Developmentally regulated molecular programmes in human fetal haematopoietic cells cooperate with MLL-AF4 to drive infant acute lymphoblastic leukaemia



Siobhan Rice

The Queen's College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2021

This thesis is dedicated to my parents, Donna and Padraig Rice,
whose endless support and encouragement has made this possible.
Trí na chéile a thógtar na cáisléain

Acknowledgements

I would like to begin by acknowledging my supervisors, Irene Roberts, Anindita Roy and Thomas Milne. My thanks to Irene for sharing your infectious optimism and passion for discovery. Thank you to Andi for being the kind of mentor who always goes out of your way to highlight and celebrate the achievements of everyone in your lab, even when we think they're small. Thank you to both Irene and Andi for being such inspirational examples of what it is to be a scientist; creating a challenging yet supportive lab environment, fostering collaboration, and always remembering that our work should be, first and foremost, for the good of patients and their families. My thanks to Tom for making me feel an equal part of your lab, without which I would not have gained the same breadth of experience during my DPhil. Thank you to Tom for always being there to listen, and for having enough confidence in me to push me to achieve more than I ever thought I could.

My heartfelt thanks to all the members of the Roberts/Roy and Milne labs, past and present. Special thanks go to Sorcha O'Byrne, who welcomed me to the lab with open arms, shared her expertise without hesitation and became like a sister to me. Thank you to Nick Fordham for being my saviour in all things mouse model-related, and for teaching me the art of being zen in the face of experimental chaos. Thank you to Natalina Elliott for taking care of all of us; for always being there to help a friend, personally and professionally, no questions asked. Thank you to Nick Crump for never being too busy to help (even though you definitely always were too busy). Thank you to Marcela Mansur from your "Ella"; while we did not get to work together very long, you have made my final months in the lab a joy. Further thanks to Kate Murnane, Tom Jackson, Lucy Field, Rebecca Ling, Laura Godfrey, Dimitra Dimou, Ross Thorne, Joe Harman, Alastair Smith, I-Jun Lau and Nicole Jackson, who have all generously given their time to help me over the course of my DPhil.

I would like to gratefully acknowledge the wider community of the WIMM. In particularly, I would like to thank all members of the flow cytometry facility, not only for their fantastic technical help but also for being the WIMM's unofficial resident therapists during sorts. Thank you also to our collaborators on this project. Special thanks to everyone at Great Ormond Street Hospital for Children and to Prof Chris Halsey, whose work made all the difference in highlighting the translational benefit of this model. Further thanks to Prof Pablo Menendez and his lab for sharing their expertise so openly.

Finally, I would like to thank my family. Thank you to my parents, Donna and Padraig, who taught me to work hard and to dedicate my life to trying to help others. Without you supporting me, in so many ways, I would never have made it here. Thank you to my sister, Michelle, for always being my biggest cheerleader. Thank you to my partner, Jamie, for being there by my side through all of the ups and downs, for doing everything you possibly can to facilitate me following my dreams, and for filling my life with fun and adventure.

Declaration

I would like to gratefully acknowledge all collaborators who have generated data for this project. The following is a full list of data included in this thesis that was generated by collaborators, both inside and outside of the Roberts/Roy and Milne laboratories:

Figure 3.3c, 3.12a, 3.12b and 3.12e; Table 3.2: All FISH and karyotyping carried out by Dr Dariusz Ladon, Great Ormond Street Hospital for Children, London.

Figure 3.6: *MLL* intron 9 and alternative *AF4* intron 3 sgRNA sequences provided by Prof Pablo Menendez, Barcelona University.

Figure 3.11: Diagnostic flow cytometry work-up of some of the patient samples performed by Sarah Inglott, Great Ormond Street Hospital for Children, London.

Figure 3.13d: Tissue embedding, sectioning, mounting and staining carried out by the Translational Histopathology Lab (THL), CRUK Oxford Cancer Centre, University of Oxford.

Figure 3.13e; Table 3.2: CNS tissue processing, sectioning, mounting, staining, imaging and analysis carried out in the lab of Prof Christina Halsey, Wolfson Wohl Cancer Research Centre, University of Glasgow.

Table 3.2: IgH VDJ rearrangement analysis performed by Gary Wright, Great Ormond Street Hospital for Children, London.

Figure 4.2-4.4: ^{CRISPR}*MLL-AF4+* ALL RNA-seq data generated in collaboration with Dr Thomas Jackson, Roy laboratory.

Figure 4.3a: UMAP generated by Dr Thomas Jackson, Roy laboratory.

Figure 4.5c-4.5e and 5.2: Primary patient and PDX MLL-N and AF4-C ChIP-seq carried out by Dr Nicholas Crump, Milne laboratory. SEM cell ChIP-seq carried out by Dr Erica Ballabio and Dr Nicholas Crump, Milne laboratory.

Figure 4.8a, 4.9, 5.5a and 5.5b: Sorted FL HSPC RNA-seq previously generated by Dr Gillian Cowan, Dr Katerina Goudevenou and Dr Anindita Roy, Roberts and Roy laboratory.

Figure 4.10: GSEA carried out by Dr Thomas Jackson, Roy laboratory.

Figure 5.2: HEK-293T cell BioCAP and MLL CXXC domain ChIP-seq generated by Dr Jon Kerry, Milne Laboratory.

Figure 5.3: All FL HSPC sorts (ATAC-seq, sorted $^{CRISPR}MLL\text{-}AF4+$ MS-5 co-culture and sorted $^{CRISPR}MLL\text{-}AF4+$ transplants) carried out by Dr Natalina Elliott, Roberts and Roy Laboratory.

Figure 5.4a, 5.5c and 5.5d: Sorted FL HSPC ATAC-seq generated in collaboration with Dr Thomas Jackson and Dr Natalina Elliott, Roberts and Roy laboratory.

Figure 5.5c and 5.5d: Primary patient ATAC-seq generated by Dr Nicholas Crump and Dr Alastair Smith, Milne laboratory.

Abstract

Although advances in the treatment of childhood acute lymphoblastic leukaemia (ALL) mean that about 90% of children with ALL are now leukaemia-free 5 years on from their diagnosis, the prognosis for the subset of ALL that occurs within the first year of life, referred to as infant ALL, remains dismal. The most common genetic driver of infant ALL is a chromosomal translocation that leads to rearrangement of the *Mixed Lineage Leukaemia (MLL)* gene. In infant ALL, the *MLL*-rearrangement invariably arises *in utero*, and this single mutational hit in fetal life is sufficient to give rise to a leukaemia that is uniquely aggressive and treatment-refractory, even when compared to leukaemia in older children with the same *MLL*-rearrangement. The reasons for disparate outcomes in patients of different ages with identical driver mutations are unknown. Moreover, the contribution of the particular human fetal haematopoietic progenitor in which the translocation occurs remains unexplored. In this DPhil project, I aimed to address the hypothesis that fetal-specific gene expression programmes cooperate with the most common *MLL*-rearrangement in infants, *MLL-AF4*, to initiate and maintain infant ALL. I further aimed to identify the human fetal haematopoietic progenitor(s) that provide the permissive cellular context for leukaemia initiation. I found that CRISPR/Cas9 editing of primary human fetal liver (FL) cells to produce an *MLL-AF4*/t(4;11) chromosomal translocation gave rise to a leukaemia that replicated the aggressive phenotypic features of infant ALL, designated ^{CRISPR}*MLL-AF4+* ALL. Using direct comparison of the transcriptomes of human fetal and adult haematopoietic progenitor cells, I identified a fetal-specific gene signature and further showed that ^{CRISPR}*MLL-AF4+* ALL and *MLL-AF4* infant ALL patients, but not *MLL-AF4* childhood ALL patients, maintain the expression of these fetal genes. Finally, I found that, in-keeping with the B-ALL phenotype of *MLL-AF4* infant ALL, lymphoid primed progenitors provided the most permissive cellular context to give rise to ^{CRISPR}*MLL-AF4+* ALL.

Contents

List of Figures **xii**

List of Tables **xv**

List of Abbreviations **xvi**

1 Introduction **1**

1.1 *MLL*-rearranged infant acute lymphoblastic leukaemia (ALL) represents a unique subset of childhood ALL 2

1.1.1 Infant ALL 2

1.1.2 *MLL*-rearranged infant ALL 3

1.1.3 Outcomes for *MLL*-rearranged infant ALL 4

1.2 MLL fusion proteins assemble a large multi-protein complex to drive aberrant gene expression 6

1.2.1 Wild-type MLL in embryogenesis and haematopoiesis 6

1.2.2 Transcriptional regulation by the MLL fusion protein complex 9

1.3 *MLL*-rearranged infant ALL invariably arises *in utero* 12

1.4 Fetal haematopoiesis displays distinct developmentally-regulated characteristics 12

1.4.1 The haematopoietic hierarchy 12

1.4.2 Development of the haematopoietic system in fetal life . . . 14

1.4.3 Distinct characteristics of fetal HSPCs 15

1.5 Modelling *MLL*-rearranged infant ALL 18

1.5.1 Previous models of *MLL*-rearranged leukaemia 18

1.5.2 Modelling the fetal origins of *MLL*-rearranged infant leukaemia 20

1.5.3 Cell type-specific models of *MLL*-rearranged leukaemia . . . 21

1.6 Hypotheses and aims 23

2 Methods **25**

2.1 Human samples 26

2.1.1 Fetal samples 26

2.1.2 ALL patient samples 26

2.1.3 Cell lines 27

2.2 Animals . . . 27
2.3 Sample processing, cryopreservation and thawing . . . 27
2.4 Cell line maintenance . . . 28
2.5 Flow cytometry analysis and sorting . . . 28
2.6 CRISPR/Cas9 induction of the *MLL-AF4*/t(4;11) chromosomal translocation . . . 31
2.6.1 Single guide RNA (sgRNA) design . . . 31
2.6.2 Primary human FL HSPC electroporation . . . 31
2.7 MS-5 stromal co-culture assays . . . 33
2.8 NSG xenograft transplantation assays . . . 34
2.9 Histopathology . . . 35
2.10 *In vitro* drug sensitivity assays . . . 36
2.11 Plasmid cloning . . . 36
2.11.1 FLAG-tagged *MLL-AF4* plasmid cloning . . . 36
2.11.2 Bacterial transformation . . . 39
2.12 HEK-293T cell transfections . . . 39
2.13 Molecular assays . . . 40
2.13.1 DNA and RNA extraction . . . 40
2.13.2 cDNA synthesis . . . 40
2.13.3 PCR and Sanger sequencing . . . 41
2.13.4 IgH rearrangement analysis . . . 42
2.13.5 qPCR . . . 42
2.13.6 Western blotting . . . 42
2.13.7 RNA-seq . . . 43
2.13.8 Assay for Transposase-Accessible Chromatin (ATAC)-seq . . . 44
2.13.9 Chromatin Immunoprecipitation (ChIP) . . . 45
2.14 Bioinformatic analyses . . . 47
2.14.1 RNA-seq analysis . . . 47
2.14.2 ATAC-seq analysis . . . 47
2.14.3 ChIP-seq analysis . . . 48
2.15 Statistics . . . 48

3 Modelling *MLL-AF4* infant ALL using primary human FL HSPCs 49
3.1 Introduction . . . 50
3.1.1 CRISPR/Cas9 as a method for inducing chromosomal translocations . . . 50
3.1.2 The *MLL-AF4* breakpoint varies with age . . . 52
3.1.3 Human FL HSPCs as a target for leukaemic transformation 52
3.2 Aims . . . 53

3.3 A CRISPR/Cas9 editing approach to induce an *MLL-AF4* /t(4;11) translocation . . . 53
3.4 ^{CRISPR}*MLL-AF4+* FL HSPC undergo dramatic B lineage expansion *in vitro* . . . 54
3.5 ^{CRISPR}*MLL-AF4+* FL cells with an *MLL* intron 9 breakpoint do not consistently induce a B lineage expansion *in vitro* . . . 60
3.6 ^{CRISPR}*MLL-AF4+* FL HSPCs cause aggressive B-ALL *in vivo* . . . 64
3.7 CRISPR/Cas9-induced off-target editing was not detected in ^{CRISPR}*MLL-AF4+* ALL . . . 71
3.8 ^{CRISPR}*MLL-AF4+* ALL recapitulates key clinico-pathological features of *MLL-AF4* infant ALL . . . 73
3.9 Discussion . . . 78

4 ^{CRISPR}*MLL-AF4+* ALL recapitulates the molecular profile of *MLL-AF4* infant ALL 81
4.1 Introduction . . . 82
4.1.1 Distinct gene expression profiles in *MLL*-rearranged infant ALL 82
4.1.2 Fetal gene expression programmes in leukaemia . . . 83
4.2 Aims . . . 85
4.3 ^{CRISPR}*MLL-AF4+* ALL recapitulates the gene expression profile of *MLL-AF4* ALL patients . . . 86
4.3.1 Known *MLL*-rearranged leukaemia-associated genes are up-regulated in ^{CRISPR}*MLL-AF4+* ALL . . . 86
4.3.2 ^{CRISPR}*MLL-AF4+* ALL is transcriptomically similar to *MLL-AF4* ALL patients . . . 89
4.3.3 ^{CRISPR}*MLL-AF4+* ALL recapitulates the poor prognosis subset of *HOXA*^{lo} *MLL-AF4* ALL . . . 93
4.4 ^{CRISPR}*MLL-AF4+* ALL recapitulates the chromatin profile of *MLL-AF4* ALL . . . 94
4.4.1 The MLL-AF4 binding profile of ^{CRISPR}*MLL-AF4+* ALL recapitulates that of *MLL-AF4* ALL patients . . . 94
4.4.2 MLL-AF4 binding leads to increased H3K79me2 deposition during ^{CRISPR}*MLL-AF4*-induced leukaemic transformation . 98
4.5 ^{CRISPR}*MLL-AF4+* ALL specifically recapitulates the molecular profile of *MLL-AF4* infant ALL . . . 101
4.5.1 *MLL-AF4* infant ALL is transcriptomically distinct from *MLL-AF4* childhood ALL . . . 101
4.5.2 Fetal gene expression programmes drive the distinct transcriptomic profile of *MLL-AF4* infant ALL . . . 103

4.5.3 Human FL-derived ^{CRISPR}*MLL-AF4+* ALL specifically recapitulates *MLL-AF4* infant ALL . . . 106
4.6 Discussion . . . 109

5 Investigating the cellular origin of *MLL-AF4* infant ALL 113
5.1 Introduction . . . 114
5.1.1 Previous studies of the cellular origin of *MLL-AF4* infant ALL 114
5.1.2 What defines the MLL-AF4 binding profile in the earliest stages of leukaemogenesis? . . . 116
5.2 Aims . . . 118
5.3 The CXXC domain plays a major role in directing MLL-AF4 binding 118
5.3.1 Generating a FLAG-tagged *MLL-AF4* plasmid . . . 118
5.3.2 In an *MLL*wt cell line, MLL-AF4 binds accessible, unmethylated CpG regions . . . 121
5.4 The chromatin profile of ^{CRISPR}*MLL-AF4+* ALL resembles that of FL B progenitors . . . 124
5.5 ^{CRISPR}*MLL-AF4+* FL LMPP and GMP consistently give rise to a dramatic B lineage expansion *in vitro* . . . 132
5.6 ^{CRISPR}*MLL-AF4+* FL ELP and PPB give rise to ALL *in vivo* . . . 139
5.7 Discussion . . . 142

6 Discussion 145
6.1 Limitations . . . 147
6.2 Future work . . . 149

Bibliography 151

List of Figures

1.1 MLL wild-type and MLL fusion protein domains 8
1.2 Interactions between key members of the MLL fusion protein complex at target genes 11
1.3 Human fetal and post-natal B cell differentiation pathways 14
1.4 Stages of haematopoietic development 16

2.1 Biased distribution of of *MLL* breakpoint with age 32
2.2 ^{CRISPR}*MLL-AF4+* editing strategy 33

3.1 CRISPR/Cas9 function 51
3.2 Efficiency of *MLL*- and *AF4*-targeting sgRNAs 55
3.3 ^{CRISPR}*MLL-AF4+* FL cells undergo a dramatic B lineage expansion *in vitro* 56
3.4 58
3.4 Virtually all cells in ^{CRISPR}*MLL-AF4+* cultures are CD19+CD10+ preB cells 59
3.5 ^{CRISPR}*MLL-AF4+* FL cells form distinctive leukaemic cell clusters *in vitro* 60
3.6 Comparison of *MLL* intron 11 and *MLL* intron 9 breakpoints *in vitro* 61
3.7 63
3.7 ^{CRISPR}*MLL-AF4+* FL cells with an *MLL* intron 9 breakpoint cause a dramatic, mixed lineage expansion *in vitro* 64
3.8 65
3.8 ^{CRISPR}*MLL-AF4+* cells are detectable in the peripheral blood of recipient mice from week 9 post-transplant 66
3.9 ^{CRISPR}*MLL-AF4+* FL cells give rise to an aggressive B-ALL *in vivo* 67
3.10 69
3.10 ^{CRISPR}*MLL-AF4+* FL cells can give rise to both proB and preB ALL *in vivo* 70
3.11 ^{CRISPR}*MLL-AF4+* proB ALL recapitulates the immunophenotype of *MLL-AF4* infant ALL 70
3.12 No off-target CRISPR/Cas9 edits were detected in ^{CRISPR}*MLL-AF4+* ALL 72

3.13 ^{CRISPR}*MLL-AF4+* ALL recapitulates the phenotypic features of *MLL-AF4* infant ALL . . . 75
3.14 ^{CRISPR}*MLL-AF4+* ALL shows chemo-resistance *in vitro* . . . 77

4.1 . . . 87
4.1 ^{CRISPR}*MLL-AF4+* ALL sort strategy for RNA-seq . . . 88
4.2 . . . 90
4.2 ^{CRISPR}*MLL-AF4+* ALL shows upregulation of known *MLL*-rearranged leukaemia-associated genes . . . 91
4.3 . . . 92
4.3 ^{CRISPR}*MLL-AF4+* ALL recapitulates the transcriptomic profile of *MLL-AF4* ALL patients . . . 93
4.4 ^{CRISPR}*MLL-AF4+* ALL model poor prognosis *HOXA*^{lo} *MLL-AF4* ALL 94
4.5 . . . 97
4.5 ^{CRISPR}*MLL-AF4+* ALL recapitulates the chromatin profile of *MLL-AF4* ALL patients . . . 98
4.6 Increased H3K79me2 is a direct consequence of MLL-AF4 binding during ^{CRISPR}*MLL-AF4*-driven leukaemic transformation . . . 100
4.7 *MLL-AF4* infant ALL is transcriptomically distinct from *MLL-AF4* childhood ALL . . . 102
4.8 Fetal gene expression programmes drive the distinct molecular profile of *MLL-AF4* infant ALL . . . 104
4.9 Expression of fetal-specific genes in primary human HSPCs and *MLL-AF4* ALL patients . . . 106
4.10 . . . 108
4.10 FL-derived ^{CRISPR}*MLL-AF4+* ALL specifically recapitulates the molecular profile of *MLL-AF4* infant-ALL . . . 109

5.1 . . . 120
5.1 Generating a FLAG-tagged *MLL-AF4* plasmid . . . 121
5.2 . . . 123
5.2 MLL-AF4 binds accessible, unmethylated CpG regions following transfection in HEK-293T cells . . . 124
5.3 . . . 125
5.3 Sorting strategy for FL HSPC subpopulation ATAC-seq . . . 126
5.4 . . . 127
5.4 The chromatin profile of ^{CRISPR}*MLL-AF4+* ALL resembles committed FL B progenitors . . . 128
5.5 . . . 130

5.5 Transcriptomic and epigenetic analysis of *MLL-AF4* ALL based on differentially expressed genes in FL HSPC subpopulations 131
5.6 Schematic representation of edited FL HSPC subpopulation experiments 133
5.7 FL LMPP and GMP consistently transform *in vitro* 133
5.8 135
5.8 ^{CRISPR}*MLL-AF4+* FL LMPP and GMP consistently give rise to a dramatic B lineage expansion *in vitro* 136
5.9 137
5.9 Representative flow cytometry plots for FL LMPP and GMP co-cultures 138
5.10 Expression of *MLL-AF4* and *AF4-MLL* in edited FL HSPC subpopulations *in vitro* 139
5.11 141
5.11 ^{CRISPR}*MLL-AF4+* ELP and PPB give rise to leukaemic engraftment *in vivo* 142

List of Tables

2.1 HDBR staging guide. 26
2.2 Antibodies used. 30
2.3 Bulk FL CD34+ HSPC xenografts. 35
2.4 Primers used. 38
2.5 Plasmid cloning PCR master mix. 38
2.6 Standard PCR parameters. 39
2.7 InFusion cloning reaction master mix. 39
2.8 cDNA synthesis master mix. 41
2.9 cDNA synthesis parameters. 41
2.10 ATAC-sequencing indexing PCR parameters. 45

3.1 Analysis of off-target editing in ^{CRISPR}*MLL-AF4+* ALL. 73
3.2 Summary of *in vivo* experiments. 76

List of Abbreviations

AGM . . . . . Aorta-gonad-mesonephros

ALL . . . . . Acute lymphoblastic leukaemia

AML . . . . . Acute myeloid leukaemia

ATAC . . . . . Assay for Transposase-Accessible Chromatin

BM . . . . . Bone marrow

CB . . . . . Cord blood

ChIP . . . . . Chromatin Immunoprecipitation

CLP . . . . . Common lymphoid progenitor

CNS . . . . . Central nervous system

COG . . . . . Children's Oncology Group

CRISPR . . . Cluster of regularly interspaced short palindromic repeats

DSB . . . . . Double-strand break

EFS . . . . . Event-free survival

EMP . . . . . Erythro-myeloid progenitor

FBM . . . . . Fetal bone marrow

FISH . . . . . Fluorescent *in situ* hybridisation

FL . . . . . Fetal liver

FMO . . . . . Fluorescence-minus-one

hCD45+ . . . Human CD45+

HDBR . . . . Human Developmental Biology Resource

HI-FBS . . . . Heat-inactivated fetal bovine serum

HSC . . . . . Haematopoietic stem cell

HSPC . . . . . Haematopoietic stem and progenitor cell

HTA . . . . . Human Tissue Authority

indel . . . . . insertion/deletion

LMPP . . . . . Lymphoid-primed multipotent progenitor

mCD45+ . . . Mouse CD45+

mESC . . . . . Murine embryonic stem cell

***MLL*wt** . . . . MLL wild-type

MNC . . . . . Mononuclear cell

MPP . . . . . Multipotent progenitor

MRD . . . . . Minimal residual disease

NHEJ . . . . . Non-homologous end-joining

NK . . . . . Natural killer

nt . . . . . Nucleotide

PAM . . . . . Protospacer-adjacent motif

PB . . . . . ProB

PBS-EF . . . . Phosphate-buffered saline + 2% HI-FBS + 1mM EDTA

pcw . . . . . Post-conception weeks

PDX . . . . . Patient-derived xenograft

PPB . . . . . PreProB

qPCR . . . . . Quantitative polymerase chain reaction

RNP . . . . . Ribonucleoprotein

sgRNA . . . . Single guide RNA

VAF . . . . . Variant allele frequency

WBC . . . . . White blood cell

1 Introduction

Contents

1.1 *MLL*-rearranged infant acute lymphoblastic leukaemia (ALL) represents a unique subset of childhood ALL 2
1.1.1 Infant ALL 2
1.1.2 *MLL*-rearranged infant ALL 3
1.1.3 Outcomes for *MLL*-rearranged infant ALL 4
1.2 MLL fusion proteins assemble a large multi-protein complex to drive aberrant gene expression 6
1.2.1 Wild-type MLL in embryogenesis and haematopoiesis . 6
1.2.2 Transcriptional regulation by the MLL fusion protein complex 9
1.3 *MLL*-rearranged infant ALL invariably arises *in utero* 12
1.4 Fetal haematopoiesis displays distinct developmentally-regulated characteristics 12
1.4.1 The haematopoietic hierarchy 12
1.4.2 Development of the haematopoietic system in fetal life . 14
1.4.3 Distinct characteristics of fetal HSPCs 15
1.5 Modelling *MLL*-rearranged infant ALL 18
1.5.1 Previous models of *MLL*-rearranged leukaemia 18
1.5.2 Modelling the fetal origins of *MLL*-rearranged infant leukaemia 20
1.5.3 Cell type-specific models of *MLL*-rearranged leukaemia 21
1.6 Hypotheses and aims 23

1.1 *MLL*-rearranged infant acute lymphoblastic leukaemia (ALL) represents a unique subset of childhood ALL

1.1.1 Infant ALL

Acute leukaemia is the most common cancer in childhood [Ward et al., [2014]]. This can result in abnormal expansion of myeloid progenitors (acute myeloid leukaemia; AML) or lymphoid progenitors (acute lymphoblastic leukaemia; ALL) in the bone marrow (BM), with ALL being the more common in children [Ward et al., [2014]]. Remarkable progress has been made in the treatment of childhood ALL, such that about 90% of patients in high income countries now survive beyond 5 years from their diagnosis [Vora et al., [2013]]. In contrast, infant ALL, or ALL that presents within the first year of life, remains a disease with high relapse rates and a dismal prognosis. 5-year event-free survival (EFS) for infant ALL is <50% [Pieters et al., [2007], Dreyer et al., [2015], Pieters et al., [2019], Tomizawa et al., [2020]]. Hereafter, I will use "infant ALL" to refer to patients ≤1 year of age and "childhood ALL" to refer to patients >1 year of age.

Current treatment protocols for infant ALL are mainly carried out as part of large international collaborative study groups. In both the Interfant and Children's Oncology Group (COG) trials, first remission was achieved in about 90% of infant ALL patients. However, this was followed by relapse in 45-50% of patients with a median latency of 8-10 months [Pieters et al., [2007], Dreyer et al., [2015], Pieters et al., [2019]]. This suggests that relapse, possibly due to the emergence of a chemo-resistant clone, is a major driver of poor outcomes in infant ALL.

Not only does infant ALL have a worse prognosis compared to childhood ALL, but it also typically presents with more aggressive clinico-pathological features at diagnosis. These include high white blood cell (WBC) counts in the peripheral blood, hepatosplenomegaly and leukaemic infiltration into the skin and central nervous system (CNS) [Silverman, [2007], Brown, [2013], Roberts & Fordham, [2018]]. This CNS infiltration can lead to CNS relapse in about 25% of infant

ALL patients [Pieters et al., [2019]]. Immunophenotypically, infant ALL blasts are predominantly CD19+CD10- B progenitor-like cells that can aberrantly co-express myeloid cell surface markers, such as CD15 and CD65 [Borkhardt et al., [2002], Hilden et al., [2006], Jansen et al., [2007]].

1.1.2 *MLL*-rearranged infant ALL

The most common genetic driver of infant ALL is a balanced chromosomal translocation between the *Mixed Lineage Leukaemia (MLL)* gene, also known as the *Lysine Methyltransferase 2A (KMT2A)* gene, on human chromosome 11 and one of up to 120 different fusion partner genes [Meyer et al., [2018]]. These *MLL*-rearrangements account for about 70-80% of infant ALL [Hilden et al., [2006], Pieters et al., [2007]], compared to only 2-5% of childhood ALL [Moorman et al., [2010], Iacobucci & Mullighan, [2017]]. Despite the large number of identified fusion partner genes in *MLL*-rearranged leukaemia, just 9 specific fusions account for >90% of *MLL*-rearrangements [Meyer et al., [2018]]. *MLL-AF4* is the most common (49%) in infant ALL, followed by *MLL-ENL* (22%) and *MLL-AF9* (16%) [Meyer et al., [2018]]. This is in contrast to infant AML, where the most common driver genes are *MLL-AF10* (27%), *MLL-AF9* (25%) and *MLL-ELL* (15%) [Harrison et al., [2010], Creutzig et al., [2012]].

In a subset of *MLL*-rearranged leukaemia patients, the reciprocal fusion gene is also expressed. For example, about 80% of *MLL-AF4* ALL patients express the *AF4-MLL* fusion gene [Downing et al., [1994], Marschalek, [2020], Wilhelm & Marschalek, [2021]]. As the reciprocal fusion gene is not detected in all *MLL*-rearranged leukaemia patients, it is hypothesised to be dispensable for leukaemia initiation and maintenance. However, this continues to be a topic of debate in the field [Kumar et al., [2011], Sanders et al., [2011], Marschalek, [2020], Wilhelm & Marschalek, [2021]], as does the functional role of the *AF4-MLL* fusion protein in the 80% of *MLL-AF4* ALL patients that express it [Gaussmann et al., [2007], Bursen et al., [2010], Benedikt et al., [2011], Wilkinson et al., [2013], Wilhelm & Marschalek, [2021]].

Apart from the chromosomal translocation, *MLL*-rearranged infant ALL displays a largely silent mutational landscape [Bardini et al., [2010], Bardini et al., [2011], Dobbins et al., [2013], Andersson et al., [2015]]. In fact, with a somatic mutation rate of about 1.3 mutations/case, *MLL*-rearranged infant ALL has one of the lowest reported mutation frequencies of all cancers [Andersson et al., [2015]]. Moreover, *MLL*-rearranged infant ALL has a lower mutation rate than *MLL*-rearranged childhood ALL (6.5 mutations/case) [Mullighan et al., [2007], Andersson et al., [2015]]. When additional mutations are detected, they are most often found in genes encoding members of the RAS signalling pathway, such as *KRAS, NRAS, FLT3, NF1* and *PTPN11*. However, these are predominantly sub-clonal, with a variant allele frequency (VAF) of <30%, and can be lost at relapse [Liang et al., [2006], Andersson et al., [2015], Trentin et al., [2016], Agraz-Doblas et al., [2019]].

1.1.3 Outcomes for *MLL*-rearranged infant ALL

The presence of an *MLL*-rearrangement is itself an independent predictor of poor outcome in ALL, particularly in infant ALL. EFS was found to be about 66% at 3 years and 34-37% at 6 years in *MLL*-rearranged infant ALL, compared to about 93% at 3 years and 60-74% at 6 years in *MLL* wild-type (*MLL*wt) infant ALL [Hilden et al., [2006], Dreyer et al., [2015], Pieters et al., [2019], Tomizawa et al., [2020]]. However, while *MLL*-rearrangements are much more prevalent in infant ALL [Hilden et al., [2006], Pieters et al., [2007], Moorman et al., [2010], Iacobucci & Mullighan, [2017]], it is unlikely that the worse outcomes in infant ALL are solely due to the higher proportion of *MLL*-rearrangements because, even when driven by the same *MLL* fusion gene, *MLL*-rearranged infant ALL is still uniquely aggressive and treatment-refractory compared to *MLL*-rearranged childhood ALL (5-6 year EFS of 34-37% in *MLL*-rearranged infant ALL compared to 50-59% in *MLL*-rearranged childhood ALL) [Hilden et al., [2006], Mann et al., [2010], Moorman et al., [2010], Dreyer et al., [2015], Iacobucci & Mullighan, [2017], Pieters et al., [2019]]. Other important prognostic factors in *MLL*-rearranged infant ALL include high WBC counts, younger age at diagnosis (6-year EFS 20% for

patients aged 0-3 months compared to 68% for patients aged 9-12 months) and poor early response to treatment with glucocorticoids, such as prednisone [Pieters et al., [1991], Ramakers-van-Woerden et al., [2004], Pieters et al., [2019]]. Blast cells in *MLL*-rearranged infant ALL, especially *MLL-AF4* infant ALL, are more resistant *in vitro* to prednisolone (the metabolite of prednisone) compared to *MLL*wt ALL blasts [Ramakers-van-Woerden, [2004]]. This suggests that chemo-resistance contributes to poor outcomes in *MLL*-rearranged infant ALL [Hilden et al., [2006], Pieters et al., [2007], Mann et al., [2010], Pieters et al., [2019]].

Interestingly, in a significant proportion of cases, *MLL*-rearranged ALL relapsed after treatment as AML with the same *MLL*-rearrangement that was detected as diagnosis [Rossi et al., [2012], Gardner et al., [2016], Jacoby et al., [2016] Rayes et al., [2016]]. This, along with the aberrant expression of myeloid surface markers in 34-48% of infant ALL cases [Borkhardt et al., [2002]], led to the hypothesis that infant ALL has "myeloid-like" properties and the ability to switch lineage upon relapse, which might limit the effectiveness of ALL-based treatment protocols. In light of this, the most recently published Interfant-06 trial introduced a consolidation randomisation, with one arm receiving chemotherapy normally used to treat AML [Pieters et al., [2019]]. However, there was no significant difference in EFS in the more intensive "myeloid-type" treatment arm compared to the "lymphoid-type" treatment arm; or compared to the previous Interfant-99 trial [Pieters et al., [2007]].

The results from these trials highlight that outcomes for *MLL*-rearranged infant ALL have not improved significantly in the last 20 years. Moreover, these trials utilise intensive chemotherapeutic and haematopoietic transplantation strategies such that the resultant acute and long-term treatment-related morbidity is considerable. More recently, *MLL*-rearranged leukaemia research has begun to focus on the development of targeted therapies in the hope that they will more effectively eradicate the leukaemia with reduced treatment-related morbidity. However, molecules targeting proteins that were initially thought to be promising targets in this subset of leukaemia, such as DOT1L and FLT3 inhibitors, have so far shown disappointing results in clinical trials [Brown et al., [2005], Stein et al., [2018]]. Moreover, the

ability of *MLL*-rearranged infant ALL to switch to a myeloid lineage upon relapse has so far stalled any meaningful progress in the use of immunotherapies targeting B-ALL cell surface markers [Jacoby et al., [2016], Rayes et al., [2016]]. It is clear that in order to design more effective therapies for the treatment of *MLL*-rearranged infant ALL, we must better understand the mechanism by which MLL fusion proteins initiate and maintain this aggressive leukaemia.

1.2 MLL fusion proteins assemble a large multi-protein complex to drive aberrant gene expression

1.2.1 Wild-type MLL in embryogenesis and haematopoiesis

MLL was originally identified as a mammalian homologue of the *trithorax* gene in *Drosophila melanogaster* that represented a common target of chromosomal translocations in human acute leukaemia [Djabali et al., [1992], Nakamura et al., [2002], Tkachuk et al., [1992]]. Wild-type MLL plays an essential role in embryogenesis and haematopoiesis by maintaining expression of the axial patterning *HOX* gene clusters [Yu et al., [1998]]. Mice that are heterozygous for *Mll (Mll+/-)* display retarded growth, haematopoietic abnormalities and homeotic transformation due to a posterior shift in *Hox* gene expression [Yu et al., [1995]]. Mice that are homozygous null for *Mll (Mll-/-)* die at embryonic day (E)12.5-E16.5. Although *Hox* gene expression is initiated in *Mll-/-* embryos in a similar manner to wild-type controls, it is not maintained [Yu et al., [1998]]. Analysis of the haematopoietic compartment showed that *Mll-/-* mice have defects in the haematopoietic stem and progenitor cell (HSPC) pool, including a decrease in overall haematopoietic stem cell (HSC) numbers and in the quiescent HSC fraction [Jude et al., [2007], McMahon et al., [2007]]. When *Mll* was conditionally knocked out in adult mice, haematopoietic cells showed reduced colony-forming capacity *in vitro* and a dramatic reduction in their ability to compete in haematopoietic reconstitution assays *in vivo* [Jude et al., [2007], McMahon et al., [2007]].

The MLL protein is a member of a family of histone methyltransferases that plays a role in chromatin remodelling and transcription [Ballabio & Milne, [2012]]. Wild-type MLL is proteolytically cleaved by Taspase I to create 2 fragments, the N-terminal portion (MLL-N) and the C-terminal portion (MLL-C), which form a heterodimer with one another [Nakamura et al., [2002]] (Figure 1.1). MLL-N is made up of several functional domains involved in "reading" chromatin, including 3 AT hooks that bind AT-rich DNA [Zeleznik-Le et al., [1994]], a CXXC domain that binds unmethylated CpG regions of DNA [Birke et al., [2002]], 4 PHD fingers, one of which (PHD3) that recognise histone-3-lysine-4 di- and trimethylation (H3K4me2/3) (Wang et al., [2009], Wang, et al., [2010]], and an atypical bromodomain [Wang, et al., [2010]] (Figure 1.1). Proteins involved in transcription elongation, such as PAF1, have also been reported to interact with MLL-N around the CXXC domain [Milne et al., [2010], Muntean, et al., [2010]] (Figure 1.2). Milne, et al., [2010] described a minimal domain, including the CXXC domain and the third PHD finger that was capable of targeting MLL to the *HOXA9* locus, demonstrating the role of MLL-N in targeting the MLL complex to certain loci. Two particularly important proteins that bind to MLL-N are Menin and LEDGF [Yokoyama et al., [2005], Yokoyama & Cleary, [2008] (Figure 1.2). Initially believed to be involved in MLL recruitment to gene targets, Menin has since been seen to facilitate the interaction between MLL and LEDGF [Yokoyama & Cleary, [2008]], which recognises H3K36me and plays a direct role in transcription activation [Yokoyama & Cleary, [2008], Pradeepa et al., [2012], Pradeepa et al., [2014], Okuda et al., [2014]]. A recent publication provides promising preclinical data for an inhibitor that blocks the interaction between Menin and MLL in an *MLL*-rearranged leukaemia model and patient-derived xenografts (PDX) [Krivtsov et al., [2019]] and, encouragingly, early results from the AUGMENT-101 phase 1 clinical trial (NCT04065399) presented at the most recent American Society of Hematology (ASH) conference suggest that the overall response rate in patients with relapsed or refractory *MLL*-rearranged leukaemia was 54% with an undetectable minimal residual disease (MRD) rate of 67% [Stein et al., [2021]].

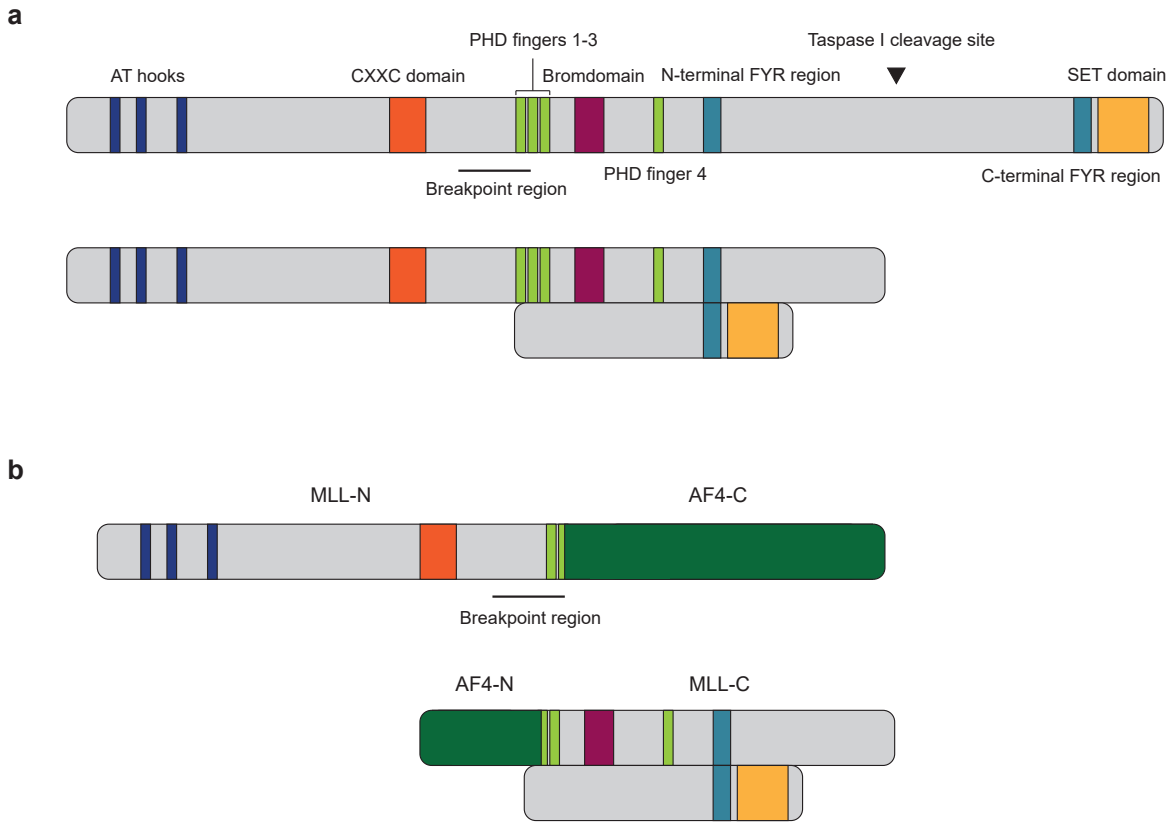


Figure 1.1: MLL wild-type and MLL fusion protein domains. a. (top) MLL protein domains before cleavage by Taspase I. Translocations have been found to occur within the labelled breakpoint region. (bottom) Cleaved MLL-N and AF4-C interacting with one another. **b.** (top) MLL-AF4 fusion protein with infant ALL breakpoint. (bottom) Reciprocal AF4-MLL protein based on infant ALL MLL-N breakpoint. (Adapted from Ballabio & Milne, [2012]).

MLL-C is involved in the chromatin modifying function of the MLL complex as it contains a SET domain at the C-terminal end that mediates deposition of H3K4me2/3 and histone-4-lysine-16 acetylation (H4K16ac) [Milne et al., [2002], Nakamura et al., [2002]] (Figure 1.1). MLL-C forms a complex with WDR5, stabilised by RBBP5, which recruits the MLL complex to H3K4me2, allowing the addition of another methyl group [Wysocka et al., [2005]]. MLL-C also interacts with MOF, which mediates H4K16ac [Dou et al., [2005]] and CBP, which acetylates H3 and is important for the transcriptional activation of a number of specific target genes [Ernst et al., [2001]].

1.2.2 Transcriptional regulation by the MLL fusion protein complex

The breakpoint of *MLL* spans an 8.3kb region [Thirman et al., [1993]], including exons which predominantly encode the intrinsically disordered region of MLL-N between the CXXC domain and the PHD fingers (Figure 1.1). Therefore, regardless of the exact *MLL* breakpoint, all MLL fusion proteins maintain aspects of the chromatin binding ability of MLL-N, through the AT hooks, CXXC domain and Menin interaction site, but they lose the PHD3 H3K4me3 binding and the H3K4 catalytic activity of the SET domain of MLL-C (Figure 1.1). Instead, this truncated MLL-N is most often fused with proteins that form part of what is known as the "super elongation complex" (SEC), including AF4, AF9 and ENL [Erfurth et al., [2004], Zeisig et al., [2005], Mueller et al., [2007], He et al., [2010], Lin et al., [2010]]. As part of the MLL fusion protein, these SEC members act as a platform to assemble a large complex of transcriptional activators and epigenetic regulators at MLL fusion protein-bound sites to drive aberrant gene expression [Yokoyama et al., [2010], Biswas et al., [2011], Takahashi & Yokoyama, [2020]] (Figure 1.2). Precisely how these complex components interact and the significance of each component's role in driving aberrant gene expression remains poorly understood. However, it appears that most, if not all, MLL fusion protein complexes utilise the AF4/ENL/P-TEFb (AEP) complex and DOT1L to some extent [Takahashi & Yokoyama, [2020]].

P-TEFb is comprised of Cyclin-Dependent Kinase 9 (CDK9) and Cyclin T1 [Marshall & Price, [1995], Yang et al., [2005], Yamada et al., [2006]]. It is recruited to target genes as part of the AEP complex [Mueller et al., [2007], Lin et al., [2010], Yokoyama et al., [2010]], or via an interaction with BRD4 [Yang et al., [2005]]. CDK9 acts to phosphorylate serine-2 of the RNA Polymerase II (RNAPII) C-terminal domain [Marshall et al., [1996]], as well as DSIF and NELF [Yamaguchi et al., [1999]], to release paused RNAPII, leading to transcription elongation. P-TEFb can be recruited to target genes as part of the MLL fusion protein complex through its direct interaction with AF4 [Mueller et al., [2007], Yokoyama et al., [2010]] (Figure 1.2), and *MLL*-rearranged cell lines have been reported to be sensitive to CDK9

inhibition [Mueller et al., [2007]], suggesting that P-TEFb may play a role in MLL fusion protein function. However, when an *MLL-AF4* fusion containing only the P-TEFb interaction domain of AF4 was transduced into mouse BM HSPCs, it was unable to drive leukaemic transformation [Yokoyama et al., [2010]]. Moreover, when *MLL-AF4* was knocked down in the SEM (*MLL-AF4+*) cell line, P-TEFb binding at key MLL-AF4 target genes was unaffected [Wilkinson et al., [2013], Benito et al., [2015]], suggesting its binding was independent of the MLL fusion protein.

DOT1L is another histone methyltransferase, which catalyses the deposition of histone-3-lysine-79 di- and trimethylation (H3K79me2/3) [Frederiks et al., [2008], Feng et al., [2010]]. This histone modification is often found in the body of active genes [Steger et al., [2008]], and it has been observed that >95% of MLL fusion protein target genes display aberrantly high levels of H3K79me2/3 [Krivtsov et al., [2008], Bernt et al., [2011]]. This high H3K79me2/3 appears to be dependent upon the presence of the MLL fusion protein complex, as Mll-AF4 [Krivtsov et al., [2008]] and MLL-ENL [Milne et al., [2005]] binding has been shown to lead to significant increases in H3K79me2/3 levels at target genes (Figure 1.2). Moreover, knockdown of *Dot1l* in a mouse model of *MLL-AF9* leukaemia led to reduced H3K79me2/3 and downregulation of some key MLL fusion protein target genes, such as *Hoxa9* and *Meis1* [Bernt et al., [2011]]. Therefore, high H3K79me2/3 is thought to be a hallmark of *MLL*-rearranged leukaemia and to contribute directly to the profound changes in gene expression that occur during leukaemogenesis. Unfortunately, a small molecule inhibitor of DOT1L (EPZ-5676) has so far not shown promising results in clinical trials for relapsed/refractory adult and childhood *MLL*-rearranged ALL, with only a transient reduction in peripheral blood and BM blasts detected in about 40% of children and no objective response observed [Stein et al., [2018]], although it is worth noting that the pharmacokinetics of EPZ-5676 may also play a role in these disappointing results.

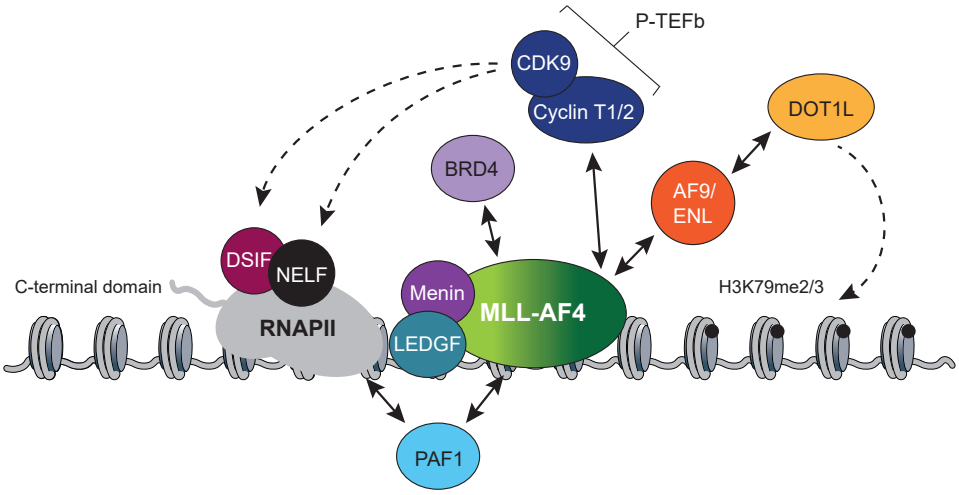


Figure 1.2: Interactions between key members of the MLL fusion protein complex at target genes. MLL-AF4 binds chromatin, predominantly at promoter regions, due to the MLL-N portion interacting with unmethylated CpG regions and Menin/LEDGF. Potential protein:protein interactions based on previous biochemical studies are shown with double-headed arrows. Catalytic activity is shown in dashed arrows: P-TEFb phosporylating the RNAPII C-terminal domain as well as DSIF and NELF, and DOT1L catalysing H3K79me2/3 deposition at the nucleosome core. (Nucleosome diagrams kindly provided by Dr Thomas Milne).

These molecular and biochemical studies suggest that the most common MLL fusion proteins assemble a similar complex of transcriptional activators and epigenetic regulators, and therefore function through similar mechanisms. However, studies modelling *MLL*-rearranged leukaemic transformation in mice indicate that MLL fusion proteins differ in their ability to transform particular haematopoietic stem and progenitor cell (HSPC) subpopulations [Drynan et al., [2005], Lobato et al., [2008]], suggesting a complex interplay between the particular *MLL*-rearrangement and the cellular context in which it first arises. In addition, different *MLL*-rearrangements show a clear lineage skew. For example, *MLL-AF4* almost invariably gives rise to ALL and, while *MLL-AF9* is capable of giving rise to both ALL and AML, there is a skew towards ALL among younger patients [Meyer et al., [2018]]. Moreover, as mentioned previously, *MLL*-rearranged infant ALL has a worse outcome than *MLL*-rearranged childhood ALL with the same *MLL* fusion gene [Biondi et al., [2000], Pui, [2000], Pui & Campana, [2007], Mann et al., [2010], Iacobucci & Mullighan,

[2017]]. Could these age-related differences be driven by the developmental stage- and site-specific context of the unique cell in which the leukaemia first arises?

1.3 *MLL*-rearranged infant ALL invariably arises *in utero*

There is clear evidence that *MLL*-rearranged infant ALL arises *in utero*. The *MLL*-rearrangement has been tracked back to its fetal origin through retrospective detection of the *MLL* fusion gene in neonatal blood spots [Ford et al., [1993], Gale et al., [1997], Hunger et al., [1998], Greaves, [2005]] and twin studies [Greaves et al., [2003]]. A characteristic and baffling feature of *MLL*-rearranged infant ALL is the fact that this single mutational "hit" before birth seems to be sufficient to induce a rapidly-proliferating, therapy-resistant leukaemia without the need for additional cooperating mutations, unlike many cases of childhood ALL, which also originate *in utero* but only develop into full-blown leukaemia after a second post-natal hit [Wiemels et al., [1999], Greaves, [2018]]. One reason for this could be that developmentally-regulated functional and molecular characteristics of fetal progenitors provide the permissive cellular context necessary to cooperate with the MLL fusion protein to induce infant ALL.

1.4 Fetal haematopoiesis displays distinct developmentally-regulated characteristics

1.4.1 The haematopoietic hierarchy

The traditional model of haematopoiesis, based on extensive immunophenotypic and functional studies carried out in both mice and humans, is a branching hierarchical structure whereby HSCs at the top of the hierarchy both self-renew and differentiate into increasingly restricted progenitor cell types to give rise to all mature haematopoietic cells [Eaves, [2015], Hofer & Rodewald, [2018]]. In this traditional model, HSCs differentiate into specific progenitor cell types with restricted lineage output, which are defined by the cell surface markers they express. More recently, mainly due

to the increased use of single-cell technologies in the study of haematopoiesis, this model has been revised. Haematopoietic differentiation is now hypothesised to be a continuous rather than a rigid step-wise process, with molecularly heterogeneous single cells making up the previously immunophenotypically-defined progenitor populations [Carrelha et al., [2018], Karamitros et al., [2018], Laurenti & Göttgens, [2018], Jacobsen & Nerlov, [2019]]. Nonetheless, whether defined by the continuum of single-cell gene expression data or by the traditional immunophenotypic progenitor types, it is accepted that HSCs first differentiate into progenitors that lose their megakaryocyte/erythroid potential, and then into progenitors that lose their myeloid potential in order to give rise to lymphoid-committed progenitors capable of giving rise to B, T and natural killer (NK) cells [Jacobsen & Nerlov, [2019]].

In the case of functional studies, the traditional immunophenotypically-defined subpopulations of HSPCs are still particularly useful to provide a "snapshot" of this continuous differentiation process [Notta et al., [2016], Laurenti & Göttgens, [2018]]. In adult BM, lymphoid commitment involves HSC differentiation into multipotent progenitor (MPP) cells, which give rise to lymphoid-primed multipotent progenitor (LMPP) cells and, further, to common lymphoid progenitor (CLP) cells [Karamitros et al., [2018], Jacobsen & Nerlov, [2019]] (Figure 1.3). Once committed to the B lineage (marked by expression of CD19 in humans and B220 in mice) [Ichii et al., [2014]] murine B progenitors are divided into 5 fractions based on their cell surface marker expression, immunoglobulin heavy chain (IgH) rearrangement and growth dependence on IL-7 [Hardy et al., [1991]].

