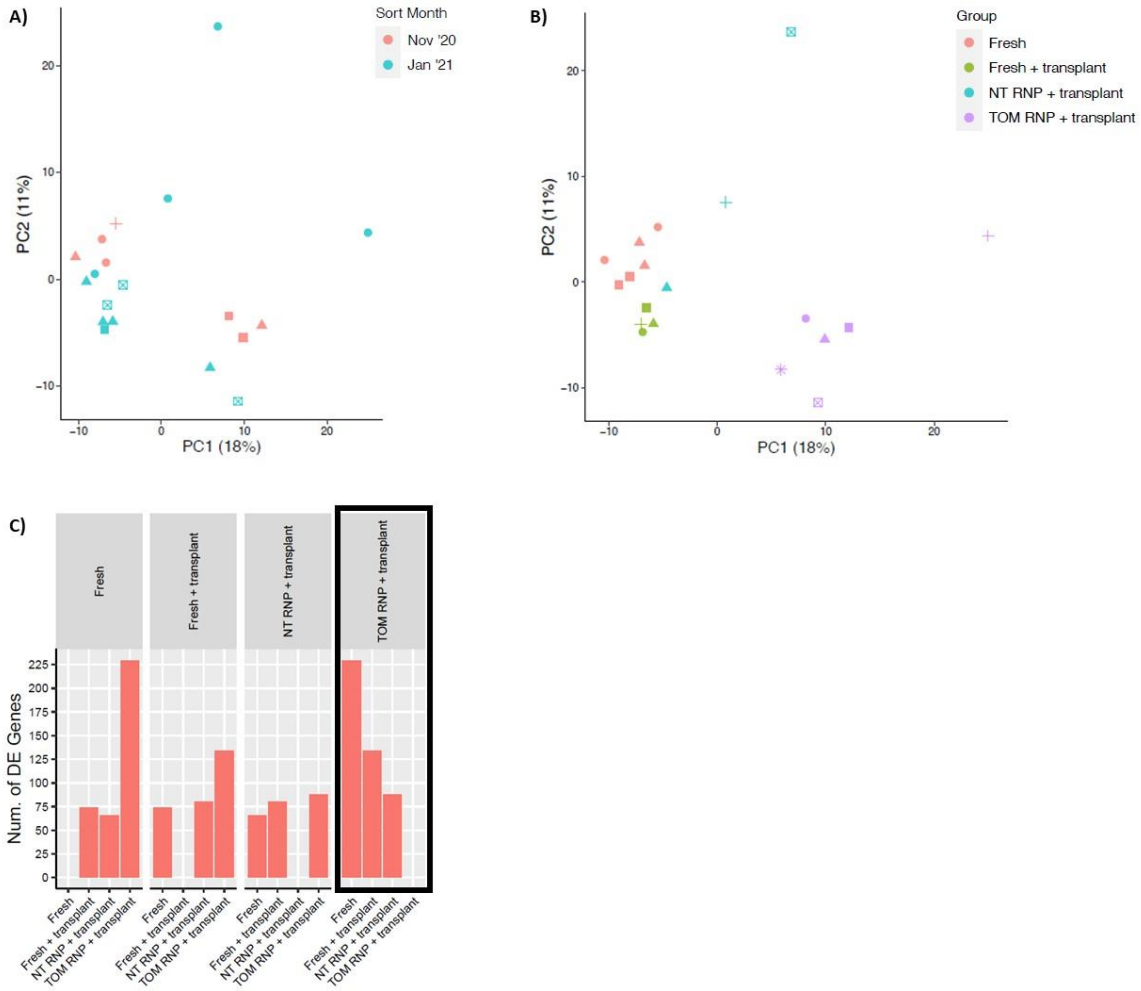


Figure 3.60: Supervised principal component analysis by differential gene expression separated TOM RNP targeted samples from transplant and fresh controls Principal component analysis by differential gene expression of fresh and transplanted samples labelled by **A)** batch **B)** sample. **C)** Log10 (no. of differentially expressed genes) for each condition (bottom x-axis) compared to every other condition as a baseline (top x-axis). Differentially expressed genes in all conditions compared to TOM RNP are outlined in black. N=6 (Fresh), 4 (Fresh + transplant), 3 (NT RNP + transplant), 6 (TOM RNP + transplant) biological replicates per group.

As shown in **Figure 3.61**, differential gene expression was highest in TOM RNP samples compared to fresh un-transplanted controls however there was a small group of genes overlapping between the control groups indicating that these genes were concomitantly up-regulated in TOM RNP samples compared to

multiple control groups. These commonly upregulated genes are labelled in the volcano plots shown in **Figure 3.62**.

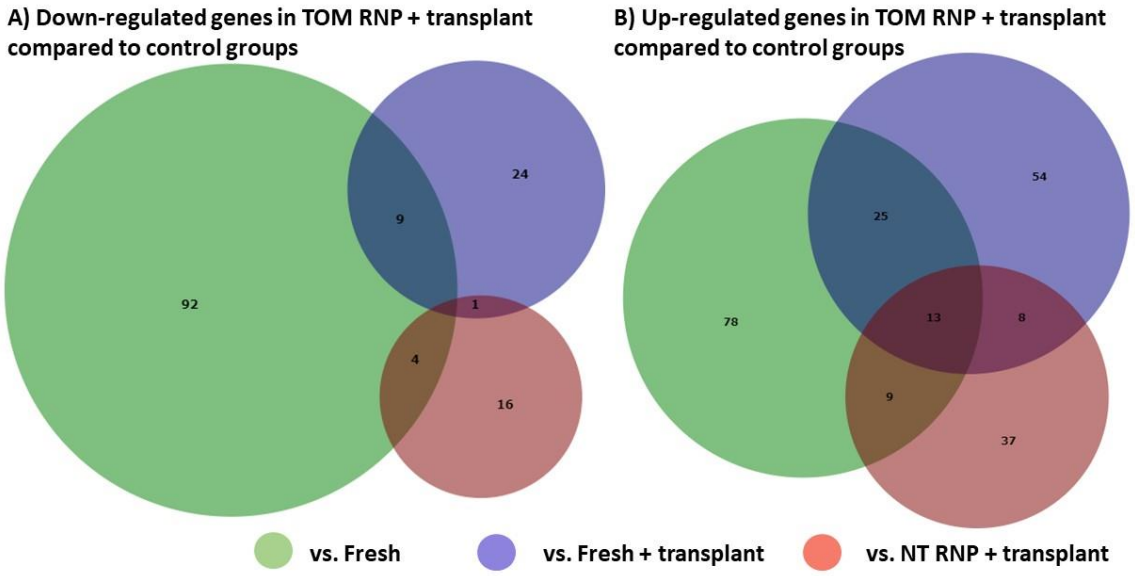


Figure 3.61: Differential gene expression displayed by TOM RNP samples showed some overlap between control conditions Venn diagrams representing overlapping differential gene expression for **A)** Genes down-regulated in transplanted TOM RNP targeted samples compared to controls **B)** Genes up-regulated in transplanted TOM RNP targeted samples compared to controls. N=6 (Fresh), 4 (Fresh + transplant), 3 (NT RNP + transplant), 6 (TOM RNP + transplant) biological replicates per group.

Of the small group of genes which showed overlapping up-regulation in TOM RNP samples compared to at least one control, these included *Cxcr4*, *Clec7a* and *Ms4a1* (**Figure 3.62**). All of these genes have been reported by Capone and colleagues as part of the group of pro-inflammatory markers which are up-regulated in senescent human haematopoietic progenitors (Capone, Connor et al. 2018). Of the two common genes down-regulated in edited samples compared to at least two controls, *Inka1* has recently been reported to be

critical in maintaining the long-term capabilities of stem cells (Kaufmann, Zeng et al. 2021).

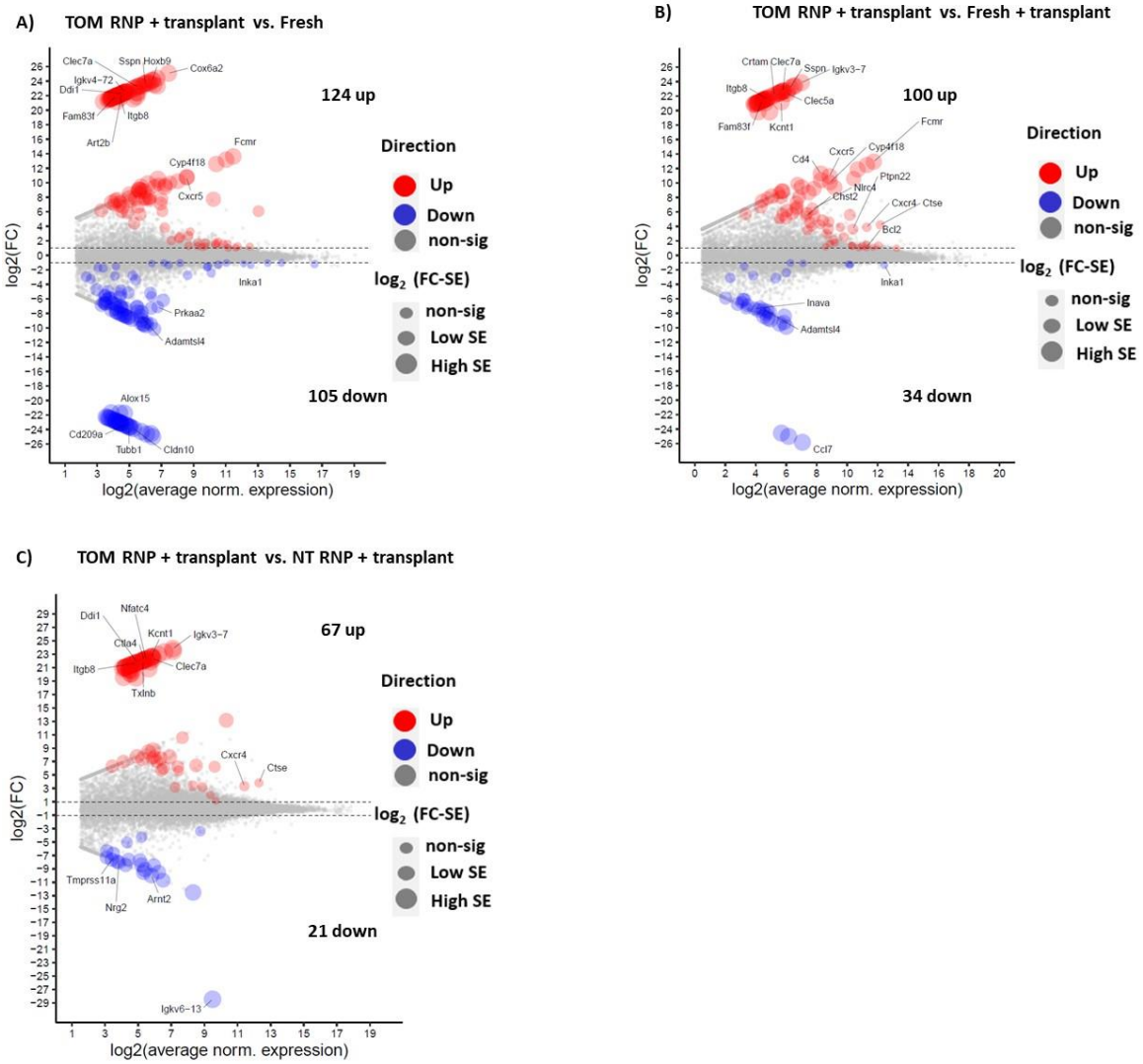


Figure 3.62: Differential gene expression identified transcriptional changes suggesting implications on HSC function Volcano plots showing up- and down- regulated genes in transplanted TOM RNP samples compared to **A)** Fresh **B)** Fresh + transplant **C)** NT RNP + transplant **D)** Plots represent log_2 average normalised expression against log_2 fold change. Up-regulated genes are labelled in red, down-regulated genes are labelled in blue. Dot sizes represent level of standard error. Smaller dots have lower standard error and vice versa. N=6 (Fresh), 4 (Fresh + transplant), 3 (NT RNP + transplant), 6 (TOM RNP + transplant) biological replicates per group.

Of the gene sets which were shown to be enriched in the transplanted TOM RNP samples compared to controls, one which was particularly surprising was allograft rejection which was enriched compared to NT RNP samples (**Figure 3.63**). However when this gene list was interrogated and each gene categorised by other related cell processes, the majority of these genes fell into one of three groups: the innate immune response, the interferon response and negative cell cycle regulation. *Appendix 7* provides an annotated list of the genes in this hallmark set. This mirrored results seen for differential gene expression (**Figure 3.62**) and random forest analysis (**Figure 3.64**) in which genes related to the immune response and cell cycle control were upregulated in transplanted TOM RNP samples. Interestingly, both innate immune response and interferon gamma response genes have been shown to be up-regulated in human HSPCs in a state of senescence (Capone, Connor et al. 2018).

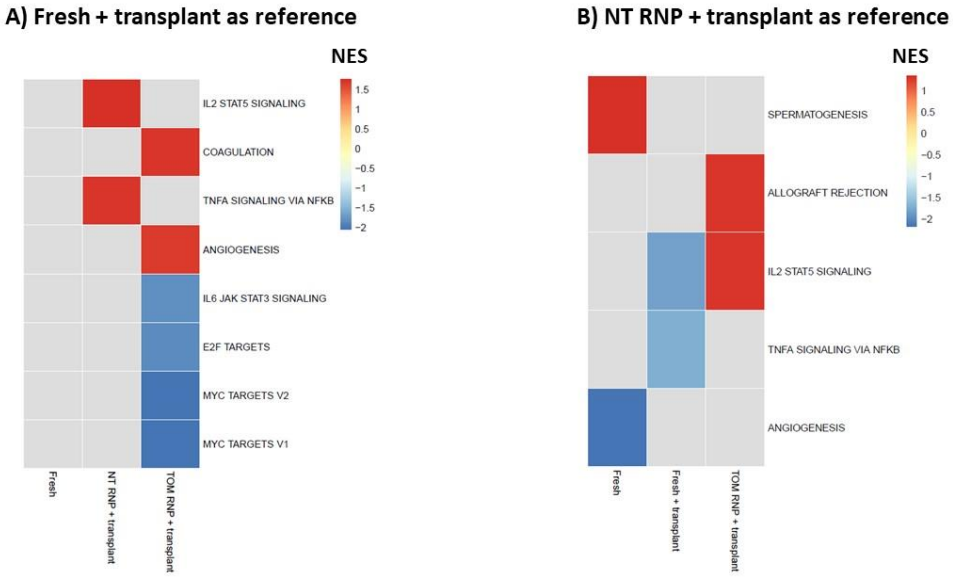


Figure 3.63: Gene set enrichment analysis of transplanted TOM RNP targeted samples compared to sham-edited and fresh transplant controls
Gene set enrichment analysis using hallmark gene sets heat map with reference set as **A)** Fresh + transplant **B)** NT RNP + transplant controls. NES: normalised enrichment score. **Note:** Fresh as a reference is not shown as it did not reveal any significant gene set enrichment. N=6 (Fresh), 4 (Fresh + transplant), 3 (NT RNP + transplant), 6 (TOM RNP + transplant) biological replicates per group.

Random forest analysis (**Figure 3.64**) also revealed upregulation of genes associated with innate immune and stress responses in edited samples (**Figure 3.65 A**). **Figure 3.65 B** illustrates upregulation of multiple genes essential for pluripotency in edited samples, including *Utp6* and *Cops2*. Interestingly, *Clpb1*, involved in mitochondrial mediated innate immunity, and up-regulated in edited samples, has been shown to be up-regulated in HSPCs with deficient megakaryopoiesis (Furusawa, Sadashivaiah et al. 2016). This is particularly interesting as platelet chimerism was seen to diminish over the course of transplantation in edited samples, in contrast to the control samples (see **Figure 3.51** and **Figure 3.52**).

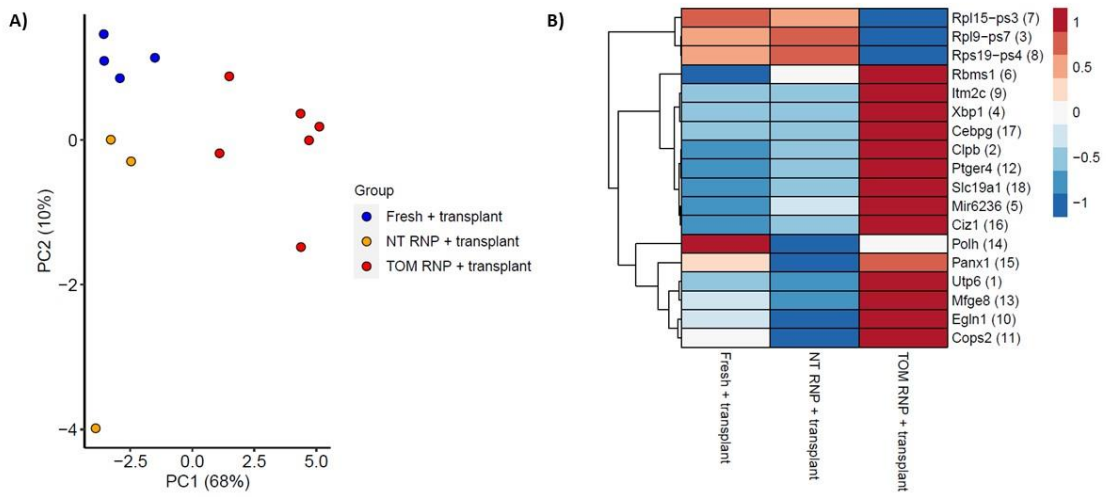


Figure 3.64: Random forest analysis separated transplanted TOM RNP targeted samples from sham-edited and fresh transplanted controls
A) Principal component analysis of transplant groups clustered by differentially expressed genes by random forest. **B)** Heat map of differentially expressed genes between transplant groups by random forest. N=6 (Fresh), 4 (Fresh + transplant), 3 (NT RNP + transplant), 6 (TOM RNP + transplant) biological replicates per group.

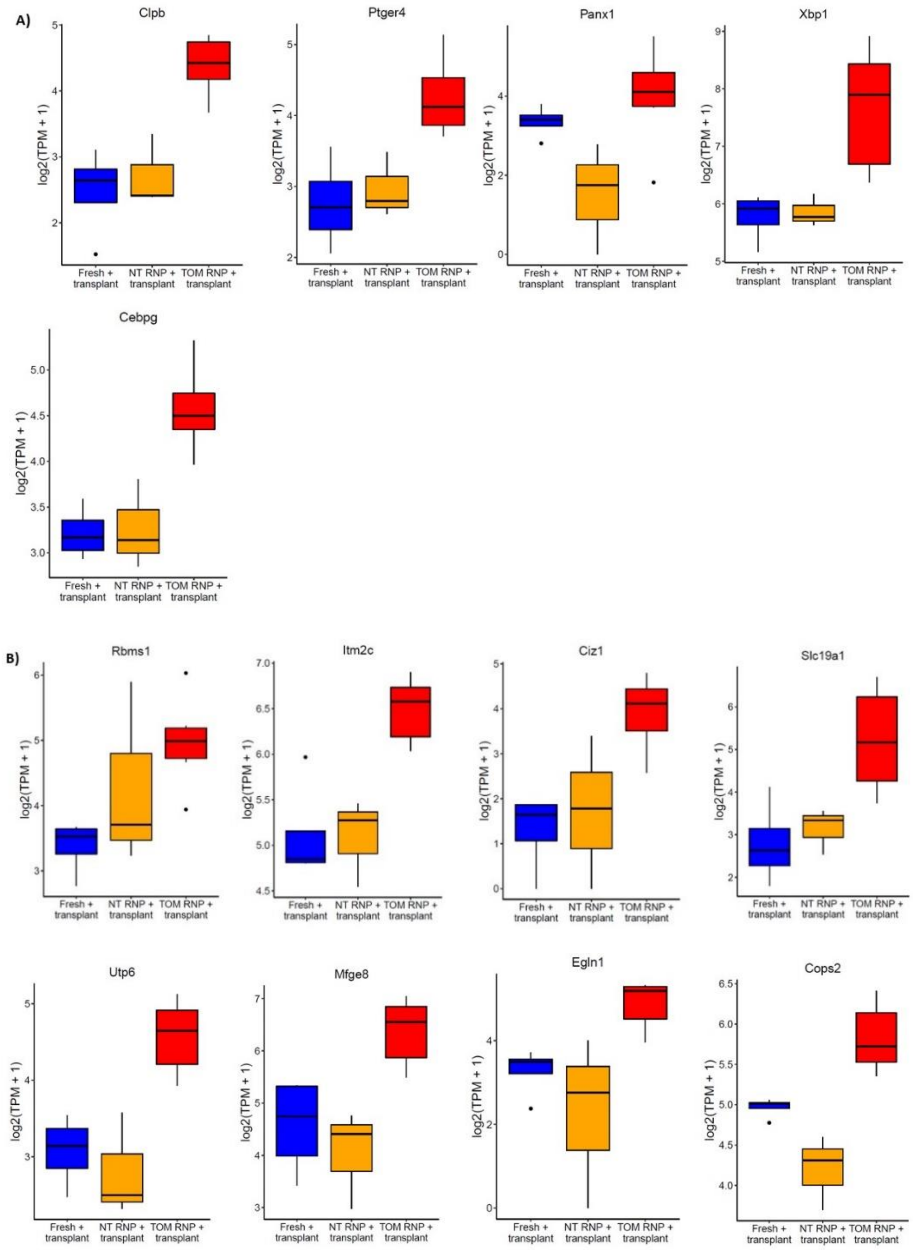


Figure 3.65: Random forest analysis revealed transcriptomic up-regulation of genes associated with innate immune and stress responses, as well as cell cycle control Boxplots of genes related to **A)** innate immune and stress responses **B)** haematopoietic cell cycle regulation and homeostasis up-regulated in TOM RNP samples according to random forest analysis. N=6 (Fresh), 4 (Fresh + transplant), 3 (NT RNP + transplant), 6 (TOM RNP + transplant) biological replicates per group. Boxplots represent expression as log_2 transformed transcripts per million.

Having established that there was evidence of differential gene expression between TOM RNP targeted samples and control groups both 24 hours post-nucleofection and 18 weeks post-transplant, it was hypothesised that a transcriptional signature, which was established at the onset of CRISPR editing, was maintained within the HSC compartment long-term. **Figure 3.66** clearly demonstrates however that this was not the case, as genes commonly up- and down-regulated in TOM RNP samples compared to at least two control groups were not shared between 24-hour post-nucleofection and 18 weeks post-transplant samples. There was no overlap for either differential expression (parts **A** and **B**) or random forest (parts **C** and **D**) analysis. Therefore, although analysis of the immediate and long-term transcriptionally consequences highlighted up-regulation of innate immune/stress associated genes as well as those with potential impacts on stem cell function, these signatures were distinctly separate. While it is possible that the long-term gene expression changes were a consequence of the immediate effects of genome editing, the immediate expression changes were not a permanent signature of edited HSCs.

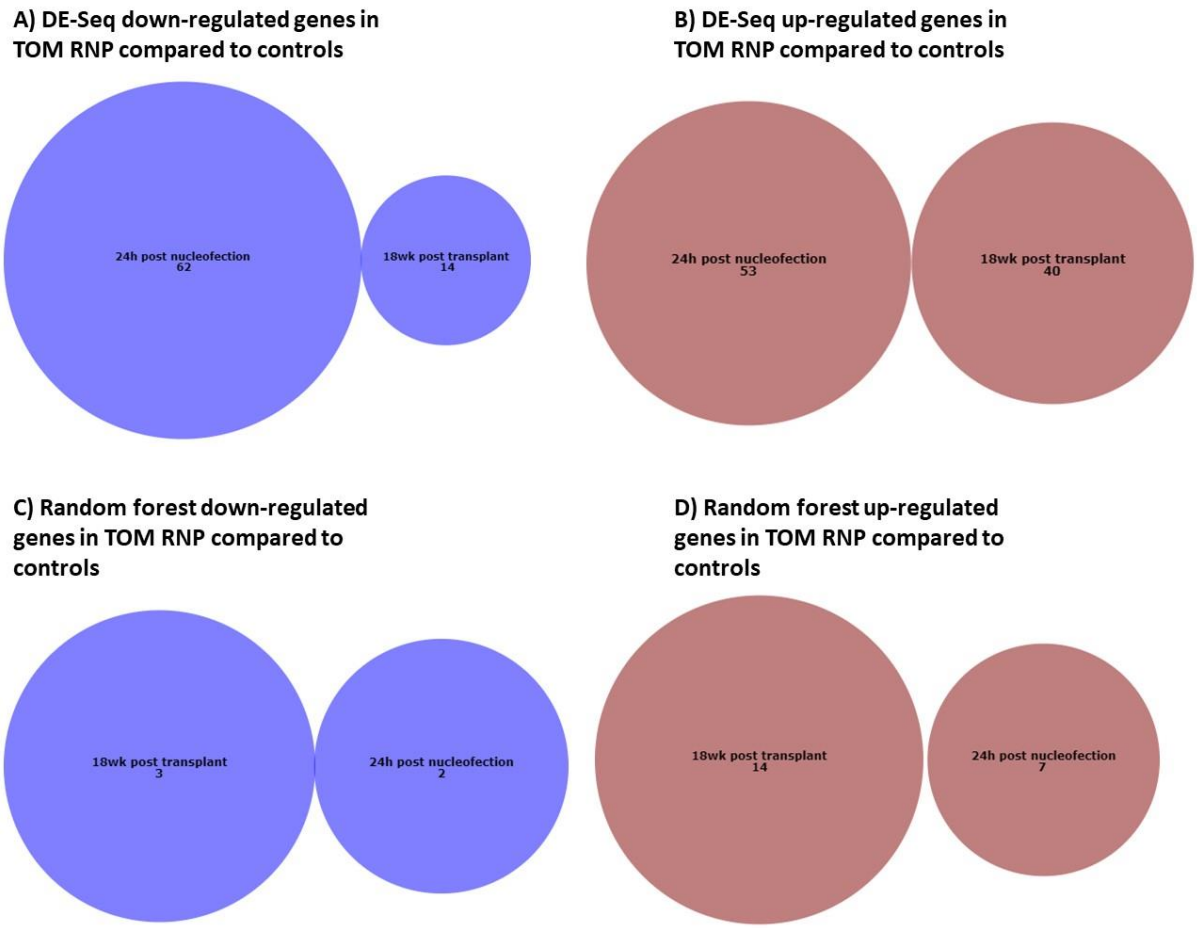


Figure 3.66: Differentially expressed genes in TOM RNP targeted samples 18 weeks post-transplant showed no overlap with differential gene expression patterns 24 hours post-nucleofection Venn diagrams representing 24-hour post-nucleofection and 18-week post-transplant overlap for **A)** down-regulated **B)** up-regulated genes in TOM RNP samples by differential gene expression and **C)** down-regulated **D)** up-regulated genes in TOM RNP samples by random forest compared to controls.

Interestingly, the increased LT-HSC transcriptional signature seen at 24 hours post-nucleofection was also not maintained 18 weeks post-transplant.

In order to determine whether there was a difference in transcriptional HSC scores, LT and ST-HSC signatures (Ivanova, Dimos et al. 2002) were interrogated for each group. A higher HSC score would predict higher long-term re-populating capacity of the stem cells. Therefore it was predicted that, for the edited samples, which showed lower engraftment and reduction in red blood cell

and platelet chimersim over time, there would be a lower HSC score. **Figure 3.67** demonstrates that, although the overall HSC score was highest for edited samples within the transplant group, this difference was not significant and, unlike 24-hour post-nucleofection samples, it was more driven by ST-HSC genes than LT-HSC genes. Indeed, there were no single genes particularly driving this augmented signature compared to 24-hour samples (**Figure 3.45 B**).

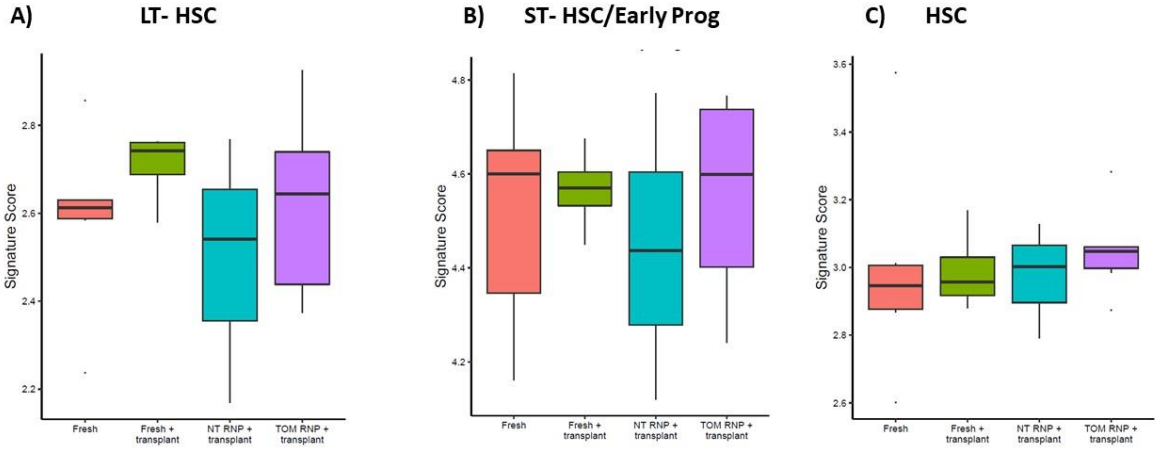


Figure 3.67: HSC signature analysis revealed the LT-HSC signature seen in 24h post-nucleofection was not sustained post-transplant although the overall stem signature was strong in edited HSCs Box plots of **A)** LT-HSC **B)** ST-HSC/Early progenitor **C)** HSC scores for edited samples compared to controls. N=6 (Fresh), 4 (Fresh + transplant), 3 (NT RNP + transplant), 6 (TOM RNP + transplant) biological replicates per group.

3.2.6 Summary

Overall, the aim of this chapter was to interrogate the short-term and long-term consequences of genome editing on HSCs both functionally and transcriptionally. The key questions to be addressed were whether genome editing affected the proportion of cell types within the BM HSPC compartment, whether the cell cycle status of edited HSCs was changed, and whether gene

expression changes occurred as an immediate consequence of genome editing. In the long-term it was hypothesised that genome editing alters the lineage output of HSC in an *in vivo* setting and that a new transcriptional signature established immediately post-editing could be propagated long-term.

By interrogating each HSPC population in the context of the proportion it contributed to the previously gated (parental) population, it was demonstrated that the only population in which there were clear differences between edited and control samples was the LS population within the Lin- fraction of the BM (**Figure 3.30 C**). LS cells contain both MPP and HSC, the most primitive cell types within the HSPC compartment, and in the case of HSC, those with the ability for long-term repopulation *in vivo*. This suggests that there could have been a selective loss of LS cells upon editing or, alternatively, induction of a more quiescent state in which there was less proliferation of cells in this compartment. However, assessment of the cell cycle status of all HSPC compartments did not suggest that LS cells were more quiescent in the edited samples (**Figure 3.32 D**) which points towards the former hypothesis of selective loss.

Interestingly, the cell cycle status was not notably changed in any HSPC compartment assessed, including the HSC (**Figure 3.32**). Despite the lack of difference in cell cycle status between edited and unedited samples, transcriptional analysis of HSC signatures revealed that edited HSCs possessed a higher LT-HSC score, indicative of a stronger level of quiescence but also greater ability for long-term repopulation *in vivo*, than un-edited controls, with up-regulation of genes associated with metabolic quiescence (**Figure 3.45**). It could be that the functional consequence of this transcriptional

change was delayed and that at the 24-hour post-nucleofection time-point, changes seen in gene expression had not yet impacted potential cell cycle changes, due to a delay in changes of protein levels.

What was quite striking from the differential gene expression data was that the majority of the changes in gene expression separated the fresh and cultured only groups from the nucleofection groups (**Figure 3.39**). This suggests that the acute transcriptional changes were predominantly a result of the nucleofection procedure, as otherwise a higher level of gene expression changes would have been seen between N + H_2O, N + NT RNP and N + TOM RNP. However, gene set enrichment analysis revealed that *TP53* and DNA damage responses were more enriched across the nucleofection groups with increasing intervention (**Figure 3.42**) which indicates that although the majority of changes appeared to occur as a result of nucleofection itself, the additional presence of guide RNA + Cas9 and DSB induction had an additive effect on gene expression changes.

In terms of the specific genes that were differentially expressed, at 24 hours, there was a significant enrichment of LT-HSC genes expressed in the tdTomato group (**Figure 3.45**), for example upregulation of *Ak1*, *Hsd11b2* and *Lactb*. All these three genes play key roles in energy metabolism promoting a more metabolically inactive state and reduced cell proliferation (Dzeja and Terzic 2009) (Chapman, Holmes et al. 2013) (Keckesova, Donaher et al. 2017). Interestingly, these genes are also implicated in responses to cellular stress, with *LACTB* acting as a tumour suppressor preventing the proliferation of damaged cells in certain cancers (Keckesova, Donaher et al. 2017). Thus, it could be argued that although the "LT-HSC transcriptional signature" was increased in the tdTomato edited groups, this could have in fact been a

reactionary response to DNA damage as some of the upregulated genes are shared with signatures derived from relatively metabolically inactive LT-HSCs (Ivanova, Dimos et al. 2002) (Keckesova, Donaher et al. 2017).

Strikingly, despite showing minimal transcriptional differences 24 hours post-nucleofection, NT RNP and TOM RNP groups separated from each other much more distinctly in the long-term *in vivo* setting, with NT RNP samples clustering much closer to fresh and fresh transplanted controls than to transplanted TOM RNP samples (**Figure 3.60**). Indeed, there was more overlap in differential gene expression in TOM RNP samples compared to NT RNP and fresh controls at the 18-week analysis time-point compared to the 24-hour time-point (**Figures 3.40** and **3.61**). This suggests that the transcriptional consequences of genome editing may have had more of a long-term impact than the process of nucleofection in the presence of guide RNA/Cas9 with non-targeting groups clustering more closely with control groups post-transplant.

This observation was supported by the functional data in which, although NT RNP samples showed lower overall engraftment than fresh controls (**Figures 3.50** and **3.51**), this reduction was more pronounced in TOM RNP samples and this group also showed a unique negative trajectory of platelet and red blood cell engraftment, compared to the initial four-week level (**Figure 3.52**). Further, at the transcriptional level, there was an upregulation of genes including *Cxcr4, Clec7a* and *Ms4a1* all of which have been reported by Capone and colleagues as part of the group of pro-inflammatory markers which are up-regulated in senescent human haematopoietic progenitors (Capone, Connor et al. 2018). (**Figure 3.62**).

Cxcr4 was also implicated in the reduced hypoxia response associated with senescence in this study. Notably, it has been reported by Nie and colleagues that CXCR4 maintains stem cells in a quiescent state by inhibiting cell cycle progression (Nie, Han et al. 2008). *Clec7a* upregulation was also reported in mouse BM cells induced into a senescent state in response to DNA damage induced by activation of the *MLL-ENL* oncogene (Takacova, Slany et al. 2012). In addition to these genes, both *Ctse* and *Cyp4f18* were found to be upregulated in HSCs at 18 weeks post-transplant in tdTomato edited cells (**Figure 3.62**). These genes have both been reported as part of the signature of cellular deep quiescence, *Ctse* belonging to the group of lysosomal genes upregulated in this state and *Cyp4f18* to the group of oxidative phosphorylation genes (Fujimaki, Li et al. 2019).

Of the two common genes down-regulated in edited samples compared to at least two controls, *Inka1* has recently been reported to be critical in maintaining the long-term capabilities of HSCs (Kaufmann, Zeng et al. 2021) supporting the argument that genome editing has an impact on LT-HSC function. The small number of genes which were shown to be upregulated in edited samples by random forest analysis were also enriched for those involved in the immune response and stem cell senescence (**Figure 3.65**). These included *Xbp1* and *Ptger4* both of which contribute to the immune response group up-regulated in cellular senescence (Capone, Connor et al. 2018). *Xbp1* has been implicated in the ER stress response and its associated pathway has been shown to be upregulated by *Ras* mutations to promote stem cell survival (Liu, Zhao et al. 2019). *Ptger4* has been shown to be a target for transcriptional repression by

ETV6 in pre-B acute lymphoblastic leukaemia, potentially playing a role in leukemic cell survival (Boily, Larose et al. 2007).

A large portion of genes found to be also upregulated in tdTomato edited samples by random forest were those involved in cell cycle control (**Figure 3.65**). These included *Mfge8* and *Egln1*. Conversely, there was also an upregulation of essential genes required for pluripotency including *Utp6* and *Cops2* (**Figure 3.65**).

Although there is little evidence for the role of *Mfge8* in HSCs, it has been reported to maintain neural stem cells in a quiescent state by inhibiting their activation through the PI3K-Akt pathway (Okazaki and Gotoh 2018). *Egln1* encodes the protein PHD2 which plays an important role in stem cell maintenance. It has been shown that PHD2 knockout induces self-renewal in MPPs but not HSCs (Singh, Franke et al. 2013). Given that it was upregulated in HSCs at 18 weeks post-transplant, it could be predicted that this was maintained through to MPPs and could partially explain the lower level of this population in proportion to HSCs in edited compared to control samples (**Figures 3.54-55**). Further, it has also been shown that PHD2 knockout leads to elevated erythroid levels in adult mice (Takeda, Aguila et al. 2008). This is consistent with the evidence of loss of erythroid chimerism in edited samples in which HSCs overexpressed *Elgn1* (**Figure 3.65**).

Taken together, this transcriptional profile unique to long-term edited HSCs suggests reactionary gene expression changes which, although they didn't correspond to a specific subtype of HSCs, showed overlap with signatures of stem cell quiescence and senescence. This is similar to studies of other forms of

acute DNA damage. Wang and colleagues demonstrated that full body ionising radiation led to reduced numbers of LSKs, reduced PB lineage chimerism and gene expression characteristic of senescence in sorted LSKs (Wang, Schulte et al. 2006).

The “LT-HSC signature" was not significantly enriched in edited HSCs 18 weeks post-transplant (**Figure 3.67**) although it is possible that this was because of the greater variability between technical replicates. What was striking from this transcriptional data however was that, despite a lack of enrichment for the "LT-HSC signature", a high proportion of upregulated genes in HSCs 18 weeks post-transplant were enriched for those involved in cellular quiescence and senescence (**Figures 3.62 and 3.65**).

In conclusion, this chapter has revealed there were potentially important long-term functional consequences of genome editing on HSC output, as demonstrated by marked reduction in red blood cell and platelet chimerism over time, although this would require secondary transplantation to confirm a long-term stem cell defect. Further, these functional differences were also associated with transcriptional changes most highly enriched for genes reported to be up-regulated in states of senescence. However, acute transcriptional changes were distinct and separate from gene expression patterns that separated edited samples in the long-term.

3.3 Introduction of the JMML associated *NF-1* loss of function mutation in human cell lines and primary haematopoietic cells

As described in *Introduction section 1.5*, the broad aim of this thesis was to gain an in-depth understanding of the CRISPR approach to model normal and malignant haematopoiesis. Having interrogated the effect of the genome editing process in a normal murine haematopoietic setting using NHEJ based targeting of the tdTomato reporter, I then studied the application of the CRISPR approach to model malignant haematopoiesis in a human haematopoietic setting. JMML is a rare childhood leukaemia that has been studied widely in our lab. In this chapter of my thesis, I wanted to take advantage of the CRISPR NHEJ approach to allow introduction of a loss of function mutation in normal CB primary human samples to model disease development from the point of initiation in a fetal-like context.

3.3.1 Experimental approach

The aim of this chapter was to use a CRISPR targeting approach to model JMML *in vitro* and *in vivo*, by targeting one of the initiating mutations –*NF-1*– whose loss of function is associated with predisposition to the rare childhood leukaemia, JMML. As detailed in *Introduction section 2*, various murine models currently exist for this disease however they lack certain features, including introduction of initiating events in primitive haematopoietic cells within a human cell context. This means that although they recapitulate the myeloid expansion and GM-CSF hypersensitivity characteristic of JMML, they do not provide insight into how the leukaemia develops from the initiating event.

To address this question, the spectrum of *NF-1* mutations in patients was assessed and guide RNAs were designed using the Synthego design tool (Thomas, Parry-Smith et al. 2019) against exon 2 within the *NF-1* gene. This approach was chosen such that the edit would occur close to the transcriptional start site, increasing the chances of an early premature stop codon and loss of protein translation. As described in *Methods section 2.4*, multiple guide RNAs were tested in the human erythroleukaemia cell line –TF-1– to determine the efficiency of editing and confirm loss of protein function as well as cytokine sensitivity. The TF-1 cell line was chosen for its intact *NF-1* gene, as well as GM-CSF cytokine dependence; this meant it could be used to assess whether the single edit was sufficient to render the cells cytokine independent. Guide RNAs which proved efficient at the level of editing, loss of protein expression and the bestowing of cytokine dependence were taken forward into human CB HSPCs. CD34+ bead selected human CB HSPCs were targeted with the optimal *NF-1* guide RNAs, followed by single cell cloning in minimal GM-CSF concentrations to determine whether the independence seen in cell lines was reproducible in primary cells. Finally, to determine whether a compound heterozygous loss-of-function mutation in *NF-1* alone was sufficient to cause a JMML-like disease in primary human haematopoietic cells, CD34+ CB HSPCs, edited at the *NF-1* locus, were transplanted into sub-lethally irradiated NSG recipients.

3.3.2 *NF-1* guide RNA selection

i) Patient mutations and guide RNA design

In order to determine the most appropriate *NF-1* gene region to target, the spectrum of initiating mutations was assessed from the two largest published series of JMML patients where genomic analysis was performed. **Table 3.1** and **Figure 3.68** summarise the data for *NF-1* initiating mutations compiled from these two in-depth studies of the mutational landscape of JMML (Stieglitz, Taylor-Weiner et al. 2015) (Caye, Strullu et al. 2015).

Exon/Intron	Nucleotide position	Protein position	Co-operating mutations
Exon 18 Exon 46	29553477 c.2026insC 29665757 c.6855C>A	p.T676fs p.Y2285*	*NF-1* ASXL1 SH2B3 PTPN11 JAK3
Exon 18	c.2033dupC	p.Ile679Aspfs*21	
Exon 9	29527569 c.1015delCT	p.N339fs	SETBP1 ASXL1 GATA2
Exon 12	29533315 c.1318C>T	p.R440*	PTPN11 SH2B3
Exon 12 Exon 28	29533330 c.1333G>T 29562746 c.3826C>T	p.E445* p.R1276*	EZH2 RUNX1
Intron 11-12	29533315 c.C1318T	p.R440*	PTPN11 SH2B3
Exon 11 Intron 32.33	29528489 c.1246C>T 29586048 c.e33	p.R440* p.R416	
Intron 14-15	29546175 c.2041C>T	p.R681*	SETBP1
Exon 16	c.1756_1759delACTA	p.Thr586Valfs*18	
Exon 28	c.3721C>T	p.Arg1241*	
Exon 29 Intron 11-12	29562981 c.C3916T 29533315 c.C1318T	p.R1306* p.R440*	PTPN11 SH2B3
Exon 29	c.3892C>T	p.Gln1298*	CBL ASXL1
Exon 32	29585473 c.4285insA	p.K1429fs	CBL
Intron 34-35	c.4577+1G>C	p.? (splice effect?)	ASXL1 SETBP1
Exon 38 Intron 53-54	c.5283delT c.7869+5G>A	p.Val1762Serfs*3 p.? (splice effect?)	ASXL1
Exon 40	c.5959C>T	p.Gln1987*	ZRSR2
Exon 45	29665110 c.C6772T	p.R2258*	
Exon 50	c.7420_7421delGA	p.Asp2474Tyrfs*8	NRAS

Table 3.1: *NF-1* germline initiating mutations genomic and protein regions. If additional mutations are present in at least one patient, these are detailed in the far right column. Mutations are listed in order of frequency. Corroborated from in-depth sequencing data generated by Stieglitz, Caye and colleagues (Stieglitz, Taylor-Weiner et al. 2015) (Caye, Strullu et al. 2015).

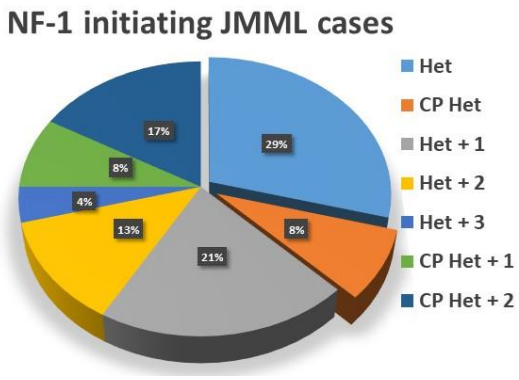


Figure 3.68: Frequency of additional events in *NF-1* initiating JMML
Patients with *NF-1* initiating mutations subdivided into categories depending on mutation type and additional events: Heterozygous only (Het), Compound heterozygous only (CP Het), + additional corroborating mutations (+ n).

The data in **Table 3.1** and **Figure 3.68** show a wide spectrum of *NF-1* mutations, as well as the number and type of co-operating mutations. 37% of *NF-1* mutated patients had no detectable additional mutations (29% heterozygous, 8% compound heterozygous) and the least common event was heterozygous *NF-1* with three additional mutations (4%). What was common, however, between the mutations was that they caused loss of NF-1 protein or truncation. Therefore, taking into account the wide spectrum of mutations but common translation outcome, guide RNAs were designed with the highest probability of inducing loss of protein function. Given the large size of *NF-1*, and presence of multiple transcript isoforms, an early exon, common to all transcripts was targeted with the aim of introducing a premature stop codon as close to the transcriptional start site (TSS) as possible. **Figure 3.69** summarises the targeted *NF-1* region and guide RNAs designed.

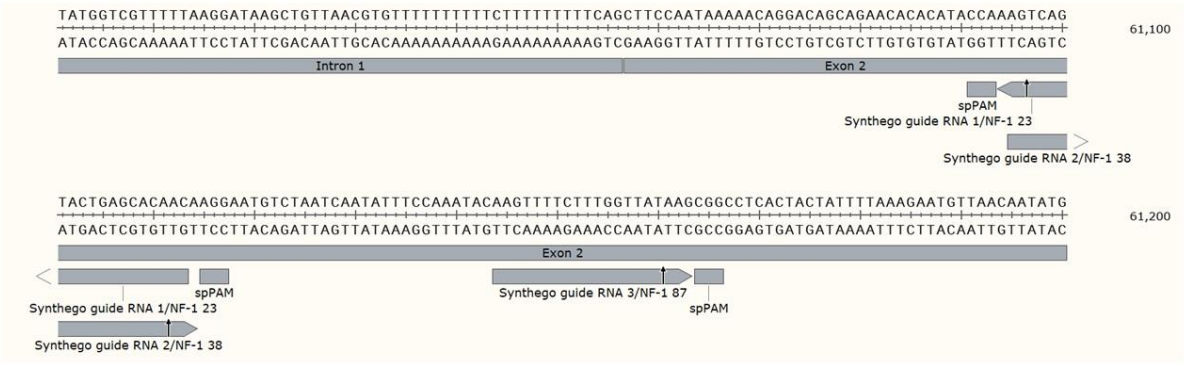


Figure 3.69: *NF-1* exon 2 region guide RNAs Plot shows sequence of flanking intron 1 and exon 2, with sequences labelled in grey of three guide RNAs designed. PAM (5'-NGG-3') are labelled spPAM, required for Cas9 binding, but not included in the guide RNA sequence.

3.3.3 TF-1 cell line selection and nucleofection optimisation

Given that the TF-1 cell line requires GM-CSF for optimal growth, it was hypothesised that growth would be considerably restricted in the absence of GM-CSF; and that *NF-1* editing could rescue this. TF-1 cells were therefore grown with and without GM-CSF, and counted at regular intervals, to determine the level of growth restriction induced by the absence of GM-CSF. Given the previous optimisation results (*Results section 3.1*), the RNP nucleofection approach was also chosen for this study. Due to the absence of a commercially optimised program for this particular cell line, experiments were carried out to test a range of conditions and select a program. Using the matrix detailed in *Appendix 5*, the centre column programs were first tested after which program optimisation by row was carried out to fine tune the balance between cell viability and efficiency. Initially for column optimisation a single buffer (SG) was tested. For further fine tuning by row, for which there was a more limited number of programs, both SG and SF buffers were tested. **Figures 3.70-3.71** summarise the results.

In the absence of GM-CSF, TF-1 cells proliferated very slowly and significantly less than in the presence of GM-CSF (**Figure 3.70 A**, p=0.0321). For the first round of optimisation, using the centre column of the matrix (*Appendix 5*), the two programs which gave the highest viability post-nucleofection were DN-100 and EO-100 (**Figure 3.70 B**) although this wasn't statistically significant. In terms of total number of GFP positive cells, which took into account both the viability and efficiency (% GFP positive), DN-100 gave the highest average number, although this was also not statistically significant. This was most likely due to the high variability and limited number of technical replicates.

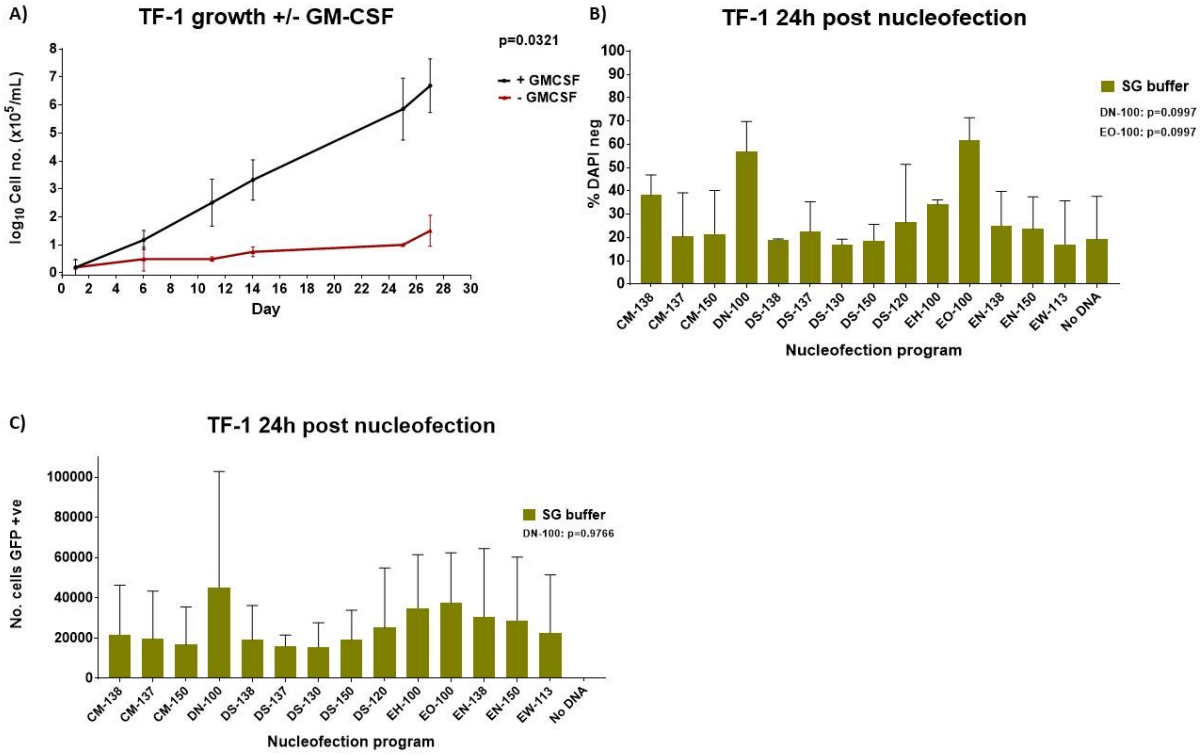


Figure 3.70: GM-CSF significantly reduced TF-1 growth. DN100 was the optimal column-optimised nucleofection program for TF-1 cells in SG buffer A) TF-1 cells grown in RPMI supplemented with 10% FCS, 1% Penicillin/Streptomycin, and 1% sodium pyruvate with/without 5ng/mL GM-CSF. 24h post-nucleofection column matrix optimisation: **B)** Percentage DAPI negative **C)** Estimated no. of cells GFP positive. N=2 biological replicates. Significance in **A)** calculated using paired parametric t-test. Significance in **B)** and **C)** calculated using a one-way ANOVA setting the labelled condition as the control.

Having chosen DN-100 as the optimal program by column optimisation, all programs within the same row as DN-100 were tested with two buffers: SG and SF, to fine tune the optimal conditions. As shown in **Figure 3.71 A**, for all conditions with the exception of the no DNA control, SG buffer resulted in higher post-nucleofection viability than SF buffer (p=0.0013). In contrast, for most conditions, SF buffer resulted in higher total numbers of GFP positive cells (**Figure 3.71 B**, p=0.0002) indicating that the reduced viability caused by this buffer was outweighed by the increased transfection efficiency. Therefore SF was

chosen as the optimal buffer. For the SF buffer, nucleofection programs DA-100, DN-100, EH-100 and FF-120 all resulted in high levels of GFP positive cells, therefore it could be concluded that all four would have been viable options for downstream TF-1 transfection. DN-100 was chosen due to having the highest consistency between replicates.

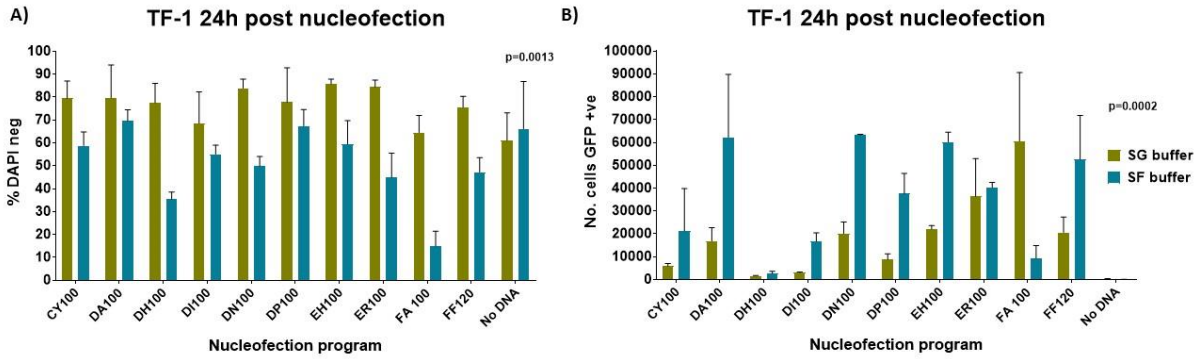


Figure 3.71: SG buffer maintained higher viability but SF buffer gave higher transfection efficiency in nucleofection row-optimisation for TF-1 cells 24h post-nucleofection row matrix optimisation: **A)** Percentage DAPI negative **B)** Estimated no. of cells GFP positive. Significance between SG and SF buffers was calculated using a two-way RM ANOVA.

3.3.4 Establishment of GM-CSF independent TF-1 clones edited at the *NF-1* locus using the optimised SF DN-100 nucleofection program

The optimised program was then used to compare the editing efficiency for the three *NF-1* gRNAs and to test whether the absence of GM-CSF could select for *NF-1* edited cells. The *AAVS1* gRNA, targeting the safe harbour viral integration intronic region, was used as a non-targeting control. Guides 1 and 2 were significantly more efficient than guide 3 or all three guides in combination (**Figure 3.72**); and growing the cells in the absence of GM-CSF selected for *NF-1* edited

cells, although this did not reach statistical significance at the bulk level (**Figure 3.72**).

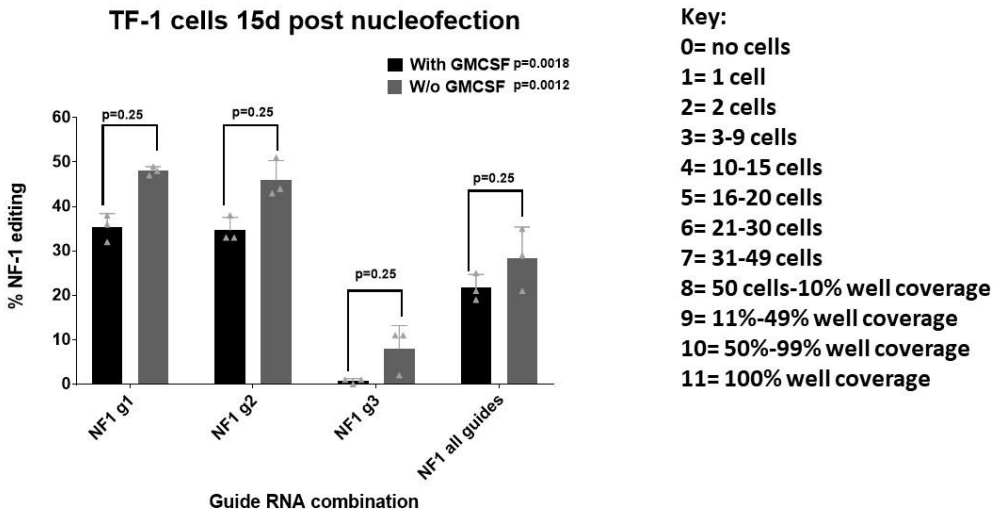


Figure 3.72: NF-1 guides 1 and 2 were significantly more efficient than guide 3 or triple guide combination Percentage NF-1 exon 2 editing. N= 3 technical replicates. Data represent mean with standard deviation. Significance with/without GM-CSF calculated using Wilcoxon test. Significance between guide combinations calculated using Kruskal-Wallis test.

To establish the cell numbers and conditions to best assess the impact of *NF-1* editing on TF-1 cells, 1-500 TF-1 cells were then sorted into conditioned or non-conditioned medium (both with 5ng/mL GM-CSF). Conditioned medium was prepared from the filtrate of cultured TF-1 cells two days post medium change and unconditioned medium was prepared with fresh reagents in the absence of any prior cell exposure. Cells sorted into these two medium types were grown for 22 days and the colony size was assessed by light microscopy and manual counting (**Figure 3.73**, colony counting key depicted in **Figure 3.72**). This showed that Terasaki plates with non-conditioned medium were the most efficient for producing clones from low numbers of sorted cells (**Figure 3.73**). As this was

the only condition in which clones greater than 10 cells were produced from a single cell, this condition was used to grow edited clones.

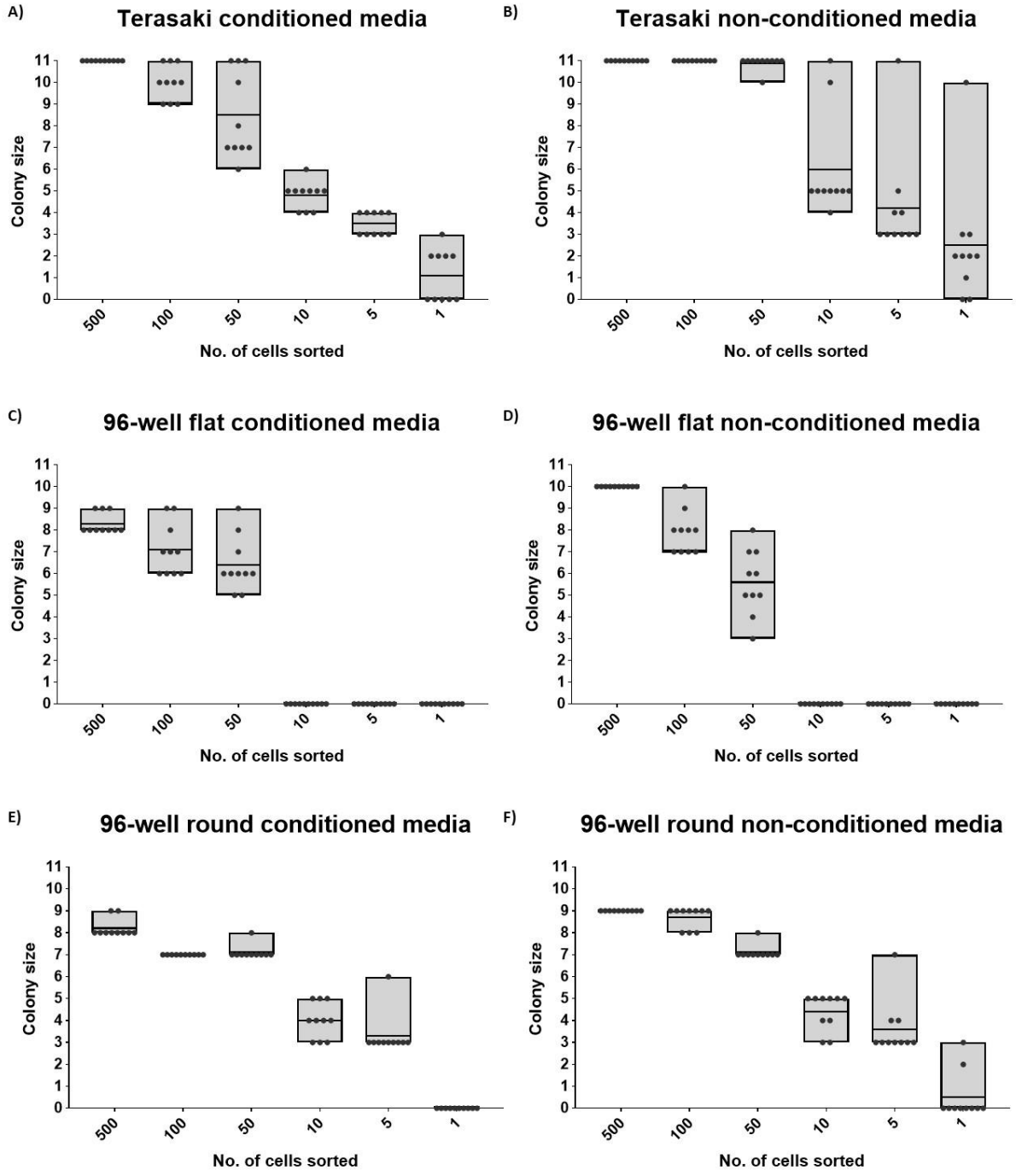


Figure 3.73: Single cell TF-1 colony growth was supported in Terasaki non-conditioned medium Size of colony produced when sorted in **A)** Terasaki conditioned **B)** Tersakai non-conditioned **C)** 96-well flat conditioned **D)** 96-well flat non-conditioned **E)** 96-well round conditioned **F)** 96-well round non-conditioned. N=10 technical replicates. Bars show mean with individual colonies plotted, no error bars shown due to data taking discrete form (non-continuous, numbers on the y-axis don't indicate increases by equal proportions).

Using these optimised colony growth conditions, *NF-1* or *AAVS-1* targeted TF-1 cells were then sorted to assess their clonogenic output. *AVVS-1* is a safe harbour intronic region in which DSB induction should not have any consequences other than the DNA damage caused at the site. In contrast to chapter 2, in which a non-targeting guide RNA was used as a control for the presence of Cas9 and a guide RNA without a DSB, it was more important in the context of modelling cancer using CRISPR that it was possible to conclude with confidence that any phenotype seen was a direct consequence of causing a DSB in a tumour suppressor (in this case *NF-1*) and not simply from the presence of DNA damage. Therefore, DSB induced DNA damage at the *AAVS-1* safe harbour site was used; if no proliferative advantage or disease development was seen for this group it would be possible to interpret the *NF-1* results with confidence that the site of DSB induction was the cause of potential phenotypes. To account for the low clonogenicity from single cells seen in the unedited TF-1 cells (**Figure 3.73**), the edited TF-1 cells were sorted at 5 or 1 cell(s) per well. **Figures 3.74-75** summarise the results.

Figure 3.74 clearly demonstrates an increased clonogenic potential was bestowed by *NF-1* editing both in the presence or absence of GM-CSF, although it appears that single cell sorting combined with no GM-CSF conditions was too stringent a selective pressure and under these conditions there was no significant difference between *AAVS1* and *NF-1* edited samples (**Figure 3.74 F**). In the 5-cell sort condition, both *NF-1* guides caused TF-1 cells to produce significantly more colonies than *AAVS1* edited, and this was more marked when selective pressure was introduced by removing GM-CSF from the medium (**Figure 3.74 E**). This correlates with the data at the bulk level in which the percentage edited

cells increased within the population in the absence of GM-CSF (**Figure 3.72**). In the single cell sort condition only *NF-1* guide 2 produced significantly more colonies than *AAVS1*, and this was only seen in the with GM-CSF condition. However, it should be noted that only the *NF-1* edited samples produced colonies from single sorting into medium without GM-CSF, indeed no colonies were produced under this high selection pressure for *AAVS1* targeted samples (**Figure 3.74 F**).

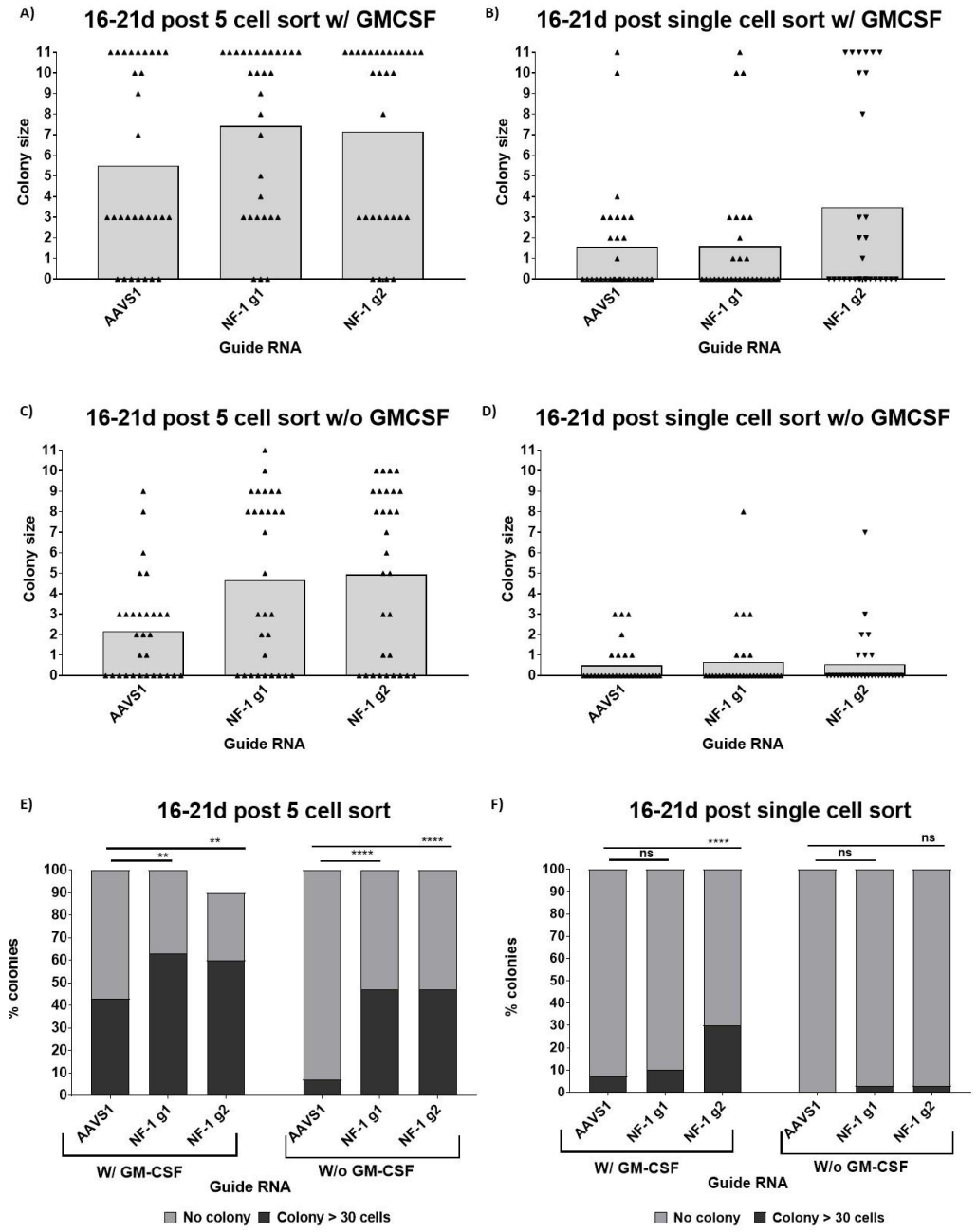


Figure 3.74: *NF-1* targeting increased the clonogenicity of TF-1 cells, particularly in the absence of GM-CSF Size of colonies produced for **A)** 5 cells with GM-CSF **B)** 1 cell with GM-CSF **C)** 5 cells without GM-CSF **D)** 1 cell without GM-CSF. Data represent individual discrete data points, bars represent mean. Percentage colonies above 30 cells produced for **E)** 5 cells sorted **F)** 1 cell sorted. N=30 wells per guide/condition from 3 biological replicates. Significance in **E)** and **F)** calculated using Fisher's exact test.

Colonies grown with GM-CSF in Terasaki plates were then propagated as clones and confirmed for editing at the *NF-1* locus prior to further *in vitro* study.

NF-1 protein expression was assessed in order to determine whether the exon 2 edit was sufficient to prevent translation. Clones were also grown in the absence of GM-CSF to assess whether the *NF-1* edit bestowed cytokine independence without prior exposure to GM-CSF absent conditions.

Figures 3.75 and 3.76 confirm that the colonies produced were unedited at the *NF-1* locus for *AAVS1* targeted samples and *NF-1* targeted colonies were edited at 85% or above, even for colonies produced from 5 cells sorted. This suggests that either all 5 cells received the same edit, or more probable, that a single cell within the group dominated the colony growth.

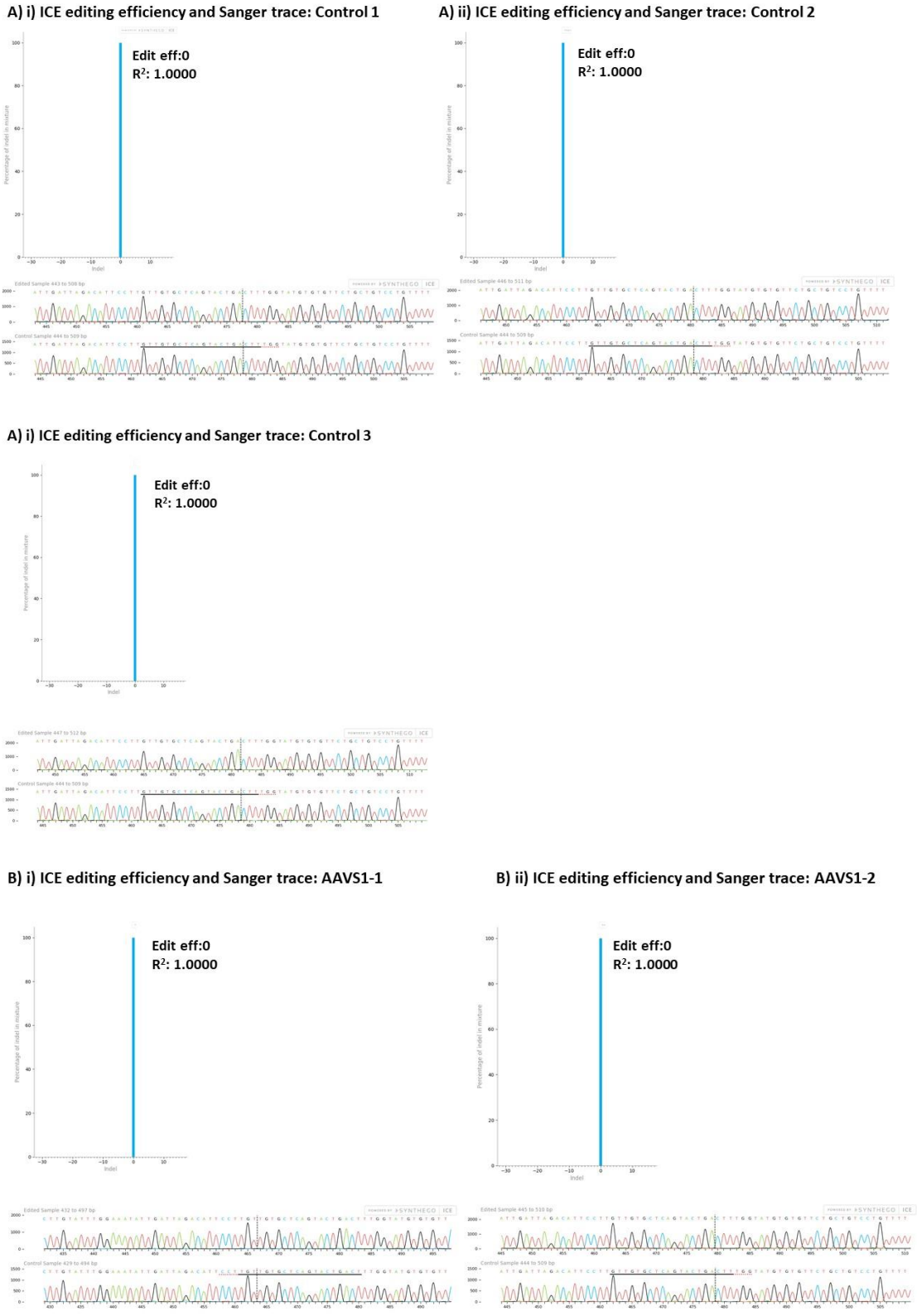


Figure 3.75: TF-1 control and AAVS1 non-targeted clones maintained an intact *NF-1* locus Sanger ICE plots and traces for *NF-1* PCR on TF-1 cells (21d post cell line growth in GM-CSF) used for proliferation assessment **A)** Controls **i)** 1 **ii)** 2 **iii)** 3 **B)** *AAVS1* 5 cell clones **i)** 1 **ii)** 2. N=2-3 biological replicates.

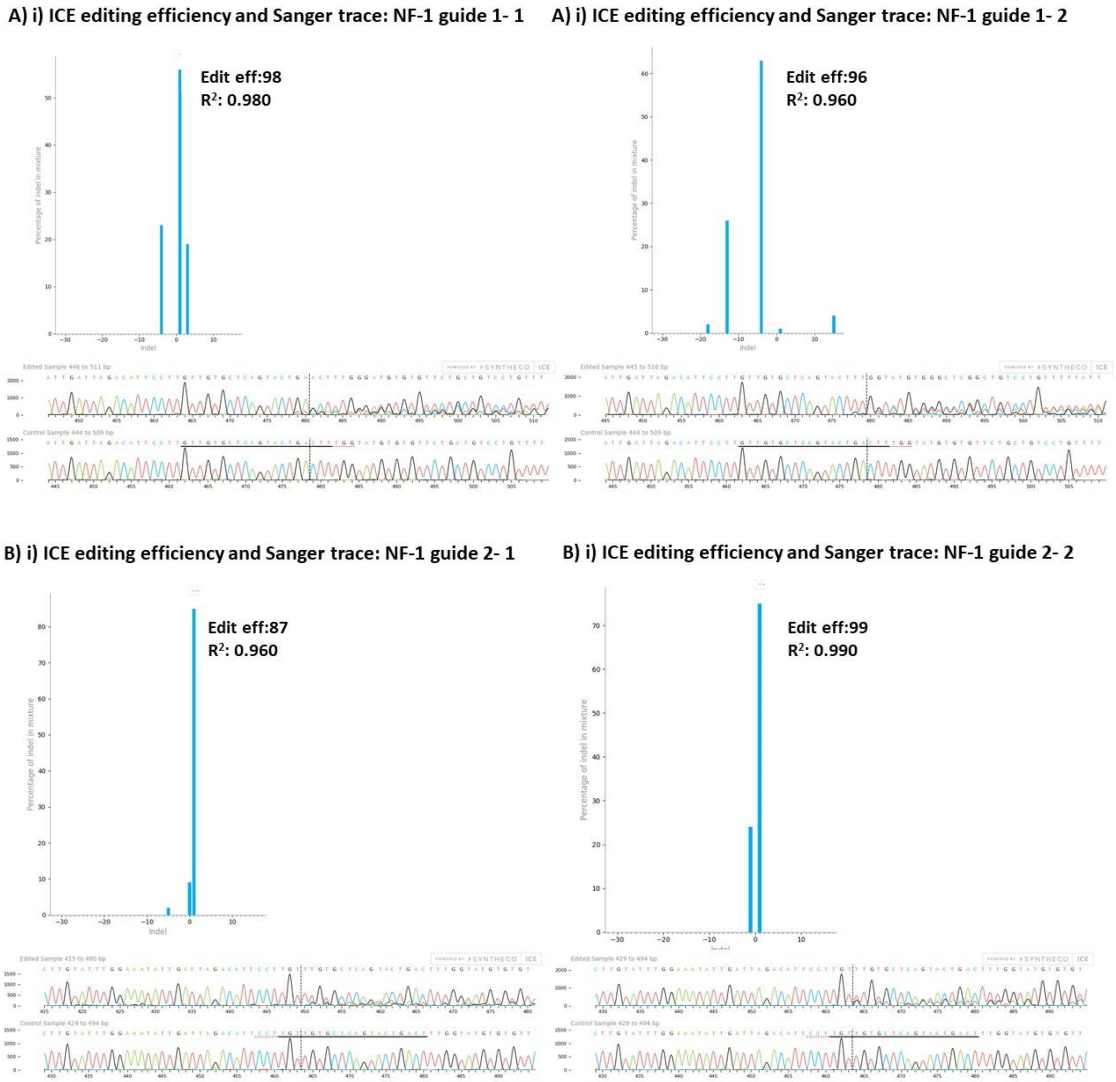


Figure 3.76: *NF-1* edited clones harboured 87-99% editing at the *NF-1* locus Sanger ICE plots and traces for *NF-1* PCR on TF-1 cells (21d post cell line growth in GM-CSF) used for proliferation assessment. **A)** *NF-1* guide 1 5 cell clone **i)** 1 **ii)** 2 **B)** *NF-1* guide 2 single cell clone **i)** 1 **ii)** 2. N= 2 biological replicates.

Figure 3.77 confirms complete loss of NF-1 protein in all samples except *NF-1* g1-1 which retained some expression, although not as high as the controls. It is known that there are multiple transcript isoforms of *NF-1* (Chen, Kostadima et al. 2014), therefore it could be possible that an alternative transcript was still able to be translated despite the introduced edit. Beta-tubulin and GAPDH internal

controls confirm equal protein loading, meaning differences in NF-1 expression levels were true biological differences.

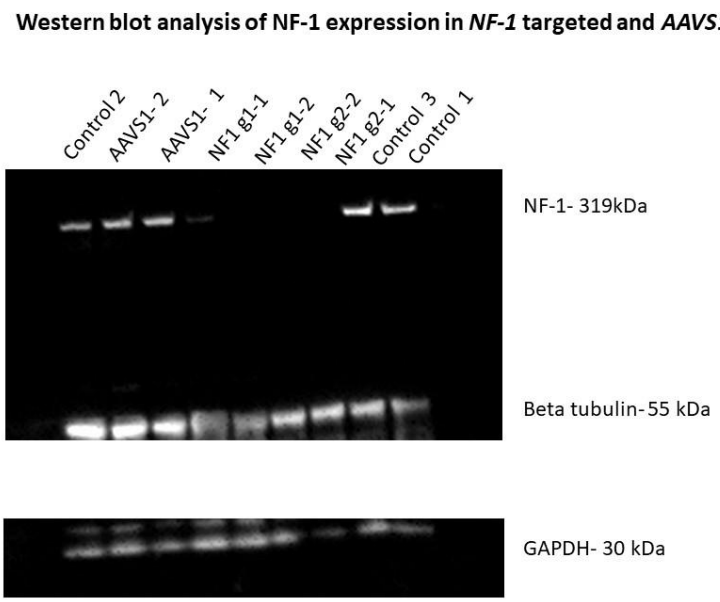


Figure 3.77: *NF-1* editing caused loss of protein expression in targeted clones Western blot probing for NF-1 in TF-1 cells (21d post cell line growth in GM-CSF) used for proliferation assessment. Beta-tublin and GAPDH used as loading controls. N=2-3 biological replicates.

Cell lines grown up from these clones showed similar proliferation ability in the presence of GM-CSF, whereas *NF-1* guide 1-2 and guide 2-2 clones showed significantly higher proliferation in the absence of GM-CSF compared to control 3 (**Figure 3.78 A) ii)**, p= 0.0273) and this proliferation level was equivalent to the control level in the presence of GM-CSF (**Figure 3.78 B**). This indicates that *NF-1* editing at exon 2 was sufficient to cause loss of protein expression and bestow GM-CSF independent growth on TF-1 cells. Interestingly, *NF-1* guide 1-1 and guide 2-1 clones didn't show GM-CSF independent growth despite NF-1 protein full or partial loss. It is possible that *NF-1* editing produces an environment which facilitates further genetic modifications, which occurred at a level enough to

induce GM-CSF independence in some, but not all, clones. Whole genome sequencing of these cell lines would be required to determine whether certain clones acquired additional genetic hits, as is often the case in patients.

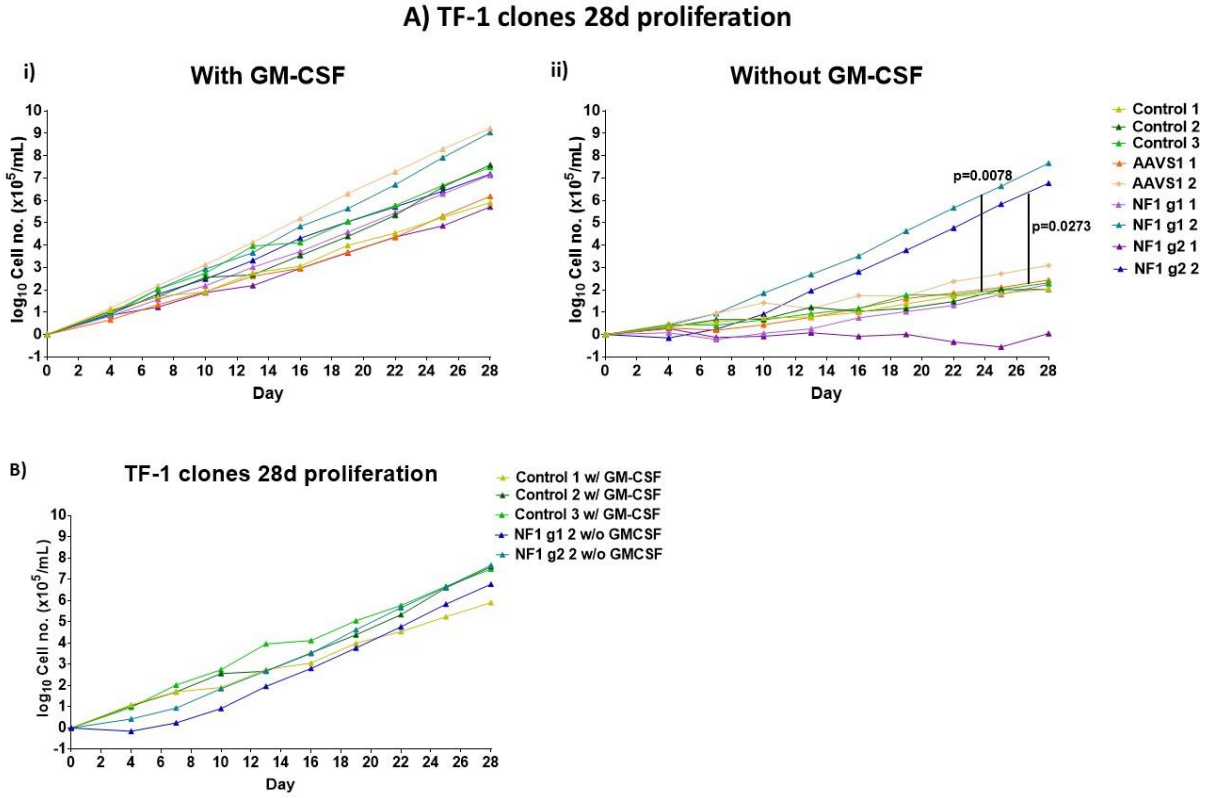


Figure 3.78: *NF-1* editing caused GM-CSF independent proliferation in targeted clones Proliferation of edited cell lines for 28d in TF-1 non-conditioned medium **A) i)** with GM-CSF **ii)** without GM-CSF **B)** Control and GM-CSF independent NF-1 clones only. N= 2-3 biological replicates. Significance in **A) ii)** calculated using Wilcoxon paired t-test comparing NF-1 g1-2 and NF-1 g2-2 to control 3.

3.3.5 Introduction of *in vitro* GM-CSF hypersensitivity in human CB cells through *NF-1* editing using RNP nucleofection

Having established in the TF-1 cell line that *NF-1* editing at the exon 2 locus was sufficient to cause complete loss of protein expression in some clones, as well as rendering some clones able to sustain normal growth rates independent

of GM-CSF, it was hypothesised that similar phenotypic consequences could be achieved by targeting the same locus in human CB. The rationale behind choosing this primary cell type for further *in vitro* studies was for its relevance in JMML. As discussed in *Introduction section 1.3*, there is evidence that supports a fetal stem cell origin of this haematological malignancy (Loh 2011) (Nakamura, Ito et al. 2005) (Louka, Povinelli et al. 2021). Therefore, in order to accurately model the disease course, it was important to introduce the JMML disease initiating *NF-1* mutation in a cell context that most closely recapitulated what is predicted to occur in patients.

Similar to the approach taken for TF-1 cells, experiments were first carried out to determine the optimal cell culture conditions that would enable identification of GM-CSF independence in the context of *NF-1* editing without completely abrogating colony production in control groups. Using a simple growth medium containing only SCF and GM-CSF, instead of a full cocktail, significantly reduced colony production in wild type CB (**Figure 3.79**), however a further reduction to 1ng/mL GM-CSF did not reduce the clonogenicity further, suggesting at this concentration it was not limiting. Therefore the GM-CSF concentrations were reduced even further to grow edited clones: at 1ng/mL and 0.1ng/mL.

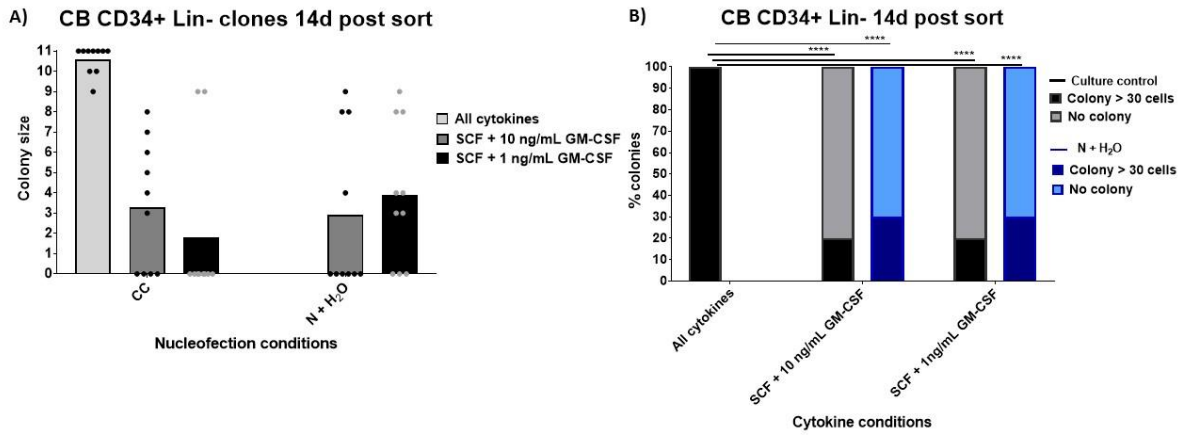


Figure 3.79: Wild type CB clonogenicity was reduced in minimal cytokine (SCF + GM-CSF) growth 14d post control CB sort into full cytokine cocktail or SCF with decreasing GM-CSF **A)** colony size (bars represent mean colony size) **B)** percentage colonies > 30 cells. N=10. Significance calculated using Fisher's exact test.

For *NF-1* or *AAVS1* targeted CB colonies grown in minimal (SCF and GM-CSF) cytokine conditions, the average colony size produced for both *AAVS1* and *NF-1* guide RNAs was similar (**Figure 3.80 B, C**), however, above a colony size of 8 (50 cells), which was required for genotyping, there were far fewer in the *AAVS1* targeted group (**Figure 3.80 D**) and the greatest number in the *NF-1* guide 1 group. Indeed, visual inspection showed the overall colony size was smaller in *AAVS1* targeted clones grown at 0.1ng/mL GM-CSF compared to their counterparts grown at 1ng/mL GM-CSF. On the other hand, *NF-1* guide 1 targeted clones produced colonies at 0.1ng/mL GM-CSF closer in size to those produced at 1ng/mL GM-CSF (**Figure 3.80**, representative images).