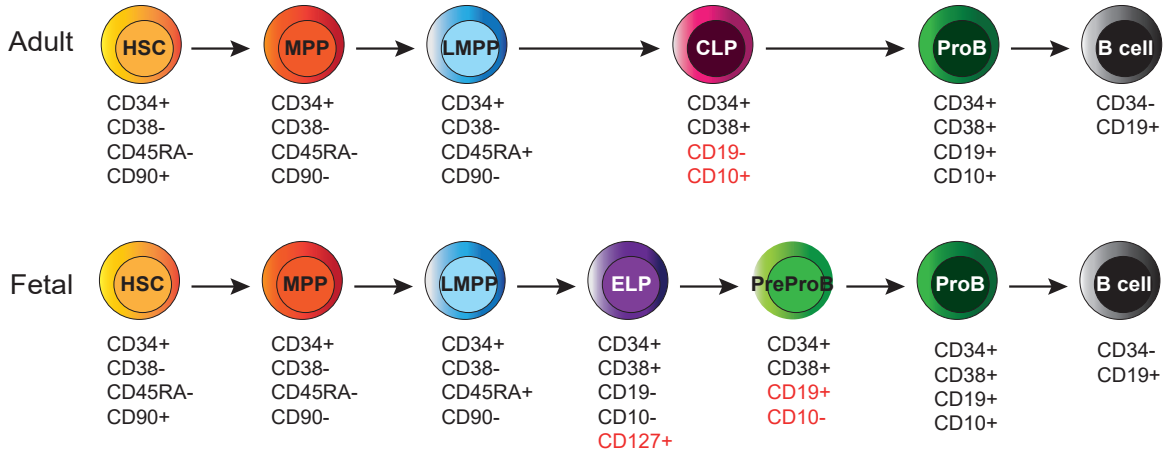


Figure 1.3: Human fetal and post-natal B cell differentiation pathways. Immunophenotypically-defined HSPC subpopulations along the B cell differentiation trajectory, including the cell surface markers used to define each subpopulation, for (top) adult BM and (bottom) fetal liver (FL) and fetal BM (FBM). Early lymphoid progenitor (ELP) and PreProB (PPB) cells represent fetal-specific HSPC subpopulations, virtually absent in adult BM [Roy et al., [2012], O'Byrne et al., [2019]]. (Adapted from a figure originally created by Dr Sorcha O'Byrne).

1.4.2 Development of the haematopoietic system in fetal life

Much of our knowledge of haematopoietic development comes from studies carried out in mice. During murine development, there are 3 waves of haematopoiesis [Dzierzak & Bigas, [2018]]. The first primitive wave involves haemogenic angioblasts giving rise to erythroid cells in the yolk sac, beginning at E7.5. This is followed by a second wave of haematopoiesis in the yolk sac at E9, giving rise to erythro-myeloid progenitor (EMP) cells [Dzierzak & Bigas, [2018]]. There is also evidence that haematopoietic cells arising from the yolk sac at this early developmental stage may have some primitive B lineage potential [Yoshimoto et al., [2011]]. Finally, the third, definitive wave of haematopoiesis begins when HSCs that immunophenotypically and functionally resemble HSCs found in adult BM emerge from the aorta-gonad-mesonephros (AGM) region at E10.5 [Medvinsky & Dzierzak, [1996], Sanchez et al., [1996], de Bruijn et al. [2002]]. These AGM-derived HSCs migrate and colonise the

fetal liver (FL), which remains the main site of haematopoiesis until birth when the BM takes over [Dzierzak & Bigas, [2018]].

In humans, primitive erythro-myelopoiesis has also been observed in the yolk sac during early development [Tavian et al., [1999], Popescu et al., [2019]], and definitive haematopoiesis begins when HSCs emerge from the AGM at 4 post-conception weeks (pcw) [Tavian et al., [1999], Tavian et al., [2001], Ivanovs et al., [2011]]. These AGM-derived HSCs colonise the FL at 5pcw [Tavian et al., [1999]], which remains a major site of haematopoiesis throughout prenatal development. In humans, haematopoietic cells are detectable in the fetal BM (FBM) from 10-12pcw [Charbord et al., [1996], O'Byrne et al., [2019], Jardine et al., [2021]]. By the time of birth, the BM has become the main site of haematopoiesis and this will remain the case throughout post-natal life (Figure 1.4).

1.4.3 Distinct characteristics of fetal HSPCs

Fetal haematopoiesis is distinct from haematopoiesis in the adult BM in a number of important ways, and these functional and molecular differences may underlie the uniquely aggressive phenotype of *MLL*-rearranged infant ALL.

Firstly, fetal HSPCs are more proliferative than adult BM and cord blood (CB) HSPCs [Lansdorp et al., [1993], Muench et al., [1994], Holyoake et al., [1999], Bowie et al., [2006]] (Figure 1.4), and they have better long-term repopulating ability in xenograft models [Rebel et al., [1996], Lepus et al., [2009]]. These differences may be driven by the fetal-specific gene expression programmes (Figure 1.4) that have been detected by comparing fetal and post-natal/neonatal HSPCs [O'Byrne et al., [2019], Copley et al., [2013], Popescu et al., [2019], Symeonidou et al., [2021]]. One well-studied example is the *Lin28b-let-7-Hmga2* axis [Copley et al., [2013]]. Expression of *Lin28b* in fetal HSPCs represses the expression of the micro-RNA *let-7*, which in turn leads to de-repression of *Hmga2* [Copley & Eaves, [2013]]. *Hmga2* has been reported to drive fetal HSPC self-renewal, as *Hmga2* knockout in FL HSPCs eliminated their proliferative advantage over adult BM HSPCs [Copley et al., [2013]]. Moreover, ectopic expression of either *Lin28b* or *Hmga2* in adult BM

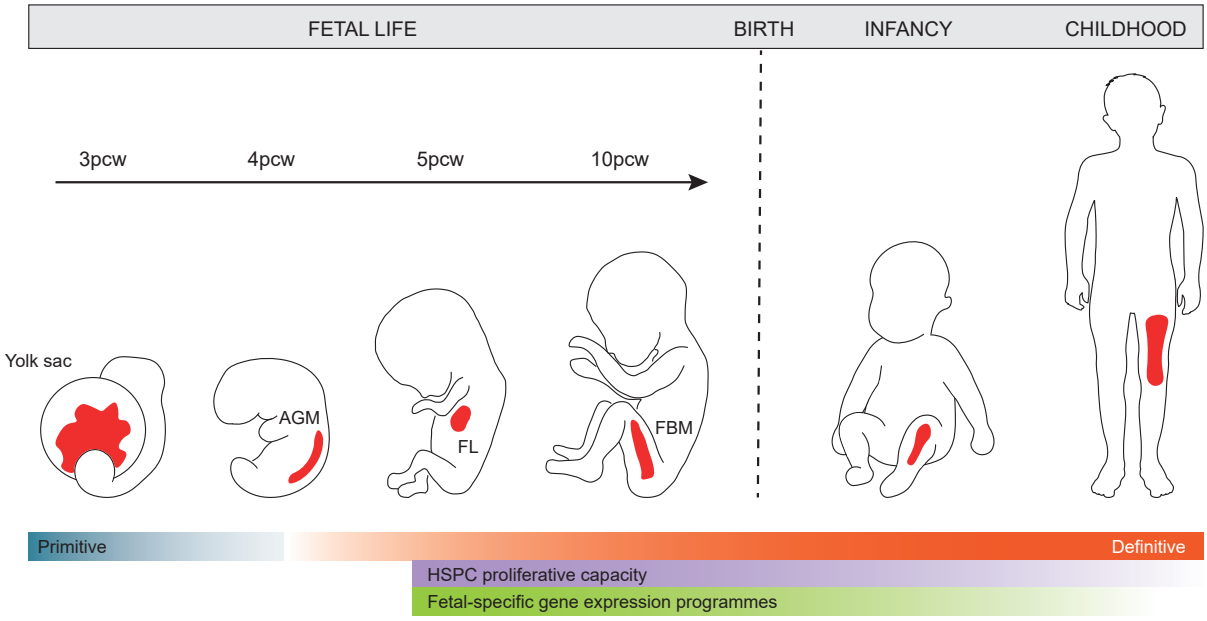


Figure 1.4: Stages of haematopoietic development. Schematic diagram showing haematopoietic development, with primitive haematopoiesis first occuring in the yolk sac from 3pcw in humans, followed by definitive haematopoiesis after HSC emergence in the AGM between 4-6pcw in humans [Tavian et al., [1999], Tavian et al., [2001]]. AGM-derived HSCs colonise the FL beginning at 5pcw and haematopoiesis is detected in the FBM from 10pcw [Charbord et al., [1996]]. At the time of birth (dashed line), FBM has become the main site of haematopoiesis. Previous studies have found that HSPC proliferative capacity is highest in FL compared to CB and adult BM, and that these fetal HSPCs have a distinct fetal-specific gene expression programme, which is lost after birth. (FL and FBM diagrams kindly provided by Dr Sorcha O'Byrne. Childhood diagram kindly provided by Dr Anindita Roy).

HSPCs recapitulated the highly proliferative phenotype of fetal HSPCs [Yuan et al., [2012], Copley et al., [2013]]. These fetal-specific gene expression programmes have also been reported to play a role in oncogenesis [Viswanathan et al., [2009], Zhou et al., [2013], Balzeau et al., [2017], Meder et al., [2018], Jackson et al., [2021]]. It is therefore possible that fetal HSPCs are intrinsically more susceptible to transformation following an *MLL*-rearrangement.

Secondly, B progenitors make up a higher proportion of the fetal BM HSPC compartment compared to adult BM [Rossi et al., [2003], O'Byrne et al., [2019]],

suggesting fetal haematopoiesis is skewed toward the lymphoid lineage. This is particularly interesting given the predominance of ALL in infant and childhood cancer. Not only this, but fetal HSPCs differentiate down a distinct developmental pathway, virtually absent in post-natal life, to give rise to B progenitors. Recent work from the Roberts and Roy laboratory and others has shed light on the distinct B lineage differentiation pathway in fetal life and its relevance to childhood ALL [Jackson et al., [2021]]. Roy et al., [2012] first reported the presence of a CD127+ (IL-7R+) early lymphoid progenitor (ELP) cell with lympho-myeloid output, as well as a CD34+CD19+CD10- PreProB (PPB) progenitor in the human FL. Further work by O'Byrne et al., [2019] ordered the distinct B lineage developmental pathway in the FBM to show that fetal ELP gain CD19 surface expression to become the first committed B progenitor in fetal life, the PPB. PPB progenitors then differentiate to become CD34+CD19+CD10+ ProB (PB) progenitors (Figure 1.3). This is in contrast with adult BM haematopoiesis, where CD10 expression is gained before CD19 during lymphoid differentiation [Eaves, [2015], Jacobsen & Nerlov, [2019]]; and in fact, these PPB progenitors appear to be a fetal-specific cell type, as they are virtually absent in adult BM [O'Byrne et al., [2019]] (Figure 1.3). This is particularly interesting as fetal PPB cells bear striking immunophenotypic and molecular similarities to infant ALL blasts. For example, although they are functionally committed to the B lineage, PPB cells maintain chromatin accessibility at several myeloid genes including *MPO, CSF1R* and *CD7*, as well as stem cell- and leukaemia-associated genes like *LIN28B* [O'Byrne et al., [2019]], which are often found to be upregulated in infant ALL blasts and may, in part, underlie their ability to switch to the myeloid lineage upon relapse. However, while these similarities make PPB progenitors an interesting potential target cell type in *MLL*-rearranged infant ALL, the exact cell(s) of origin of *MLL-AF4* infant ALL, for example, remain unknown.

1.5 Modelling *MLL*-rearranged infant ALL

While the phenotypic and molecular features of *MLL*-rearranged leukaemia at diagnosis are well described, little is known about the early stages of leukaemogenesis driven by the MLL fusion protein. Even less is known about the extent to which the fetal-specific and target cell-type-specific context in which *MLL*-rearranged infant leukaemia arises contributes to its aggressive phenotype and poor clinical outcome. In an attempt to determine the factors that are sufficient for *MLL*-rearranged leukaemogenesis, a variety of mouse models of the most common *MLL*-rearranged leukaemias have been generated over the years [reviewed in Milne et al., [2017]]. However, while mouse models of certain *MLL* fusion genes like *MLL-AF9*, which is common in both infant and childhood AML, are capable of recapitulating myeloid disease phenotypes observed in patients, attempts to model *MLL-AF4* ALL have been less successful.

1.5.1 Previous models of *MLL*-rearranged leukaemia

One of the first models of *MLL*-rearranged leukaemia involved insertion of the *AF9* sequence into the endogenous *Mll* locus in mouse embryonic stem cells (mESC), and injection of these cells into a blastocyst to create a chimeric mouse [Corral et al., [1996]]. One disadvantage of this approach is that *Mll-AF9* was expressed in other cell types outside of the haematopoietic system, which could lead to other developmental defects. However, after a myeloproliferative phase, these mice ultimately succumb to AML [Corral et al., [1996]]. To target the haematopoietic system more specifically, retroviral transduction has been used to insert *MLL-ENL* and a chimeric human/mouse *MLL-Af4* fusion gene into mouse BM HSPCs [Lavau et al., [1997], Lin et al., [2016]]. In both cases, mice transplanted with these transduced cells developed AML. Interestingly, when *MLL-ENL*-transduced cells were first cultured in B-lineage-promoting culture conditions, they were capable of giving rise to ALL in recipient mice [Zeisig et al., [2003]], suggesting some lineage plasticity and a role for the microenvironment in driving the leukaemia phenotype.

Both *MLL-AF9* and *MLL-ENL* are capable of giving rise to ALL in patients, and *MLL-AF4* almost invariably gives rise to ALL [Meyer et al., [2018]], and yet these early mouse models predominantly succumb to AML. It was hypothesised that a human cell context may be required for an *MLL*-rearrangement to give rise to ALL. Lentiviral transduction of *MLL-AF9* into human CB HSPCs mainly gave rise to AML; however, it was also capable of giving rise to some B lineage ALL in recipient mice [Barabe et al., [2007], Wei et al., [2008]]. Transduction of *MLL-ENL* [Barabe et al., [2007]] and *MLL-Af4* [Lin et al., [2016]] into human CB HSPCs gave rise predominantly to ALL but, similarly to the earlier *MLL-ENL* mouse model, these cells showed a certain level of lineage plasticity. When cultured in myeloid lineage-promoting culture conditions before transplant, both *MLL-ENL* and *MLL-Af4*-transduced CB cells could give rise to AML in recipient mice [Barabe et al., [2007], Lin et al., [2017]].

While it has proven to be an effective method of expressing *MLL* fusion genes in both mouse and human HSPCs, lentiviral transduction cannot model certain key molecular characteristics of *MLL*-rearranged leukaemia. For example, expression of the *MLL* fusion as a transgene often results in higher expression levels compared to *MLL* fusion genes in patients [Chen et al., [2008]]. Moreover, both wild-type alleles of *MLL* and the fusion partner gene are intact and expressed in transduced cells. Finally, the consequences of the significant structural change caused by the chromosomal translocation on chromatin cannot be explored. To model a bona fide *MLL*-rearrangement, the Rabbitts laboratory developed an elegant "translocator" model that utilised the *Cre-loxP* system to conditionally induce an *Mll-Af9* or *Mll-Enl* chromosomal translocation in mouse HSPCs [Drynan et al., [2005], Lobato et al., [2008]]. Due to the orientation of *Mll* and *Af4* in the mouse genome, they were unable to make an *Mll-Af4* translocator model [Lobato et al., [2008]]. More recently, CRISPR/Cas9 genome editing has been used to simultaneously create double-strand breaks at the *MLL* and *ENL, AF9* or *AF4* loci to induce a translocation in a subset of edited cells through non-homologous end-joining (NHEJ) [Reimer et al., [2017], Secker et al., [2019]]. CRISPR/Cas9-induced *MLL-ENL* in human

CB HSPCs gave rise to a monocytic leukaemia-like expansion in recipient mice [Reimer et al., [2017]]. CRISPR/Cas9-induced *MLL-AF9* and *MLL-AF4* in human CB HSPCs gave rise to a myeloid expansion *in vitro* but leukaemic potential *in vivo* was not analysed [Secker et al., [2019]].

1.5.2 Modelling the fetal origins of *MLL*-rearranged infant leukaemia

The majority of mouse models that consider the role of the age of the initiating cell have used *MLL-AF9*, as it is known to give rise to either AML or ALL but with a clear skew towards the myeloid lineage in infants and children compared to adults. When *MLL-AF9* was transduced into adult mouse BM HSPCs, it gave rise only to AML in recipient mice [Somervaille & Cleary, [2006], Krivtsov et al., [2013]]. In contrast, when *MLL-AF9* expression was induced in mouse FL HSPCs, it was capable of giving rise to B-ALL, as well as AML and leukaemias with an undifferentiated immunophenotype, though it showed a long latency [Chen et al., [2011]]. In similar experiments using human cells, lentiviral transduction was used to directly compare the outcome of *MLL-AF9* expression in adult BM HSPCs and CB HSPCs. While *MLL-AF9* expression in adult BM HSPCs gave rise to non-leukaemic myeloid-biased engraftment in recipient mice [Horton et al., [2013]], CB HSPCs expressing *MLL-AF9* gave rise to ALL, as well as some AML and mixed lineage disease [Barabe et al., [2007], Horton et al., [2013]]. Finally, a model in which *Mll-AF4* was conditionally expressed from the endogenous locus in mouse fetal HSPCs generated a B-cell lymphoma from mid-gestation (E12-14) FL HSPCs, albeit with a long latency [Barrett et al., [2016]]. Interestingly, a more recent study from the same group has shown that overexpression of miR130b or miR128a, microRNAs upregulated in infant and childhood *MLL-AF4* ALL blasts, in mouse fetal HSPC-derived *Mll-AF4+* cells resulted in proB, myeloid or mixed phenotype acute leukaemia that recapitulated the CNS infiltration of *MLL*-rearranged infant leukaemia [Malouf et al., [2021]].

Others have hypothesised that it is the fetal/neonatal BM niche that drives the uniquely aggressive and lymphoid-biased phenotype of *MLL*-rearranged infant leukaemia. Rowe et al., [2019] transduced adult mouse BM HSPCs with *MLL-AF9* and transplanted these cells into either adult or neonatal recipient mice. They found that while *MLL-AF9* gave rise exclusively to AML in adult recipients, the neonatal recipients showed a mixed leukaemia phenotype, with a small proportion of B220+ B progenitor cells. While neonatal recipient mice were clearly more permissive to a mixed leukaemia phenotype, *MLL-AF9* expression in adult BM HSPCs still gave rise to a predominantly myeloid leukaemia in both cases. This would suggest that, at least in mice, the influence of the niche might be less important than intrinsic properties of fetal HSPCs.

Taken together these results suggest that fetal and neonatal HSPCs in the presence of a neonatal haematopoietic microenvironment are more likely to give rise to lymphoid leukaemias upon expression of an *MLL* fusion gene. Thus, the lymphoid bias of normal fetal haematopoiesis may well be a key factor in the predominance of ALL among *MLL*-rearranged infant leukaemia patients. Importantly, it appears that human cells, even when derived from CB, more closely recapitulate *MLL*-rearranged infant leukaemia phenotypes upon expression of an *MLL* fusion gene. It is important to note that so far, none of these models have used human fetal cells transformed by a human MLL fusion protein; and better understanding of the importance of the fetal context for *MLL*-rearranged leukaemogenesis in patients is likely to require models derived from human fetal HSPCs and/or niche.

1.5.3 Cell type-specific models of *MLL*-rearranged leukaemia

In addition to a human fetal cell context, it has been hypothesised that the particular haematopoietic progenitor cell type in which the *MLL*-rearrangement first arises may also affect the likelihood of leukaemogenesis and the phenotype of the resultant leukaemia. The *Mll-Af9* and *Mll-Enl* translocator models developed by the Rabbitts laboratory conditionally induced *Cre* expression under the control of promoters for several lineage-specific genes. In this way, they could examine the effect of the

MLL-rearrangement arising in specific progenitor cell types. When the *Mll-Af9* translocation was induced under the control of *Lmo2-Cre* (expressed in HSCs and other early progenitors), it gave rise to AML. However, when it was induced under the control of *Lck-Cre* (expressed in T progenitors), it did not give rise to leukaemia. In contrast, when the *Mll-Enl* translocation was induced under the control of *Lck-Cre*, it gave rise to both AML and ALL [Drynan et al., [2005], Lobato et al., [2008]]. While it was not possible to generate an *Mll-Af4* translocator model, the Rabbitts laboratory did generate a conditional *Mll-AF4* "invertor" model, which utilized the *Cre-loxP* system to conditionally induce *Mll-AF4* expression from the endogenous *Mll* locus under the control of lineage-specific promoters [Metzler et al., [2006]]. When *Mll-AF4* expression was induced under the control of *Lmo2-Cre*, it led to embryonic lethality. When its expression was induced under the control of *Lck-Cre*, or *Rag-Cre* or *Cd19-Cre* (both expressed in B progenitors), it gave rise to B cell lymphoma [Metzler et al. [2006]].

Lentiviral transduction has been used to introduce *MLL-AF9* and *MLL-ENL* into purified mouse HSPC subpopulations. These fusion genes were both found to be capable of transforming Lineage-Sca1+c-kit+ (LSK) early progenitor cells, common myeloid progenitor (CMP) cells and granulocyte/macrophage progenitor (GMP) cells to give rise to AML [Cozzio et al., [2003], Krivtsov et al., [2006], Somervaille & Cleary, [2006]]. However, LSK-derived *MLL-AF9* AML appeared to be more aggressive [Krivtsov et al., [2013]]. Similarly, using a doxycycline-inducible expression system, *MLL-AF9* expression in mouse long-term HSCs (CD34-CD150+CD48- LSKs) produced a highly aggressive AML with a shorter latency and a gene expression profile similar to patients with a poor prognosis compared to GMP-derived *MLL-AF9* AML [Stavropoulou et al., [2016]]. Interestingly, a doxycycline-inducible *MLL-ENL* model, which could induce AML when expressed in T progenitors, appeared to be incapable of giving rise to leukaemia when expressed in mouse HSCs [Ugale et al., [2014], Ugale et al., [2017]].

Taken together, these results highlight two key points. Firstly, they suggest that the haematopoietic progenitor cell type in which the *MLL*-rearrangement arises

plays an important role in leukaemogenesis. In these mouse studies, some progenitor cell types are incapable of giving rise to leukaemia, even when *MLL* fusion gene expression can be detected in these cells, which would suggest that a permissive cell context is required to cooperate with the MLL fusion protein to drive leukaemia. Moreover, the same MLL fusion protein, such as MLL-AF9 for example, appears to give rise to leukaemia with distinct characteristics depending on the progenitor cell type in which it is first expressed. Secondly, they suggest that the permissive cell context required to cooperate with the MLL fusion protein differs between different *MLL*-rearrangements. However, as MLL fusion proteins appear to assemble very similar multi-protein complexes to one another, the mechanism underlying these differences remains unknown. So far, target cell type-specific *MLL*-rearranged leukaemia models have not been generated from primary human HSPCs and, even in mice, studying the role of the cell-of-origin in *MLL-AF4* infant ALL has been hindered by the overall lack of an accurate human model of the disease.

1.6 Hypotheses and aims

The main aim of this project was to investigate how the human fetal haematopoietic cell context in which *MLL-AF4* infant ALL, the most common infant ALL, first arises may cooperate with the MLL fusion protein to give rise to the uniquely aggressive phenotype of infant ALL and, in so doing, generate the first bona fide model of human fetal HSPC-derived *MLL-AF4* infant ALL. I have divided this aim into 3 questions that I set out to answer:

1. Can induction of an *MLL-AF4*/t(4;11) chromosomal translocation in primary human FL HSPCs give rise to a leukaemia that recapitulates the aggressive phenotypic features of *MLL-AF4* infant ALL? I addressed this by:
    - developing a method to induce an *MLL-AF4*/t(4;11) translocation in primary human FL HSPCs, designated ^{CRISPR}*MLL-AF4+*
    - characterising ^{CRISPR}*MLL-AF4+* FL HSPCs *in vitro*

- characterising the infant ALL-like phenotypic features of ^{CRISPR}*MLL-AF4+* ALL in xenografted recipient mice

2. Does human FL-derived ^{CRISPR}*MLL-AF4+* ALL recapitulate the molecular characteristics of poor prognosis *MLL-AF4* infant ALL? I addressed this by:

   - comparing the gene expression and chromatin profile of ^{CRISPR}*MLL-AF4+* ALL to primary *MLL-AF4* ALL patients
   - deriving an *MLL-AF4* infant ALL-specific and human fetal HSPC-specific gene expression signature, using primary patient/HSPC RNA-seq data
   - analysing the maintenance of *MLL-AF4* infant ALL-specific and fetal-specific gene expression programmes in ^{CRISPR}*MLL-AF4+* ALL compared to a previous, CB-derived *MLL-Af4* ALL model [Lin et al., [2016]]

3. Which human FL HSPC subpopulation provides a permissive cell context to cooperate with *MLL-AF4* to give rise to infant ALL? I addressed this by:

   - expressing *MLL-AF4* in an *MLL*wt cell line to determine the key factors that direct MLL-AF4 binding in the earliest stages of leukaemogenesis
   - comparing the chromatin environment of sorted subpopulations of FL HSPCs
   - assessing ^{CRISPR}*MLL-AF4*-induced transformation in sorted subpopulations of FL HSPCs *in vitro*
   - assessing ^{CRISPR}*MLL-AF4*-induced ALL in sorted subpopulations of FL HSPCs *in vivo*

2 Methods

Contents

2.1 Human samples 26
2.1.1 Fetal samples 26
2.1.2 ALL patient samples 26
2.1.3 Cell lines 27
2.2 Animals 27
2.3 Sample processing, cryopreservation and thawing . . . 27
2.4 Cell line maintenance 28
2.5 Flow cytometry analysis and sorting 28
2.6 CRISPR/Cas9 induction of the *MLL-AF4* /t(4;11) chromosomal translocation 31
2.6.1 Single guide RNA (sgRNA) design 31
2.6.2 Primary human FL HSPC electroporation 31
2.7 MS-5 stromal co-culture assays 33
2.8 NSG xenograft transplantation assays 34
2.9 Histopathology 35
2.10 *In vitro* drug sensitivity assays 36
2.11 Plasmid cloning 36
2.11.1 FLAG-tagged *MLL-AF4* plasmid cloning 36
2.11.2 Bacterial transformation 39
2.12 HEK-293T cell transfections 39
2.13 Molecular assays 40
2.13.1 DNA and RNA extraction 40
2.13.2 cDNA synthesis 40
2.13.3 PCR and Sanger sequencing 41
2.13.4 IgH rearrangement analysis 42
2.13.5 qPCR 42
2.13.6 Western blotting 42
2.13.7 RNA-seq 43

2.13.8 Assay for Transposase-Accessible Chromatin (ATAC)-seq 44
2.13.9 Chromatin Immunoprecipitation (ChIP) 45
2.14 Bioinformatic analyses 47
2.14.1 RNA-seq analysis 47
2.14.2 ATAC-seq analysis 47
2.14.3 ChIP-seq analysis 48
2.15 Statistics 48

2.1 Human samples

2.1.1 Fetal samples

Fetal tissues were provided by the Human Developmental Biology Resource (HDBR, `www.hdbr.org`), regulated by the Human Tissue Authority (HTA, `www.hta.gov.uk`) and covered under ethics (REC: 18/NE/2090 and 18/LO/0822). The tissues were collected during elective termination of pregnancy after obtaining written informed consent for the collection and use of fetal tissue in accordance with the Declaration of Helsinki and staged [Hern, [1984]] (Table 2.1) by the HDBR.

Table 2.1: HDBR staging guide.

Post-conception weeks (pcw)	Foot length (mm)	Knee – heel length (mm)	Gestational age in weeks
Late 8	Above 5 and less than 6	8	10
9	6 or above and less than 8	11	11
10	8 or above and less than 10	13	12
11	10 or above and less than 13	17	13
12	13 or above and less than 17	24	14
13	17 or above and less than 20	31	15
14	20 or above and less than 23	36	16
15	23 or above and less than 25	41	17
16	25 or above and less than 27	43	18
17	27 or above and less than 31	51	19
18	31 or above and less than 33	56	20
19	33 or above and less than 36	60	21
20	36 or above and less than 40	66	22

2.1.2 ALL patient samples

Cryopreserved *MLL-AF4* infant and childhood ALL diagnostic samples used for immunophenotypic and molecular assays were obtained from Blood Cancer UK Childhood Leukaemia CellBank (REC: 16/SW/0219). Patient samples were anonymised at source, assigned a unique study number and linked.

Cryopreserved *MLL*-rearranged and *ETV6-RUNX1+* childhood ALL PDX samples used for *in vitro* drug sensitivity assays were generously provided by Prof Christina Halsey, Wolfson Wohl Cancer Research Centre, University of Glasgow.

Additional immunophenotypic data generated by Sarah Inglott, Great Ormond Street Hospital for Children, London: *MLL*-rearranged infant and childhood ALL samples from patients being treated at Great Ormond Street Hospital for Children, London were analysed by flow cytometry as part of their diagnostic work-up as standard of care; and the immunophenotypic data was shared with us in an anonymised manner as part of an ongoing collaboration.

2.1.3 Cell lines

SEM (*MLL-AF4+* B-ALL cell line) and KOPN8 (*MLL-ENL+* B-ALL cell line) cells were purchased from DSMZ (`www.cell-lines.de`). HEK-293T (human embryonic kidney cell line) cells were purchased from ATCC (`www.atcc.org`). MS-5 murine stromal cells were generously provided by Professor Paresh Vyas, MRC Weatherall Institute of Molecular Medicine, University of Oxford.

2.2 Animals

NOD.Cg-Prkdc^{scid}Il2rg^{tm1Wjl}/SzJ (NSG) mice were purchased from Charles River Laboratories (`www.criver.com`). All experiments involving NSG mice were performed under a project licence (PPL 303442 or 2666723) approved by the UK Home Office under the Animal (Scientific Procedures) Act 1986 and in accordance with the principles of the 3Rs (replacement, reduction and refinement) in animal research.

2.3 Sample processing, cryopreservation and thawing

Fetal tissue was transported to the laboratory at 4°C and processed immediately as previously described [Roy et al., [2012]]. Briefly, FL tissue was passed through a 70µm filter, washed with PBS supplemented with 2% heat-inactivated fetal

bovine serum (HI-FBS; Sigma) and 1mM EDTA (PBS-EF) to obtain a single-cell suspension. FL cells were then red cell- and granulocyte-depleted using Ficoll Paque PLUS (GE Healthcare). Freshly isolated MNC then underwent CD34 enrichment using a CD34 MicroBead kit (Milltenyi). Cells were aliquoted and cryopreserved in liquid nitrogen in HI-FBS supplemented with 10% DMSO. Post-enrichment CD34+ cell purity was consistently >95%.

Cells were thawed in a water bath at 37°C for 3 minutes and then diluted in an excess of HI-FBS, spun down at 300g for 5 minutes and resuspended in IMDM (Gibco) supplemented with 10% HI-FBS and 100ng/ml DNase (Sigma), or other media appropriate to the particular assay. Cell counts and viability were determined using Trypan Blue (Gibco) staining and a haemocytometer.

2.4 Cell line maintenance

SEM and KOPN8 cells were grown in suspension in DMEM supplemented with 10% HI-FBS at a density of 0.5-2x10^6 cells/ml. Cells were counted and passaged into fresh media every 2-3 days.

HEK-293T cells were grown as a monolayer in DMEM supplemented with 10% HI-FBS. When HEK-293T cell monolayers reached 80% confluency, they were gently washed with PBS and incubated with 0.05% trypsin/EDTA (Gibco, 0.5% trypsin/EDTA diluted with PBS) at 37°C and 5% CO_2 for 5 minutes. Trypsinisation was stopped by adding an excess of media, spinning at 300g for 5 minutes and resuspending cells in fresh media. Cells were then split approximately 1 in 5 and replated in a fresh flask.

2.5 Flow cytometry analysis and sorting

Cells were stained for flow cytometry with fluorophore-conjugated monoclonal antibodies and viability dye (Table 2.2) in 100µl PBS-EF at 4°C for 30 minutes. Cells were then resuspended in 1ml PBS-EF and spun at 300g for 5 minutes twice to wash away excess antibodies. Cells were analysed on a BD Fortessa X20 or BD

Fortessa X50, and sorted using a BD Fusion or Sony MA900 instrument. Gates were set using unstained and fluorescence-minus-one (FMO) controls. Data were analysed using FlowJo V10.0 software.

Table 2.2: Antibodies used.

Antibody	Fluorophore	Clone	Company	Catalogue number
CD2	PerCP-Cy5.5	RPA-2.10	BioLegend	300216
CD3	PerCP-Cy5.5	OKT3	BioLegend	317336
CD3	BV711	OKT3	BioLegend	317328
CD10	PE-Cy7	ebioCB-CALLA	Life Technologies	25-0106-42
CD10	FITC	ebioCB-CALLA	Life Technologies	11-0106-42
CD11b	FITC	ICRF44	Life Technologies	11-0118-42
CD14	PerCP-Cy5.5	M5E2	BioLegend	301824
CD14	FITC	61D3	Life Technologies	11-0149-42
CD16	PerCP-Cy5.5	3G8	BioLegend	302028
CD19	APC	HIB19	BioLegend	302212
CD20	ef450	2H7	Life Technologies	48-0209-42
CD20	APC	2H7	BD Bioscience	641414
CD24	BV510	ML5	BioLegend	311126
CD33	FITC	P67.6	BioLegend	366620
CD34	BV421	561	BioLegend	343610
CD34	PE-Cy7	8G12	Life Technologies	25-0349-42
CD34	PerCP	8G12	BD Bioscience	345803
CD38	BV605	HIT2	BioLegend	303532
CD38	AF700	HIT2	Life Technologies	56-0389-42
CD38	FITC	HIT2	Life Technologies	11-0389-42
CD45	AF700	2D1	Life Technologies	56-9459-42
CD45	BUV395	HI30	BD Bioscience	563792
CD45	V450	30-F11	BD Bioscience	642275
CD45RA	FITC	HI100	BioLegend	304106
CD45RA	APC-ef780	HI100	Life Technologies	47-0458-42
CD56	PerCP-Cy5.5	HCD56	BioLegend	318322
CD56	BV605	HCD56	BioLegend	318334
CD90	BV421	5E10	BioLegend	328122
CD123	PE	6H6	BioLegend	306006
CD123	BV650	6H6	BioLegend	306020
CD127	PE	ebioRDR5	Life Technologies	12-1278-42
CD133	PE	AC133	Miltenyi	130-113-108
CD235a	PerCP-Cy5.5	HIR2	BioLegend	306614
AF4-C	unconjugated	-	Abcam	ab31812
FLAG-tag	unconjugated	M2	Sigma	F1804
H3K27ac	unconjugated	-	Diagenode	C15410196
H3K79me2	unconjugated	NL59	Millipore	04-835
IgD	FITC	IA6-2	BioLegend	348206
IgM	FITC	MHM-88	BioLegend	314506
Lineage	APC	-	Life Technologies	22-7776-72
MLL-N	unconjugated	-	Bethyl	A300-086A
Mouse CD45.1	APC-Cy7	30-F11	BioLegend	103116
NG2	PE	7.1	Beckman Coulter	B92429
Viability	ef506	-	Life Technologies	65-0866-18

2.6 CRISPR/Cas9 induction of the *MLL-AF4* / t(4;11) chromosomal translocation

2.6.1 Single guide RNA (sgRNA) design

sgRNAs were designed to target the intronic breakpoints of *MLL* and *AF4* most commonly found in infant ALL: intron 11 of *MLL* and intron 3 of *AF4* [Reichel et al., [2001]] (Figure 2.1). sgRNAs were chosen using the CRISPRScan online tool (`www.crisprscan.org`) based on having both high *in silico*-predicted on-target editing potential and low *in silico*-predicted off-target editing potential. sgRNAs were further validated *in silico* using the Synthego CRISPR guide verification online tool (`http://design.synthego.com/#/validate`). sgRNAs were purchased lyophilised from Synthego, resuspended at 100mM with nuclease-free water, aliquoted and stored at -80°C.

2.6.2 Primary human FL HSPC electroporation

CRISPR/Cas9 editing of primary human FL HSPC was carried out using a previously described protocol [Gundry et al., [2016]]. Briefly, CD34+ cells from an individual 12-20pcw human FL sample were cultured in suspension in StemLine II (Sigma) supplemented with 100ng/ml SCF, 100ng/ml FLT3L and 100ng/ml TPO (all Peprotech) at a density of 2.5x10^5 cells/ml at 37°C and 5% CO_2 for 10-12 hours. Cells were harvested, and resuspended in excess PBS and spun at 300g for 7 minutes three times to wash away culture medium. Cells were then resuspended at a density of 2x10^7 cells/ml in buffer T (Neon^{TM} Transfection kit, Thermo Fisher). In parallel, for every 2x10^5 cells, 1.5µg Cas9 protein (Alt-R® S.p. Cas9 Nuclease v3, Integrated DNA Technologies (IDT)) was incubated with either (i) no, (ii) one (1µg each) or (iii) both (0.5µg each) sgRNAs targeting *MLL* and/or *AF4* at room temperature for 20-30 minutes before being added to the cell suspension in buffer T. In this way, cells were electroporated with either (i) Cas9 protein only as a mock-edited control, (ii) a single Cas9/sgRNA ribonucleoprotein (RNP) targeting either *MLL* or *AF4* to test single sgRNA efficiencies, or (iii) a 1:1 mix of *MLL*- and

AF4-targeting RNPs to induce an *MLL-AF4* /t(4;11) chromosomal translocation (Figure 2.2). Electroporations were carried out using a Neon^{TM} Transfection System and a Neon^{TM} Transfection Pipette (Thermo Fisher) with the following parameters: 1600V, 10 milliseconds, 3 pulses. Electroporated cells were immediately placed in fresh StemLine II (Sigma) supplemented with 100ng/ml SCF, 100ng/ml FLT3L and 100ng/ml TPO (Peprotech) and incubated at 37°C and 5% CO_2 overnight (14-16 hours) to recover before being harvest for subsequent *in vitro*, *in vivo* and molecular assays. Cells with a CRISPR/Cas9-induced *MLL-AF4* /t(4;11) translocation were designated ^{CRISPR}*MLL-AF4+* (Figure 2.2).

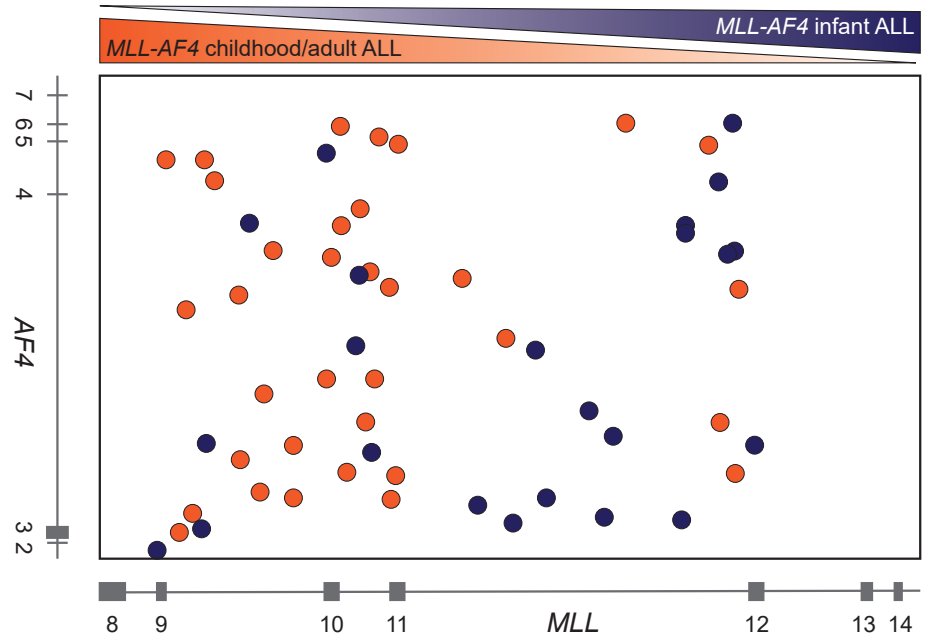


Figure 2.1: Biased distribution of of *MLL* breakpoint with age. Plot showing *MLL* (x-axis) and *AF4* (y-axis) genomic breakpoints in *MLL-AF4* translocation for *MLL-AF4* infant ALL (blue) and *MLL-AF4* childhood/adult ALL (orange) patients. Grey bars on each axis represent exonic regions of *MLL* and *AF4*. The *MLL-AF4* childhood ALL *MLL* breakpoint region is skewed toward introns 8-9, while the *MLL-AF4* infant ALL *MLL* breakpoint is skewed toward intron 11. (Plot adapted from Reichel et al., [2001]).

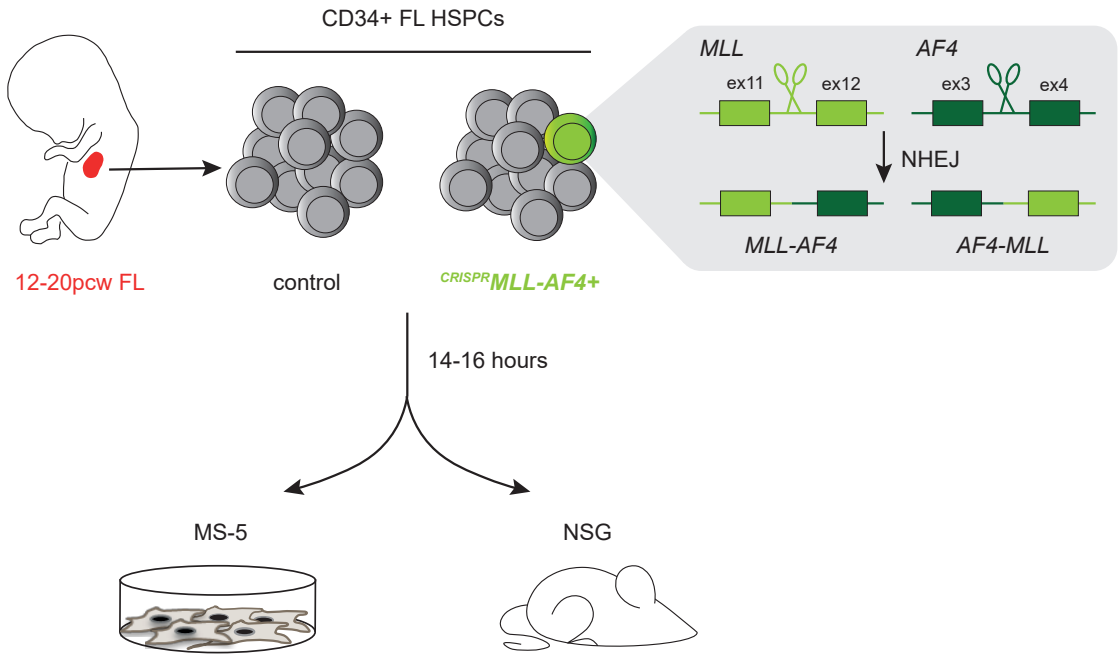


Figure 2.2: ^{CRISPR}*MLL-AF4+* editing strategy. Cells were harvest from individual 12-20pcw human FL samples and CD34-enriched. For each FL sample, CD34+ cells were split into 2 groups: the ^{CRISPR}*MLL-AF4+* group, which underwent CRISPR/Cas9 editing with both *MLL*- and *AF4*-targeting sgRNAs such that a proportion of cells would develop an *MLL-AF4* /t(4;11) translocation (light green cell); and the control group, which was electroporated with Cas9 protein only without sgRNAs to guide it to a particular genomic locus. Both edited and mock-edited control cells were allowed to recover overnight (14-16 hours) before being seeded onto MS-5 stromal co-cultures or transplanted into sub-lethally irradiated NSG mice.

2.7 MS-5 stromal co-culture assays

MS-5 stromal cells were grown to confluency and passaged as previously described [Goardon et al., [2011], O'Byrne et al., [2019]]. Briefly, MS-5 cells were grown as monolayers in α-MEM supplemented with 10% HI-FBS, 100U/ml Penicillin, 100µg/ml Streptomycin, 2mM L-glutamine, 50µM 2-Mercaptoethanol, and 10mM HEPES (α-MEM complete). When cells reached 80% confluency, they were gently washed with PBS and incubated with 0.05% trypsin/EDTA (Gibco, 0.5% trypsin/EDTA diluted with PBS) at 37°C and 5% CO_2 for 5 minutes. Trypsinisation was stopped by adding an excess of media, spinning at 300g for 5 minutes and resuspending cells in fresh media. Cells were then split approximately 1 in 5

and replated in a fresh flask.

When required for a co-culture assay, $3x10^4$ MS-5 cells were seeded in each well of a 24-well plate and incubated at 37°C and 5% CO_2 for 24 hours prior to initiation of the co-culture with primary human haematopoietic cells as described in [O'Byrne et al., [2019]]. 2000 unsorted FL HSPC or 1000 sorted subpopulations of FL HSPC (both ^{CRISPR}*MLL-AF4+* and control cells) were seeded onto the confluent MS-5 stromal layer in α-MEM complete supplemented with 20ng/ml SCF, 10ng/ml FLT3L, 10ng/ml IL-2 and 5ng/ml IL-7 (all Peprotech). Co-cultures were fed with a half media change every 3-4 days. At weekly timepoints, HSPC were replated onto a fresh MS-5 stromal layer. Cells were detached from the well by repeated pipetting and passed through a 70µm filter washed with PBS to obtain a single cell suspension depleted of MS-5 cells. Haematopoietic differentiation and *MLL-AF4/AF4-MLL* expression were assessed during replating by retaining half of the harvested cell suspension for flow cytometry and quantitative polymerase chain reaction (qPCR) assays. When absolute cell numbers are quoted, CountBright beads (Invitrogen) were used in the flow cytometry assay as per the manufacturer's instructions to calculate the total number of cells present in each well at each timepoint.

2.8 NSG xenograft transplantation assays

8-12 week old female NSG mice were sub-lethally irradiated with 2 doses of 1.25Gy 6 hours apart (2.5Gy total) and injected via the tail vein with $3x10^4$ unsorted FL HSPC (both ^{CRISPR}*MLL-AF4+* and control cells, with or without $3x10^4$ sex-mismatched FL CD34+ carrier cells) (Table 2.3) or 500-2000 sorted subpopulations of FL HSPC (^{CRISPR}*MLL-AF4+* cells with $3x10^4$ sex-mismatched FL MNC carrier cells). Human cell engraftment was monitored by peripheral blood sampling every 3 weeks. Human CD45+ (hCD45+) cells were sorted from peripheral blood samples to carry out *MLL-AF4* and *AF4-MLL* qPCRs to detect leukaemic cell engraftment. Animals were monitored regularly using a standardised physical scoring system, and any mouse found to be in distress was humanely killed. Mice were considered to be leukaemic if they met at least 3 of the following criteria: (i) overt signs of disease (hunching,

lack of movement, severe weight loss and/or paralysis), (ii) splenomegaly, (iii) PB blast count over 50%, (iv) peripheral organ (spleen, liver and/or CNS) infiltration, (v) detection of ^{CRISPR}*MLL-AF4+* cells in BM, spleen and peripheral blood.

Table 2.3: Bulk FL CD34+ HSPC xenografts.

FL samples	Primary recipient mice	Secondary recipients mice	Tertiary recipient mice
1	^{CRISPR}*MLL-AF4+* (n=1; preB), control (n=1)	^{CRISPR}*MLL-AF4+* (n=2; preB and mixed preB/proB); control (n=1)	^{CRISPR}*MLL-AF4+* (n=1), control (n=1)
2	^{CRISPR}*MLL-AF4+* (n=2; proB), control (n=1)	^{CRISPR}*MLL-AF4+* (n=2; proB), control (n=1)	^{CRISPR}*MLL-AF4+* (n=2), control (n=1)
3	^{CRISPR}*MLL-AF4+* (n=4; proB), control (n=3)	ND	ND

2.9 Histopathology

On termination, 0.5-1cm^2 samples were taken from the spleen and liver of ^{CRISPR}*MLL-AF4+* and control mice and fixed in 10% formaldehyde (FA). All subsequent processing steps were carried out by the Translational Histopathology Laboratory (THL), CRUK Oxford Centre, Department of Oncology, University of Oxford. After fixation, tissue samples were processed and paraffin-embedded. 4µm paraffin sections were cut onto SuperFrost Plus adhesive slides (VWR). Haematoxylin and Eosin (H&E) staining was performed using the Vector Laboratories H&E kit as per the manufacturer's instructions and mounted using Vectamount (Vector Laboratories).

Heads were harvested from ^{CRISPR}*MLL-AF4+* and control mice and fixed in 10% FA to analyse leukaemic infiltration into the CNS. All subsequent processing steps were carried out in the laboratory of Professor Christina Halsey, Wolfson Wohl Cancer Research Centre, University of Glasgow as previously described [Shi et al., [2020]]. Briefly, following paraffin embedding, 2.5µm sections were cut onto Poly-L-silane-coated slides and stained with Gill's Haematoxylin and Putt's Eosin (both made in-house). Slides were imaged on a NanoZoomer Digital Pathology (NDP) slide scanner (Hamamatsu) and analysed with NDPview 2 software.

2.10 *In vitro* drug sensitivity assays

Leukaemic cells (^{CRISPR}*MLL-AF4+* cells, PDX samples and cell lines) were assayed *in vitro* for their response to prednisolone and L-asparaginase treatment using the MTT assay (Cell Proliferation kit I, Roche) as previously described [Pieters et al., [1990], Pieters et al., [1991]]. Briefly, cells were cultured in suspension at a density of 2x10^6 cells/ml in a 96-well flat-bottom plate (100µl/well, or 2x10^5 cells/well) in RPMI supplemented with 10% HI-FBS, 100U/ml Penicillin, 100ng/ml Streptomycin, 1% ITS liquid media supplement (Sigma) and either a range of concentrations (0-900µg/ml) of prednisolone (Sigma, resuspended in DMSO) or a range of concentrations (0-10IU/ml) of L-asparaginase (Cambridge Biosciences, resuspended in water). For prednisolone, DMSO equivalent to the highest drug concentration was added to untreated control cells. Cells were incubated with drug at 37°C and 5% CO_2 for 48-96 hours, followed by addition of 10µl MTT reagent (Cell Proliferation kit I, Roche) and incubation for a further 4 hours. During this time, the tetrazolium salt MTT is reduced to a purple-coloured formazan product by living but not dead cells. 100µl solubilisation solution (Cell Proliferation kit I, Roche) was added to each well to dissolve the formazan crystals. The optical density (OD) of each well was measured using a SPECTROStar Nano microplate reader (BMG Labtech) at 570nm. For each drug concentration, leukaemic cell survival (LCS) was calculated using the following equation:

$$LCS = (ODtreatedwell/ODuntreatedwell)x100\%$$

Drug resistance was expressed as LC50, the drug concentration lethal to 50% of the cells.

2.11 Plasmid cloning

2.11.1 FLAG-tagged *MLL-AF4* plasmid cloning

A plasmid backbone, including a pSFFV promoter sequence upstream of the clone site and an ampicillin resistance cassette, was generously provided by Dr Philip

Hublitz, Genome Engineering Facility, MRC Weatherall Institute of Molecular Medicine, University of Oxford (originally purchased from Addgene, `www.addgene.org`). *MLL* (exon 1-11) and *AF4* (exon 4-20, followed by a T2A separator sequence and a *NeonGreen* reporter) transcripts were codon-optimised and ordered using the GeneArt online tool (Invitrogen, `www.thermofisher.com/order/geneartgenes`). Forward and reverse oligonucleotides encoding 3 repeat FLAG-tag sequences (Table 2.4) were ordered lyophilised from IDT, resuspended at 100mM in nuclease-free water and annealed by heating to 94°C and gradually cooling. All transcripts were inserted stepwise into the plasmid backbone using the InFusion cloning kit (Takara) as per the manufacturer's instructions. Briefly, the plasmid was linearised by digestion with restriction enzymes (NEB) at 37°C for 30 minutes, purified by gel electrophoresis (0.8-2% agarose gel, 1µg/ml ethidium bromide) and extracted from the gel using a Gel Extraction kit (Qiagen). PCR was used to add 15 nucleotide (nt) overhangs to each end of the inserts complementary to the 15nt sequences flanking the insertion site of the plasmid (Takara, Table 2.5). The reaction was run using a thermocycler (T100 Thermal Cycler, Bio-RAD) (Table 2.6). Amplicons were purified using a PCR Purification kit (Qiagen). The InFusion cloning reaction (Table 2.7) was incubated at 50°C for 15 minutes, followed by immediate transformation into competent bacteria. During the stepwise insertion of *MLL*, *AF4*-T2A-*NeonGreen* and FLAG-tag transcripts, plasmids were screened after each cloning reaction and positive clones were used in the subsequent step.

Table 2.4: Primers used.

Gene	Application	Sequence
MLL sgRNA	CRISPR	GGAGUGACAGACUGCACCAC
AF4 sgRNA	CRISPR	UGAGAUGGAACUGGCGACAU
MLL intron 11 forward	PCR	AGGTGTGGTGGTGGGC
AF4 intron 3 reverse	PCR	GCCCTATCAGTTCCCTGCCA
AF4 intron 3 forward	PCR	GAGAGGCTGAGGTTGTAGGG
MLL intron 11 reverse	PCR	GGCAACTGATTACACCCAATT
MLL-AF4 cDNA forward	qPCR	GTGGAAGGCAACATCAGGC
MLL-AF4 cDNA reverse	qPCR	TATTGCTGTCAAAGGAGGCGG
AF4-MLL cDNA forward	qPCR	CAGAAGCCCACGGCTTATGT
AF4-MLL cDNA reverse	qPCR	AGGCACTCAGGGTGATAGC
GAPDH cDNA forward	qPCR	AACAGCGACACCCATCCTC
GAPDH cDNA reverse	qPCR	CATACCAGGAAATGAGCTTGACAA
KDM6B off-target forward	PCR	CTGTACGAGTCAGAGCACGAT
KDM6B off-target reverse	PCR	GCTCTCTCTAAGGTCACTTCCG
CTNNDA off-target forward	PCR	GCAGAGGCATAAAGGGTCAACA
CTNNDA off-target reverse	PCR	AAAGTGCCCTCCTCCCTACTA
ATRNL1 off-target forward	PCR	CATAGCAGTTCACAGGGGATTCT
ATRNL1 off-target reverse	PCR	AAATGCAACTCATCTCCACTCGTA
SMR3A off-target forward	PCR	AGATGTGACAACAACCGCAG
SMR3A off-target reverse	PCR	GATTTCTACTCCCTGCTGGCAC
PARVA off-target forward	PCR	ATGTGAGCAGGAGAGAGGGG
PARVA off-target reverse	PCR	GTGTTCCAGAGGCCATCCAG
DCC off-target forward	PCR	AGCTAGGCACAAAAGAGAGATCC
DCC off-target reverse	PCR	CACAGCTTATGTGCAGTCATCC
TRGC2 off-target forward	PCR	ATCTGTGGAGTGTGGCTCTAC
TRGC2 off-target reverse	PCR	CTCCTGTATGGTATTCACTGCCC
TULP3 off-target forward	PCR	GGGGCTATGGGAATGCTCAA
TULP3 off-target reverse	PCR	GGGGAGGAGTGCCCTGAATA
FOXO1 off-target forward	PCR	TGGCAACAAACTCCAATAGCC
FOXO1 off-target reverse	PCR	GGTCTGTGACCATAAAGCAGGA
KCNQ2 off-target forward	PCR	TTTCTCCTCCACCTGCGTTT
KCNQ2 off-target reverse	PCR	GGAAGGGATTGGCCTGTCTT
HOXA9 MLL-AF4 binding site forward	ChIP-qPCR	GCACAAAGCGGCTCTAAACCTCAG
HOXA9 MLL-AF4 binding site reverse	ChIP-qPCR	ATACCTCCTCCATCAAAGCGGCAG
MEIS1 MLL-AF4 binding site forward	ChIP-qPCR	TGACAACCTCGCCTGTGATTGAC
MEIS1 MLL-AF4 binding site reverse	ChIP-qPCR	TGACAACCTCGCCTGTGATTGAC
MEF2C MLL-AF4 binding site forward	ChIP-qPCR	GACATCTTCGGGCTGCTATT
MEF2C MLL-AF4 binding site reverse	ChIP-qPCR	TACTTCGTCCAGCGTTGAAG
Negative control region forward	ChIP-qPCR	GGCTCCTGTAACCAACCACTACC
Negative control region reverse	ChIP-qPCR	CCTCTGGGCTGGCTTCATTC
3x FLAG-tag forward	annealing	CACAAAGGCGGCGATCGACTACAAAGACCATGACGG
3x FLAG-tag reverse	annealing	CTGGGGATCCGCGATCTTGTCATCGTCATCCTTGTA

Table 2.5: Plasmid cloning PCR master mix.

Reagent	Amount
2x CloneAmp HiFi PCR premix	10µl
Forward/reverse primer mix (10mM)	1µl
Purified DNA template	100ng
Nuclease-free water	to 20µl

Table 2.6: Standard PCR parameters.

Step	Tempterature (°C)	Time (seconds)	Cycles
1	95	180	1
2	98	20	-
3	58-63	15	-
4	72	60-180	repeat steps 2-4 x30
5	72	300	1
6	4	HOLD	-

Table 2.7: InFusion cloning reaction master mix.

Reagent	Amount
5x InFusion HD Enzyme Premix	2µl
Linearised plasmid backbone	200ng
Purified PCR product	100ng
Nuclease-free water	to 10µl

2.11.2 Bacterial transformation

Ligation products and plasmids were transformed into Stable Competent E. coli (NEB) as per the manufacturer's instructions. Transformed bacteria were plated onto ampicillin-containing agarose plates and incubated overnight at 37°C. For plasmid screening, 10 clones were picked and grown in LB media supplemented with ampicillin for 6-8 hours at 37°C, rotating at 250rpm. Plasmid clones were then purified using a Miniprep kit (Qiagen) as per the manufacturer's instructions and screened for successful insertion by diagnostic restriction digest, followed by gel electrophoresis (0.8-2% agarose gel, 1µg/ml ethidium bromide). For amplification of previously validated plasmids, 1 positive clone was picked and grown in LB media supplemented with ampicillin overnight at 37°C, rotating at 250rpm, for purification using an Endotoxin-free Maxiprep kit (Qiagen) as per the manufacturer's instructions.

2.12 HEK-293T cell transfections

When required for transfections, 5x10^6 HEK-293T cells were seeded onto a 10cm tissue culture-treated dish in 10ml DMEM supplemented with 10% HI-FBS and incubated at 37°C and 5% CO_2 for 24 hours prior to transfection. After 24 hours,

media was removed such that 3ml remained in the dish. In parallel, for every 10cm dish of HEK-293T cells, 200µl Opti-MEM media (Gibco) was mixed with 12µl Transit-LT1 transfection reagent (Mirus) and incubated at room temperature for 5 minutes, followed by addition of 4µg plasmid DNA. This Opti-MEM/Transit-LT2/plasmid mix was then added dropwise evenly across each 10cm dish of cells and incubated at 37°C and 5% CO_2 for 24 hours. After 24 hours, cells were checked for successful transfection (NeonGreen positivity) under a fluorescence microscope (Leica), then gently washed with PBS and incubated with 0.05% trypsin/EDTA (Gibco, 0.5% trypsin/EDTA diluted with PBS) at 37°C and 5% CO_2 for 5 minutes. Trypsinisation was stopped by adding an excess of media, spinning at 300g for 5 minutes and resuspending cells in PBS-EF. Cells were stained with viability dye at 4°C for 30 minutes, resuspended in 1ml PBS-EF and spun at 300g for 5 minutes twice to wash away excess dye. Viable NeonGreen+ HEK-293T cells were sorted for subsequent molecular assays.

2.13 Molecular assays

2.13.1 DNA and RNA extraction

DNA was extracted from $5x10^4$-$2x10^6$ cells using the DNeasy Blood and Tissue kit (Qiagen) as per the manufacturer's instructions, quantified using the NanoDrop (Thermo Fisher) and stored at -20°C. Whole cell RNA was extracted from 500-$5x10^4$ cells using the RNeasy Micro Plus kit (Qiagen) and $5x10^4$-$2x10^6$ cells using the RNeasy Mini kit (Qiagen) as per the manufacturer's instructions. RNA was quantified using the NanoDrop (Thermo Fisher) and either taken immediately forward for cDNA synthesis or stored at -80°C.

2.13.2 cDNA synthesis

RNA was mixed with 10mM dNTP mix (Invitrogen) and 50µM random hexamers (Roche) and placed at 65°C for 5 minutes before being cooled on ice for 2 minutes. When working with multiple samples, starting RNA was diluted such that the same RNA concentration was used for all samples. This mixture was then added to the

cDNA reaction mix (Table 2.8, Invitrogen). The reaction was run using a thermocycler (T100 Thermal Cycler, Bio-RAD) (Table 2.9). cDNA was stored at -20°C.

Table 2.8: cDNA synthesis master mix.

Reagent	Amount
5x First Strand buffer	4µl
0.1M DTT	2µl
RNaseOUT™	1µl
SuperScript III enzyme	1µl
Nuclease-free water	2µl

Table 2.9: cDNA synthesis parameters.

Step	Tempterature (°C)	Time (minutes)	Cycles
1	25	5	1
2	50	60	1
3	15	70	1
4	4	HOLD	-

2.13.3 PCR and Sanger sequencing

All PCR in this project unrelated to plasmid cloning was performed using KAPA HiFi HotStart master mix (Roche). 10-100ng DNA was mixed with 1µl (10mM) forward/reverse primer mix (IDT, 2.4) and 10µl 2x KAPA HiFi HotStart mix. Nuclease-free water was added to bring the reaction volume up to 20µl. PCR reactions were run on a thermocycler (T100 Thermal Cycler, Bio-RAD) (Table 2.6).

Sanger sequencing was carried out using the Eurofins Mix2Seq kit as per the manufacturer's protocol. Briefly, 50-200ng DNA was mixed with 1µl (10mM) sequencing primer (IDT, 2.4) in a Mix2Seq barcoded tube and sent for sequencing by the manufacturer. Data were downloaded from the Eurofins website (`www.eurofins.co.uk`) and visualised using SnapGene v4.1.7 software. When testing sgRNA on- and off-target editing potential, Sanger sequencing traces around the cut site from edited and unedited control FL HSPC were uploaded to the Synthego ICE analysis online tool (`http://ice.synthego.com`). Editing potential was expressed

as the percentage of reads in the edited sample that contained insertions and/or deletions (indels) compared to the unedited control sample.

2.13.4 IgH rearrangement analysis

Samples were screened for IgH complete (VH-DH-JH) and IgH incomplete (DH-JH) rearrangements by Gary Wright, Great Ormond Street Hospital for Children, London using BIOMED-2 protocols to detect clonality as described previously [O'Byrne et al., [2019], Wright et al., [2019]]. Briefly, DNA was extracted from cells from the BM of 3 primary ^{CRISPR}*MLL-AF4+* recipient mice and sequenced using the Thermo Fisher BigDye® Terminator v3.1 Cycle Sequence Kit protocol. Capillary electrophoresis was performed on a Thermo Fisher Scientific 24-capillary 3500xL Genetic Analyser. Sequences were entered into the National Center for Biotechnology Information (NCBI) IgBLAST tool for immunoglobulin domain identification and alignment and annotated using IMGT/V-QUEST [Brochet et al., [2008]].

2.13.5 qPCR

All qPCR in this project was carried out using Fast SYBR^{TM} Green Master Mix (Applied Biosystems). 0.1-10ng cDNA was mixed with 1µl (10mM, IDT) forward/reverse primer mix (Table 2.4) and 10µl Fast SYBR^{TM} Green Master Mix (Applied Biosystems) in a 96-well optical plate (Applied Biosystems). Nuclease-free water was added to bring the reaction volume up to 20µl. qPCR reactions were run on a QuantStudio3 real-time PCR system (Thermo Fisher). ΔCt values were calculated by subtracting the mean Ct value of housekeeping gene, *GAPDH*, from the gene-of-interest. ΔCt values were transformed ($2^{-\Delta Ct}$) and plotted in GraphPad Prism v7.04.

2.13.6 Western blotting

1-5x10^6 cells were resuspended in 50µl BC300 (20mM Tris-HCl pH 7.5, 20% glycerol, 30mM KCl, 5mM EDTA) plus 0.5% NP40 and protease inhibitor cocktail (Roche) and incubated on ice for 30 minutes. Cells were then spun at 16,200g at 4°C for

10 minutes and the supernatant was transferred to a fresh tube. 50µl NuPAGE LDS buffer (Life Technologies, 2x master mix) plus 0.1% 2-Mercaptoethanol was added to the supernatant. Samples were run on pre-cast 3-8% Tris-Acetate gel (Novex, Life Technologies) in Tris-Acetate Running Buffer (Life Technologies, 20x master mix). Gels were run at 100V for 1-2 hours. The protein molecular weight was estimated using a Precision Plus Protein Kaleidoscope (Bio-RAD) ladder.

Proteins were transferred from gel to a PVDF membrane (Immobilon®-P Transfer Membrane, EMD Millipore) using a wet transfer setup. Briefly, gel and membrane were sandwiched between filter paper and sponges and submerged in a tank filled with transfer buffer (192mM glycine, 25mM Trizma® base, 10% methanol) and run at 100V and 4°C for 1 hour. Following transfer, membranes were incubated with Ponceau stain (Sigma) for 5 minutes and washed 3 times in 100% methanol. The membranes were then dried, re-hydrated and blocked in 5% milk for 1 hour. Membranes were incubated with monoclonal antibodies (Table 2.2) at 4°C overnight (14-18 hours), washed 3 times in TBS plus 0.05% Tween-20 for 30 minutes and subsequently incubated with the appropriate secondary antibody (Table 2.2). Membranes were then washed a further 3 times in TBS plus 0.05% Tween-20 for 30 minutes and probed using ECL (ECL Prime Western Blotting Detection Reagent, Amersham, GE Healthcare). Solutions A and B were mixed in a 1:1 ration and added over the membrane and incubated for 5 minutes at room temperature. Excess solution was removed and the membrane was imaged using a Chemi-Doc Imaging System (Bio-RAD).

2.13.7 RNA-seq

$1x10^6$ unsorted cells were used for RNA-sequencing of *MLL-AF4* infant and childhood ALL patient samples. For ^{CRISPR}*MLL-AF4+* samples, approximately $3x10^5$ hCD45+CD19+ cells were sorted from the BM of ^{CRISPR}*MLL-AF4+* and control primary recipient mice. Total RNA was extracted using an RNeasy Mini Kit (Qiagen). Poly(A) purification was conducted using the NEB Poly(A) mRNA magnetic isolation module as per the manufacturer's instructions. Library

preparation was carried out using the Ultra II Directional RNA Library Prep Kit (NEB, E7765) as per the manufacturer's instructions. Prior to sequencing, libraries were quality controlled by analysis on a 2200 TapeStation instrument (Agilent Technologies) as per the manufacturer's instructions to check for amplification and lack of adapter dimers. Libraries were quantified using the KAPA Universal Complete Kit (Kappa Biosystems) and pooled for sequencing. RNA libraries were sequenced by paired-end sequencing using a 150 cycle high output kit on a Nextseq 500 (Illumina).

2.13.8 Assay for Transposase-Accessible Chromatin (ATAC)-seq

ATAC-seq was carried out using a published protocol, adapted for small cell numbers as described previously [Corces et al., [2016]]. 500-2000 cells per subpopulation of FL HSPC were sorted directly into 23.75µl lysis buffer (2x TD buffer (Illumina) plus 0.01% digitonin (Promega)) at 4°C. 1.25µl Tn5 enzyme was then added to the lysis buffer and immediately incubated at 37°C for 30 minutes, rotating at 300rpm. DNA was purified using a Reaction Clean-up kit (Qiagen) and eluted in 10µl elution buffer (EB). Transposed fragments were amplified and custom sequencing indexes (IDT) added by PCR (Table 2.10). Following PCR, an equal volume of room temperature Ampure XP beads (Ambion) was added to the mix and incubated at room temperature for 5 minutes. Beads were gently washed with 80% ethanol on a magnetic stand three times. Ethanol was then removed and beads were allowed to dry on the magnetic stand. DNA was eluted from the beads in 10µl EB at room temperature for 2 minutes. Tubes were placed back on the magnetic stand and the supernatant was transferred to a fresh tube. Prior to sequencing, libraries were quality controlled by analysis on a 2200 TapeStation instrument (Agilent Technologies) as per the manufacturer's instructions to check for amplification and lack of adapter dimers. Libraries were quantified using a Qubit system (Thermo Fisher) and pooled for sequencing. ATAC libraries were sequenced by paired-end sequencing using a 150 cycle high output kit on a Nextseq 500 (Illumina).

Table 2.10: ATAC-sequencing indexing PCR parameters.

Step	Tempterature (°C)	Time (seconds)	Cycles
1	72	500	1
2	98	30	1
3	98	10	-
4	63	30	-
5	72	60	repeat steps 3-5 x12
6	4	HOLD	-

2.13.9 Chromatin Immunoprecipitation (ChIP)

ChIP was carried out as previously described [Kerry et al., [2017]]. Briefly, cells were either double-fixed (for transcription factors and other chromatin-binding proteins) by incubation in 1ml 2mM disuccinimidyl glutarate (DSG, Sigma) at room temperature for 30 minutes with rotation, followed by resuspension in PBS and spinning at 10,000g for 1 minutes and incubation in 1ml 1% FA at room temperature for 30 minutes with rotation, or single-fixed (for histone modifications) in 1ml 1% FA at room temperature for 10 minutes with rotation. Fixed cells were stored at -80°C.

Fixed cells were thawed on ice and resuspended in 100µl SDS lysis buffer (50mM Tris-HCl pH 8.1, 10mM EDTA, 1% SDS) plus 1:200 protease inhibitor cocktail (Roche), and syringe-passaged through a 27 gauge (27G) needle. Lysed samples were transferred to a pre-slit snap-cap microTUBE (Covaris Inc) and sonicated in 4°C degassed water using a S220 Covaris sonicator (Covaris Inc) with the following parameters: power = 75, duty = 25, cycles/burst = 500, duration = 600 seconds (double-fixed sample) or 300 seconds (single-fixed sample). Sonicated samples were transferred to a fresh tube and spun at 20,000g for 10 minutes at 4°C. The supernatant was transferred to a fresh tube and diluted in excess ChIP dilution buffer (CDB, 16.7mM Tris-HCl pH8.1, 167mM NaCl, 1.2mM EDTA; 0.01% SDS and 1.1% Triton X-100 (Sigma)) plus 1:1000 protease inhibitor cocktail. 10µl of a 1:1 mix of Protein A and Protein G DynaBeads (Novex, Life Technologies) was added to the sample and rotated at 4°C for 30 minutes, followed by spinning at 1000g for 1 minute and transfer of the supernatant to a fresh tube. 5% of the supernatant was stored at 4°C to act as an input DNA control. 2µl of the appropriate monoclonal

antibody was added to the remainder of the supernatant and incubated at 4°C overnight (14-18 hours) with rotation, followed by a further incubation with 20µl of a 1:1 mix of Protein A and Protein G DynaBeads (Novex, Life Technologies) for a further 3-5 hours. Beads were then washed three times on a magnetic stand with 1ml RIPA buffer (50mM HEPES-KOH pH7.6, 500mM LiCl; 1mM EDTA, 1% NP-40 and 0.7% Na-Deoxycholate), switching to a fresh tube between washes, and once with 1ml Tris-EDTA (TE) plus 50mM NaCl. Beads were then spun at 960g for 3 minutes and eluted in 105µl ChIP elution buffer (50mM Tris-HCl pH8.1, 10mM EDTA and 1% SDS) at 65°C for 30 minutes. Samples were then spun at 16000g at room temperature for 1 minute and the supernatant was transferred to a new tube for incubation at 65°C overnight to remove cross-links. Input DNA controls were diluted to 100µl with ChIP elution buffer and also incubated at 65°C overnight to remove cross-links. ChIP and input DNA control samples were incubated with 1µl 500ng/ml RNase at 37°C for 30 minutes and 1µl 14-22mg/ml Proteinase K at 45°C for 30 minutes before purification with the ChIP Clean and Concentrator kit (Zymo Research) as per the manufacturer's instructions.

For ChIP-qPCR, samples were diluted to 200µl with nuclease-free water and used for subsequent qPCR assays. For ChIP-seq, samples were eluted in 55µl nuclease-free water. 5µl was used for quality control analysis by ChIP-qPCR and quantification by Qubit. The remaining 50µl was used to generate DNA libraries using the NEBNext Ultra DNA library preparation kit for Illumina (NEB, E7103). Prior to sequencing, libraries were quality controlled by analysis on a 2200 TapeStation instrument (Agilent Technologies) as per the manufacturer's instructions to check for amplification and lack of adapter dimers. Libraries were quantified using a Qubit system (Thermo Fisher) and pooled for sequencing. ChIP libraries were sequenced by paired-end sequencing using a 75 cycle high output kit on a Nextseq 500 (Illumina).

2.14 Bioinformatic analyses

2.14.1 RNA-seq analysis

Following sequencing, read quality was analysed using the fastQC package. Adaptor and quality trimming was carried out using the Trim Galore package. Trimmed reads were mapped to the human genome assembly (hg38) using STAR and the featureCounts function of the Subread package was used to quantify gene expression levels using default parameters. The resultant expression counts tables were used to identify differentially expressed genes in R v4.0.1 using the edgeR package. Differential gene expression was defined as an adjusted P-value (FDR) of <0.05. When counts tables from multiple sources were combined for analysis, counts were first batch-corrected in R v4.0.1 using the RUVSeq package to normalise based on the following housekeeping genes: *GAPDH*, *SDHA*, *B2M* and *HMBS*. To derive a FL vs adult BM gene signature, bulk RNA-seq for sorted subpopulations of FL HSPC were compared to matched sorted subpopulations of adult BM HSPC (FL HSC vs adult BM HSC, FL MPP vs adult BM MPP, FL LMPP vs adult BM LMPP and FL committed B progenitors (CBP) vs adult BM CLP). Genes that were differentially expressed between FL and adult BM in at least one matched HSPC subpopulation were included in the gene signature. Genes that showed a significant change in opposite directions in different HSPC subtypes (e.g. upregulated in FL HSC vs adult BM HSC, but downregulated in FL LMPP vs adult BM LMPP) or in the normal vs leukemic setting (e.g. upregulated in FL HSPC vs adult BM HSPC, but downregulated in *MLL-AF4* infant ALL vs *MLL-AF4* childhood ALL) were filtered out of the gene signature.

2.14.2 ATAC-seq analysis

Following sequencing, read quality was analysed using the fastQC package. Adaptor and quality trimming was carried out using the Trim Galore package. Trimmed reads were mapped to the human genome assembly (hg38) using Bowtie2. ATAC peaks were called using the MACS2 package with default parameters. The makeBigWig.pl

function of the HOMER package was used to generate bigwig files for visualization in UCSC, normalizing tag counts to tags per 1x10^7 total tags. Peak accessibility counts tables were generated in R v4.0.1 using the DiffBind package and differentially accessible peaks were identified using either the DiffBind package or the edgeR package. Differentially accessible peaks were defined by an FDR <0.05. When peak accessibility counts tables from multiple sources were combined for analysis, counts were first batch-corrected in R v4.0.1 using the RUVr function of the RUVSeq package. Peaks were linked to their nearest gene in R v4.0.1 using the Genomic Ranges package.

2.14.3 ChIP-seq analysis

For ChIP-seq, quality control of FASTQ reads, alignment, PCR duplicate filtering, blacklisted region filtering and UCSC data hub generation was performed using an in-house pipeline as previously described [Telenius & Hughes, [2018]]. The makeBigWig.pl function of the HOMER package was used to generate bigwig files for visualization in UCSC, normalizing tag counts to tags per 1x10^7 total tags. ChIP-seq peaks were called using the findPeaks.pl function of the HOMER package with ChIP input sample used to estimate background signal. Gene profiles were generated using the annotatePeaks.pl function of the HOMER package. MLL-AF4 peaks were defined as MLL-N and AF4-C peaks within 1kb of one another.

2.15 Statistics

Two-tailed Mann-Whitney, Log-rank (Mantel-Cox) tests and ANOVA followed by multiple comparisons testing were used to compare experimental groups as indicated in the figure legends. Statistical analyses were performed using GraphPad Prism v7.04 or R v4.0.1. Data are expressed as mean ± SEM unless otherwise indicated.

3

Modelling *MLL-AF4* infant ALL using primary human FL HSPCs

Contents

3.1 Introduction 50
3.1.1 CRISPR/Cas9 as a method for inducing chromosomal translocations 50
3.1.2 The *MLL-AF4* breakpoint varies with age 52
3.1.3 Human FL HSPCs as a target for leukaemic transformation 52
3.2 Aims 53
3.3 A CRISPR/Cas9 editing approach to induce an *MLL-AF4*/t(4;11) translocation 53
3.4 ^{CRISPR}*MLL-AF4*+ FL HSPC undergo dramatic B lineage expansion *in vitro* 54
3.5 ^{CRISPR}*MLL-AF4*+ FL cells with an *MLL* intron 9 breakpoint do not consistently induce a B lineage expansion *in vitro* 60
3.6 ^{CRISPR}*MLL-AF4*+ FL HSPCs cause aggressive B-ALL *in vivo* 64
3.7 CRISPR/Cas9-induced off-target editing was not detected in ^{CRISPR}*MLL-AF4*+ ALL 71
3.8 ^{CRISPR}*MLL-AF4*+ ALL recapitulates key clinico-pathological features of *MLL-AF4* infant ALL 73
3.9 Discussion 78

3.1 Introduction

MLL-AF4 infant ALL represents a distinct subset of ALL with poor prognosis and uniquely aggressive clinico-pathological features [Pui & Campana, [2007], Roberts & Fordham, [2018], Pieters et al., [2019]]. In order to gain a greater understanding of the factors that contribute to this aggressive phenotype, a faithful human fetal HSPC-derived model of the disease is needed. However, this has so far proven difficult to generate [reviewed by Milne, [2017]]. I set out to develop a method to generate a faithful model of *MLL-AF4* infant ALL and discuss the results in this chapter. To achieve this aim, I hypothesised that there were two key requirements. Firstly, the most common driver mutation found in infant ALL, an *MLL-AF4*/t(4;11) chromosomal translocation, should be induced; and secondly, that the cell context in which the driver mutation first arises should be recapitulated by targeting primary human fetal HSPCs. The results presented in this chapter have recently been published [Rice et al., [2021]].

3.1.1 CRISPR/Cas9 as a method for inducing chromosomal translocations

The clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated (Cas) system was first identified as an RNA-mediated adaptive defence system in bacteria and archaea [Wiedenheft et al., [2012]]. Cas proteins are endonucleases that function in complex with small RNA molecules to cleave the DNA of invading viruses. The CRISPR/Cas9 system requires a base-pairing structure between an activating tracrRNA and a targeting crRNA, which recognises a target (protospacer) DNA sequence, for Cas9 to cause a double-strand break (DSB) at the protospacer-adjacent motif (PAM) sequence (NGG) [Figure 3.1a) [Jinek et al., [2012], Wiedenheft et al., [2012]]. The first use of an engineered tracrRNA:crRNA chimeric sgRNA to target the Cas9 endonuclease to specific regions of DNA (Figure 3.1b) was described in 2012 [Jinek et al., [2012], Gasiunas et al., [2012]]. Shortly after the initial report, Cong et al., [2013] and Mali et al., [2013] further adapted this system to function as a genome editing tool in eukaryotes.

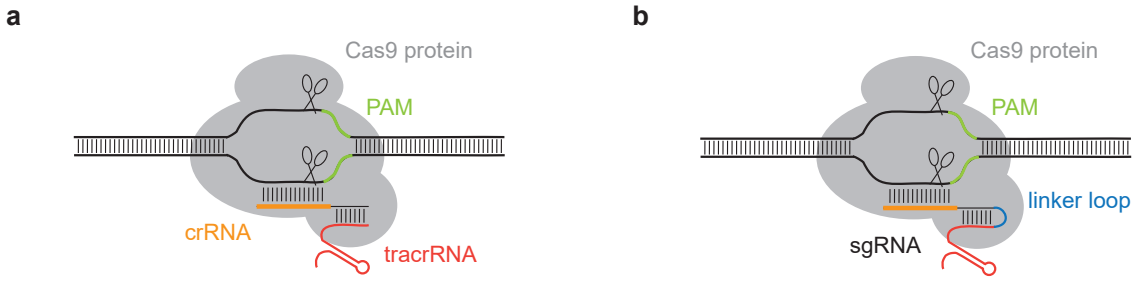


Figure 3.1: CRISPR/Cas9 function. a. Schematic representation of Cas9 protein (grey) recruited to a locus of double-stranded DNA (black lines) by a tracrRNA:crRNA pair (orange and red lines), leading to a DSB (scissors) at the PAM sequence (NGG; green line). **b.** Optimised CRISPR/Cas9 system for genome editing. Schematic representation of Cas9 protein (grey) recruited to a locus of double-stranded DNA (black lines) by a chimeric sgRNA pair (orange and red lines), leading to a DSB (scissors) at the PAM sequence (NGG; green line).

The precise mechanism by which chromosomal translocations arise *in vivo* is not fully understood. However, DNA sequencing of leukaemia patients has revealed that the translocation breakpoint often has deletions, suggesting that the NHEJ DNA repair system, which is often associated with deletions at the repair site, is likely to underlie these translocation events [Torres et al., [2014]]. In light of this, it was hypothesised that CRISPR/Cas9-induced DSBs at genes commonly found to be fused in cancer could generate cancer-causing chromosomal translocations by NHEJ. This method was found to efficiently induce translocations between the targeted regions, resulting in expression of the expected fusion transcript, first in cell lines [Choi & Meyerson, [2014], Torres et al., [2014]] and subsequently in primary human cells [Reimer et al., [2017], Secker et al., [2019]].

While CRISPR/Cas9 editing could represent a new affordable, intuitive and efficient system for modelling cancer-causing translocation events compared to previous methods, it can also lead to DSBs at off-target sites throughout the genome due to partial recognition of off-target sequences by the sgRNA [Cho et al., [2014]]. These off-target DSBs could result in mutations in the coding regions of off-target genes, complex chromosomal translocations between target and off-target chromosomes or other major structural abnormalities. Therefore, it is essential to evaluate the specificity as well as the efficiency of editing when using CRISPR/Cas9.

3.1.2 The *MLL-AF4* breakpoint varies with age

The breakpoint of *MLL* found in all major *MLL*-rearrangements spans an 8.3kb region [Thirman et al., [1993]]. However, in *MLL-AF4* ALL, the exact location of the breakpoint within this region shows a biased distribution with age. In infants, the breakpoint most commonly occurs around intron 11, whereas in older children and adults, the breakpoint most commonly occurs upstream, around intron 8-9 [Reichel et al., [2001], Meyer et al., [2018]] (Figure 2.1) . The underlying mechanism for this age-related bias in breakpoint regions is unknown, and the role this may play in driving the *MLL-AF4* infant ALL phenotype is also unclear, as the additional *MLL* sequence encodes only the remainder of the intrinsically disordered region and the first PHD finger, which has no reported function as part of the MLL fusion protein [Muntean et al., [2008], Ali et al., [2014]]. However, an association between the infant ALL breakpoint and the poor prognosis $HOXA^{lo}$ gene expression signature has previously been reported in *MLL*-rearranged leukaemia [Meyer et al., [2013]], and the most successful previous attempt at modelling *MLL-AF4* infant ALL involved retroviral transduction of an *MLL-Af4* fusion gene that recapitulated this infant ALL-biased *MLL* breakpoint in intron 11 [Lin et al., [2016]], suggesting that, while we do not currently fully understand why, this breakpoint region may be significant.

3.1.3 Human FL HSPCs as a target for leukaemic transformation

Fetal HSPCs have a number of unique molecular and functional characteristics compared to post-natal HSPCs. It seems plausible that these properties may cooperate with an *MLL-AF4*/t(4;11) translocation to drive the aggressive phenotype of infant ALL. These characteristics, including an increased proliferative capacity compared to post-natal HSPCs and fetal-specific gene expression programmes, have predominantly been characterised in the FL [Lansdorp et al., [1993], Muench et al., [1994], Harrison et al., [2010], Copley et al., [2013], Copley & Eaves, [2013],Symeonidou et al., [2021]]. By contrast, much less is known about the fetal BM particularly in humans [O'Byrne et al., [2019]]. Therefore, in order to

model the fetal-specific cell context in which *MLL-AF4* infant ALL first arises, I chose the more thoroughly characterised primary human FL HSPCs as the target for a CRISPR/Cas9-induced *MLL-AF4*/t(4;11) translocation.

3.2 Aims

The overall aim of this chapter was to generate a faithful model of *MLL-AF4* infant ALL by targeting an *MLL-AF4*/t(4;11) translocation to primary human FL HSPCs, which I split into the following specific objectives:

- Develop a CRISPR/Cas9-based method to induce an *MLL-AF4*/t(4;11) translocation in primary human FL HSPCs
- Characterise these edited FL HSPCs, designated ^{CRISPR}*MLL-AF4+* cells, *in vitro*
- Test the significance of the *MLL* breakpoint on ^{CRISPR}*MLL-AF4+* cell phenotype *in vitro*
- Develop and characterise ^{CRISPR}*MLL-AF4+* ALL in a xenograft model
- Investigate the presence of infant ALL-specific clinico-pathological features in ^{CRISPR}*MLL-AF4+* ALL

3.3 A CRISPR/Cas9 editing approach to induce an *MLL-AF4*/t(4;11) translocation

To directly induce an *MLL-AF4*/t(4;11) translocation using a plasmid-free and virus-free CRISPR/Cas9 RNP approach, sgRNAs were designed to target the most common intronic breakpoints found in infant ALL: intron 11 of *MLL* and intron 3 of *AF4* [Reichel et al., [2001]]. While some *MLL-AF4* fusion genes do have exonic breakpoints [Reichel et al., [2001]] (Figure 2.1), introns were targeted so that transcription of the fusion gene would not be disrupted by CRISPR/Cas9-induced

indels at the breakpoint, thus resulting in a consistent fusion gene transcript between CRISPR/Cas9 editing replicates (Figure 3.2a and 3.2b).

sgRNAs that were predicted *in silico* to be both highly efficient and specific to the *MLL* and *AF4* target regions were chosen using the CRISPRscan online tool. Using a previously described protocol [Gundry et al., [2016]], Cas9/*MLL*-sgRNA RNPs, or Cas9 protein only as a mock-edited control, were electroporated into primary human FL CD34+ cells. 24 hours post-electroporation, genomic DNA was isolated from edited and control cells and the region around the *MLL* cut site was sequenced. Based on the proportion of sequencing reads with indels detected in edited cells compared to biologically matched, mock-edited control cells, the editing efficiency of the Cas9/*MLL*-sgRNA RNP was estimated to be 77% (estimated based on Sanger sequencing data (Figure 3.2a) using the ICE analysis tool (Synthego)). The same approach was used to test the editing efficiency of the Cas9/*AF4*-sgRNA RNP, which was also estimated to be approximately 77% (estimated based on Sanger sequencing data (Figure 3.2b) using the ICE analysis tool (Synthego)). Therefore, these highly efficient sgRNAs were taken forward for use in all subsequent experiments.

A 1:1 mix of Cas9/*MLL*-sgRNA and Cas9/*AF4*-sgRNA RNPs was electroporated into 13-20pcw human FL CD34+ cells to induce the translocation by NHEJ. Edited samples (n=6) and biologically matched mock-edited controls (Cas9 protein only, n=6) were transferred to MS-5 stromal co-cultures to facilitate expansion of successfully edited cells along the B lineage. I designated the edited FL cells as ^{CRISPR}*MLL-AF4+* and the mock-edited cells as controls (Figure 2.2).

3.4 ^{CRISPR}*MLL-AF4+* FL HSPC undergo dramatic B lineage expansion *in vitro*

By week 3 of MS-5 stromal co-culture, the total number of human haematopoietic cells (hCD45+) per well was about 60-fold higher in ^{CRISPR}*MLL-AF4+* cultures compared to controls ($p<0.05$, n=6; approximately $3.1\text{x}10^6 \pm 1.7\text{x}10^6$ in ^{CRISPR}*MLL-AF4+* cultures vs $5\text{x}10^4 \pm 2\text{x}10^4$ in controls) (Figure 3.3a), and the total number of CD19+ B lineage cells was about 600-fold higher ($p<0.05$, n=6) (Figure 3.3b).

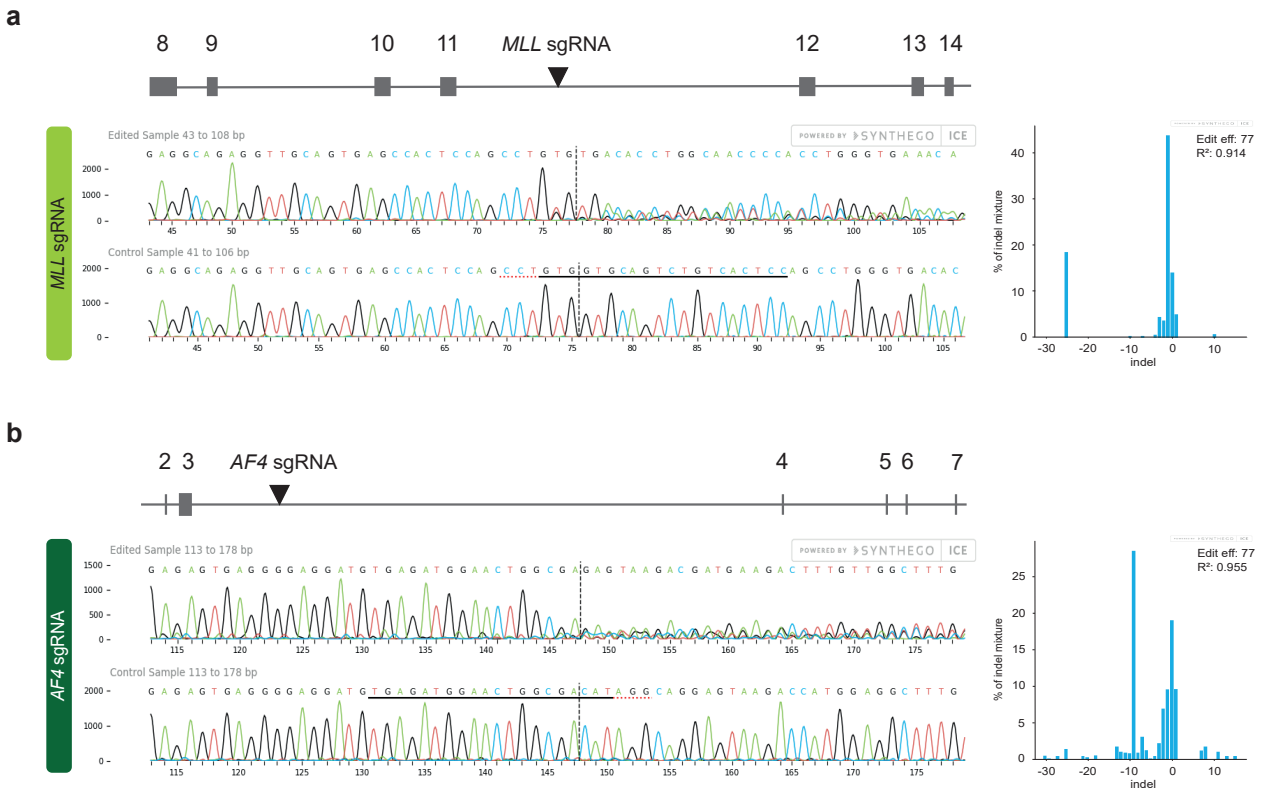


Figure 3.2: Efficiency of *MLL*- and *AF4*-targeting sgRNAs a. (top) position of the *MLL*-targeting sgRNA in the *MLL* gene. Grey bars represent exonic regions. (bottom left) Sanger sequencing traces for edited (above) and mock-edited control (below) cells. (bottom right) Quantification of indels in edited cells based on Sanger sequencing results, using the ICE analysis tool (Synthego). **b.** (top) position of the *AF4*-targeting sgRNA in the *AF4* gene. Grey bars represent exonic regions. (bottom left) Sanger sequencing traces for edited (above) and mock-edited control (below) cells. (bottom right) Quantification of indels in edited cells based on Sanger sequencing results, using the ICE analysis tool (Synthego).

Fluorescent in situ hybridisation (FISH) analysis using *MLL* and *AF4* probes confirmed the presence of a heterozygous *MLL-AF4*/t(4;11) translocation in >80% of hCD45+ cells in ^{CRISPR}*MLL-AF4*+ cultures at week 4 (n=2) (Figure 3.3c), and RT-qPCR confirmed that this resulted in expression of both *MLL-AF4* and the reciprocal fusion gene, *AF4-MLL* (n=3) (Figure 3.3d).

Virtually all human haematopoietic cells generated from ^{CRISPR}*MLL-AF4*+ cultures were CD19+ B lineage cells, compared to <20% in control cultures (Figure 3.4a and 3.4b). Moreover, while the absolute number of CD34+ progenitor cells was

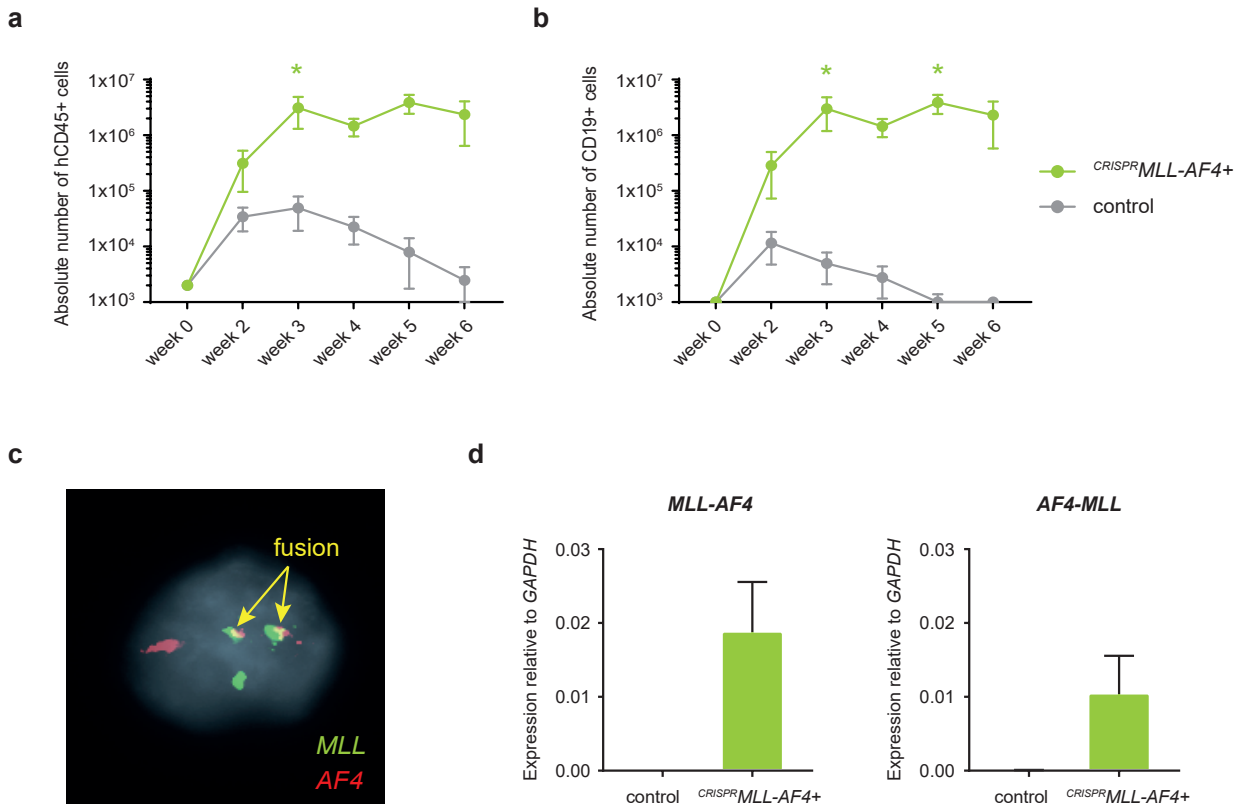


Figure 3.3: ^{CRISPR}*MLL-AF4+* FL cells undergo a dramatic B lineage expansion *in vitro*. **a.** Cumulative absolute number of hCD45+ cells per well over time during ^{CRISPR}*MLL-AF4+* (light green) and control (grey) MS-5 co-culture (n=6). * =p<0.05 (Two-way ANOVA with Sidak correction for multiple comparisons). Data shown as mean ± SEM. **b.** Cumulative absolute number of CD19+ cells per well over time during ^{CRISPR}*MLL-AF4+* (light green) and control (grey) MS-5 co-culture (n=6). * =p<0.05 (Two-way ANOVA with Sidak correction for multiple comparisons). Data shown as mean ± SEM. **c.** Representative FISH image, showing reciprocal *MLL-AF4*/t(4;11) translocation. *MLL* probe shown in green and *AF4* probe shown in red. Combined yellow signal (arrows) represent *MLL-AF4* and *AF4-MLL* fusion genes. One wild-type allele remains for both *MLL* and *AF4*. Data generously provided by Dariusz Ladon, Great Ormond Street Hospital for Children, London. **d.** qPCR for hCD45+ cells harvested from co-culture, showing expression of *MLL-AF4* and *AF4-MLL* relative to *GAPDH* at week 4 of MS-5 co-culture (n=3). Data shown as mean ± SEM.

comparable between ^{CRISPR}*MLL-AF4+* and control cultures (data not shown), the majority of ^{CRISPR}*MLL-AF4+* CD34+ cells were CD19+ B progenitors, suggesting that *MLL-AF4*-driven B lineage specification may occur at a progenitor stage (Figure 3.4b). More detailed immunophenotyping showed that the majority of ^{CRISPR}*MLL-AF4+* cells were CD34-CD19+CD10+IgM/IgD- preB cells, of which

5-10% aberrantly expressed the leukaemia-associated cell surface marker CD133 (Figure 3.4b), which is encoded by the *PROM1* gene, a known direct target of the MLL-AF4 fusion protein [Godfrey et al., [2020]]. Together, these data suggested that inducing an *MLL-AF4*/t(4;11) translocation alone was sufficient to transform primary human FL HSPCs *in vitro*.

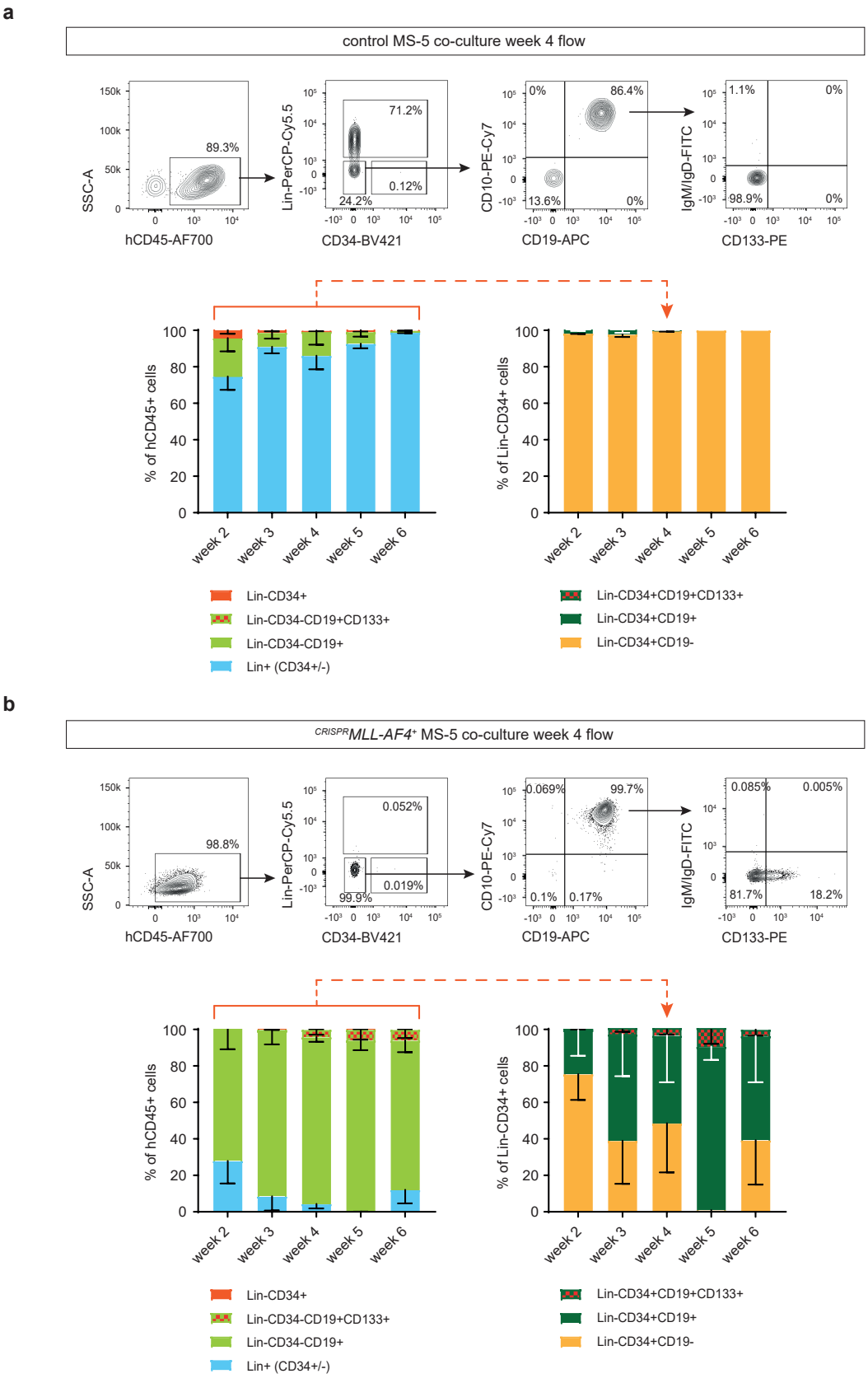

a
control MS-5 co-culture week 4 flow
SSC-A
hCD45-AF700
89.3%
Lin-PerCP-Cy5.5
CD34-BV421
71.2%
0.12%
24.2%
CD10-PE-Cy7
CD19-APC
0%
86.4%
13.6%
0%
IgM/IgD-FITC
CD133-PE
1.1%
0%
98.9%
0%
% of hCD45+ cells
% of Lin-CD34+ cells
week 2
week 3
week 4
week 5
week 6
Lin-CD34+
Lin-CD34-CD19+CD133+
Lin-CD34-CD19+
Lin+ (CD34+/-)
Lin-CD34+CD19+CD133+
Lin-CD34+CD19+
Lin-CD34+CD19-
b
CRISPRMLL-AF4+ MS-5 co-culture week 4 flow
SSC-A
hCD45-AF700
98.8%
Lin-PerCP-Cy5.5
CD34-BV421
0.052%
0.019%
99.9%
CD10-PE-Cy7
CD19-APC
0.069%
99.7%
0.1%
0.17%
IgM/IgD-FITC
CD133-PE
0.085%
0.005%
81.7%
18.2%
% of hCD45+ cells
% of Lin-CD34+ cells
week 2
week 3
week 4
week 5
week 6
Lin-CD34+
Lin-CD34-CD19+CD133+
Lin-CD34-CD19+
Lin+ (CD34+/-)
Lin-CD34+CD19+CD133+
Lin-CD34+CD19+
Lin-CD34+CD19-

Figure 3.4

Figure 3.4: Virtually all cells in ^{CRISPR}*MLL-AF4+* cultures are CD19+CD10+ preB cells. a. (top) Representative flow cytometry plots for control MS-5 co-cultures at week 4. Arrows represent sequential gating strategy. Percentage of parent population in each gate is shown in each sequential plot. (bottom left) Quantification of human haematopoietic cell populations in control MS-5 co-cultures over time, shown as a percentage of hCD45+ cells (n=6). (bottom right) Quantification of haematopoietic progenitor cells in control MS-5 co-cultures over time, shown as a percentage of Lin-CD34+ cells (orange population in left-hand plot) (n=6). Data shown as mean ± SEM. **b.** (top) Representative flow cytometry plots for ^{CRISPR}*MLL-AF4+* MS-5 co-cultures at week 4. Arrows represent sequential gating strategy. Percentage of parent population in each gate is shown in each sequential plot. (bottom left) Quantification of human haematopoietic cell populations in ^{CRISPR}*MLL-AF4+* MS-5 co-cultures over time, shown as a percentage of hCD45+ cells (n=6). (bottom right) Quantification of haematopoietic progenitor cells in ^{CRISPR}*MLL-AF4+* MS-5 co-cultures over time, shown as a percentage of Lin-CD34+ cells (orange population in left-hand plot) (n=6). Data shown as mean ± SEM. Lin = CD2, CD3, CD14, CD16, CD56, CD235a.

While transformation was documented for every FL sample tested (n=8; n=6 co-cultures presented here, n=2 co-cultures subsequently carried out by Dr Thomas Jackson, Roy laboratory), not all wells with ^{CRISPR}*MLL-AF4+* FL cells in a particular experiment showed evidence of proliferation. Under the microscope, ^{CRISPR}*MLL-AF4+* cells formed distinctive leukaemic cell clusters in the wells that included transformed cells (Figure 3.5). Based on this visual analysis, the proportion of edited FL CD34+ co-culture wells plated that gave rise to transformed ^{CRISPR}*MLL-AF4+* cells was recorded. 26/52 wells had transformed (n=8). As 2,000 edited cells were plated per well, if we assume the B lineage expansion in each well arose from a single transformed clone, the efficiency of ^{CRISPR}*MLL-AF4+* leukaemic transformation would be a minimum of about 1:4,000 edited cells.

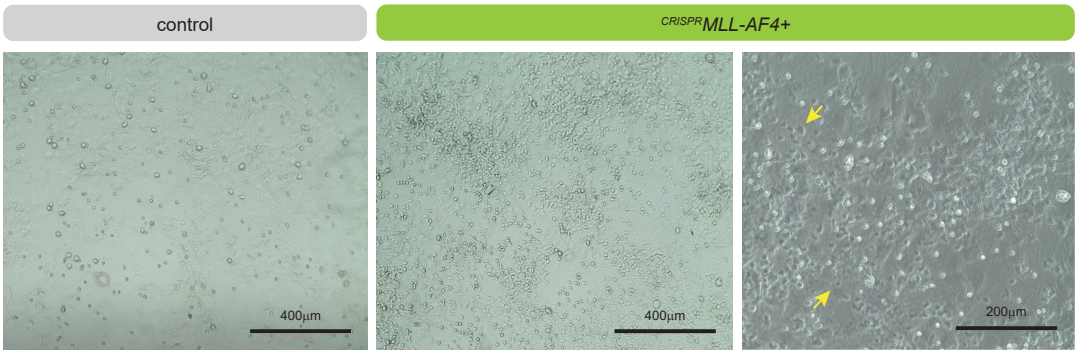


Figure 3.5: ^{CRISPR}*MLL-AF4+* FL cells form distinctive leukaemic cell clusters *in vitro*. (left) Representative brightfield image of a control MS-5 co-culture well at week 3. The haematopoietic cells are the circular cells sitting on top of the MS-5 stromal layer. Scale bar = 400µm. (right) representative brightfield images of a ^{CRISPR}*MLL-AF4+* MS-5 co-culture well at week 3 at increasing magnifications. Yellow arrows point to examples of ^{CRISPR}*MLL-AF4+* cells, which, unlike untransformed cells that appear as bright, circular cells sitting on top of the stromal layer, appear to be darker, with a halo, suggesting they may be burrowing underneath the MS-5 stromal layer. Scale bars = 400µm (left) and 200µm (right).