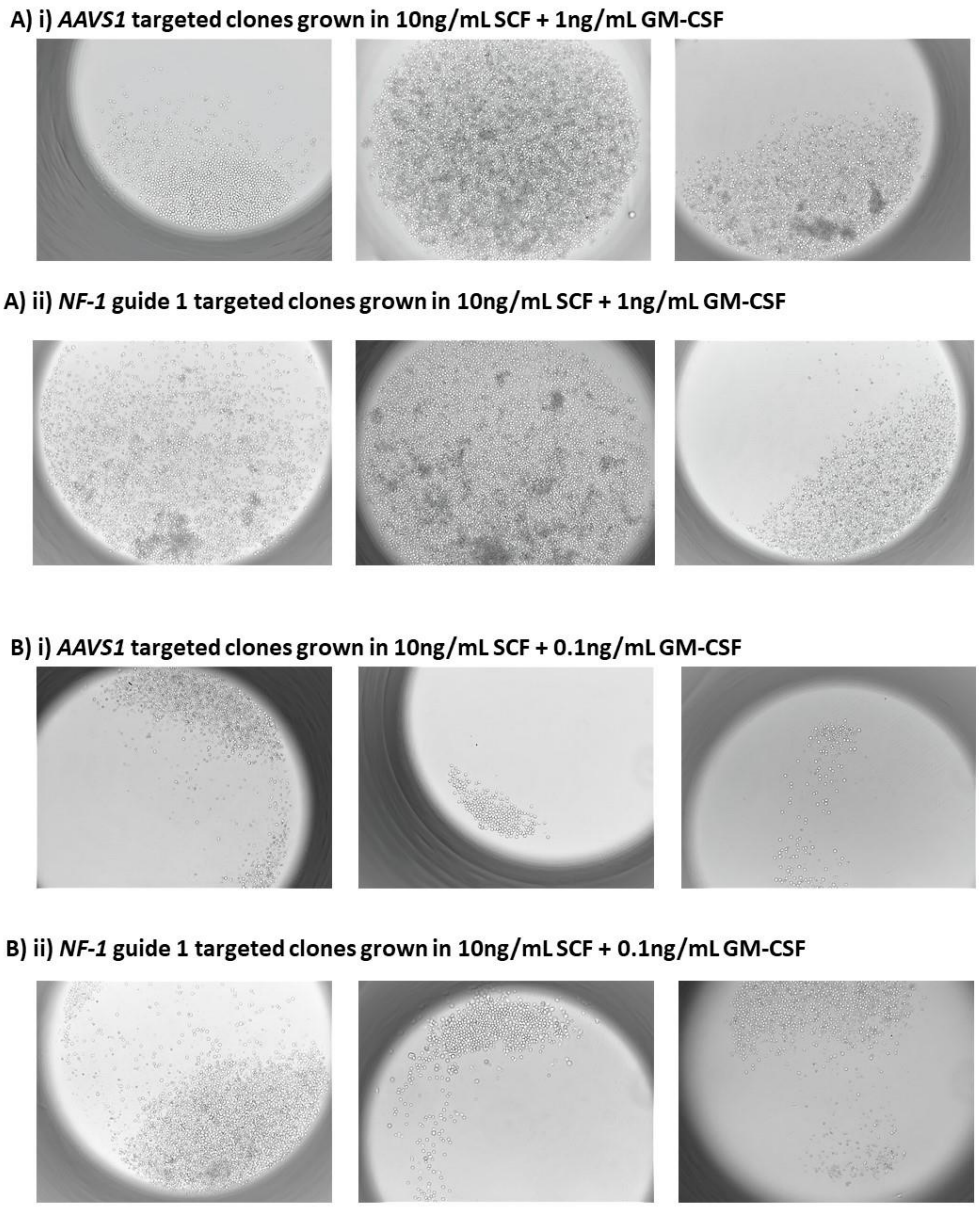


Figure 3.80: *NF-1* edited clones produced bigger colonies than AAVS1 controls under the most stringent cytokine conditions (**10ng/mL SCF + 0.1ng/mL GM-CSF**) Representative images 14d post sort for colonies grown in **A)** 10ng/mL SCF and 1ng/mL GM-CSF or **B)** 10ng/mL SCF and 0.1ng/mL GM-CSF for **i)** *AAVS1* targeted CB clones **ii)** *NF-1* guide 1 targeted CB clones. All images taken using 10X magnification.

In the *AAVS1* group, 17% of single sorted cells produced colonies above 50 cells in the full cytokine condition, which dropped to 7% in 10ng/mL SCF + 1ng/mL GM-CSF and 4.5% in 10ng/mL SCF + 0.1ng/mL GM-CSF. In contrast, *NF-1*

guide 1 targeted CB produced 13% colonies above 50 cells for full cytokine conditions and 10ng/mL SCF + 1ng/mL GM-CSF and 11.5% for 10ng/mL SCF + 0.1ng/mL GM-CSF. Although *NF-1* guide 1 targeted CB produced more colonies at lower GM-CSF concentrations compared to *AAVS1* targeted CB, this was not statistically significant (**Figure 3.81 D**). It is possible that 0.1ng/mL GM-CSF was not limiting in this experiment and even lower concentrations would have given statistical significance. It is also possible that the editing efficiency was low and therefore a large proportion of the *NF-1* targeted CB cells were wild type hence possessing the same propensity for growth in low GM-CSF concentrations as the *AAVS1* targeted cells. *NF-1* guide 2 produced 10% colonies above 50 cells for full cytokine conditions and SCF + 0.1ng/mL GM-CSF but only 3% for 10ng/mL SCF + 1ng/mL GM-CSF (**Figure 3.81 D**). This could be due to not enough cells being sorted for this condition to sample the true variation, or lower editing efficiency in this particular experiment. This result supported the approach of targeting *NF-1* in CB to develop an *in vivo* model of JMML as a key phenotypic characteristic of patient cells is *in vitro* GM-CSF hypersensitivity (Emanuel, Bates et al. 1991) (Metzner, Horstmann et al. 2007) (Kotecha, Flores et al. 2008). This phenotype was recapitulated by *NF-1* targeting in healthy CD34+ selected CB, although not a level that reached statistical significance (**Figure 3.81**). Interestingly, in TF-1 single cell sorts, *NF-1* guide 2 outperformed guide 1 (**Figure 3.74 F**) although this was not the case at the bulk level with both guides performing equally (**Figure 3.72**). In contrast, *NF-1* guide 1 was optimal in single cell CB cloning experiments (**Figure 3.81 D**). This highlighted the cell type and guide specificity of the nucleofection approach in terms of editing efficiency. Given the consistency of *NF-1* guide 1 in bestowing a growth advantage on CB

cells in minimal GM-CSF concentrations, this guide was taken forward for *in vivo* studies.

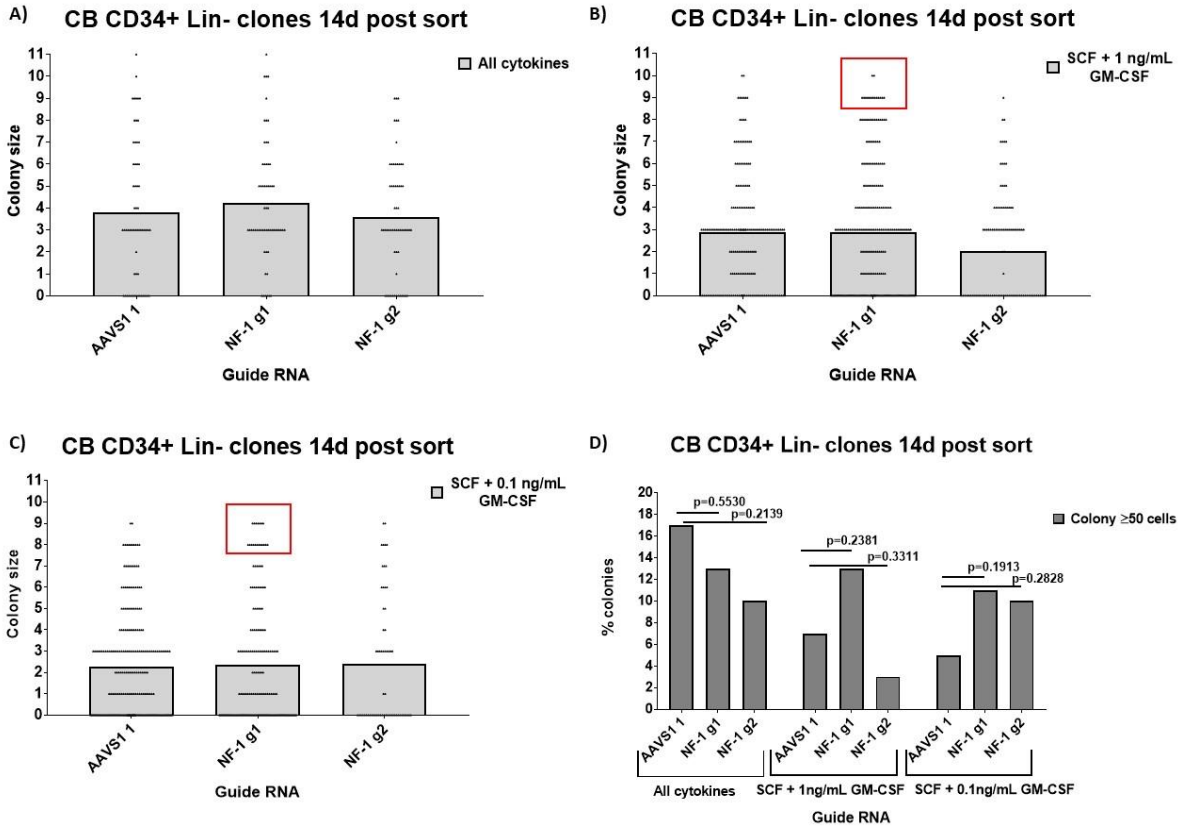


Figure 3.81: Effect of *NF-1* editing on CB clonogenicity at low GM-CSF concentrations Colony size of *AAVS1* or *NF-1* targeted CB 14d post sort into **A)** full cytokine cocktail **B)** SCF + 1ng/mL GM-CSF **C)** SCF + 0.1ng/mL GM-CSF. Bars represent mean colony size. **D)** Percentage colonies ≥ 50 cells for *AAVS1* or *NF-1* targeted CB. N= 60-240 single cells sorted per guide RNA/culture condition. Significance in **D)** calculated using Fisher's exact test.

In order to establish whether CB clones grown under minimal GM-CSF concentrations were in fact edited at the *NF-1* locus and harboured disease relevant allelic aberrations (**Figure 3.68**, heterozygous and compound heterozygous), colonies were genotyped at the *NF-1* locus 14 days post single cell sorting. As predicted from evidence that *AAVS1* targeted CB was still able to produce some colonies in the most stringent cytokine conditions tested

(**Figure 3.81 D**, 10ng/mL SCF + 0.1ng/mL GM-CSF), some *NF-1* targeted CB clones maintained an intact *NF-1* locus with no evidence of editing (**Figure 3.82 A and B**). These Sanger traces matched results from control *AAVS1* targeted samples (**Figure 3.82 C**). Importantly, for *NF-1* edited CB, the colonies produced represented all types of possible editing (**Figure 3.83 A** and **B**: heterozygous, **C** and **D**: compound heterozygous, **E** and **F**: homozygous).

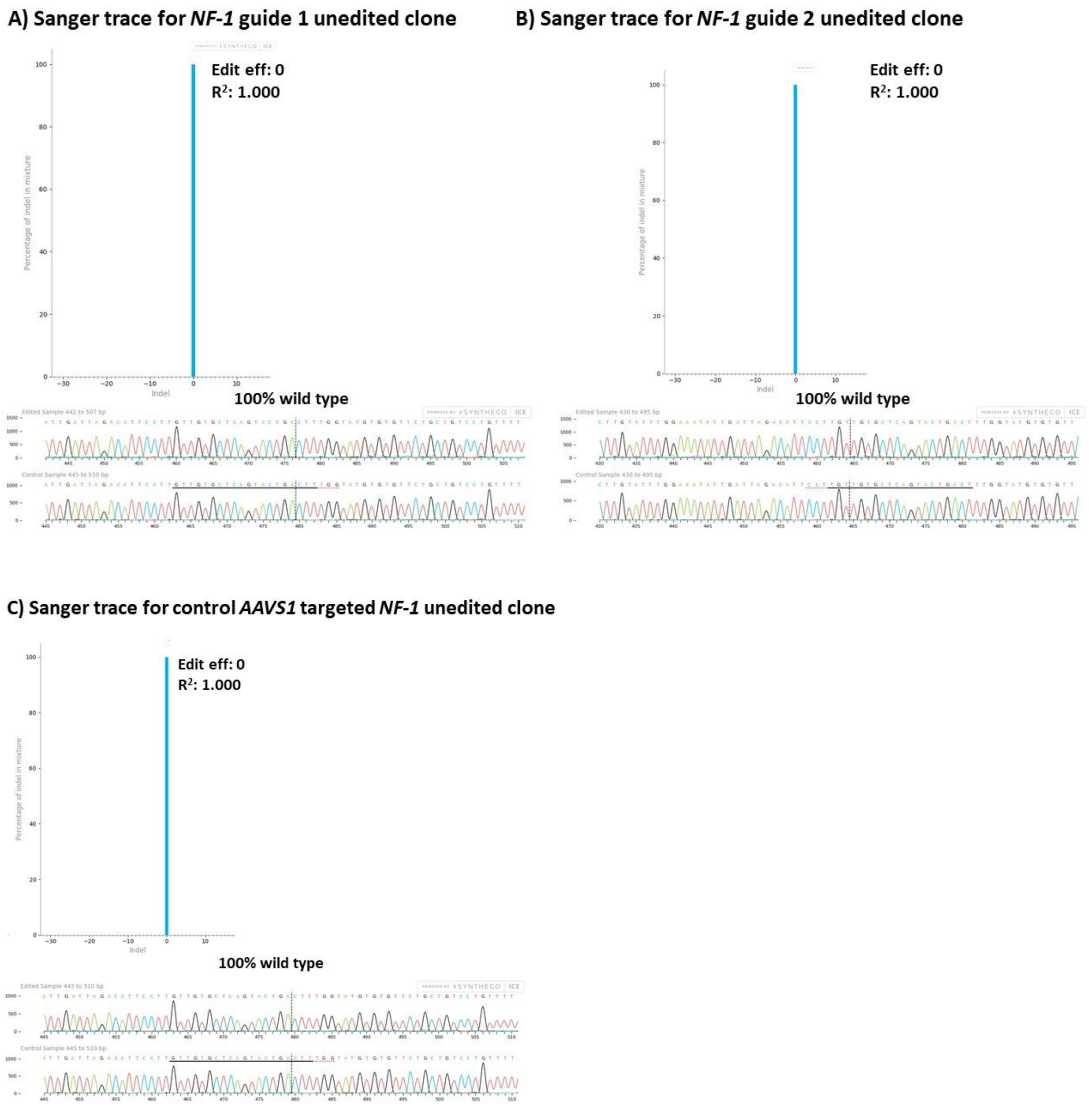


Figure 3.82: *NF-1* CB targeting was not 100% efficient in producing edited colonies in minimal cytokine conditions 1Sanger sequencing trace aligned to control wild type and indel score computed using Synthego ICE tool for representative clones of **A)** *NF-1* guide 1 unedited **B)** *NF-1* guide 2 unedited **C)** *AAVS1* targeted control.

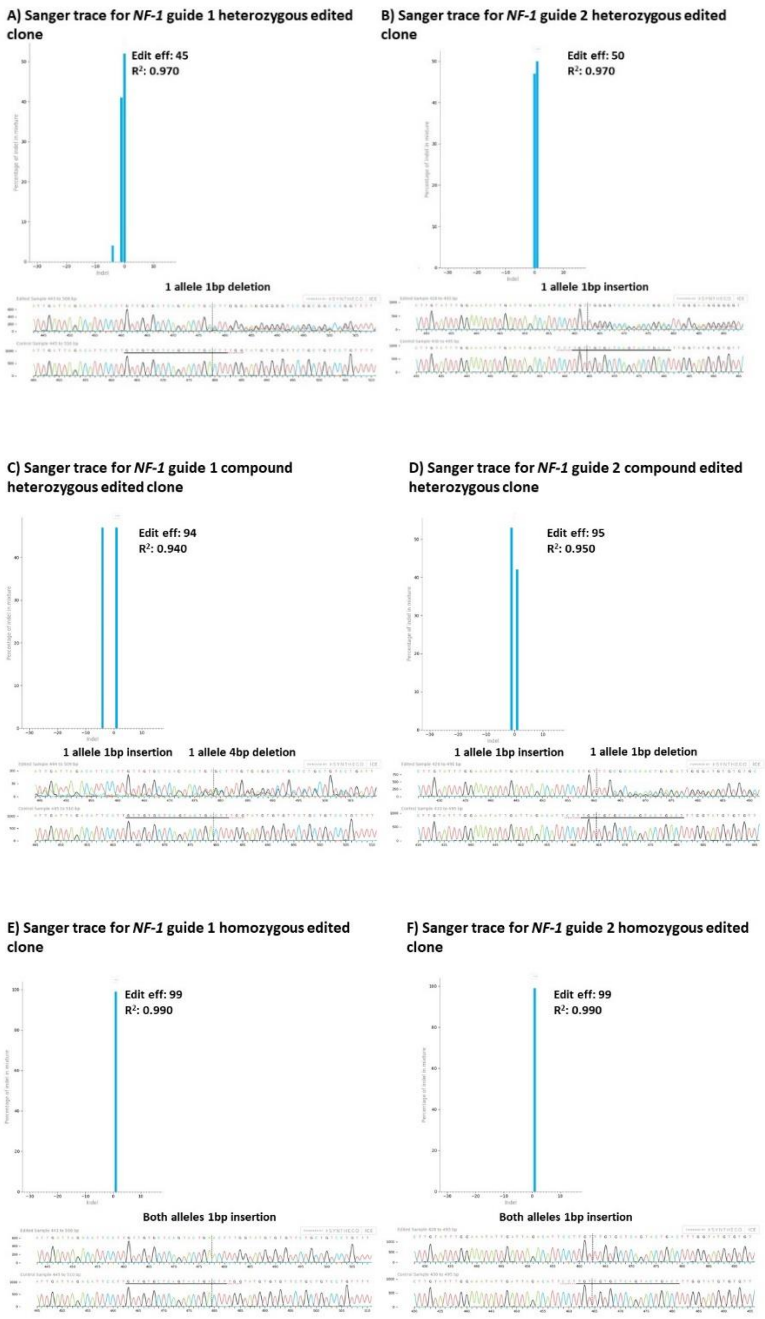


Figure 3.83: *NF-1* editing produced heterozygous, compound heterozygous and homozygous mutated clones Sanger sequencing trace aligned to control wild type and indel score computed using the Synthego ICE tool for representative clones of **A)** *NF-1* guide 1 heterozygous **B)** *NF-1* guide 2 heterozygous **C)** *NF-1* guide 1 compound heterozygous **D)** *NF-1* guide 2 compound heterozygous **E)** *NF-1* guide 1 homozygous **F)** *NF-1* guide 2 homozygous.

In the context of developing a CB derived model of JMML that was most likely to cause a clonal expansion in NSG mice, it was predicted that clones in which both

NF-1 alleles were mutated would have the highest fitness advantage *in vivo*. Therefore, the *NF-1* targeted colonies were assessed for the proportion of each type of mutation they harboured. As demonstrated in **Figure 3.84**, all types of mutations were induced by both *NF-1* guides. Analysing all colonies targeted with the same guide revealed that compound heterozygous mutations were most common for *NF-1* guide 1, whereas heterozygous mutations were most common for guide 2 (**Figure 3.84**). From these data it was predicted therefore that *NF-1* guide 1 would give the greatest fitness advantage to CB in *in vivo* NSG transplants.

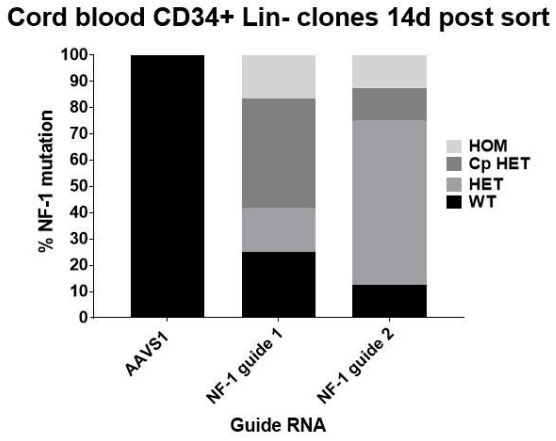


Figure 3.84: *NF-1* guide 1 produced more compound heterozygous clones than *NF-1* guide 2 Percentage of each type of edit for representative colonies genotyped 14d post sort at the *NF-1* exon 2 locus. N= 4 (AAVS1 guide), 12 (*NF-1* guide 1), 8 (*NF-1* guide 2).

3.3.6 NSG transplantation of *NF-1* targeted CB to evaluate the propensity for *in vivo* clonal expansion

Having established *in vitro* that *NF-1* guide 1 produced the greatest number of CB colonies in the most limiting GM-CSF concentration (0.1ng/mL), as well as the highest proportion of bi-allelic editing, this guide was taken forward for *in*

vivo studies. CD34+ selected CB was nucleofected with *NF-1* guide 1 or *AAVS1* control RNPs and transplanted into sub-lethally irradiated NSG recipients 24 hours post-nucleofection. CB samples from multiple donors were pooled during CD34+ selection to ensure comparable CB CD34+s were used for control *AAVS1* and *NF-1* targeting. **Table 3.2** summarises the mice and experimental groups used as well as their disease progression trajectory.

NSG gender	**NSG age at time of transplant**	**CD34+ cord blood target guide RNA**	**Age at culling**	**Time to disease phenotype**	**Disease phenotype**
Female	10 wk	*AAVS1*-1	28 wk	No disease	No disease
Female	10 wk	*AAVS1*-2	28 wk	No disease	No disease
Female	10 wk	*AAVS1*-3	36 wk	No disease	No disease
Female	10 wk	*AAVS1*-4	36 wk	No disease	No disease
Female	10 wk	*NF-1*-1	27 wk	17 wk	Anaemia, thrombocytopenia
Female	10 wk	*NF-1*-2	36 wk	No disease or human engraftment	No disease
Female	10 wk	*NF-1*-3	36 wk	No disease	No disease
Female	10 wk	*NF-1*-4	23 wk	13 wk	Pancytopenia

Table 3.2: Summary of NSG mouse CB recipients and timeline of disease onset

In order to be able to measure the expansion of *NF-1* edited clones in the recipient mice, it was important to determine the starting mutational level at the time of transplant. Therefore a small number of cells were spared from transplantation for genotyping purposes. As expected, the *NF-1* targeting level was 0% in *AAVS1* controls with *NF-1* guide 1 targeted samples harbouring 8%

editing (**Figure 3.85**). This was considerably lower than the bulk level achieved for TF-1 cells (**Figure 3.72**) which is perhaps unsurprising given primary stem cells are known to be less amenable to editing than cell lines.

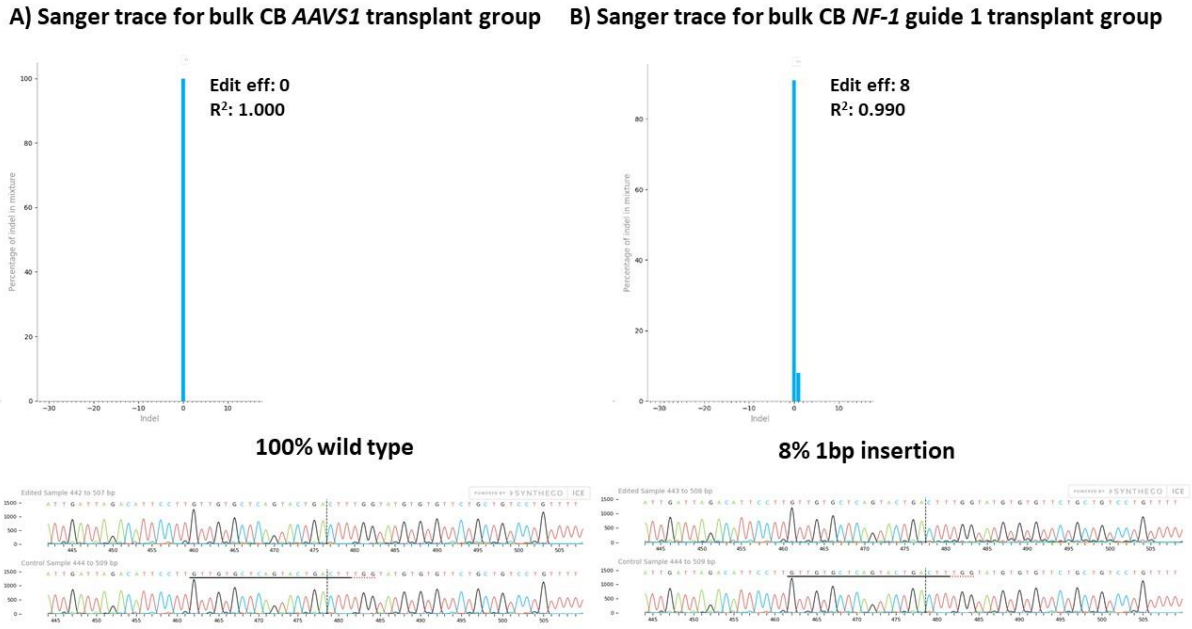


Figure 3.85: CB *NF-1* targeting showed a single base insertion dominated at the bulk level ICE analysis of NF-1 exon 2 Sanger traces 24 hours post-nucleofection for **A)** *AAVS1* and **B)** *NF-1* targeted bulk CB samples.

In order to assess the lineage contribution of both edited and un-edited CB samples to engraftment in NSG mice, blood counts and flow cytometric analyses were carried out on tail-vein PB at four-five week intervals post-transplant. Mice were also monitored by BMS staff for signs of ill health. Total human blood cell and specific B and myeloid lineage engraftment was assessed by staining for human CD45 as well as lineage specific markers (**Figure 3.86**).

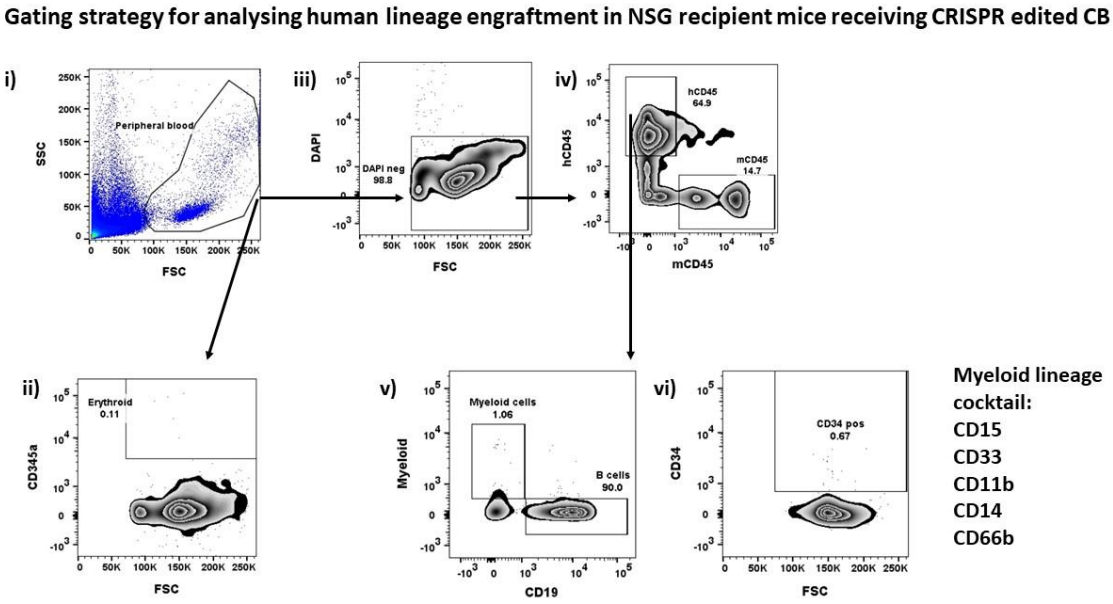


Figure 3.86: Gating strategy for assessing human mature lineage engraftment in NSG CB CRISPR edited recipients 13wk AAVS1-1 example flow cytometry gating for peripheral blood chimerism: **i)** initial morphology gate used as parental population for **ii)** red blood cells, **iii)** DAPI negative cells separate for **iv)** human and mouse CD45, human CD45 separated for **v)** lymphoid/myeloid and **vi**) CD34+ contribution.

The level of total hCD45 engraftment was similar between the *AAVS1* and *NF-1* groups, with minimal engraftment at four weeks but substantial (60-70%) hCD45 chimerism by 13 weeks (**Figure 3.87 A**). This trajectory is unsurprising as CB HSPC require time to home to the BM and establish lineage output in NSG recipients. Interestingly the proportion of B and myeloid output from the hCD45 compartment was also similar between *AAVS1* and *NF-1* groups (**Figure 3.87 B**, no significant differences). Equally, the percentage of human CD34+ cells in the PB was not statistically significant between the control and edited groups; CD34+ fractions were higher at 4 weeks then gradually reduced in the PB (**Figure 3.87 C**). Taken together, these data showed no sign of aberrant expansion of a

specific human blood cell lineage, nor did they suggest the presence of a population of aberrant immature cells in the PB.

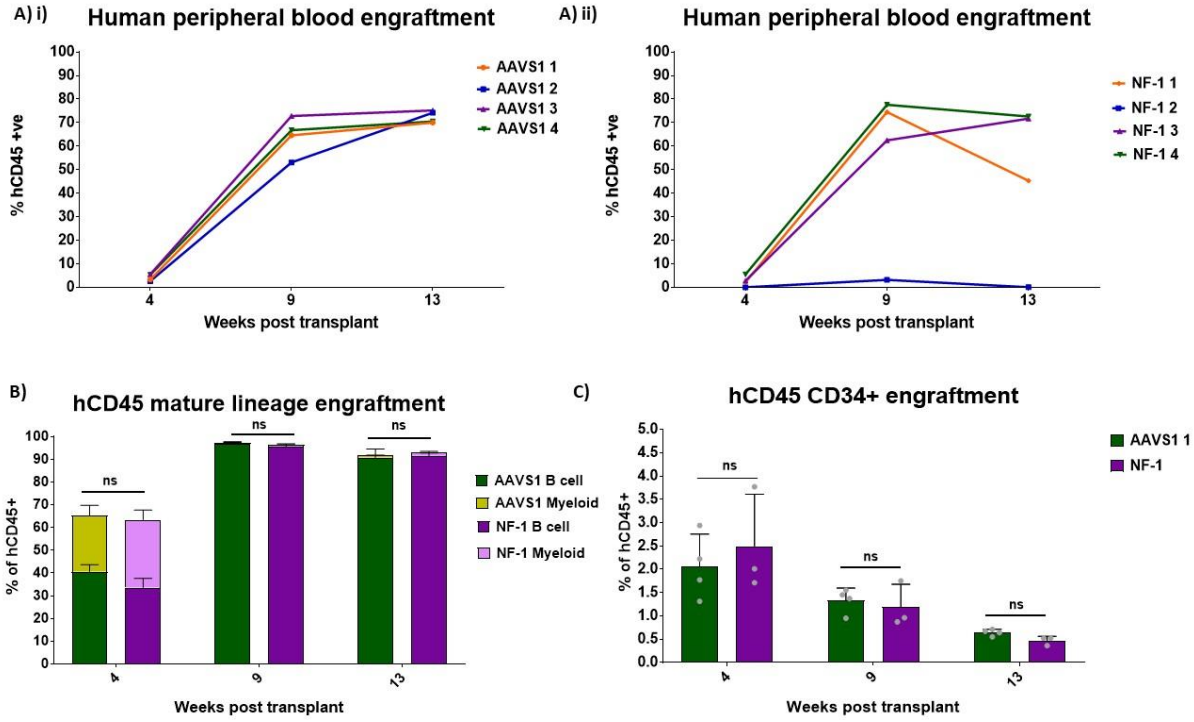


Figure 3.87: Human CD45 peripheral blood lineage engraftment showed a similar pattern in *NF-1* and *AAVS1* targeted CB NSG recipients A) Percentage total human CD45 engraftment for **i)** *AAVS1* **ii)** *NF-1* CB targeted NSG recipients. **B)** Contribution of human cell peripheral blood engraftment to myeloid and B cell compartments. **C)** Contribution of human cell peripheral blood engraftment to the CD34 compartment. Significance in **B)** and **C)** between *AAVS1* and *NF-1* groups was calculated separately for each lineage and time-point using a non-parametric unpaired Mann-Whitney test.

Given the absence of differences in human CD45 engraftment between the *AAVS1* and *NF-1* groups, it was hypothesised that a defect could be seen within the mCD45 chimerism, given that in the NSG setting, a proportion of peripheral blood reconstitution remains mouse derived in a healthy setting. As expected from the hCD45 trajectory (**Figure 3.87 A**), the mCD45 chimerism started at a high level initially (at the four-week time-point) and gradually reduced as human engraftment took over (**Figure 3.88**). Of note, in 2 mice (*NF-*

1 1 and *NF-1* 4), mCD45 chimerism in the PB was less than 5% by 13 weeks post-transplant, whereas for the other two *NF-1* recipients and all the *AAVS1* controls, it remained above 5% (**Figure 3.88**).

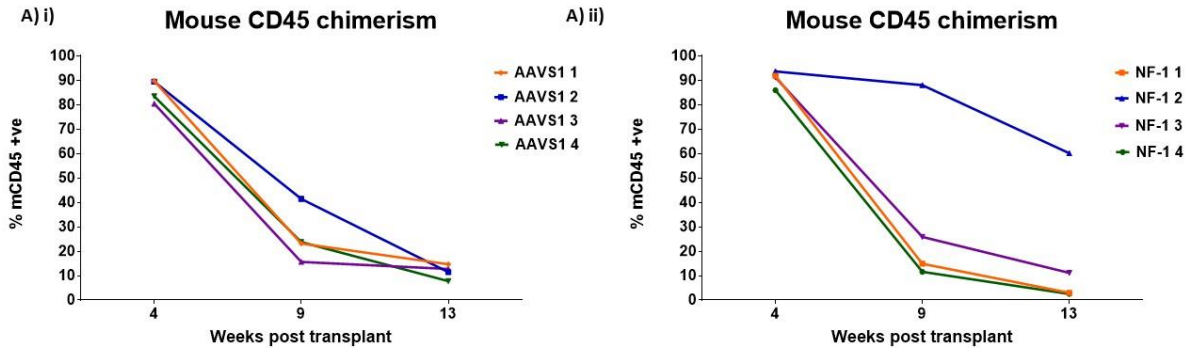


Figure 3.88: Mouse CD45 chimerism reduced over the course of transplant in *NF-1* and *AAVS1* targeted CB NSG recipients A) Percentage total mouse CD45 engraftment for **i)** *AAVS1* **ii)** *NF-1* CB targeted NSG recipients.

Given the slightly lower total mouse CD45 chimerism in certain *NF-1* recipients, PB counts were analysed in order to determine whether the reduced output from the mouse BM compartment had an effect on overall lineage output, including platelets and red blood cells. While the total white blood cell count remained stable across the course of the 13-week readout for both control and edited groups (**Figure 3.89 A**), one mouse (*NF-1* 4) developed anaemia and thrombocytopenia: by the week 13 readout, the red blood cell count was 1.62 $x10^6$/mL and the platelet count was 318 $x10^3$/mL compared to an average of 8.15 $x10^6$/mL (RBC) and 645 $x10^3$/mL (platelet) for the control group (**Figure 3.89 B**, **C**).

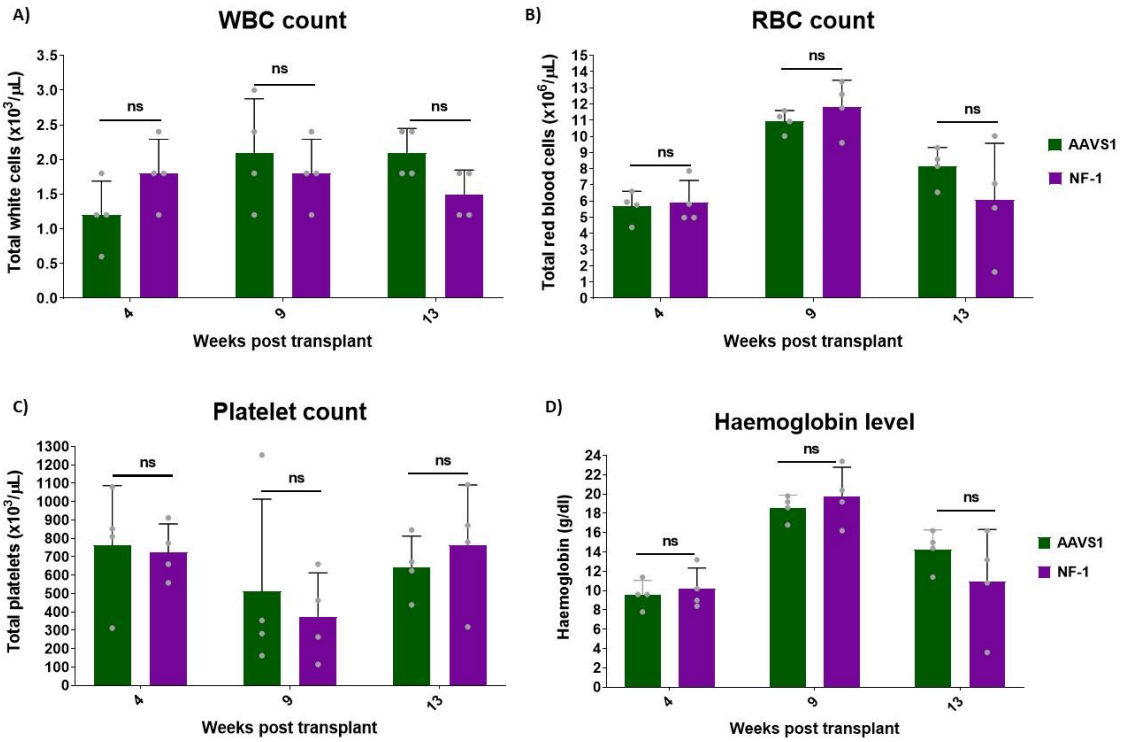


Figure 3.89: Peripheral blood counts of *NF-1* CB NSG recipients Serial **A)** total white cell **B)** red blood cell **C)** platelet and **D)** haemoglobin level counts for **i)** *AAVS1* and **ii)** *NF-1* edited CB NSG recipients. Significance between *AAVS1* and *NF-1* groups was calculated separately for each time-point using a non-parametric unpaired Mann-Whitney test.

There was no difference in the overall differential WBC count between the *NF-1* and *AAVS-1* groups (**Figure 3.90**). There was a slightly higher myeloid contribution in the *NF-1* group at 13 weeks but this was not statistically significant. All other lineages at this time-point were similar. Taken together, analyses of the PB counts and hCD45 lineage contribution revealed little difference between control and edited groups.

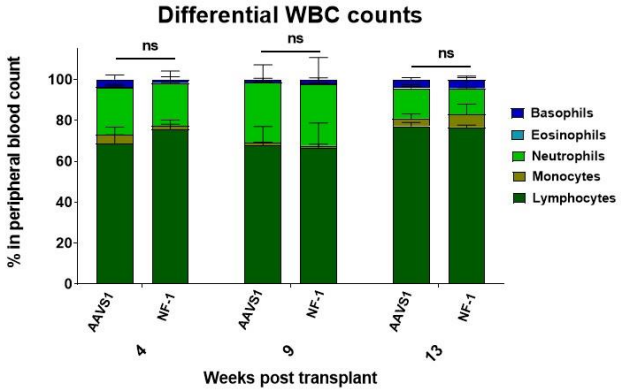


Figure 3.90: Differential WBC counts were similar between *AAVS1* control and *NF-1* targeted groups Peripheral blood counts analysed 4, 9 and 13wk post-transplant. Significance between *AAVS1* and *NF-1* groups calculated using a two-way ANOVA, p value calculated using column factor for each time point.

Two of the mice that had received *NF-1*-edited cells (*NF-1* 1 and *NF-1* 4) fell sick with weight loss and pale limbs at 17 weeks and 13 weeks post-transplant respectively. One of the mice (*NF-1 4*) was pancytopenic while *NF-1 1* was anaemic and thrombocytopenic with a WBC count at the lower end of normal (**Table 3.3**, 1.2 x10^6/µL). These results indicated a potential haematopoietic defect although these symptoms could have also resulted from infection.

Peripheral blood count parameter	NSG CB recipient ID		Normal range (female NOD/SCID 8-10wk mouse)
	NF-1 1	*NF-1* 4	
Total white blood cells ($x10^3/\mu L$)	1.2	0.6	0.96-4.68 ($x10^3/\mu L$)
Red blood cells ($x10^6/\mu L$)	0.54	0.78	8.2-10.48 ($x10^6/\mu L$)
Haemoglobin (g/dl)	1.2	1.8	12.1-17.6 (g/dl)
Platelets ($x10^3/\mu L$)	6	6	651-1878 ($X10^3/\mu L$)

Table 3.3: White cell, red blood cell, haemoglobin and platelet counts for NSG recipients of *NF-1* 1 and *NF-1* 4 edited CB at time-point culled sick- 17 and 13 weeks post-transplant respectively. Normal ranges obtained from Charles River NOD SCID Mouse Clinical Pathology Data sheet (River 2021)

In order to elucidate the cause of the haematological abnormalities, the PB, BM and spleen were analysed and compared to two mice in the control *AAVS1* group (culled at 18 weeks). The final two *AAVS1* controls and *NF-1* recipient 3 which didn't fall sick were culled at the experiment end 26 weeks post-transplant. For analysis of human lineage engraftment, *NF-1* recipient 2 was excluded as no human engraftment was seen for this recipient. The proportion of human cells (hCD45+) in all 3 tissues was higher in the *NF-1* compared to the *AAVS1* CB recipient mice (**Figure 3.91 A**). Although there was a slightly higher proportion of myeloid cells in the PB of the *NF-1* 1 recipient, this was not seen in *NF-1 4* or in the BM or spleen. (**Figure 3.91 B**). CD34+ engraftment was similar for all groups within the BM, spleen and PB although it was variable between replicates in the PB (**Figure 3.91 C**).