3.5 ^{CRISPR}*MLL-AF4+* FL cells with an *MLL* intron 9 breakpoint do not consistently induce a B lineage expansion *in vitro*

I next wanted to test the role of the infant ALL-biased *MLL* intron 11 breakpoint in driving leukaemic transformation of FL HSPCs. Cas9/sgRNA RNPs targeting intron 9 of *MLL* and intron 3 of *AF4* (sgRNA sequences provided by Prof Pablo Menéndez, Barcelona University) were electroporated into 15-19pcw FL CD34+ cells using the same protocol as before (Figure 3.6). Edited and biologically matched mock-edited control cells (n=3) were transferred to MS-5 stromal co-cultures.

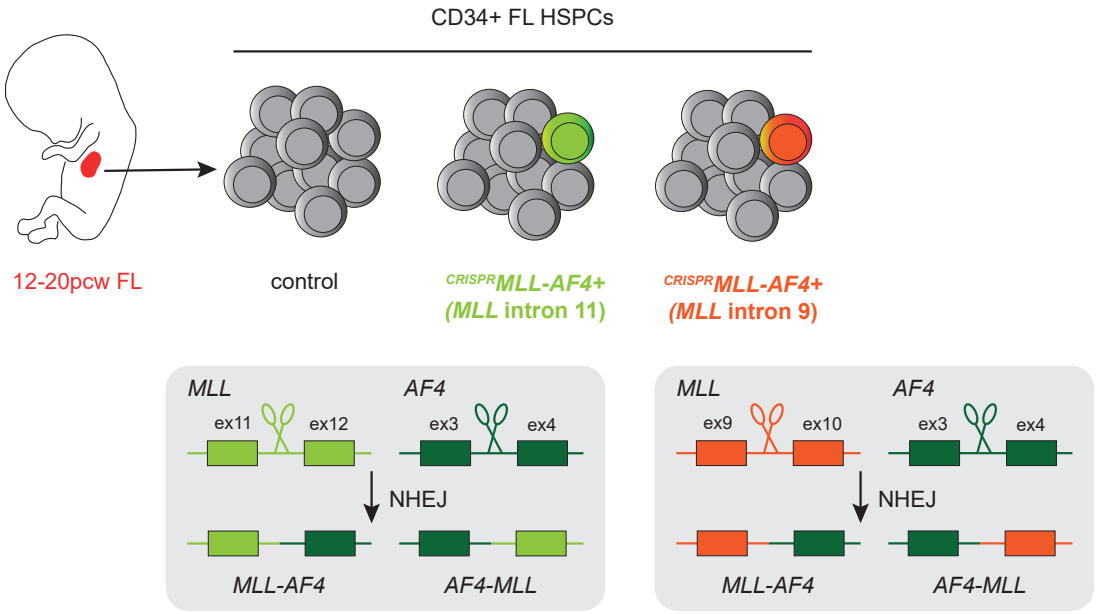


Figure 3.6: Comparison of *MLL* intron 11 and *MLL* intron 9 breakpoints *in vitro*. CD34+ FL cells, biologically matched to FL samples with existing *MLL* intron 11 ^{CRISPR}*MLL-AF4+ in vitro* data, were electroporated with Cas9/sgRNA RNPs targeting the *MLL* intron 9 and *AF4* intron 3 breakpoints to cause an *MLL-AF4*/t(4;11) translocation more commonly observed in *MLL-AF4* childhood and adult ALL patients (n=3), or with Cas9 protein only as a mock-edited control. These cells were then seeded onto MS-5 stromal co-cultures. *MLL* intron 9 and alternative *AF4* intron 3 sgRNA sequences were provided by Prof Pablo Menedez, Barcelona University.

Similarly to ^{CRISPR}*MLL-AF4+* cells with an *MLL* intron 11 breakpoint, all ^{CRISPR}*MLL-AF4+* cells with an *MLL* intron 9 breakpoint (3/3) transformed to give rise to a dramatic increase in hCD45+ cells by week 3 of culture (Figure 3.7a), which expressed *MLL-AF4* (Figure 3.7b). In fact, throughout the culture period, there was no significant difference in the total number of hCD45+ cells per well between *MLL* intron 11 ^{CRISPR}*MLL-AF4+* cultures and *MLL* intron 9 ^{CRISPR}*MLL-AF4+* cultures (Figure 3.7a). Interestingly however, while 2/3 *MLL* intron 9 ^{CRISPR}*MLL-AF4+* cultures gave rise to the characteristic B lineage immunophenotype of *MLL-AF4* infant ALL, 1/3 *MLL* intron 9 ^{CRISPR}*MLL-AF4+* cultures predominantly gave rise to a Lin+CD19- haematopoietic cell expansion (Figure 3.7c and 3.7d). The nature of this cellular expansion was not clear from these data. Subsequent immunophenotyping revealed that CD56 represented one of the lineage markers

expressed by the Lin+ cells, possibly suggesting a CD56+ AML-like expansion, but this was not further investigated due to time limitations. Taken together, these preliminary data suggested that an *MLL-AF4* /t(4;11) translocation with an infant ALL-biased *MLL* intron 11 breakpoint may more efficiently drive B lineage specification of FL HSPCs. I therefore used sgRNAs targeting intron 11 of *MLL* and intron 3 of *AF4* to generate ^{CRISPR}*MLL-AF4+* cells in all subsequent experiments.

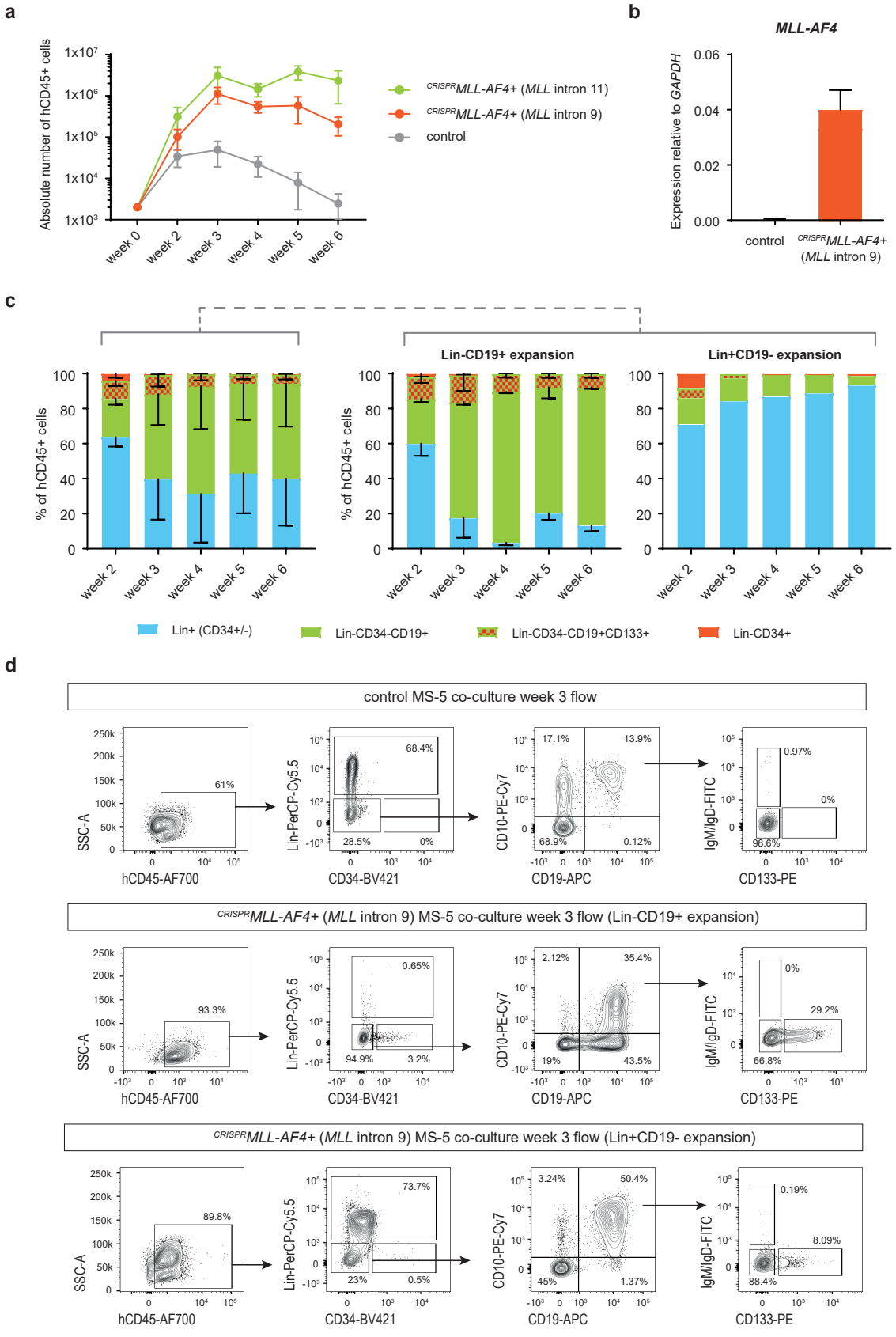

a
Absolute number of hCD45+ cells
1x10^7
1x10^6
1x10^5
1x10^4
1x10^3
week 0
week 2
week 3
week 4
week 5
week 6
CRISPR MLL-AF4+ (MLL intron 11)
CRISPR MLL-AF4+ (MLL intron 9)
control
b
MLL-AF4
Expression relative to GAPDH
0.06
0.04
0.02
0.00
control
CRISPR MLL-AF4+ (MLL intron 9)
c
Lin-CD19+ expansion
Lin+CD19- expansion
% of hCD45+ cells
Lin+ (CD34+/-)
Lin-CD34-CD19+
Lin-CD34-CD19+CD133+
Lin-CD34+
d
control MS-5 co-culture week 3 flow
SSC-A
hCD45-AF700
61%
Lin-PerCP-Cy5.5
CD34-BV421
68.4%
28.5%
0%
CD10-PE-Cy7
CD19-APC
17.1%
13.9%
68.9%
0.12%
IgM/IgD-FITC
CD133-PE
0.97%
0%
98.6%
CRISPR MLL-AF4+ (MLL intron 9) MS-5 co-culture week 3 flow (Lin-CD19+ expansion)
93.3%
0.65%
94.9%
3.2%
2.12%
35.4%
19%
43.5%
0%
29.2%
66.8%
CRISPR MLL-AF4+ (MLL intron 9) MS-5 co-culture week 3 flow (Lin+CD19- expansion)
89.8%
73.7%
23%
0.5%
3.24%
50.4%
45%
1.37%
0.19%
8.09%
88.4%

Figure 3.7

Figure 3.7: ^{CRISPR}*MLL-AF4+* FL cells with an *MLL* intron 9 breakpoint cause a dramatic, mixed lineage expansion *in vitro*. a. Cumulative absolute number of hCD45+ cells per well over time during *MLL* intron 11 ^{CRISPR}*MLL-AF4+* (light green; n=6), *MLL* intron 9 ^{CRISPR}*MLL-AF4+* (orange; n=3) and control (grey; n=6) MS-5 co-culture. Data shown as mean ± SEM. **b.** qPCR for *MLL* intron 9 ^{CRISPR}*MLL-AF4+* or control hCD45+ cells harvested from co-culture, showing expression of *MLL-AF4* relative to *GAPDH* at week 4 of MS-5 co-culture (n=2). Data shown as mean ± SEM. **c.** (left) Quantification of human haematopoietic cell populations in all ^{CRISPR}*MLL-AF4+* *MLL* intron 9 MS-5 co-cultures over time, shown as a percentage of hCD45+ cells (n=3). Data shown as mean ± SEM. (right) Separate plots showing quantification of human haematopoietic cell populations in ^{CRISPR}*MLL-AF4+* *MLL* intron 9 MS-5 co-cultures with a Lin-CD19+ expansion (n=2), and a Lin+CD19- expansion (n=1) over time. Lin = CD2, CD3, CD14, CD16, CD56, CD235a. **d.** Representative flow cytometry plots for control (top), Lin-CD19+ ^{CRISPR}*MLL-AF4+* *MLL* intron 9 (centre) and Lin+CD19- ^{CRISPR}*MLL-AF4+* *MLL* intron 9 (bottom) MS-5 co-cultures at week 3. Arrows represent sequential gating strategy. Percentage of parent population in each gate is shown in each sequential plot.

3.6 ^{CRISPR}*MLL-AF4+* FL HSPCs cause aggressive B-ALL *in vivo*

To test whether ^{CRISPR}*MLL-AF4+* cells could cause leukaemia *in vivo*, human FL CD34+ cells (n=3; 13pcw) were edited as before and transplanted into sub-lethally irradiated NSG mice (^{CRISPR}*MLL-AF4+* , n=7 mice; biologically matched mock-edited controls. n=5 mice) 14-16 hours after electroporation (Figure 2.2).

Engraftment of human cells was monitored by peripheral blood sampling every 3 weeks. Both ^{CRISPR}*MLL-AF4+* cells and control cells engrafted efficiently in NSG mice, and there was no significant difference in the proportion of hCD45+ cells in the peripheral blood between ^{CRISPR}*MLL-AF4+* and control mice (Figure 3.8a). However, there were notable differences in the phenotype of these cells. At week 12 post-transplant, expression of both *MLL-AF4* and *AF4-MLL* was clearly detectable by qPCR in hCD45+ cells sorted from the peripheral blood of ^{CRISPR}*MLL-AF4+* mice (n=6) (Figure 3.8b). Moreover, from week 12 onward, there was a clear enrichment of human CD34+ progenitor cells in the peripheral blood of ^{CRISPR}*MLL-AF4+* mice compared to controls (Figure 3.8c), and more detailed immunophenotyping revealed that these were CD34+CD19+ B progenitor cells. Finally, morphological analysis revealed a high proportion of blast cells in

the peripheral blood of $^{CRISPR}MLL\text{-}AF4+$ mice, which represented virtually all human cells by week 18 post-transplant (Figure 3.8d).

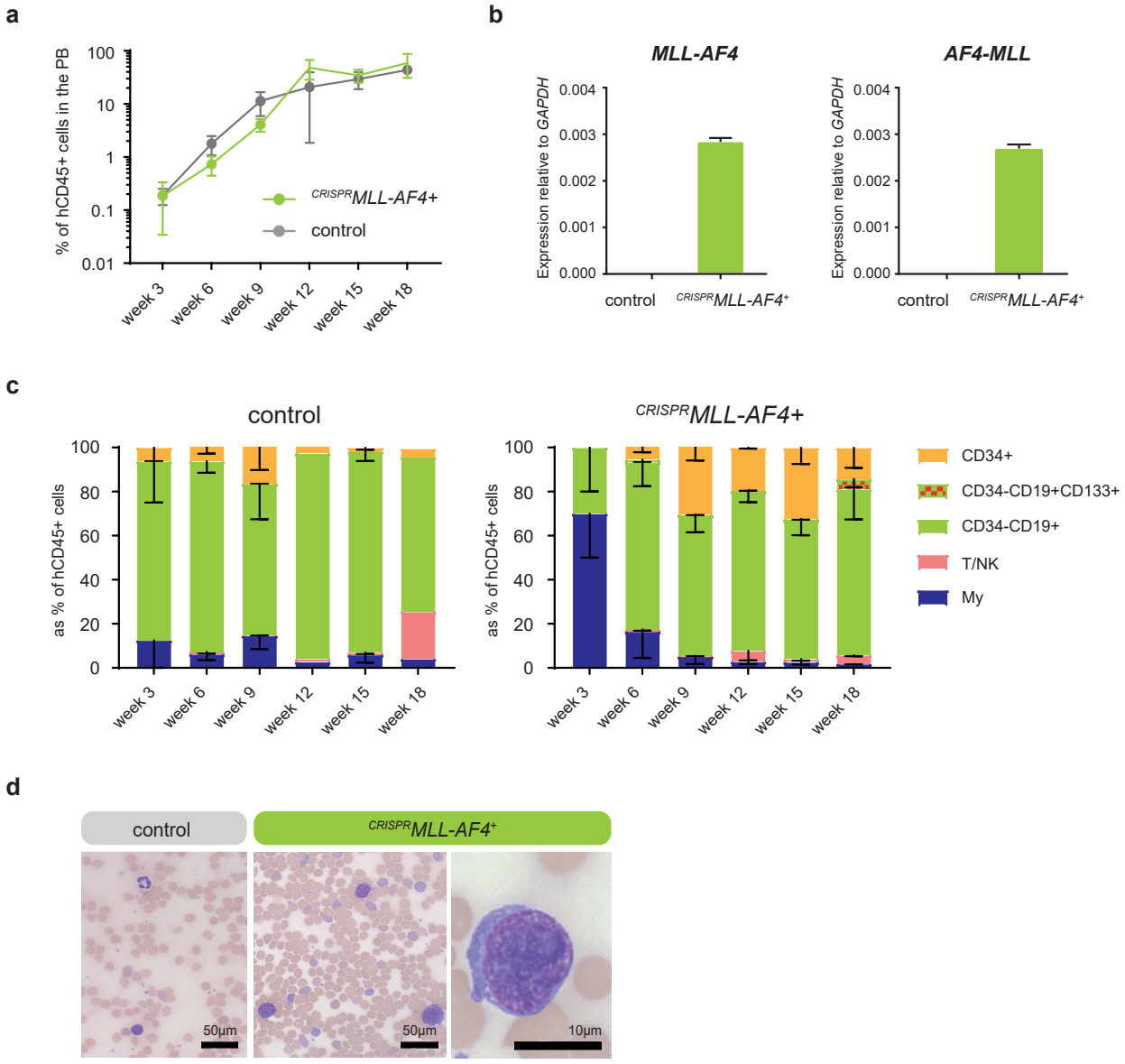


Figure 3.8

Figure 3.8: ^{CRISPR}*MLL-AF4+* cells are detectable in the peripheral blood of recipient mice from week 9 post-transplant. a. Peripheral blood engraftment of hCD45+ cells over time in primary ^{CRISPR}*MLL-AF4+* (light green; n=7) and control (grey; n=5) recipient mice, quantified as a percentage of all CD45+ cells (mouse CD45.1+ (mCD45+) and hCD45+). Data shown as mean ± SEM (Two-way ANOVA with Sidak correction for multiple comparisons. ns; p>0.5). **b.** qPCR showing expression of *MLL-AF4* and *AF4-MLL* relative to *GAPDH* in hCD45+ cells isolated from peripheral blood at week 12 post-engraftment (n=6). Data shown as mean ± SEM. **c.** Quantification of human haematopoietic cell populations in control (left, n=5) and ^{CRISPR}*MLL-AF4+* (right, n=7) primary recipient mouse peripheral blood over time, shown as a percentage of hCD45+ cells. T/NK = CD3, CD56. My = CD11b, CD14, CD33. **d.** Representative H&E-stained peripheral blood films for control (left) and ^{CRISPR}*MLL-AF4+* (right) primary recipient mice. Low magnification images (scale bar = 50µm) show multi-lineage cells in the peripheral blood in controls and predominantly circulating blast cells in ^{CRISPR}*MLL-AF4+* mice. High magnification image (scale bar = 10µm) shows a representative blast cell from ^{CRISPR}*MLL-AF4+* peripheral blood.

B-ALL rapidly developed in all (7/7) ^{CRISPR}*MLL-AF4+* mice with a median latency of 18 weeks post-transplant, whereas no control mice (0/5) developed any form of leukaemia (Figure 3.9a). On termination, when ^{CRISPR}*MLL-AF4+* mice displayed signs of ill-health, hCD45+ cells engraftment in the BM was significantly greater in ^{CRISPR}*MLL-AF4+* mice compared to controls (p<0.01), accounting for >90% of all haematopoietic cells in the BM (Figure 3.9b). Virtually all cells in the ^{CRISPR}*MLL-AF4+* BM were CD19+CD20-CD34+/- B-ALL blasts (Figure 3.9c), and sequencing of the IgH locus revealed that this expansion was clonal (Table 3.2). On average, about 50% (range = 0.01-97.7%) of blasts co-expressed CD34 (Figure 3.9c) but this varied between individual recipient mice, as is the case in *MLL-AF4* infant ALL patients.

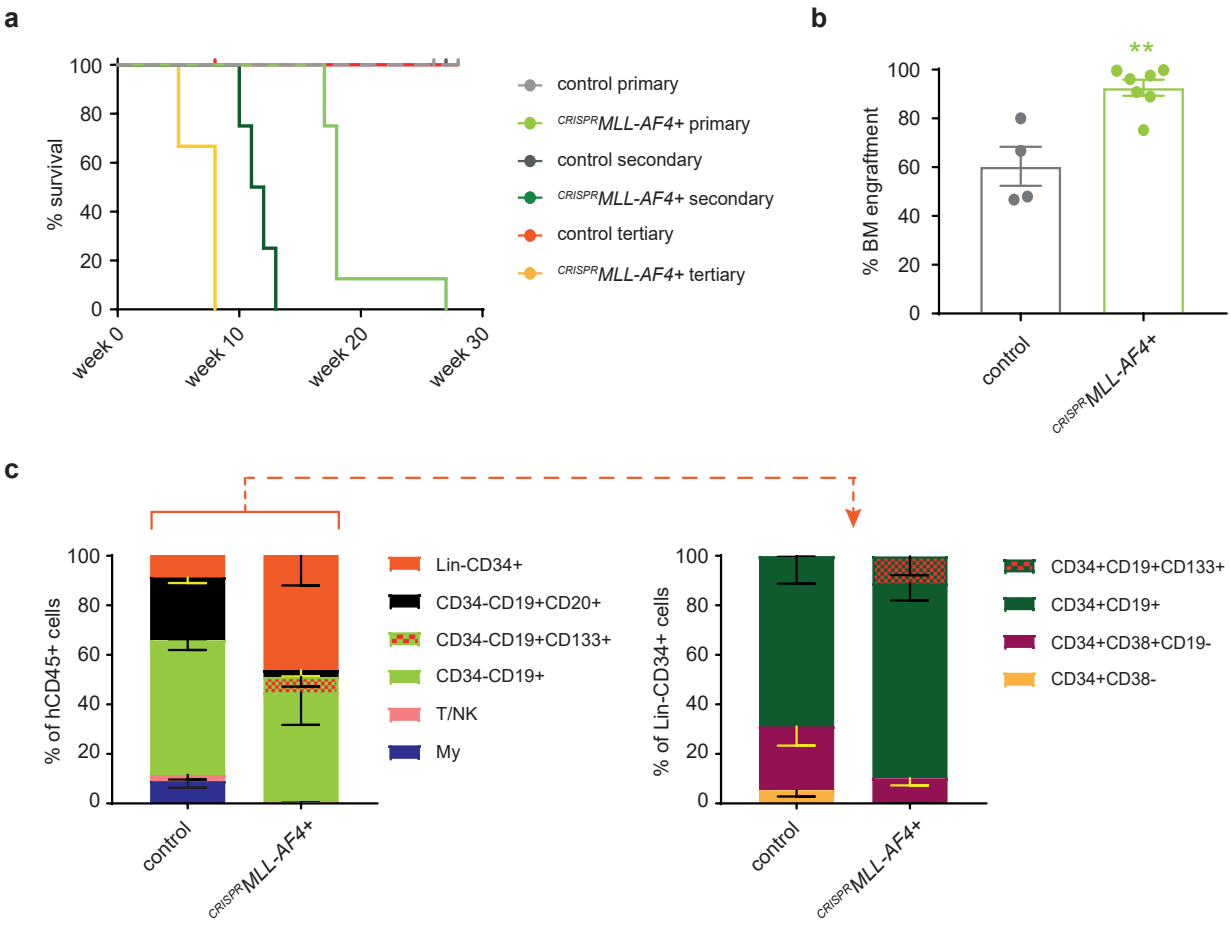


Figure 3.9: *CRISPR MLL-AF4+* FL cells give rise to an aggressive B-ALL *in vivo*. **a.** Leukaemia-free survival for primary (*CRISPR MLL-AF4+* n=7; control n=5), secondary (*CRISPR MLL-AF4+* n=4; control n=1) and tertiary (*CRISPR MLL-AF4+* n=3; control n=2) recipient mice. Mice culled with no signs of leukaemia (see Section 2.8) are censored (shown as tick above line). Latency significantly reduced for secondary and tertiary *CRISPR MLL-AF4+* compared to primary *CRISPR MLL-AF4+* ($p<0.05$, Log-Rank (Mantel-Cox) test). **b.** BM engraftment of hCD45+ cells at termination in primary *CRISPR MLL-AF4+* (light green; n=7) and control (grey; n=5) recipient mice, quantified as a percentage of all CD45+ cells (mCD45+ and hCD45+). Data shown as mean ± SEM. ** = $p<0.01$ (unpaired t-test). **c.** (left) Quantification of human haematopoietic cell populations in control (n=5) and *CRISPR MLL-AF4+* (n=7) primary recipient mouse BM, shown as a percentage of hCD45+ cells. Data shown as mean ± SEM. T/NK = CD3, CD56. My = CD14, CD16, CD235a (erythroid cells excluded in CD45+ gate). (right) Quantification of haematopoietic progenitor cells in control (n=5) and *CRISPR MLL-AF4+* (n=7) primary recipient mouse BM, shown as a percentage of Lin-CD34+ cells (orange population in left-hand plot). Data shown as mean ± SEM. Lin = CD3, CD14, CD16, CD56, CD235a.

6/7 ^{CRISPR}*MLL-AF4+* mice had the CD19+CD10-CD20-IgM/IgD- proB ALL immunophenotype most commonly seen in *MLL-AF4* infant ALL (Figure 3.10a). In contrast, one ^{CRISPR}*MLL-AF4+* primary recipient mouse (the only recipient mouse for 1/3 edited FL samples (Table 2.3) had a preB ALL immunophenotype, with the majority of cells being CD19+CD10+CD20-IgM/IgD- (Figure 3.10b). Secondary (n=4 mice from n=2 primary recipients (Table 2.3) and tertiary (n=3 mice from n=2 secondary recipients (Table 2.3)) all developed B-ALL with significantly reduced latency compared to primary recipients (median latency 11.5 weeks port-transplant ($p<0.02$) in secondary and 8 weeks post-transplant ($p<0.03$) in tertiary) (Figure 3.9a). The proportion of CD34+ blasts varied between primary and secondary recipients, as well as among secondary recipients from the same primary ^{CRISPR}*MLL-AF4+* mouse donor, which may be a consequence of different proportions of CD34+ blasts being transplanted to each secondary recipient, as these mice were transplanted with whole ^{CRISPR}*MLL-AF4+* primary mouse BM. Interestingly, while one secondary recipient mouse for the primary ^{CRISPR}*MLL-AF4+* preB ALL maintained the predominantly CD10+ preB immunophenotype of the primary mouse, the other had 50% CD19+CD10- proB blasts.

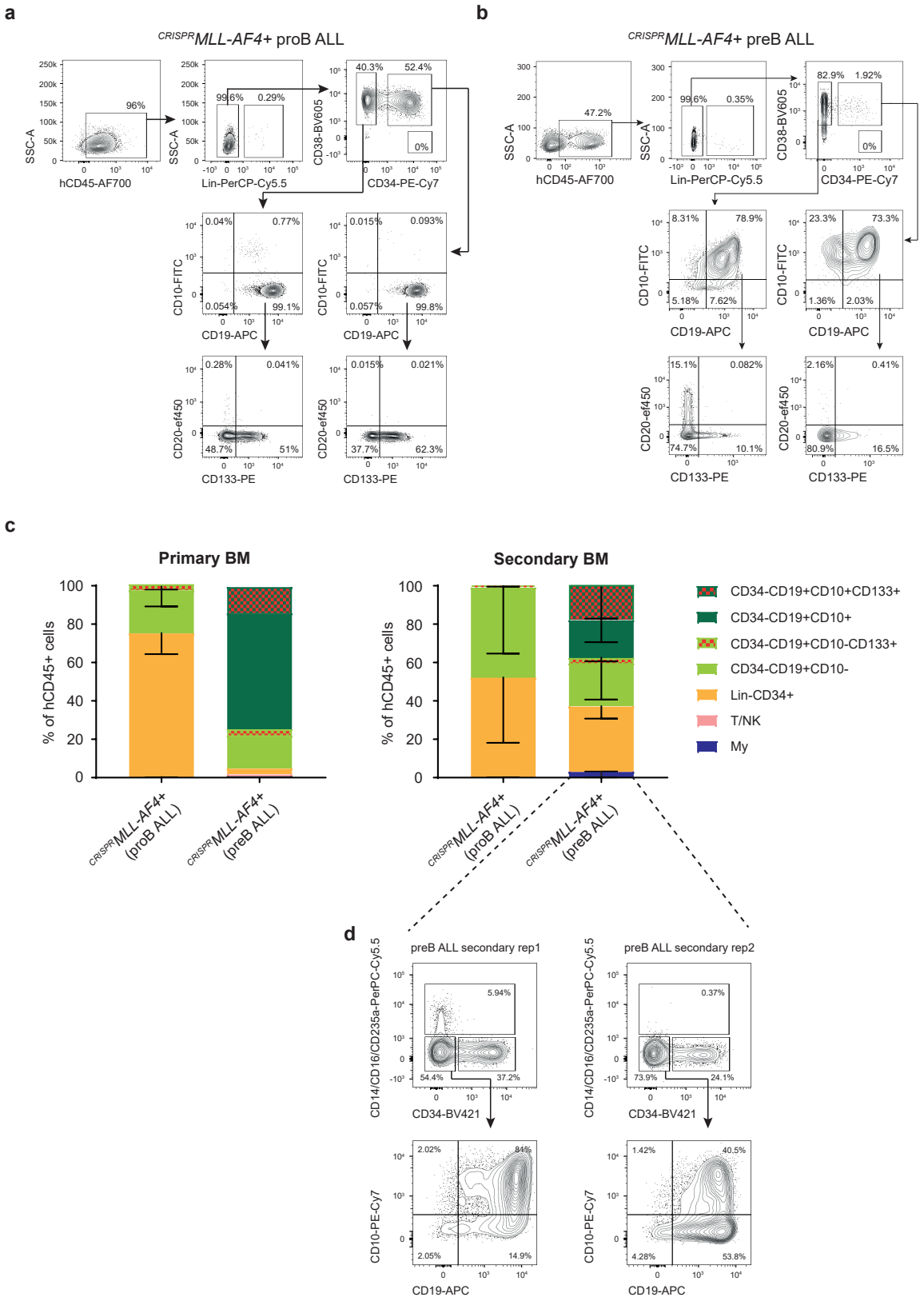

a
CRISPR MLL-AF4+ proB ALL
SSC-A
hCD45-AF700
96%
Lin-PerCP-Cy5.5
99.6%
0.29%
CD38-BV605
CD34-PE-Cy7
40.3%
52.4%
0%
CD10-FITC
CD19-APC
0.04%
0.77%
99.1%
0.015%
0.093%
99.8%
CD20-ef450
CD133-PE
0.28%
0.041%
48.7%
51%
0.057%
0.021%
37.7%
62.3%
b
CRISPR MLL-AF4+ preB ALL
47.2%
99.6%
0.35%
82.9%
1.92%
8.31%
78.9%
5.18%
7.62%
23.3%
73.3%
1.36%
2.03%
15.1%
0.082%
74.7%
10.1%
2.16%
0.41%
80.9%
16.5%
c
Primary BM
Secondary BM
% of hCD45+ cells
CRISPR MLL-AF4+ (proB ALL)
CRISPR MLL-AF4+ (preB ALL)
CD34-CD19+CD10+CD133+
CD34-CD19+CD10+
CD34-CD19+CD10-CD133+
CD34-CD19+CD10-
Lin-CD34+
T/NK
My
d
preB ALL secondary rep1
preB ALL secondary rep2
CD14/CD16/CD235a-PerPC-Cy5.5
CD34-BV421
5.94%
54.4%
37.2%
0.37%
73.9%
24.1%
CD10-PE-Cy7
CD19-APC
2.02%
84%
2.05%
14.9%
1.42%
40.5%
4.28%
53.8%

Figure 3.10

Figure 3.10: ^{CRISPR}*MLL-AF4+* FL cells can give rise to both proB and preB ALL *in vivo*. **a.** Representative flow cytometry plots for ^{CRISPR}*MLL-AF4+* proB ALL primary recipient mouse BM on termination. Arrows represent sequential gating strategy. Percentage of parent population in each gate is shown in each sequential plot. **b.** Representative flow cytometry plots for ^{CRISPR}*MLL-AF4+* preB ALL primary recipient mouse BM on termination. Arrows represent sequential gating strategy. Percentage of parent population in each gate is shown in each sequential plot. **c.** (left) Quantification of human haematopoietic cell populations in ^{CRISPR}*MLL-AF4+* proB ALL (n=6) and ^{CRISPR}*MLL-AF4+* preB ALL (n=1) primary recipient mouse BM, shown as a percentage of hCD45+ cells. Data shown as mean ± SEM. (right) Quantification of human haematopoietic cell populations in ^{CRISPR}*MLL-AF4+* proB ALL (n=2) and ^{CRISPR}*MLL-AF4+* preB ALL (n=2) primary recipient mouse BM, shown as a percentage of hCD45+ cells. Data shown as mean ± SEM. T/NK = CD3, CD56. My = CD14, CD16, CD235a (erythroid cells excluded in CD45+ gate). Lin = CD3, CD14, CD16, CD56, CD235a. **d.** Representative flow cytometry plots for the BM of 2 ^{CRISPR}*MLL-AF4+* preB ALL secondary recipient mice on termination. Arrows represent sequential gating strategy. Percentage of parent population in each gate is shown in each sequential plot.

Sarah Inglott, Great Ormond Street Hospital for Children, London, carried out flow cytometry analysis of primary ^{CRISPR}*MLL-AF4+* proB ALL alongside an *MLL-AF4* infant ALL patient using an ALL diagnostic panel and found that the model faithfully recapitulated the immunophenotype of *MLL-AF4* infant ALL, with heterogeneous expression of CD133, NG2 and CD24 (Figure 3.11).

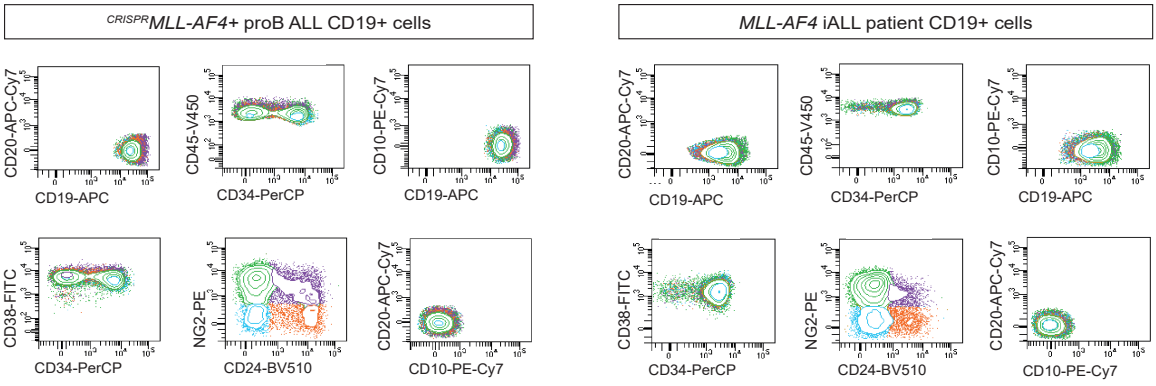


Figure 3.11: ^{CRISPR}*MLL-AF4+* proB ALL recapitulates the immunophenotype of *MLL-AF4* infant ALL. (left) Representative flow cytometry plots for a primary ^{CRISPR}*MLL-AF4+* proB ALL mouse BM on termination. Dots are coloured based on their CD34/NG2 expression status. (right) Representative flow cytometry plots for a *MLL-AF4* infant ALL patient at diagnosis. Dots are coloured based on their CD34/NG2 expression status. Data generously provided by Sara Inglott, Great Ormond Street Hospital for Children, London.

3.7 CRISPR/Cas9-induced off-target editing was not detected in ^{CRISPR}*MLL-AF4+* ALL

I next wanted to validate that ^{CRISPR}*MLL-AF4+* ALL had been caused by the CRISPR/Cas9-induced *MLL-AF4* /t(4;11) translocation alone. FISH analysis confirmed the presence of a heterozygous *MLL-AF4* /t(4;11) translocation in >87% of cells in ^{CRISPR}*MLL-AF4+* mice (n=3, range = 87-99%) (Figure 3.12a). Karyotyping further showed that no other major structural abnormalities had been caused by the CRISPR/Cas9 editing (n=3) (Figure 3.12b). However, it did highlight a previously unidentified constitutional der(14;21) translocation in 1/3 FL donor samples, accounting for 2/7 ^{CRISPR}*MLL-AF4+* mice (both proB ALLs) and 1/5 control mice. Sanger sequencing of the *MLL-AF4* and *AF4-MLL* breakpoints (n=2) showed that there were no large indels at either breakpoint (Figure 3.12c). Sanger sequencing of the remaining wild-type *MLL* allele showed that, while deletions had occurred in intron 11 (n=3, range = -1bp to -16bp) (Figure 3.12d), they were not large enough to disrupt transcription of *MLL* by, for example, changing the sequence at splice junctions. Additional FISH analysis of the next most common *MLL*-rearrangements (*MLL-ENL, MLL-AF9, MLL-AF6* and *MLL-AF10*) further showed that *MLL* had not translocated with any other fusion partner in ^{CRISPR}*MLL-AF4+* cells (n=3) (Figure 3.12e). Finally, Sanger sequencing of the 9 *in silico*-predicted off-target sites for both the *MLL*- and *AF4*-targeting sgRNAs that had <4 mismatches from the target sequence found that no off-target editing had occurred at any of these sites (Table 3.1).

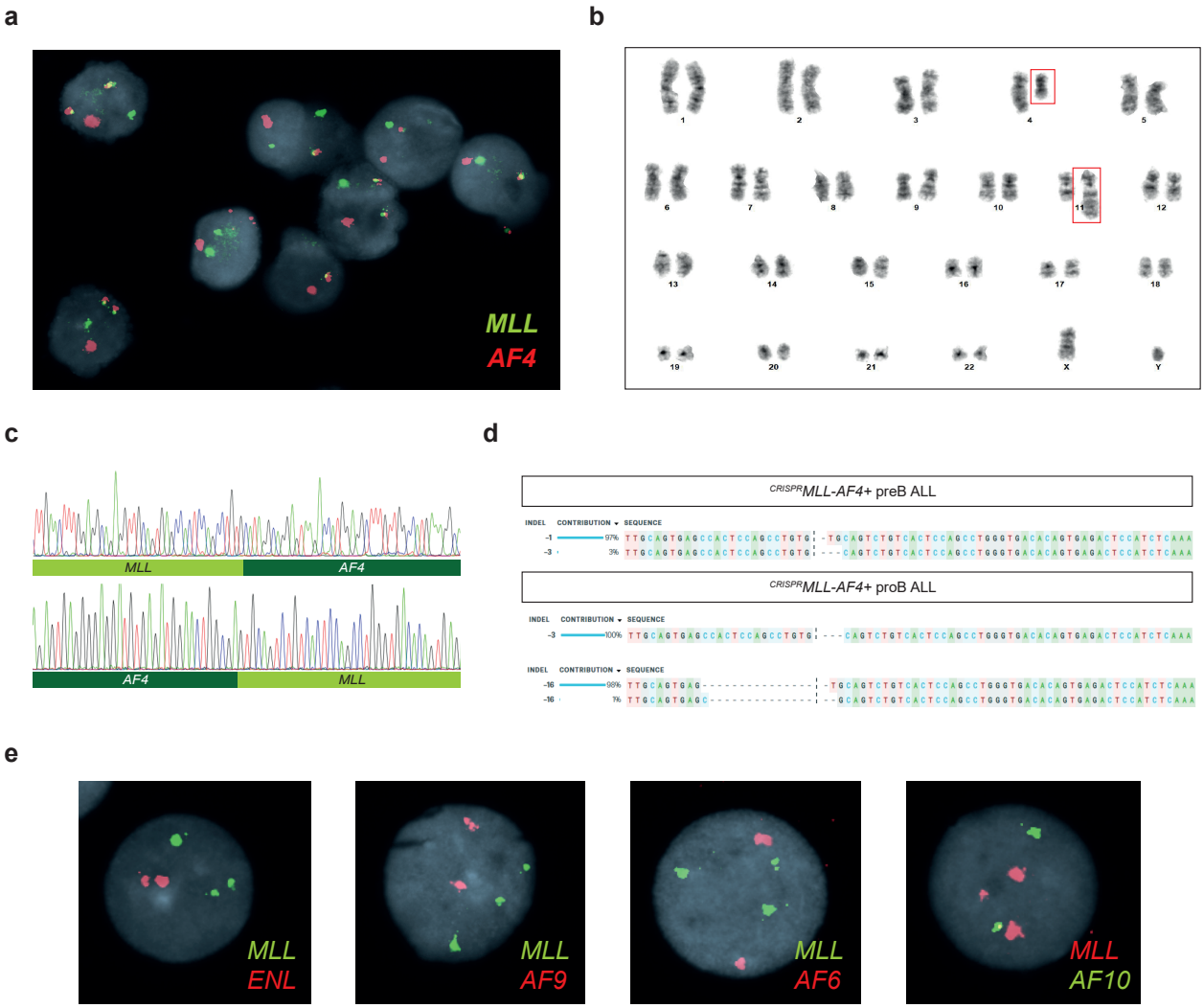


Figure 3.12: No off-target CRISPR/Cas9 edits were detected in *CRISPR*MLL-AF4+ ALL. a. Representative *MLL-AF4* FISH image showing multiple nuclei for *CRISPR*MLL-AF4+ ALL spleen. Data generously provided by Dariusz Ladon, Great Ormond Street Hospital for Children, London. **b.** Representative karyotype data for *CRISPR*MLL-AF4+ ALL. The translocated chromosomes are shown in red boxes. Data generously provided by Dariusz Ladon, Great Ormond Street Hospital for Children, London. **c.** Representative Sanger sequencing traces at the *MLL-AF4* and *AF4-MLL* breakpoints. In this example, no indels were detected at the breakpoint. **d.** Sanger sequencing data showing indels present in the wild-type allele of *MLL* in each *CRISPR*MLL-AF4+ ALL (n=3) compared to matched original, unedited FL cells. Contribution = percentage of cells with each indel. Deletions are shown as horizontal dashed lines in the sequence. The CRISPR/Cas9 cut sites is shown as a vertical dashed line in the sequence. **e.** Representative FISH for the top 4 other *MLL* translocations (*MLL-ENL, MLL-AF9, MLLAF6* and *MLL-AF10*) in the *CRISPR*MLL-AF4+ ALL spleen (n=3). 3 signals can be detected for *MLL* in each case due to the presence of the *MLL-AF4*/t(4;11) translocation. Data generously provided by Dariusz Ladon, Great Ormond Street Hospital for Children, London.

Table 3.1: Analysis of off-target editing in ^{CRISPR}*MLL-AF4+* ALL.

			% of Sanger sequencing reads matching wild-type, unedited sequence		
gene name	**PAM location**	**sgRNA**	**primary ^{CRISPR}*MLL-AF4+* ALL 1 (preB)**	**primary ^{CRISPR}*MLL-AF4+* ALL 2 (preB)**	**primary ^{CRISPR}*MLL-AF4+* ALL 3 (preB)**
SMR3A	intron	MLL	100%	100%	100%
CTNNDA	intron	MLL	100%	100%	100%
PARVA	intron	MLL	100%	100%	100%
KDM6B	intron	MLL	100%	100%	100%
DCC	intron	MLL	100%	100%	100%
TRGC2	intron	MLL	100%	100%	100%
ATRNL1	intron	MLL	100%	100%	100%
FOXO1	intron	AF4	100%	100%	100%
KCNQ2	exon	AF4	100%	100%	100%
MLL wild-type allele	intron	MLL	0% (100% with 1bp deletion)	0% (100% with 3bp deletion)	1% (99% with 16bp deletion)

3.8 ^{CRISPR}*MLL-AF4+* ALL recapitulates key clinico-pathological features of *MLL-AF4* infant ALL

As well as having circulating blasts in the peripheral blood, ^{CRISPR}*MLL-AF4+* mice were found to have blast infiltration into the spleen, leading to splenomegaly (Figure 3.13a). hCD45+ cell engraftment was significantly greater in the spleen of ^{CRISPR}*MLL-AF4+* mice compared to controls (Figure 3.13b). While the majority of hCD45+ cells in control mouse spleens were CD19+CD20+ mature B cells, as would be expected, the immunophenotype of hCD45+ cells in the spleen of ^{CRISPR}*MLL-AF4+* mice matched that of the blasts found in the BM (Figure 3.9c), including a large proportion (50% on average; range = 2.5-70%) of CD19+CD34+ cells (Figure 3.13c). Blast infiltration was also observed in the liver, recapitulating the hepatomegaly that is common in infant ALL (Figure 3.13d). CNS infiltration is a key feature of *MLL*-rearranged infant ALL, which can lead to relapses involving the CNS, either alone or in addition to BM relapse, in about 25% of patients [Pieters et al., [2007], Pieters et al., [2019]]. Histopathology analysis was carried out on ^{CRISPR}*MLL-AF4+* and control mice in the laboratory of Prof Christina Halsey, Wolfson Wohl Cancer Research Centre, University of Glasgow, to determine whether this CNS phenotype was recapitulated by ^{CRISPR}*MLL-AF4+* mice. Striking parameningeal blast cell infiltration was observed in ^{CRISPR}*MLL-AF4+* mice but not in controls (Figure 3.12e), suggesting that these mice had leukaemic infiltration

into the CNS. These infant ALL-like clinico-pathological features were maintained in secondary recipients (Table 3.2).

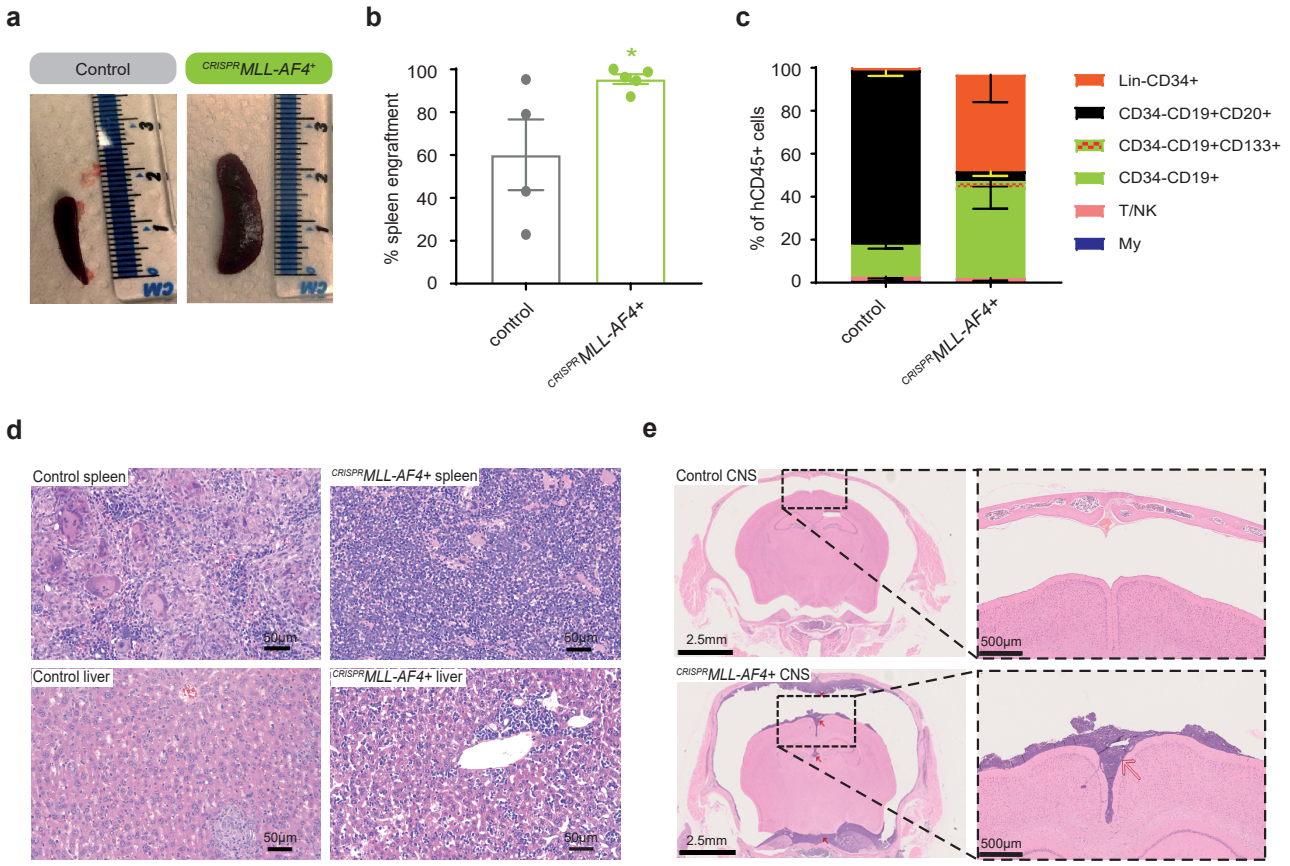


Figure 3.13: *CRISPR MLL-AF4+* ALL recapitulates the phenotypic features of *MLL-AF4* infant ALL. a. Representative images of the spleens of control and *CRISPR MLL-AF4+* primary recipient mice on termination. **b.** Spleen engraftment of hCD45+ cells at termination in primary *CRISPR MLL-AF4+* (light green; n=7) and control (grey; n=5) recipient mice, quantified as a percentage of all CD45+ cells (mCD45+ and hCD45+). Data shown as mean ± SEM. ** = p<0.01 (unpaired t-test). **c.** Quantification of human haematopoietic cell populations in control (n=5) and *CRISPR MLL-AF4+* (n=7) primary recipient mouse spleens, shown as a percentage of hCD45+ cells. Data shown as mean ± SEM. T/NK = CD3, CD56. My = CD14, CD16, CD235a (erythroid cells excluded in CD45+ gate). Lin = CD3, CD14, CD16, CD56, CD235a. **d.** Representative H&E staining of spleen and liver from control and *CRISPR MLL-AF4+* primary recipient mice. Scale bar = 50µm. Tissue embedding, sectioning and staining carried out by the Translational Histopathology Laboratory (THL), CRUK Oxford Cancer Centre, University of Oxford. **e.** Representative H&E staining of control and *CRISPR MLL-AF4+* primary recipient mouse heads (scale bar = 2.5mm; red arrows = regions of concentrated blast cell infiltration). High magnification images (scale bar = 500µm) highlight striking parameningeal blast cell infiltration in *CRISPR MLL-AF4+* mice (red arrow) but not control. Data generously provided by Prof Christina Halsey, Wolfson Wohl Cancer Research Centre, University of Glasgow.

Table 3.2: Summary of *in vivo* experiments.