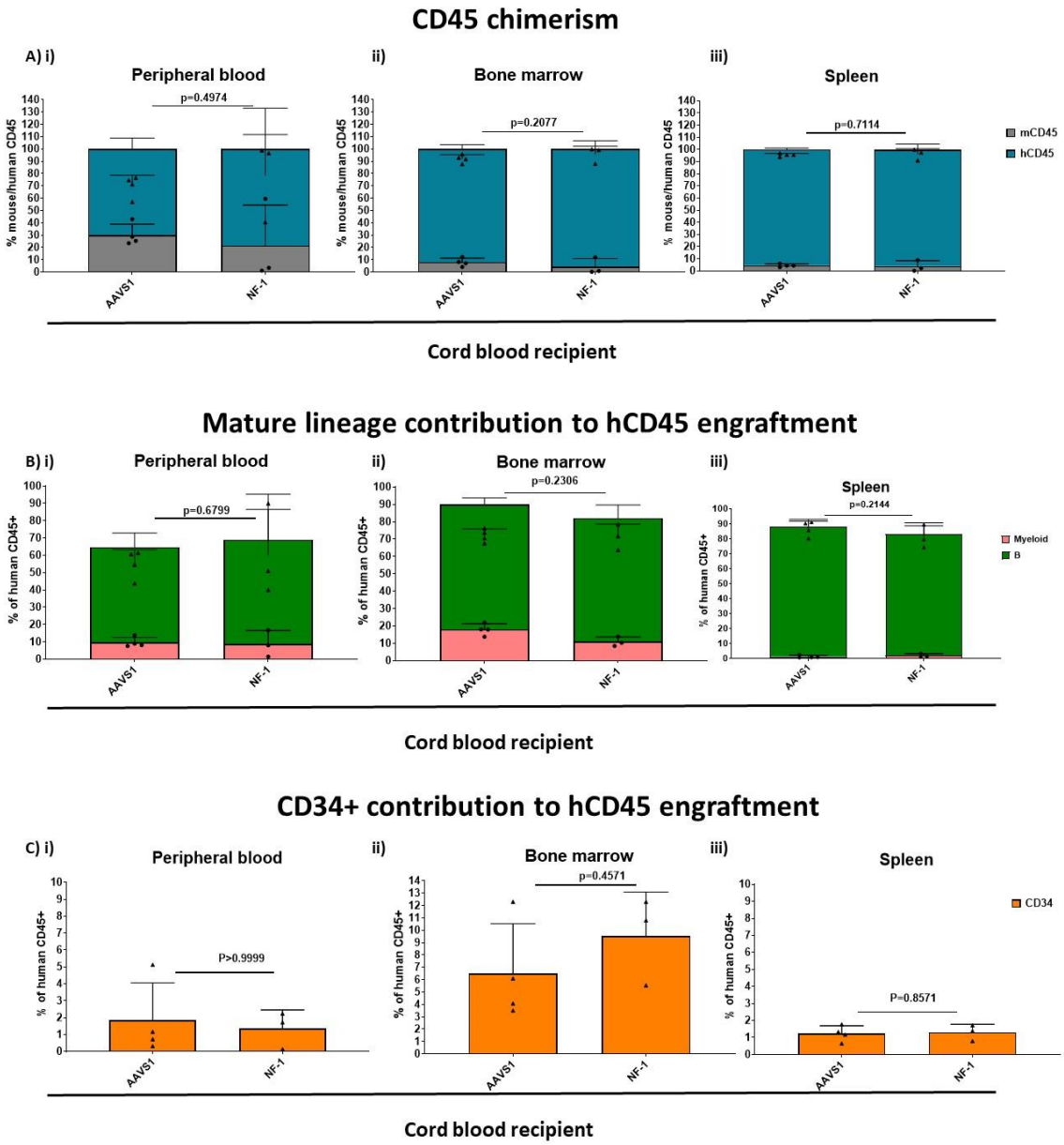


Figure 3.91: Engraftment of *NF-1*- and control-edited hCD45+ cells in peripheral blood, bone marrow and spleen at 13-26 weeks post-transplant **A)** Mouse and human CD45 chimerism **B)** Mature B and myeloid cell **C)** CD34+ cell engraftment in **i)** peripheral blood **ii)** bone marrow **iii)** spleen. Significance between *AAVS1* (N=4) and *NF-1* (N=3) groups in mouse/human CD45 and mature lineage graphs calculated using two-way ANOVA, interaction between row factor and column factor used for p value. Significance between *AAVS1* and *NF-1* groups in CD34+ graph calculated using non-parametric unpaired Mann-Whitney t-test.

The human stem and progenitor compartments were then analysed in the BM and spleen. As shown in the gating example in **Figure 3.92**, live cells were gated

for hCD45, lineage negative followed by separating the CD34+ gate into CD38 +/-. In the CD34+ CD38+ gate, CMPs were gated by CD123+ CD45RA-, MEPs by CD123-CD45RA- and GMPs by CD45RA+ CD123-. In the CD34+ CD38- gate, HSCs were gated by CD45RA- CD90+, MPPs by CD45RA- CD90- and LMPPs by CD45RA+ CD90-.

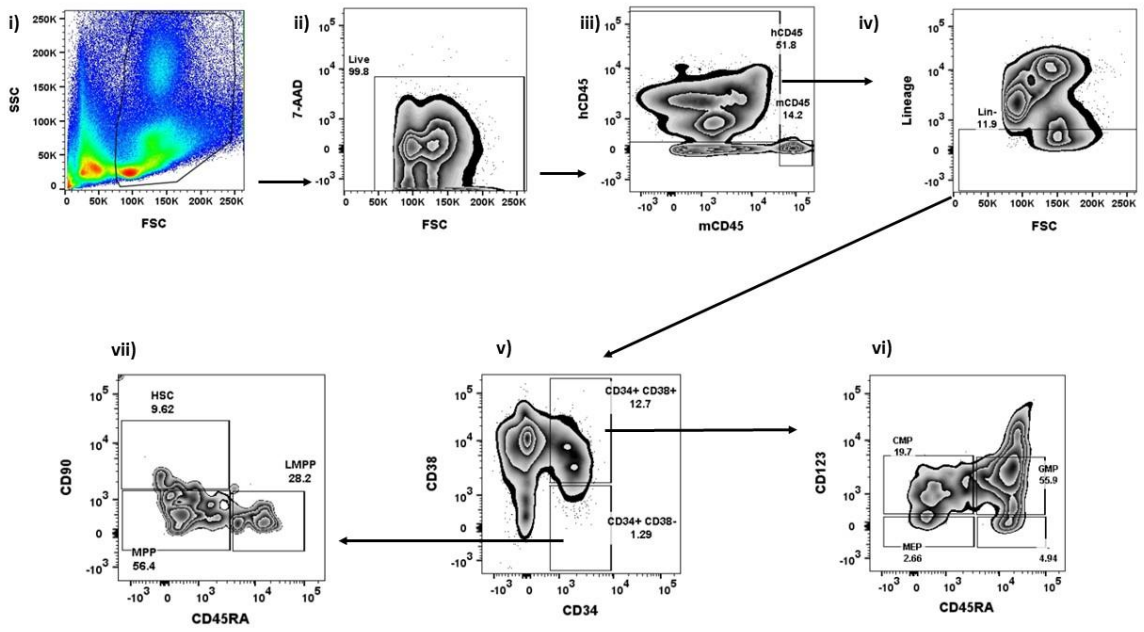


Figure 3.92: Flow cytometry gating strategy for analysis of engraftment of human HSPC populations in the bone marrow and spleen of *NF-1* and *AAVS1* edited CB recipient mice Gating strategy for 18-week BM AAVS1-1 example showing **i)** total cells **ii)** Live **iii)** mCD45 and human CD45 chimerism **iv)** Lineage (CD8, CD20, CD2, CD3, CD16, CD19, CD235a, CD66b, CD10, CD127) negative **v)** CD34 vs CD38 **vi)** GMP (CD45RA+ CD123+), MEP (CD45RA- CD123-), CMP (CD45RA-CD123+) **vii)** HSC (CD45RA- CD90+), MPP (CD45RA- CD90-), LMPP (CD45RA+ CD90-).

Within the BM stem and progenitor compartment, there were two expanded populations in both *NF-1* mice which fell sick compared to the *AAVS1* controls. Normal human HSPC engraftment was seen for all four *AAVS1* controls (**Figure 3.93**). On the other hand, within the CD34+ CD38+ compartment, a population of CD45RA+ CD123- cells were present at ~20-30% in the *NF-1* group (**Figure**

3.94) for the two recipients which fell sick compared to 4-6% in the controls (**Figure 3.93**). This population was not however present in the *NF-1* recipient (3) which had no sign of illness by week 26 post-transplant. There was also an expansion of the LMPP population within the CD34+CD38- fraction for the *NF-1* group (**Figure 3.94**). Interestingly, this was seen in all three *NF-1* groups, including the recipient which didn't fall sick, suggesting a potential initiating population that might require further mutations to cause disease development. In the case of the CD45RA+ CD123- expansion, this was associated with a reduction in CMPs (**Figure 3.95 A**, p=0.0058) and in the case of the LMPP expansion, this was associated with a reduction in HSCs and MPPs (**Figure 3.95 B**, p= 0.0002). The aberrantly expanded CD45RA+ CD123- population represented a GMP-like population which suggested a block in differentiation at the pre-GMP stage. The LMPP expansion was interesting as our lab recently reported the presence of expanded LMPP, and/or a CD90+ LMPP population, in many patients with JMML (Louka, Povinelli et al. 2021) (Louka et al, 2021). These aberrantly expanded progenitor populations could explain the downstream consequences of anaemia and thrombocytopenia, and reduced mouse CD45 chimerism. Further, the reduction in human HSC and MPP in favour of the aberrantly expanded populations could have resulted in reduced human platelet and erythroid output.

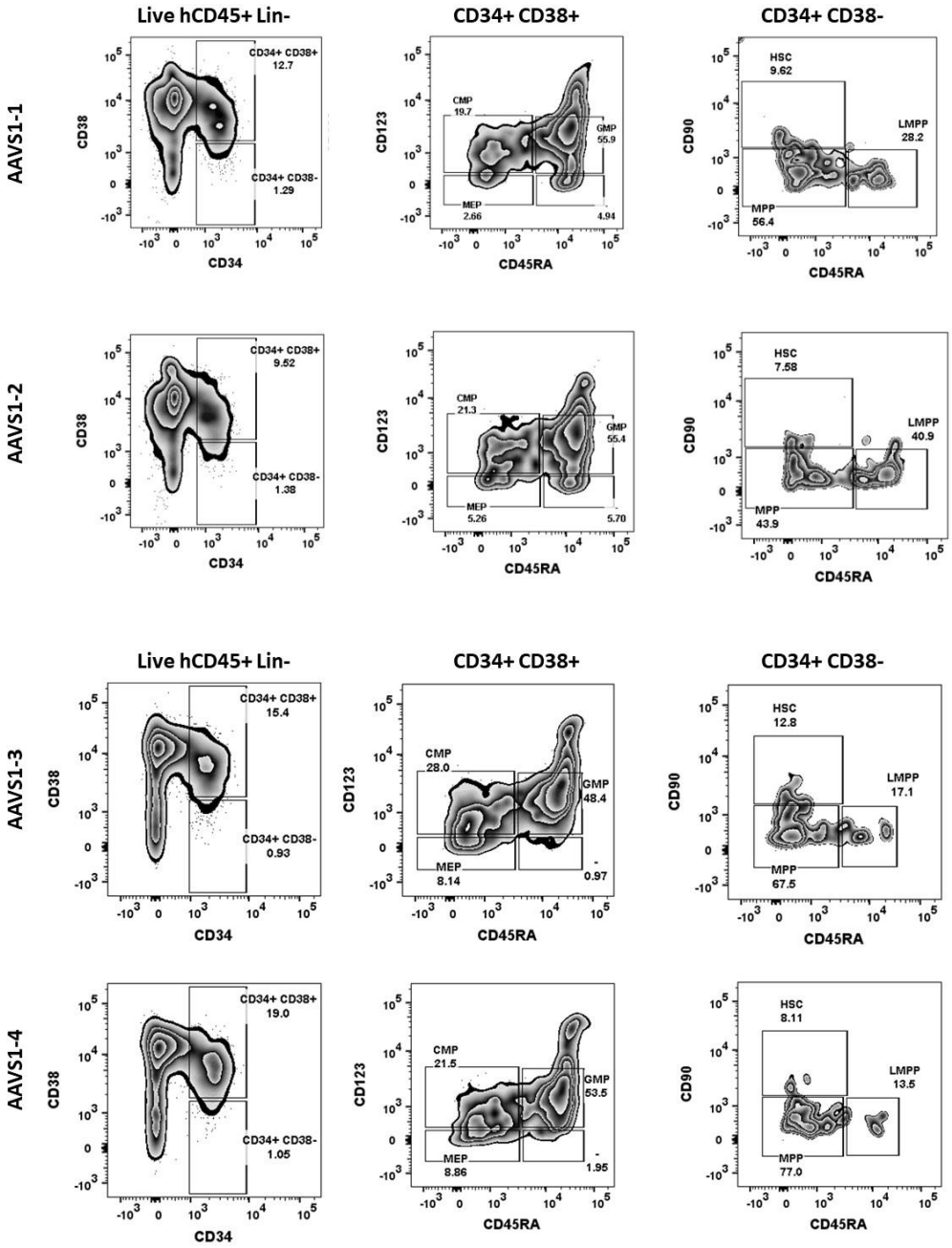


Figure 3.93: *AAVS1* edited CB NSG recipient mice showed a normal human bone marrow composition Flow cytometry panels of CD34+ CD38+ and CD34+ CD38- primitive gates for *AAVS1* CB recipients

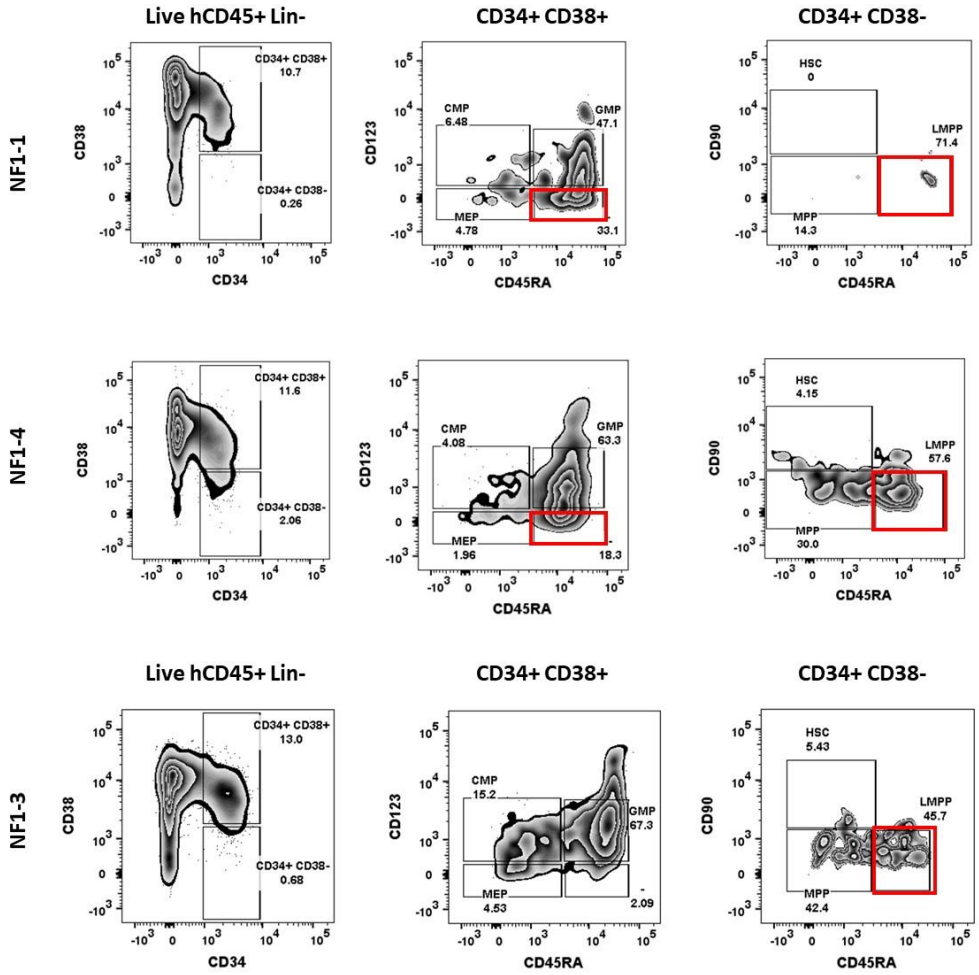


Figure 3.94: *NF-1* edited CB NSG recipient mice showed an expansion of human CD45RA+ CD123- and LMPP in the BM Flow cytometry panels of CD34+ CD38+ and CD34+ CD38- primitive gates for *NF-1* cord blood recipients.

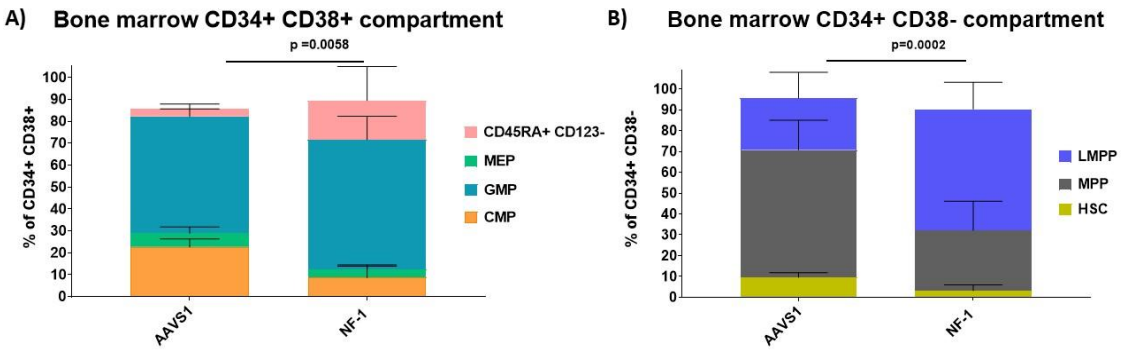


Figure 3.95: *NF-1* edited CB NSG recipient mice showed an expansion of CD45RA+ CD123- and LMPP in the BM Percentage contribution of each subgroup to the **A)** CD34+ CD38+ and **B)** CD34+ CD38- compartments in *AAVS1* and *NF-1* recipient BM. Significance calculated using a two-way ANOVA.

In order to establish whether the aberrant expansion could correspond to an expansion of the *NF-1* edited population, Sanger sequencing at the exon 2 *NF-1* locus was carried out on total BM samples. As expected, *AAVS1* recipients showed no editing at the *NF-1* locus, with 100% wildtype sequence in the total BM (**Figure 3.96)**. Strikingly, recipients of *NF-1* 1, *NF-1* 4 and *NF-1* 3 CB showed 37%, 31% and 32% editing respectively at the *NF-1* exon 2 locus in DNA extracted from the total BM at the time of culling, despite only *NF-1* 4 and *NF-1* 1 recipients falling sick. (**Figure 3.97**). This was almost four times the level of editing in the population that was transplanted (**Figure 3.85 B**, 8%) which demonstrated an expansion of the edited population. Most of the edited population in *NF-1* 4 harboured a 1bp insertion, at the same position as the 8% population transplanted. On the other hand, the predominant edit in *NF-1* 1 was a 2bp deletion, not detectable at the bulk level in the initial transplant, indicative of a smaller starting group, which could explain why the disease development had a longer latency in this recipient. There was evidence of several additional edits in both *NF-1* targeted samples although most of these were below 5% which is the threshold of sensitivity for the Sanger ICE tool (Synthego 2021). On the other hand, these could indicate additional expanded populations which were below detection at the bulk level at the time of transplant.

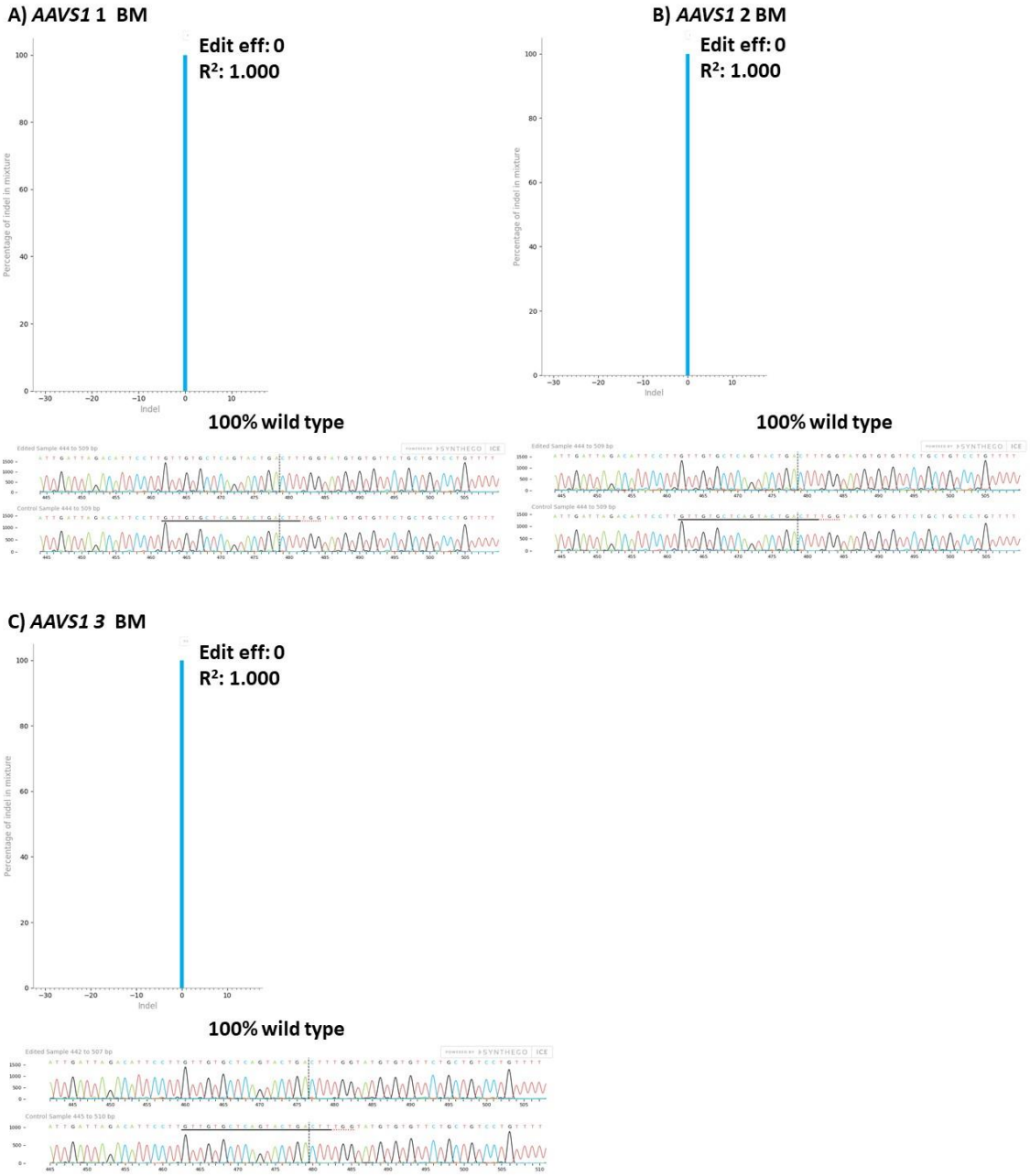


Figure 3.96: *AAVS1* CB recipient BM showed no expansion of an *NF-1* edited population Sanger sequencing trace and ICE (Synthego) editing score for DNA extracted from total bone marrow of **A)** *AAVS1* 1 recipient **B)** *AAVS1* 2 recipient **C)** *AAVS1* 3 at the time of culling

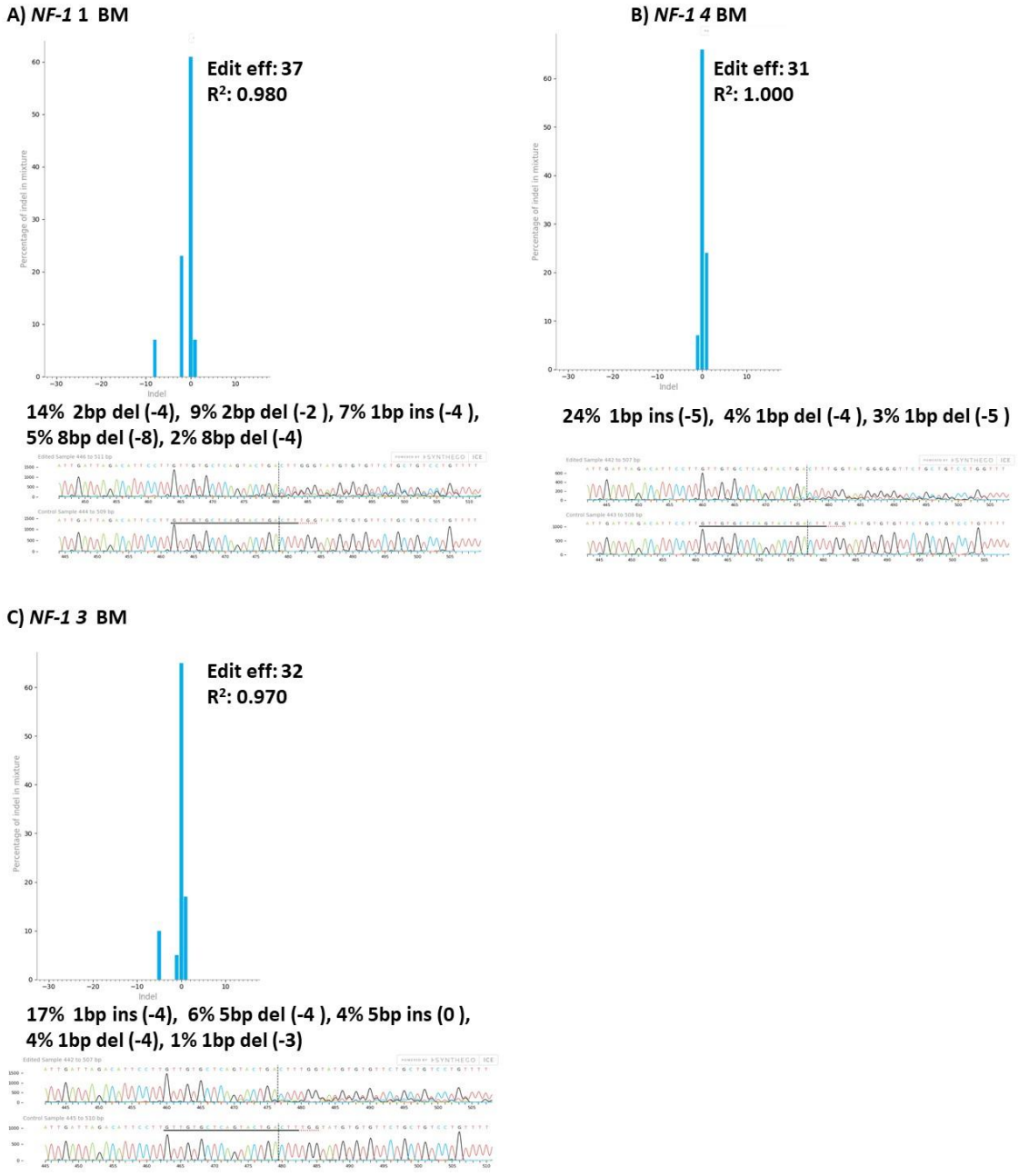


Figure 3.97: *NF-1* CB recipient bone marrow showed an expansion of the edited population compared to at the time of transplant Sanger sequencing trace and ICE (Synthego) editing score for DNA extracted from total bone marrow of **A)** *NF-1* 1 recipient **B)** *NF-1* 4 recipient **C)** *NF-1* 3 recipient and **D)** *AAVS1* 2 at the time of culling. Numbers in brackets represent the distance of the insertion/deletion from the PAM (5'-NGG-3') site.

Having identified the presence of an aberrant population within the BM, also previously described in JMML patient xenografts (Caye, Rouault-Pierre et al. 2020) (Louka et al, 2021), I also evaluated the spleen of the mice; an enlarged

spleen is common in JMML (Niemeyer and Flotho 2019) and was also seen in JMML xenografts (Louka et al, 2021). Although the spleens of *NF1-1* and *NF-1* 3 were slightly larger than the *AAVS1* controls, overall there was no significant difference in spleen weight. (**Figure 3.98**). Morphological analysis of BM smears showed a clear lack of megakaropoiesis, one of the clinical features of JMML (Niemeyer and Flotho 2019), in the *NF-1* recipients which fell sick compared to *AAVS1* controls (**Figure 3.99**, representative images). There were increased numbers of abnormal, monocytoid blast-like cells in the bone marrow of *NF-1 4* and to a lesser extent in *NF-1 1* but there was no evidence of diffuse leukaemic infiltration of the BM, which would be expected for a JMML phenotype (**Figure 3.99**).

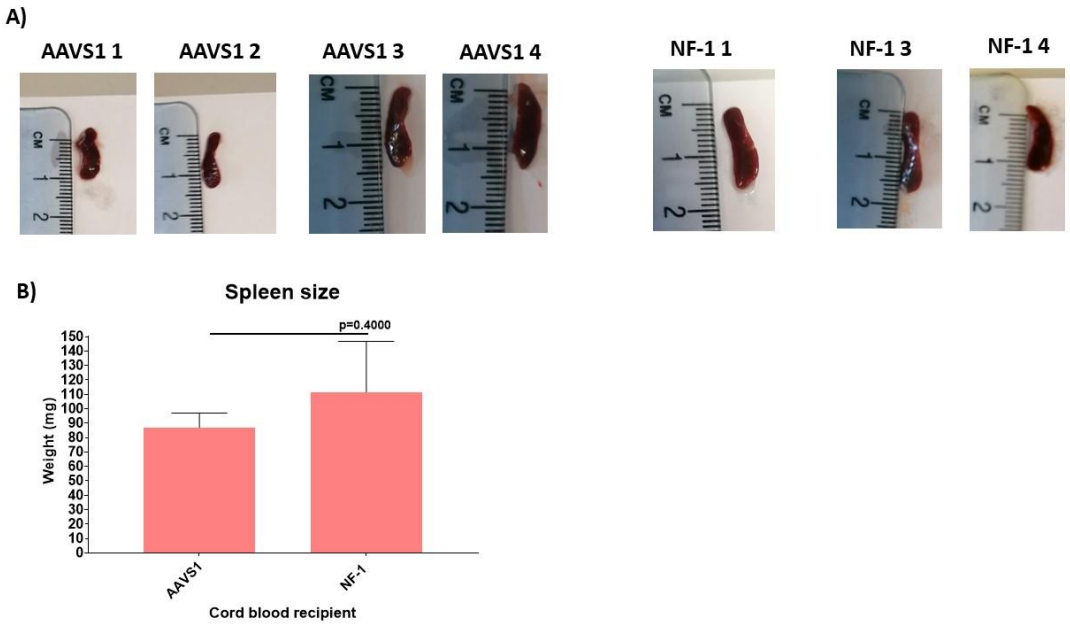


Figure 3.98: *NF-1* CB NSG recipients did not show an enlarged spleen A) Spleen images of *NF-1* and *AAVS1* CB recipients **B)** Spleen weight of *NF-1* and *AAVS1* CB recipients.

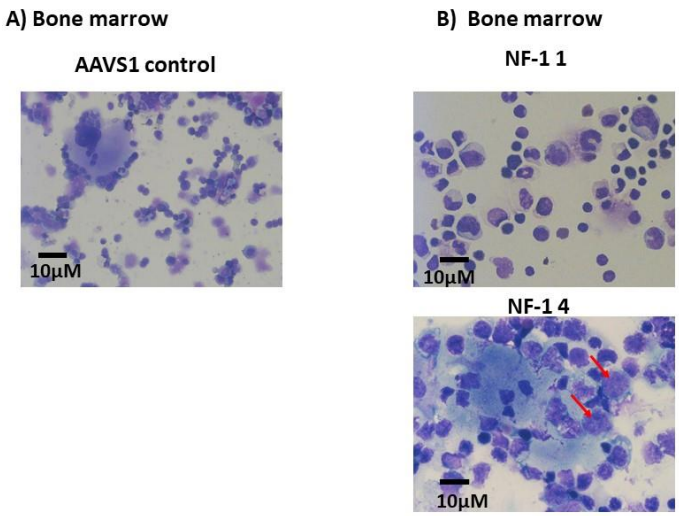


Figure 3.99: *NF-1* CB NSG recipients showed reduced megakaryopoiesis but no significant infiltrations of blast populations in the bone marrow Microscopy images of Giemsa-May-Grünwald stained bone marrow (smear *NF-1* 1, cytospin *NF-1* 4: both 100x magnification) **A)** *AAVS1* **B)** *NF-1* CB recipients

3.3.7 Summary

The overall aim of this chapter was to establish using an NHEJ based CRISPR editing approach whether induction of *NF-1* mutations in a fetal-like context were sufficient to recapitulate *in vitro* and *in vivo* characteristics of the rare childhood leukaemia JMML, and thus provide a valid xenograft model of the disease.

The spectrum of mutations and consequence of these at the protein level was first summarised from the two largest genomic analyses carried out on JMML patients by Stieglitz, Caye and colleagues (Stieglitz, Taylor-Weiner et al. 2015) (Caye, Strullu et al. 2015). This highlighted the broad mutational landscape of *NF-1* initiating mutations, as well as the presence of a range of additional secondary mutations within the RAS pathway and components of the PRC2 repressive complex. Importantly, there was evidence of patients with *NF-1* as the only detectable mutation, both heterozygous and compound heterozygous

(**Figure 3.68**). Further, the loss of function mutations abrogated protein function. These findings validated targeting *NF-1* using an NHEJ approach.

Guide RNAs were first tested in the TF-1 cell line which established at the bulk level that *NF-1* was successfully targeted at exon 2 and GM-CSF absent conditions slightly increased the proportion of edited cells within the population, although this was not statistically significant (**Figure 3.72**). At the clonal level, *NF-1* editing of the TF-1 cell line was sufficient to bestow a growth advantage both in the presence and absence of GM-CSF in five-cell sorts and in the presence of GM-CSF in single-cell sorts (**Figure 3.74**) where *NF-1* targeted cells produced significantly more colonies above 50 cells. Sanger sequencing at the targeted locus confirmed editing at *NF-1* exon 2 in TF-1 clones with no evidence of editing in the *AAVS1* control (**Figures 3.75 and 3.76**). Indeed, clones which showed editing by ICE (Synthego 2021) also demonstrated loss of *NF-1* at the protein level (**Figure 3.77**).

Similar *in vitro* approaches were then taken for the fetal-like primary human tissue CB. In support of the results found for TF-1 cells, *NF-1* editing of primary human CD34+ selected CB was sufficient to bestow GM-CSF hypersensitivity, in which *NF-1* guide 1 targeted single cells produced more colonies at 1ng/mL and 0.1ng/mL GM-CSF than control *AAVS1* targeted counterparts, although this was not statistically significant (**Figure 3.81**). This was in contrast to TF-1 cells in which growth of *NF-1* targeted colonies was significantly higher than controls, most likely due to the higher editing efficiency of TF-1 cells (**Figure 3.72**, 35% bulk level) compared to CB (**Figure 3.85**, 8% bulk level) meaning more cells would have received the edit and therefore be at an advantage of producing colonies under limiting cytokine conditions. Importantly, in the CB targeted

clones, there was evidence of all possible mono- and bi-allelic aberrations (**Figure 3.83**) and compound heterozygous mutations dominated in *NF-1* guide 1 targeted clones (**Figure 3.84**). Loss of both wild type *NF-1* alleles was crucial to increase the chance of expansion of an *NF-1* edited population *in vivo* as this meant increased chances of complete loss of protein function.

In vivo, two of four *NF-1* targeted CB NSG recipients succumbed to illness showing weight loss and pale limbs at 13 and 17 weeks post-transplant, a much shorter latency than seen for somatic inactivation of the *NF-1* locus in haematopoietic cells in mice (Le, Kong et al. 2004). Although these mice were severely anaemic and thrombocytopenic (**Table 3.3**), there was no definite evidence of a myeloid lineage expansion in the PB over the course of serial readouts up to 13 weeks (**Figure 3.87**). Interestingly, although the spleen size was not significantly increased in the *NF-1* group (**Figure 3.98 B**), reduced megakarypoiesis, determined by the absence of megakaryocytes and platelets in the BM, was seen for both *NF-1* recipients that fell sick (**Figure 3.99**) as well as an expansion of the CD45RA+ CD123- and LMPP populations in both *NF-1* samples compared to controls (**Figure 3.95**) consistent with data from xenografts of JMML patient samples (Caye, Rouault-Pierre et al. 2020) (Louka, Povinelli et al. 2021). Interestingly, the *NF-1* recipient that didn't fall sick had an expanded LMPP population but not the CD45RA+ CD123- population present in the two *NF-1* recipients that fell sick (**Figure 3.94**). It is therefore possible that the expansion of these abnormal progenitor populations represented an early, pre-leukaemic state and that additional genetic changes would be needed for full blown JMML to develop, as reported in some patients with JMML. Importantly, there was definite evidence of an expansion of the *NF-1* edited population in both

recipients which succumbed to illness, as well as the *NF-1* recipient that remained healthy at 26 weeks post-transplant (**Figure 3.97**) which contributed 37%, 31% and 32% of the total BM for *NF-1* 1, *NF-1* 4 and *NF-1* 3 respectively, which would be consistent with a pre-leukaemic population. Further experiments would be necessary to establish the significance of these findings. In order to establish which populations are capable of disease propagation, it would be interesting to firstly sort the two expanded populations and analyse these specific groups for whether they are enriched for *NF-1* editing compared to the total BM. Secondly, it would be interesting to transplant both these groups into secondary recipients and monitor their ability to engraft and cause disease development in the recipients. Ability to sustain the disease development seen in primary recipients would indicate either of these groups as the primary disease propagating population.

In conclusion, I have shown that *in vitro* in both the TF-1 cell line and human primary CB that *NF-1* editing alone was sufficient to drive GM-CSF independence (TF-1) and hypersensitivity (CB), a key characteristic of JMML. I have also shown that *in vivo* targeting of fetal-like HSPCs was sufficient to cause malignant expansion in the BM and cause a failure to sustain normal haematopoietic output, characterised by severe anaemia and thrombocytopenia in diseased recipients.

CHAPTER 4: CONCLUSIONS AND FUTURE DIRECTIONS

4. CONCLUSIONS AND FUTURE DIRECTIONS

4.1 Overview

My thesis took the form of both a method development and discovery project with one broad aim: to gain an in-depth understanding of the use of CRISPR genome editing approaches to study both normal and malignant haematopoiesis. This broad aim was divided into three principal objectives. The first was to establish an approach to achieve 90-100% CRISPR genome editing in the HSPC compartment of mouse BM cells, targeting a genetic locus that would allow edited cells to be readily tracked for their functional output and molecular profile. CRISPR is an exciting and versatile approach for both therapeutic genome correction (Wu, Zeng et al. 2019) (Khosravi, Abbasalipour et al. 2019) and disease modelling (Heckl, Kowalczyk et al. 2014) (Tothova, Krill-Burger et al. 2017) however it is well documented that efficiency varies greatly between cell types, with primary cells exhibiting lower editing efficiencies than their cell line counterparts. Therefore I tested three different modes and methods of CRISPR delivery in order to establish the optimal approach for primary mouse HSPCs. I also tested two editing strategies, one in which gRNA targeted cutting led to induction of expression of the tdTomato fluorescent reporter, the other in which guide RNA targeted cutting led to loss of tdTomato reporter expression.

In the second part of the project, I took the optimised mouse HSPC genome editing approach and applied it to investigate the short-term and long-term implications, on mouse HSCs, of CRISPR targeting a gene not associated with haematopoiesis. I targeted constitutively tdTomato-expressing mouse HSPCs with a single gRNA inducing loss of tdTomato expression and investigated the

changes in the transcriptional profile of edited HSCs compared to sham-edited, mock-nucleofected and fresh controls 24 hours post-nucleofection by RNA-Seq. This allowed me to investigate in detail the acute response to genome editing and separate each stage of the editing process into the unique changes in gene expression they caused. I then investigated whether there were sustained effects of genome editing on the function of HSCs by transplanting the edited HSPCs 24 hours post-nucleofection and tracking their lineage output for 18 weeks. Similarly, in order to establish whether any of the acute transcriptional changes were sustained long-term, I performed RNA-Seq on edited HSCs purified from the BM of recipients 18 weeks post-transplant. By targeting a fluorescent transgene reporter, I was able to track the contribution of edited cells to all peripheral blood lineages, including platelets and red blood cells, the output of which is one of the key factors in assessing the function of stem cells. Such an analysis would have not been possible if the usual pan-haematopoietic cell surface marker CD45 had been used as this is not expressed on platelets or red blood cells. Further, targeting the fluorescent transgene allowed for the removal of any potential changes seen resulting from the chosen target gene, since tdTomato has no function in maintenance of normal haematopoietic output.

In the final part of my project, I focused on using the optimised CRISPR RNP nucleofection approach towards developing an *in vivo* model of the childhood leukaemia, JMML, which can be caused by mutations in the *NF-1* gene. *NF-1* is a tumour suppressor and both heterozygous/homozygous loss of function is associated with lymphoid and myeloid leukaemias and many solid cancers (Yap, McPherson et al. 2014), as well as with therapy resistance (Philpott, Tovell et al. 2017) (Cohen 2021) (Caye, Strullu et al. 2015, Stieglitz, Taylor-Weiner et al.

2015). In the case of germline *NF-1* heterzygous mutations, these are often accompaned with somatic mutation of the other allele leading to compound heterzygous mutations as well as loss of heterozygosity causing a loss of the remaning wild type allele (Eisenbarth, Beyer et al. 2000). *NF-1* mutations are particularly common in JMML, with 15% of patients harbouring *NF-1* loss of function mutations (Caye, Strullu et al. 2015) (de Vries, Zwaan et al. 2010). In order to establish NHEJ based genome editing at the *NF-1* locus as a viable approach to model key features of JMML, from its initiation in a humanised setting, I targeted this locus in human CB. I initially used *in vitro* approaches to determine whether *NF-1* editing alone was sufficient to establish GM-CSF independence, a characteristic feature of JMML. Having established a mild GM-CSF sensitivity phenotype, I then transplanted *NF-1* edited CB into NSG recipient mice and monitored human cell chimerism and the development of disease.

4.2 Optimisation of CRISPR in primary mouse HSPCs using *in vitro* methods

Using both cell lines and primary cells, I systematically analysed the efficiency of three modes of delivery of the CRISPR machinery: plasmid vector, RNP and viral vector; and three methods of CRISPR delivery: electroporation, nucleofection and transduction. Extensive analysis of both electroporation and nucleofection methods in the Ba/F3 cell line established nucleofection to be the optimal approach at lower plasmid concentrations (**Figure 3.10**, pg.171). This is most likely because the nucleofection buffer delivers the CRISPR machinery directly to the nucleus without requiring the host cell machinery (Lonza 2021). This was

an important comparison, particularly in the context of primary cells which are more challenging to transfect efficiently. Indeed, other studies have compared lipofection with electroporation (Maurisse, De Semir et al. 2010) in the HEK293T cell line as well as investigations into the most optimal nucleofection buffer (Flanagan, Gimble et al. 2011).

In primary cells, the observation that nucleofection was superior to electroporation was even more evident (**Figure 3.13**, pg.175). By directly comparing viral transduction, plasmid and RNP nucleofection using the same gRNA, it was possible to evaluate the key advantages and disadvantages of each approach within the context of primary mouse HSPCs. In this way, I found that while high transfection efficiencies could be achieved using plasmid delivery, this came at a high cost for cell viability (**Figure 3.19**, pg.186). On the other hand, while viral transduction didn't reduce cell viability to the same extent as plasmid nucleofection, the efficiency was particularly low for gRNA + Cas9 vectors despite small vectors demonstrating high efficiency. This was most likely due to the larger size of the Cas9 gRNA vectors. Indeed, others have used approaches in which the Cas9 and gRNA are delivered on separate vectors in order to reduce the size of the construct, as well as allow delivery of guide pools into the same cell (Heckl, Kowalczyk et al. 2014). What was evident was that the most efficient approach for targeting the tdTomato locus in mouse HSPCs was RNP nucleofection (**Figure 3.23**, pg.192) with 90-100% editing in the LSK compartment. Indeed, this method is the one that was selected for the therapeutic strategy using NHEJ mediated disruption of the *BCL11a* enhancer for treatment of sickle cell disease and beta thalassaemia (Wu, Zeng et al. 2019) (Frangoul, Altshuler et al. 2021). On the other hand, in cases where HDR is required for successful therapeutic

editing, virus transduction alone or in combination with RNP nucleofection remains the optimal method due to the large size of the donor template and advantage of sustained expression (Dever, Bak et al. 2016); although for small donor templates it is possible to introduce these by nucleofection as ssODNs together with the RNP (Salisbury-Ruf and Larochelle 2021).

The efficiency of HDR and NHEJ approaches in primary HSPCs reported by others (Bak, Dever et al. 2017) (Mettananda, Fisher et al. 2017) were in line with the results seen for targeting tdTomato induction and knockdown in this section of the thesis. Using a dual guide approach to target removal of the STOP cassette within the Ai9-Rosa26 mouse strain proved to be an inefficient process despite high transfection and transduction efficiency (**Figures 3.14** pg.178 and **3.15** pg.180). This was likely due to the fact that in order to ensure that the tdTomato gene remained in frame, editing without out of frame insertions/deletions was required and this is unlikely to occur with an NHEJ approach. This strategy could have been optimised with the introduction of a donor template although the efficiency would most likely have been lower than the knockdown approach as it would have relied on HDR. Overall, a highly efficient genome editing approach was achieved targeting the tdTomato locus for knockdown which is in line with current methods for therapeutic targeting (Wu, Zeng et al. 2019) (Frangoul, Altshuler et al. 2021) (Mettananda, Fisher et al. 2017).

4.3 Molecular and functional characterisation of CRISPR edited mouse HSCs

To assess the molecular and functional effects of CRISPR-mediated gene editing, I interrogated each stage of the nucleofection process for its specific effect on the cell cycle status, proportion of cells in each HSPC compartment and transcriptional profile of mouse HSCs at 24 hours after nucleofection. I then assessed the long-term lineage output and transcriptional profile of engrafted, edited HSCs at 18 weeks after transplantation.