	Primary		Secondary		Tertiary	
	control	^{CRISPR}*MLL-AF4+*	control	^{CRISPR}*MLL-AF4+*	control	^{CRISPR}*MLL-AF4+*
Engrafted	5/5	7/7	1/1	4/4	2/2	3/3
ALL	0/5	7/7	0/1	4/4	0/2	3/3
Median latency	n/a	18 weeks	n/a	11.5 weeks	n/a	8 weeks
BM engraftment	5/5	7/7	1/1	4/4	2/2	3/3
Splenomegaly	0/5	7/7	0/1	4/4	0/2	3/3
CNS infiltration (/of those tested)	0/1	2/2	0/1	3/3	ND	ND
Pred resistant *in vitro* (LC50 ≥100µg/ml)	n/a	1/1	n/a	2/2	ND	ND
L-asp resistant *in vitro* (LC50 >0.1 IU/ml)	n/a	1/1	n/a	3/3	ND	ND
proB (CD19+CD10-CD20-IgM/IgD-)	n/a	6/7	n/a	2/4	n/a	3/3
preB (CD19+CD10+CD20-IgM/IgD-)	n/a	1/7	n/a	2/4	n/a	0/3
CD34 positive (>20% of blasts)	n/a	ProB (5/6); PreB (0/1)	n/a	ProB (1/2); PreB (2/2)	n/a	ND
Karyotype	ND	FL sample 1 and 3: 46 XY t(4;11); FL sample 2: 45 XY t(4;11) der(14;21)(q10;q10)	ND	ND	ND	ND
MLL-AF4 /t(4;11) positive by FISH	n/a	80-97% (n=3)	ND	ND	ND	ND
MLL-AF4 expression	0/2	7/7	ND	ND	ND	ND
AF4-MLL expression	0/2	7/7	ND	ND	ND	ND
VDJ rearrangement (/of those tested)	ND	2/3 (ProB and PreB) clonal; 1/3 (ProB) non-rearranged	ND	ND	ND	ND

Finally, as chemo-resistance is an important feature of *MLL*-rearranged infant ALL that contributes to its poor prognosis [Hilden et al., [2006], Pieters et al., [2007], Pieters et al., [2019]], I compared the response of ^{CRISPR}*MLL-AF4+* ALL blasts to prednisolone and L-asparaginase with ALL PDX models (*ETV6-RUNX1+* and *MLL*-rearranged childhood ALLs, kindly provided by Prof Christina Halsey, Wolfson Wohl Cancer Research Centre, University of Glasgow) and the SEM (*MLL-AF4+*) and KOPN8 (*MLL-ENL+*) cell lines *in vitro*. Cells were cultured in increasing concentration of each drug for 48-96 hours and the proportion of dead cells was quantified using the MTT assay according to a previously described protocol [Pieters et al., [1990], Pieters et al., [1991], Pieters et al., [1994], Ramakers-van Woerden et al., [2004]]. These data were used to calculate the LC50 values (the drug concentration at which 50% of the cells die) for each sample (Figure 3.14a). ^{CRISPR}*MLL-AF4+* ALL and the SEM cell line showed a similar level of *in vitro* resistance to both prednisolone (LC50 of ≥100µg/ml) (Figure 3.14b) and L-asparaginase (LC50 of >1IU/ml) (Figure 3.14c) to previous reports of treatment-resistant patient samples [Ramakers-van Woerden et al., [2004]]. *MLL*-rearranged childhood ALL PDXs showed slightly reduced levels of resistance to prednisolone but were both L-asparaginase-resistant (Figure 3.14b and 3.14c). *ETV6-RUNX1+* PDXs were used as a control because *MLL*wt childhood ALL have been shown to be more chemo-sensitive both *in vitro* and *in vivo* compared to *MLL*-rearranged

childhood ALL [Ramakers-van Woerden et al., [2004]]. *ETV6-RUNX1+* PDXs were less prednisolone-resistant compared to ^{CRISPR}*MLL-AF4+* ALL (Figure 3.14b), and showed varied responses to L-asparaginase that on average were equivalent to all other samples assayed (Figure 3.14c).

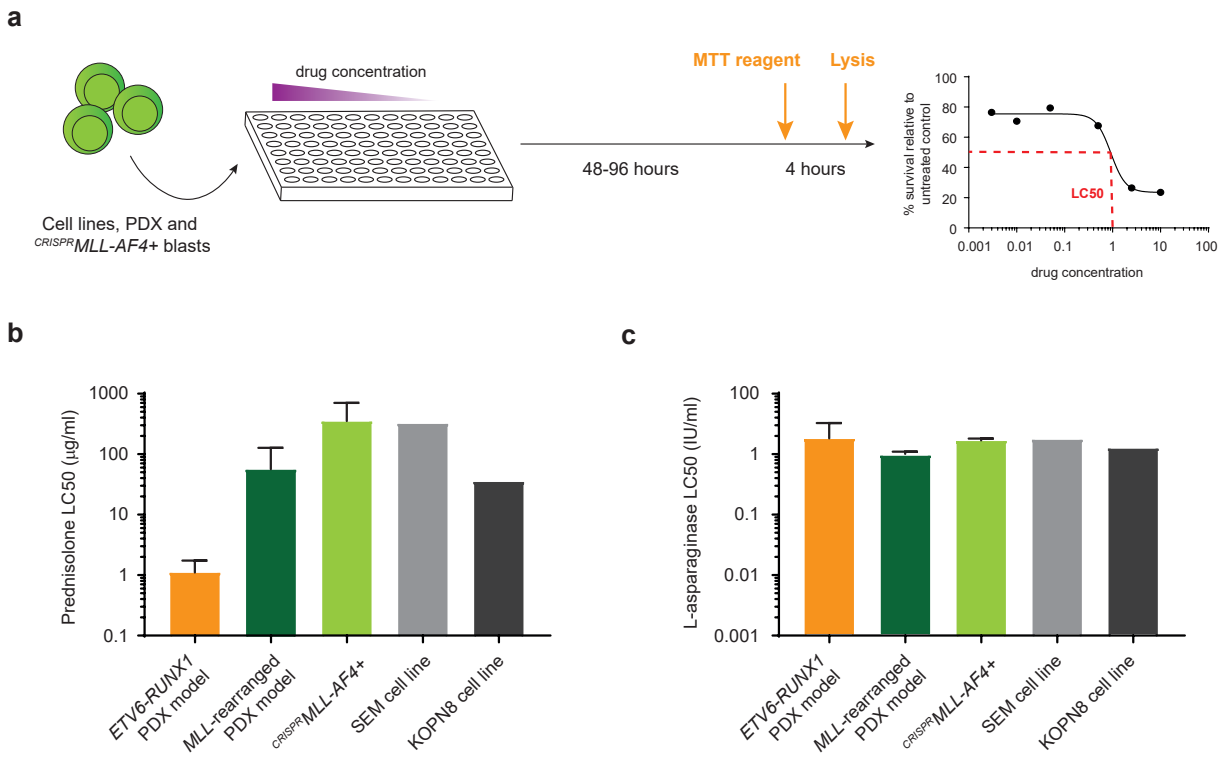


Figure 3.14: ^{CRISPR}*MLL-AF4+* ALL shows chemo-resistance *in vitro*. **a.** Schematic representation of *in vitro* drug sensitivity assay. ^{CRISPR}*MLL-AF4+* blasts, PDX-derived blasts (*ETV6-RUNX1+* and *MLL*-rearranged childhood ALL) and cell lines (SEM and KOPN8) were plated into multiple wells of a 96-well plate, with each well containing an increasing concentration of either prednisolone or L-asparaginase. DMSO (prednisolone) or water (L-asparaginase) was used as an untreated control for each sample. Cells were incubated with drug for 48 (primary blasts) or 96 (cell lines) hours, then incubated for a further 4 hours with MTT reagent before lysis and viability (purple formazan) readout using a spectrophotometer. Cell viability relative to untreated control at each drug concentration was plotted in order to calculate the LC50 (drug concentration at which 50% of cells die). Representative data for *MLL*-rearranged childhood ALL treated with L-asparaginase shown. **b.** LC50 values for *EVT6-RUNX1+* childhood ALL (n=4), *MLL*-rearranged childhood ALL (n=2), ^{CRISPR}*MLL-AF4+* ALL (n=3), SEM cells (n=1) and KOPN8 cells (n=1) treated with prednisolone. Data shown as mean ± SEM. **c.** LC50 values for *EVT6-RUNX1+* childhood ALL (n=3), *MLL*-rearranged childhood ALL (n=2), ^{CRISPR}*MLL-AF4+* ALL (n=4), SEM cells (n=1) and KOPN8 cells (n=1) treated with L-asparaginase. Data shown as mean ± SEM.

3.9 Discussion

In this chapter, I have developed a method to induce an *MLL-AF4*/t(4;11) translocation in primary human FL HSPCs, which is capable of driving leukaemic transformation both *in vitro* and *in vivo*.

In vitro, the CRISPR/Cas9-induced *MLL-AF4*/t(4;11) translocation leads to a dramatic expansion of B lineage cells by week 2-3 of MS-5 stromal co-culture. However, these were virtually all CD19+CD10+ preB cells, unlike the CD19+CD10- proB ALL immunophenotype that is much more commonly observed in patients [Basso et al., [1992], Borkhardt et al., [2002]]. This was true even for FL samples that gave rise to proB ALL *in vivo*. These data are in-keeping with previous observations that MS-5 stromal co-culture results in differentiation toward a CD19+CD10+ preB immunophenotype [Richardson et al., [2021]]. Moreover, by week 5-6, the total number of hCD45+ cells in ^{CRISPR}*MLL-AF4+* cultures began to decline. Together, these data suggest that, while MS-5 stromal co-culture provides a clear, short-term readout of ^{CRISPR}*MLL-AF4*-driven leukaemic transformation, it is not optimal for long-term maintenance of ^{CRISPR}*MLL-AF4+* cells. In an interesting parallel, primary ALL blasts are incredibly difficult to culture *ex vivo* [Boutter et al., [2014], Pal et al., [2016]] but can be maintained and expanded in PDX models [le Viseur et al., [2008], Bardini et al., [2015]]. This suggests that the growth and survival of ^{CRISPR}*MLL-AF4+* cells may depend on as yet unidentified factors present in the *in vivo* BM microenvironment that are not present in the MS-5 stromal co-culture system. For example, a variety of other cell types present in the BM milieu have been suggested to play a key role in the survival of B-ALL blasts [Dander et al., [2021]].

In vivo, ^{CRISPR}*MLL-AF4+* cells gave rise to an aggressive B-ALL in all recipient mice, which recapitulated key infant ALL-like clinico-pathological features, including circulating blasts in the peripheral blood, blast infiltration into the liver, spleen and CNS, and *in vitro*-assayed chemo-resistance. Therefore, ^{CRISPR}*MLL-AF4+* ALL provides a previously lacking, preclinical model for translational studies that specifically recapitulates poor prognosis infant-ALL. For example, the ability of novel treatments to target steroid-resistant blasts or to eradicate blasts from the

CNS are important considerations in the case of infant ALL, and these can now be tested in ^{CRISPR}*MLL-AF4+* ALL.

^{CRISPR}*MLL-AF4+* ALL showed some immunophenotypic heterogeneity between recipient mice. Most strikingly, 1/7 ^{CRISPR}*MLL-AF4+* ALL had preB ALL immunophenotype, while the other 6/7 ^{CRISPR}*MLL-AF4+* ALLs had a proB immunophenotype. Interestingly however, no other significant clinical differences were observed between proB and preB ^{CRISPR}*MLL-AF4+* ALL, and both proB and preB ^{CRISPR}*MLL-AF4+* ALL recapitulated key infant ALL-like features. ^{CRISPR}*MLL-AF4+* mice also differed in the proportion of blasts that co-expressed CD34 and CD133, both of which are found to be variably expressed in *MLL-AF4* infant ALL patients [Borkhardt et al., [2002], Godfrey et al., [2020]]. In this chapter, I have induced the *MLL-AF4* /t(4;11) translocation in FL CD34+ cells, which represent a mixture of different HSPC types. Based on these data, I hypothesise that the immunophenotypic heterogeneity observed between ^{CRISPR}*MLL-AF4+* mice may be a consequence of the translocation occurring in different progenitor cell types.

The expression of both reciprocal fusion genes, *MLL-AF4* and *AF4-MLL*, was detected in all ^{CRISPR}*MLL-AF4+* cells. As 20% of *MLL-AF4* ALL patients do not express *AF4-MLL*, it is hypothesised not to be required for leukaemic transformation [Downing et al., [1994]]. However, the role that this reciprocal fusion gene may play in the 80% of *MLL-AF4* ALL patients that do express it is a topic of debate in the field and requires further study [Bursen et al., [2010], Benedikt et al., [2011], Kumar et al., [2011], Sanders et al., [2011], Wilkinson et al., [2013]]. This CRISPR/Cas9-based method of inducing the *MLL-AF4* /t(4;11) translocation seems unlikely to be useful for such studies. In the future, it may be possible to adapt this method to simultaneously induce the translocation and silence one of the fusion genes. For example, if new sgRNAs were designed in exonic regions of *MLL* and *AF4*, homology-directed repair could be used to introduce a premature STOP codon in either of the fusion genes.

Taken together, these data suggest that inducing an *MLL-AF4* /t(4;11) translocation in primary human fetal cells is sufficient to give rise to an aggressive leukaemia

that recapitulates the phenotypic features of infant ALL. What is not clear from these data is whether a similar gene editing approach in post-natal cells would generate a similar proB ALL model. The natural history and molecular profile of *MLL-AF4* ALL in older patients suggests that the fetal cell context is likely to be the crucial factor that allowed the successful recapitulation of infant-like *MLL-AF4* B-ALL in our model. Therefore, I hypothesise that the distinct molecular profile of fetal HSPCs compared to post-natal HSPCs is likely to play a key role in their ability to give rise to an infant ALL-like disease, and that this fetal-specific molecular profile is likely maintained to some extent in *MLL-AF4* infant ALL patients.

4

^{CRISPR}*MLL-AF4+* ALL recapitulates the molecular profile of *MLL-AF4* infant ALL

Contents

4.1 Introduction 82
4.1.1 Distinct gene expression profiles in *MLL*-rearranged infant ALL 82
4.1.2 Fetal gene expression programmes in leukaemia 83
4.2 Aims 85
4.3 ^{CRISPR}*MLL-AF4+* ALL recapitulates the gene expression profile of *MLL-AF4* ALL patients 86
4.3.1 Known *MLL*-rearranged leukaemia-associated genes are upregulated in ^{CRISPR}*MLL-AF4+* ALL 86
4.3.2 ^{CRISPR}*MLL-AF4+* ALL is transcriptomically similar to *MLL-AF4* ALL patients 89
4.3.3 ^{CRISPR}*MLL-AF4+* ALL recapitulates the poor prognosis subset of *HOXA*^{lo} *MLL-AF4* ALL 93
4.4 ^{CRISPR}*MLL-AF4+* ALL recapitulates the chromatin profile of *MLL-AF4* ALL 94
4.4.1 The MLL-AF4 binding profile of ^{CRISPR}*MLL-AF4+* ALL recapitulates that of *MLL-AF4* ALL patients 94
4.4.2 MLL-AF4 binding leads to increased H3K79me2 deposition during ^{CRISPR}*MLL-AF4*-induced leukaemic transformation 98
4.5 ^{CRISPR}*MLL-AF4+* ALL specifically recapitulates the molecular profile of *MLL-AF4* infant ALL 101
4.5.1 *MLL-AF4* infant ALL is transcriptomically distinct from *MLL-AF4* childhood ALL 101

4.5.2 Fetal gene expression programmes drive the distinct transcriptomic profile of *MLL-AF4* infant ALL 103
4.5.3 Human FL-derived ^{CRISPR}*MLL-AF4+* ALL specifically recapitulates *MLL-AF4* infant ALL 106
4.6 Discussion 109

4.1 Introduction

By inducing an *MLL-AF4*/t(4;11) translocation in primary human FL HSPCs, I have generated a model, ^{CRISPR}*MLL-AF4+* ALL, that specifically recapitulates the phenotypic features of *MLL-AF4* infant ALL, suggesting that the fetal cell context in which the translocation first arises plays a key role in driving the distinct phenotype of infant ALL. In order to more rigorously test this idea, I set out to derive one of the first human fetal-specific gene expression signatures, and to determine to what extent this may contribute to the unique phenotype of *MLL-AF4* infant ALL. Therefore, in this chapter, I first ask whether *MLL-AF4* infant ALL has a distinct molecular profile compared to *MLL-AF4* childhood ALL, and define a set of infant ALL-specific genes. Next, I compare FL HSPCs to adult BM HSPCs to define a set of human fetal-specific genes and ask to what extent the expression of these fetal-specific genes may drive the infant ALL phenotype. Finally, I test whether ^{CRISPR}*MLL-AF4+* ALL is more successful than the best previous humanised *MLL-Af4* ALL mouse model [Lin et al., [2016]] at recapitulating this infant ALL-specific molecular profile. The data presented in this chapter have recently been published [Rice et al., [2021]].

4.1.1 Distinct gene expression profiles in *MLL*-rearranged infant ALL

MLL-rearranged infant ALL is not a molecularly homogeneous subset of patients. Several genes have been found to have varied expression levels among *MLL*-rearranged infant ALL patients, which can have a significant impact on prognosis. For example, high *MCL-1* expression is associated with increased resistance to

prednisone treatment both *in vitro* and *in vivo* [Stam et al., [2010]]. In addition, high level expression of wild-type *FLT3* correlates with poor outcomes for *MLL*-rearranged infant ALL patients [Armstrong et al., [2003], Stam et al., [2007]]. Based on promising preclinical data [Armstrong et al., [2003], Brown et al., [2005]], the COG trial AALL0631 with 218 participants incorporated lestaurtinib, a FLT3 inhibitor, along with chemotherapy into upfront treatment of *MLL*-rearranged infant ALL but, unfortunately, the addition of lestaurtinib in this trial did not demonstrate a benefit (EFS of 38.8% in the control arm versus 35.82% in treatment arm)[Brown et al., [2021]].

The best characterised molecular subsets of *MLL*-rearranged infant ALL are the *HOXA*$^{\text{hi}}$ and *HOXA*$^{\text{lo}}$ subsets. Using microarray data, Stam et al., [2010] found that *MLL-AF4* infant ALL patients could be subdivided into 2 groups. In one group, the cluster of *HOXA* genes, including *HOXA3, HOXA5, HOXA7, HOXA9* and *HOXA10* were highly expressed. In the other group, the *HOXA* cluster genes were lowly expressed and another set of homeobox genes, *IRX1* and *IRX2*, were highly expressed [Stam et al., [2010]]. The *HOXA*$^{\text{lo}}$ gene expression signature was found to be more common in *MLL*-rearranged infant ALL compared to *MLL*-rearranged childhood ALL [Trentin et al., [2009], Stam et al., [2010], Symeonidou & Ottersbach, [2021]]. Moreover, *HOXA*$^{\text{lo}}$ *MLL-AF4* infant ALL was found to have a worse outcome compared to *HOXA*$^{\text{hi}}$ *MLL-AF4* infant ALL [Stam et al., [2010]]. Therefore, *HOXA*$^{\text{lo}}$ *MLL-AF4* infant ALL represents a subset of *MLL*-rearranged ALL with the worst prognosis.

4.1.2 Fetal gene expression programmes in leukaemia

It is clear that *MLL*-rearranged infant ALL is uniquely aggressive compared to *MLL*-rearranged childhood ALL; and so far, I have shown that FL-derived ^{CRISPR}*MLL-AF4+* ALL recapitulates key infant ALL-specific phenotypic features. I hypothesise that this is a result of distinct cell-intrinsic molecular features of the fetal HSPCs in which the *MLL-AF4*/t(4;11) translocation first arises. While a range of functional differences between fetal and adult HSPCs have been described in both mice and

humans [Lansdorp et al., [1993], Notta et al., [2016], O'Byrne et al., [2019]], much less was known about the gene expression programmes that drive these differences at the time that this project began. However, a number of papers have since been published that shed important light on these molecular changes through ontogeny [Popescu et al., [2019], Roy et al., [2021]].

One well-characterised fetal-specific gene expression programme is the *LIN28B* axis. By comparing FL LSKs to adult BM LSKs in mice, *Lin28b* was found to be among the most significantly upregulated genes in FL [Copley et al., [2013]]. The same was found when mouse FL proB cells were compared to adult BM proB cells [Yuan et al., [2012]]. *LIN28B* inhibits *let7* micro-RNA biogenesis, which in turn leads to upregulation of *let7*-repressed oncogenes such as *HMGA2* and the *IGF2BP* genes (*IGF2BP1, IGF2BP2* and *IGF2BP3*) [Copley et al., [2013]]. When *Lin28b* was ectopically expressed in mouse adult BM HSPCs, they underwent fetal-like B-lymphopoiesis, differentiating into fetal-type B-1a cells [Yuan et al., [2012]].

LIN28B expression has been found to be enriched in a number of infant and childhood cancers, including juvenile myelo-monocytic leukaemia (JMML) [Helsmoortel et al., [2016]], and neuroblastoma, where *LIN28B* expression was predictive of a worse outcome [Molenaar et al., [2012]]. However, it should also be acknowledged that *LIN28B* has been found to be expressed in a range of other malignancies, such as ovarian cancer and hepatocellular carcinoma [Zhou et al., [2013]]. Therefore, while *LIN28B* expression may be maintained in childhood cancers with fetal origins, there is also evidence to suggest in may be reactivated in adult malignancies. In any case, this fetal-specific gene appears to play a role in cancer progression [Viswanathan et al., [2009], Zhou et al., [2013], Meder et al., [2018]].

Despite clear clinico-pathological differences, a systematic analysis of age-related transcriptomic differences between infant ALL and childhood ALL with the same driver mutations, and to what extent the fetal origin of infant ALL may underlie these differences, had not been carried out at the time that I began this project. However, the Ottersbach lab has since explored these questions using primary RNA-seq data [Symeonidou & Ottersbach, [2021], Symeonidou et al., [2021]]. Stam et al.,

[2010] used microarray data to compare *MLL*-rearranged infant ALL to childhood B-ALL, but any age-related differences were obscured by the confounding factors of varied driver mutations and phenotypic differences (the childhood B-ALL patients had a more mature B progenitor phenotype compared to *MLL*-rearranged infant ALL patients). In this chapter, I compare infant ALL and childhood ALL patients with an identical driver mutation, *MLL-AF4*, to derive a novel infant ALL-specific gene expression signature in *MLL*-rearranged ALL; and ask to what extent it is characterised by maintenance of fetal gene expression programmes.

4.2 Aims

The overall aim of this chapter was to use transcriptomic and epigenetic profiling to determine whether ^{CRISPR}*MLL-AF4+* ALL specifically recapitulates the molecular profile of *MLL-AF4* infant ALL, which I split into the following sub-aims:

- Determine whether ^{CRISPR}*MLL-AF4+* ALL recapitulates the gene expression profile of *MLL-AF4* ALL patients
- Determine whether ^{CRISPR}*MLL-AF4+* ALL recapitulates the chromatin profile of *MLL-AF4* ALL patients
- Derive a human fetal HSPC-specific and an *MLL-AF4* infant ALL-specific gene expression signature
- Investigate the potential maintenance of human fetal-specific gene expression programmes in *MLL-AF4* infant ALL
- Test whether FL derived ^{CRISPR}*MLL-AF4+* ALL is a more faithful model of *MLL-AF4* infant ALL than the best previous model (CB derived *MLL-Af4+* ALL) [Lin et al., [2016]]

4.3 ^{CRISPR}*MLL-AF4+* ALL recapitulates the gene expression profile of *MLL-AF4* ALL patients

4.3.1 Known *MLL*-rearranged leukaemia-associated genes are upregulated in ^{CRISPR}*MLL-AF4+* ALL

I first set out to characterise the gene expression profile of ^{CRISPR}*MLL-AF4+* ALL, and how this differed from xenografted control FL cells, to understand the transcriptomic changes underlying leukaemic transformation in ^{CRISPR}*MLL-AF4+* ALL. hCD45+CD19+ cells were sorted from the BM of primary ^{CRISPR}*MLL-AF4+* (n=3; 1 preB ALL and 2 proB ALLs) and control (n=3) mice for bulk RNA-seq analysis (Figure 4.1). CD19 was chosen as the only immunophenotypic marker on which blasts were sorted from ^{CRISPR}*MLL-AF4+* mouse BM because this marker was common to all blasts in all recipient mice. As such, these data represent all CD19+ cells, including all subpopulations defined by CD34, CD133 and, in the case of the preB ALL, CD10 expression (Figure 4.1). hCD45+CD19+ B lineage cells were also sorted from the BM of control mice as I hypothesised that they would represent the best non-leukaemic equivalent of B-ALL blasts.

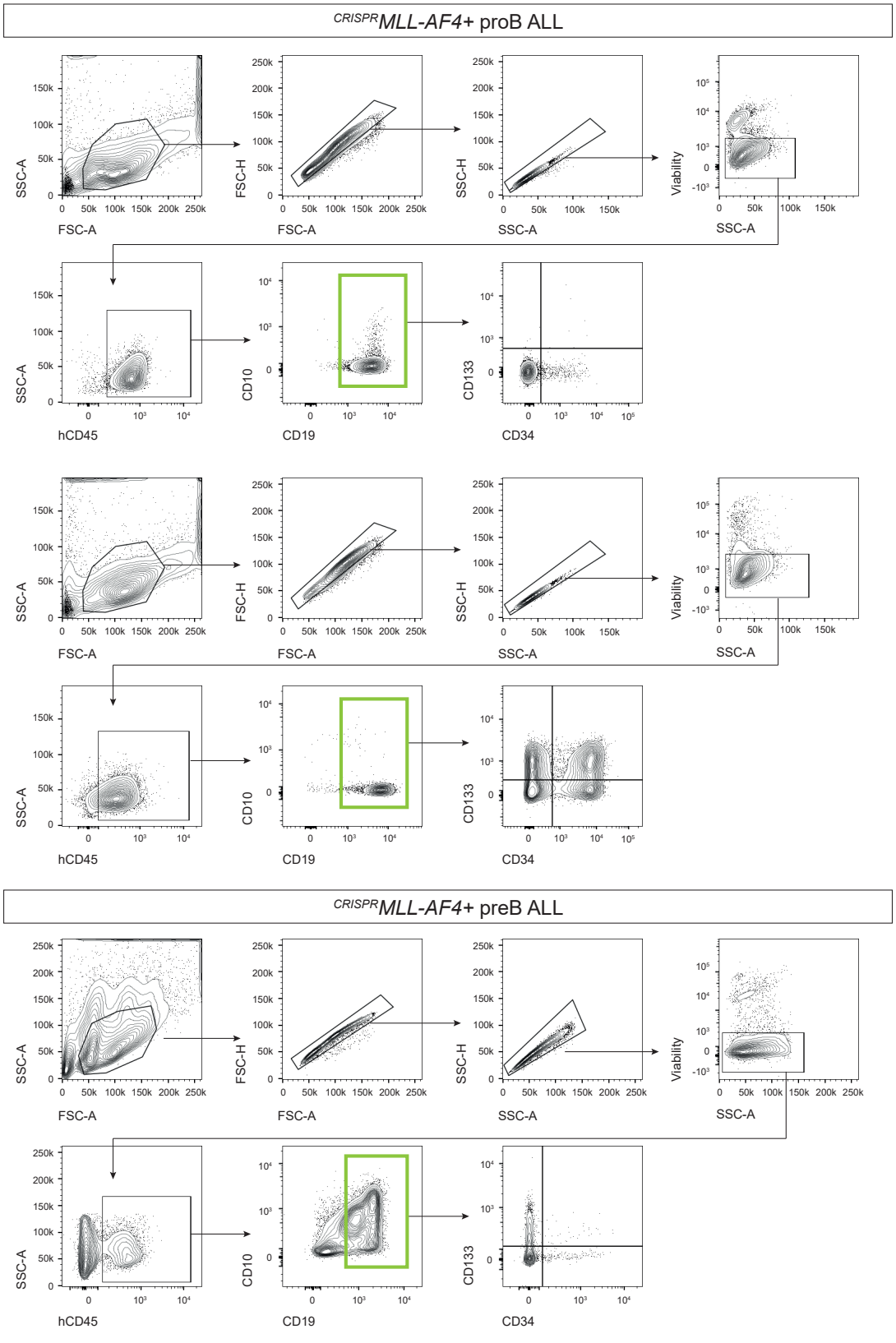

CRISPR MLL-AF4+ proB ALL
SSC-A
FSC-A
FSC-H
SSC-H
Viability
hCD45
CD19
CD10
CD34
CD133
CRISPR MLL-AF4+ preB ALL

Figure 4.1

Figure 4.1: ^{CRISPR}*MLL-AF4+* ALL sort strategy for RNA-seq. Representative flow cytometry plots showing sequential gating strategy, where gates and arrows represent the population taken forward to the subsequent plot, to sort hCD45+CD19+ (green gate) from the BM of primary recipient ^{CRISPR}*MLL-AF4+* mice for two ^{CRISPR}*MLL-AF4+* proB ALLs (top) and one ^{CRISPR}*MLL-AF4+* preB ALL (bottom). Gates were set based on FMO controls.

An initial principal component (PC) analysis showed that ^{CRISPR}*MLL-AF4+* ALL and control FL CD19+ cells clustered away from one another on PC1 (Figure 4.2a), suggesting that the greatest variance in the dataset was explained by the difference between normal and leukaemic cells. This separation was driven by higher expression levels of known stem cell and *MLL*-rearranged leukaemia-associated genes like *MEIS1* in the ^{CRISPR}*MLL-AF4+* ALL cells and by higher expression levels of genes associated with B lineage maturation, such as *MME* that encodes CD10, *MS4A1* that encodes CD20, and *JCHAIN*, in the control FL CD19+ cells (Figure 4.2b). ^{CRISPR}*MLL-AF4+* preB ALL fell mid-way between ^{CRISPR}*MLL-AF4+* proB ALL and control FL CD19+ cells on PC1, which likely reflects the more mature B progenitor phenotype of preB ALL compared to proB ALL, but could also reflect a greater proportion of residual *MLL-AF4-* B cells (see CD19+/CD19^{hi} populations in Figure 3.10b; *MLL-AF4* status of each individual population unknown). However, ^{CRISPR}*MLL-AF4+* preB ALL also clustered separately from control FL CD19+ cells on PC2, suggesting that, while it may show some expression of B lineage maturation markers, ^{CRISPR}*MLL-AF4+* preB ALL is abnormal compared to control FL CD19+ cells (Figure 4.2a).

I next carried out differential expression analysis between ^{CRISPR}*MLL-AF4+* ALL and control FL CD19+ cells to identify 1,068 significantly differentially expressed genes (FDR<0.05) (Figure 4.2c). Genes upregulated in ^{CRISPR}*MLL-AF4+* ALL included previously reported MLL-AF4 target genes such as *MEIS1, FLT3, MEF2C* and *RUNX1* [Kerry et al., [2017], Godfrey et al., [2019], Krivtsov et al., [2019], Harman et al., [2021]], stem cell-associated genes such as *LMO2* and *PROM1*, which encodes CD133, and myeloid lineage-associated genes such as *CSF1R* (Figure 4.2c and 4.2d). Based on this geneset, all ^{CRISPR}*MLL-AF4+* samples (both preB and

proB ALL) clustered together, and separately from control FL CD19+ cells. When visualised as a heatmap, it is clear that the majority of genes up- or downregulated in ^{CRISPR}*MLL-AF4+* proB ALL are also up- or downregulated in ^{CRISPR}*MLL-AF4+* preB ALL. However, there is a subset of genes strongly downregulated in ^{CRISPR}*MLL-AF4+* proB ALL but relatively unchanged in ^{CRISPR}*MLL-AF4+* preB ALL (Figure 4.2c). These genes included *MME, JCHAIN* and *CD79B*, all of which reflect the more mature B progenitor phenotype of ^{CRISPR}*MLL-AF4+* preB ALL (Figure 4.2d). Nonetheless, key *MLL*-rearranged leukaemia genes such as *FLT3, MEF2C* and *RUNX1* showed similar expression levels in all ^{CRISPR}*MLL-AF4+* samples (Figure 4.2d).

4.3.2 ^{CRISPR}*MLL-AF4+* ALL is transcriptomically similar to *MLL-AF4* ALL patients

I next wanted to compare the ^{CRISPR}*MLL-AF4+* ALL RNA-seq data to primary patient datasets to determine whether ^{CRISPR}*MLL-AF4+* ALL faithfully recapitulates the gene expression profile of *MLL-AF4* ALL patient blasts. Using a previously published dataset that included bulk RNA-seq data for samples from *MLL-AF4* infant ALL patients and *MLL*wt infant ALL patients [Agraz-Doblas et al., [2019]], I carried out a 4-way comparison between control FL CD19+ cells, ^{CRISPR}*MLL-AF4+* ALL, *MLL-AF4* infant ALL patient samples and *MLL*wt infant ALL patient samples to identify marker genes that could separate the populations from one another. In this way, I identified 7,041 significantly differentially expressed genes (FDR<0.05). I used UMAP analysis, a nearest neighbour dimensionality reduction approach, to ask which samples were most similar to one another based on this set of 7,041 marker genes and found that, unlike control FL CD19+ cells and *MLL*wt infant ALL, which formed separate clusters, ^{CRISPR}*MLL-AF4+* ALL clustered with *MLL-AF4* infant ALL (Figure 4.3a). As before, the two ^{CRISPR}*MLL-AF4+* proB ALLs clustered next to one another, while the ^{CRISPR}*MLL-AF4+* preB ALL did not. The *MLL-AF4* infant ALL samples formed a large, diffuse cluster, with some

patients closer to ^{CRISPR}*MLL-AF4+* proB ALL and some closer to ^{CRISPR}*MLL-AF4+* preB ALL, suggesting that heterogeneity exists in the transcriptomic profiles of different *MLL-AF4* infant ALL patients, and ^{CRISPR}*MLL-AF4+* ALL appears to capture some of this heterogeneity (Figure 4.3a).

Taking this analysis a step further, I next wanted to ask whether ^{CRISPR}*MLL-AF4+* ALL specifically recapitulated *MLL-AF4* ALL rather than ALL with other *MLL*-rearrangements. I used another published patient RNA-seq dataset which

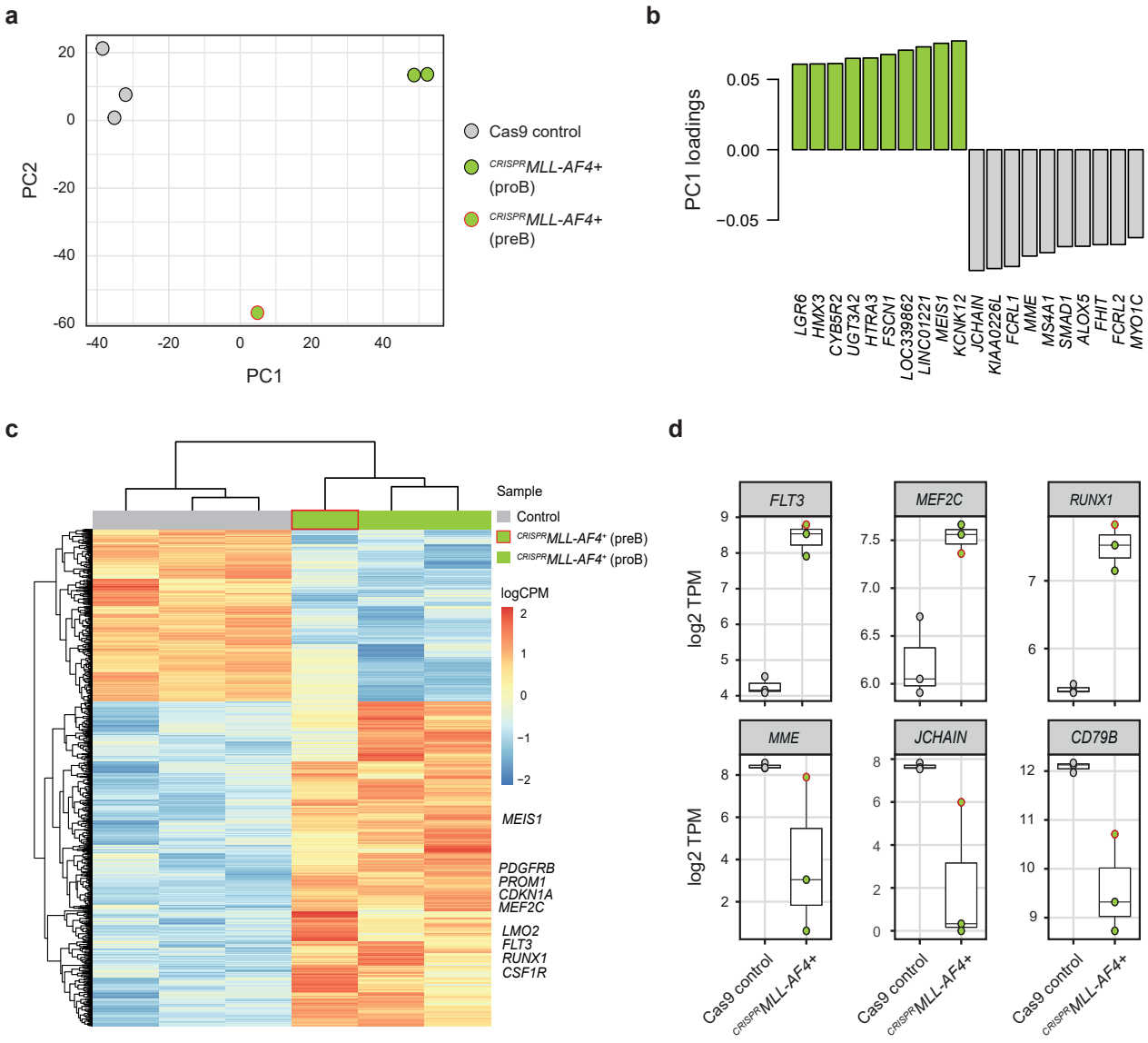


Figure 4.2

Figure 4.2: ^{CRISPR}*MLL-AF4+* ALL shows upregulation of known *MLL*-rearranged leukaemia-associated genes. a. PC analysis based on the top 1000 most variable genes among ^{CRISPR}*MLL-AF4+* ALL (n=3) and control FL CD19+ cells (n=3). ^{CRISPR}*MLL-AF4+* samples are shown in green. ^{CRISPR}*MLL-AF4+* proB ALLs have a black border, while ^{CRISPR}*MLL-AF4+* preB ALL has a red border. **b.** Plot of the top 20 most variable genes among ^{CRISPR}*MLL-AF4+* ALL (n=3) and control FL CD19+ cells (n=3), and their relative contribution to the separation of PC1 (PC1 loadings). Genes with a positive PC1 loading are more highly expressed in ^{CRISPR}*MLL-AF4+* samples (green), while genes with a negative PC1 loading are more highly expressed in control FL CD19+ cells (grey). **c.** Heatmap showing 1,068 significantly differentially expressed genes between ^{CRISPR}*MLL-AF4+* ALL (n=3) and control FL CD19+ cells (n=3) (FDR<0.05). Dendrograms represent Euclidean clusters of samples (top) and genes (left). ^{CRISPR}*MLL-AF4+* preB ALL is shown with a red border. Colour scale represent relative log2-transformed counts per million (log2 CPM) for each gene. A selection of known *MLL*-rearranged leukaemia-associated, stem cell-associated and myeloid lineage-associated genes are labelled. **d.** Boxplots showing expression (log2 TPM) of 3 known MLL-AF4 target genes, *FLT3, MEF2C* and *RUNX1* (top), and 3 known B lineage maturation genes, *MME, JCHAIN* and *CD79B* (bottom), in ^{CRISPR}*MLL-AF4+* ALL (n=3) and control FL CD19+ cells (n=3). ^{CRISPR}*MLL-AF4+* samples are shown in green. ^{CRISPR}*MLL-AF4+* proB ALLs have a black border, while ^{CRISPR}*MLL-AF4+* preB ALL has a red border.

included patients of all ages with a variety of *MLL*-rearrangements, as well as *MLL*wt ALL patients [Andersson et al., [2015]]. PC analysis revealed that *MLL-AF4* ALL patients clustered together, while ALL patients with other *MLL*-rearrangements and *MLL*wt ALL patients clustered away from *MLL-AF4* ALL (Figure 4.3b). Unlike *MLL-AF4* ALL, patients with other *MLL*-rearrangements did not appear to form distinct clusters based on their *MLL* fusion partner gene. Once again, ^{CRISPR}*MLL-AF4+* ALL clustered with *MLL-AF4* ALL patients specifically (Figure 4.3b), whereas control FL CD19+ cells clustered with *MLL*wt ALL patients, likely due to the more mature B lineage phenotype of these ALLs, as stem cell- and leukaemia-associated genes like *FLT3, PROM1* and *CCNA1* and B lineage maturation gene like *MME* played a major role in defining PC1 (Figure 4.3c).

On PC2, two genes that played a major role were *HOXA10* and *IRX1.* Samples with high *HOXA10* expression showed low *IRX1* expression and vice-versa (Figure 4.3d). This would suggest that the well-characterised *HOXA*$^{\text{hi}}$ vs *HOXA*$^{\text{lo}}$ gene expression signatures were a major source of variance between these patient samples. I therefore set out to investigate in more detail whether ^{CRISPR}*MLL-AF4+* ALL

recapitulated *HOXA*^{hi} *MLL-AF4* ALL or *HOXA*^{lo} *MLL-AF4* ALL.

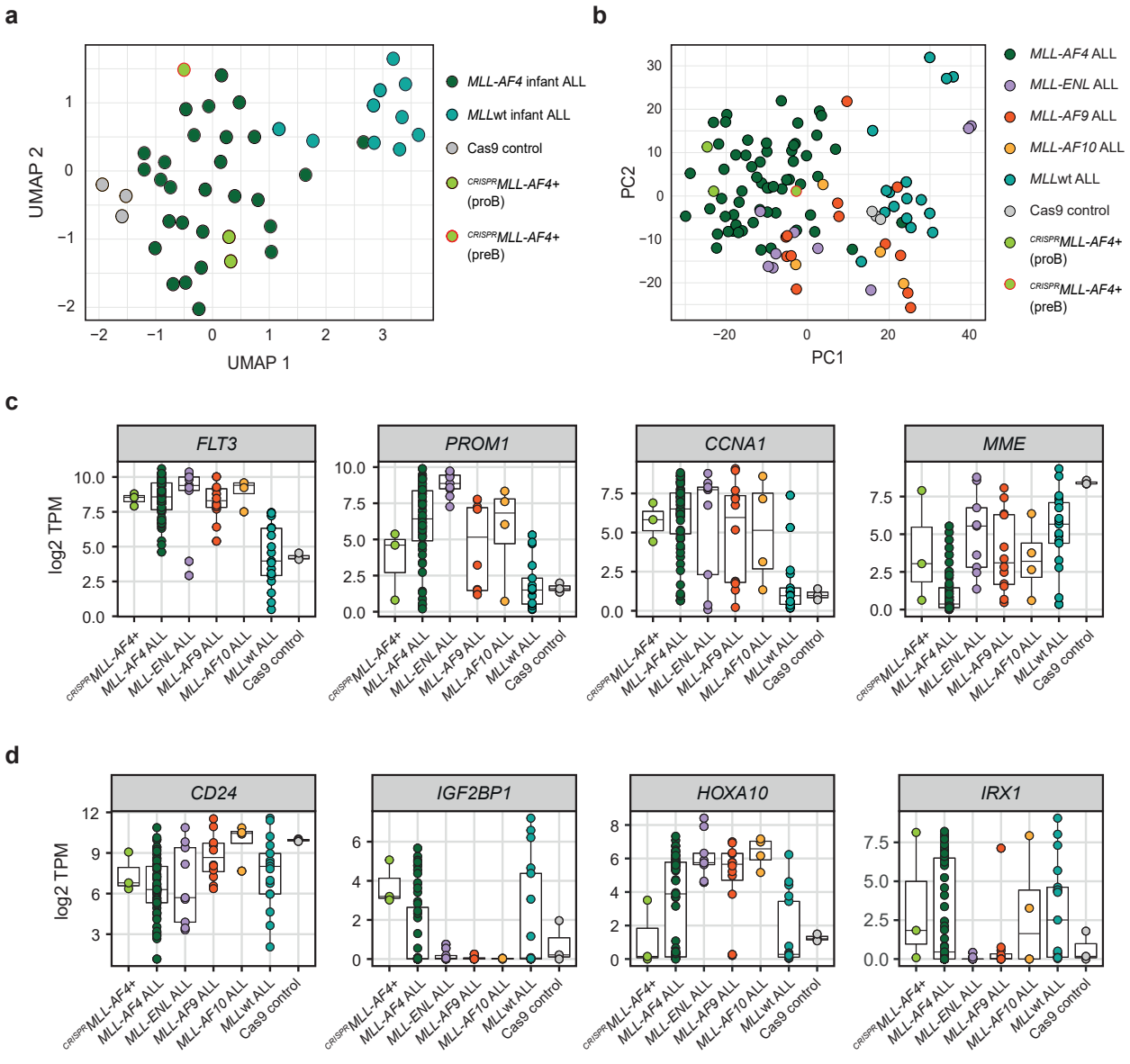


Figure 4.3

Figure 4.3: ^{CRISPR}*MLL-AF4+* ALL recapitulates the transcriptomic profile of *MLL-AF4* ALL patients. a. UMAP showing clustering of ^{CRISPR}*MLL-AF4+* ALL (light green (preB ALL = red border), n=3) and control FL CD19+ cells (grey, n=3) with *MLL-AF4* (dark green, n=31) and *MLL*wt (blue, n=9) infant-ALL patient samples from a publicly available dataset [Agraz-Doblas et al., [2019]] based on 7,041 significantly differentially expressed genes (FDR<0.05) between all 4 groups. **b.** PC analysis of ^{CRISPR}*MLL-AF4+* ALL (light green (preB ALL = red border, n=3), control FL CD19+ cells (grey, n=3), *MLL-AF4* (dark green, n=19), *MLL-ENL* (purple, n=9), *MLL-AF9* (orange, n=13), *MLL-AF10* (yellow, n=4) and *MLL*wt (blue, n=16) ALL patients [Andersson et al., [2015]] based on the top 1000 most variable genes in the combined dataset. **c.** Boxplots showing expression (log2-transformed transcripts per million (log2 TPM)) of 4 of the top 10 genes contributing to the separation of PC1 (PC1 loadings) for ^{CRISPR}*MLL-AF4+* ALL (light green (preB ALL = red border, n=3), control FL CD19+ cells (grey, n=3), *MLL-AF4* (dark green, n=19), *MLL-ENL* (purple, n=9), *MLL-AF9* (orange, n=13), *MLL-AF10* (yellow, n=4) and *MLL*wt (blue, n=16) ALL patients [Andersson et al., [2015]]. **d.** Boxplots showing expression (log2 TPM) of 4 of the top 10 genes contributing to the separation of PC2 (PC2 loadings) for ^{CRISPR}*MLL-AF4+* ALL (light green (preB ALL = red border, n=3), control FL CD19+ cells (grey, n=3), *MLL-AF4* (dark green, n=19), *MLL-ENL* (purple, n=9), *MLL-AF9* (orange, n=13), *MLL-AF10* (yellow, n=4) and *MLL*wt (blue, n=16) ALL patients [Andersson et al., [2015]].

4.3.3 ^{CRISPR}*MLL-AF4+* ALL recapitulates the poor prognosis subset of *HOXA*^{lo} *MLL-AF4* ALL

Taking the patient RNA-seq data [Andersson et al., [2015]], I classified *MLL-AF4* ALL patients into *HOXA*^{hi} (*HOXA9* expression level >15 transcripts per million (TPM)) or *HOXA*^{lo} (*HOXA9* expression level <15 TPM) subsets and carried out differential expression analysis between these 2 subsets to identify 588 significantly differentially expressed genes (FDR<0.05). In-keeping with all previous reports of the *HOXA*^{hi} vs *HOXA*^{lo} gene expression signatures, *HOXA*^{hi} *MLL-AF4* ALL overexpressed several *HOXA* cluster genes, while *HOXA*^{lo} *MLL-AF4* ALL overexpressed *IRX1* and *IRX2* (Figure 4.4a). I therefore carried out UMAP analysis of *MLL-AF4* ALL patients and ^{CRISPR}*MLL-AF4+* ALL based on this set of 588 genes and found that all ^{CRISPR}*MLL-AF4+* samples (both preB and proB ALL) clearly clustered with *HOXA*^{lo} *MLL-AF4* ALL patients (Figure 4.4b), indicating that ^{CRISPR}*MLL-AF4+* ALL not only recapitulates the gene expression profile of *MLL-AF4* ALL patients, but it recapitulates the poor prognosis subset of *HOXA*^{lo} *MLL-AF4* ALL, irrespective of its immunophenotype.

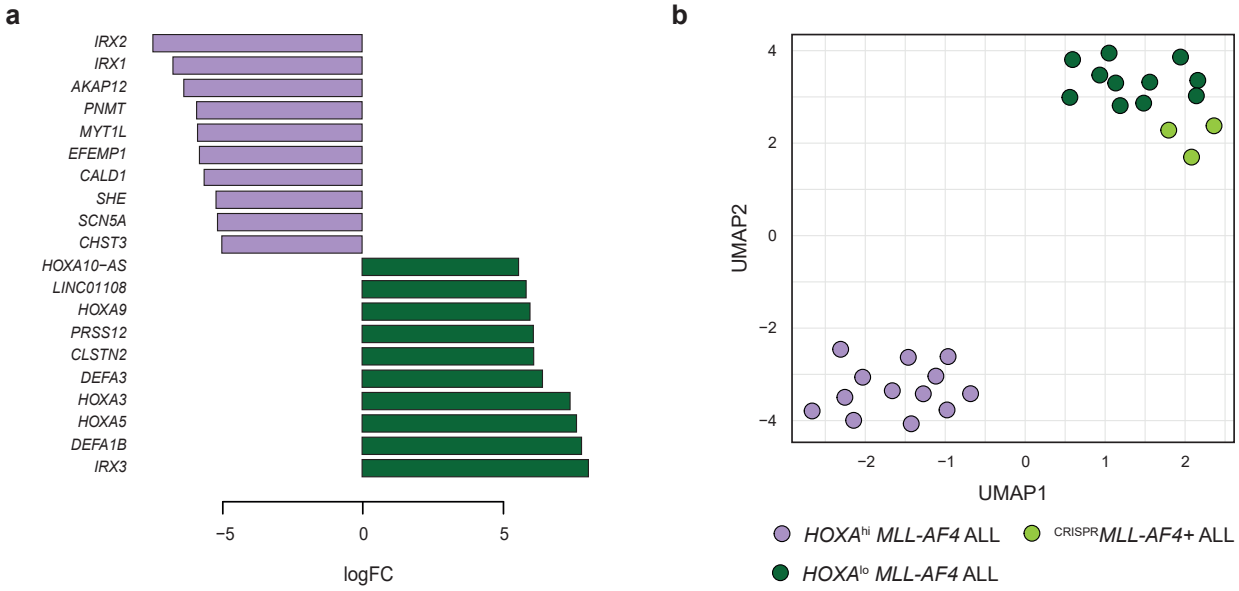


Figure 4.4: *CRISPR*MLL-AF4+ ALL model poor prognosis *HOXA*^{lo} *MLL-AF4* ALL. a. Top 20 significantly differentially expressed genes (FDR<0.05) with the greatest log fold change (logFC) in *HOXA*^{hi} *MLL-AF4* infant ALL (purple, negative logFC) and *HOXA*^{lo} *MLL-AF4* infant ALL (dark green, positive logFC) [Andersson et al., [2015]]. **b.** UMAP showing clustering of *CRISPR*MLL-AF4+ ALL with *HOXA*^{lo} *MLL-AF4* infant ALL patients and away from *HOXA*^{hi} *MLL-AF4* infant ALL patients [Andersson et al., [2015]] based on 588 significantly differentially expressed genes between *HOXA*^{lo} *MLL-AF4* infant ALL patients and *HOXA*^{hi} *MLL-AF4* infant ALL (FDR<0.05).

4.4 *CRISPR*MLL-AF4+ ALL recapitulates the chromatin profile of *MLL-AF4* ALL

4.4.1 The MLL-AF4 binding profile of *CRISPR*MLL-AF4+ ALL recapitulates that of *MLL-AF4* ALL patients

Strong upregulation of previously reported MLL-AF4 target genes in *CRISPR*MLL-AF4+ ALL suggested that the CRISPR/Cas9-induced MLL-AF4 fusion protein was functioning in a similar manner to the MLL-AF4 fusion protein in patients. To test this, I used ChIP-seq to analyse the chromatin binding profile of MLL-AF4 in proB ALL blasts harvested from the spleen of 2 primary *CRISPR*MLL-AF4+ recipient mice. As there is no antibody specific to the fusion protein, I carried out ChIP-seq for the N-terminus of MLL (MLL-N) and the C-terminus of AF4 (AF4-C) (an

approach that has been well-established in the field, [Guenther et al., [2008], Kerry et al., [2017]]) and defined an MLL-AF4 binding site as a region with directly overlapping MLL-N and AF4-C ChIP-seq peaks. There was a high degree of overlap between ^{CRISPR}*MLL-AF4+* mice, with 5,944 MLL-AF4 target genes common to both mice (Figure 4.5a), including well-characterised target genes such as *FLT3, RUNX1* and *PROM1*. MLL-AF4 peaks common to both ^{CRISPR}*MLL-AF4+* mice were taken forward for all subsequent analyses.

In order to better characterise the chromatin environment at MLL-AF4 binding sites, I carried out H3K27ac ChIP-seq in blasts harvested from the spleen of 2 ^{CRISPR}*MLL-AF4+* recipient mice. I then split the MLL-AF4 peaks into the following subsets:

- Promoter peak (MLL-AF4 ChIP-seq peak within 2kb of the gene transcriptional start site (TSS))
- Intragenic enhancer peak (MLL-AF4 ChIP-seq peak occurring inside the gene body (>2kb away from the TSS) and overlapping with a H3K27ac peak)
- Intergenic enhancer peak (MLL-AF4 ChIP-seq peak occurring outside the gene body (>2kb away from the TSS) and overlapping with a H3K27ac peak)
- Intragenic (other) peak (MLL-AF4 ChIP-seq peak occurring inside the gene body (>2kb away from the TSS) and not overlapping with a H3K27ac peak)
- Intergenic (other) peak (MLL-AF4 ChIP-seq peak occurring outside the gene body (>2kb away from the TSS) and not overlapping with a H3K27ac peak)

In keeping with previous reports [Kerry et al., [2017], Godfrey et al., [2019]], the majority (43%) of MLL-AF4 peaks were promoter peaks, while a smaller proportion (12%) occurred at enhancers (Figure 4.5b). 24% of MLL-AF4 peaks were found in intragenic regions that were not classified as enhancers (Figure 4.5b). This likely represents MLL-AF4 spreading from the promoter into the gene body, which has previously been described in detail in the SEM cell line by the Milne laboratory [Kerry et al., [2017]]. Therefore, the regions at which MLL-AF4 binds to

chromatin in $^{CRISPR}MLL\text{-}AF4+$ mice replicated previous findings from *MLL-AF4* patients and patient-derived cell lines.