There was no significant effect of gene editing on the cell cycle status for all HSPC compartments tested at 24 hours (**Figure 3.32**, pg.210) in line with the findings of Schiroli and colleagues. They reported a change in the proportion of cells at each stage of the cell cycle for low specificity, but not high specificity, RNPs and this was a transient change which had resolved by 24 hours (Schiroli, Conti et al. 2019). Amplification of the top, off-target tdTomato sites confirmed this guide to be of high specificity which supports the lack of effect on cell cycle status also seen for high specificity RNPs by Schiroli and colleagues (Schiroli, Conti et al. 2019). Despite no difference seen in cell cycle status, the frequency of LSs was reduced (**Figure 3.30**, pg.206) suggesting a reduced output of progenitors from the HSC compartment. A similar finding was reported by Wang and colleagues in mouse LSKs subsequent to acute DNA damage caused by whole body ionising irradiation (Wang, Schulte et al. 2006). These findings support a change in the functional output of HSCs post DNA damage induction of multiple origins.

Analysis of the transcriptomic data also separated each stage of the nucleofection and editing process by differential gene expression. Unexpectedly, the most prominent segregation came when fresh cells were compared to all other controls (**Figure 3.38**, pg.220) and there was a graded enrichment for genes associated with the *TP53* and DNA damage response across the nucleofection and editing procedure (**Figure 3.42**, pg.228). This suggests an additive effect on transcriptional response, in which nucleofection itself elicits a stress response which is amplified with the addition of Cas9, a gRNA, and a DSB. Interestingly, Schiroli and colleagues also reported the highest *TP53* response emanating from cells subject to transfection/transduction with low specificity RNPs in combination with AAV6 virus (Schiroli, Conti et al. 2019) and that this response was partially abrogated by overexpression of the cell cycle survival promotion protein E4orf6/7 (Ferrari, Jacob et al. 2020).

To assess whether the acute effects of gene editing on HSCs were maintained long-term, I measured the lineage output of mature blood cells serially in the PB and in the BM, PB and spleen 18 weeks after transplantation with gene edited HSCs. This showed that HSCs targeted with a high specificity RNP were capable of contributing to all blood lineages in the mouse: B cells, T cells, myeloid cells, platelets and red blood cells (**Figure 3.51**, pg.244) although red blood cell and platelet chimerism reduced slightly over a time in a way that was not evident in the other lineages or the control conditions (**Figure 3.52**, pg.246). Indeed, a similar lineage defect was reported by Wang and colleagues in response to DNA damage in mice caused by irradiation (Wang, Schulte et al. 2006). This might suggest a partial defect in stem cell output, or a higher engraftment of ST-HSCs vs LT-HSCs in the edited condition although this could also be a result of the low

overall engraftment of the edited cells compared to control conditions, which exaggerated the apparent platelet/red blood cell drop off. Indeed, Wilkinson and colleagues demonstrated robust engraftment of HDR edited HSCs at similar levels in all lineages 20 weeks post-transplant when LT-HSCs were sorted and pre-cultured for seven days in F12 medium with specific supplements, TPO and SCF. This suggested a pre-enrichment of LT-HSCs could rescue the lower level engraftment of edited cells by transfecting a higher amount of LT-HSCs at the initial step hence increasing the chances of high numbers of LT-HSC survival (Wilkinson, Dever et al. 2021).

What was most striking from the transcriptional data 18 weeks post-transplant was that tdTomato edited samples remained distinctly separated from their control counterparts by differential gene expression analysis (**Figure 3.60**, pg.256) which demonstrates that transcriptional changes associated with gene editing were not all together transient. Interestingly, numbers of differentially expressed genes between tdTomato edited and fresh controls were much higher at the 24-hour time-point than at the 18-week time-point. By contrast, in the NT RNP group, the pattern was the opposite with higher numbers of differentially expressed genes at 18 weeks post-transplant compared to 24 hours post-nucleofection (**Figures 3.39**, pg.221 and **3.60**, pg.256). Two main hypotheses could explain the transcriptional segregation of the tdTomato targeted group. The first is that, within the heterogeneous population of HSCs, a specific sub-type selectively survived the editing process and were therefore enriched in the tdTomato group. Alternatively, the altered transcriptional profile may have been a reactionary change induced by the editing process itself.

In terms of the genes specifically upregulated in the tdTomato targeted group 18 weeks post-transplant, within the differential expression and random forest analysis, upregulated genes were enriched for those that have been reported in states of haematopoietic senescence (*Cxcr4*, *Clec7a*, *Ms4a1, Xbp1* and *Ptger4*) (**Figures 3.62**, p.258 and **Figure 3.65**, pg.262) (Capone, Connor et al. 2018) as well as deep quiescence (*Ctse* and *Cyp4f18*) (Fujimaki, Li et al. 2019). Interestingly *Clec7a* has also been reported in another study to be upregulated in response to DNA damage induction (Takacova, Slany et al. 2012). This pattern of upregulated genes does not fit a signature of a specific stem cell sub-population and was therefore more likely to be a reactionary change as opposed to selection by genome editing of the survival of a specific sub-population. Similarly, the lack of overlap between transcriptomic signatures of tdTomato targeted samples 24 hours post-nucleofection and 18 weeks post-transplant (**Figure 3.66**, pg.264) also suggests a differential short-term and long-term reactionary response as, if selection of a specific sub-population occurred *in vitro* at the time of editing, it would be expected that this sub-population and therefore transcriptomic signature would be propagated *in vivo*.

Therefore, it can be concluded that nucleofection altered the HSC transcriptional profile of targeted murine HSCs which in the short-term was enriched for *TP53* and DNA damage response pathways. These pathways were more strongly enriched with the introduction of a non-targeting RNP and even further enriched with a targeting RNP. In the long-term, the transcriptional profile of targeted murine HSCs showed greater diversion from non-targeted HSCs with enriched gene expression for signatures of stem cell quiescence and senescence. These long-term molecular changes were reflected at the functional level by reduced

overall lineage engraftment and reduced lineage chimerism particularly in red blood cells and platelets. These observations could be interpreted as a unique transcriptional signature resulting from the CRISPR induced DSB which altered the overall molecular and functional characteristic of targeted HSCs. Another explanation for these observations could be that certain sub-populations within the HSC compartment, particularly LT-HSCs, could be less resistant to DNA damage induced by CRISPR, and that these less resistant sub-populations were less likely to survive engraftment. Hence, proportionally they were reduced in the targeting group compared to other nucleofection groups. On the other hand, nucleofection itself although causing changes in transcription at the acute phase, had a lesser impact on the overall fitness and engraftment potential of HSCs.

4.4 Introduction of the JMML associated *NF-1* loss of function mutation in human cell lines and primary haematopoietic cells

Using an RNP nucleofection approach, I targeted the *NF-1* loss of function mutation which is particularly common in patients with JMML (15%) (Caye, Strullu et al. 2015). As a first step, I analysed the mutational spectrum of a cohort of patients with *NF-1* as the initiating genetic lesion to identify potential hotspots for gRNA targeting. Although there were a number of mutated regions shared among patients, there wasn't a specific hotspot (**Table 3.1**, pg.275) however in each case the consequence was truncation or loss of protein function. This analysis also confirmed cases of JMML in which the only genetic lesion was *NF-1*, confirming this CRISPR approach as a viable system for modelling the disease (**Figure 3.68**, pg.276). Initial targeting in the TF-1 human erythroleukaemia cell

line confirmed 30-35% editing at the *NF-1* locus for optimal guides which, at the bulk level, increased to 40-45% in the absence of GM-CSF suggesting a selection for edited cells under these conditions (**Figure 3.72**, pg.281). Through limited cell number sorting (1 or 5 cells) it was demonstrated that *NF-1* editing increased the clonogenicity of TF-1 cells both in the presence and absence of GM-CSF (**Figure 3.74**, pg.286). Given that the TF1-1 cell line is dependent on GM-CSF (Kitamura, Tange et al. 1989) and that cell growth is significantly abrogated in the absence of the cytokine (**Figure 3.70 A**, pg.279), the clonogenic output of *NF-1* targeted cells in the absence of GM-CSF offered the first strong evidence for induction of one of the diagnostic JMML phenotypes. To validate this, clones were genotyped for *NF-1* editing and grown up as cell lines in which the entire population harboured the desired edit (**Figure 3.76**, pg.289). Most mutations were single base pair insertions causing a frameshift; this type of edit is the most common seen in *NF-1* mutated JMML patients although at varying genomic loci (Caye, Strullu et al. 2015) (**supplementary Table 6**). Importantly, all clones were grown exclusively in GM-CSF containing medium such that no GM-CSF hypersensitivity was pre-established. Clonal lines harbouring the *NF-1* lesion were confirmed for loss of function at the protein level (**Figure 3.77**, pg.290) which validated the edited region in exon 2 as sufficient to prevent translation. Consistent with this result, two mutated clones challenged with medium lacking GM-CSF showed cytokine independent growth at a rate equivalent to un-edited controls in the presence of GM-CSF (**Figure 3.78**, pg.291). This confirmed that the *NF-1* edit was sufficient to rescue GM-CSF dependence in the TF-1 cell line.

GM-CSF hypersensitivity is a key phenotypic characteristic of JMML patient cells, irrespective of the initiating mutation (Kotecha, Flores et al. 2008) (Emanuel, Bates et al. 1991) (Niemeyer, Kang et al. 2010) (Matsuda, Shimada et al. 2007) (Metzner, Horstmann et al. 2007) (Schubbert, Zenker et al. 2006) (Schubbert, Bollag et al. 2007). The validated *NF-1* guides were then taken forward into human CB cells in order to model the fetal-like context predicted to support the progression of JMML from RAS pathway mutations (Loh 2011) (Helsmoortel, De Moerloose et al. 2016). Having established the minimum cytokine combination (10ng/mL SCF, 1ng/mL GM-CSF) required to significantly reduce but not completely abrogate single cell colony growth from CD34+Lin- CB HSPCs (**Figure 3.79**, pg.293), CD34-selected CB HSPCs were targeted with the *NF-1* exon 2 guides and subject to single cell colony growth. *NF-1* edited CD34+Lin- CB HSPCs demonstrated higher colony output compared to *AAVS1* targeted controls of which the difference increased with reducing GM-CSF concentration (**Figure 3.81**, pg.296), although at both GM-CSF concentrations tested this was not statistically significant. This higher colony output in low GM-CSF concentrations suggested the generation of a characteristic JMML phenotype through *NF-1* targeting of CB HSPCs. Crucially, both heterozygous and compound heterozygous clones were generated (**Figure 3.83** pg.298) which recapitulated the spectrum of *NF-1* only JMML patient data (**Figure 3.68**, pg.276).

In order to determine whether edited *NF-1* was able to transform CB cells into a JMML-like disease and to model JMML *in vivo*, *AAVS1* control and *NF-1* edited CB HSPCs were transplanted into sub-lethally irradiated NSG mice. Of the four *NF-1* mice transplanted, two had to be culled after 13 and 17 weeks because of

severe anaemia, thrombocytopenia and weight loss (**Table 3.3**, pg.307). This was a much shorter disease latency than seen by Doan and colleagues using somatic Mx1-Cre mediated *Nf-1* loss in mouse HSCs (Le, Kong et al. 2004) although not as acute as seen for somatic inactivation of *Nf-1* in a *Kras*-G12D background in mouse HSCs (Cutts, Sjogren et al. 2009). Notably, Mx1-Cre systems are generally ~95% efficient in targeted haematopoietic tissues whereas, in the NSG experiment described in *Results Chapter 3*, 8% of transplanted CB harboured the *NF-1* mutation (**Figure 3.85**, pg.301) thus it was necessary for the small edited clone to outcompete a much larger proportion of normal cells to cause disease. Sanger sequencing in the BM of the 13wk and 17wk diseased mice showed 31% and 37% editing respectively at the time of culling (**Figure 3.97**, pg.315). Indeed, the most abundant edit for the *NF-1* 4 recipient was a single base pair insertion, also seen at the time of transplant, with lower levels of single or double base pair deletions. This indicated the presence of a dominant clone with potential sub-clones although these were on the threshold of the ICE sensitivity limit (Synthego 2021). It is not known, however, whether the *NF-1* edited populations in either of these cases acquired additional mutations upon transplantation as deep-sequencing of haematopoietic tissue post-transplant in the diseased mice was not carried out. It is highly possible that additional mutations were acquired given that the *NF-1* CB recipient that didn't fall sick (recipient 3) had similar levels of *NF-1* editing in the BM as the two that fell sick (32% **Figure 3.97**, pg.315). The most evident changes in haematopoietic output between the two culled *NF-1* recipients and the *AAVS1* controls were within the PB and BM CD45 chimerism and the BM HSPC compartment. In the PB the human CD45 compartment was increased compared to controls (**Figure**

3.91 A, pg.308). In addition, within the CD34+ CD38+ compartment, there was a significant increase in the CD45RA+ CD123- population with a concomitant decrease in CMPs (**Figure 3.95**, pg.312). Further, within the CD34+ CD38- compartment, there was a significant increase in the LMPP population with a concomitant decrease in HSCs and MPPs (**Figure 3.95**, pg.312). Interestingly, only the LMPP population and not the CD45RA+ CD123- was expanded in the *NF-1* recipient that didn't fall sick (**Figure 3.94**, pg.312). Indeed, this CD45RA+ CD123- population has been observed in NSG recipients of JMML patient BM HSC in previous work carried out by Eleni Louka in the lab of Adam Mead (Louka, Povinelli et al. 2021). LMPP populations from JMML patients have also been shown to be able to propagate the disease *in vivo* (Caye, Rouault-Pierre et al. 2020). Despite these aberrant expansions, thrombocytopenia and anaemia, the spleen size of *NF-1* recipients was not significantly larger (**Figure 3.98**, pg.316) and there was only low level detection of blasts in the bone marrow of *NF-1* 4 (**Figure 3.99**, pg.317). There was over a clear lack of megakaryocytes and platelets in the BM of both *NF-1* recipients that fell sick compared to the *AAVS1* control (**Figure 3.99**, pg.317) which is a clinical feature of JMML patients (Niemeyer and Flotho 2019) and correlates with the thrombocytopenia observed in these mice. Due to the fact that both lymphoid and myeloid lineages can arise from the CD45RA+ CD123- population, as well as lack of evidence of aberrant B or myeloid expansion in the PB and BM of edited groups (**Figure 3.91**, pg.308), with the exception of *NF-1* 4 which showed slightly higher B cell engraftment in the PB, it is difficult to conclude with the current data whether this was a lymphoid or myeloid driven malignancy. A more extensive BM panel including a wider range of mature lineage and progenitor markers would be required to pin-point

the defect. Indeed, it is not impossible that in this setting a lymphoid driven malignancy could have arisen from *NF-1* CB editing and NSG transplant. This is due to the fact that there is evidence of *NF-1* mutations in children with B-acute lymphoblastic leukaemia (Galbiati, Lettieri et al. 2013) and B-cell lymphoma in an adult patient harbouring an *NF-1* mutation (Kim, Seo et al. 2003).

4.5 Implications of this work

4.5.1 Direct comparison of three modes and methods of CRISPR delivery in primary HSPCs targeting the same non-haematopoietic genomic locus with the same guide RNA

In this project I have extensively compared the currently available CRISPR editing approaches in mouse HSPCs, a group of primary cells notoriously difficult to target. By using the same guide RNA for all delivery approaches tested (transduction, RNP and plasmid electroporation/nucleofection), I have provided a detailed comparison of the advantages and disadvantages of each approach *in vitro*. While the different approaches have been compared across different studies in several reviews (Lino, Harper et al. 2018) (Jensen, Axelgaard et al. 2019), this is the first study in which the three major approaches have been directly compared in the same primary cell type targeting the same locus. This has allowed more accurate conclusions about the optimal approach to be drawn as results were not confounded by the genomic locus targeted or gRNA sequence, which can have a big impact on efficiency (Jensen, Fløe et al. 2017). Further, targeting the tdTomato fluorescent reporter as opposed to an endogenous gene has allowed the effect of genome editing to be studied without the potential confounding effects of the specific locus targeted.

4.5.2 Use of tdTomato RNP editing as a proxy for editing at another locus

Another advantage of targeting the tdTomato reporter in a constitutive background is that this optimised RNP approach can now be used to enrich for editing at other loci. Given that RNP editing efficiency has been shown to be coupled to transfection efficiency in this study (**Figures 3.14** pg.178 and **3.18** pg.184), it can be predicted that in the context of delivery two RNPs, one targeting tdTomato and the other an endogenous genomic region, loss of tdTomato expression could predict editing at the other locus. In this way this method could be used as a tool, for example where high levels of edited cells are required to develop an *in vivo* disease model, to pre-select for cells that are more likely to be editing by sorting for tdTomato negative cells prior to transplantation. This would of course require waiting until 48-72 hours post-nucleofection when tdTomato protein expression would be lost.

4.5.3 Evidence for potential enrichment of TP53 mutated cells by CRISPR editing

The RNA-Seq data at 24 hours post-nucleofection has provided evidence that in the short-term CRISPR editing increased the DNA damage and *TP53* responses in mouse HSCs, which supports previously published data in human CD34+s demonstrating the same transcriptomic effect in HSCs (Schiroli, Conti et al. 2019), as well as data demonstrating that introduction of a *TP53* dominant negative mutant transgene increased HDR and stem cell survival in human

HSPCs (Ihry, Worringer et al. 2018). These results provide additional supporting evidence for a need to take precaution using this approach in the therapeutic setting as both in human and mouse HSCs it has been found that *TP53* pathway upregulation was a common response to editing. This implies that cells in which *TP53* is already mutated might be more likely to survive the process and therefore CRISPR genome editing could enrich for *TP53* mutated populations. This question would need to be addressed by editing an inert locus, for example tdTomato in a background of *TP53* mutated cells, transplanting these and comparing whether *TP53* edited cells expand in the engraftment setting at a higher rate than *TP53* wildtype cells. If this was the case, it would of paramount importance in the clinical setting to screen patients, particularly elderly patients with clonal haematopoiesis of indeterminate potential (CHIP) for any sign of *TP53* mutations before undergoing CRISPR editing based therapy and set a threshold where this is too high to safely undergo CRISPR therapy. It would also be important to monitor patients currently on CRISPR therapy for emergence of *TP53* clones to be able to address these as early as possible if this occurs. Further investigation required to prove this link is discussed in *Section 4-6 Future directions*.

4.5.4 Evidence for changes in HSPC composition and function

This is the first study which has reported serial contributions of edited HSCs to lineage output and therefore allowed the chimerism to be tracked over a considerable period. The reduced platelet and red blood cell chimerism over the course of transplant provide additional evidence supporting a possible HSC

defect following genome editing. In the context of mouse competitive transplant, the reduced platelet and red blood cell chimerism didn't result in thrombocytopenia or anaemia due to wild type competitors being co-transplanted therefore contributing to haematopoietic output. However, in the context of patients, this would not be the case- haematopoietic output would be entirely reliant on the edited HSC and therefore if this phenotype was seen in patient recipients of CRISPR edited HSCs it could lead to thrombocytopenia and anaemia. This interpretation however is speculative and would require secondary transplants to test whether the lower platelet and red blood cell chimerism is a long-term defect seen for these edited HSCs. In addition, a similar experiment performed in human HSCs in a xenograft setting would also be required to confirm that this phenotype is not limited to murine HSCs. Further, there is evidence in murine HSCs that reduced chimersim could be avoided by pre-enriching for LT-HSCs as demonstrated by Wilkinson and colleagues. In a murine HSC transplant study they demonstrated rescue of the lower level engraftment of edited cells by transfecting a higher amount of LT-HSCs at the initial step (Wilkinson, Dever et al. 2021). Another important consideration is the possible selection of small *TP53* mutated populations by CRISPR leading to their clonal expansion, due to evidence of up-regulation of the *TP53* pathway as an immediate consequence of genome editing. This transcriptomic response has been reported in other primary cell studies (Schiroli, Conti et al. 2019) (Conti and Micco 2018) and it has been shown that Cas9 is more active in *TP53* mutated cell lines (Enache, Rendo et al. 2020). Given the tragic consequences of multiple previous gene therapy approaches (Cavazzana-Calvo, Hacein-Bey et al. 2000) (Check 2002) (Lehrman 1999), it is critical to prove that CRISPR genome editing

does not enrich for *TP53* mutated populations in primary human cells, as this could potentially lead to expansion of malignant populations.

4.5.5 Direct comparison of short-term and long-term transcriptomic consequences of CRISPR editing

While several studies have analysed the short-term transcriptional consequences of CRISPR editing on HSCs at 12 hours (Ferrari, Jacob et al. 2020) and 24 hours (Schiroli, Conti et al. 2019) post-transfection, this is the first study in which CRISPR editing at the same locus in the same cell type has been analysed for the changes in gene expression both at the acute phase and in the long-term. These data support the findings of Schiroli and colleagues that the *TP53* response to genome editing was transient (Schiroli, Conti et al. 2019) as well as providing new evidence that edited HSCs were transcriptionally changed and crucially that this transcriptional change was not seen in non-targeting controls, indicating a direct impact of genome editing itself. The enrichment of genes associated with senescence and deep quiescence provide new evidence that edited HSCs might not function in exactly the same way, which suggests additional caution should be taken in the therapeutic context. However, this study did not provide evidence that the transcriptional signatures seen in edited mouse HSCs would be recapitulated in their human counterpart; given that human and mouse HSCs don't always respond in the same way to DNA damage (Biechonski and Michael 2013) it could be the case that long-term transcriptional consequences of genome editing are different in human HSCs.

4.5.6 *NF-1* targeting as an *in vitro* JMML model

In vitro editing of *NF-1* in human CB has provided important evidence that this single genetic aberration is sufficient *in vitro* to increase the clonogenic output of a fetal-like cell type in low GM-CSF levels although this was not statistically significant at the concentrations used. This provides evidence that this single RAS pathway associated genetic lesion could be sufficient to recapitulate a key characteristic of JMML, although additional data at lower GM-CSF concentrations down to 0.01ng/mL, as other studies have used (Kotecha, Flores et al. 2008) (Schubbert, Lieuw et al. 2005), would increase the confidence of the hypersensitivity phenotype suggested by the current data. While this GM-CSF hypersensitivity has been shown in other studies using patient cells as well as introduction of JMML-associated mutations in murine fetal liver (Emanuel, Bates et al. 1991) (Metzner, Horstmann et al. 2007) (Kotecha, Flores et al. 2008) (Donovan, See et al. 2002) (Chan, Leedy et al. 2005), this study has allowed assessment of the effect of the single initiating event *NF-1* and a human fetal context on the ability to drive GM-CSF independence. On the other hand, previous studies on human patient cells didn't provide evidence of the minimum initiating genetic aberrations required to drive this cytokine independence and the murine derived models lack the human cell context.

4.5.7 *NF-1* targeting as an *in vivo* JMML model

In vivo engraftment data provided evidence that *NF-1* initiating mutations were sufficient to drive disease development in NSG recipients leading to severe thrombocytopenia and anaemia. Indeed there was also evidence of an aberrantly

expanded CD45RA+ CD123- population in the BM which has previously been seen in NSG transplant of JMML BM HSCs. Further the LMPP population, which was shown to be expanded in *NF-1* CB recipient BM compared to *AAVS1* CB recipient BM, has also been shown to be capable of *in vivo* JMML disease propagation in patient derived xenograft models (Caye, Rouault-Pierre et al. 2020). Although based on the current evidence, the disease development didn't completely recapitulate the *in vivo* phenotype of JMML, it is the first study in which *NF-1* targeting alone in the context of human fetal- like CB has caused disease development and abnormal haematopoiesis in a xenograft setting. Indeed previous studies have targeted *NF-1* alone but this was in a mouse haematopoietic context (Le, Kong et al. 2004). Tothova and colleagues have also reported a myeloid neoplasm in *NF-1* edited CB however this was in the context of multiple concurrent genetic aberrations (Tothova, Krill-Burger et al. 2017). Therefore this study has provided a system in which to study disease development from an *NF-1* initiating event in fetal-like HSPCs. Further experimentation opportunities are discussed in *4.7 Future directions*.

4.6 Limitations of the experimental approaches

4.6.1 Refining editing techniques for different cell types

Although the *in vitro* optimisation approaches were extensive, they only addressed specific cell lines and primary cell types and therefore conclusions of optimal CRISPR approaches can only be deduced for these specific cell types tested. Further, the conclusions drawn surrounding the optimal CRISPR approach were only analysed in an *in vitro* context, in order to reduce the *in vivo* experimental burden. In order to more robustly address this question, all

transfection (plasmid and RNP, electroporation and nucleofection) and transduction approaches could have been evaluated *in vivo*. It is possible to hypothesise that although a certain approach was more efficient *in vitro*, another approach could have led to more stable engraftment and therefore be a better method in the context of sustained long-term lineage output.

4.6.2 Time-points of transcriptomic evaluation

Although both a short-term and long-term time-point were taken to analyse the transcriptomic effects of gene editing, it is possible that some acute changes were missed by not taking earlier time-points. Given that editing by RNP CRISPR has been shown to occur as early as two hours post-transfection (Wu, Zeng et al. 2019), it is possible that the peak in transcriptional differences occur earlier. On the other hand, it is also possible that the 24-hour time-point was too early and transcriptional changes are delayed compared to the point of DSB induction. Schiroli and colleagues reported the highest level of DSB associated protein at 24 hours post-transfection (Schiroli, Conti et al. 2019). Therefore it could be hypothesised that significant transcriptional changes occur subsequent to this. Repeating this experiment at multiple earlier and later time-points would provide valuable data to address these further questions. Additional long-term time-points, for example every four weeks after transplant would also have provided additional evidence of the potential accumulation of transcriptional changes within the edited population.

Further, at the 24-hour time-point, cells were stringently sorted for viability which could have led to the omission of certain transcriptional changes, given the

evidence of apoptosis induction by genome editing. A more accurate depiction of the short-term impact may have been revealed by performing RNA-Seq directly on the HSC population without sorting based on viability, as other studies have done (Ferrari, Jacob et al. 2020).

4.6.3 Impact of species differences

An additional limitation of this approach was the use of mouse HSPCs in order to predict the therapeutic implication of this technology. Murine HSPCs were used as the model system in this study due to the major benefit of targeting a fluorescent reporter for tracking of edited cells. In addition, primary murine HSPCs are much more readily available than primary human HSPCs which allowed the study of a wide range of controls to dissect each separate stage of the transfection and editing process on the molecular and functional properties of HSCs. However, given that the double strand breaks, which are initiated by CRISPR and the consequence of which are the cause for concern and study, are not always repaired in the same way by mouse and human HSCs (Biechonski and Michael 2013), it can only be hypothesised that the transcriptional and functional changes seen for mouse HSCs would be recapitulated in the human context. Therefore, it would be necessary to use a HDR based fluorescence integration approach in human HSCs in a xenograft setting to determine whether the consequence of editing on human HSCs are the same.

4.6.4 Examination of off-target editing

Although top off-target cutting sites were predicted for the tdTomato gRNA *in silico* and interrogated *in vitro* for the presence of editing at these sites, it is possible that this evaluation didn't represent all the off-target events. It is possible that there were other off-target edits not predicted by the *in silico* Cosmid tool (Cradick, Qiu et al. 2014). Whole genome deep sequencing before and after editing would be required to reliably confirm that the single off-target event detected by Sanger sequencing was indeed the only one present.

4.6.5 Modelling childhood human leukaemias using gene editing of normal CB and limited recipient numbers

A major limitation of the *NF-1* targeting approach was that it did not induce overt myeloid leukaemia, nor did it show distinct domination of blast cells by morphology in the BM, despite causing disease on-set in the mice. This could be due to the timing of sacrifice, at which the disease on-set was not progressive enough to present a dominant blast infiltration or the presence of only a single gene edit. It is possible that in this transplant setting *NF-1* was sufficient to develop a disease initiation state but that a second hit, or secondary transplantation would be required to recapitulate the full characteristics of JMML. It is also possible that despite the primitive fetal-like HSPCs in CB, this context was not enough to drive a JMML phenotype but that this would only be possible in the context of fetal liver. Another limitation of this study was the use of only four replicate recipients per group. This is a particularly small number to be able to reach accurate conclusions about the effect of *NF-1* editing in cord blood. This

number of recipients was chosen initially due to empirical data from previous experiments in the lab (not shown) that engraftment of cord blood in NSG recipients can be low when less than 30,000 CD34+ cells are injected. Due to the limited amount of cord blood available, this meant only four recipients could be transplanted per group to ensure 30,000 or above CD34+ cells were injected. In hindsight, given the high level of engraftment for both *NF-1* and *AAVS1* edited cord blood, it would have been possible to halve the amount of starting material injected and therefore use double the number of recipients per group. This would have generated a more accurate result of the effect of *NF-1* cord blood editing as it would have increased the numbers of possible diseased mice and increased the chances of replicating the CD45RA expanded phenotype seen in two of the *NF-1* recipients. This limited number of CB CD34+ cells also meant that there weren't enough to confirm *NF-1* protein loss by western blot, as was performed in the TF-1 cells, due to this approach requiring ≥ 1 million cells. Having enough cells to confirm protein loss by western blot would have validated that the successful editing seen by Sanger sequencing did indeed lead to loss of protein function. Finally, higher numbers of CB CD34+ cells would have allowed additional transplant controls, for example nucleofection without a guide RNA and Cas9, as performed in results chapter 2, which would have complemented the *AAVS1* safe harbour control in confirming the phenotype of *NF-1* editing resulting from loss of tumour suppressor function as opposed to a potential consequence of the nucleofection process.

4.6.6 NSG mice as a xenograft model for myeloid malignancy

Further, the use of NSG mice as recipients for modelling a myeloid malignancy may have contributed to the lack of complete recapitulation of a JMML phenotype. This is due to the fact that NSG mice are optimised to support human lymphoid engraftment (Theocharides, Rongvaux et al. 2016). Indeed, Tothova and colleagues reported generation of a myeloid neoplasia by targeting CD34+ CB with four genetic lesions (*TTP53, RUNX1*, *EZH2* and *NF-1*) (Tothova, Krill-Burger et al. 2017) and transplanting into NSG-S recipients, which support more balanced human lineage engraftment (Wunderlich, Chou et al. 2018). As an alternative approach, co-operating mutations could be introduced into *NF-1* edited human HSCs, for example *ASXL-1* or *EZH2*, post short-term xenograft transplant, in order to model the accumulation of genetic aberrations, followed by secondary transplanting into NSG-S recipients. It can be hypothesised that both the multiple genetic aberrations and improved support of myeloid engraftment could lead to an aggressive myeloid expansion.

4.7 Future directions

4.7.1 Use of tdTomato reporter to enrich for edited populations

As alluded to in *Section 4.5.2*, the optimised protocol for tracking edited mouse HSPCs has the potential to be used in combination with targeting of additional genes in order to easily enrich for edited samples. For example, in the context of modelling haematological malignancies in mouse HSPCs, it will be possible to use the constitutive Rosa26-tdTomato reporter to introduce RNPs targeting the tdTomato locus as well as specific tumour suppressor genes associated with malignancy. These edited populations will then be able to be sorted for loss of

tdTomato expression, a proxy for successful editing at the tumour suppressor locus, and thus enriched prior to transplantation. This will increase the chances of developing a disease phenotype by selecting for edited cells at this early stage.

4.7.2 Assessment of the effect of TP53 mutations on the engraftment of edited HSCs

The optimised tdTomato targeting protocol could also be used to test whether, as speculated, *TP53* mutations could be selected for by the CRISPR editing approach. In the context of assessing the transcriptional and functional consequences of edited HSCs, the short-term RNA-Seq provided evidence that the *TP53* pathway was one of the major targeted pathways for upregulation in the context of genome editing. By targeting tdTomato in HSCs, with a background of *TP53* homozygous and heterozygous mutation, would elucidate whether these cells circumvent this transcriptional response and importantly whether this has a functional consequence on their engraftment and lineage output. Experiments to answer this question would involve targeting the tdTomato locus in a mouse strain that already also harbours a *TP53* mutation. Similar to the *in vivo* experiments described in this study, tdTomato HSCs on a background of wild type *TP53* would be compared to tdTomato HSCs on a background of mutant *TP53*. Assuming the editing efficiency would be the same in both conditions, lineage engraftment and transcriptomic profiles could be compared between the two. If the *TP53* mutant tdTomato HSCs showed a higher level of engraftment than wild type *TP53* tdTomato edited HSCs, it would suggest that *TP53* mutant HSCs are therefore more likely to survive editing and contribute a higher level to

the lineage output. Indeed, this is something that is currently being pursued in the lab. Comparison of the transcriptomic profiles between the tdTomato edited *TP53* wild type and *TP53* mutants would also provide evidence as to whether the enrichment for senescence associated genes is abrogated by *TP53* mutations. This would provide important information as to whether the altered transcriptional signature seen in long-term edited HSCs was driven by the short-term *TP53* response.

Further to this, secondary transplants of purified edited HSCs would provide key evidence as to whether the apparent increased transcriptional senescence/quiescence has an impact on further long-term lineage output or whether this is rescued. It would be interesting to evaluate whether this transcriptional state is representative of partial stem cell exhaustion and therefore sustained limited lineage output. In order to address this question, edited stem cells would be purified from primary recipient BM and transplanted into secondary recipients in equal numbers to the primary transplants. Lineage output would then be serially tracked to assess whether the reduced platelet and red blood cell output was sustained in secondary transplants.

4.7.3 Assessment of potential disease propagating populations

As briefly discussed in *3.3.7* it would be interesting to evaluate the two expanded populations in the *NF-1* CB targeted recipients as well as the remaining transplanted mice. By sorting the two populations and performing *NF-1* targeted Sanger sequencing, as well as whole genome sequencing, it would be possible to determine whether each population was enriched for mutated *NF-1* compared to total BM. This would provide evidence that indeed mutated *NF-1* was

implicated in the aberrant expansion. Whole genome sequencing would establish whether additional mutations were acquired in these expanded populations; it would be particularly interesting to see, if additional mutations were present, whether these matched those secondary mutations seen in JMML patients (Caye, Strullu et al. 2015) (Stieglitz, Taylor-Weiner et al. 2015). Further, by transplanting these sorted populations into secondary NSG recipients, it would be possible to determine whether they are disease propagating populations by their ability to cause disease development in a more aggressive manner compared to primary transplants.

4.7.4 Assessment of the importance of a fetal-like context in *NF-1* driven malignancy

In order to assess whether a fetal-like cell of origin is required for *NF-1* mutations to drive aberrant haematopoiesis in a xenograft model, similar approaches could be taken in adult BM CD34+s and compared to the disease development seen for *NF-1* targeted CB. These experiments would involve targeting the same *NF-1* exon 2 locus with equal efficiency in human CD34+ selected adult BM and CB followed by transplantation of equal numbers into NSG recipients. If only the *NF-1* edited CB led to disease development, it would provide additional evidence supporting the hypothesis that the fetal cell of origin is crucial for *NF-1* disease initiation in patients. It would also be interesting to compare CB to fetal liver in terms of whether they lead to different disease development.

4.7.5 Assessment of the importance of timing of acquisition of additional mutations in *NF-1* driven malignancy

In addition to analysing the importance of the cell of origin for *NF-1*, the RNP nucleofection approach could also be used to assess the impact of secondary mutations on disease progression. Experiments to address this question would include targeting *NF-1* alone, as described in this study, and analysing *NF-1* expanded populations post-transplant in order to establish whether additional mutations have been accumulated in the course of transplant. This would provide evidence as to whether *NF-1* provides a context which promotes the accumulation of additional mutations. In addition, experiments could be carried out to target multiple mutations simultaneously or consequentially post short-term engraftment in xenograft recipients and disease development could be analysed in an *in vivo* context. For example, *NF-1* could be introduced first, followed by a *ZRSR2* mutation, a known co-operating mutation in *NF-1* initiating JMML (Caye, Strullu et al. 2015). Disease progression could then be compared for simultaneous vs. consequential editing of *NF-1* and *ZRSR2* to see if disease progression is different. This is an important question because some JMML patients spontaneously resolve (Matsuda, Shimada et al. 2007) whereas others show a more aggressive disease course particularly with the presence of secondary mutations (Caye, Strullu et al. 2015). Due to the fact that when patients present, the secondary mutations are already present, it is difficult to understand the disease development from the initiating event. The advantage of the CB RNP editing approach is that these sequential mutations could be targeted in a controlled manner to allow careful assessment of the importance and timing of secondary mutations. In addition, given that secondary mutations

very often occur in the repressive epigenetic complex, PRC2, it would be interesting to edit CB with the *NF-1* initiating mutation, analyse the transcriptomic and chromatin profile by RNA-Seq and ATAC-Seq, followed by the same analysis after the introduction of a secondary mutation, for example *ZRSR2*, to assess how this might alter gene expression and promoter accessibility and thus cause disease progression.

4.8 Conclusions

Overall, this project has applied the CRISPR editing technology to optimise a 90-100% efficient approach to target and track edited mouse HSPCs. It has provided insights into the short-term transcriptional consequences of edited HSCs, as well as both long-term functional and transcriptional consequences. It has also provided a basis on which it is possible to readily assess editing efficiency of other targets by proxy as well as interrogate the potential selection of underlying genetic aberrations (for example *TP53*) by CRISPR editing. Finally, it has provided a proof of principle in targeting and transplanting *NF-1* edited CB cells to assess the abnormal haematopoietic consequences of inducing a JMML-associated initiating event in a human fetal-like context. Taken together, the work detailed in this thesis has allowed an in-depth study of the CRISPR editing approach as a tool to model both aspects of normal and malignant haematopoiesis.

REFERENCES

Addgene. (2021). "57818 EFS-Cas9 vector." from https://www.addgene.org/57818/.
Addgene. (2021). "57827 SFFV-Cas9 vector." from https://www.addgene.org/57827/.
Addgene. (2021). "Ai9 plasmid vector." from https://www.addgene.org/22799/.
Addgene. (2021). "PBSK vector." from https://www.addgene.org/vector-database/6263/.
Addgene. (2021). "pMD2.G." from https://www.addgene.org/12259/.
Addgene. (2021). "psPAX2 vector." from https://www.addgene.org/12260/.
Addgene. (2021). "PX458." from https://www.addgene.org/48138/.
Adli, M. (2018). "The CRISPR tool kit for genome editing and beyond." Nat Commun **9**(1): 1911.
Adolfsson, J., R. Mansson, N. Buza-Vidas, A. Hultquist, K. Liuba, C. T. Jensen, D. Bryder, L. Yang, O. J. Borge, L. A. Thoren, K. Anderson, E. Sitnicka, Y. Sasaki, M. Sigvardsson and S. E. Jacobsen (2005). "Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment." Cell **121**(2): 295-306.
Akashi, K., D. Traver, T. Miyamoto and I. L. Weissman (2000). "A clonogenic common myeloid progenitor that gives rise to all myeloid lineages." Nature **404**(6774): 193-197.
Ali, M. A. E., K. Fuse, Y. Tadokoro, T. Hoshii, M. Ueno, M. Kobayashi, N. Nomura, H. T. Vu, H. Peng, A. M. Hegazy, M. Masuko, H. Sone, F. Arai, A. Tajima and A. Hirao (2017). "Functional dissection of hematopoietic stem cell populations with a stemness-monitoring system based on NS-GFP transgene expression." Sci Rep **7**(1): 11442.
Asai, T., Y. Liu, N. Bae and S. D. Nimer (2011). "The p53 tumor suppressor protein regulates hematopoietic stem cell fate." J Cell Physiol **226**(9): 2215-2221.
Aubrey, B. J., G. L. Kelly, A. Janic, M. J. Herold and A. Strasser (2018). "How does p53 induce apoptosis and how does this relate to p53-mediated tumour suppression?" Cell Death Differ **25**(1): 104-113.
Bak, R. O., D. P. Dever, A. Reinisch, D. Cruz Hernandez, R. Majeti and M. H. Porteus (2017). "Multiplexed genetic engineering of human hematopoietic stem and progenitor cells using CRISPR/Cas9 and AAV6." Elife **6**.
Baldridge, M. T., K. Y. King, N. C. Boles, D. C. Weksberg and M. A. Goodell (2010). "Quiescent haematopoietic stem cells are activated by IFN-gamma in response to chronic infection." Nature **465**(7299): 793-797.
Barrangou, R., C. Fremaux, H. Deveau, M. Richards, P. Boyaval, S. Moineau, D. A. Romero and P. Horvath (2007). "CRISPR provides acquired resistance against viruses in prokaryotes." Science **315**(5819): 1709-1712.
Benchling. (2018). "tdTomato guide RNA design platform." from https://benchling.com/lucy_f/f/lib_eExzOKLF-example-project/seq_N5fhL54E-tdtomato-dna-1431-bp/edit.
Benchling. (2019). "Benchling [Biology Software]."
Benveniste, P., C. Frelin, S. Janmohamed, M. Barbara, R. Herrington, D. Hyam and N. N. Iscove (2010). "Intermediate-term hematopoietic stem cells with extended but time-limited reconstitution potential." Cell Stem Cell **6**(1): 48-58.
Benz, C., M. R. Copley, D. G. Kent, S. Wohrer, A. Cortes, N. Aghaeepour, E. Ma, H. Mader, K. Rowe, C. Day, D. Treloar, R. R. Brinkman and C. J. Eaves (2012). "Hematopoietic stem cell subtypes expand differentially during development and display distinct lymphopoietic programs." Cell Stem Cell **10**(3): 273-283.
Bhatia, M., J. C. Wang, U. Kapp, D. Bonnet and J. E. Dick (1997). "Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice." Proc Natl Acad Sci U S A **94**(10): 5320-5325.
Biechonski, S. and M. Michael (2013). "Differences between human and rodent DNA-damage response in hematopoietic stem cells: at the crossroads of self-renewal, aging and leukemogenesis." Translational Cancer Research **Vol 2, No.5**.

Boily, G., J. Larose, S. Langlois and D. Sinnett (2007). "Identification of transcripts modulated by ETV6 expression." Br J Haematol **136**(1): 48-62.
Boisset, J. C. and C. Robin (2012). "On the origin of hematopoietic stem cells: progress and controversy." Stem Cell Res **8**(1): 1-13.
Bowie, M. B., D. G. Kent, B. Dykstra, K. D. McKnight, L. McCaffrey, P. A. Hoodless and C. J. Eaves (2007). Identification of a new intrinsically timed developmental checkpoint that reprograms key hematopoietic stem cell properties. Proc Natl Acad Sci U S A. **104:** 5878-5882.
Brendel, C., B. Goebel, A. Daniela, M. Brugman, S. Kneissl, J. Schwäble, K. B. Kaufmann, U. Müller-Kuller, H. Kunkel, L. Chen-Wichmann, T. Abel, H. Serve, L. Bystrykh, C. J. Buchholz and M. Grez (2015). "CD133-targeted gene transfer into long-term repopulating hematopoietic stem cells." Mol Ther **23**(1): 63-70.
Bult, C. J., O. White, G. J. Olsen, L. Zhou, R. D. Fleischmann, G. G. Sutton, J. A. Blake, L. M. FitzGerald, R. A. Clayton, J. D. Gocayne, A. R. Kerlavage, B. A. Dougherty, J. F. Tomb, M. D. Adams, C. I. Reich, R. Overbeek, E. F. Kirkness, K. G. Weinstock, J. M. Merrick, A. Glodek, J. L. Scott, N. S. Geoghagen and J. C. Venter (1996). "Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii." Science **273**(5278): 1058-1073.
Cao, A. and R. Galanello (2010). "Beta-thalassemia." Genet Med **12**(2): 61-76.
Capone, S., K. M. Connor, A. Colombo, X. Li, T. J. Triche, Jr. and G. Ramsingh (2018). "Senescent human hematopoietic progenitors show elevated expression of transposable elements and inflammatory genes." Exp Hematol **62**: 33-38.e36.
Carrelha, J., Y. Meng, L. M. Kettyle, T. C. Luis, R. Norfo, V. Alcolea, H. Boukarabila, F. Grasso, A. Gambardella, A. Grover, K. Högstrand, A. M. Lord, A. Sanjuan-Pla, P. S. Woll, C. Nerlov and S. E. W. Jacobsen (2018). "Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells." Nature **554**(7690): 106-111.
Cavazzana-Calvo, M., S. Hacein-Bey, G. de Saint Basile, F. Gross, E. Yvon, P. Nusbaum, F. Selz, C. Hue, S. Certain, J. L. Casanova, P. Bousso, F. L. Deist and A. Fischer (2000). "Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease." Science **288**(5466): 669-672.
Caye, A., K. Rouault-Pierre, M. Strullu, E. Lainey, A. Abarrategi, O. Fenneteau, C. Arfeuille, J. Osman, B. Cassinat, S. Pereira, F. Anjos-Afonso, E. Currie, L. Ariza-McNaughton, V. Barlogis, J. H. Dalle, A. Baruchel, C. Chomienne, H. Cavé and D. Bonnet (2020). "Despite mutation acquisition in hematopoietic stem cells, JMML-propagating cells are not always restricted to this compartment." Leukemia **34**(6): 1658-1668.
Caye, A., M. Strullu, F. Guidez, B. Cassinat, S. Gazal, O. Fenneteau, E. Lainey, K. Nouri, S. Nakhaei-Rad, R. Dvorsky, J. Lachenaud, S. Pereira, J. Vivent, E. Verger, D. Vidaud, C. Galambrun, C. Picard, A. Petit, A. Contet, M. Poiree, N. Sirvent, F. Mechinaud, D. Adjaoud, C. Paillard, B. Nelken, Y. Reguerre, Y. Bertrand, D. Haussinger, J. H. Dalle, M. R. Ahmadian, A. Baruchel, C. Chomienne and H. Cave (2015). "Juvenile myelomonocytic leukemia displays mutations in components of the RAS pathway and the PRC2 network." Nat Genet **47**(11): 1334-1340.
Caye, A., M. Strullu, F. Guidez, B. Cassinat, S. Gazal, O. Fenneteau, E. Lainey, K. Nouri, S. Nakhaei-Rad, R. Dvorsky, J. Lachenaud, S. Pereira, J. Vivent, E. Verger, D. Vidaud, C. Galambrun, C. Picard, A. Petit, A. Contet, M. Poirée, N. Sirvent, F. Méchinaud, D. Adjaoud, C. Paillard, B. Nelken, Y. Reguerre, Y. Bertrand, D. Häussinger, J. H. Dalle, M. R. Ahmadian, A. Baruchel, C. Chomienne and H. Cavé (2015). "Juvenile myelomonocytic leukemia displays mutations in components of the RAS pathway and the PRC2 network." Nat Genet **47**(11): 1334-1340.
Cazzola, A., G. Cazzaniga, A. Biondi, R. Meneveri, S. Brunelli and E. Azzoni (2020). "Prenatal Origin of Pediatric Leukemia: Lessons From Hematopoietic Development." Front Cell Dev Biol **8**: 618164.
Challen, G. A., N. Boles, K. K. Lin and M. A. Goodell (2009). "Mouse Hematopoietic Stem Cell Identification And Analysis." Cytometry A **75**(1): 14-24.

Chan, I. T., J. L. Kutok, I. R. Williams, S. Cohen, L. Kelly, H. Shigematsu, L. Johnson, K. Akashi, D. A. Tuveson, T. Jacks and D. G. Gilliland (2004). "Conditional expression of oncogenic K-ras from its endogenous promoter induces a myeloproliferative disease." J Clin Invest **113**(4): 528-538.
Chan, R. J., M. B. Leedy, V. Munugalavadla, C. S. Voorhorst, Y. Li, M. Yu and R. Kapur (2005). "Human somatic PTPN11 mutations induce hematopoietic-cell hypersensitivity to granulocyte-macrophage colony-stimulating factor." Blood **105**(9): 3737-3742.
Chapman, K., M. Holmes and J. Seckl (2013). "11β-hydroxysteroid dehydrogenases: intracellular gate-keepers of tissue glucocorticoid action." Physiol Rev **93**(3): 1139-1206.
Check, E. (2002). "A tragic setback." Nature **420**(6912): 116-118.
Chen, L., M. Kostadima, J. H. A. Martens, G. Canu, S. P. Garcia, E. Turro, K. Downes, I. C. Macaulay, E. Bielczyk-Maczynska, S. Coe, S. Farrow, P. Poudel, F. Burden, S. B. G. Jansen, W. J. Astle, A. Attwood, T. Bariana, B. de Bono, A. Breschi, J. C. Chambers, B. Consortium, F. A. Choudry, L. Clarke, P. Coupland, M. van der Ent, W. N. Erber, J. H. Jansen, R. Favier, M. E. Fenech, N. Foad, K. Freson, C. van Geet, K. Gomez, R. Guigo, D. Hampshire, A. M. Kelly, H. H. D. Kerstens, J. S. Kooner, M. Laffan, C. Lentaigne, C. Labalette, T. Martin, S. Meacham, A. Mumford, S. Nurnberg, E. Palumbo, B. A. van der Reijden, D. Richardson, S. J. Sammut, G. Slodkowicz, A. U. Tamuri, L. Vasquez, K. Voss, S. Watt, S. Westbury, P. Flicek, R. Loos, N. Goldman, P. Bertone, R. J. Read, S. Richardson, A. Cvejic, N. Soranzo, W. H. Ouwehand, H. G. Stunnenberg, M. Frontini and A. Rendon (2014). "Transcriptional diversity during lineage commitment of human blood progenitors." Science **345**(6204): 1251033.
Chen, X. and M. A. Gonçalves (2016). "Engineered Viruses as Genome Editing Devices." Mol Ther **24**(3): 447-457.
Civin, C. I., L. C. Strauss, C. Brovall, M. J. Fackler, J. F. Schwartz and J. H. Shaper (1984). "Antigenic analysis of hematopoiesis. III. A hematopoietic progenitor cell surface antigen defined by a monoclonal antibody raised against KG-1a cells." J Immunol **133**(1): 157-165.
Cohen, P. R. (2021). "Chronic Lymphocytic Leukemia in Neurofibromatosis Type 1 Patients: Case Report and Literature Review of a Rare Occurrence." Cureus **13**(4): e14258.
Conti, A. and R. D. Micco (2018). "p53 activation: a checkpoint for precision genome editing?" Genome Medicine **10**(1): 1-4.
Cooper, L. J., K. M. Shannon, M. R. Loken, M. Weaver, K. Stephens and E. L. Sievers (2000). "Evidence that juvenile myelomonocytic leukemia can arise from a pluripotential stem cell." Blood **96**(6): 2310-2313.
Copley, M. R., S. Babovic, C. Benz, D. J. H. F. Knapp, P. A. Beer, D. G. Kent, S. Wohrer, D. Q. Treloar, C. Day, K. Rowe, H. Mader, F. Kuchenbauer, R. K. Humphries and C. J. Eaves (2013). "The Lin28b–let-7–Hmga2 axis determines the higher self-renewal potential of fetal haematopoietic stem cells." Nature Cell Biology **15**: 916-925.
Cradick, T. J., P. Qiu, C. M. Lee, E. J. Fine and G. Bao (2014). "COSMID: A Web-based Tool for Identifying and Validating CRISPR/Cas Off-target Sites." Mol Ther Nucleic Acids **3**(12): e214-.
Cutts, B. A., A. K. Sjogren, K. M. Andersson, A. M. Wahlstrom, C. Karlsson, B. Swolin and M. O. Bergo (2009). "Nf1 deficiency cooperates with oncogenic K-RAS to induce acute myeloid leukemia in mice." Blood **114**(17): 3629-3632.
Dahlin, J. S., F. K. Hamey, B. Pijuan-Sala, M. Shepherd, W. W. Y. Lau, S. Nestorowa, C. Weinreb, S. Wolock, R. Hannah, E. Diamanti, D. G. Kent, B. Gottgens and N. K. Wilson (2018). "A single-cell hematopoietic landscape resolves 8 lineage trajectories and defects in Kit mutant mice." Blood **131**(21): e1-e11.
de Vries, A. C., C. M. Zwaan and M. M. van den Heuvel-Eibrink (2010). Molecular basis of juvenile myelomonocytic leukemia. Haematologica. **95:** 179-182.
Dever, D. P., R. O. Bak, A. Reinisch, J. Camarena, G. Washington, C. E. Nicolas, M. Pavel-Dinu, N. Saxena, A. B. Wilkens, S. Mantri, N. Uchida, A. Hendel, A. Narla, R. Majeti, K. I. Weinberg and M. H. Porteus (2016). "CRISPR/Cas9 β-globin gene targeting in human haematopoietic stem cells." Nature **539**: 384-389.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson and T. R. Gingeras (2013). "STAR: ultrafast universal RNA-seq aligner." Bioinformatics **29**(1): 15-21.
Donovan, S., W. See, J. Bonifas, D. Stokoe and K. M. Shannon (2002). "Hyperactivation of protein kinase B and ERK have discrete effects on survival, proliferation, and cytokine expression in Nf1-deficient myeloid cells." Cancer Cell **2**(6): 507-514.
Dow, L. E., J. Fisher, K. P. O'Rourke, A. Muley, E. R. Kastenhuber, G. Livshits, D. F. Tschaharganeh, N. D. Socci and S. W. Lowe (2015). "Inducible in vivo genome editing with CRISPR-Cas9." Nat Biotechnol **33**(4): 390-394.
Dykstra, B., D. Kent, M. Bowie, L. McCaffrey, M. Hamilton, K. Lyons, S. J. Lee, R. Brinkman and C. Eaves (2007). "Long-term propagation of distinct hematopoietic differentiation programs in vivo." Cell Stem Cell **1**(2): 218-229.
Dzeja, P. and A. Terzic (2009). "Adenylate kinase and AMP signaling networks: metabolic monitoring, signal communication and body energy sensing." Int J Mol Sci **10**(4): 1729-1772.
Eisenbarth, I., K. Beyer, W. Krone and G. Assum (2000). "Toward a survey of somatic mutation of the NF1 gene in benign neurofibromas of patients with neurofibromatosis type 1." Am J Hum Genet **66**(2): 393-401.
Ema, H., Y. Morita and T. Suda (2014). "Heterogeneity and hierarchy of hematopoietic stem cells." Exp Hematol **42**(2): 74-82.e72.
Emanuel, P. D., L. J. Bates, R. P. Castleberry, R. J. Gualtieri and K. S. Zuckerman (1991). "Selective hypersensitivity to granulocyte-macrophage colony-stimulating factor by juvenile chronic myeloid leukemia hematopoietic progenitors." Blood **77**(5): 925-929.
Enache, O. M., V. Rendo, M. Abdusamad, D. Lam, D. Davison, S. Pal, N. Currimjee, J. Hess, S. Pantel, A. Nag, A. R. Thorner, J. G. Doench, F. Vazquez, R. Beroukhim, T. R. Golub and U. Ben-David (2020). "Cas9 activates the p53 pathway and selects for p53-inactivating mutations." Nat Genet **52**(7): 662-668.
Ferrari, S., A. Jacob, S. Beretta, G. Unali, L. Albano, V. Vavassori, D. Cittaro, D. Lazarevic, C. Brombin, F. Cugnata, A. Kajaste-Rudnitski, I. Merelli, P. Genovese and L. Naldini (2020). "Efficient gene editing of human long-term hematopoietic stem cells validated by clonal tracking." Nat Biotechnol.
Flaherty, S. and J. M. Reynolds (2015). "Mouse Naïve CD4+ T Cell Isolation and In vitro Differentiation into T Cell Subsets." J Vis Exp(98).
Flanagan, M., J. M. Gimble, G. Yu, X. Wu, X. Xia, J. Hu, S. Yao and S. Li (2011). "Competitive electroporation formulation for cell therapy." Cancer Gene Ther **18**(8): 579-586.
Flemming, A. (2015). "Cancer stem cells: Targeting the root of cancer relapse." Nat Rev Drug Discov **14**(3): 165.
Flotho, C., D. Steinemann, C. G. Mullighan, G. Neale, K. Mayer, C. P. Kratz, B. Schlegelberger, J. R. Downing and C. M. Niemeyer (2007). "Genome-wide single-nucleotide polymorphism analysis in juvenile myelomonocytic leukemia identifies uniparental disomy surrounding the NF1 locus in cases associated with neurofibromatosis but not in cases with mutant RAS or PTPN11." Oncogene **26**(39): 5816-5821.
Frangoul, H., D. Altshuler, M. D. Cappellini, Y. S. Chen, J. Domm, B. K. Eustace, J. Foell, J. de la Fuente, S. Grupp, R. Handgretinger, T. W. Ho, A. Kattamis, A. Kernytsky, J. Lekstrom-Himes, A. M. Li, F. Locatelli, M. Y. Mapara, M. de Montalembert, D. Rondelli, A. Sharma, S. Sheth, S. Soni, M. H. Steinberg, D. Wall, A. Yen and S. Corbacioglu (2021). "CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia." N Engl J Med **384**(3): 252-260.
Fu, Y., J. D. Sander, D. Reyon, V. M. Cascio and J. K. Joung (2014). "Improving CRISPR-Cas nuclease specificity using truncated guide RNAs." Nat Biotechnol **32**(3): 279-284.
Fujimaki, K., R. Li, H. Chen, K. Della Croce, H. H. Zhang, J. Xing, F. Bai and G. Yao (2019). "Graded regulation of cellular quiescence depth between proliferation and senescence by a lysosomal dimmer switch." Proc Natl Acad Sci U S A **116**(45): 22624-22634.

Furusawa, A., K. Sadashivaiah, Z. N. Singh, C. I. Civin and A. Banerjee (2016). "Inefficient megakaryopoiesis in mouse hematopoietic stem-progenitor cells lacking T-bet." Exp Hematol **44**(3): 194-206.e117.
Galbiati, M., A. Lettieri, C. Micalizzi, S. Songia, C. Morerio, A. Biondi, C. Dufour and G. Cazzaniga (2013). Natural history of acute lymphoblastic leukemia in neurofibromatosis type 1 monozygotic twins. Leukemia. England. **27:** 1778-1781.
Galán-Díez, M., Á. Cuesta-Domínguez and S. Kousteni (2018). "The Bone Marrow Microenvironment in Health and Myeloid Malignancy." Cold Spring Harb Perspect Med **8**(7).
Garneau, J. E., M. E. Dupuis, M. Villion, D. A. Romero, R. Barrangou, P. Boyaval, C. Fremaux, P. Horvath, A. H. Magadan and S. Moineau (2010). "The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA." Nature **468**(7320): 67-71.
Gasiunas, G., R. Barrangou, P. Horvath and V. Siksnys (2012). "Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria." Proc Natl Acad Sci U S A **109**(39): E2579-2586.
Gazit, R., P. K. Mandal, W. Ebina, A. Ben-Zvi, C. Nombela-Arrieta, L. E. Silberstein and D. J. Rossi (2014). "Fgd5 identifies hematopoietic stem cells in the murine bone marrow." J Exp Med **211**(7): 1315-1331.
Gerber, H. P., A. K. Malik, G. P. Solar, D. Sherman, X. H. Liang, G. Meng, K. Hong, J. C. Marsters and N. Ferrara (2002). "VEGF regulates haematopoietic stem cell survival by an internal autocrine loop mechanism." Nature **417**(6892): 954-958.
Gravitz, L. and S. Pincock (2014). "Sickle-cell disease." Nature **515**(7526): S1.
Grover, A., A. Sanjuan-Pla, S. Thongjuea, J. Carrelha, A. Giustacchini, A. Gambardella, I. Macaulay, E. Mancini, T. C. Luis, A. Mead, S. E. W. Jacobsen and C. Nerlov (2016). "Single-cell RNA sequencing reveals molecular and functional platelet bias of aged haematopoietic stem cells." Nature Communications, Published online: 24 March 2016; | doi:10.1038/ncomms11075.
Gul-Uludag, H., P. Xu, L. A. Marquez-Curtis, J. Xing, A. Janowska-Wieczorek and J. Chen (2012). "Cationic Liposome-Mediated CXCR4 Gene Delivery into Hematopoietic Stem/Progenitor Cells: Implications for Clinical Transplantation and Gene Therapy." Stem Cells Dev **21**(10): 1587-1596.
Gundry, M. C., L. Brunetti, A. Lin, A. E. Mayle, A. Kitano, D. Wagner, J. I. Hsu, K. A. Hoegenauer, C. Rooney, M. A. Goodell and D. Nakada (2016). "Highly efficient genome editing of murine and human hematopoietic progenitor cells by CRISPR/Cas9." Cell Rep **17**(5): 1453-1461.
Gupta, D., H. P. Shah, K. Malu, N. Berliner and P. Gaines (2014). "Differentiation and characterization of myeloid cells." Curr Protoc Immunol **104**: 22f.25.21-22f.25.28.
Haapaniemi, E., S. Botla, J. Persson, B. Schmierer and J. Taipale (2018). "CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response." Nature Medicine **24**(7): 927-930.
Haeussler, M., K. Schönig, H. Eckert, A. Eschstruth, J. Mianné, J. B. Renaud, S. Schneider-Maunoury, A. Shkumatava, L. Teboul, J. Kent, J. S. Joly and J. P. Concordet (2016). "Evaluation of off-target and on-target scoring algorithms and integration into the guide RNA selection tool CRISPOR." Genome Biol **17**(1): 148.
Hale, C. R., P. Zhao, S. Olson, M. O. Duff, B. R. Graveley, L. Wells, R. M. Terns and M. P. Terns (2009). "RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex." Cell **139**(5): 945-956.
He, S., I. Kim, M. S. Lim and S. J. Morrison (2011). "Sox17 expression confers self-renewal potential and fetal stem cell characteristics upon adult hematopoietic progenitors." Genes Dev **25**(15): 1613-1627.
Heckl, D., M. S. Kowalczyk, D. Yudovich, R. Belizaire, R. V. Puram, M. E. McConkey, A. Thielke, J. C. Aster, A. Regev and B. L. Ebert (2014). "Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing." Nature Biotechnology **32**: 941-946.

Helsmoortel, H. H., S. Bresolin, T. Lammens, H. Cavé, P. Noellke, A. Caye, F. Ghazavi, A. d. Vries, H. Hasle, V. Labarque, R. Masetti, J. Stary, M. M. v. d. Heuvel-Eibrink, J. Philippé, N. V. Roy, Y. Benoit, F. Speleman, C. Niemeyer, C. Flotho, G. Basso, G. t. Kronnie, P. V. Vlierberghe and B. D. Moerloose (2016). "LIN28B overexpression defines a novel fetal-like subgroup of juvenile myelomonocytic leukemia."
Helsmoortel, H. H., B. De Moerloose, T. Pieters, F. Ghazavi, S. Bresolin, H. Cavé, A. de Vries, V. de Haas, C. Flotho, V. Labarque, C. Niemeyer, P. De Paepe, N. Van Roy, J. Stary, M. M. van den Heuvel-Eibrink, Y. Benoit, J. Schulte, S. Goossens, G. Berx, J. J. Haigh, F. Speleman, P. Van Vlierberghe and T. Lammens (2016). LIN28B is over-expressed in specific subtypes of pediatric leukemia and regulates lncRNA H19. Haematologica. **101:** e240-244.
Hendel, A., R. O. Bak, J. T. Clark, A. B. Kennedy, D. E. Ryan, S. Roy, I. Steinfeld, B. D. Lunstad, R. J. Kaiser, A. B. Wilkens, R. Bacchetta, A. Tsalenko, D. Dellinger, L. Bruhn and M. H. Porteus (2015). "Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells." Nature Biotechnology **33**(9): 985.
Heyer, W. D., K. T. Ehmsen and J. Liu (2010). "Regulation of homologous recombination in eukaryotes." Annu Rev Genet **44**: 113-139.
Hille, F., H. Richter, S. P. Wong, M. Bratovič, S. Ressel and E. Charpentier (2018). "The Biology of CRISPR-Cas: Backward and Forward." Cell **172**(6): 1239-1259.
Hoffmann-Fezer, G., J. Mysliwietz, W. Mortlbauer, H. J. Zeitler, E. Eberle, U. Honle and S. Thierfelder (1993). "Biotin labeling as an alternative nonradioactive approach to determination of red cell survival." Ann Hematol **67**(2): 81-87.
http://www.lonzabio.com, L. C. A.-. (2017). "Cell & Transfection Database A database containing data on cells as well as experimental data on transfections performed with Lonza technologies. Contains data from Lonza as well as customer experiments. - Lonza AG." from http://bio.lonza.com/6.html?tx_lzcelldatabase_pi_view%5Bid%5D=745&logintype=login&tx_lzcelldatabase_pi_view[id]=204.
Hu, J. H., S. M. Miller, M. H. Geurts, W. Tang, L. Chen, N. Sun, C. M. Zeina, X. Gao, H. A. Rees, Z. Lin and D. R. Liu (2018). "Evolved Cas9 variants with broad PAM compatibility and high DNA specificity." Nature **556**(7699): 57-63.
Höfer, T. and H. R. Rodewald (2016). "Output without input: the lifelong productivity of hematopoietic stem cells." Curr Opin Cell Biol **43**: 69-77.
Ihry, R. J., K. A. Worringer, M. R. Salick, E. Frias, D. Ho, K. Theriault, S. Kommineni, J. Chen, M. Sondey, C. Ye, R. Randhawa, T. Kulkarni, Z. Yang, G. McAllister, C. Russ, J. Reece-Hoyes, W. Forrester, G. R. Hoffman, R. Dolmetsch and A. Kaykas (2018). "p53 inhibits CRISPR–Cas9 engineering in human pluripotent stem cells." Nature Medicine **24**(7): 939-946.
Iscove, N. N. (2010). "Stability of Self-Renewal in Hematopoietic Stem Cells." Blood **116**(21): SCI-42.
Ishino, Y., M. Krupovic and P. Forterre (2018). "History of CRISPR-Cas from Encounter with a Mysterious Repeated Sequence to Genome Editing Technology." J Bacteriol **200**(7).
Ishino, Y., H. Shinagawa, K. Makino, M. Amemura and A. Nakata (1987). "Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product." J Bacteriol **169**(12): 5429-5433.
Ivanova, N. B., J. T. Dimos, C. Schaniel, J. A. Hackney, K. A. Moore and I. R. Lemischka (2002). "A stem cell molecular signature." Science **298**(5593): 601-604.
Ivanovs, A., S. Rybtsov, E. S. Ng, E. G. Stanley, A. G. Elefanty and A. Medvinsky (2017). "Human haematopoietic stem cell development: from the embryo to the dish." Development **144**(13): 2323-2337.
JacksonLaboratory, T. (2021). "Ai9." from https://www.jax.org/strain/007909.
JacksonLaboratory, T. (2021). "C57BL/6J." from https://www.jax.org/strain/000664.
JacksonLaboratory, T. (2021). "CMV-Cre." from https://www.jax.org/strain/006054.
JacksonLaboratory, T. (2021). "NSG." from https://www.jax.org/strain/005557.

Jafri, S., S. D. Moore, N. W. Morrell and M. L. Ormiston (2017). "A sex-specific reconstitution bias in the competitive CD45.1/CD45.2 congenic bone marrow transplant model." Sci Rep **7**(1): 3495.
Jansen, R., J. D. Embden, W. Gaastra and L. M. Schouls (2002). "Identification of genes that are associated with DNA repeats in prokaryotes." Mol Microbiol **43**(6): 1565-1575.
Jensen, K. T., L. Fløe, T. S. Petersen, J. Huang, F. Xu, L. Bolund, Y. Luo and L. Lin (2017). "Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency." FEBS Lett **591**(13): 1892-1901.
Jensen, T. I., E. Axelgaard and R. O. Bak (2019). "Therapeutic gene editing in haematological disorders with CRISPR/Cas9." Br J Haematol **185**(5): 821-835.
Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna and E. Charpentier (2012). "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." Science **337**(6096): 816-821.
Johnson, W. E., C. Li and A. Rabinovic (2006). "Adjusting batch effects in microarray expression data using empirical Bayes methods." Biostatistics **8**(1): 118-127.
Kaina, B. (2003). "DNA damage-triggered apoptosis: critical role of DNA repair, double-strand breaks, cell proliferation and signaling." Biochem Pharmacol **66**(8): 1547-1554.
Kaufmann, K. B., A. G. X. Zeng, E. Coyaud, L. Garcia-Prat, E. Papalexi, A. Murison, E. M. N. Laurent, M. Chan-Seng-Yue, O. I. Gan, K. Pan, J. McLeod, H. Boutzen, S. Zandi, S. I. Takayanagi, R. Satija, B. Raught, S. Z. Xie and J. E. Dick (2021). "A latent subset of human hematopoietic stem cells resists regenerative stress to preserve stemness." Nat Immunol.
Keckesova, Z., J. L. Donaher, J. De Cock, E. Freinkman, S. Lingrell, D. A. Bachovchin, B. Bierie, V. Tischler, A. Noske, M. C. Okondo, F. Reinhardt, P. Thiru, T. R. Golub, J. E. Vance and R. A. Weinberg (2017). "LACTB is a tumour suppressor that modulates lipid metabolism and cell state." Nature **543**(7647): 681-686.
Khosravi, M. A., M. Abbasalipour, J. P. Concordet, J. V. Berg, S. Zeinali, A. Arashkia, K. Azadmanesh, T. Buch and M. Karimipoor (2019). "Targeted deletion of BCL11A gene by CRISPR-Cas9 system for fetal hemoglobin reactivation: A promising approach for gene therapy of beta thalassemia disease." Eur J Pharmacol **854**: 398-405.
Kiel, M. J., O. H. Yilmaz, T. Iwashita, C. Terhorst and S. J. Morrison (2005). "SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells." Cell **121**(7): 1109-1121.
Kim, I., T. L. Saunders and S. J. Morrison (2007). "Sox17 dependence distinguishes the transcriptional regulation of fetal from adult hematopoietic stem cells." Cell **130**(3): 470-483.
Kim, S., D. Kim, S. W. Cho, J. Kim and J. S. Kim (2014). "Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins." Genome Res **24**(6): 1012-1019.
Kim, S. J., J. H. Seo, S. W. Lee, E. Han, E. S. Lee, S. H. Cha and B. K. Seo (2003). "A case of non-Hodgkin's lymphoma in a patient with neurofibromatosis type 1." Korean J Intern Med **18**(3): 202-205.
Kimura, A., C. Martin, G. W. Robinson, J. M. Simone, W. Chen, M. C. Wickre, J. J. O'Shea and L. Hennighausen (2010). "The gene encoding the hematopoietic stem cell regulator CCN3/NOV is under direct cytokine control through the transcription factors STAT5A/B." J Biol Chem **285**(43): 32704-32709.
Kitamura, T., T. Tange, T. Terasawa, S. Chiba, T. Kuwaki, K. Miyagawa, Y. F. Piao, K. Miyazono, A. Urabe and F. Takaku (1989). "Establishment and characterization of a unique human cell line that proliferates dependently on GM-CSF, IL-3, or erythropoietin." J Cell Physiol **140**(2): 323-334.
Kleinstiver, B. P., V. Pattanayak, M. S. Prew, S. Q. Tsai, N. T. Nguyen, Z. Zheng and J. K. Joung (2016). "High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects." Nature **529**(7587): 490-495.

Kondo, M., I. L. Weissman and K. Akashi (1997). "Identification of clonogenic common lymphoid progenitors in mouse bone marrow." Cell **91**(5): 661-672.
Kotecha, N., N. J. Flores, J. M. Irish, E. Simonds, D. S. Sakai, S. Archambeault, E. Diaz-Flores, M. Coram, K. M. Shannon, G. P. Nolan and M. L. Loh (2008). "Single Cell Profiling Identifies Aberrant STAT5 Activation in Myeloid Malignancies with Specific Clinical and Biologic Correlates." Cancer Cell **14**(4): 335-343.
Kotini, A. G., C. J. Chang, A. Chow, H. Yuan, T. C. Ho, T. Wang, S. Vora, A. Solovyov, C. Husser, M. Olszewska, J. Teruya-Feldstein, D. Perumal, V. M. Klimek, A. Spyridonidis, R. K. Rampal, L. Silverman, E. P. Reddy, E. Papaemmanuil, S. Parekh, B. D. Greenbaum, C. S. Leslie, M. G. Kharas and E. P. Papapetrou (2017). "Stage-Specific Human Induced Pluripotent Stem Cells Map the Progression of Myeloid Transformation to Transplantable Leukemia." Cell Stem Cell **20**(3): 315-328.e317.
Kratz, C. P., C. M. Niemeyer, R. P. Castleberry, M. Cetin, E. Bergsträsser, P. D. Emanuel, H. Hasle, G. Kardos, C. Klein, S. Kojima, J. Stary, M. Trebo, M. Zecca, B. D. Gelb, M. Tartaglia and M. L. Loh (2005). "The mutational spectrum of PTPN11 in juvenile myelomonocytic leukemia and Noonan syndrome/myeloproliferative disease." Blood **106**(6): 2183-2185.
Krombholz, C. F., K. Aumann, M. Kollek, D. Bertele, S. Fluhr, M. Kunze, C. M. Niemeyer, C. Flotho and M. Erlacher (2016). "Long-term serial xenotransplantation of juvenile myelomonocytic leukemia recapitulates human disease in Rag2-/-γc-/- mice." Haematologica **101**(5): 597-606.
Krombholz, C. F., K. Aumann, M. Kollek, D. Bertele, S. Fluhr, M. Kunze, C. M. Niemeyer, C. Flotho and M. Erlacher (2016). Long-term serial xenotransplantation of juvenile myelomonocytic leukemia recapitulates human disease in Rag2−/−γc−/− mice. Haematologica. **101:** 597-606.
Lapidot, T., C. Sirard, J. Vormoor, B. Murdoch, T. Hoang, J. Caceres-Cortes, M. Minden, B. Paterson, M. A. Caligiuri and J. E. Dick (1994). "A cell initiating human acute myeloid leukaemia after transplantation into SCID mice." Nature **367**(6464): 645-648.
Laurenti, E. and B. Göttgens (2018). "From haematopoietic stem cells to complex differentiation landscapes." Nature **553**(7689): 418-426.
Le, D. T., N. Kong, Y. Zhu, J. O. Lauchle, A. Aiyigari, B. S. Braun, E. Wang, S. C. Kogan, M. M. Le Beau, L. Parada and K. M. Shannon (2004). "Somatic inactivation of Nf1 in hematopoietic cells results in a progressive myeloproliferative disorder." Blood **103**(11): 4243-4250.
Lebois, M. and E. C. Josefsson (2016). "Regulation of platelet lifespan by apoptosis." Platelets **27**(6): 497-504.
Lehrman, S. (1999). "Virus treatment questioned after gene therapy death." Nature **401**(6753): 517-518.
Lewis, M. D., E. de Leenheer, S. Fishman, L. K. Siew, G. Gross and F. S. Wong (2015). "A reproducible method for the expansion of mouse CD8+ T lymphocytes." J Immunol Methods **417**: 134-138.
Li, H. and R. Durbin (2009). "Fast and accurate short read alignment with Burrows–Wheeler transform." Bioinformatics **25**(14): 1754-1760.
Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis and R. Durbin (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.
Li, Y., H. A. Rogoff, S. Keates, Y. Gao, S. Murikipudi, K. Mikule, D. Leggett, W. Li, A. B. Pardee and C. J. Li (2015). "Suppression of cancer relapse and metastasis by inhibiting cancer stemness." Proc Natl Acad Sci U S A **112**(6): 1839-1844.
Liao, X. Y., K. Y. Qiu, J. P. Fang, R. H. Wu, S. Y. Guo, K. Huang and D. H. Zhou (2019). "Diagnosis and treatment of juvenile myelomonocytic leukemia." Hematology **24**(1): 577-582.
Liao, Y., G. K. Smyth and W. Shi (2014). "featureCounts: an efficient general purpose program for assigning sequence reads to genomic features." Bioinformatics **30**(7): 923-930.

Liberzon, A., C. Birger, H. Thorvaldsdóttir, M. Ghandi, J. P. Mesirov and P. Tamayo (2015). "The Molecular Signatures Database (MSigDB) hallmark gene set collection." Cell Syst **1**(6): 417-425.
Lim, J. F., H. Berger and I. H. Su (2016). "Isolation and Activation of Murine Lymphocytes." J Vis Exp(116).
Lin, S., B. T. Staahl, R. K. Alla and J. A. Doudna (2014). "Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery." Elife **3**: e04766.
Lino, C. A., J. C. Harper, J. P. Carney and J. A. Timlin (2018). "Delivering CRISPR: a review of the challenges and approaches." Drug Deliv **25**(1): 1234-1257.
Liu, L., M. Zhao, X. Jin, G. Ney, K. B. Yang, F. Peng, J. Cao, T. Iwawaki, J. Del Valle, X. Chen and Q. Li (2019). "Adaptive endoplasmic reticulum stress signalling via IRE1α-XBP1 preserves self-renewal of haematopoietic and pre-leukaemic stem cells." Nat Cell Biol **21**(3): 328-337.
Liu, Y. Z. a. S. G. a. J. X. a. F. (2018). "Hematopoietic Hierarchy – An Updated Roadmap." Trends in Cell Biology **28**(12): 976-986.
Loh, M. L. (2011). "Recent advances in the pathogenesis and treatment of juvenile myelomonocytic leukaemia." Br J Haematol **152**(6): 677-687.
Lonza. (2021). "Ba/F3 optimised nucleofection protocol." from https://knowledge.lonza.com/cell?id=693&search=baf3.
Lonza. (2021). "https://bioscience.lonza.com/lonza_bs/GB/en/nucleofector-technology."
Louka, E., B. Povinelli, A. Rodriguez-Meira, G. Buck, W. X. Wen, G. Wang, N. Sousos, N. Ashley, A. Hamblin, C. A. G. Booth, A. Roy, N. Elliott, D. Iskander, J. de la Fuente, N. Fordham, S. O'Byrne, S. Inglott, R. Norfo, M. Salio, S. Thongjuea, A. Rao, I. Roberts and A. J. Mead (2021). "Heterogeneous disease-propagating stem cells in juvenile myelomonocytic leukemia." J Exp Med **218**(2).
Lu, R., A. Czechowicz, J. Seita, D. Jiang and I. L. Weissman (2019). "Clonal-level lineage commitment pathways of hematopoietic stem cells in vivo." Proc Natl Acad Sci U S A **116**(4): 1447-1456.
Luo, Z., M. Gao, N. Huang, X. Wang, Z. Yang, H. Yang, Z. Huang and W. Feng (2019). "Efficient disruption of bcr-abl gene by CRISPR RNA-guided FokI nucleases depresses the oncogenesis of chronic myeloid leukemia cells." J Exp Clin Cancer Res **38**.
Madisen, L., T. A. Zwingman, S. M. Sunkin, S. W. Oh, H. A. Zariwala, H. Gu, L. L. Ng, R. D. Palmiter, M. J. Hawrylycz, A. R. Jones, E. S. Lein and H. Zeng (2010). "A robust and high-throughput Cre reporting and characterization system for the whole mouse brain." Nat Neurosci **13**(1): 133-140.
Makarova, K. S., D. H. Haft, R. Barrangou, S. J. Brouns, E. Charpentier, P. Horvath, S. Moineau, F. J. Mojica, Y. I. Wolf, A. F. Yakunin, J. van der Oost and E. V. Koonin (2011). "Evolution and classification of the CRISPR-Cas systems." Nat Rev Microbiol **9**(6): 467-477.
Mandal, P. K., L. M. Ferreira, R. Collins, T. B. Meissner, C. L. Boutwell, M. Friesen, V. Vrbanac, B. S. Garrison, A. Stortchevoi, D. Bryder, K. Musunuru, H. Brand, A. M. Tager, T. M. Allen, M. E. Talkowski, D. J. Rossi and C. A. Cowan (2014). "Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9." Cell Stem Cell **15**(5): 643-652.
Mansson, R., A. Hultquist, S. Luc, L. Yang, K. Anderson, S. Kharazi, S. Al-Hashmi, K. Liuba, L. Thoren, J. Adolfsson, N. Buza-Vidas, H. Qian, S. Soneji, T. Enver, M. Sigvardsson and S. E. Jacobsen (2007). "Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors." Immunity **26**(4): 407-419.
Marraffini, L. A. and E. J. Sontheimer (2008). "CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA." Science **322**(5909): 1843-1845.
Martin, M. (2011). "Cutadapt removes adapter sequences from high-throughput sequencing reads." 2011 **17**(1): 3.
Matsuda, K., A. Shimada, N. Yoshida, A. Ogawa, A. Watanabe, S. Yajima, S. Iizuka, K. Koike, F. Yanai, K. Kawasaki, M. Yanagimachi, A. Kikuchi, Y. Ohtsuka, E. Hidaka, K. Yamauchi, M. Tanaka, R. Yanagisawa, Y. Nakazawa, M. Shiohara, A. Manabe and S. Kojima (2007). "Spontaneous

improvement of hematologic abnormalities in patients having juvenile myelomonocytic leukemia with specific RAS mutations." Blood **109**(12): 5477-5480.
Maurisse, R., D. De Semir, H. Emamekhoo, B. Bedayat, A. Abdolmohammadi, H. Parsi and D. C. Gruenert (2010). "Comparative transfection of DNA into primary and transformed mammalian cells from different lineages." BMC Biotechnol **10**: 9.
Mayani, H., W. Dragowska and P. M. Lansdorp (1993). "Characterization of functionally distinct subpopulations of CD34+ cord blood cells in serum-free long-term cultures supplemented with hematopoietic cytokines." Blood **82**(9): 2664-2672.
McGee, R. B. and K. E. Nichols (2016). "Introduction to cancer genetic susceptibility syndromes." Hematology Am Soc Hematol Educ Program **2016**(1): 293-301.
Mettananda, S., C. A. Fisher, D. Hay, M. Badat, L. Quek, K. Clark, P. Hublitz, D. Downes, J. Kerry, M. Gosden, J. Telenius, J. A. Sloane-Stanley, P. Faustino, A. Coelho, J. Doondeea, B. Usukhbayar, P. Sopp, J. A. Sharpe, J. R. Hughes, P. Vyas, R. J. Gibbons and D. R. Higgs (2017). "Editing an α-globin enhancer in primary human hematopoietic stem cells as a treatment for β-thalassemia." Nat Commun **8**(1): 424.
Metzner, A., M. A. Horstmann, B. Fehse, G. Ortmeyer, C. M. Niemeyer, C. Stocking, G. W. Mayr and M. Jücker (2007). "Gene transfer of SHIP-1 inhibits proliferation of juvenile myelomonocytic leukemia cells carrying KRAS2 or PTPN11 mutations." Gene Ther **14**(8): 699-703.
Mojica, F. J., C. Ferrer, G. Juez and F. Rodriguez-Valera (1995). "Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning." Mol Microbiol **17**(1): 85-93.
Mulero-Navarro, S., A. Sevilla, A. C. Roman, D. F. Lee, S. L. D'Souza, S. Pardo, I. Riess, J. Su, N. Cohen, C. Schaniel, N. A. Rodriguez, A. Baccarini, B. D. Brown, H. Cavé, A. Caye, M. Strullu, S. Yalcin, C. Y. Park, P. S. Dhandapany, G. Yongchao, L. Edelmann, S. Bahieg, P. Raynal, E. Flex, M. Tartaglia, K. A. Moore, I. R. Lemischka and B. D. Gelb (2015). "Myeloid dysregulation in a human induced pluripotent stem cell model of PTPN11-associated juvenile myelomonocytic leukemia." Cell Rep **13**(3): 504-515.
Muller-Sieburg, C. E., R. H. Cho, L. Karlsson, J. F. Huang and H. B. Sieburg (2004). "Myeloid-biased hematopoietic stem cells have extensive self-renewal capacity but generate diminished lymphoid progeny with impaired IL-7 responsiveness." Blood **103**(11): 4111-4118.
Nakamura, Y., M. Ito, T. Yamamoto, X. Y. Yan, H. Yagasaki, Y. Kamachi, K. Kudo and S. Kojima (2005). "Engraftment of NOD/SCID/gammac(null) mice with multilineage neoplastic cells from patients with juvenile myelomonocytic leukaemia." Br J Haematol **130**(1): 51-57.
Nasri, M., P. Mir, B. Dannenmann, D. Amend, T. Skroblyn, Y. Xu, K. Schulze-Osthoff, M. Klimiankou, K. Welte and J. Skokowa (2019). "Fluorescent labeling of CRISPR/Cas9 RNP for gene knockout in HSPCs and iPSCs reveals an essential role for GADD45b in stress response." Blood Adv **3**(1): 63-71.
Nestorowa, S., F. K. Hamey, B. Pijuan Sala, E. Diamanti, M. Shepherd, E. Laurenti, N. K. Wilson, D. G. Kent and B. Göttgens (2016). "A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation." Blood **128**(8): e20-31.
Ng, A. P. and W. S. Alexander (2017). "Haematopoietic stem cells: past, present and future." Cell Death Discov **3**: 17002.
Nie, Y., Y. C. Han and Y. R. Zou (2008). "CXCR4 is required for the quiescence of primitive hematopoietic cells." J Exp Med **205**(4): 777-783.
Niemeyer, C. M. (2018). "JMML genomics and decisions." Hematology Am Soc Hematol Educ Program **2018**(1): 307-312.
Niemeyer, C. M. and C. Flotho (2019). "Juvenile myelomonocytic leukemia: who's the driver at the wheel?" Blood **133**(10): 1060-1070.
Niemeyer, C. M., M. W. Kang, D. H. Shin, I. Furlan, M. Erlacher, N. J. Bunin, S. Bunda, J. Z. Finklestein, T. A. Gorr, P. Mehta, I. Schmid, G. Kropshofer, S. Corbacioglu, P. J. Lang, C. Klein, P.

G. Schlegel, A. Heinzmann, M. Schneider, J. Starý, M. M. van den Heuvel-Eibrink, H. Hasle, F. Locatelli, D. Sakai, S. Archambeault, L. Chen, R. C. Russell, S. S. Sybingco, M. Ohh, B. S. Braun, C. Flotho and M. L. Loh (2010). "Germline CBL mutations cause developmental abnormalities and predispose to juvenile myelomonocytic leukemia." Nat Genet **42**(9): 794-800.
Notta, F., S. Doulatov, E. Laurenti, A. Poeppl, I. Jurisica and J. E. Dick (2011). "Isolation of Single Human Hematopoietic Stem Cells Capable of Long-Term Multilineage Engraftment."
Notta, F., S. Zandi, N. Takayama, S. Dobson, O. I. Gan, G. Wilson, K. B. Kaufmann, J. McLeod, E. Laurenti, C. F. Dunant, J. D. McPherson, L. D. Stein, Y. Dror and J. E. Dick (2016). "DISTINCT ROUTES OF LINEAGE DEVELOPMENT RESHAPE THE HUMAN BLOOD HIERARCHY ACROSS ONTOGENY." Science **351**(6269): aab2116.
Oguro, H., L. Ding and S. J. Morrison (2013). "SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors." Cell Stem Cell **13**(1): 102-116.
Okazaki, T. and Y. Gotoh (2018). "An Unexpected Calm: Mfge8 Controls Stem Cell Quiescence and Maintenance." Cell Stem Cell **23**(3): 311-312.
Osawa, M., K. Hanada, H. Hamada and H. Nakauchi (1996). "Long-term lymphohematopoietic reconstitution by a single CD34-low/negative hematopoietic stem cell." Science **273**(5272): 242-245.
Paul, F., Y. Arkin, A. Giladi, D. A. Jaitin, E. Kenigsberg, H. Keren-Shaul, D. Winter, D. Lara-Astiaso, M. Gury, A. Weiner, E. David, N. Cohen, F. K. Lauridsen, S. Haas, A. Schlitzer, A. Mildner, F. Ginhoux, S. Jung, A. Trumpp, B. T. Porse, A. Tanay and I. Amit (2015). "Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors." Cell **163**(7): 1663-1677.
Pei, W., F. Shang, X. Wang, A. K. Fanti, A. Greco, K. Busch, K. Klapproth, Q. Zhang, C. Quedenau, S. Sauer, T. B. Feyerabend, T. Höfer and H. R. Rodewald (2020). "Resolving Fates and Single-Cell Transcriptomes of Hematopoietic Stem Cell Clones by PolyloxExpress Barcoding." Cell Stem Cell **27**(3): 383-395.e388.
Petrov, I., M. Suntsova, O. Mutorova, M. Sorokin, A. Garazha, E. Ilnitskaya, P. Spirin, S. Larin, O. Kovalchuk, V. Prassolov, A. Zhavoronkov, A. Roumiantsev and A. Buzdin (2016). "Molecular pathway activation features of pediatric acute myeloid leukemia (AML) and acute lymphoblast leukemia (ALL) cells." Aging (Albany NY) **8**(11): 2936-2947.
Philpott, C., H. Tovell, I. M. Frayling, D. N. Cooper and M. Upadhyaya (2017). "The NF1 somatic mutational landscape in sporadic human cancers." Hum Genomics **11**(1): 13.
Pietras, E. M., D. Reynaud, Y. A. Kang, D. Carlin, F. J. Calero-Nieto, A. D. Leavitt, J. M. Stuart, B. Gottgens and E. Passegue (2015). "Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions." Cell Stem Cell **17**(1): 35-46.
Pilzecker, B., O. A. Buoninfante, P. van den Berk, C. Lancini, J. Y. Song, E. Citterio and H. Jacobs (2017). "DNA damage tolerance in hematopoietic stem and progenitor cells in mice." Proc Natl Acad Sci U S A **114**(33): E6875-e6883.
Popescu, D. M., R. A. Botting, E. Stephenson, K. Green, S. Webb, L. Jardine, E. F. Calderbank, K. Polanski, I. Goh, M. Efremova, M. Acres, D. Maunder, P. Vegh, Y. Gitton, J. E. Park, R. Vento-Tormo, Z. Miao, D. Dixon, R. Rowell, D. McDonald, J. Fletcher, E. Poyner, G. Reynolds, M. Mather, C. Moldovan, L. Mamanova, F. Greig, M. D. Young, K. B. Meyer, S. Lisgo, J. Bacardit, A. Fuller, B. Millar, B. Innes, S. Lindsay, M. J. T. Stubbington, M. S. Kowalczyk, B. Li, O. Ashenberg, M. Tabaka, D. Dionne, T. L. Tickle, M. Slyper, O. Rozenblatt-Rosen, A. Filby, P. Carey, A. C. Villani, A. Roy, A. Regev, A. Chédotal, I. Roberts, B. Göttgens, S. Behjati, E. Laurenti, S. A. Teichmann and M. Haniffa (2019). "Decoding human fetal liver haematopoiesis." Nature **574**(7778): 365-371.