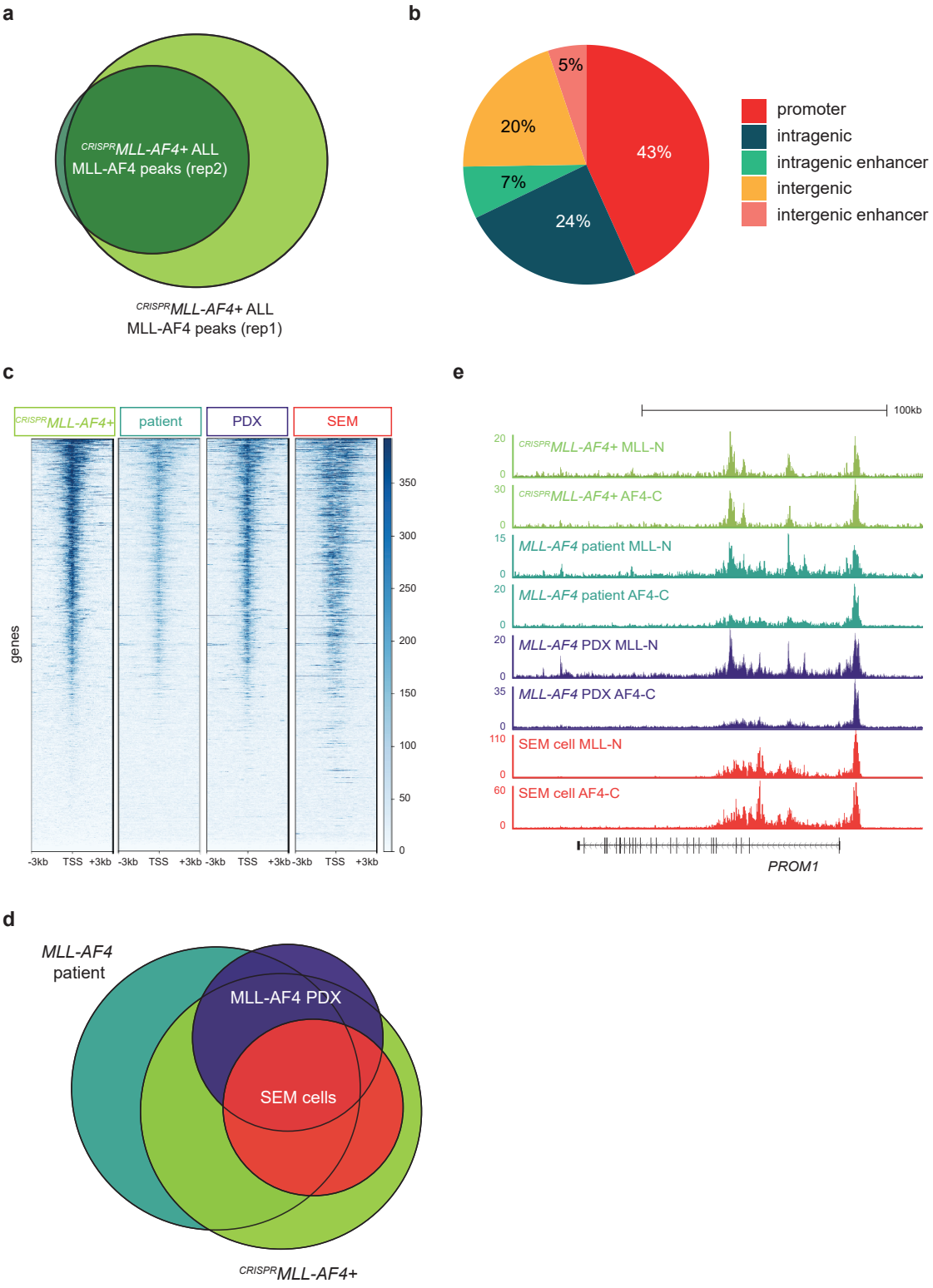

a
CRISPR MLL-AF4+ ALL
MLL-AF4 peaks (rep2)
CRISPR MLL-AF4+ ALL
MLL-AF4 peaks (rep1)
b
5%
20%
43%
7%
24%
promoter
intragenic
intragenic enhancer
intergenic
intergenic enhancer
c
CRISPR MLL-AF4+
patient
PDX
SEM
genes
-3kb TSS +3kb
350
300
250
200
150
100
50
0
d
MLL-AF4
patient
MLL-AF4 PDX
SEM cells
CRISPR MLL-AF4+
e
100kb
CRISPR MLL-AF4+ MLL-N
CRISPR MLL-AF4+ AF4-C
MLL-AF4 patient MLL-N
MLL-AF4 patient AF4-C
MLL-AF4 PDX MLL-N
MLL-AF4 PDX AF4-C
SEM cell MLL-N
SEM cell AF4-C
PROM1

Figure 4.5

Figure 4.5: ^{CRISPR}*MLL-AF4+* ALL recapitulates the chromatin profile of *MLL-AF4* ALL patients. a. Venn diagram showing overlap in MLL-AF4-bound genes (gene with at least one overlapping MLL-N and AF4-C peak in their gene body) between two ^{CRISPR}*MLL-AF4+* proB ALL replicates. 5,944 genes were bound by MLL-AF4 in both replicates. These genes were taken forward as ^{CRISPR}*MLL-AF4+* ALL MLL-AF4 binding sites for subsequent analyses. **b.** Pie chart showing genomic regions of MLL-AF4 binding in ^{CRISPR}*MLL-AF4+* ALL. Promoter = ± 2kb from a gene TSS. Intragenic = a region inside of a gene body with no enhancer marks. Intragenic enhancer = a region inside of a gene body with an overlapping H3K27ac peak. Intragenic = a region outside of a gene body with no enhancer marks. Intragenic enhancer = a region outside of a gene body with an overlapping H3K27ac peak. **c.** Heatmap showing MLL-N ChIP-seq enrichment for a 6kb region centered on the TSS of all genes in ^{CRISPR}*MLL-AF4+* ALL, the SEM cell line, an *MLL-AF4* ALL PDX model and a primary *MLL-AF4* ALL patient sample, sorted by MLL-N ChIP-seq signal in ^{CRISPR}*MLL-AF4+* ALL. Scale = reads/bp normalized to 10^7 total reads. **d.** Representative ChIP-seq tracks at the MLL-AF4 target gene, *PROM1*, in ^{CRISPR}*MLL-AF4+* ALL, the SEM cell line, an *MLL-AF4* ALL PDX model and a primary *MLL-AF4* ALL patient sample. Scale = reads/bp normalized to 10^7 total reads.

I next carried out a direct comparison of MLL-AF4 binding sites in ^{CRISPR}*MLL-AF4+* ALL, the SEM cell line, a primary *MLL-AF4* ALL patient sample and an *MLL-AF4* ALL PDX model. I observed a clear genome-wide correlation between all 4 samples (Figure 4.5c) and a substantial overlap between MLL-AF4 target genes (Figure 4.5d). Moreover, within common target genes that are upregulated in *MLL-AF4* ALL like *PROM1*, the MLL-AF4 binding profile was strikingly similar between all 4 samples (Figure 4.5e), suggesting that ^{CRISPR}*MLL-AF4+* ALL recapitulates the MLL-AF4 binding profile of *MLL-AF4* ALL patients, leading to strong upregulation of known MLL-AF4 target genes.

4.4.2 MLL-AF4 binding leads to increased H3K79me2 deposition during ^{CRISPR}*MLL-AF4*-induced leukaemic transformation

I next wanted to investigate the molecular mechanisms by which MLL-AF4 binding could drive these profound changes in gene expression at MLL-AF4 target genes during ^{CRISPR}*MLL-AF4*-induced leukaemic transformation. The high levels of H3K79me2/3 observed in *MLL-AF4* ALL patients is widely believed to underlie upregulated expression of MLL-AF4 target genes. However, the lack of appropriate

model systems has made prospective studies of this phenomenon difficult. I therefore took advantage of a unique feature of the ^{CRISPR}*MLL-AF4+* ALL model by carrying out H3K79me2 ChIP-seq in the same primary human FL sample before and after ^{CRISPR}*MLL-AF4*-driven leukaemic transformation. I carried out H3K79me2 ChIP-seq in human CD19+ cells sorted from BM of a ^{CRISPR}*MLL-AF4+* mouse and unedited CD19+ cells sorted from the same primary FL sample and found that H3K79me2 was massively enriched at MLL-AF4 target genes, like *FLT3* and *PAN3* (Figure 4.6a and 4.6b), compared to non-MLL-AF4-bound genes in ^{CRISPR}*MLL-AF4+* blasts, as well as all genes in the matched, unedited FL control (Figure 4.6b). I further confirmed these findings using ChIP-qPCR in 3 ^{CRISPR}*MLL-AF4+* ALL samples and 3 matched FL controls at MLL-AF4 target genes, *BCL11A* and *CDK6* (Figure 4.6c). These data confirm that increased H3K79me2 is a direct consequence of MLL-AF4 binding during leukemic transformation and suggest that this is an important mechanism by which MLL-AF4 drives aberrant expression of target genes.

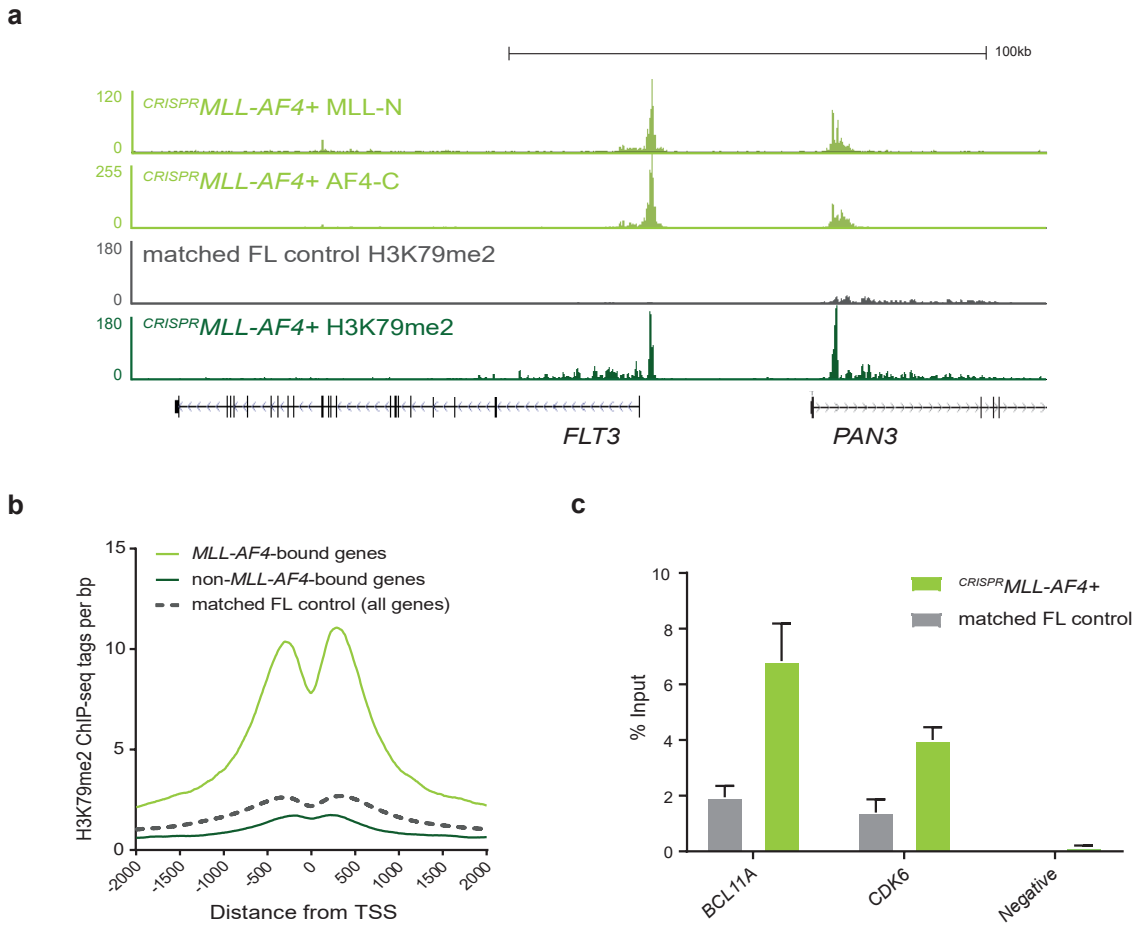


Figure 4.6: Increased H3K79me2 is a direct consequence of MLL-AF4 binding during ^{CRISPR}*MLL-AF4*-driven leukaemic transformation. a. Representative ChIP-seq tracks for MLL-N, AF4-C (light green) and H3K79me2 (dark green) in ^{CRISPR}*MLL-AF4+* ALL and H3K79me2 (grey) in primary unedited FL CD19+ cells from the sample FL sample at MLL-AF4 target genes, *FLT3* and *PAN3*. Scale = reads/bp normalized to 10^7 total reads. **b.** Metaplot of H3K79me2 ChIP-seq signal centered on the TSS of MLL-AF4-bound genes (light green) and non-MLL-AF4-bound genes (dark green) in ^{CRISPR}*MLL-AF4+* ALL, and all genes in matched control FL CD19+ cells (derived from the same FL samples; grey, dashed). **c.** ChIP-qPCR data showing MLL-N enrichment (as percentage of total input DNA) at the promoters of MLL-AF4 bound genes, *BCL11A* and *CDK6*, in ^{CRISPR}*MLL-AF4+* ALL (light, green, n=3) and matched control FL CD19+ cells (derived from the same FL samples; grey, n=3).

4.5 ^{CRISPR}*MLL-AF4+* ALL specifically recapitulates the molecular profile of *MLL-AF4* infant ALL

Taken together, these data indicate that ^{CRISPR}*MLL-AF4+* ALL faithfully recapitulates the molecular profile of *MLL-AF4* ALL patients, both in terms of gene expression and chromatin environment. However, the aim of targeting an infant ALL-specific *MLL-AF4*/t(4;11) translocation to primary human FL HSPCs was to generate a model of the poor prognosis subset of *MLL-AF4* infant ALL. Therefore, I next set out to determine firstly, whether *MLL-AF4* infant ALL was molecularly distinct from *MLL-AF4* ALL in older children; secondly, to what extent human fetal-specific gene expression programmes drive this distinct molecular signature; and finally, whether ^{CRISPR}*MLL-AF4+* ALL specifically recapitulated the molecular profile of *MLL-AF4* infant ALL.

4.5.1 *MLL-AF4* infant ALL is transcriptomically distinct from *MLL-AF4* childhood ALL

Returning to the published patient RNA-seq dataset that included patients of all ages with varied *MLL*-rearrangements [Andersson et al., [2015]], I carried out differential gene expression analysis between all *MLL-AF4* infant ALL and *MLL-AF4* childhood ALL samples to identify 617 significantly differentially expressed genes (FDR <0.05), 193 of which were upregulated in *MLL-AF4* infant ALL and therefore represented an infant ALL-specific gene expression profile (Figure 4.7a). I again subset these patients based on the *HOXA*^{hi} vs *HOXA*^{lo} gene expression signature and found that *HOXA*^{hi} *MLL-AF4* infant ALL and *HOXA*^{lo} *MLL-AF4* infant ALL did not separate based on these 617 genes (Figure 4.7a), suggesting that this was a true infant ALL-specific gene expression profile, irrespective of other molecular characteristics of the leukaemia. In fact, of the top 10 most significantly upregulated genes that could separate *MLL-AF4* infant ALL from *MLL-AF4* childhood ALL (Figure 4.7b), no significant differences were observed between *HOXA*^{lo} and *HOXA*^{hi} *MLL-AF4* infant ALL subsets (Figure 4.7c).

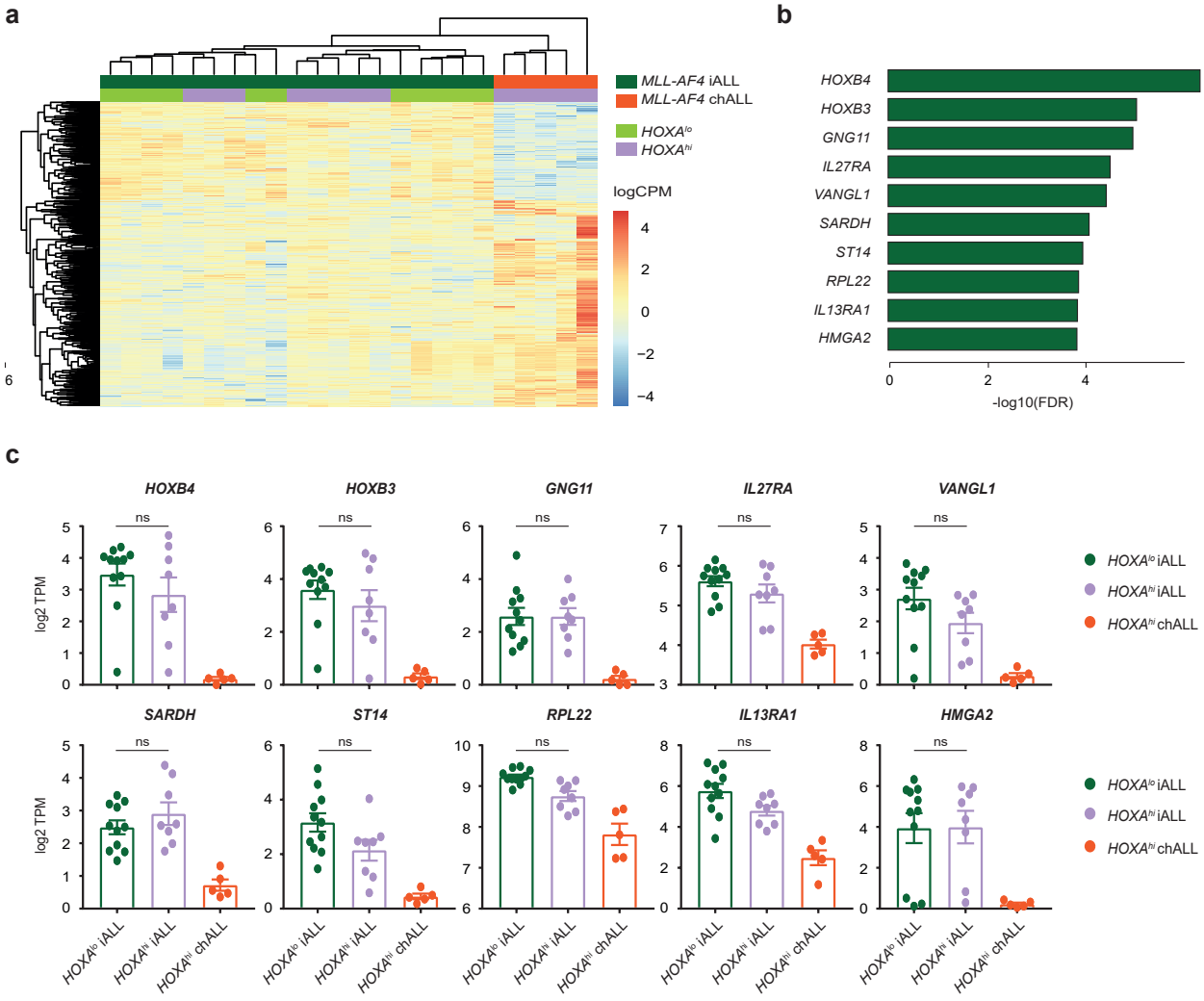


Figure 4.7: *MLL-AF4* infant ALL is transcriptomically distinct from *MLL-AF4* childhood ALL. a. Heatmap showing clustering of *MLL-AF4* infant ALL (iALL (dark green), n=19) and *MLL-AF4* childhood ALL (chALL (orange), n=5) based on 617 significantly differentially expressed genes (FDR<0.05). *HOXA*$^{\text{lo}}$ (light green, n=11) and *HOXA*$^{\text{hi}}$ (purple, n=13) *MLL-AF4* subsets of infant ALL and childhood ALL are annotated. Color scale = log2 CPM. **b.** Barplot showing significance (-log10(FDR)) for the 10 most significantly upregulated genes in *MLL-AF4* infant-ALL. **c.** Expression (log2 TPM) of the top 10 most significantly upregulated genes in *MLL-AF4* infant-ALL in *HOXA*$^{\text{lo}}$ *MLL-AF4* infant-ALL (iALL (light green), n=11), *HOXA*$^{\text{hi}}$ *MLL-AF4* infant-ALL (iALL (purple), n=8) and *HOXA*$^{\text{hi}}$ *MLL-AF4* childhood-ALL (chALL (orange), n=5). Data shown as mean ± SEM (ns: p>0.05).

4.5.2 Fetal gene expression programmes drive the distinct transcriptomic profile of *MLL-AF4* infant ALL

I next sought to determine the extent to which normal fetal gene expression programmes contribute to the distinct molecular profile of *MLL-AF4* infant ALL. I compared bulk RNA-seq data for sorted human FL HSPC subpopulations previously generated in the Roberts and Roy laboratory [Agraz-Doblas et al., [2019]] to a human adult BM HSPC RNA-seq dataset [Corces et al., [2016]]. I carried out differential gene expression analysis between comparable subpopulations of FL and adult BM HSPCs along the B lineage differentiation pathway (Figure 4.8a (left)). The HSC subpopulation showed the greatest number of differentially expressed genes between FL and adult BM (3,787 genes), reducing at each subsequent stage of B lineage differentiation (1,509 genes differentially expressed between FL committed B progenitors (CBP; including PPB and PB subpopulations) and adult BM CLP) (Figure 4.8a (right)). A total of 5,709 genes were differentially expressed between FL and adult BM in at least one HSPC subpopulation when I combined all differentially expressed gene lists.

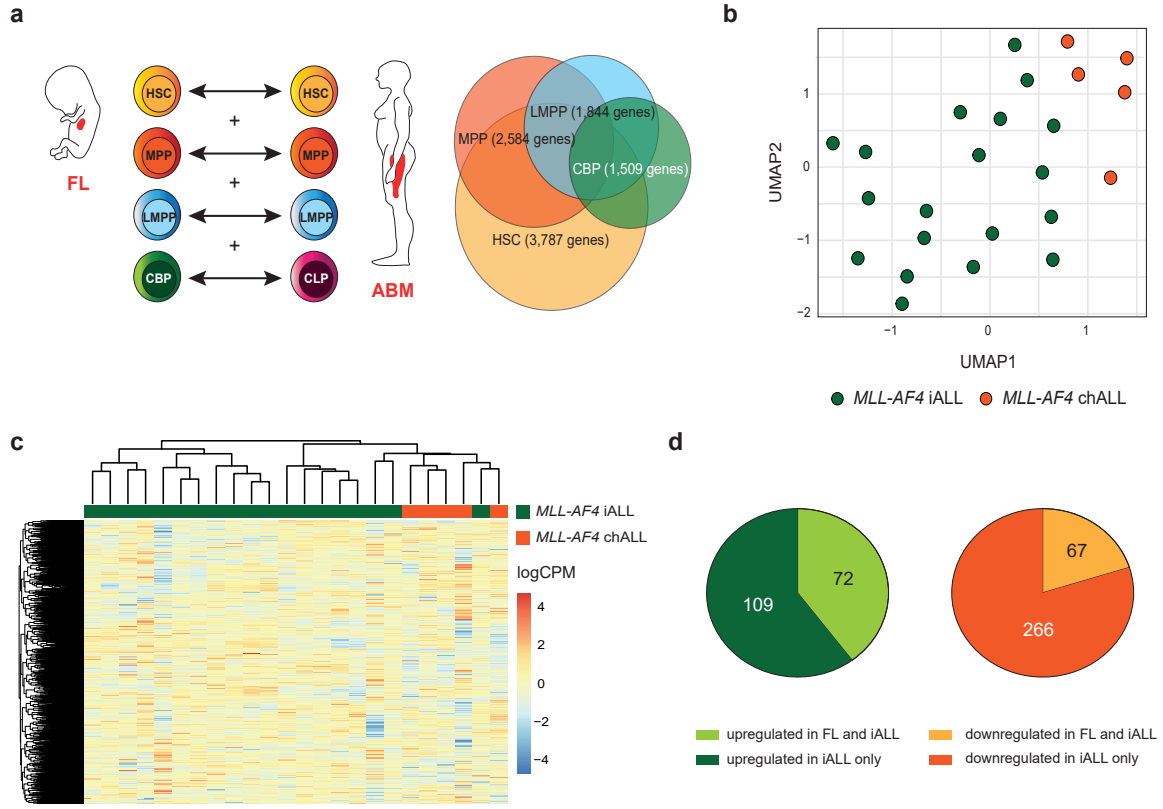


Figure 4.8: Fetal gene expression programs drive the distinct molecular profile of *MLL-AF4* infant ALL. a. (left) Schematic representation of differential gene expression analysis between FL and adult BM. Equivalent HSPC subpopulations were compared and significantly differentially expressed genes (FDR<0.05) [Corces et al., [2016], Agraz-Doblas et al., [2019]] for all 4 comparisons were combined into a master list of genes that were differentially expressed in at least 1 HSPC subpopulation (HSC = haematopoietic stem cell, MPP = multipotent progenitor cell, LMPP = lymphoid-primed multipotent progenitor cell, CBP = committed B progenitor, CLP = common lymphoid progenitor). (right) Venn diagram showing overlap of differentially expressed genes for each HSPC subpopulation. **b.** UMAP showing clustering of *MLL-AF4* infant ALL (iALL (dark green), n=19) and *MLL-AF4* childhood ALL (chALL (orange), n=5) from a previously published patient dataset [Andersson et al., [2015]] based on 5,709 genes differentially expressed between FL HSPCs and adult BM HSPCs. **c.** Heatmap showing clustering of *MLL-AF4* infant ALL (iALL (green), n=19) and *MLL-AF4* childhood ALL (chALL (orange), n=5) [Andersson et al., [2015]] based on 5,709 significantly differentially expressed genes between FL [Agraz-Doblas et al., [2019]] and adult BM [Corces et al., [2016]] HSPCs (FDR<0.05). Color scale = log2 CPM. **d.** Pie charts showing proportion of genes upregulated in *MLL-AF4* infant ALL (compared to *MLL-AF4* childhood ALL; dark green) [Andersson et al., [2015]] that are also upregulated in FL (compared to adult BM; light green) [Corces et al., [2016], Agraz-Doblas et al., [2019]], and the proportion of genes downregulated in *MLL-AF4* infant ALL (compared to *MLL-AF4* childhood ALL; orange) that are also downregulated in FL (compared to adult BM; yellow). Values shown as number of genes.

I carried out clustering analysis of the patient dataset based on these 5,709 genes and found that they were capable of separating *MLL-AF4* infant ALL from *MLL-AF4* childhood ALL (Figure 4.8b and 4.8c). Comparing differentially expressed genes in both the normal and leukemic setting, I found 72 genes that were significantly upregulated in at least one normal FL HSPC subpopulation and also in *MLL-AF4* infant ALL (40% of all genes upregulated in *MLL-AF4* infant ALL compared to *MLL-AF4* childhood ALL) (Figure 4.8d). These included some of the most significantly upregulated genes in *MLL-AF4* infant ALL such as *HOXB4* and *IGF2BP1*, a member of the *LIN28B* axis that has previously been reported to positively regulate *HOXB4* expression (Figure 4.9). Together these data suggest that the molecular profile of human fetal HSPCs plays a key role in determining the distinct gene expression profile of *MLL-AF4* infant ALL.

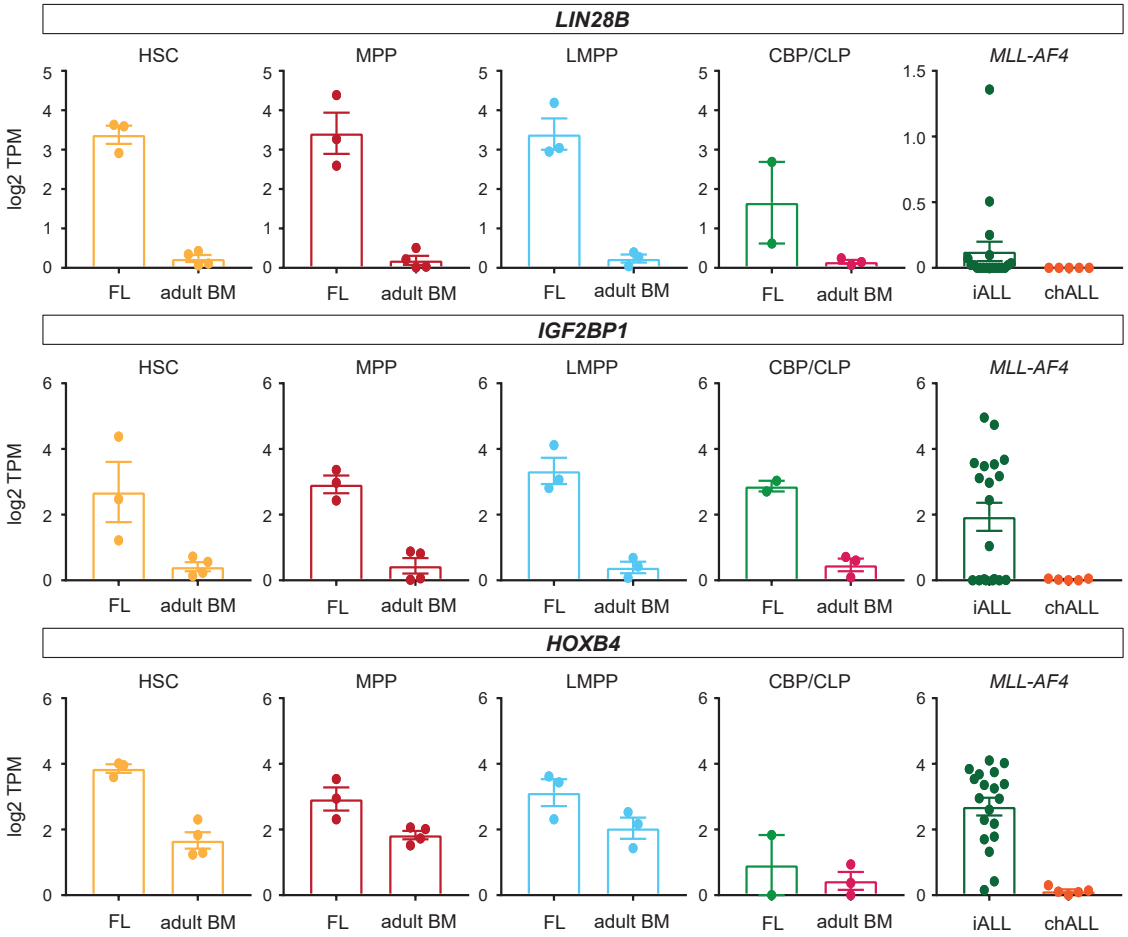


Figure 4.9: Expression of fetal-specific genes in primary human HSPCs and *MLL-AF4* ALL patients. Barplot showing expression of *LIN28B*, *IGF2BP1* and *HOXB4* in FL and adult BM HSPC subpopulations (HSC = haematopoietic stem cell (n=3 FL, n=4 adult BM), MPP = multipotent progenitor cell (n=3 FL, n=4 adult BM), LMPP = lymphoid-primed multipotent progenitor cell (n=3 FL, n=4 adult BM), CBP = committed B progenitor (n=2 FL), CLP = common lymphoid progenitor (n=4 adult BM)) [Corces et al., [2016], Agraz-Doblas et al., [2019]], as well as *MLL-AF4* infant ALL (iALL; n=19) and *MLL-AF4* childhood ALL (chALL; n=5) [Andersson et al., [2015]]. Data shown as mean ± SEM.

4.5.3 Human FL-derived ^{CRISPR}*MLL-AF4*+ ALL specifically recapitulates *MLL-AF4* infant ALL

Finally, I wanted to use these infant ALL-specific and fetal-specific signatures to ask whether inducing an *MLL-AF4*/t(4;11) translocation in primary human FL HSPCs gave rise to a model that specifically recapitulated the molecular profile of *MLL-AF4*

infant ALL. The only other humanised mouse model of *MLL-AF4* proB-ALL that has previously been published retrovirally introduced a chimeric human/mouse *MLL-Af4* fusion gene into human CB HSPCs (hereafter referred to as CB *MLL-Af4+* ALL) [Lin et al., [2016]]. While it must be acknowledged that this CB model was generated using a different technique, and did not harbour the reciprocal *AF4-MLL* fusion gene, I hypothesised that it may represent a neonatal HSPC-derived (non-fetal) *MLL-Af4+* ALL to which my fetal HSPC-derived ^{CRISPR}*MLL-AF4+* ALL could be compared.

To examine the fetal and post-natal gene expression programmes that are key to determining the age-related differences in *MLL-AF4* ALL, I used the 139 genes up- or downregulated in both FL (compared to adult BM) and *MLL-AF4* infant ALL (compared to *MLL-AF4* childhood ALL). Clustering analysis based on this core fetal-specific infant ALL gene list showed that ^{CRISPR}*MLL-AF4+* ALL clustered with *MLL-AF4* infant ALL, whereas both *MLL-AF4* childhood ALL and CB *MLL-Af4+* ALL formed their own, separate clusters (Figure 4.10a). To explore this in more detail, I carried out differential gene expression analysis between ^{CRISPR}*MLL-AF4+* ALL and CB *MLL-Af4+* ALL followed by gene set enrichment analysis (GSEA). I found that ^{CRISPR}*MLL-AF4+* ALL was significantly enriched for genes upregulated in both *MLL-AF4* infant ALL ($p<0.03$) and FL HSPCs ($p<0.001$) compared to CB *MLL-Af4+* ALL (Figure 4.10b).

Comparing MLL-AF4 binding at promoters genome-wide in both models, I found that MLL-AF4 in ^{CRISPR}*MLL-AF4+* ALL showed greater enrichment (normalized ChIP-seq reads/bp) at the promoters of infant-ALL-specific and FL-specific genes compared to MLL-Af4 in CB *MLL-Af4+* ALL. However, at all other genes, MLL-AF4/MLL-Af4 enrichment was comparable (Figure 4.10c). At infant ALL-specific and FL-specific genes *IGF2BP1* (Figure 4.10d) and *HOXB4* (Figure 4.10e), I observed an MLL-AF4 peak at the promoter in ^{CRISPR}*MLL-AF4+* ALL but not CB *MLL-Af4+* ALL. These data suggested that MLL-AF4 may play an active role in maintaining fetal gene expression programmes in infant ALL. To investigate this further, I examined the H3K79me2 ChIP-seq data before and after leukaemic transformation and found that H3K79me2 levels were increased at MLL-AF4 peaks

in infant ALL-specific and fetal-specific genes, *IGF2BP1* (Figure 4.10d) and *HOXB4* (Figure 4.10e), in ^{CRISPR}*MLL-AF4+* ALL, further suggesting that MLL-AF4 actively maintains the expression of these fetal-specific genes in *MLL-AF4* infant ALL. Together, these results suggest that *MLL*-rearranged infant ALL patients have a unique gene expression profile that is derived from the fetal origin of the disease and may contribute to the especially aggressive phenotype of the leukaemia.

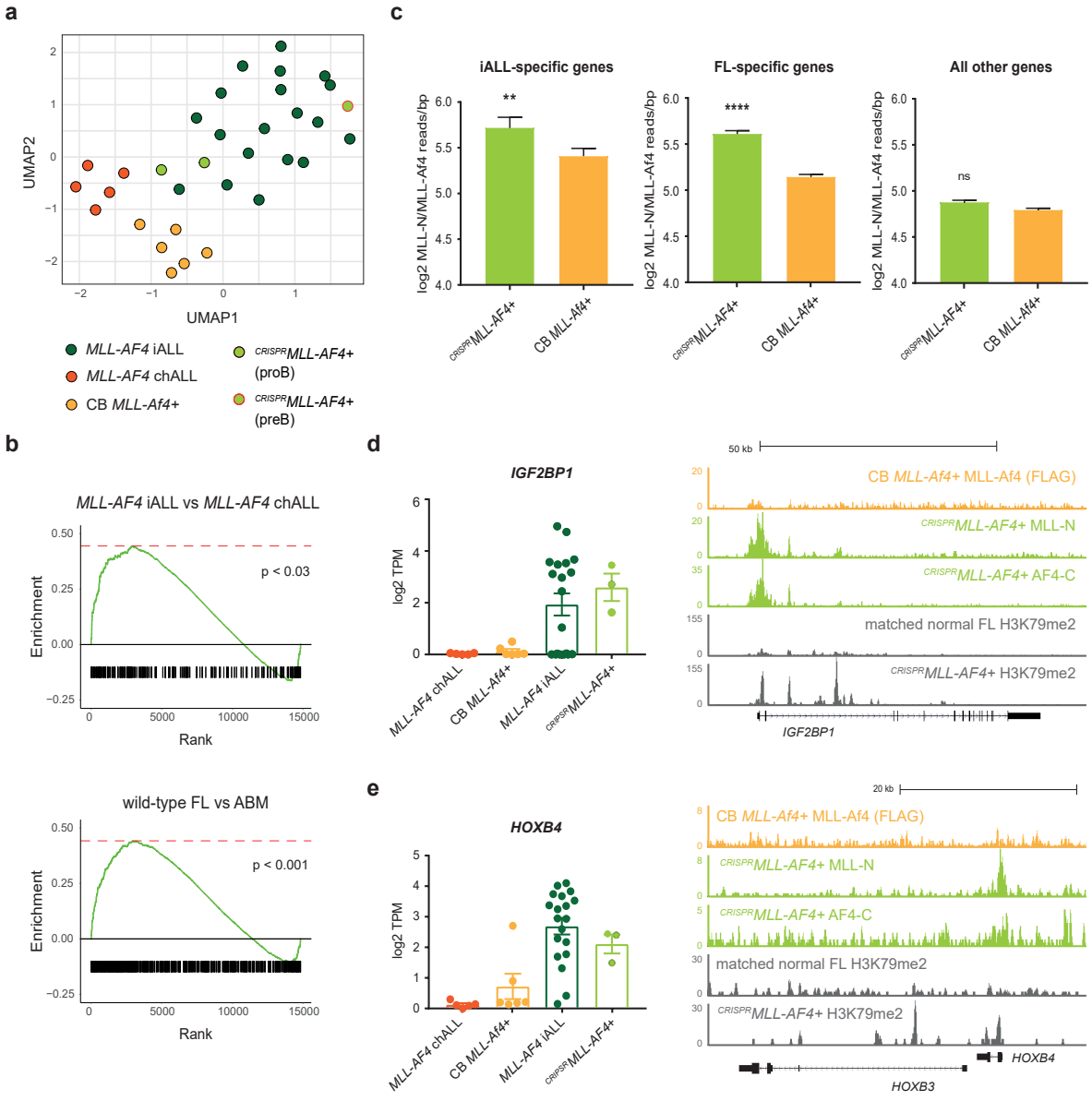


Figure 4.10

Figure 4.10: FL-derived ^{CRISPR}*MLL-AF4+* ALL specifically recapitulates the molecular profile of *MLL-AF4* infant-ALL. a. UMAP showing clustering of ^{CRISPR}*MLL-AF4+* ALL (light green, red border = preB ALL), *MLL-AF4* infant ALL (iALL (dark green), n=19), *MLL-AF4* childhood ALL (chALL (orange), n=5) [Andersson et al., [2015]] and CB *MLL-Af4+* ALL (yellow, n=6) [Lin et al., [2016]] based on a set of genes differentially expressed in both FL (compared to adult BM) and *MLL-AF4* infant-ALL (compared to *MLL-AF4* childhood-ALL) (139 genes). **b.** Gene set enrichment analyses (GSEA) [Korotkevich et al., [2018]] showing ^{CRISPR}*MLL-AF4+* ALL is significantly more enriched for genes upregulated in *MLL-AF4* infant ALL (compared to *MLL-AF4* childhood ALL) (617 genes, p<0.03) and in FL (compared to adult BM) (5,709 genes, p<0.001) when compared to CB *MLL-Af4+* ALL. **c.** Bar plots showing *MLL-AF4* enrichment (MLL-N ChIP-seq reads/bp normalised to 10^7 total reads) for ^{CRISPR}*MLL-AF4+* ALL and CB *MLL-Af4+* ALL (FLAG ChIP-seq) at infant-ALL-specific genes (193 genes, two-tailed Mann-Whitney test, p=0.012) and FL-specific genes (3,949 genes, two-tailed Mann-Whitney test, p<0.0001). Data shown as mean ± SEM. **d.** (left) Barplot showing expression of *IGF2BP1* in *MLL-AF4* childhood ALL (chALL (orange), n=5), CB *MLL-Af4+* ALL (yellow, n=6), *MLL-AF4* infant ALL (iALL (dark green), n=19) and ^{CRISPR}*MLL-AF4+* ALL (light green, red border = preB ALL). Data shown as mean ± SEM. (right) MLL-Af4 (FLAG ChIP-seq) in CB *MLL-Af4+* ALL, MLL-N and AF4-C ChIP-seq in ^{CRISPR}*MLL-AF4+* ALL and H3K79me2 ChIP-seq in ^{CRISPR}*MLL-AF4+* ALL and matched unedited FL sample at *IGF2BP1*. ChIP-seq data shown is normalized to 10^7 total reads. **e.** (left) Barplot showing expression of *HOXB4* in *MLL-AF4* childhood ALL (chALL (orange), n=5), CB *MLL-Af4+* ALL (yellow, n=6), *MLL-AF4* infant ALL (iALL (dark green), n=19) and ^{CRISPR}*MLL-AF4+* ALL (light green, red border = preB ALL). Data shown as mean ± SEM. (right) MLL-Af4 (FLAG ChIP-seq) in CB *MLL-Af4+* ALL, MLL-N and AF4-C ChIP-seq in ^{CRISPR}*MLL-AF4+* ALL and H3K79me2 ChIP-seq in ^{CRISPR}*MLL-AF4+* ALL and matched unedited FL sample at *HOXB3/HOXB4*. ChIP-seq data shown is normalized to 10^7 total reads.

4.6 Discussion

In this chapter, I have confirmed that ^{CRISPR}*MLL-AF4+* ALL recapitulates both the transcriptomic and epigenetic profile of *MLL-AF4* ALL in patients. Furthermore, I have demonstrated that *MLL-AF4* infant ALL shows significant gene expression differences compared to *MLL-AF4* childhood ALL, suggesting that *MLL*-rearranged ALL presenting in the first year of life is a molecularly distinct subset of this leukaemia. By deriving a human fetal-specific gene expression signature through the comparison of primary human HSPC RNA-seq data, I was able to show that a major driver of the *MLL-AF4* infant ALL-specific molecular profile (accounting for approximately 40% of all significantly upregulated genes compared to *MLL-AF4* childhood ALL) was maintained expression of human fetal-specific gene expression

programmes, likely mediated by direct binding of the MLL-AF4 fusion protein to fetal-specific genes during leukaemia initiation. Finally, I have found that ^{CRISPR}*MLL-AF4+* ALL represents a more faithful model of this infant ALL-specific molecular profile than the previous CB derived *MLL-Af4+* humanised mouse model. Overall, these data suggest that the transcriptomic and epigenetic environment of the fetal target cell in which the *MLL-AF4*/t(4;11) translocation first arises plays a key role in defining the molecular profile of the resultant ALL, both in FL-derived ^{CRISPR}*MLL-AF4+* ALL and in infant ALL patients.

While all ^{CRISPR}*MLL-AF4+* ALL samples showed high level expression of well-known *MLL*-rearranged leukaemia-associated genes, they also demonstrated some heterogeneity in their gene expression profile. By PC analysis, ^{CRISPR}*MLL-AF4+* preB ALL clustered separately from ^{CRISPR}*MLL-AF4+* proB ALL, and this appeared to be driven by higher expression levels of genes associated with B lineage maturation, in-keeping with the more mature immunophenotype of preB ALL. Unfortunately, of all primary ^{CRISPR}*MLL-AF4+* mice generated in this project, only 1/7 had this preB immunophenotype. Therefore, there were not enough preB ALL replicates available for an in-depth analysis of the significance of these transcriptomic differences. When clustering analysis was carried out with primary patient RNA-seq data, ^{CRISPR}*MLL-AF4+* ALL samples did not form one tight cluster of their own, but rather spread out between different *MLL-AF4* ALL patient samples, suggesting that transcriptomic difference between ^{CRISPR}*MLL-AF4+* ALL replicates may model some of the transcriptomic heterogeneity among patients. As the data in this chapter suggest that the molecular profile of the target cells in which the *MLL-AF4*/t(4;11) translocation first arises can be maintained to a certain extent in the leukaemic blasts, one possible explanation for transcriptomic heterogeneity among *MLL-AF4* infant ALL patients is that the leukaemias arose in varied fetal HSPC subtypes. For example, the earliest FL progenitors show the highest expression of fetal-specific genes such as *LIN28B* and *HOXB4*. Could it be then that the *MLL-AF4* infant ALL patients with the highest expression of these genes have a leukaemia derived from an early progenitor such as a HSC, whereas patients with

low expression have a leukaemia derived from more mature progenitor types, like B progenitor cells? Due to the lack of appropriate model systems, the cell-of-origin in *MLL-AF4* infant ALL remains unknown.

Of the ^{CRISPR}*MLL-AF4+* ALL samples analysed in this chapter, 3/3 represented the *HOXA*^{lo} *MLL-AF4* infant ALL subset. This is in-keeping with the higher frequency of this molecular subtype in *MLL*-rearranged infant ALL compared to *MLL*-rearranged childhood ALL; however, it precludes any analysis of the factors that drive these molecular subtypes. These factors may include target HSPC subpopulation, *MLL-AF4*/t(4;11) breakpoint region or gestational age/haematopoietic site of the target cells. In the future, it will be interesting to generate ^{CRISPR}*MLL-AF4+* ALL from different sources (FL, FBM, CB) as well as different HSPC subpopulations to determine if this approach is capable of giving rise to *HOXA*^{hi} ALL in any of these contexts.

High level H3K79me2/3 at MLL fusion protein target genes has long been recognised as an epigenetic hallmark of *MLL*-rearranged leukaemia. While this mark is generally found at expressed genes, [Krivtsov et al., [2008]] showed that H3K79me2 levels were higher at Mll-AF4 target genes in a mouse model of *Mll-Af4* leukaemia compared to wild-type mouse proB cells. Moreover, previous work from the Milne laboratory has demonstrated that H3K79me2/3 marks a key subset of MLL-AF4-bound enhancers [Godfrey et al., [2019]], that inhibition of DOT1L (the only known H3K79 methyltransferase) leads to downregulation of key MLL-AF4 target genes [Godfrey et al., [2017], Godfrey et al., [2019]] and that H3K79me2/3 promotes enhancer-promoter interactions that drive expression of *PROM1* in *MLL-AF4* ALL [Godfrey et al., [2020]]. However, these studies were carried out in established leukaemias or cell lines, and a prospective analysis of H3K79me2 changes in a single human sample before and after leukaemic transformation was lacking. Therefore, using this new model system, I analysed H3K79me2 levels genome-wide in primary human FL cells before and after ^{CRISPR}*MLL-AF4*-driven leukaemic transformation. I found that H3K79me2 was in fact strikingly increased specifically at MLL-AF4 binding sites, suggesting this increased H3K79me2 was a direct consequence of

MLL-AF4 binding rather than an indirect effect of high expression levels of MLL-AF4-bound genes. However, from a biochemical perspective, these data were unexpected. Structural analyses have shown that DOT1L interacts with the ANC Homology Domain (AHD) of MLL-AF4 complex members, ENL and AF9 [Mueller et al., [2007], Leach et al., [2013], Li et al., [2014]] but this domain is also the site of interaction with the AF4 fusion partner, meaning that MLL-AF4 and DOT1L form mutually exclusive complexes with ENL/AF9 [Yokoyama et al., [2010], Biswas et al., [2011]]. Thus, ENL and AF9 cannot facilitate direct interaction between the MLL-AF4 complex and DOT1L, and in fact, there is no known mechanism by which the MLL-AF4 fusion protein is capable of directly recruiting DOT1L. Therefore, while these data suggest that the MLL-AF4 complex is somehow recruiting DOT1L to induce increased H3K79me2 specifically at MLL-AF4 binding sites, further work is required to understand the biochemical mechanism underlying this activity.

5

Investigating the cellular origin of *MLL-AF4* infant ALL

Contents

5.1 Introduction 114
5.1.1 Previous studies of the cellular origin of *MLL-AF4* infant ALL 114
5.1.2 What defines the MLL-AF4 binding profile in the earliest stages of leukaemogenesis? 116
5.2 Aims 118
5.3 The CXXC domain plays a major role in directing MLL-AF4 binding 118
5.3.1 Generating a FLAG-tagged *MLL-AF4* plasmid 118
5.3.2 In an *MLL*wt cell line, MLL-AF4 binds accessible, unmethylated CpG regions 121
5.4 The chromatin profile of ^{CRISPR}*MLL-AF4+* ALL resembles that of FL B progenitors 124
5.5 ^{CRISPR}*MLL-AF4+* FL LMPP and GMP consistently give rise to a dramatic B lineage expansion *in vitro* . . 132
5.6 ^{CRISPR}*MLL-AF4+* FL ELP and PPB give rise to ALL *in vivo* 139
5.7 Discussion 142

5.1 Introduction

By inducing an *MLL-AF4*/t(4;11) translocation in primary human FL cells, I have shown that ^{CRISPR}*MLL-AF4+* ALL retains human fetal-specific molecular signatures, likely in part due to MLL-AF4 fusion protein binding at fetal-specific genes, recapitulating what is seen in *MLL-AF4* infant ALL patients. These data suggest that *MLL-AF4* infant ALL maintains some of the epigenetic and transcriptomic features of the cells in which it first arises. This therefore raises the question, does the particular fetal HSPC subpopulation in which the leukaemia first arises also affect its phenotypic and molecular features? For example, all ^{CRISPR}*MLL-AF4+* ALLs derived from bulk CD34+ FL HSPCs were of the *HOXA*$^{\text{lo}}$ *MLL-AF4* infant ALL subtype (Figure 3.4). However, it may be that the *HOXA*$^{\text{hi}}$ and *HOXA*$^{\text{lo}}$ signatures are a product of the HSPC subpopulation in which the leukaemia first arises. As such, the more aggressive *HOXA*$^{\text{lo}}$ blasts, derived from a HSPC subpopulation which had already shut down *HOXA* expression, may have out-competed *HOXA*$^{\text{hi}}$ blasts derived from a *HOXA*-expressing HSPC subpopulation when engrafting in NSG mice. The cellular origins of *MLL-AF4* infant ALL, and what effect heterogeneous cells-of-origin may have on the clinico-pathological features of different patients, remain unknown, largely due to the lack of appropriate model systems to address this question. In this chapter, I set out to investigate the cellular origin of *MLL-AF4* infant ALL using our new CRISPR/Cas9-based modelling approach.

5.1.1 Previous studies of the cellular origin of *MLL-AF4* infant ALL

There are a small number of previous studies that have aimed to identify the cellular origin of *MLL-AF4* infant ALL. As mentioned in Chapter 1, the Rabbitts laboratory used the *Mll-AF4* "invertor" model to induce *Mll-AF4* expression in early progenitors (*Lmo2-Cre*) and lymphoid progenitors (*Lck-Cre, CD19-Cre* and *Rag-Cre*), and found that, while expression in early progenitors led to embryonic lethality, expression in any lymphoid progenitor tested led to a long-latency B cell lymphoma [Metzler et al., [2006]]. They therefore concluded that *Mll-AF4* must arise in a

HSPC subpopulation with lymphoid potential to give rise to ALL. [Barrett et al., [2016]] and [Malouf & Ottersbach, [2018]] used the same *Mll-AF4* invertor model to induce *Mll-AF4* expression in all definitive haematopoietic cells generated during mouse embryonic development (*VE-Cadherin-Cre*) and subsequently sorted 3 groups of HSPC subpopulations from the FL of these mouse embryos (HSC/MPPs, LMPPs, and Lin-ckit+ (LK)/CLPs). They found that in methylcellulose assays, *Mll-AF4+* HSC/MPP and LMPPs showed greater proB cell output and HSC self-renewal compared to wild-type HSC/MPPs and LMPPs and to *Mll-AF4+* LK/CLP. When engrafted into recipient mice, *Mll-AF4+* LMPPs showed significantly increased engraftment and higher WBC counts in the peripheral blood, although this decreased over time. None of these transplants gave rise to leukaemia in recipient mice. Despite the lack of full-blown leukaemia, [Malouf & Ottersbach, [2018]] concluded that FL LMPP represent the cell-of-origin of *MLL-AF4* infant ALL, in-keeping with the conclusion from [Metzler et al., [2006]] that the target cell was likely to be lymphoid-primed but multipotent.

Josef Vormoor's group has utilised PDX models of infant and childhood ALL in an attempt to determine the frequency and immunophenotype of leukaemia-initiating cells. [le Viseur et al., [2008]] intrafemorally transplanted CD34+CD19-, CD34+CD19+ and CD34-CD19+ populations into NOD/SCID mice. All three subpopulations were found to engraft and reconstitute the leukaemia with the same pattern of surface markers seen in the original, unsorted patient sample for at least four sequential transplantations. A minimum cell dose of 2,000 cells was shown to be sufficient to re-establish the leukaemia in NOD/SCID mice. Rehe et al., [2013] further examined whether CD10 or CD20 could act as markers to enrich for leukaemia propagating activity. Again, they found leukaemia-initiating cells in all immunophenotypic subpopulations. Bardini et al., [2015] sorted six immunophenotypic subpopulations from diagnostic BM or peripheral blood samples from *MLL-AF4* infant ALL patients: CD34+CD19+NG2+, CD34+CD19+NG2-, CD34-CD19+NG2+, CD34-CD19+NG2-, CD34-CD19-, CD34+CD19- and transplanted into NSG mice. They found that all CD19+ cells fractions contained

leukaemia-initiating cells. CD34-CD19- cells did not engraft and CD34+CD19- cells from one of seven patients gave rise to *MLL*wt multilineage reconstitution. NG2 showed no ability to enrich for leukaemia-initiating cells. Therefore, none of the surface markers examined so far (CD34, CD19, CD10, CD20 and NG2) have been found to enrich for leukaemia-initiating cells in infant ALL.

Taking a different approach, Tirtakusuma et al., [2021] examined *MLL-AF4* expression in lineage switch acute leukaemia (LSAL) and MPAL patients, including infants, children and adults. In one LSAL patient, they observed *MLL-AF4* expression in MPPs and all subsequent haematopoietic cell types, including fully differentiated monocytes, NK, B and T cells. In another MPAL patient, they observed *MLL-AF4* expression in HSC/MPPs and blasts only, and these *MLL-AF4+* HSCs were maintained through serial xenotransplantations. Based on these data, they suggest that *MLL-AF4* can arise in an early haematopoietic progenitor that acts as a pre-leukaemic cell capable of giving rise to B/myeloid blasts, as well as fully differentiated haematopoietic cells. These somewhat conflicting data from *de novo* murine models, PDXs and primary patient studies highlight the need for further study of the cellular origins of *MLL-AF4* infant ALL.

5.1.2 What defines the MLL-AF4 binding profile in the earliest stages of leukaemogenesis?

To understand how the cell-of-origin may influence the molecular profile of *MLL-AF4* infant ALL, it is important to investigate the key factors that define MLL-AF4 fusion protein binding in the earliest stages of leukaemogenesis. Exactly, how MLL fusion proteins are recruited to target genes remains an open question in the field. Some suggest that Menin/LEDGF recruit the fusion protein to target genes [Yokoyama et al., [2005], Yokoyama & Cleary, [2008], Krivtsov et al., [2019]]. However, recent data using a small molecule inhibitor of the Menin/MLL interaction show that disrupting this interaction does not lead to loss of MLL fusion protein binding at all target genes [Krivtsov et al., [2019]]. Moreover, Milne et al., [2010]

showed that a minimal MLL CXXC domain could be recruited to *Hoxa9* in the absence of the Menin interaction site.

Kerry et al., [2017] combined the use of BioCAP and ATAC-seq assays to describe the chromatin landscape of the SEM cell line. BioCAP is an assay originally developed by the Klose laboratory to detect unmethylated CpG regions of DNA [Long et al., [2018]]. A purified biotinylated ZF-CXXC domain is used to bind and pull down all unmethylated CpG regions of purified (non-chromatinised) genomic DNA, regardless of whether these regions had an accessible chromatin profile in the cell [Long et al., [2018]]. On the other hand, ATAC-seq uses the Tn5 transposase to identify all accessible regions of chromatin in the cell [Buenrostro et al., [2013]]. Therefore, combining these techniques can identify all accessible, unmethylated CpG regions in a cell that could be bound by a CXXC domain-containing protein. Comparing these data to MLL-AF4 ChIP-seq in SEM cells, Kerry et al., [2017] found that the vast majority of MLL-AF4 binding sites were at accessible, unmethylated CpG regions, further suggesting that the CXXC domain plays an important role in defining the MLL-AF4 binding profile. Recent studies have found that ATAC-seq is a useful assay for defining haematopoietic cell hierarchies, potentially even superior to transcriptomic analysis [Corces et al., [2016]]. If MLL-AF4 binding is driven by the accessible, unmethylated CpG regions present in a particular HSPC subpopulation in the earliest stages of leukaemogenesis, this may explain how the target cell context could influence leukaemia initiation and the phenotypic and molecular features of the resultant *MLL-AF4* infant ALL. However, SEM cells represent an established leukaemia, derived from an *MLL-AF4* ALL patient at relapse, and therefore cannot be used to study the binding profile of MLL-AF4 during leukaemia initiation. In this chapter, I introduce MLL-AF4 into an *MLL*wt cell line to study its binding profile in the earliest stages following its expression, and investigate the chromatin architecture of these early MLL-AF4 binding sites.

5.2 Aims

The overall aim of this chapter was to identify the cellular origin of *MLL-AF4* infant ALL, which I split into the following sub-aims:

- Identify key factors directing MLL-AF4 fusion protein binding in the earliest stages of its expression
- Compare the gene expression and chromatin profile of ^{CRISPR}*MLL-AF4+* ALL to sorted FL HSPC subpopulations to generate hypotheses about possible cell(s)-of-origin
- Test these hypotheses, both *in vitro* and *in vivo*, to identify the cellular origin of *MLL-AF4* infant ALL

5.3 The CXXC domain plays a major role in directing MLL-AF4 binding

In order to investigate the key factors that direct MLL-AF4 fusion protein binding in the earliest stages of leukaemogenesis, I set out to express MLL-AF4 in an *MLL*wt cell line and characterise the chromatin landscape at MLL-AF4 binding sites. I chose the HEK-293T cell line due to its high efficiency of transfection and because Dr Jon Kerry of the Milne laboratory had previously generated BioCAP and ChIP-seq datasets in this cell line.

5.3.1 Generating a FLAG-tagged *MLL-AF4* plasmid

Using the GeneArt service (Thermo Fisher) I designed plasmids containing the fragments of *MLL* and *AF4* present in ^{CRISPR}*MLL-AF4+* ALL (*MLL* exons 1-10, *AF4* exons 4-21), and a NeonGreen fluorescent reporter. In addition, I designed oligonucleotides encoding 3 repeating FLAG-tag sequences. Using the InFusion cloning method, I cloned *MLL-AF4*, with a FLAG-tag in the breakpoint region between *MLL* and *AF4* and a NeonGreen reporter downstream of *AF4*, separated by a T2A sequence, into a plasmid containing an SFFV promoter (generously

provided by Dr Philip Hublitz, WIMM genome engineering facility) (Figure 5.1a). As an empty vector (EV) control, I cloned the NeonGreen fluorescent reporter alone downstream of the SFFV promoter (Figure 5.1b). Successful cloning was detected using diagnostic restriction digest.

To further validate both *MLL-AF4* and EV plasmids, I transfected each plasmid into HEK-293T cells, which I then harvest 24 hours post-transfection. Fluorescent microscopy confirmed expression of NeonGreen in both *MLL-AF4*- (Figure 5.1c) and EV-transfected (Figure 5.1d) cells, while RT-qPCR showed *MLL-AF4* was only expressed in *MLL-AF4*-transfected cells (Figure 5.1e). Western blot confirmed the presence of the MLL-AF4 fusion protein in *MLL-AF4*-transfected but not EV-transfected cells (Figure 5.1f). MLL-AF4 protein abundance in *MLL-AF4*-transfected HEK-293T cells was strikingly higher than in the SEM cell line (Figure 5.1f).

Finally, to test the functionality of the transfected MLL-AF4 fusion protein, and to determine the optimum time-point at which to assess transfected cells, I carried out an MLL-AF4 (anti-FLAG antibody) ChIP-qPCR time-course. I found that I could detect some MLL-AF4 binding from 16 hours post-transfection, which further increased by 24 hours post-transfection, at known MLL-AF4 target genes, *HOXA9, MEIS1* and *MEF2C* (Figure 5.1g).

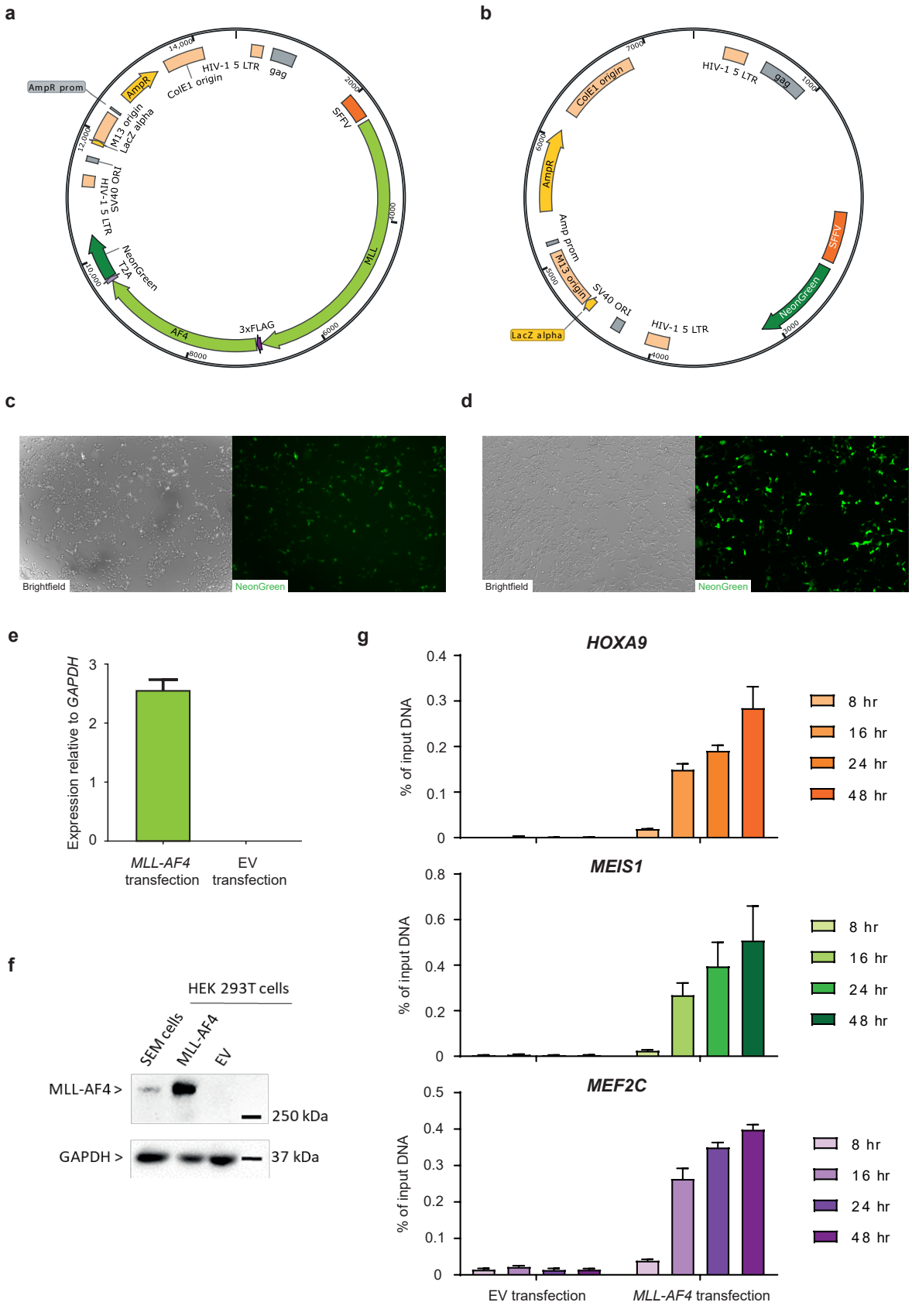

a
AmpR prom
AmpR
ColE1 origin
HIV-1 5 LTR
gag
SFFV
MLL
3xFLAG
AF4
NeonGreen T2A
HIV-1 5 LTR
SV40 ORI
M13 origin
LacZ alpha
b
ColE1 origin
HIV-1 5 LTR
gag
AmpR
Amp prom
M13 origin
SV40 ORI
LacZ alpha
HIV-1 5 LTR
SFFV
NeonGreen
c
Brightfield
NeonGreen
d
Brightfield
NeonGreen
e
Expression relative to GAPDH
MLL-AF4 transfection
EV transfection
f
HEK 293T cells
SEM cells
MLL-AF4
EV
MLL-AF4 >
250 kDa
GAPDH >
37 kDa
g
HOXA9
MEIS1
MEF2C
% of input DNA
8 hr
16 hr
24 hr
48 hr
EV transfection
MLL-AF4 transfection

Figure 5.1

Figure 5.1: Generating a FLAG-tagged *MLL-AF4* plasmid. a. Map of *MLL-AF4* plasmid showing SFFV promoter in orange, *MLL-AF4* in light green (with 3 repeating FLAG-tag sequences in the breakpoint region shown in dark purple), the T2A sequence in light purple, followed by the NeonGreen reporter in dark green. This plasmid contained an ampicillin resistance genes (AmpR in yellow). **b.** Map of EV plasmid showing SFFV promoter in orange, followed by the NeonGreen reporter in dark green. This plasmid contained an ampicillin resistance genes (AmpR in yellow). **c.** Representative microscopy images for *MLL-AF4*-transfected HEK-293T cells 24 hours post-transfection. The panel on the left shows a brightfield image, while the panel on the right shows NeonGreen fluorescence for the same cells. **d.** c. Representative microscopy images for EV-transfected HEK-293T cells 24 hours post-transfection. The panel on the left shows a brightfield image, while the panel on the right shows NeonGreen fluorescence for the same cells. **e.** RT-qPCR data showing *MLL-AF4* expression in *MLL-AF4*- and EV-transfected HEK-293T cells 24 hours post-transfection relative to *GAPDH* expression (n=3). Data shown as mean ± SEM. **f.** Representative Western blot showing MLL-AF4 fusion protein abundance in *MLL-AF4*-transfected HEK293T cells 24 hours post-transfection compared to the SEM cell line (positive control) and EV-transfected HEK-293T cells (negative control). The MLL-AF4 fusion protein is >250kDa, while the 37kDa GAPDH protein was used as a loading control. **g.** ChIP-qPCR data for MLL-AF4 (anti-FLAG antibody) at the promoters of *HOXA9* (orange), *MEIS1* (green) and *MEF2C* (purple) in *MLL-AF4*- and EV-transfected HEK-293T cells at 8, 16, 24 and 48 hours post-transfection (n=3). Data shown as mean ± SEM.

5.3.2 In an *MLL*wt cell line, MLL-AF4 binds accessible, unmethylated CpG regions

Having validated the *MLL-AF4* plasmid, I again transfected HEK-293T cells with *MLL-AF4* and sorted NeonGreen+ cells 24 hours post-transfection. I then carried out MLL-AF4 (anti-FLAG antibody) ChIP-seq in these sorted cells. Previously, Dr Jon Kerry in the Milne laboratory transfected a plasmid encoding a FLAG-tagged MLL CXXC domain into HEK-293T cells and carried out ChIP-seq 24 hours post-transfection. Therefore, these data could be used to compare the binding profile of MLL-AF4 to the binding profile of a minimal MLL CXXC domain in the same cell type. I further compiled a dataset of untransfected HEK-293T cell BioCAP data (previously generated by Dr Jon Kerry), and publicly available HEK-293T cell ATAC-seq, and H3K27ac and H3K4me3 ChIP-seq data (`https://www.encodeproject.org/`).

Combining these datasets, I found that all MLL-AF4 and MLL CXXC domain binding detected 24 hours post-transfection occurred at unmethylated CpG regions.

Moreover, there was a striking overlap between the binding profiles of MLL-AF4 and the MLL CXXC domain (Figure 5.2a left). However, there were many more BioCAP peaks not bound by MLL-AF4 or the MLL CXXC domain. Further analysis revealed that the vast majority of MLL-AF4 and MLL CXXC domain ChIP-seq peaks overlapped with ATAC-seq peaks (Figure 5.2a left), suggesting that MLL-AF4 binds most accessible, unmethylated CpGs in HEK-293T cells in a manner that is almost identical to the MLL CXXC domain alone. Looking in more detail at these MLL-AF4 and MLL CXXC domain binding sites, I found that there was also a striking overlap with active histone marks, H3K27ac and H3K4me3 (Figure 5.2a right). To try to understand which factors were most important for MLL-AF4 binding, I carried out a correlation analysis between all datasets, taking into account the intensity of the signal at each peak, and found that MLL CXXC domain ChIP-seq and BioCAP data were most highly correlated with MLL-AF4 ChIP-seq, followed by ATAC-seq and the active histone marks (Figure 5.2b).

I next characterised the types of regions at which MLL-AF4 was binding, and found that 80% of MLL-AF4 binding sites were occurring at promoters (±2kb from a gene TSS) (Figure 5.2c and 5.2d), while only 11% were occurring at enhancers (H3K27ac peaks >2kb from a gene TSS, both intragenic and intergenic) (Figure 5.2c and 5.2e). This is in contrast to my previous findings from ^{CRISPR}*MLL-AF4+* ALL in 4, where MLL-AF4 binding sites occurred less frequently at promoters (43%) and more often at regions distal to the promoter (both intragenic and intergenic) (Figure 4.5b), further suggesting that the MLL-AF4 binding profile is, in part, determined by the cell type in which it arises. In-keeping with this idea, there was very little overlap between the MLL-AF4 binding profile in *MLL-AF4*-transfected 293T cells and the SEM cell line (Figure 5.2f).

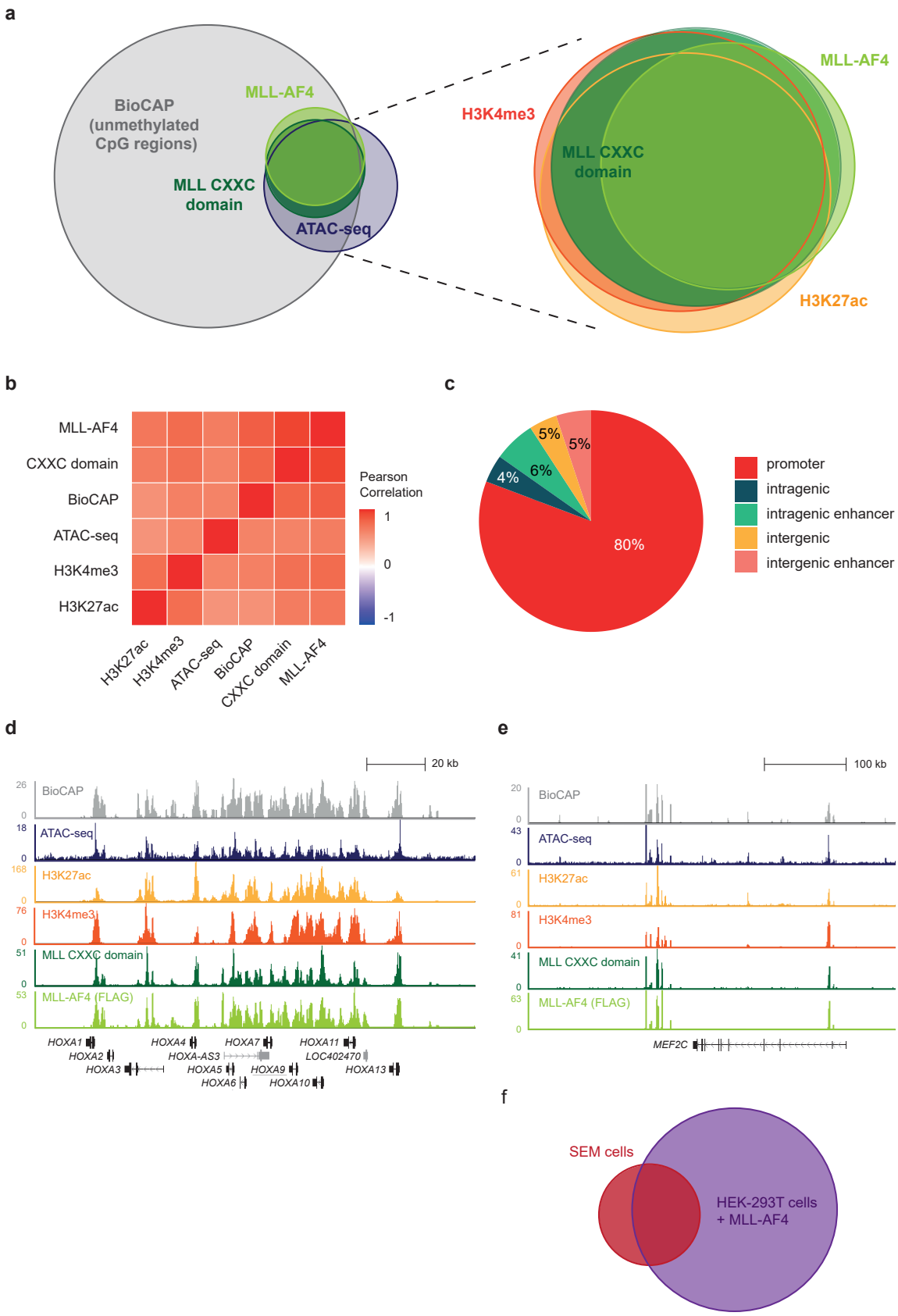

a
BioCAP
(unmethylated
CpG regions)
MLL-AF4
MLL CXXC
domain
ATAC-seq
H3K4me3
MLL-AF4
MLL CXXC
domain
H3K27ac
b
MLL-AF4
CXXC domain
BioCAP
ATAC-seq
H3K4me3
H3K27ac
Pearson
Correlation
1
0
-1
H3K27ac
H3K4me3
ATAC-seq
BioCAP
CXXC domain
MLL-AF4
c
5%
5%
4%
6%
80%
promoter
intragenic
intragenic enhancer
intergenic
intergenic enhancer
d
20 kb
BioCAP
ATAC-seq
H3K27ac
H3K4me3
MLL CXXC domain
MLL-AF4 (FLAG)
HOXA1
HOXA2
HOXA3
HOXA4
HOXA-AS3
HOXA5
HOXA6
HOXA7
HOXA9
HOXA10
HOXA11
LOC402470
HOXA13
e
100 kb
BioCAP
ATAC-seq
H3K27ac
H3K4me3
MLL CXXC domain
MLL-AF4 (FLAG)
MEF2C
f
SEM cells
HEK-293T cells
+ MLL-AF4

Figure 5.2

Figure 5.2: MLL-AF4 binds accessible, unmethylated CpG regions following transfection in HEK-293T cells. a. Venn diagrams showing overlap between BioCAP, ATAC-seq and ChIP-seq peaks in HEK-293T cells. (left) Venn diagram showing overlap between BioCAP and ATAC-seq peaks in untransfected HEK-293T cells, and MLL-AF4 and the MLL CXXC domain ChIP-seq peaks in HEK-293T cells 24 hours post-transfection. (right) Venn diagram showing overlap between H3K27ac and H3K4me3 ChIP-seq peaks in untransfected HEK-293T cells, and MLL-AF4 and the MLL CXXC domain ChIP-seq peaks specifically at accessible, unmethylated CpG regions (overlapping BioCAP and ATAC-seq peaks only) in HEK-293T cells 24 hours post-transfection. **b.** Correlation matrix of all HEK-293T cell datasets. The colour scale represents a Pearson correlation score, and datasets are order based on degree of correlation with one another. **c.** Pie chart representing proportion of MLL-AF4 (anti-FLAG antibody) ChIP-seq peaks at each region subtype in HEK-293T cells 24 hours post-transfection. Promoter = peaks ±2kb from a gene TSS. Intragenic = peaks inside the gene body >2kb from the TSS with no overlapping H3K27ac peak. Intragenic enhancer = peaks inside the gene body > 2kb from the TSS with overlapping H3K27ac peak. Intergenic = peaks outside the gene body >2kb from the TSS with no overlapping H3K27ac peak. Intergenic enhancer = peaks outside the gene body > 2kb from the TSS with overlapping H3K27ac peak. **d.** Representative BioCAP, ATAC-seq and ChIP-seq traces showing MLL-AF4 (light green) and MLL CXXC domain (dark green) binding at the promoters of *HOXA* cluster genes. Scale = reads/bp normalised to 10^7 total reads. **e.** Representative BioCAP, ATAC-seq and ChIP-seq traces showing MLL-AF4 (light green) and MLL CXXC domain (dark green) binding at an intergenic enhancer (overlapping ATAC-seq (dark blue) and H3K27ac ChIP-seq (orange) peaks) downstream of the *MEF2C* gene. Scale = reads/bp normalised to 10^7 total reads. **f.** Venn diagram showing overlap between MLL-AF4 ChIP-seq peaks in the SEM cell line (overlapping MLL-N and AF4-C ChIP-seq; red) and MLL-AF4 (anti-FLAG antibody) ChIP-seq in HEK-293T cells 24 hours post-transfection (purple).

These data best fit a model whereby, in the earliest stages of leukaemogenesis, MLL-AF4 is unlikely to significantly alter the chromatin landscape of the cell but rather binds to unmethylated CpG regions that are already accessible in that cell type. In this way, the chromatin landscape of the cell-of-origin may play a major role in defining the initial binding profile of MLL-AF4, which could then be maintained in the established leukaemia.

5.4 The chromatin profile of ^{CRISPR}*MLL-AF4+* ALL resembles that of FL B progenitors

As data from the HEK-293T cell transfection experiments suggested that analysing the chromatin landscape of HSPC subpopulations may provide insight into the possible cellular origin of *MLL-AF4* infant ALL, I next generated ATAC-seq data

for sorted FL HSPC subpopulations of interest, with a focus on the fetal B lineage differentiation pathway (Figure 5.3a and 5.3b), as well as for hCD45+CD19+ ^{CRISPR}*MLL-AF4*+ proB ALL blasts sorted from the BM of primary recipient mice.

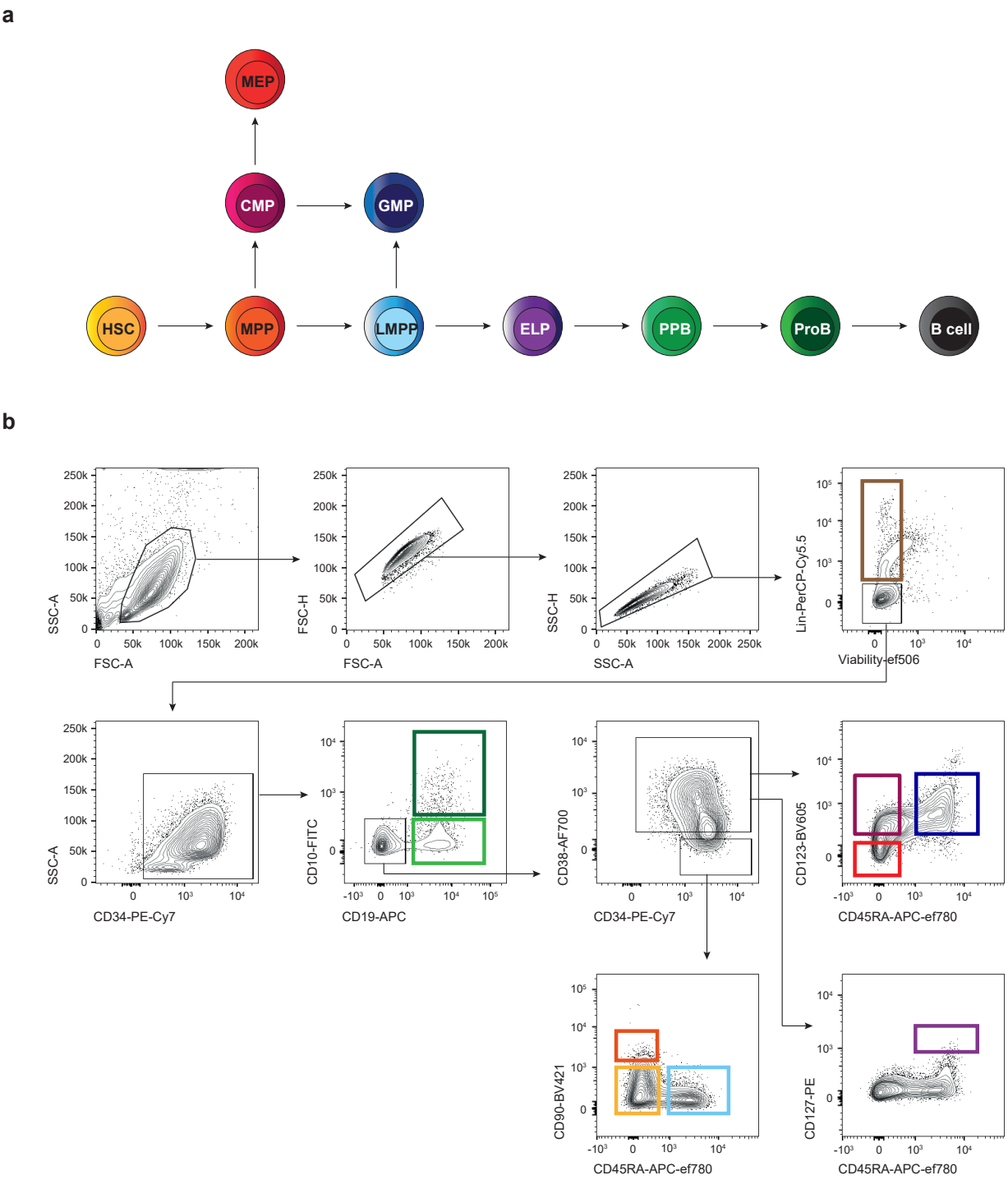


Figure 5.3

Figure 5.3: Sorting strategy for FL HSPC subpopulation ATAC-seq. a. Schematic representation of FL HSPC subpopulations of interest. Arrows represent the fetal haematopoietic differentiation trajectory based on previous functional analyses [O'Byrne et al., [2019]], whereby HSC can give rise to all progenitor types, whereas PB can only give rise to mature B cells. **b.** Representative flow cytometry plots showing gating strategy to sort an FL HSPC subpopulations of interest. The sort gate for each subpopulation shares the same colour as the schematic representation above. Arrows represent sequential gating strategy. Lin = CD2, CD3, CD14, CD16, CD56, CD235a.

I carried out PC analysis and found that PC1 represented the B lineage differentiation trajectory from HSCs through to mature B cells (Figure 5.4a), mirroring the FL haematopoietic hierarchy as previously determined by FACS and functional analyses [Roy et al., [2012], O'Byrne et al., [2019]]. PC2 appeared to represent the FL myeloid/erythroid differentiation pathway. Interestingly, GMPs clustered closely with LMPPs and away from CMPs, suggesting they had a more lymphoid-primed chromatin landscape (Figure 5.4a). ^{CRISPR}*MLL-AF4+* samples clustered with the committed B progenitors and B cells, in-keeping with their CD19+CD34+/- immunophenotype (Figure 5.4a). I next wanted to investigate how FL HSPC ATAC-seq profiles correlated with ^{CRISPR}*MLL-AF4+* MLL-AF4 ChIP-seq. Based on their clustering in the PC analysis (Figure 5.4a), I generated shared peaksets for HSCs and MPPs (HSC/MPP), LMPPs and GMPs (LMPP/GMP), and PPBs and PBs (PPB/PB). I then examined the overlap between HSC/MPP, LMPP/GMP, ELP, PPB/PB ATAC-seq peaks and ^{CRISPR}*MLL-AF4+* MLL-AF4 ChIP-seq peaks. This analysis was based on a binary readout of whether a peak was present or not in a particular cell type, and did not take into account intensity of the signal at that peak. There was substantial overlap between all ATAC-seq datasets. However, the only ATAC-seq dataset that overlapped with 100% of ^{CRISPR}*MLL-AF4+* MLL-AF4 ChIP-seq peaks were the shared PPB/PB peaks (Figure 5.4b), further suggesting that, of all FL HSPCs, ^{CRISPR}*MLL-AF4+* ALL is epigenetically most similar to committed FL B progenitors.

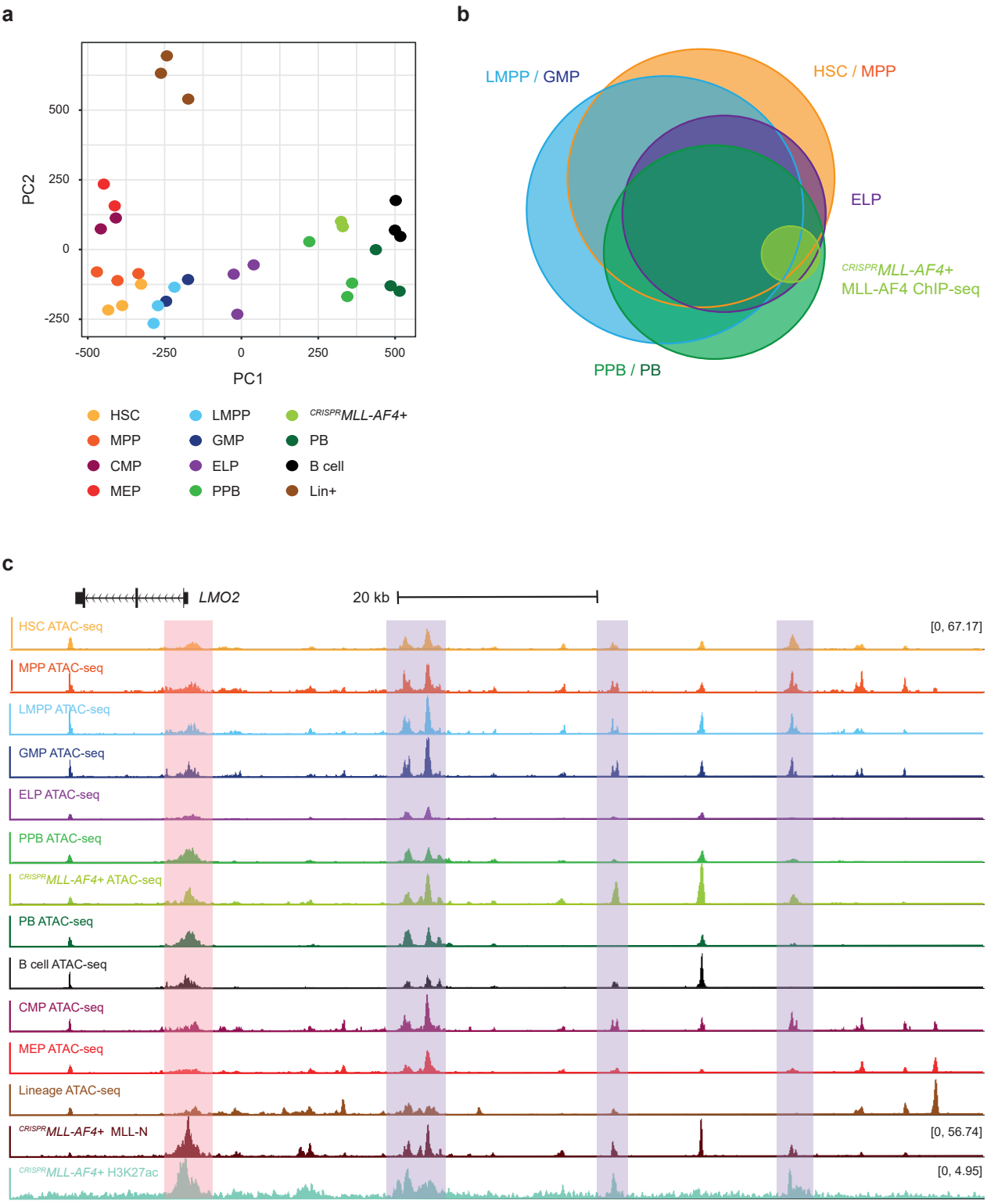

a
500
250
0
-250
PC2
-500
-250
0
250
500
PC1
HSC
MPP
CMP
MEP
LMPP
GMP
ELP
PPB
CRISPRMLL-AF4+
PB
B cell
Lin+
b
LMPP / GMP
HSC / MPP
ELP
CRISPRMLL-AF4+
MLL-AF4 ChIP-seq
PPB / PB
c
LMO2
20 kb
HSC ATAC-seq
[0, 67.17]
MPP ATAC-seq
LMPP ATAC-seq
GMP ATAC-seq
ELP ATAC-seq
PPB ATAC-seq
CRISPRMLL-AF4+ ATAC-seq
PB ATAC-seq
B cell ATAC-seq
CMP ATAC-seq
MEP ATAC-seq
Lineage ATAC-seq
CRISPRMLL-AF4+ MLL-N
[0, 56.74]
CRISPRMLL-AF4+ H3K27ac
[0, 4.95]

Figure 5.4

Figure 5.4: The chromatin profile of ^{CRISPR}*MLL-AF4+* ALL resembles committed FL B progenitors. a. PC analysis of ATAC-seq data for all FL HSPC subpopulations and ^{CRISPR}*MLL-AF4+* ALL blasts, based on ATAC-seq signal at 183,905 peaks present in any and/or all samples quantified using the R DiffBind package. **b.** Venn diagram showing overlap between ATAC-seq peaks in HSC/MPP (shared HSC and MPP peaks, orange), LMPP/GMP (shared LMPP and GMP peaks, blue), ELP (purple), PPB/PB (shared PPB and PB peaks, dark green), and MLL-AF4 ChIP-seq peaks in ^{CRISPR}*MLL-AF4+* ALL (light green). **c.** Representative ATAC-seq and ChIP-seq traces at the *LMO2* gene, showing peaks at the promoter (gene TSS; highlighted in pink) and at upstream intergenic enhancers (overlapping ATAC-seq and H3K27ac ChIP-seq peaks in ^{CRISPR}*MLL-AF4+* ALL; highlighted in purple). Scale (in square brackets) = reads/bp normalised to 10^7 total reads.

Interestingly, at stem cell-associated genes such as *LMO2*, which is highly expressed in early FL progenitors (HSCs, MPPs and LMPPs) as well as ^{CRISPR}*MLL-AF4+* ALL but not in FL CBPs, there were clear ATAC-seq peaks at the promoter as well as several upstream intergenic enhancers in FL PPBs, PBs and B cells (Figure 5.4c), likely contributing to the substantial overlap in ATAC-seq peaks between different FL HSPC subpopulations. Therefore, ATAC-seq signal alone cannot give a comprehensive picture of the molecular profile of the cell. I hypothesised that combining RNA-seq and ATAC-seq data may give additional insight into the molecular profile of each HSPC subpopulation, and further help to investigate the cellular origin of *MLL-AF4* infant ALL.

Returning to the sorted FL HSPC subpopulation RNA-seq data previously generated in the Roberts and Roy laboratory, I carried out an ANOVA-style comparison of all FL HSPC subpopulations to identify 729 significantly differentially expressed genes (FDR<0.05). Based on this gene set, ^{CRISPR}*MLL-AF4+* ALL clustered with LMPPs and GMPs in a heatmap analysis (dendrogram based in euclidean distance), and were clearly separate from CBPs (representing a mixture of PPB and PB progenitors) (Figure 5.5a). In addition, I carried out PC analysis of FL HSPCs, ^{CRISPR}*MLL-AF4+* ALL and *MLL-AF4* infant ALL patients [Agraz-Doblas et al., [2019]] based on these 729 genes. Again, I observed ^{CRISPR}*MLL-AF4+* ALL and *MLL-AF4* infant ALL patients clustering with progenitors upstream in the haematopoietic hierarchy from CBPs (Figure 5.5b). While the *MLL-AF4* infant

ALL patients clustered tightly with early progenitors (HSCs, MPPs and LMPPs), both proB and preB ^{CRISPR}*MLL-AF4+* ALLs fell between these early progenitors and the fetal-specific ELPs (Figure 5.5b). A contributing factor to this may be that Agraz-Doblas, et al., [2019] sorted CD19+CD34+ blasts from *MLL-AF4* infant ALL patients to generate their RNA-seq data, potentially enriching for blasts with a more stem-like molecular profile.

Interestingly, in contrast with the RNA-seq data, ^{CRISPR}*MLL-AF4+* ALL clustered with PPB and PB progenitors in a heatmap analysis based on ATAC-seq peaks associated with these 729 genes (dendrogram based on Euclidean distance) (Figure 5.5c). Therefore, even though the expression levels of these genes places ^{CRISPR}*MLL-AF4+* ALL higher in the FL haematopoietic hierarchy than CBPs, the ^{CRISPR}*MLL-AF4+* chromatin landscape at these genes correlates most with PPB and PB progenitors. As a further confirmation, I included *MLL-AF4* infant ALL and *MLL-AF4* childhood ALL patient ATAC-seq data (generated as part of a collaboration between the Roberts, Roy and Milne laboratories), in a PC analysis based on peaks associated with these 729 genes and found that the patients, as well as ^{CRISPR}*MLL-AF4+* ALL, clustered with PPB and PB progenitors (Figure 5.5d). However, a small proportion of peaks enriched in early progenitors (HSCs, MPPs and LMPPs) were also enriched in ^{CRISPR}*MLL-AF4+* ALL compared to PPB and PB (Figure 5.5c). These peaks were associated with stem-like genes such as *LMO2, NOG* and *CD34* and myeloid genes such as *MPO*. As a further confirmation, I included *MLL-AF4* infant ALL and *MLL-AF4* childhood ALL patient ATAC-seq data (generated as part of a collaboration between the Roberts, Roy and Milne laboratories), in a PC analysis based in peaks associated with these 729 genes and found that the patients, as well as ^{CRISPR}*MLL-AF4+* ALL, clustered with PPB and PB progenitors (Figure 5.5d).

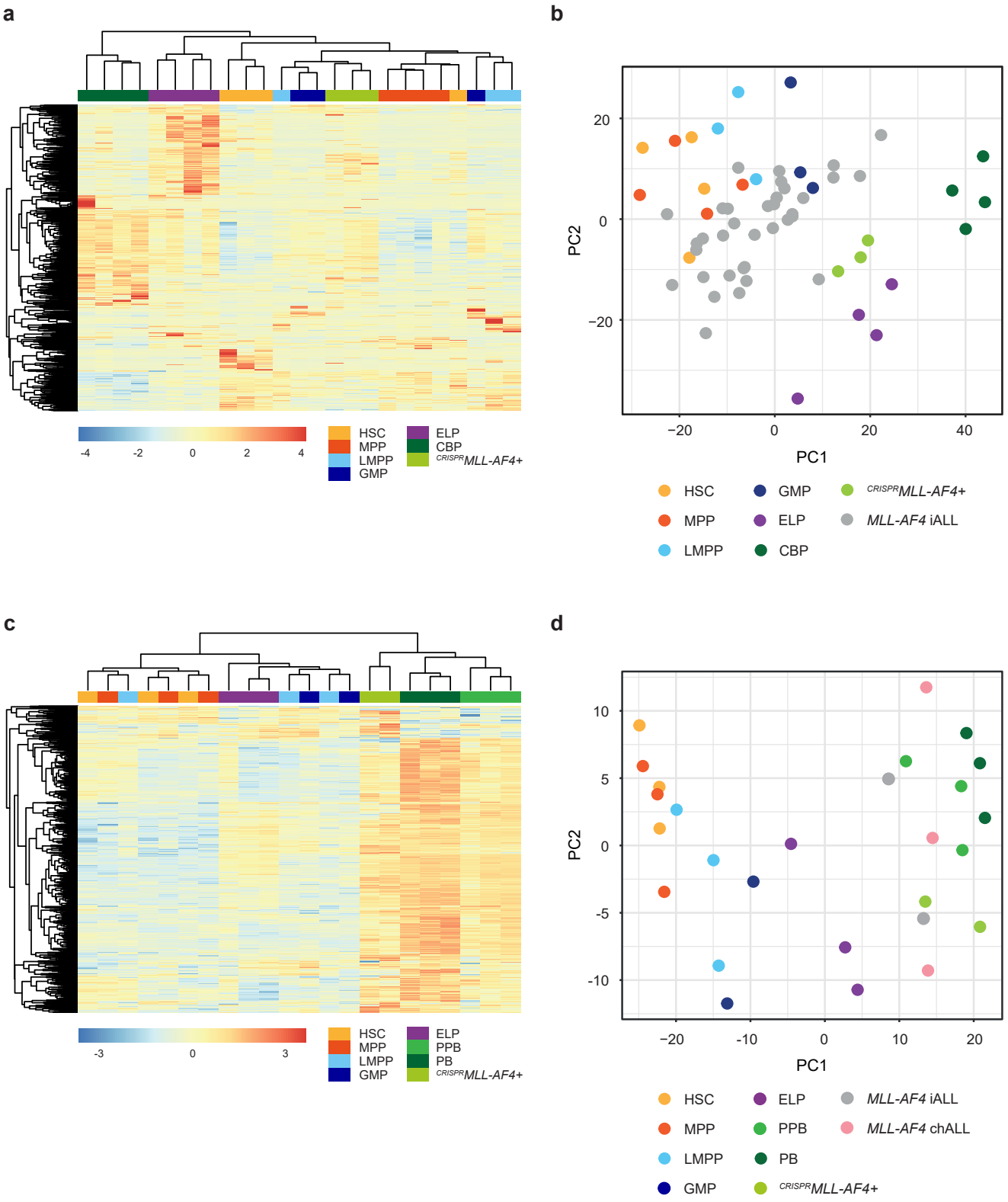

a
-4
-2
0
2
4
HSC
MPP
LMPP
GMP
ELP
CBP
CRISPRMLL-AF4+
b
20
0
−20
PC2
−20
0
20
40
PC1
HSC
MPP
LMPP
GMP
ELP
CBP
CRISPRMLL-AF4+
MLL-AF4 iALL
c
-3
0
3
HSC
MPP
LMPP
GMP
ELP
PPB
PB
CRISPRMLL-AF4+
d
10
5
0
-5
-10
PC2
−20
-10
0
10
20
PC1
HSC
MPP
LMPP
GMP
ELP
PPB
PB
CRISPRMLL-AF4+
MLL-AF4 iALL
MLL-AF4 chALL

Figure 5.5

Figure 5.5: Transcriptomic and epigenetic analysis of *MLL-AF4* ALL based on differentially expressed genes in FL HSPC subpopulations. a. Heatmap showing clustering of FL HSPC subpopulations and ^{CRISPR}*MLL-AF4+* ALL based on 729 differentially expressed genes between all 6 FL HSPC subpopulations analysed (HSC, MPP, LMPP, GMP, ELP, CBP) derived from an ANOVA-style statistical analysis (FDR<0.05, edgeR glm test). The dendrogram is based on Euclidean distance between samples (above) and genes (left). Scale = log2 CPM. **b.** PC analysis of FL HSPC subpopulations, ^{CRISPR}*MLL-AF4+* ALL and *MLL-AF4* infant ALL patients based on 729 significantly differentially expressed genes in the FL HSPC subpopulations. **c.** Heatmap showing clustering of FL HSPC subpopulations and ^{CRISPR}*MLL-AF4+* ALL based on the intensity of ATAC-seq signal at ATAC-seq peaks associated with 729 genes significantly differentially expressed in the FL HSPC subpopulations. The dendrogram is based on Euclidean distance between samples (above) and genes (left). Scale = log2(ATAC-seq reads/bp). **d.** PC analysis of FL HSPC subpopulations, ^{CRISPR}*MLL-AF4+* ALL and *MLL-AF4* infant ALL patients based on the intensity of ATAC-seq signal at ATAC-seq peaks associated with 729 significantly differentially expressed genes in the FL HSPC subpopulations.

Taken together, these data suggest that the epigenetic profile of *MLL-AF4* infant ALL resembles that of FL committed B progenitors. There was a clear correlation between ^{CRISPR}*MLL-AF4+* , PPB and PB ATAC-seq data, and 100% of ^{CRISPR}*MLL-AF4+* MLL-AF4 binding sites occur at regions that have an open chromatin profile in PPB/PB progenitors. However, the effect of MLL-AF4 binding at these sites may vary between different genes, as the gene expression profile of ^{CRISPR}*MLL-AF4+* ALL and *MLL-AF4* infant ALL patients is more similar to early, lymphoid-primed progenitors like LMPPs. Therefore, one could hypothesise a scenario where the *MLL-AF4* /t(4;11) translocation arises in an early progenitor, such as an LMPP, whereby the initial MLL-AF4 binding profile would be directed by accessible, unmethylated CpG regions, possibly in highly expressed genes, in the early progenitor, aberrantly maintaining their expression, but as the cells differentiates part of the way down the B lymphopoietic pathway before stalling, MLL-AF4 is able to bind B progenitor-associated accessible regions. Alternatively, the *MLL-AF4* /t(4;11) translocation may arise in a committed FL B progenitor, and MLL-AF4 binding at lowly expressed but accessible genes, such as *LMO2*, could lead to massive upregulation of these genes and a more stem-like transcriptomic profile. Overall, based on these data, I reached the hypothesis that the cellular

origin of *MLL-AF4* infant ALL was likely to be a FL HSPC with some level of lymphoid priming, anywhere from LMPPs to committed B progenitors. Therefore, I set out to test this functionally.

5.5 *CRISPR*MLL-AF4+ FL LMPP and GMP consistently give rise to a dramatic B lineage expansion *in vitro*

In order to test the leukaemic transformation potential of different FL HSPC subpopulations of interest, I edited bulk FL CD34+ cells as before and then sorted the edited cells after 14-16 hours into their respective FL HSPC subpopulations for co-culture on MS-5 stroma (n=3, 12pcw) (Figure 5.6). 1,000 cells were seeded per well for each HSPC subpopulation, as well as bulk CD34+ cells as a positive control. It was not possible to sort sufficient numbers of edited FL ELP for co-culture assays. Sorted unedited, biologically matched FL HSPC populations were used as a negative control.

By week 2 of co-culture, 6/6 edited LMPP wells (n=3 FL samples) and 5/5 edited GMP wells (n=3 FL samples) showed a clear expansion of CD19+ B lineage cells compared to controls (Figure 5.7), and this expansion was maintained for up to 6 weeks of MS-5 stromal co-culture, at which time the unedited control cell numbers had dropped to zero (Figure 5.8 left). Over the culture period, a small Lin+ population was maintained in edited LMPP wells, while edited GMP wells were >95% CD19+ B lineage cells (Figure 5.8 right). Further immunophenotypic analysis revealed that these B lineage cells in both edited LMPP and GMP wells were CD19+CD10+IgM/IgD- preB cells, as had previously been observed in edited bulk FL CD34+ co-cultures. While the proportion of CD19+CD133+ cells was low in all wells, it appeared to be slightly higher in edited LMPP wells (Figure 5.9a) compared to edited GMP wells (Figure 5.9b), perhaps reflecting the greater *PROM1* expression in the target LMPPs compared to GMPs.

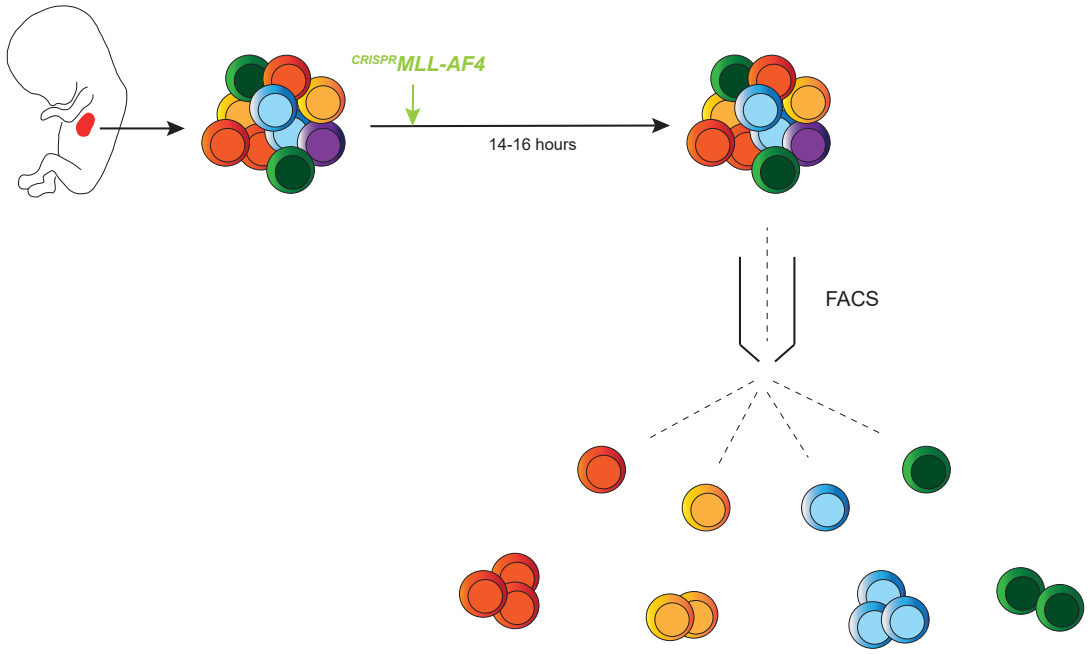


Figure 5.6: Schematic representation of edited FL HSPC subpopulation experiments. Bulk FL CD34+ cells, representing a mixture of different HSPC subpopulations, (mixed colours) were electroporated with Cas9/*MLL*-sgRNA and Cas9/*AF4*-sgRNA RNPs as before, followed by overnight recovery (14-16 hours) before being sorted by FACS into immunophenotypically-defined FL HSPC subpopulations (see Figure 5.3b) for MS-5 stromal co-culture or *in vivo* transplantation into sublethally irradiated NSG mice.