Pourcel, C., G. Salvignol and G. Vergnaud (2005). "CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies." Microbiology **151**(Pt 3): 653-663.
Primer3. (2021). "Primer 3 web design tool." from https://primer3.ut.ee/.
Psaila, B., N. Barkas, D. Iskander, A. Roy, S. Anderson, N. Ashley, V. S. Caputo, J. Lichtenberg, S. Loaiza, D. M. Bodine, A. Karadimitris, A. J. Mead and I. Roberts (2016). "Single-cell profiling of human megakaryocyte-erythroid progenitors identifies distinct megakaryocyte and erythroid differentiation pathways." Genome Biol **17**: 83.
Ramalho-Santos, M., S. Yoon, Y. Matsuzaki, R. C. Mulligan and D. A. Melton (2002). ""Stemness": transcriptional profiling of embryonic and adult stem cells." Science **298**(5593): 597-600.
Ran, F. A., P. D. Hsu, C. Y. Lin, J. S. Gootenberg, S. Konermann, A. E. Trevino, D. A. Scott, A. Inoue, S. Matoba, Y. Zhang and F. Zhang (2013). "Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity." Cell **154**(6): 1380-1389.
Ran, F. A., P. D. Hsu, J. Wright, V. Agarwala, D. A. Scott and F. Zhang (2013). "Genome engineering using the CRISPR-Cas9 system." Nature Protocols **8**: 2281-2308.
Rao, S., K. Gharib and A. Han (2019). "Cancer Immunosurveillance by T Cells." Int Rev Cell Mol Biol **342**: 149-173.
Rath, D., L. Amlinger, A. Rath and M. Lundgren (2015). "The CRISPR-Cas immune system: biology, mechanisms and applications." Biochimie **117**: 119-128.
Rieger, M. A. and T. Schroeder (2012). "Hematopoiesis." Cold Spring Harb Perspect Biol **4**(12).
River, C. (2021). "https://www.criver.com/sites/default/files/resources/NODSCIDMouseClinicalPathologyData.pdf." from https://www.criver.com/sites/default/files/resources/NODSCIDMouseClinicalPathologyData.pdf.
Rodriguez-Meira, A., G. Buck, S. A. Clark, B. J. Povinelli, V. Alcolea, E. Louka, S. McGowan, A. Hamblin, N. Sousos, N. Barkas, A. Giustacchini, B. Psaila, S. E. W. Jacobsen, S. Thongjuea and A. J. Mead (2019). "Unravelling Intratumoral Heterogeneity through High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing." Mol Cell **73**(6): 1292-1305.e1298.
Rosendaal, M., G. S. Hodgson and T. R. Bradley (1979). "Organization of haemopoietic stem cells: the generation-age hypothesis." Cell Tissue Kinet **12**(1): 17-29.
Salisbury-Ruf, C. T. and A. Larochelle (2021). "Advances and Obstacles in Homology-Mediated Gene Editing of Hematopoietic Stem Cells." J Clin Med **10**(3).
Sanjuan-Pla, A., I. C. Macaulay, C. T. Jensen, P. S. Woll, T. C. Luis, A. Mead, S. Moore, C. Carella, S. Matsuoka, T. B. Jones, O. Chowdhury, L. Stenson, M. Lutteropp, J. C. A. Green, R. Facchini, H. Boukarabila, A. Grover, A. Gambardella, S. Thongjuea, J. Carrelha, P. Tarrant, D. Atkinson, S.-A. Clark, C. Nerlov and S. E. W. Jacobsen (2013). "Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy." Nature **502**: 232-236.
Schiroli, G., A. Conti, S. Ferrari, L. Della Volpe, A. Jacob, L. Albano, S. Beretta, A. Calabria, V. Vavassori, P. Gasparini, E. Salataj, D. Ndiaye-Lobry, C. Brombin, J. Chaumeil, E. Montini, I. Merelli, P. Genovese, L. Naldini and R. Di Micco (2019). "Precise Gene Editing Preserves Hematopoietic Stem Cell Function following Transient p53-Mediated DNA Damage Response." Cell Stem Cell **24**(4): 551-565.e558.
Schubbert, S., G. Bollag, N. Lyubynska, H. Nguyen, C. P. Kratz, M. Zenker, C. M. Niemeyer, A. Molven and K. Shannon (2007). "Biochemical and functional characterization of germ line KRAS mutations." Mol Cell Biol **27**(22): 7765-7770.
Schubbert, S., K. Lieuw, S. L. Rowe, C. M. Lee, X. Li, M. L. Loh, D. W. Clapp and K. M. Shannon (2005). "Functional analysis of leukemia-associated PTPN11 mutations in primary hematopoietic cells." Blood **106**(1): 311-317.

Schubbert, S., M. Zenker, S. L. Rowe, S. Böll, C. Klein, G. Bollag, I. van der Burgt, L. Musante, V. Kalscheuer, L. E. Wehner, H. Nguyen, B. West, K. Y. Zhang, E. Sistermans, A. Rauch, C. M. Niemeyer, K. Shannon and C. P. Kratz (2006). "Germline KRAS mutations cause Noonan syndrome." Nat Genet **38**(3): 331-336.
Schulte, R., N. K. Wilson, J. C. Prick, C. Cossetti, M. K. Maj, B. Gottgens and D. G. Kent (2015). Index sorting resolves heterogeneous murine hematopoietic stem cell populations. Exp Hematol. **43:** 803-811.
Seita, J., H. Ema, J. Ooehara, S. Yamazaki, Y. Tadokoro, A. Yamasaki, K. Eto, S. Takaki, K. Takatsu and H. Nakauchi (2007). "Lnk negatively regulates self-renewal of hematopoietic stem cells by modifying thrombopoietin-mediated signal transduction." Proc Natl Acad Sci U S A **104**(7): 2349-2354.
Seita, J. and I. L. Weissman (2010). "Hematopoietic stem cell: self-renewal versus differentiation." Wiley Interdiscip Rev Syst Biol Med **2**(6): 640-653.
Sell, S. (2005). "Leukemia: stem cells, maturation arrest, and differentiation therapy." Stem Cell Rev **1**(3): 197-205.
Shah, F. T., F. Sayani, S. Trompeter, E. Drasar and A. Piga (2019). "Challenges of blood transfusions in β-thalassemia." Blood Rev **37**: 100588.
Shalem, O., N. E. Sanjana, E. Hartenian, X. Shi, D. A. Scott, T. Mikkelson, D. Heckl, B. L. Ebert, D. E. Root, J. G. Doench and F. Zhang (2014). "Genome-scale CRISPR-Cas9 knockout screening in human cells." Science **343**(6166): 84-87.
Shaner, N. C., R. E. Campbell, P. A. Steinbach, B. N. Giepmans, A. E. Palmer and R. Y. Tsien (2004). "Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein." Nat Biotechnol **22**(12): 1567-1572.
Shiba, N., M. Kato, M. J. Park, M. Sanada, E. Ito, K. Fukushima, M. Sako, H. Arakawa, S. Ogawa and Y. Hayashi (2010). CBL mutations in juvenile myelomonocytic leukemia and pediatric myelodysplastic syndrome. Leukemia. England. **24:** 1090-1092.
Silva, G., L. Poirot, R. Galetto, J. Smith, G. Montoya, P. Duchateau and F. Pâques (2011). "Meganucleases and Other Tools for Targeted Genome Engineering: Perspectives and Challenges for Gene Therapy." Curr Gene Ther **11**(1): 11-27.
Singh, R. P., K. Franke, J. Kalucka, S. Mamlouk, A. Muschter, A. Gembarska, T. Grinenko, C. Willam, R. Naumann, K. Anastassiadis, A. F. Stewart, S. Bornstein, T. Chavakis, G. Breier, C. Waskow and B. Wielockx (2013). "HIF prolyl hydroxylase 2 (PHD2) is a critical regulator of hematopoietic stem cell maintenance during steady-state and stress." Blood **121**(26): 5158-5166.
Slaymaker, I. M., L. Gao, B. Zetsche, D. A. Scott, W. X. Yan and F. Zhang (2016). "Rationally engineered Cas9 nucleases with improved specificity." Science **351**(6268): 84-88.
Spangrude, G. J., S. Heimfeld and I. L. Weissman (1988). "Purification and characterization of mouse hematopoietic stem cells." Science **241**(4861): 58-62.
Spits, H. (2002). "Development of alphabeta T cells in the human thymus." Nat Rev Immunol **2**(10): 760-772.
Sterner, R. C. and R. M. Sterner (2021). "CAR-T cell therapy: current limitations and potential strategies." Blood Cancer J **11**(4): 69.
Stieglitz, E., A. N. Taylor-Weiner, T. Y. Chang, L. C. Gelston, Y. D. Wang, T. Mazor, E. Esquivel, A. Yu, S. Seepo, S. Olsen, M. Rosenberg, S. L. Archambeault, G. Abusin, K. Beckman, P. A. Brown, M. Briones, B. Carcamo, T. Cooper, G. V. Dahl, P. D. Emanuel, M. N. Fluchel, R. K. Goyal, R. J. Hayashi, J. Hitzler, C. Hugge, Y. L. Liu, Y. H. Messinger, D. H. Mahoney, P. Monteleone, E. R. Nemecek, P. A. Roehrs, R. J. Schore, K. C. Stine, C. M. Takemoto, J. A. Toretsky, J. F. Costello, A. B. Olshen, C. Stewart, Y. Li, J. Ma, R. B. Gerbing, T. A. Alonzo, G. Getz, T. Gruber, T. Golub, K. Stegmaier and M. L. Loh (2015). "The Genomic Landscape of Juvenile Myelomonocytic Leukemia." Nat Genet **47**(11): 1326-1333.

Suzuki, M., T. Moriguchi, K. Ohneda and M. Yamamoto (2009). "Differential Contribution of the Gata1 Gene Hematopoietic Enhancer to Erythroid Differentiation ▿ ��." Mol Cell Biol **29**(5): 1163-1175.
Synthego. (2021). "ICE Sensitivity." from https://www.synthego.com/help/ice-sensitivity.
SynthegoGuideDesign. (2021). "Synthego Guide RNA design tool." from https://www.synthego.com/products/bioinformatics/crispr-design-tool.
Takacova, S., R. Slany, J. Bartkova, V. Stranecky, P. Dolezel, P. Luzna, J. Bartek and V. Divoky (2012). "DNA damage response and inflammatory signaling limit the MLL-ENL-induced leukemogenesis in vivo." Cancer Cell **21**(4): 517-531.
Takeda, K., H. L. Aguila, N. S. Parikh, X. Li, K. Lamothe, L. J. Duan, H. Takeda, F. S. Lee and G. H. Fong (2008). "Regulation of adult erythropoiesis by prolyl hydroxylase domain proteins." Blood **111**(6): 3229-3235.
Tartaglia, M. and B. D. Gelb (2005). "Noonan syndrome and related disorders: genetics and pathogenesis." Annu Rev Genomics Hum Genet **6**: 45-68.
Thein, S. L. (2013). "The Molecular Basis of β-Thalassemia." Cold Spring Harb Perspect Med **3**(5).
Theocharides, A. P., A. Rongvaux, K. Fritsch, R. A. Flavell and M. G. Manz (2016). "Humanized hemato-lymphoid system mice." Haematologica **101**(1): 5-19.
Thomas, M., D. Parry-Smith and V. Iyer (2019). "Best practice for CRISPR design using current tools and resources." Methods **164-165**: 3-17.
Tothova, Z., J. M. Krill-Burger, K. D. Popova, C. C. Landers, Q. L. Sievers, D. Yudovich, R. Belizaire, J. C. Aster, E. A. Morgan, A. Tsherniak and B. L. Ebert (2017). "Multiplex CRISPR/Cas9-Based Genome Editing in Human Hematopoietic Stem Cells Models Clonal Hematopoiesis and Myeloid Neoplasia." Cell Stem Cell **21**(4): 547-555.e548.
Vakulskas, C. A., D. P. Dever, G. R. Rettig, R. Turk, A. M. Jacobi, M. A. Collingwood, N. M. Bode, M. S. McNeill, S. Yan, J. Camarena, C. M. Lee, S. H. Park, V. Wiebking, R. O. Bak, N. Gomez-Ospina, M. Pavel-Dinu, W. Sun, G. Bao, M. H. Porteus and M. A. Behlke (2018). "A high-fidelity Cas9 mutant delivered as a ribonucleoprotein complex enables efficient gene editing in human hematopoietic stem and progenitor cells." Nat Med **24**(8): 1216-1224.
Valletta, S., H. Dolatshad, M. Bartenstein, B. H. Yip, E. Bello, S. Gordon, Y. Yu, J. Shaw, S. Roy, L. Scifo, A. Schuh, A. Pellagatti, T. A. Fulga, A. Verma and J. Boultwood (2015). "ASXL1 mutation correction by CRISPR/Cas9 restores gene function in leukemia cells and increases survival in mouse xenografts." Oncotarget **6**(42): 44061-44071.
Vazquez, S. E., M. A. Inlay and T. Serwold (2015). "CD201 and CD27 identify hematopoietic stem and progenitor cells across multiple murine strains independently of Kit and Sca-1." Exp Hematol **43**(7): 578-585.
Velten, L., S. F. Haas, S. Raffel, S. Blaszkiewicz, S. Islam, B. P. Hennig, C. Hirche, C. Lutz, E. C. Buss, D. Nowak, T. Boch, W. K. Hofmann, A. D. Ho, W. Huber, A. Trumpp, M. A. Essers and L. M. Steinmetz (2017). "Human haematopoietic stem cell lineage commitment is a continuous process." Nat Cell Biol **19**(4): 271-281.
Vijey, P., B. Posorske and K. R. Machlus (2018). "In vitro culture of murine megakaryocytes from fetal liver-derived hematopoietic stem cells." Platelets **29**(6): 583-588.
Wagenblast, E., M. Azkanaz, S. A. Smith, L. Shakib, J. L. McLeod, G. Krivdova, J. Araujo, L. D. Shultz, O. I. Gan, J. E. Dick and E. R. Lechman (2019). "Functional profiling of single CRISPR/Cas9-edited human long-term hematopoietic stem cells." Nat Commun **10**(1): 4730.
Wang, D., H. Mou, S. Li, Y. Li, S. Hough, K. Tran, J. Li, H. Yin, D. G. Anderson, E. J. Sontheimer, Z. Weng, G. Gao and W. Xue (2015). "Adenovirus-Mediated Somatic Genome Editing of Pten by CRISPR/Cas9 in Mouse Liver in Spite of Cas9-Specific Immune Responses." Hum Gene Ther **26**(7): 432-442.
Wang, Y., B. A. Schulte, A. C. LaRue, M. Ogawa and D. Zhou (2006). "Total body irradiation selectively induces murine hematopoietic stem cell senescence." Blood **107**(1): 358-366.

Weissman, I. L. and J. A. Shizuru (2008). "The origins of the identification and isolation of hematopoietic stem cells, and their capability to induce donor-specific transplantation tolerance and treat autoimmune diseases." Blood **112**(9): 3543-3553.
Wilkinson, A. C., D. P. Dever, R. Baik, J. Camarena, I. Hsu, C. T. Charlesworth, C. Morita, H. Nakauchi and M. H. Porteus (2021). "Cas9-AAV6 gene correction of beta-globin in autologous HSCs improves sickle cell disease erythropoiesis in mice." Nat Commun **12**(1): 686.
Wilkinson, A. C., R. Ishida, H. Nakauchi and S. Yamazaki (2020). "Long-term ex vivo expansion of mouse hematopoietic stem cells." Nat Protoc **15**(2): 628-648.
Wilson, A., E. Laurenti, G. Oser, R. C. van der Wath, W. Blanco-Bose, M. Jaworski, S. Offner, C. F. Dunant, L. Eshkind, E. Bockamp, P. Lio, H. R. Macdonald and A. Trumpp (2008). "Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair." Cell **135**(6): 1118-1129.
Wilson, N. K., D. G. Kent, F. Buettner, M. Shehata, I. C. Macaulay, F. J. Calero-Nieto, M. Sanchez Castillo, C. A. Oedekoven, E. Diamanti, R. Schulte, C. P. Ponting, T. Voet, C. Caldas, J. Stingl, A. R. Green, F. J. Theis and B. Gottgens (2015). "Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations." Cell Stem Cell **16**(6): 712-724.
Wu, Y., J. Zeng, B. P. Roscoe, P. Liu, Q. Yao, C. R. Lazzarotto, K. Clement, M. A. Cole, K. Luk, C. Baricordi, A. H. Shen, C. Ren, E. B. Esrick, J. P. Manis, D. M. Dorfman, D. A. Williams, A. Biffi, C. Brugnara, L. Biasco, C. Brendel, L. Pinello, S. Q. Tsai, S. A. Wolfe and D. E. Bauer (2019). "Highly efficient therapeutic gene editing of human hematopoietic stem cells." Nat Med **25**(5): 776-783.
Wunderlich, M., F. S. Chou, C. Sexton, P. Presicce, C. A. Chougnet, J. Aliberti and J. C. Mulloy (2018). "Improved multilineage human hematopoietic reconstitution and function in NSGS mice." PLoS One **13**(12): e0209034.
Xavier-Ferrucio, J. and D. S. Krause (2018). "Concise Review: Bipotent Megakaryocytic-Erythroid Progenitors: Concepts and Controversies." Stem Cells **36**(8): 1138-1145.
Xie, F., L. Ye, J. C. Chang, A. I. Beyer, J. Wang, M. O. Muench and Y. W. Kan (2014). "Seamless gene correction of β-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac." Genome Res **24**(9): 1526-1533.
Yamamoto, R., Y. Morita, J. Ooehara, S. Hamanaka, M. Onodera, K. L. Rudolph, H. Ema and H. Nakauchi (2013). "Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells." Cell **154**(5): 1112-1126.
Yamamoto, R., A. C. Wilkinson, J. Ooehara, X. Lan, C. Y. Lai, Y. Nakauchi, J. K. Pritchard and H. Nakauchi (2018). "Large-Scale Clonal Analysis Resolves Aging of the Mouse Hematopoietic Stem Cell Compartment." Cell Stem Cell **22**(4): 600-607.e604.
Yamauchi, T., T. Masuda, M. C. Canver, M. Seiler, Y. Semba, M. Shboul, M. Al-Raqad, M. Maeda, V. A. C. Schoonenberg, M. A. Cole, C. Macias-Trevino, Y. Ishikawa, Q. Yao, M. Nakano, F. Arai, S. H. Orkin, B. Reversade, S. Buonamici, L. Pinello, K. Akashi, D. E. Bauer and T. Maeda (2018). "Genome-wide CRISPR-Cas9 Screen Identifies Leukemia-Specific Dependence on a Pre-mRNA Metabolic Pathway Regulated by DCPS." Cancer Cell.
Yap, Y. S., J. R. McPherson, C. K. Ong, S. G. Rozen, B. T. Teh, A. S. Lee and D. F. Callen (2014). "The NF1 gene revisited - from bench to bedside." Oncotarget **5**(15): 5873-5892.
Yilmaz Ö, H., M. J. Kiel and S. J. Morrison (2006). SLAM family markers are conserved among hematopoietic stem cells from old and reconstituted mice and markedly increase their purity. Blood. **107:** 924-930.
Yokota, T., K. Oritani, S. Butz, K. Kokame, P. W. Kincade, T. Miyata, D. Vestweber and Y. Kanakura (2009). "The endothelial antigen ESAM marks primitive hematopoietic progenitors throughout life in mice." Blood **113**(13): 2914-2923.
Yoshimi, A., M. E. Balasis, A. Vedder, K. Feldman, Y. Ma, H. Zhang, S. C. Lee, C. Letson, S. Niyongere, S. X. Lu, M. Ball, J. Taylor, Q. Zhang, Y. Zhao, S. Youssef, Y. R. Chung, X. J. Zhang, B.

H. Durham, W. Yang, A. F. List, M. L. Loh, V. Klimek, M. F. Berger, E. Stieglitz, E. Padron and O. Abdel-Wahab (2017). "Robust patient-derived xenografts of MDS/MPN overlap syndromes capture the unique characteristics of CMML and JMML." Blood **130**(4): 397-407.
Zhang, Y., G. Parmigiani and W. E. Johnson (2020). "ComBat-seq: batch effect adjustment for RNA-seq count data." NAR Genom Bioinform **2**(3): lqaa078.
Zonari, E., G. Desantis, C. Petrillo, F. E. Boccalatte, M. R. Lidonnici, A. Kajaste-Rudnitski, A. Aiuti, G. Ferrari, L. Naldini and B. Gentner (2017). "Efficient Ex Vivo Engineering and Expansion of Highly Purified Human Hematopoietic Stem and Progenitor Cell Populations for Gene Therapy." Stem Cell Reports **8**(4): 977-990.

Appendix 1: List of monoclonal antibodies used in this project

Antibody	Anti-mouse/human	Company/Clone	Catalogue number
7-AAD	N/A	Thermofisher	A1310
B220 FITC	Mouse	BioLegend/RA3-6B2	11-0452-85
B220 PECY5	Mouse	BioLegend/RA3-6B2	103210
CD10 FITC	Human	BioLegend/HI10a	312208
CD11b FITC	Human	Invitrogen/ICRF44	11-0118-42
CD123 PECY7	Human	BioLegend/6H6	306010
CD127 FITC	Human	Invitrogen/eBioRDR5	11-1278-42
CD14 FITC	Human	Invitrogen/61D3	11-0149-42
CD15 FITC	Human	BioLegend/MMA	394706
CD150 APC	Mouse	BioLegend/TC15-12F12.2	115910
CD150 PECY7	Mouse	BioLegend/TC15-12F12.2	115714
CD16 FITC	Human	Invitrogen/eBioCB16	11-0168-41
CD19 FITC	Human	Invitrogen/HIB19	11-0199-42
CD2 FITC	Human	BD/RPA-2.10	555326
CD20 FITC	Human	BioLegend/2H7	302304
CD235a FITC	Human	eBioscience/10F7MN	11-9886-42
CD235a PerCP Cy5.5	Human	BioLegend/HIR2	306614
CD3 FITC	Mouse	BD/145-2C11	553062
CD3 FITC	Human	BioLegened/SK7 (Leu-4)	345763
CD3 PECY5	Mouse	BioLegend/145-2C11	100310
CD33 FITC	Human	BioLegend/HIM3-4	303304
CD34 APC-ef780	Human	eBioscience/4H11	47-0349-42
CD38 PETx Red	Human	Invitrogen/HIT2	MHCD3817
CD41 BV421	Mouse	BioLegend/MWReg30	133911
CD45RA BV650	Human	BioLegend/HI100	304136
CD48 APC	Mouse	BioLegend/HM48-1	103412
CD66b FITC	Human	BioLegend/G10F5	305104
CD8 FITC	Human	BioLegend/RPA-T8	301006
CD90 BV421	Human	BioLegend/5E10	328122
c-Kit APC-ef780	Mouse	eBioscience/2B8	47-1171-82
DAPI	N/A	ThermoFisher	62248
F4/80 FITC	Mouse	BioLegend/BM8	123108

F4/80 PECY5	Mouse	BioLegend/BM8	123112
Gr1 FITC	Mouse	eBioscience/RB6-8C5	11-5931-85
Gr1 PECY5	Mouse	BioLegend/RB6-8C5	108410
hCD45 AF700	Human	BioLegend/HI30	304024
Ki-67 AF700	N/A	BioLegend/B56	652420
Mac-1 FITC	Mouse	eBioscience/M1-70	11-0112-85
Mac-1 PECY5	Mouse	BioLegend/M1-70	101210
mCD45 PE	Mouse	eBioscience/30-F11	12-0451
mCD45 PETx Red	Mouse	Invitrogen/30-F11	MCD4517
NK1.1 FITC	Mouse	BD/PK136	553164
NK1.1 PECY5	Mouse	BioLegend/PK136	108716
Sca-1 BV605	Mouse	BioLegend/D7	108133
Sca-1 PB	Mouse	BioLegend/E13-161.7	122520
Ter119 BV650	Mouse	BioLegend/TER-119	116235
Ter119 FITC	Mouse	BioLegend/TER-119	11-5921-85
Ter119 PECY5	Mouse	BioLegend/TER-119	116210

Appendix 2: Detailed large cell number RNP nucleofection protocol for tdTomato knockdown in mouse HSPCs

Mouse c-Kit cell preparation

For all aspiration steps use vacuum pump

1. Obtain the tibia and femur from one adult (8-12 weeks) CD45.1 TOMC female (TOMC60.1e) and one adult (8-12 weeks) CD45.2 female. Obtain the bone marrow by crushing the bones with a pestle and mortar in MACS buffer. LS columns: $1x10^8$ max labelled cells. $2x10^9$ max total cells per column. As long as total cells doesn't exceed $2x10^9$ use 1 column per strain

2. Put the cells through a 70µm cell strainer placed on a 50mL falcon tube

3. Count cells diluted with trypan blue 1:10 cell count

4. Pellet the cells at 500g for 5 mins at room temperature

5. Aspirate supernatant and resuspend cells in 100µl/100 million cells MACS BUFFER. NB: If cells not re-suspending easily increase this to 300uL/100 million cells.

6. Add 2.5µl CD117 beads/100 million cells. (Shake tube well before). NB: increase to 7.5uL/100 million cells if have to increase MACS buffer volume.

7. Transfer to an eppendorf tube.

8. Incubate for 20 minutes on ice with tube lying flat and mix every 2mins during incubation by carefully inverting tube.

9. Wash with 3ml of MACS buffer (transfer after 1mL to 15mL falcon and re-wash eppendorf with another 2mL using same tip, use fresh tip though to add MACS buffer) and spin down at 500g for 5mins.

10. In the meantime, wash the column with 3ml of MACS buffer

11. Aspirate supernatant and resuspend the cells into 3ml of MACS buffer and add them to the column through a Falcon filter

12. Perform 1 washes of 3ml

13. Remove the column from the magnet. Place it into a 15ml Falcon tube. Flush out the cells with 5ml, with the help of the plunger, and then rewash column with 5ml. So ckit+ cells are eluted into a volume of 10ml

14. Count cells diluted 1:2 with trypan blue

15. Pellet cells 500g 5 mins at room temperature

16. Resuspend in 1mL StemSpan per million cells with appropriate cytokines (50 ng/ml SCF, 50 ng/ml TPO, 10 ng/ml IL-3, and 10 ng/ml IL-6, 50ng/mL Flt3L), place in a 12 well plate at $1x10^6$ cells per mL and incubate 24hrs in an incubator at 37ºC 5% CO2 to allow cells to recover

Day 2

17. 30 mins before nucleofection prepare StemSpan media with cytokines (50 ng/ml SCF, 50 ng/ml TPO, 10 ng/ml IL-3, and 10 ng/ml IL-6, 50ng/mL Flt3-ligand, 1% P/S) for transfer post transfection. Prepare 1mL per sample in a 12 well and place in an incubator at 37ºC 5% CO2. Remove Primary Cell Buffer P3 from 4 degrees to allow to reach room temperature. If new kit, add supplement to buffer. Turn on PCR machine. Collect 4D nucleofector from clean TC.

18. Remove all c-kit enriched cells from 12 well plate and wash with MACS buffer. Pellet cells at 350g for 10mins room temperature.

• Dilute Cas9 in Neb Diluent B to 6ug/uL.

• Clean PCR tube rack with RNaseZap

• Prepare guide RNA @6.4ug per sample

• Combine Cas9 with RNA

• Place PCR tubes in PCR machine and incubate at 21 degrees for 10 minutes

19. Resuspend cells @1,000,000 per 100uL in Primary Cell buffer P3.

20. To each RNP (guide + Cas9) tube at 100uL cells and mix by pipetting up and down

21. Transfer 100uL of cell/RNP to separate large cuvettes and tap briefly to ensure equal cell distribution. Run Primary Cell P3 DK100 program on 4D nucleofector. Immediately transfer the cells to the pre-warmed 12 well media plate and incubate at 37 degrees 5% CO2. 1 sample per well.

Appendix 3: RNA library indexing using the Nextera XT (Illumina, FC-131-1096) kit

1. Take reagents and samples out:

- Qubit Ampure cleaned cDNA and dilute to 0.7ng/1.25uL
- Make sure beads for PCR purification are brought to room temperature 15-30min before.
- PCR plates: make sure they fit in the PCR machine, use semi-skirted plates (Axygen, 96-M2-HS-C)
- Seals, use AB0558 from Thermofisher.

On ice (reagents from -20 box):

- isofreeze rack
- Tagment DNA buffer
- Resuspension Buffer
- Indexes
- Samples
- Index holder rack

On RT:

- Neutralization buffer (NT) in 4C fridge.

2. Turn the PCR ON

Use program that has two ‘hold’ steps and that includes both ‘PCR’ steps, start program (10C on hold, lid will be warm when placing the plate)

3. Prepare Sample dilutions

4. Tagmentation (work at RT for this step)

- Move empty PCR plate to isofreeze rack

- Vortex and spin **Tagmentation buffer** and add 2.5ul to each well
- Add 1.25ul undiluted DNA samples
- Take **ATM** (amplicon tagment mix) out from -20°C, agitate (do not vortex) and spin down. Add 1.25ul to each tube.
- Put ATM back to -20°C, seal plate, vortex and spin down PCR tubes/plates.
- Run Nextera Tag program (55°C 6 min, 10°C), reaction volume 5ul. Let temperature come up before adding samples to the machine. Take plate out immediately when temperature reaches 10°C.

5. Neutralization

- Add 1.25ul of NT buffer (vortex/spin) to each tube and move sample to RT.
- Vortex and spin tubes
- Incubate 5 minutes

Once the neutralization buffer has been added, there is 'no rush'

6. Indexing (keep working at RT)

- Vortex and spin all indexes
- Add 1.25ul of white index (S50X)
- Add 1.25ul of orange index (N70X)
- Move tubes to isofreeze rack, add 3.75ul of **NPM (-20°C) buffer (take out last)** to each tube.
- Vortex and spin tubes
- Run Nextera XT PCR program 12.5 ul volume

Program:	72°C	3 min	
	95°C	30 sec	
	95°C	10 sec|	
	55°C	30 sec|	12 cycles
	72°C	30sec |	
	72°C	5 min	
	10°C	hold	

Ideally continue with PCR purification, however, plate can be kept at 4°C for up to 2 days (or put it in -20°C)

For next step:

- Put Ampure beads at RT

- Put Resuspension Buffer at RT (use resuspension buffer from kit)

7. PCR Purification

If possible, do this on the same day

- Set up 96-well V bottom plate, by adding 25µl of beads in as many wells as the samples processed
- Add to each well containing beads the (**12.5µL of PCR product +12.5µL water**) coming from the previous step
- Mix with multichannel 10X
- Incubate 5-8 minutes to let the bids bind to DNA
- During incubation prepare fresh Ethanol 80%
- After incubation put 96 well V bottom plate on the magnet
- Incubate 5 minutes
- Aspirate supernatant
- Add 200ul of 80% EtOH on the opposite side of the pellet
- Repeat EtOH wash
- Aspirate EtOH twice, also using individual tip
- Let the pellet dry for 5- **be careful not to let it over dry**
- Add **17.5**µl of resuspension buffer
- Move the plate away from the magnet
- Let pellet resuspend for 2 minutes without pipetting
- Mix with multichannel
- Put the plate back to magnet
- Collect 15µl of the supernatant
- Plate can be stored in -20°C freezer before Bioanalyser/Qubit
- Run Bioanalyzer and Qubit

Appendix 4: TF-1 home-brew SF buffer

Solution I (supplement) (store at -20°C):

- Dissolve 2g ATP•Disodium Salt (Sigma-Aldrich, cat. no. A3377) and 1.2g $MgCl_2$• $6H_2O$ (Sigma-Aldrich, cat. no. M2393) in 10mL nuclease-free UltraPure water
- Filter sterilize, make 20µL aliquots, and store at -20°C for up to two months.

Solution II (store at 4°C):

- Dissolve 6g KH_2PO_4 (0.044 mol) (Sigma-Aldrich, cat. no. P5655), 0.6g NaHCO3 (Sigma-Aldrich, cat. no.S5761), 0.2g glucose (Sigma-Aldrich, cat. no. G8270) in 500mL nuclease-free UltraPure water
- Adjust pH to 7.4, (3M NaOH)
- Filter-sterilize, make 1mL aliquots, and store at 4°C for up to 3 months.

Appendix 5: Lonza 4D nucleofection matrix optimisation table

← Increasing viability				Best program(s)	Increasing efficiency →			
CA-113	CA-123	CA-132	CA-139	**CA-137**	CA-158	CA-189	CA-167	CA-201
CL-138	CM-113	CM-119	CM-132	**CM-138**	CM-134	CM-167	CU-138	DG-138
CL-135	CL-137	CM-134	CM-135	**CM-137**	CM-158	CM-189	CU-137	DG-137
CL-120	CL-150	CM-116	CM-120	**CM-150**	CM-198	CM-156	CU-150	DG-150
CY-100	DA-100	DH-100	DI-100	**DN-100**	DP-100	EH-100	ER-100	FA-100
DG-138	DS-113	DS-119	DS-132	**DS-138**	DS-134	DS-167	DT-138	ED-138
DG-135	DG-137	DS-134	DS-135	**DS-137**	DS-158	DS-189	DT-137	ED-137
DG-130	DS-128	DS-107	DS-118	**DS-130**	DS-155	DS-139	DT-130	ED-130
DG-120	DG-150	DS-116	DS-120	**DS-150**	DS-198	DS-156	DT-150	ED-150
DG-120	DS-106	DS-103	DS-111	**DS-120**	DS-141	DS-142	DT-120	ED-120
DG-113	DS-112	DS-109	DS-104	**EH-100**	DS-123	DS-118	DT-113	ED-113
CM-100	CV-100	DG-100	DR-100	**EO-100**	DT-100	DU-100	ED-100	EN-100
DZ-135	DZ-137	EH-134	EH-135	**EN-138**	EH-158	EH-189	ER-137	FA-137
DZ-120	DZ-150	EH-116	EH-120	**EN-150**	EH-198	EH-156	ER-150	FA-150
DZ-113	EH-112	EH-109	EH-104	**EW-113**	EH-123	EH-118	ER-113	FA-113

Appendix 6: Initial alignment mapping rate results for 24hr post-nucleofection and 18wk post-transplant RNA-Seq samples

Table 1: 24hr post-nucleofection samples

sample.id	PE.reads.total (Million)	PE.reads.mapped (Million)	reads.mapped.Pct
LF10TOM3	27.3	25.8	94.4
LF11TOM4	27.4	26.3	96.2
LF12TOM5	30	28.8	95.8
LF13Stop4	31.7	31	97.5
LF14Stop5	28.3	27	95.2
LF1548hrcc3	30.5	28.6	93.8
LF1648hrcc4	26.8	24.1	89.9
LF1748hrcc5	27.1	25	92.4
LF18Water4	30.1	28.5	94.7
LF19Water5	29.5	27.6	93.6
LF1TOM1	27.6	26.5	96.1
LF20fresh4	29.3	26.9	92.1
LF21Stop6	14.5	12	83.1
LF22Stop7	27.2	23.9	87.9
LF23Stop8	26.3	24.4	92.9
LF24Stop9	19.7	17.1	86.5
LF25Water6	22.4	20.1	90
LF26Water7	20.7	19.5	94.4
LF27TOM6	22.8	20.8	91.6
LF2848hrcc6	19	16.5	86.6
LF29fresh7	24.5	20.5	83.6
LF2TOM2	24.3	23.2	95.3
LF30fresh8	29	27.3	94.1
LF3Stop1	28.5	27.2	95.5
LF4Stop2	32.3	30.9	95.7
LF548hrcc1	29.3	27.4	93.5
LF648hrcc2	29.4	28.2	95.9
LF7water1	32.1	30.9	96.1
LF8fresh1	28.6	26.8	93.7
LF9fresh2	26.9	26	96.5

Table 2: 18wk post-transplant samples

sample.id	PE.reads.total (Million)	PE.reads.mapped (Million)	reads.mapped.Pct
LF1_PT_TOM1_EKD	26.6	25.49	95.79
LF1_PTTO1_1_EKD	24.3	23.14	95.17
LF10_FR_Tompos1	39	36.74	94.26
LF10_PTTO4_1_EKD	24	20.6	86.01
LF11_FR_Tompos2	34.9	32.92	94.38
LF11_PTTO4_2_EKD	21.8	18.54	85.23
LF12_PT_Fresh_1_	37.9	0.22	0.57
LF12_PTFR2_1_EKD	27.7	24.59	88.73
LF13_PT_Stop_1_2	36.1	0.09	0.26
LF13_PTFR2_2_EKD	20.8	17.94	86.16
LF14_PTNT2_1_EKD	39.4	35.42	89.9
LF15_PTNT2_2_EKD	28.9	25.31	87.47
LF16_PTNT2_3_EKD	20.4	18.67	91.33
LF17_PTNT2_4_EKD	27.6	24.52	88.79
LF18_PTFR4_2_EKD	28.8	25.47	88.44
LF19_PTFR1_1_EKD	25.1	22.15	88.44
LF2_PT_TOM2_EKD	26.4	25.13	95.3
LF2_PTTO1_2_EKD	26.9	25.54	95.06
LF20_PTTO2_3_EKD	21.8	17.85	81.96
LF21_PTTO2_1_EKD	22.5	18.07	80.25
LF3_PT_TOM2_2_E	25.9	23.92	92.51
LF3_PTFR3_3_EKDL	27.5	24.61	89.53
LF4_PT_TOM3_EKD	32.2	30.39	94.22
LF4_PTNT4_2_EKD	23.8	20.37	85.5
LF5_PTNT5_1_EKD	25.6	20.85	81.56
LF6_PT_Water1_EK	29.2	26.58	90.91
LF6_PTFR3_2_EKDL	26.9	24.4	90.59
LF7_PT_Water1_EK	36.9	35.36	95.75
LF7_TP_1_EKDL21C	22.1	19.69	88.91
LF8_FR_Tomneg1_	34.6	32.9	95.12
LF8_TP_2_EKDL21C	21.8	19.38	88.76
LF9_FR_Tomneg2_	34	31.81	93.44
LF9_TN_1_EKDL21C	19	15.98	84.25

Appendix 7: Annotated list of genes contributing to the allograft rejection hallmark gene set.

HALLMARK_ALLOGRAFT_REJECTION	Gene info
AARS1	Catalyses attachment of alanine to tRNA.
ABCE1	Co-translational quality control factor, facilitates dissociation of non-functional ribosomes and degradation of damaged mRNA, plays role in quality control of mitochondrial outer membrane mRNA translation.
ABI1	Negative regulator of cell growth and transformation by interacting with non-receptor tyrosine kinases, may play role in EGF-induced ERK pathway activation. Involved in cytoskeletal re-organisation and EGFR signalling.
ACHE	Terminates signal transduction in neuromuscular junction by hydrolysing acetylcholine released in synaptic cleft.
ACVR2A	On ligand binding, forms receptor complex of type II and type I serine/threonine receptor tyrosine kinases. Activate SMAD transcriptional regulators. SMAD: main signal transducers for receptors of TGF-beta, important in growth and development.
AKT1	One of three closely related serine/threonine protein kinases, involved in metabolims, growth and survival, phosphorylate downstream targets. Invovled in glucose uptake and storage as gylcogen, important role in regualtion of NF-kappa B depdendent transcription. Mediates effects of growth factors including PDGF, EGF and IGF.
APBB1	Amyloid beta binding.
B2M	Component of MHC class I --> presentation of peptide antigens to immune system.

BCAT1	Catalyses first reaction in catabolism of essential branched amino acids e.g. Leucine
BCL10	Involved in adaptive immune response, maturation of pro-caspase 9, activation of NF-kappa B pathway, May be an adapter protein between upstream TNFR1-TRADD-RIP complex and the downstream NIK-IKK-IKAP complex. GO biological processes: cell death, NF-kappa B signalling.
BCL3	Contributes to regulation of activation of NF-kappa B gene targets. In cytoplasm, inhibits nuclear translocation of NF-kappa B p50 subunit. In nucleus, transcriptional activator for NF-kappa B target genes, involved in controlling cellular proliferation. GO biological responses: antimicrobial humoral response, DNA damage.
BRCA1	E3 ubiquitin protein ligase, essential role in DNA repair by facilitating cellular response to DNA damage, forms heterodimer with BARD1 and together involved in co-ordinating many branches of DNA repair. Makes complex with RBBP8 which controls CHEK-1 and G2/M checkpoint on DNA damage, ubiquitinates RBBP8. Acts as transcriptional activator.
C2	Component C2, part of classical complement pathway.
CAPG	Macrophage capping protein, may play role in macrophage function, may play role in cytoplasmic and nuclear structure by interaction with actin filaments, and may bind DNA.
CARTPT	Satiety factor, anorectic peptide, inhibits normal and starvation induced feeding.
CCL11	In response to allergens, promotes accumulation of eosinophils.

CCL13	Strongly chemotactic for B lymphocytes.
CCL19	Strongly chemotactic for naïve CD4 and CD8 T cells.
CCL2	Ligand for CC chemokine receptor CCR2, chemotactic for monocytes and basophils, strong chemotactic response and mobilisation of intracellular calcium ions.
CCL22	Chemotactic for activated T lymphocytes, may play role in collaboration of dendritic cells and B lymphocytes with T cells in immune responses. GO biological response: antimicrobial humoral response and cellular response to interferon gamma.
CCL4	Chemokine with inflammatory and chemo kinetic properties. GO biological responses: cellular response to drug, cellular response to interferon-gamma.
CCL5	Chemoattractant for blood monocytes, T helper cells and eosinophils. Causes release of histamine from basophils and activation of eosinophils. GO biological processes: activation of phospholipase D, cellular response to interferon-gamma.
CCL7	Chemoattractant for monocytes and eosinophils, augments monocytes anti-tumour activity. GO biological processes: cellular response to ethanol, cellular response to interferon-gamma.
CCND2	Cyclin D2.

CCND3	Regulatory component of cyclin D3-CDK4 complex, phosphorylates and inhibits RB protein family, phosphorylation allows dissociation from E2F and its release, transcription of E2F targets and progression through G_1 phase. Also substrate for SMAD3 and represses its activity. GO biological processes: cell division, mitotic cell phase transition.
CCR1	Receptor for C-C type chemokine, binds MIP-1 alpha, transduces signal through increasing intracellular calcium ions, responsible for affecting stem cell proliferation. GO biological processes: calcium ion transport, calcium mediated signalling.
CCR2	Receptor for CCL2, can also bind CCL7 and CCL12. Mediates chemotaxis and migration and monocytes and macrophages, regulates expression of T cell inflammatory cytokines and T cell migration. GO biological process: angiogenesis, apoptotic cell clearance.
CCR5	Receptor for inflammatory cytokines including CCL3/MIP1-alpha and CCL4/MIP1-beta. May play role in granulocyte lineage proliferation and migration. Plays role in T-lymphocyte migration to infection site.
CD1D	Antigen presenting protein that presents self and non-self-glycolipids. GO biological processes: antigen presentation, innate immune response.
CD2	Interacts with lymphocyte function associated antigen CD58 (LFA-3) and CD48/BCM1 to mediate adhesion with T cells and other cell types. GO biological processes: cell-cell adhesion. Positive regulator of interferon-gamma production.

CD247	Part of TCR-CD3 receptor complex present on surface of T cells, plays important role in adaptive immune response. Transmits signals when T cell receptor activated by antigen presenting cell.
CD28	Involved in T cell activation, enhances production of IL-4 and IL-10 in T cells.
CD3D	T cell receptor T3 delta chain. T cell surface glycoprotein.
CD3E	T cell surface receptor CD3 epsilon chain. Part of TCR-CD3 complex. Contain tyrosine kinase activation motifs in cytosolic domain.
CD3G	T cell receptor T3 gamma chain.
CD4	T cell surface glycoprotein, essential role in immune response against external and internal offenses. Functions as co-receptor with MHC II peptide complex. Interacts with T cell receptor and MHC II on antigen presenting cells simultaneously.
CD40	Tumour necrosis factor receptor superfamily member 5, transduces TRAF-6 and MAP3K8 signals that activate ERK in macrophages and B cells leading to induction of immunoglobulin secretion.
CD40LG	CD40 ligand, co-stimulates T cell proliferation and cytokine production. Enhances production of IL-4 and IL-10. Induces activation of NF-kappa B.
CD47	Leukocyte surface antigen, role in cell adhesion. Adhesion receptor for THBS1 on platelets. May be involved in membrane permeability changes in virus infection.
CD7	T-cell antigen CD7, membrane receptor, adaptive immune response, homeostasis.

CD74	Critical role in MHC class II antigen processing, transports complex to endoplasmic reticulum for peptide loading. Enhances stimulation of T-cell response through interaction with CD44.
CD79A	Required in cooperation with CD79B for signal transduction in B cells on binding of antigen receptor to antigen. Leads to internalisation, trafficking to late endosomes and antigen presentation.
CD80	T lymphocyte activation antigen, T cell proliferation and cytokine production stimulated by binding of CD28 or CTLA-4 to this receptor.
CD86	Receptor involved in co-stimulatory signal required for T cell proliferation and IL-2 production. Binds CD28 or CTLA-4.
CD8A	Integral membrane protein, essential role in immune response, T-cell co-receptor for MHC class I: peptide complex.
CD8B	Integral membrane protein, eseential role in immune response, T-cell co-receptor for MHC class I: peptide complex.
CD96	T cell surface protein, interactions of activated T and NK cells in late phase of immune response. Functions when T and NK cells have migrated and are engaging with diseased cells in areas of inflammation.
CDKN2A	Capable of inducing cell cycle arrest in G_1 and G_2 phases. Tumour suppressor. Binds to MDM2 and inhibits oncogenic action by preventing it from degrading TP53. Induces G2 arrest by preventing activation of cyclin B1/CDC2
CFP	Complement factor properdin

CRTAM	Cytotoxic and regulatory T cell molecule, mediates cell-cell adhesion and involved in activation, differentiation and cell retention of T cell subsets.
CSF1	Macrophage colony stimulating factor 1, plays role in survival, differentiation and proliferation of hematopoietic precursors, especially mononuclear phagocytes, promotes release of pro-inflammatory cytokines, important role in innate immunity and inflammatory processes. plays role in lipoprotein clearance cell adhesion and migration.
CSK	Non-receptor tyrosine protein kinase, involved in cell growth, differentiation, migration and immune response, phosphorylates tyrosine residues in C terminal tails of SRC family kinases.
CTSS	Cathepsin S, adaptive immune response.
CXCL13	Strongly chemotactic for B lymphocytes.
CXCL9	Cytokine that affects growth, movement or activation state of cells involved in immune response, chemotactic for activated T cells.
CXCR3	Receptor for CXCl9 CXCL10 and CXCL11.
DARS1	Catalyses of attachment of amino acid to cognate tRNA.
DEGS1	Catalyses equilibrium isomerisation of retinols.
DYRK3	Dual specificity protein kinase, promotes disassembly of membranous organelles during mitosis.
EGFR	Receptor tyrosine kinase binding ligand of EGF family, activates four major downstream signalling cascades: RAS-RAF-MEK-ERK, PI3K-AKT, PLC-gamma-PKC and STATS modules, may also activate NF-kappa B signalling cascade.

EIF3A	RNA binding component of eukaryotic translation initiation factor, required for several steps in initiation of protein synthesis.
EIF3D	mRNA cap binding protein of eukaryotic translation initiation factor 3.
EIF3J	Eukaryotic translation initiation factor 3 30kd subunit.
EIF4G3	Probable component of eukaryotic translation initiation complex 4, involving in recognition of mRNA cap, and ATP dependent unwinding of 5' secondary structure and recruitment of mRNA to ribosome.
EIF5A	mRNA binding protein involved in translation elongation. Important function at level of mRNA turnover. Involved in stress responses, functions with SDCBP as regulator of TP53 dependent apoptosis.
ELANE	Modifies function of NK cells, monocytes and granulocytes, inhibits C5a dependent neutrophil enzyme release
ELF4	Trans-activates promoters of haematopoietic growth factor genes CSF2, IL-3, IL-8, LYZ. Controls cell senescence in TP53 dependent manner.
EREG	Ligand of EGF receptor. Contributes to inflammation and tissue repair.
ETS1	Transcription factor. Directly controls expression of cytokine and chemokine genes. May control differentiation, proliferation and survival of lymphoid cells.
F2	Prothrombin, blood homeostasis inflammation and wound healing.
F2R	High affinity receptor for activated thrombin coupled to G proteins.

FAS	Receptor for tumour necrosis factor 6. Adaptor molecule FADD recruits caspase 8 to activated receptor, resulting death inducing signalling complex mediates apoptosis.
FASLG	Tumour necrosis factor ligand
FCGR2B	Fc receptor IgG
FGR	Non-receptor tyrosine protein kinase, transmits signals from cell surface receptors devoid of kinase activity, promotes release of inflammatory cytokines and mast cell degranulation.
FLNA	Actin binding protein, anchors transmembrane proteins to actin cytoskeleton.
FYB1	Adaptor protein of FYN and LCP2 signalling cascades in T cells.
GALNT1	Catalyses initial reaction in O-linked oligosaccharide biosynthesis.
GBP2	Hydrolyses GTP to GMP, promotes oxidative killing and delivers antimicrobial peptides to auto-phagolysosomes.
GCNT1	Glycosyltransferase, catalyses the transfer of an N-acetyl glucosamine moiety onto mucin-type core 1 O-glycan to form the branched mucin-type core 2 O-glycan. Mucin-type core 2 O-glycans play an important role in leukocyte extravasation as they serve as scaffolds for the display of the selectin ligand sialyl Lewis X by leukocytes.
GLMN	Regulatory component of cullin RING based SCF (SKP1-Cullin-F-box protein).
GPR65	Receptor for glycosphingolipid psychosine, role in immune response by maintaining lysosome function, and supported phagocytosis mediated intracellular bacteria clearance. May have role in activation induced cell death.
GZMA	Protease in cytosolic granules of cytotoxic T cells.

GZMB	Abundant protease in cytosolic granules of cytotoxic T cells.
HCLS1	Substrate of antigen receptor coupled tyrosine kinase. Haematopoietic lineage specification protein, plays role in antigen receptor signalling for clonal expansion and deletion and lymphoid cells.
HDAC9	Promotes de-acetylation of lysine residues on N terminal part of core histones, histone de-acetylation gives epigenetic mark for gene repression. Plays role in transcriptional regulation and cell cycle progression.
HIF1A	Transcriptional regulator of adaptive response to hypoxia, under hypoxic conditions activates transcription of over 40 genes including erythropoietin and glucose transporters, and other genes whose protein products increase oxygen delivery. Recruits transcriptional co-activators CREBP and EP300.
HLA-A	Antigen presenting MHC I
HLA-DMA	Invovled in freeing peptide binding site in MHC II for antigen loading.
HLA-DMB	Frees antigen binding site of MHC II for antigen loading.
HLA-DOA	Important regulator in HLA Class II antigen presentation, interacts with HLA-DM in B cells for peptide exchange.
HLA-DOB	Important regulator in HLA Class II antigen presentation, interacts with HLA-DM in B cells for peptide exchange.
HLA-DQA1	Binds peptides derived from antigens that enter endocytic route of antigen presentation.
HLA-DRA	Alpha chain of MHC II molecule, in complex with beta chain of HLA-DRB presents antigenic peptides on professional antigen presenting cells.

HLA-E	Non-classical MHC Class I b involved in immune self-non self-discrimination.
HLA-G	Non-classical MHC Class Ib involved in immune regulatory processes at maternal-fetal interface.
ICAM1	Ligands for leukocyte adhesion protein 1: involved in leukocyte trans endothelial migration and acute inflammatory response to antigenic stimulus.
ICOSLG	Ligand for T cell surface receptor ICOS, co-stimulatory signal for T cell proliferation and cytokine secretion, also induces B cell differentiation to plasma cells.
IFNAR2	Not well studied.
IFNG	Interferon gamma, produced by lymphocytes in response to specific antigens or mitogens, important immune-regulatory and anti-viral activity, anti-proliferative effects on transformed cells and anti-tumour effects.
IFNGR1	Receptor for interferon gamma ligand. Ligand binding activates JAK-STAT signalling pathway.
IFNGR2	Cytokine mediated, defence response against virus.
IGSF6	Immunoglobulin superfamily member 6.
IKBKB	Inhibitor of nuclear factor kappa-B kinase subunit beta, serine kinase, plays essential role in NF-kappa B signalling and activated by inflammatory cytokines, viral components, DNA damage, cellular stress. Phosphorylates inhibitors of NF-kappa B which leads to their ubiquitination and proteasomal degradation, frees NF-kappa B to activates genes involved in immune response, growth control, or protection against apoptosis.
IL10	Immune regulatory cytokine, prevents excessive inflammation.

IL11	Cytokine that stimulates proliferation of haematopoietic stem cells and megakaryocyte progenitor cells, signalling leads to phosphorylation of intracellular protein kinases and activation of STAT3.
IL12A	Cytokine that acts as growth factor for activated T and NK cells. Can stimulate production of IFN-gamma by resting peripheral blood mononuclear cells.
IL12B	Cytokine that acts as growth factor for activated T and NK cells. Can stimulate production of IFN-gamma by resting peripheral blood mononuclear cells.
IL12RB1	Interleukin receptor which binds to IL-12. Interacts with IL-23 receptor to activate JAK-STAT signalling.
IL13	Cytokine, inhibits inflammatory cytokine production, synergises with IL-2 to regulate interferon-gamma synthesis. May be critical in regulating inflammatory and immune responses.
IL15	Cytokine that stimulates proliferation of T lymphocytes. Stimulation with IL15 requires interaction with IL-2 receptor.
IL16	Stimulates migratory response in CD4 T cells, monocytes and eosinophils. Primes CD4 T cells for IL-2 and IL-15 responsiveness.
IL18	Pro-inflammatory cytokine involved in polarised Th-1 and NK immune responses.
IL18RAP	Interleukin 18 receptor accessory protein.
IL1B	Potent pro-inflammatory cytokine, induces prostaglandin synthesis, T and B cell activation, T cell cytokine and B cell antibody production.
IL2	Produced by T cells in response to antigenic or mitogenic stimulation.

IL27RA	Receptor for IL27, signalling through STAT3 and STAT1. Appears to be involved in innate defence mechanisms.
IL2RA	Alpha subunit of receptor for il-2. Involved in regualtion of immune tolerance. Helps control TREGS. TREGS suppression activation and expansion of autoreactive T cells.
IL2RB	Beta subunit of receptor Il-2. Involved in receptor internalisation and transduction of IL-2 mitogenic signals.
IL2RG	Common subunit for receptors of variety of interleukins, in associate with IL15RA stimulates neutrophil phagocytosis.
IL4	Participates in several B cell activation responses, co-stimulator of DNA synthesis. Induces secretion and expression of IgE and IgG.
IL4R	Receptor for IL-4 and IL-13, couples to JAK-1-2-3 STAT6 signalling. IL-4 response involved in stimulating Th2 differentiation.
IL6	Cytokine with variety of functions in immunity, tissue regeneration and metabolism. Invovled in acute phase, in innate immunity synthesised by myeloid cells e.g. macrophages.
IL7	Haematopoietic growth factor capable of stimulating proliferation of lymphoid progenitors, important for proliferation in various stages of B cell maturation.
IL9	Supports Il-2 and IL-4 independent growth of helper T cells.
INHBA	Activin beta a chain, cytokine activity.
INHBB	Not well studied.

IRF4	Interferon regulatory factor 4, transcriptional activator, binds to interferon stimulatory response element of MHC class I promoter, involved in CD8+ dendritic cell differentiation.
IRF7	Key transcriptional regulator of interferon type I mediated responses, plays critical role in innate immune responses to DNA and RNA viruses. Can activate interferon alpha and beta genes.
IRF8	Transcriptional regulator that binds to upstream region of type I interferon, and interferon inducible MHC class I genes, plays negative regulatory role in cells of immune system. Positively regulates macro autophagy in dendritic cells.
ITGAL	Receptor for ICAM 1, 2, 3, 4. involved in variety of immune responses including antibody dependent killing in monocytes and granulocytes. Involved in production of common lymphoid progenitor in bone marrow.
ITGB2	Receptor for ICAM 1, 2, 3, 4. Involved in leukocyte adhesion and transmigration. Plays critical role in mast cell development.
ITK	Tyrosine kinase that plays essential role in adaptive immune response. Regulates development, function and differentiation of T cells.

JAK2	Non-receptor tyrosine kinase, involved in growth development and histone modifications. Mediates essential signalling events in innate and adaptive immunity. Pivotal role in signal transduction. Phosphorylates tyrosine on cytoplasmic tails of receptors, creating docking sites for STATs, subsequently phosphorylates STATs. Phosphorylated STATs form heterodimers and translocate to nucleus where activate gene transcription.
KLRD1	Immune receptor involved in self: non-self-discrimination, recognises non classical MHC-E loaded with self-peptides.
KRT1	May regulate activity of kinases PKT and SRC.
LCK	Tyrosine protein kinase.
LCP2	Involved in T cell receptor mediated signalling.
LIF	Capacity to induce terminal differentiation in leukemic cells. Induces haematopoietic differentiation in normal and myeloid leukaemia cells.
LTB	Cytokine binds to LTBR and TNFRSF3 (tumour necrosis factor receptor). Plays role in immune response regulation.
LY75	Endocytic receptor to direct captured antigens from extracellular space to specific antigen processing compartment, causes reduced proliferation of B lymphocytes.
LY86	May interact with CD180 and TLR4 to mediate innate immune response to bacterial LPS and cytokine production.

LYN	Non receptor tyrosine kinase, transmits signals from cell surface receptors, important role in regulation of innate and adaptive immune responses, haematopoiesis, responses to growth factors and cytokines, responses to DNA damage and genotoxic agents. Functions mainly as negative regulator but can be also activator. Important in immune self-tolerance. Mediates responses to cytokines and growth factors in hematopoietic progenitors. Down regulates pathways by phosphorylation of ITIMs.
MAP3K7	Serine/threonine kinase. Essential component of MAPK signal transduction. Important role in cellular cascades in response to changes in environment. Activated by pro-inflammatory cytokines in response to various stresses included physical, chemical and oxidative.
MAP4K1	Serine/threonine protein kinase, may play a role in response to environmental stress. May play a role in haematopoietic growth regulation and lineage fate decision.
MBL2	Calcium dependent lectin involved in innate immune defence. Binds to late apoptotic cells, apoptotic blebs and necrotic cells to facilitate their uptake by macrophages.
MMP9	Matrix metalloproteinase 9.
MRPL3	Structural component of ribosome, involved in translation.
MTIF2	Translation initiation factor 2 mitochondria.
NCF4	Neutrophil cytosol factor 4, component of NADPH oxidase, responsible for oxidative burst transferring electrons from NADPH to molecular oxygen generating reactive oxygen intermediates.

NCK1	Adaptor protein that associates with tyrosine phosphorylated growth factor receptors, such as KDR, PDGFRB, plays a role in DNA damage activating downstream factors including CHEK2.
NCR1	Cytotoxicity activating receptor, GO biological process: defence response to virus, detection of virus.
NLRP3	Sensor component of NLRP3 inflammasome, crucial role in innate immunity and inflammation, responds to pathogens and damage associated signals, inflammasomes can induce pyropoptosis, and inflammation induced programmed cell death.
NME1	Major role in synthesis of nucleoside triphosphates other than ATP. Involved in cell proliferation and development, works with TREX1 in GZMA mediated cell death.
NOS2	Produces nitric oxide, tumoricidal and bactericidal action in macrophages, enhances production of pro-inflammatory mediators IL-6 and IL-8.
NPM1	Involved in ribosome biogenesis, histone assembly, cell proliferation and regulation of tumour suppressors TP53 and ARF. Suppresses apoptosis and enhances transcription of MYC target genes.
PF4	Released during platelet aggregation. Neutralises anticoagulant effect of heparin, inhibits endothelial cell proliferation.
PRF1	Plays a role in secretory granule dependent cell death, defence against virus-infected or neoplastic cells, promotes cytolysis and apoptosis of target cells by enhancing uptake of granzymyes.

PRKCB	Involved in oxidative stress induced apoptosis, mediates activation of NF-kappa B pathway. Involved in apoptosis following oxidative damage: in case of oxidative conditions, specifically phosphorylates 'Ser-36' of isoform p66Shc of SHC1, leading to mitochondrial accumulation of p66Shc, where p66Shc acts as a reactive oxygen species producer.
PRKCG	Phosphorylates TP53/TTP53 and promotes TP53/TTP53-dependent apoptosis in response to DNA damage.
PSMB10	Proteasome subunit beta type 10, involved in peptide cleavage. Plays a role in determining the T-cell repertoire for an antiviral T-cell response.
PTPN6	Modulates signalling by tyrosine phosphorylated cell surface receptors, such as KIT and EGFR, plays a key role in haematopoiesis.
PTPRC	Protein tyrosine-protein phosphatase required for T-cell activation through the antigen receptor.
RARS1	Forms part of macromolecular complex that adds amino acids to tRNAs, modulates the secretion of AIMP1 and may be involved in generation of the inflammatory cytokine EMAP2 from AIMP1.
RIPK2	Serine/threonine/tyrosine kinase, essential role in control of innate and adaptive immune response. Promotes NF-kappa B activation.
RPL39	Smallest of one of most basic proteins in liver ribosomes.
RPL3L	Structural constituent of ribosomes.
RPL9	Structural constituent of ribosomes.
RPS19	40s ribosomal subunit protein, required for pre-RNA processing and maturation of 40s ribosome.

RPS3A	40s ribosomal protein S3a, may play a role during erythropoiesis of regulation of transcription factor DDIT3.
RPS9	40s ribosomal subunit S9, RNA processing
SIT1	Negatively regulates T cell antigen receptor signalling, involved in positive selection of T cells.
SOCS1	Part of classical feedback system that regulates cytokine induced signalling, negative regulation of cytokines that signal through JAK/STAT3. Major regulator of signalling by IL-6 and leukaemia inhibitory factor (LIF), appears to be negative regulator of IFG1R signalling pathway.
SOCS5	Part of classical negative feedback pathway in controlling cytokine signalling, inhibits EGF signalling by mediating degradation of EGF receptor.
SPI1	Binds purine rich DNA, can act as lymphoid specific enhancer, transcriptional activator, involved in differentiation or activation of B cells or macrophages.
SRGN	Plays a role in formation of mast cell secretory granules. Plays a role in cytotoxic cell granule mediated apoptosis.
ST8SIA4	Polysialic acid synthase.
STAB1	Scavenger receptor for acetylated low density lipoprotein, binds to gram positive and gram negative bacteria, may play a role in host defence against bacteria.
STAT1	Activator that mediates cellular responses to interferon, phosphorylated after activation of JAK kinases, results in transcription of IFN activated genes which drive anti-viral state.
STAT4	Signal transduction and activation of transcription.

TAP1	ABC transporter associated with antigen processing, translocation of peptides from cytoplasm to ER, for loading on MHC class I.
TAP2	ABC transporter associated with antigen processing, translocation of peptides from cytoplasm to ER, for loading on MHC class I.
TAPBP	Involved in association of MHC 1 with peptide transporter TAP proteins.
TGFB1	Transforming growth factor beta 1, involved in promotion of Th17 and Treg cell differentiation.
TGFB2	Transforming growth factor beta 2 pro-protein, involved in angiogenesis and heart development.
THY1	Membrane glycoprotein, may play role in cell-cell or cell-ligand interactions.
TIMP1	Metalloproteinase inhibitor, functions as growth factor regulating cell differentiation, migration and death, activates cellular signalling cascades via CD63 and ITGB1.
TLR1	Participates in innate immune response to microbial agents, specifically recognises di and tri-acylated lipopeptides. Acts via MYD88 and TRAF6 leading to NF-kappa B activation.
TLR2	Mediates innate immune response to bacterial lipoproteins and bacterial cell walls, acts via MYD88 and TRAF6 leading to NF-kappa- B activation.
TLR3	Key component in innate and adaptive immunity, recognises molecular patterns specific to microorganisms, activated by double stranded RNA.

TLR6	Participates in innate immune response to gram positive bacteria and fungi, recognises mainly dyacylated peptides. Leads to NF kappa-B production of cytokines, CXCL1 CXCL2 CCL9.
TNF	Tumour necrosis factor cytokine, stimulates secretion of IL-1, mainly secreted by macrophages, can induce cell death of some tumour cell lines.
TPD52	Tumour protein D52, GO biological processes: B cell differentiation, positive regulator of cell population proliferation.
TRAF2	Regulates activation of NF-kappa B and JNK, central role in regulation of cell survival and apoptosis, essential component of several E3 ubiquitin ligase complexes, regulates BIRC2 and BIRC3 protein levels and mediates activation of NF kappa-B through EIF2AK2/PKR.
TRAT1	Stabilises T cell antigen receptor and CD3 on surface of T cells.
UBE2D1	Mediates selective degradation of short lived and abnormal proteins, accepts ubiquitin from E1 and catalyses its attachment to other proteins.
UBE2N	Involved in poly-ubiquitination that doesn't lead to proteasomal degradation but regulates transcriptional activation of target genes, plays a role in cell cycle and differentiation, contributes to cell survival after DNA damage, acts with E3 ligases upon genotoxic stress which is required for DNA repair. Induces expression of NF-kappa B and MAPK inflammatory genes.

WARS1	Regulates ERK, Akt, and eNOS activation pathways that are associated with angiogenesis, cytoskeletal reorganization and shear stress-responsive gene expression.
WAS	In addition to its role in the cytoplasmic cytoskeleton, also promotes actin polymerization in the nucleus, thereby regulating gene transcription and repair of damaged DNA. Promotes homologous recombination (HR) repair in response to DNA damage by promoting nuclear actin polymerization, leading to drive motility of double-strand breaks (DSBs).
ZAP70	Tyrosine kinase that plays essential role in adaptive immune response. Regulates motility adhesions and cytokine production of mature T cells. Plays a role in development and activation of primary B-lymphocytes.

Appendix 8: HSC gene sets contributing to HSC, LT-HSC and ST-HSC/Progenitor groups for HSC score RNA-Seq analysis (Ivanova, Dimos et al. 2002)

HSC/Progenitor group	mGene Name
HSC Shared	Mm.25450
HSC Shared	Inppl1
HSC Shared	AW261460
HSC Shared	Pck1
HSC Shared	Myo10
HSC Shared	Tead1
HSC Shared	Cd2bp2
HSC Shared	Efna3
HSC Shared	Eng
HSC Shared	Tbxa2r
HSC Shared	Ercc2
HSC Shared	Pkd2
HSC Shared	Rhoip3-pending
HSC Shared	AI607462

HSC Shared	4432405B04Rik
HSC Shared	C77080
HSC Shared	Pdgfrb
HSC Shared	Emcn-pending
HSC Shared	Col4a1
HSC Shared	Scd2
HSC Shared	AI563628
HSC Shared	Dgcr2
HSC Shared	Tcf4
HSC Shared	AI451722
HSC Shared	Prkacb
HSC Shared	Nfkbie
HSC Shared	Cdv1
HSC Shared	AI461869
HSC Shared	Dnmt3a
HSC Shared	2900002H16Rik
HSC Shared	Chrnb1
HSC Shared	Sycp3
HSC Shared	Kcnj6
HSC Shared	AI035334
HSC Shared	D12Bwg0831e
HSC Shared	Rps6ka4
HSC Shared	Lrrn1
HSC Shared	Ppp2r5c
HSC Shared	Sult4a1
HSC Shared	Slit2
HSC Shared	Kcnmb4
HSC Shared	Maff
HSC Shared	Pftk1
HSC Shared	Col15a1
HSC Shared	Fmr1
HSC Shared	Dctn3
HSC Shared	Gata3
HSC Shared	5730525G14Rik
HSC Shared	AI316882
HSC Shared	AW125416
HSC Shared	1110020B03Rik
HSC Shared	2610510E10Rik
HSC Shared	AI426163
HSC Shared	Pygm
HSC Shared	Gem
HSC Shared	Sdc1
HSC Shared	Mm.87296
HSC Shared	5830413E08Rik
HSC Shared	Cpxm1
HSC Shared	AI573938
HSC Shared	Mm.52043
HSC Shared	Mm.49994

HSC Shared	Mm.34159
HSC Shared	Mm.2000
HSC Shared	Mm.23609
HSC Shared	AI844617
HSC Shared	2610528B01Rik
HSC Shared	Cpsf2
HSC Shared	Jun
HSC Shared	D4Wsu132e
HSC Shared	Cct3
HSC Shared	D5Wsu31e
HSC Shared	Ppfibp2
HSC Shared	Pde4b
HSC Shared	Rasd1
HSC Shared	Nfkbia
HSC Shared	Slc23a1
HSC Shared	Madh7
HSC Shared	Tm4sf6
HSC Shared	Mm.28483
HSC Shared	Edg1
HSC Shared	Mm.27769
HSC Shared	Rest
HSC Shared	1100001I23Rik
HSC Shared	LOC192195
HSC Shared	Mm.26486
HSC Shared	9130009C22Rik
HSC Shared	2610019N19Rik
HSC Shared	AW546647
HSC Shared	AI450991
HSC Shared	2010005I16Rik
HSC Shared	Rbx1
HSC Shared	Ptprk
HSC Shared	Map4k2
HSC Shared	AW125844
HSC Shared	Scyb14
HSC Shared	AI882582
HSC Shared	1700018F16Rik
HSC Shared	Syn2
HSC Shared	Luzp1
HSC Shared	Ephb6
HSC Shared	Tgoln2
HSC Shared	Agxt
LT-HSC Shared	Tcf3
LT-HSC Shared	D10Ertd398e
LT-HSC Shared	Htf9c
LT-HSC Shared	Rbp1
LT-HSC Shared	Ocln
LT-HSC Shared	C80638
LT-HSC Shared	Mpdz

LT-HSC Shared	Gprc5b
LT-HSC Shared	Sparcl1
LT-HSC Shared	Smpd1
LT-HSC Shared	Procr
LT-HSC Shared	1110007A14Rik
LT-HSC Shared	Pclo
LT-HSC Shared	Ptpn21
LT-HSC Shared	Mm.35515
LT-HSC Shared	Agpt
LT-HSC Shared	2210402G22Rik
LT-HSC Shared	Efnb2
LT-HSC Shared	AI837850
LT-HSC Shared	Jcam2
LT-HSC Shared	Elavl4
LT-HSC Shared	Rras
LT-HSC Shared	Mm.41339
LT-HSC Shared	AI842353
LT-HSC Shared	Mm.37710
LT-HSC Shared	Mm.74203
LT-HSC Shared	Pphn
LT-HSC Shared	5730557K01Rik
LT-HSC Shared	Jcam3
LT-HSC Shared	Mm.141984
LT-HSC Shared	AA986099
LT-HSC Shared	Ches1
LT-HSC Shared	P2rx4
LT-HSC Shared	Hoxa5
LT-HSC Shared	Fkbp7
LT-HSC Shared	Mtap7
LT-HSC Shared	Aldh2
LT-HSC Shared	Nbea
LT-HSC Shared	Slc12a2
LT-HSC Shared	AI788588
LT-HSC Shared	Mm.33054
LT-HSC Shared	4931406C07Rik
LT-HSC Shared	AI836256
LT-HSC Shared	Rps4x
LT-HSC Shared	Lama5
LT-HSC Shared	Rpl7
LT-HSC Shared	Vdr
LT-HSC Shared	1200003C15Rik
LT-HSC Shared	Evi1
LT-HSC Shared	Slc2a8
LT-HSC Shared	Matn2
LT-HSC Shared	Ptprr
LT-HSC Shared	Col18a1
LT-HSC Shared	Cyp7b1
LT-HSC Shared	Mapk13

LT-HSC Shared	Vps35
LT-HSC Shared	Chst1
LT-HSC Shared	Hn1
LT-HSC Shared	Ereg
LT-HSC Shared	Hsd11b2
LT-HSC Shared	Siat8d
LT-HSC Shared	Mta1l1
LT-HSC Shared	Scya27
LT-HSC Shared	Ltbp1
LT-HSC Shared	Fmn2
LT-HSC Shared	Large
LT-HSC Shared	Rps3
LT-HSC Shared	Tuba3
LT-HSC Shared	Lactb
LT-HSC Shared	Ahsg
LT-HSC Shared	Parva
LT-HSC Shared	Eif3s7
LT-HSC Shared	Gla
LT-HSC Shared	Wdr10
LT-HSC Shared	Ak1
LT-HSC Shared	Acadvl
LT-HSC Shared	Pln
LT-HSC Shared	AI854032
LT-HSC Shared	D15Wsu77e
LT-HSC Shared	Zfp354a
LT-HSC Shared	0610039D01Rik
LT-HSC Shared	Laf4l
LT-HSC Shared	Six5
LT-HSC Shared	Hp1bp3
LT-HSC Shared	Cntn1
LT-HSC Shared	Jag2
LT-HSC Shared	Myom1
LT-HSC Shared	Mm.826
LT-HSC Shared	Nr0b1
LT-HSC Shared	Aqp4
LT-HSC Shared	Eya2
LT-HSC Shared	Mm.33003
LT-HSC Shared	Epha5
LT-HSC Shared	5730454C12Rik
LT-HSC Shared	Krt2-18
LT-HSC Shared	4930443C24Rik
LT-HSC Shared	Mm.180366
LT-HSC Shared	Mm.178672
LT-HSC Shared	Atp1b1
LT-HSC Shared	Tie1
LT-HSC Shared	Ptprz
LT-HSC Shared	Pcp4
LT-HSC Shared	1700023A18Rik

LT-HSC Shared	Adarb1
LT-HSC Shared	Psmb4
LT-HSC Shared	1700019E19Rik
LT-HSC Shared	Xrcc1
LT-HSC Shared	Bmp8a
LT-HSC Shared	4921537D05Rik
LT-HSC Shared	Hegfl
LT-HSC Shared	Mad5
LT-HSC Shared	D8Bwg1112e
LT-HSC Shared	Crim1
LT-HSC Shared	Prkcq
LT-HSC Shared	9330160G10Rik
LT-HSC Shared	5730555F13Rik
LT-HSC Shared	D4Wsu24e
LT-HSC Shared	5730406O18Rik
LT-HSC Shared	AW108467
LT-HSC Shared	Fzd4
LT-HSC Shared	Sprr2a
LT-HSC Shared	Sdccag28
LT-HSC Shared	Nnp1
LT-HSC Shared	Ndr2
LT-HSC Shared	Krt1-c29
LT-HSC Shared	Gdap1
LT-HSC Shared	2900070I05Rik
LT-HSC Shared	Tapbp
LT-HSC Shared	Blr1
LT-HSC Shared	Igf1r
LT-HSC Shared	Cldn3
ST-HSC and Early Progenitors Shared	Ppfibp1
ST-HSC and Early Progenitors Shared	Pak1
ST-HSC and Early Progenitors Shared	Gria3
ST-HSC and Early Progenitors Shared	C87208
ST-HSC and Early Progenitors Shared	AI853608
ST-HSC and Early Progenitors Shared	Impa2
ST-HSC and Early Progenitors Shared	Gna15
ST-HSC and Early Progenitors Shared	Mm.40036
ST-HSC and Early Progenitors Shared	App
ST-HSC and Early Progenitors Shared	Cutl1

ST-HSC and Early Progenitors Shared	Slc21a11
ST-HSC and Early Progenitors Shared	AU015229
ST-HSC and Early Progenitors Shared	Smarca4
ST-HSC and Early Progenitors Shared	Rppn-pending
ST-HSC and Early Progenitors Shared	Cd34
ST-HSC and Early Progenitors Shared	Trim28
ST-HSC and Early Progenitors Shared	1700017B05Rik
ST-HSC and Early Progenitors Shared	5830401B18Rik
ST-HSC and Early Progenitors Shared	Mm.132789
ST-HSC and Early Progenitors Shared	Slc25a15
ST-HSC and Early Progenitors Shared	Mm.2415
ST-HSC and Early Progenitors Shared	1500031M22Rik
ST-HSC and Early Progenitors Shared	Rabggta
ST-HSC and Early Progenitors Shared	Mm.2930
ST-HSC and Early Progenitors Shared	Snx9
ST-HSC and Early Progenitors Shared	Dtnbp1
ST-HSC and Early Progenitors Shared	D2Ertd485e
ST-HSC and Early Progenitors Shared	Dbt
ST-HSC and Early Progenitors Shared	Prkra
ST-HSC and Early Progenitors Shared	Mpp6
ST-HSC and Early Progenitors Shared	Myb
ST-HSC and Early Progenitors Shared	Mm.253
ST-HSC and Early Progenitors Shared	2310012M18Rik
ST-HSC and Early Progenitors Shared	2410030O07Rik

ST-HSC and Early Progenitors Shared	Mm.39915
ST-HSC and Early Progenitors Shared	Mrps12
ST-HSC and Early Progenitors Shared	Bckdhb
ST-HSC and Early Progenitors Shared	Wee1
ST-HSC and Early Progenitors Shared	Nup62
ST-HSC and Early Progenitors Shared	B230118H07Rik
ST-HSC and Early Progenitors Shared	Mm.86383
ST-HSC and Early Progenitors Shared	1810015P09Rik
ST-HSC and Early Progenitors Shared	Gnb4
ST-HSC and Early Progenitors Shared	Sgcb
ST-HSC and Early Progenitors Shared	AA087357
ST-HSC and Early Progenitors Shared	6530411B15Rik
ST-HSC and Early Progenitors Shared	Mm.41568
ST-HSC and Early Progenitors Shared	2610200M23Rik
ST-HSC and Early Progenitors Shared	Psmb10
ST-HSC and Early Progenitors Shared	AI841257
ST-HSC and Early Progenitors Shared	2400010G15Rik
ST-HSC and Early Progenitors Shared	Bckdha
ST-HSC and Early Progenitors Shared	Nrbf2
ST-HSC and Early Progenitors Shared	2010103A03Rik
ST-HSC and Early Progenitors Shared	Sirt3
ST-HSC and Early Progenitors Shared	Cd3d
ST-HSC and Early Progenitors Shared	Mm.100353
ST-HSC and Early Progenitors Shared	Pinx1-pending

ST-HSC and Early Progenitors Shared	AI181996
ST-HSC and Early Progenitors Shared	Slc4a8
ST-HSC and Early Progenitors Shared	Mafb
ST-HSC and Early Progenitors Shared	Prdx1
ST-HSC and Early Progenitors Shared	Spg4
ST-HSC and Early Progenitors Shared	Poli
ST-HSC and Early Progenitors Shared	Nhlh1
ST-HSC and Early Progenitors Shared	Fut7
ST-HSC and Early Progenitors Shared	Smug1
ST-HSC and Early Progenitors Shared	Mmd
ST-HSC and Early Progenitors Shared	Hnrpab
ST-HSC and Early Progenitors Shared	Mm.96329
ST-HSC and Early Progenitors Shared	B4galt4
ST-HSC and Early Progenitors Shared	AI450544
ST-HSC and Early Progenitors Shared	Mm.32184
ST-HSC and Early Progenitors Shared	2310004L02Rik
ST-HSC and Early Progenitors Shared	S100a1
ST-HSC and Early Progenitors Shared	Mm.19962
ST-HSC and Early Progenitors Shared	1700007J06Rik
ST-HSC and Early Progenitors Shared	Mm.87461
ST-HSC and Early Progenitors Shared	Mm.216667
ST-HSC and Early Progenitors Shared	5430428K15Rik
ST-HSC and Early Progenitors Shared	Mm.45653
ST-HSC and Early Progenitors Shared	4931414P19Rik

ST-HSC and Early Progenitors Shared	Lig3
ST-HSC and Early Progenitors Shared	Cst7
ST-HSC and Early Progenitors Shared	Fzd7
ST-HSC and Early Progenitors Shared	Gpr105
ST-HSC and Early Progenitors Shared	Spr
ST-HSC and Early Progenitors Shared	Fgf3
ST-HSC and Early Progenitors Shared	Pex7
ST-HSC and Early Progenitors Shared	Slc22a3
ST-HSC and Early Progenitors Shared	Atp5g2
ST-HSC and Early Progenitors Shared	4932409F11Rik
ST-HSC and Early Progenitors Shared	Mthfd1
ST-HSC and Early Progenitors Shared	Foxp3
ST-HSC and Early Progenitors Shared	2810430O08Rik
ST-HSC and Early Progenitors Shared	Mm.69853
ST-HSC and Early Progenitors Shared	3200001I04Rik
ST-HSC and Early Progenitors Shared	1110031I02Rik
ST-HSC and Early Progenitors Shared	Dapp1
ST-HSC and Early Progenitors Shared	Mm.31961
ST-HSC and Early Progenitors Shared	C85514
ST-HSC and Early Progenitors Shared	Mm.197532
ST-HSC and Early Progenitors Shared	Glb1
ST-HSC and Early Progenitors Shared	Mm.27182
ST-HSC and Early Progenitors Shared	Mm.18654
ST-HSC and Early Progenitors Shared	2610203K23Rik

ST-HSC and Early Progenitors Shared	Vipr2
ST-HSC and Early Progenitors Shared	5830438K24Rik
ST-HSC and Early Progenitors Shared	Eif3s4
ST-HSC and Early Progenitors Shared	3000002J10Rik
ST-HSC and Early Progenitors Shared	AA408566
ST-HSC and Early Progenitors Shared	Mrpl45
ST-HSC and Early Progenitors Shared	Ywhaz
ST-HSC and Early Progenitors Shared	Rpl18
ST-HSC and Early Progenitors Shared	2610029K21Rik
ST-HSC and Early Progenitors Shared	Gcs1
ST-HSC and Early Progenitors Shared	Mm.86564
ST-HSC and Early Progenitors Shared	Fnbp3
ST-HSC and Early Progenitors Shared	Rpgr
ST-HSC and Early Progenitors Shared	4930547K05Rik
ST-HSC and Early Progenitors Shared	C530002L11Rik
ST-HSC and Early Progenitors Shared	1110001J03Rik
ST-HSC and Early Progenitors Shared	D16Ertd36e
ST-HSC and Early Progenitors Shared	2610019I03Rik
ST-HSC and Early Progenitors Shared	1810047J07Rik
ST-HSC and Early Progenitors Shared	D12Ertd7e
ST-HSC and Early Progenitors Shared	Styx
ST-HSC and Early Progenitors Shared	Rps20
ST-HSC and Early Progenitors Shared	Ppp1r14b
ST-HSC and Early Progenitors Shared	Polr2e

ST-HSC and Early Progenitors Shared	Zfp386
ST-HSC and Early Progenitors Shared	Ddx9
ST-HSC and Early Progenitors Shared	AI449786
ST-HSC and Early Progenitors Shared	Ilf3
ST-HSC and Early Progenitors Shared	Mtmr4
ST-HSC and Early Progenitors Shared	Prtn3
ST-HSC and Early Progenitors Shared	4930507D05Rik
ST-HSC and Early Progenitors Shared	AI327030
ST-HSC and Early Progenitors Shared	Cbfb
ST-HSC and Early Progenitors Shared	2610302I02Rik
ST-HSC and Early Progenitors Shared	Sec61a
ST-HSC and Early Progenitors Shared	3110052N05Rik
ST-HSC and Early Progenitors Shared	1110014L17Rik
ST-HSC and Early Progenitors Shared	1110001I24Rik
ST-HSC and Early Progenitors Shared	Idh2
ST-HSC and Early Progenitors Shared	Mm.86959
ST-HSC and Early Progenitors Shared	1200013P24Rik
ST-HSC and Early Progenitors Shared	Mm.35168
ST-HSC and Early Progenitors Shared	AI463295
ST-HSC and Early Progenitors Shared	Mm.33498
ST-HSC and Early Progenitors Shared	AW742481
ST-HSC and Early Progenitors Shared	1300013C10Rik
ST-HSC and Early Progenitors Shared	Mm.35116
ST-HSC and Early Progenitors Shared	1110014C03Rik

ST-HSC and Early Progenitors Shared	Mm.35306
ST-HSC and Early Progenitors Shared	Lbp
ST-HSC and Early Progenitors Shared	Oxct
ST-HSC and Early Progenitors Shared	Nme2
ST-HSC and Early Progenitors Shared	Als2
ST-HSC and Early Progenitors Shared	Fdx1
ST-HSC and Early Progenitors Shared	H2afy
ST-HSC and Early Progenitors Shared	Gart
ST-HSC and Early Progenitors Shared	Lxn
ST-HSC and Early Progenitors Shared	Cpsf3
ST-HSC and Early Progenitors Shared	Luc7l
ST-HSC and Early Progenitors Shared	Adprt1
ST-HSC and Early Progenitors Shared	Gtf2i
ST-HSC and Early Progenitors Shared	Mm.28981
ST-HSC and Early Progenitors Shared	Grs2-pending
ST-HSC and Early Progenitors Shared	Shmt1
ST-HSC and Early Progenitors Shared	Mm.39794
ST-HSC and Early Progenitors Shared	Mm.89847
ST-HSC and Early Progenitors Shared	1110014J01Rik
ST-HSC and Early Progenitors Shared	Hkp1
ST-HSC and Early Progenitors Shared	Mm.70065
ST-HSC and Early Progenitors Shared	AI504062
ST-HSC and Early Progenitors Shared	AW229038
ST-HSC and Early Progenitors Shared	2210402P09Rik

ST-HSC and Early Progenitors Shared	AV340375
ST-HSC and Early Progenitors Shared	Fgfr1
ST-HSC and Early Progenitors Shared	Ensa
ST-HSC and Early Progenitors Shared	Irx2
ST-HSC and Early Progenitors Shared	4733401F03Rik
ST-HSC and Early Progenitors Shared	D11Bwg1104e
ST-HSC and Early Progenitors Shared	Mm.193091
ST-HSC and Early Progenitors Shared	Ercc1
ST-HSC and Early Progenitors Shared	Sf3a2
ST-HSC and Early Progenitors Shared	Dnpep
ST-HSC and Early Progenitors Shared	Mlc1
ST-HSC and Early Progenitors Shared	AI604841
ST-HSC and Early Progenitors Shared	2700053F16Rik
ST-HSC and Early Progenitors Shared	1300001I01Rik
ST-HSC and Early Progenitors Shared	Dhodh
ST-HSC and Early Progenitors Shared	0710001K01Rik
ST-HSC and Early Progenitors Shared	4733401D09Rik
ST-HSC and Early Progenitors Shared	2410004D18Rik
ST-HSC and Early Progenitors Shared	3110003A17Rik
ST-HSC and Early Progenitors Shared	2210412D01Rik
ST-HSC and Early Progenitors Shared	2410193C02Rik
ST-HSC and Early Progenitors Shared	Mm.24729
ST-HSC and Early Progenitors Shared	6230410I01Rik
ST-HSC and Early Progenitors Shared	AI648866

ST-HSC and Early Progenitors Shared	5730591C18Rik
ST-HSC and Early Progenitors Shared	2610001J05Rik
ST-HSC and Early Progenitors Shared	Csrp
ST-HSC and Early Progenitors Shared	D18Wsu181e
ST-HSC and Early Progenitors Shared	Kdel1
ST-HSC and Early Progenitors Shared	Etv6
ST-HSC and Early Progenitors Shared	Cdk4
ST-HSC and Early Progenitors Shared	2410002F23Rik