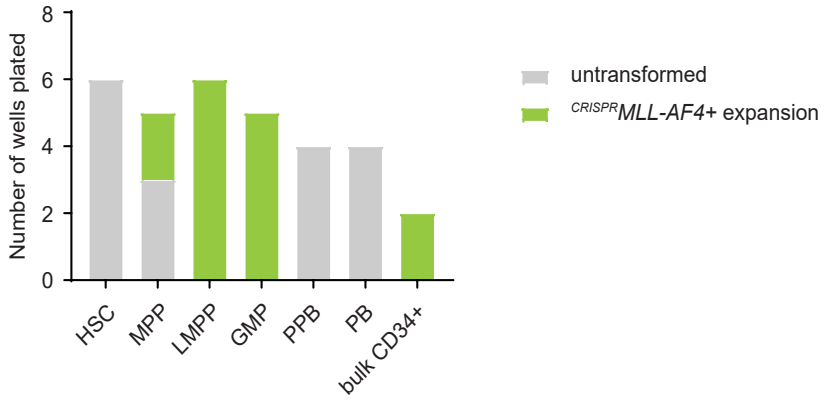


Figure 5.7: FL LMPP and GMP consistently transform *in vitro*. Bar graph showing number of wells plated for each edited FL HSPC subpopulation (n=3 FL samples), and proportion of wells that showed a $^{CRISPR}MLL\text{-}AF4+$ transformation (light green).

At week 4 of co-culture, 2/5 (n=3 FL samples) edited MPP wells showed a B lineage expansion (Figure 5.7 left). Not only was this delayed compared to edited LMPP and GMP wells but it was also less efficient, as not all edited MPP wells underwent transformation. No significant differences were observed between edited and control wells for HSC and PB (Figure 5.8). While there was not a clear B lineage expansion in the edited PPB wells compared to control PPB wells, edited PPB wells could be maintained for longer in MS-5 stromal co-culture (cell numbers dropped to zero in edited PPB wells at week 4 compared to week 3 in unedited control PPB wells) (Figure 5.8 left).

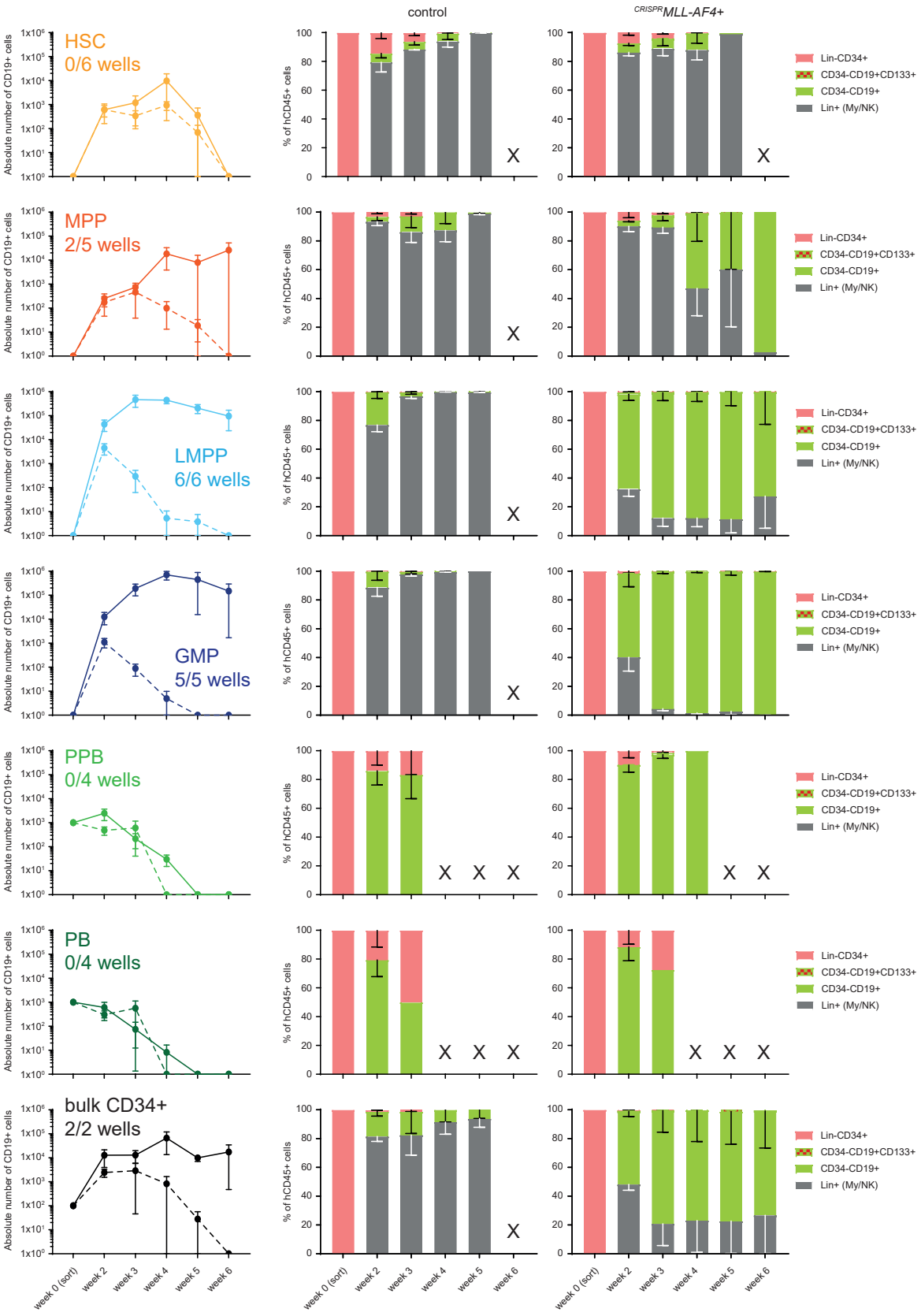

control
CRISPR MLL-AF4+
HSC
0/6 wells
MPP
2/5 wells
LMPP
6/6 wells
GMP
5/5 wells
PPB
0/4 wells
PB
0/4 wells
bulk CD34+
2/2 wells
Absolute number of CD19+ cells
% of hCD45+ cells
Lin-CD34+
CD34-CD19+CD133+
CD34-CD19+
Lin+ (My/NK)
week 0 (sort)
week 2
week 3
week 4
week 5
week 6

Figure 5.8

Figure 5.8: ^{CRISPR}*MLL-AF4+* FL LMPP and GMP consistently give rise to a dramatic B lineage expansion *in vitro*. For each FL HSPC subpopulation (colour-coded and labelled in the left panel), the left panel represent the absolute number of CD19+ B lineage cells/well over time. Solid lines represent edited well. Dashed lines represent unedited control wells (n=2-6 technical replicates from 2-3 FL samples, 12pcw). The centre panel is a quantification of the proportion of immunophenotypically-defined cell types in each unedited control well over time, as a percentage of hCD45+ cells. X = data not shown as there were <10 hCD45+ cells detected by flow cytometry. The right panel is a quantification of the proportion of immunophenotypically-defined cell types in each edited well over time, as a percentage of hCD45+ cells. X = data not shown as there were <10 hCD45+ cells detected by flow cytometry. Data shown as mean ± SEM. Lin+ (My/NK) = CD2, CD3, CD14, CD16, CD56, CD235a (T/erythroid differentiation not supported by MS-5 stromal co-culture assay).

RT-qPCR at week 4 of co-culture confirmed *MLL-AF4* (Figure 5.10a) and *AF4-MLL* (Figure 5.10b) expression in transformed LMPP and GMP wells, as well as the bulk FL CD34+ positive control wells. RT-qPCR data for MPP represents an untransformed well (transformed well data unavailable). Very low levels of *MLL-AF4* and *AF4-MLL* expression were observed in edited PPB wells (Figure 5.10a and 5.10b). This, in combination with the slightly longer survival in culture (Figure 5.8 left) may suggest that the *MLL-AF4*/t(4;11) translocation had occurred in edited PPB cells but it was unable to lead to a significant B lineage expansion in this MS-5 stromal co-culture system. As edited PPB cell numbers were very low, it was not possible to carry out FISH analysis to look for translocation events.

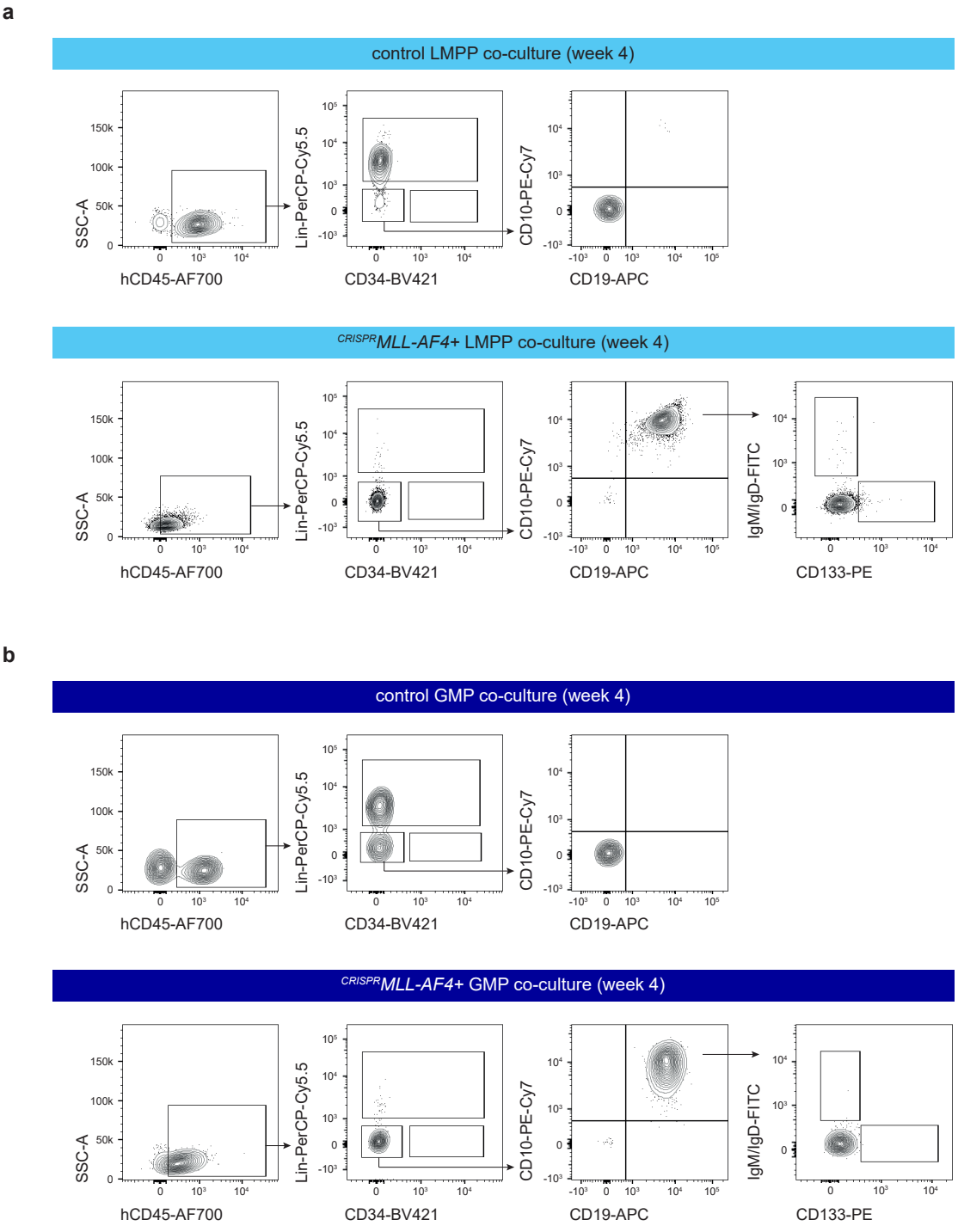

a
control LMPP co-culture (week 4)
SSC-A
hCD45-AF700
Lin-PerCP-Cy5.5
CD34-BV421
CD10-PE-Cy7
CD19-APC
CRISPR MLL-AF4+ LMPP co-culture (week 4)
IgM/IgD-FITC
CD133-PE
b
control GMP co-culture (week 4)
CRISPR MLL-AF4+ GMP co-culture (week 4)

Figure 5.9

Figure 5.9: Representative flow cytometry plots for FL LMPP and GMP co-cultures. a. (top) Representative flow cytometry plots for an unedited control FL LMPP well at week 4 of MS-5 stromal co-culture. Arrows represent sequential gating strategy. (bottom) Representative flow cytometry plots for an edited FL LMPP well at week 4 of MS-5 stromal co-culture. Arrows represent sequential gating strategy. **b.** (top) Representative flow cytometry plots for an unedited control FL GMP well at week 4 of MS-5 stromal co-culture. Arrows represent sequential gating strategy. (bottom) Representative flow cytometry plots for an edited FL GMP well at week 4 of MS-5 stromal co-culture. Arrows represent sequential gating strategy. Lin = CD2, CD3, CD14, CD16, CD56, CD235a.

Together, these data suggest that edited FL LMPP and GMP consistently give rise to a dramatic ^{CRISPR}*MLL-AF4+* B lineage expansion in vitro. MPP also appear to be capable of giving rise to this ^{CRISPR}*MLL-AF4+* expansion, although it is less frequent than in LMPP and GMP. HSC appear to be resistant, either to CRISPR/Cas9 editing or ^{CRISPR}*MLL-AF4*-driven B lineage expansion following induction of an *MLL-AF4*/t(4;11) translocation. RT-qPCR data suggest that the *MLL-AF4*/t(4;11) translocation may have occurred in edited FL PPB, but this was not capable of giving rise to the same ^{CRISPR}*MLL-AF4+* B lineage expansion in this MS-5 stromal co-culture system.

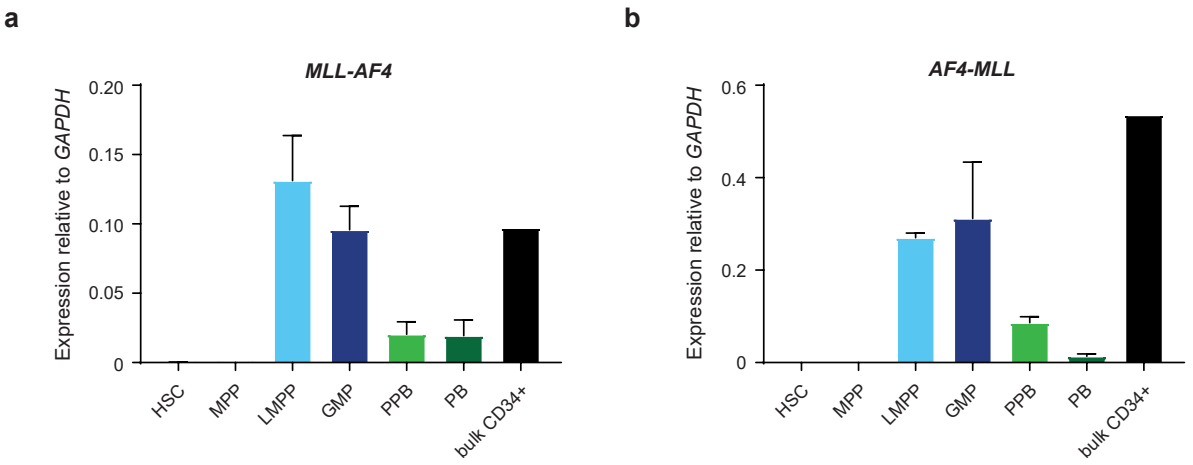


Figure 5.10: Expression of *MLL-AF4* and *AF4-MLL* in edited FL HSPC subpopulations *in vitro*. **a.** RT-qPCR showing expression of *MLL-AF4* in hCD45+ cells harvest from MS-5 stromal co-culture at week 4 relative to *GAPDH* (n=1-2). X-axis represents the FL HSPC subpopulations that were seeded into each well. Data shown as mean ± SEM. **b.** RT-qPCR showing expression of *AF4-MLL* in hCD45+ cells harvest from MS-5 stromal co-culture at week 4 relative to *GAPDH* (n=1-2). X-axis represents the FL HSPC subpopulations that were seeded into each well. Data shown as mean ± SEM.

5.6 ^{CRISPR}*MLL-AF4+* FL ELP and PPB give rise to ALL *in vivo*

I repeated the editing of bulk FL CD34+ cells (n=1, 12pcw) and sorting into their respective HSPC subpopulations (Figure 5.6). For each subpopulation, I then transplanted 2,000-3,000 edited, sorted cells plus 30,000 unedited FL MNC carrier cells into sublethally irradiated NSG mice. From week 9 post-transplant, human cell engraftment was detectable in HSC, MPP and ELP transplants but was particularly enriched in the ELP transplant (Figure 5.11a). Engraftment in the ELP transplant continued to rise steadily. In contrast, human engraftment in the PPB transplant rose rapidly between week 15 and week 21 post-transplant, overtaking ELP engraftment (Figure 5.11a). Throughout the experiment, no engraftment was observed in LMPP, GMP and PB transplants. RT-qPCR at weeks 18 and 21 post-transplant (data represent technical replicates for n=1 mouse each), showed low

level *MLL-AF4* expression in the peripheral blood of the ELP transplant and higher *MLL-AF4* expression in the peripheral blood of the PPB transplant (Figure 5.11b), perhaps reflecting the greater proportion of CD19+ cells in the PPB transplant. hCD45+ cell numbers recovered from the peripheral blood of HSC and MPP were not sufficient for reliable RT-qPCR results. In-keeping with the B lineage bias of NSG mice, most hCD45+ cells in all engrafted transplants were CD34-CD19+ B cells (Figure 5.11c). However, only in ELP and PPB transplants were CD19+CD133+ cells detected in the peripheral blood (Figure 5.11d), suggesting these may be B-ALL blasts. Therefore, based on these very preliminary data, it appears that FL ELP and PPB progenitors, both of which are human fetal-specific HSPC subpopulations, are capable of giving rise to ^{CRISPR}*MLL-AF4+* B-ALL *in vivo*.

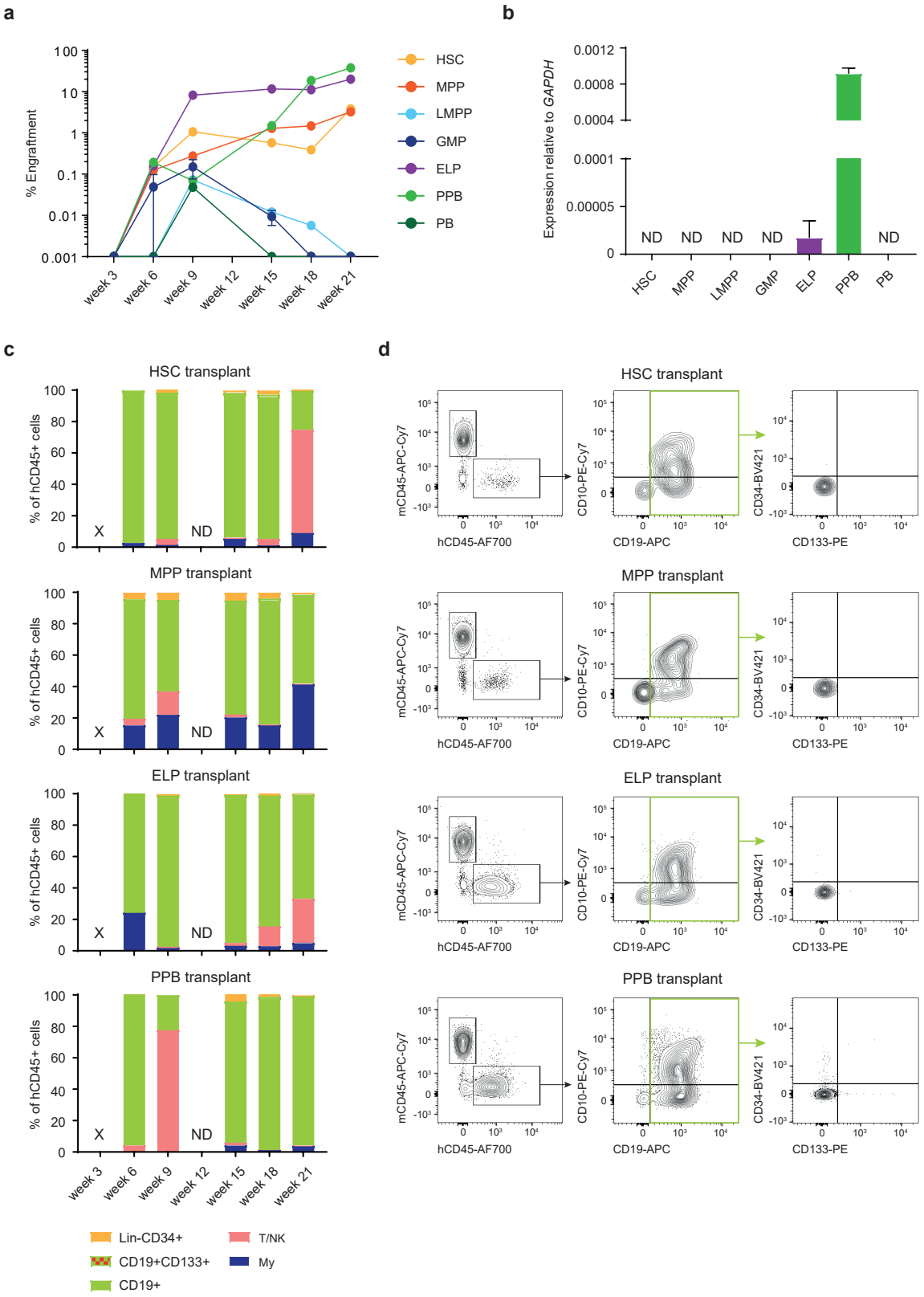

a
% Engraftment
100
10
1
0.1
0.01
0.001
week 3
week 6
week 9
week 12
week 15
week 18
week 21
HSC
MPP
LMPP
GMP
ELP
PPB
PB
b
Expression relative to GAPDH
0.0012
0.0008
0.0004
0.0001
0.00005
0
ND
HSC
MPP
LMPP
GMP
ELP
PPB
PB
c
HSC transplant
MPP transplant
ELP transplant
PPB transplant
% of hCD45+ cells
X
ND
Lin-CD34+
CD19+CD133+
CD19+
T/NK
My
d
mCD45-APC-Cy7
hCD45-AF700
CD10-PE-Cy7
CD19-APC
CD34-BV421
CD133-PE

Figure 5.11

Figure 5.11: ^{CRISPR}*MLL-AF4+* ELP and PPB give rise to leukaemic engraftment *in vivo*. **a.** Peripheral blood engraftment of hCD45+ cells over time in primary edited FL HSPC subpopulation transplants (n=1). **b.** RT-qPCR data showing expression of *MLL-AF4* in hCD45+ cells harvested from the peripheral blood of primary edited FL HSPC subpopulation transplants at week18/21 post-transplant. Data represent n=2 technical replicates from n=1 biological sample. Data are shown as mean ± SEM. ND = no data. **c.** Quantification of immunophenotypically-defined human cell types in the peripheral blood of primary edited FL HSPC subpopulation transplants over time as a percentage of hCD45+ cells (n=1). T/NK = CD3, CD56. My = CD11b, CD14, CD33. **d.** Representative flow cytometry plots showing hCD45+ peripheral blood engraftment in primary edited FL HSPC subpopulation transplants at week 21 post-transplant. Arrows represent sequential gating strategy. All CD19+ cells (light green box) were gated on and taken forward for analysis of CD133 and IgM/IgD expression.

5.7 Discussion

In this chapter, having previously shown that a CRISPR/Cas9-induced *MLL-AF4*/t(4;11) translocation in bulk primary human FL HSPCs was capable of giving rise to *MLL-AF4* infant ALL-like leukaemia, I set out to determine which FL HSPC subpopulation(s) represent the cellular origin of *MLL-AF4* infant ALL. The data included in this chapter are preliminary. Here, I will focus on discussing these preliminary data, and will discuss future directions for this project in more detail in 6.

By transfecting MLL-AF4 and a minimal MLL CXXC domain (data previously generated by Dr Jon Kerry) into an *MLL*wt cell line, I was able to show that MLL-AF4 binding in the earliest stages after its expression was largely directed toward accessible, unmethylated CpG regions in a manner that was almost identical to the MLL CXXC domain alone (Figure 5.2a). In addition to providing insight into how the cell-of-origin may influence the molecular profile of *MLL-AF4* infant ALL, these data may also shed some light on why not all MLL fusion protein target genes lose MLL fusion protein binding after treatment with a Menin inhibitor [Krivtsov et al., [2019]]. This Menin inhibitor (VTP50469) is one of the most recent novel therapies targeting the MLL fusion protein complex to enter clinical trials, following disappointing results from previous therapies such as DOT1L inhibitors [Stein et al., [2018]]. However, loss of MLL fusion protein binding was not observed at key MLL fusion protein target genes such a *HOXA* cluster genes and *MYB*

in preclinical data [Krivtsov et al., [2019]]. The data presented in this chapter suggest that the MLL CXXC domain retained in the MLL fusion protein binding to chromatin also plays an important role in MLL fusion protein recruitment to target genes, and at genes such as *HOXA9* and *MYB*, this CXXC domain binding may be sufficient to maintain fusion protein binding. These data also agree with a previous report from Milne et al., [2010] that the CXXC domain alone is sufficient to recruit MLL fusion proteins to *Hoxa9*.

Taken together, both the molecular and functional data in this chapter appear to suggest that there is a range of FL HSPC subpopulations that are capable of giving rise to a ^{CRISPR}*MLL-AF4+* B lineage expansion following an *MLL-AF4*/t(4;11) translocation event. *In vitro* data showed that 100% of edited FL LMPP and GMP wells gave rise to a ^{CRISPR}*MLL-AF4+* B lineage expansion in MS-5 stromal co-cultures, whereas only 2/5 edited MPP wells gave rise to a B lineage expansion (Figure 5.7). There is a possibility that these differences, rather than being due to the chromatin landscape of the cell in which the MLL-AF4 fusion protein is expressed, could be due to differences in CRISPR/Cas9 editing efficiency between different FL HSPC subpopulations and/or the ability of the edited cells to survive *in vitro*, and it will be important to confirm equivalent editing efficiency between all cell types in the future. However, taking these data at face value, they suggest that lymphoid-primed FL HSPCs are most permissive to ^{CRISPR}*MLL-AF4*-driven transformation. Previous data from the Roberts and Roy laboratory have shown that FL GMPs have B lineage potential *in vitro*, and the ATAC-seq data presented in this chapter further argues that the chromatin profile of FL GMPs shows a certain amount of lymphoid priming (Figure *5.4*a). These *in vitro* data, suggesting that an early progenitor with lymphoid potential may represent the cellular origin of *MLL-AF4* infant ALL, are in agreement with previous studies using mouse models that suggested the cell-of-origin of *MLL-AF4* infant ALL was a FL LMPP [Barrett et al., [2016], Malouf & Ottersbach, [2018]]. However, these data represent transformation events in an artificial MS-5 stromal co-culture assay.

In contrast with *in vitro* co-culture results, *in vivo* experiments, albeit from a single FL sample so far, suggested that the cells-of-origin of *MLL-AF4* infant ALL may be the downstream fetal-specific ELP and PPB progenitors (Figure 5.11a and 5.11b). While MS-5 stromal co-culture assays have provided a useful short-term readout of ^{CRISPR}*MLL-AF4*-driven transformation throughout this project, the more rigorous test of leukaemic potential is the ability to give rise to B-ALL *in vivo*. If future sorted *in vivo* experiments confirm the preliminary findings from this first replicate that FL ELP and PPB are the most likely progenitors capable of giving rise to a bona fide B-ALL, then these progenitors would represent a more likely candidate for the cellular origin of *MLL-AF4* infant ALL in patients, as a cell's ability to expand *in vitro* does not necessarily imply leukaemic potential.

In this first *in vivo* experiment, while ^{CRISPR}*MLL-AF4+* ELP engraftment was detected at an earlier time-point post-transplant, ^{CRISPR}*MLL-AF4+* PPB engraftment increased more rapidly and to more than double the level of engraftment observed in the ^{CRISPR}*MLL-AF4+* ELP transplant (Figure 5.11a). These preliminary data may be an early indication that there are phenotypic differences between ^{CRISPR}*MLL-AF4+* ALL derived from different FL HSPC subpopulations, or they may suggest that a secondary mutational hit is required for ^{CRISPR}*MLL-AF4+* PPB to give rise to B-ALL. In the future, it will be interesting to analyse not only the phenotypic differences between these ALLs but also the molecular differences underlying them.

In summary, FL HSPCs with lymphoid potential provide a permissive cellular context for ^{CRISPR}*MLL-AF4*-driven transformation, and FL ELP and PPB progenitors especially appear to be capable of giving rise to B-ALL *in vivo*, making them likely candidates for the cells-of-origin in *MLL-AF4* infant ALL patients. However, these data are preliminary and further study of CRISPR/Cas9 editing efficiency and *in vivo* leukaemic potential of different FL HSPC subpopulations are required to confirm these findings.

6 Discussion

Contents

6.1 Limitations 147
6.2 Future work 149

In this project, I have developed a method to induce an *MLL-AF4* /t(4;11) translocation in primary human FL HSPCs using CRISPR/Cas9 genome editing. Using this modelling approach, I have shown that an *MLL-AF4* /t(4;11) translocation arising in a human fetal cell context is sufficient to give rise to an aggressive leukaemia, ^{CRISPR}*MLL-AF4+* ALL, that recapitulates key clinico-pathological features of infant ALL. These data provide previously lacking functional evidence to support the hypothesis put forward by many in the field that the human fetal cell context in which *MLL*-rearranged infant ALL first arises plays a key role in driving the uniquely aggressive and treatment-refractory nature of this subset of leukaemia.

I next set out to explore the molecular differences underlying this unique phenotype of *MLL-AF4* infant ALL. Using primary patient RNA-seq data, I have shown that *MLL-AF4* infant ALL is transcriptomically distinct from *MLL-AF4* childhood ALL. In addition, I compared primary human FL HSPCs to adult BM HSPCs to derive one of the first human fetal gene expression signatures. In doing this, I was able to show that a substantial proportion of the genes (approximately 40%)

upregulated in *MLL-AF4* infant ALL compared to *MLL-AF4* childhood ALL could be explained by maintenance of human fetal-specific gene expression programmes. Finally, I have shown that these infant ALL-specific and fetal-specific gene expression programmes are also maintained in FL-derived ^{CRISPR}*MLL-AF4+* ALL, further supporting the hypothesis that it is the fetal cell context in which *MLL*-rearranged infant ALL first arises that plays a major role in driving its distinct molecular profile.

By introducing MLL-AF4 into an *MLL*wt cell line, I have shown that, in the earliest stages of leukaemogenesis, MLL-AF4 binds to unmethylated CpG regions that are accessible in the cell type in which MLL-AF4 is expressed, providing one possible mechanism by which the cell-of-origin can affect leukaemia initiation and the molecular profile of *MLL-AF4* infant ALL. Based on these data, I examined the chromatin profile of FL HSPC subpopulations using ATAC-seq, and found that the chromatin profile and the MLL-AF4 binding profile of ^{CRISPR}*MLL-AF4+* ALL most resembled the ATAC-seq profile of FL B progenitors. Finally, I began to assess leukaemic transformation in sorted subpopulations of FL HSPCs to identify the cellular origin of *MLL-AF4* infant ALL. I found that, *in vitro*, FL LMPP and GMP were most permissive to give rise to a ^{CRISPR}*MLL-AF4+* B lineage expansion, while, *in vivo*, ELP and PPB were permissive to give rise to ALL.

Taking all of these data together, I would propose a possible model whereby an *MLL-AF4*/t(4;11) translocation arising in either a B lineage-primed or an early B lineage-committed fetal progenitor cell leads to MLL-AF4 fusion protein binding at accessible, unmethylated CpG regions in that cell type, which, in human fetal B lineage-primed/committed progenitors, include fetal-specific genes such as *LIN28B, IGF2BP1* and *HOXB4*, stem cell-associated genes like *LMO2* and *NOG*, and myeloid-associated genes like *CSF1R* and *MPO* as well as many B lineage-associated genes. By aberrantly upregulating these target genes, MLL-AF4 is sufficient to give rise to an aggressive infant B-ALL that is distinct from B-ALL in older children and capable of switching to the myeloid lineage on relapse.

6.1 Limitations

There are a number of questions I was unable to explore in this project, and these limitations are important to acknowledge when interpreting the results presented here.

Firstly, I have not quantified the efficiency at which this CRISPR/Cas9-based method induces the *MLL-AF4* /t(4;11) translocation. Based on *in vitro* data, I have estimated that the efficiency of ^{CRISPR}*MLL-AF4+* B lineage expansion is at least 1:4,000 edited FL CD34+ cells. However, this likely does not accurately reflect the total number of cells in which a translocation event has occurred. Even more importantly, I was unable to examine whether this CRISPR/Cas9 editing was equally efficient in all FL HSPC subpopulations. Therefore, it is as yet unclear from the data presented in Chapter 5 whether HSC fail to give rise to ^{CRISPR}*MLL-AF4+* ALL because the molecular profile of this HSPC subpopulation is not permissive for MLL-AF4 to initiate leukaemogenesis, or simply because HSCs show greater resistance to, or toxicity from, CRISPR/Cas9 editing. Going forward, it will be essential to optimise a sensitive method to quantify the efficiency of CRISPR/Cas9-induced translocation events in different FL HSPC subpopulations. One possible method by which this could be achieved is droplet digital PCR, which allows absolute quantification of the number of *MLL-AF4* fusion genes present in a population of edited cells with high sensitivity.

Secondly, I was unable to investigate the role of the reciprocal fusion gene, *AF4-MLL*, in the initiation of *MLL-AF4* infant ALL. All CRISPR/Cas9 editing experiments during this project led to expression of both reciprocal fusion genes. In the future, it would be possible to knock out *AF4-MLL* in ^{CRISPR}*MLL-AF4+* blasts; however, this would not allow investigation of its role in leukaemia initiation. This is important because *AF4-MLL* has been hypothesised to be required for leukaemia initiation but dispensable for leukaemia maintenance [Marschalek, [2020], Wilhelm & Marschalek, [2021]]. One possible adjustment to the CRISPR/Cas9 editing approach presented in this thesis would be to redesign the sgRNAs to target exonic

sequences of *MLL* and *AF4*, and use homology-directed repair (HDR) to introduce a premature STOP codon in the exonic breakpoint of either *MLL-AF4* or *AF4-MLL*.

When validating this CRISPR/Cas9-based approach to induce the *MLL-AF4* /t(4;11) translocation, I was able to show that no major off-target editing had occurred; ^{CRISPR}*MLL-AF4+* blasts retained the same karyotype of the original FL sample, no other *MLL*-rearrangements occurred and *in silico*-predicted off-target cut sites did not show any indels. These data, along with the evidence that ^{CRISPR}*MLL-AF4*-driven transformation was efficient (insofar as all editing experiments gave rise to a leukaemia), suggest that the *MLL-AF4* /t(4;11) translocation alone is sufficient to give rise to *MLL-AF4* infant ALL. However, what I did not explore in detail in this project is the presence of any cooperating mutations and their possible role in leukaemogenesis and clonal evolution through serial ^{CRISPR}*MLL-AF4+* ALL transplants. While they are often sub-clonal and can be lost at relapse, we know that Ras pathway mutations can occur in *MLL*-rearranged infant ALL. Going forward, it will be important to screen all ^{CRISPR}*MLL-AF4+* ALLs for cooperating mutations. These data may be particularly useful when interpreting differences in disease latency and phenotype between different ^{CRISPR}*MLL-AF4+* ALL replicates.

Finally, I have used the MS-5 stromal co-culture system extensively in this project to assess ^{CRISPR}*MLL-AF4*-driven transformation. However, ^{CRISPR}*MLL-AF4+* blasts could not be maintained past week 6 in culture, suggesting this system is lacking key factors to support ALL blasts. Further work is required to identify factors that are missing from this *in vitro* assay, which may include other BM cell types and cytokines, and to optimise this system for ALL blasts. Furthermore, there was a discrepancy between the FL HSPC subpopulations that give rise to a ^{CRISPR}*MLL-AF4+* B lineage expansion *in vitro* and those that give rise to ALL *in vivo* in Chapter 5. Therefore, I hypothesise that aberrant expansion in this MS-5 stromal co-culture system may not accurately reflect the ability of MLL-AF4 to give rise to leukaemia in a particular HSPC subpopulation. Going forward in the study of the cellular origin of *MLL-AF4* infant ALL, *in vivo* assays of leukaemic potential should be prioritised over this *in vitro* assay.

6.2 Future work

Moving forward from this project, there are a number of avenues that I would propose exploring in further detail.

Firstly, it will be important to continue to investigate the cellular origin of *MLL-AF4* infant ALL using FL HSPCs. The data presented in this thesis are preliminary but they already suggest that a range of fetal HSPCs are permissive to give rise to infant ALL following an *MLL-AF4*/t(4;11) translocation event. This may provide an explanation for the different molecular profiles observed between *MLL-AF4* infant ALL patients, and molecular signatures arising from a particular cell-of-origin may have prognostic significance.

Secondly, while these data suggest that a CRISPR/Cas9-induced *MLL-AF4*/t(4;11) translocation in FL HSPCs gives rise to an infant ALL-like disease, it is unclear whether the same editing approach would give rise to a leukaemia in FBM, CB or adult BM HSPCs. Furthermore, it is unclear what the molecular and phenotypic characteristics of any resulting ALL would be. I would hypothesise that as fetal-specific gene expression programmes are shut down, which appears to already be occurring in late gestation FBM and CB [Roy et al., [2021]], the leukaemias derived from these HSPCs will lose infant ALL-specific characteristics and show a greater resemblance to *MLL*-rearranged childhood ALL. In addition, as these fetal-specific gene expression programmes may cooperate with the MLL-AF4 fusion protein to drive leukaemogenesis after this single mutational hit, it may be that ^{CRISPR}*MLL-AF4*-driven leukaemogenesis is less efficient in post-natal HSPCs.

Finally, I have identified overlapping fetal-specific and infant ALL-specific gene expression programmes in *MLL-AF4* infant ALL during this project, and I hypothesise that these gene expression signatures underlie the uniquely aggressive nature of *MLL-AF4* infant ALL compared to *MLL*-rearranged ALLs in older children. One striking example has been the clear representation of the fetal-specific *LIN28B* pathway in ^{CRISPR}*MLL-AF4+* ALL, with MLL-AF4 binding at, and aberrant expression of, genes like *IGF2BP1, HMGA2*, and *HOXB4*. In the future, it will be important to knock down these fetal genes in ^{CRISPR}*MLL-AF4+* ALL to test

whether this has a detrimental effect on leukaemia maintenance. Moreover, these fetal pathways could be reactivated in human post-natal HSPCs, for example by overexpression of *LIN28B*, before ^{CRISPR}*MLL-AF4*-driven leukaemogenesis to investigate whether this would lead to a more infant ALL-like leukaemia phenotype. In this way, we may identify novel dependencies specific to *MLL*-rearranged infant ALL that could be targeted to more effectively treat this poor prognosis disease; and ^{CRISPR}*MLL-AF4+* ALL represents a previously lacking *MLL-AF4* infant ALL preclinical model in which novel treatments could be tested.

Bibliography

Agraz-Doblas, A., Bueno, C., Bashford-Rogers, R., Roy, A., Schneider, P., Bardini, M., Ballerini, P., Cazzaniga, G., Moreno, T., Revilla, C., Gut, M., Valsecchi, M.G., Roberts, I., Pieters, R., De Lorenzo, P., Varela, I., Menendez, P. & Stam, R.W. (2019) "Unravelling the cellular origin and clinical prognostic markers of infant B-cell acute lymphoblastic leukemia using genome-wide analysis." *Haematologica*.

Ali, M., Hom, R.A., Blakeslee, W., Ikenouye, L. & Kutateladze, T.G. (2014) "Diverse functions of PHD fingers of the MLL/KMT2 subfamily." *Biochim Biophys Acta*, 1843(2): 366-371.

Andersson, A.K., Ma, J., Wang, J., Chen, X., Gedman, A.L., Dang, J., Nakitandwe, J., Holmfeldt, L., Parker, M., Easton, J., Huether, R., Kriwacki, R., Rusch, M., Wu, G., Li, Y., Mulder, H., Raimondi, S., Pounds, S., Kang, G., Shi, L., Becksfort, J., Gupta, P., Payne-Turner, D., Vadodaria, B., Boggs, K., Yergeau, D., Manne, J., Song, G., Edmonson, M., Nagahawatte, P., Wei, L., Cheng, C., Pei, D., Sutton, R., Venn, N.C., Chetcuti, A., Rush, A., Catchpoole, D., Heldrup, J., Fioretos, T., Lu, C., Ding, L., Pui, C.H., Shurtleff, S., Mullighan, C.G., Mardis, E.R., Wilson, R.K., Gruber, T.A., Zhang, J., Downing, J.R. & St. Jude Children's Research Hospital-Washington University Pediatric Cancer Genome Project (2015) "The landscape of somatic mutations in infant *MLL*-rearranged acute lymphoblastic leukemias." *Nat Genet*, 47(4): 330-337.

Armstrong, S.A., Kung, A.L., Mabon, M.E., Silverman, L.B., Stam, R.W., Den Boer, M.L., Pieters, R., Kersey, J.H., Sallan, S.E., Fletcher, J.A., Golub, T.R., Griffin, J.D. & Korsmeyer, S.J. (2003) "Inhibition of FLT3 in *MLL*. Validation

of a therapeutic target identified by gene expression based classification." *Cancer Cell*, 3(2): 173-183.

Ballabio, E. & Milne, T.A. (2012) "Molecular and Epigenetic Mechanisms of MLL in Human Leukemogenesis." *Cancers (Basel)*, 4(3): 904-944.

Balzeau, J., Menezes, M.R., Cao, S. & Hagan, J.P. (2017) "The *LIN28/let-7* Pathway in Cancer." *Front Genet*, 8: 31.

Barabe, F., Kennedy, J.A., Hope, K.J. & Dick, J.E. (2007) "Modeling the initiation and progression of human acute leukemia in mice." *Science*, 316(5824): 600-604.

Bardini, M., Spinelli, R., Bungaro, S., Mangano, E., Corral, L., Cifola, I., Fazio, G., Giordan, M., Basso, G., De Rossi, G., Biondi, A., Battaglia, C. & Cazzaniga, G. (2010) "DNA copy-number abnormalities do not occur in infant ALL with t(4;11)/*MLL-AF4*." *Leukemia*, 24(1): 169-176.

Bardini, M., Galbiati, M., Lettieri, A., Bungaro, S., Gorletta, T.A., Biondi, A. & Cazzaniga, G. (2011) "Implementation of array based whole-genome high-resolution technologies confirms the absence of secondary copy-number alterations in *MLL-AF4*-positive infant ALL patients." *Leukemia*, 25(1): 175-178.

Bardini, M., Woll, P.S., Corral, L., Luc, S., Wittmann, L., Ma, Z., Lo Nigro, L., Basso, G., Biondi, A., Cazzaniga, G. & Jacobsen, S.E. (2015) "Clonal variegation and dynamic competition of leukemia-initiating cells in infant acute lymphoblastic leukemia with *MLL*-rearrangement." *Leukemia*, 29(1): 38-50.

Barrett, N.A., Malouf, C., Kapeni, C., Bacon, W.A., Giotopoulos, G., Jacobsen, S.E.W., Huntly, B.J. & Ottersbach, K. (2016) "*Mll-AF4* Confers Enhanced Self-Renewal and Lymphoid Potential during a Restricted Window in Development." *Cell Rep*, 16(4): 1039-1054.

Basso, G., Putti, M.C., Cantu-Rajnoldi, A., Saitta, M., Santostasi, T., Santoro, N., Lippi, A., Comelli, A., Felici, L., Favre, C., Russo Mancuso, G., Guglielmi, C., Paolucci, P., Biondi, A., Rondelli, R. & Pession, A. (1992) "The immunophenotype

in infant acute lymphoblastic leukaemia: correlation with clinical outcome. An Italian multicentre study (AIEOP)." *Br J Haematol*, 81(2): 184-191.

Benedikt, A., Baltruschat, S., Scholz, B., Bursen, A., Arrey, T.N., Meyer, B., Varagnolo, L., Muller, A.M., Karas, M., Dingermann, T. & Marschalek, R. (2011) "The leukemogenic AF4-MLL fusion protein causes P-TEFb kinase activation and altered epigenetic signatures." *Leukemia*, 25(1): 135-144.

Benito, J.M., Godfrey, L., Kojima, K., Hogdal, L., Wunderlich, M., Geng, H., Marzo, I., Harutyunyan, K.G., Golfman, L., North, P., Kerry, J., Ballabio, E., Chonghaile, T.N., Gonzalo, O., Qiu, Y., Jeremias, I., Debose, L., O'Brien, E., Ma, H., Zhou, P., Jacamo, R., Park, E., Coombes, K.R., Zhang, N., Thomas, D.A., O'Brien, S., Kantarjian, H.M., Leverson, J.D., Kornblau, S.M., Andreeff, M., Muschen, M., Zweidler-McKay, P.A., Mulloy, J.C., Letai, A., Milne, T.A. & Konopleva, M. (2015) "*MLL*-Rearranged Acute Lymphoblastic Leukemias Activate *BCL-2* through H3K79 Methylation and Are Sensitive to the BCL-2-Specific Antagonist ABT-199." *Cell Rep*, 13(12): 2715-2727.

Bernt, K.M., Zhu, N., Sinha, A.U., Vempati, S., Faber, J., Krivtsov, A.V., Feng, Z., Punt, N., Daigle, A., Bullinger, L., Pollock, R.M., Richon, V.M., Kung, A.L. & Armstrong, S.A. (2011) "*MLL*-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L." *Cancer Cell*, 20(1): 66-78.

Biondi, A., Cimino, G., Pieters, R. & Pui, C.H. (2000) "Biological and therapeutic aspects of infant leukemia." *Blood*, 96(1): 24-33.

Birke, M., Schreiner, S., Garcia-Cuellar, M.P., Mahr, K., Titgemeyer, F. & Slany, R.K. (2002) "The MT domain of the proto-oncoprotein MLL binds to CpG-containing DNA and discriminates against methylation." *Nucleic Acids Res*, 30(4): 958-965.

Biswas, D., Milne, T.A., Basrur, V., Kim, J., Elenitoba-Johnson, K.S., Allis, C.D. & Roeder, R.G. (2011) "Function of leukemogenic mixed lineage leukemia 1 (MLL)

fusion proteins through distinct partner protein complexes." *Proc Natl Acad Sci USA*, 108(38): 15751-15756.

Borkhardt, A., Wuchter, C., Viehmann, S., Pils, S., Teigler-Schlegel, A., Stanulla, M., Zimmermann, M., Ludwig, W.D., Janka-Schaub, G., Schrappe, M. & Harbott, J. (2002) "Infant acute lymphoblastic leukemia - combined cytogenetic, immunophenotypical and molecular analysis of 77 cases." *Leukemia*, 16(9): 1685-1690.

Boutter, J., Huang, Y., Marovca, B., Vonderheit, A., Grotzer, M.A., Eckert, C., Cario, G., Wollscheid, B., Horvath, P., Bornhauser, B.C. & Bourquin, J.P. (2014) "Image-based RNA interference screening reveals an individual dependence of acute lymphoblastic leukemia on stromal cysteine support." *Oncotarget*, 5(22): 11501-11512.

Bowie, M.B., McKnight, K.D., Kent, D.G., McCaffrey, L., Hoodless, P.A. & Eaves, C.J. (2006) "Hematopoietic stem cells proliferate until after birth and show a reversible phase-specific engraftment defect." *J Clin Invest*, 116(10): 2808-2816.

Brochet, X., Lefranc, M.P. & Giudicelli, V. (2008) "IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis." *Nucleic Acids Res*, 36: 503-508.

Brown, P., Levis, M., Shurtleff, S., Campana, D., Downing, J. & Small, D. (2005) "FLT3 inhibition selectively kills childhood acute lymphoblastic leukemia cells with high levels of *FLT3* expression." *Blood*, 105(2): 812-820.

Brown, P. (2013) "Treatment of infant leukemias: challenge and promise." *Hematology Am Soc*, Hematol Educ Program 2013: 596-600.

Brown, P.A., Kairalla, J.A., Hilden, J.M., Dreyer, Z.E., Carroll, A.J., Heerema, N.A., Wang, C., Devidas, M., Gore, L., Salzer, W.L., Winick, N.J., Carroll, W.L., Raetz, E.A., Borowitz, M.J., Small, D., Loh, M.L. & Hunger, S.P. (2021). "FLT3 inhibitor lestaurtinib plus chemotherapy for newly diagnosed *KMT2A*-rearranged

infant acute lymphoblastic leukemia: Children's Oncology Group trial AALL0631." *Leukemia*, 35(5): 1279-1290.

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. (2013) "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position." *Nat Methods*, 10(12): 1213-1218.

Bursen, A., Schwabe, K., Ruster, B., Henschler, R., Ruthardt, M., Dingermann, T. & Marschalek, R. (2010) "The AF4.MLL fusion protein is capable of inducing ALL in mice without requirement of MLL.AF4." *Blood*, 115(17): 3570-3579.

Carrelha, J., Meng, Y., Kettyle, L.M., Luis, T.C., Norfo, R., Alcolea, V., Boukarabila, H., Grasso, F., Gambardella, A., Grover, A., Hogstrand, K., Lord, A.M., Sanjuan-Pla, A., Woll, P.S., Nerlov, C. & Jacobsen, S.E.W. (2018) "Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells." *Nature*, 554(7690): 106-111.

Charbord, P., Tavian, M., Humeau, L. & Peault, B. (1996) "Early ontogeny of the human marrow from long bones: an immunohistochemical study of hematopoiesis and its microenvironment." *Blood*, 87(10): 4109-4119.

Chen, W., Kumar, A.R., Hudson, W.A., Li, Q., Wu, B., Staggs, R.A., Lund, E.A., Sam, T.N. & Kersey, J.H. (2008) "Malignant transformation initiated by *Mll-AF9*: gene dosage and critical target cells." *Cancer Cell*, 13(5): 432-440.

Chen, W., O'Sullivan, M.G., Hudson, W. & Kersey, J. (2011) "Modeling human infant *MLL* leukemia in mice: leukemia from fetal liver differs from that originating in postnatal marrow." *Blood*, 117(12): 3474-3475.

Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S. & Kim, J.S. (2014) "Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases." *Genome Res*, 24(1): 132-141.

Choi, P.S. & Meyerson, M. (2014) "Targeted genomic rearrangements using CRISPR/Cas technology." *Nat Commun*, 5: 3728.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. & Zhang, F. (2013) "Multiplex genome engineering using CRISPR/Cas systems." *Science*, 339(6121): 819-823.

Copley, M.R., Babovic, S., Benz, C., Knapp, D.J., Beer, P.A., Kent, D.G., Wohrer, S., Treloar, D.Q., Day, C., Rowe, K., Mader, H., Kuchenbauer, F., Humphries, R.K. & Eaves, C.J. (2013) "The *Lin28b-let-7-Hmga2* axis determines the higher self-renewal potential of fetal haematopoietic stem cells." *Nat Cell Biol*, 15(8): 916-925.

Copley, M.R. & Eaves, C. J. (2013) "Developmental changes in hematopoietic stem cell properties." *Exp Mol Med*, 45: 55.

Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, P.G., Chan, S.M., Koenig, J.L., Snyder, M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J., Majeti, R. & Chang, H.Y. (2016) "Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution." *Nat Genet*, 48(10): 1193-1203.

Corral, J., Lavenir, I., Impey, H., Warren, A.J., Forster, A., Larson, T.A., Bell, S., McKenzie, A.N., King, G. & Rabbitts, T.H. (1996) "An *Mll-AF9* fusion gene made by homologous recombination causes acute leukemia in chimeric mice: a method to create fusion oncogenes." *Cell*, 85(6): 853-861.

Cozzio, A., Passegue, E., Ayton, P.M., Karsunky, H., Cleary, M.L. & Weissman, I.L. (2003) "Similar *MLL*-associated leukemias arising from self-renewing stem cells and short-lived myeloid progenitors." *Genes Dev*, 17(24): 3029-3035.

Creutzig, U., Zimmermann, M., Bourquin, J.P., Dworzak, M.N., Kremens, B., Lehrnbecher, T., von Neuhoff, C., Sander, A., von Stackelberg, A., Schmid, I., Stary, J., Steinbach, D., Vormoor, J. & Reinhardt, D. (2012) "Favorable outcome

in infants with AML after intensive first- and second-line treatment: an AML-BFM study group report." *Leukemia*, 26(4): 654-661.

Dander, E., Palmi, C., D'Amico, G. & Cazzaniga, G. (2021) "The Bone Marrow Niche in B-Cell Acute Lymphoblastic Leukemia: The Role of Microenvironment from Pre-Leukemia to Overt Leukemia." *Int J Mol Sci*, 22(9): 4426.

de Bruijn, M.F., Ma, X., Robin, C., Ottersbach, K., Sanchez, M.J. & Dzierzak, E. (2002) "Hematopoietic stem cells localize to the endothelial cell layer in the midgestation mouse aorta." *Immunity*, 16(5): 673-683.

Djabali, M., Selleri, L., Parry, P., Bower, M., Young, B.D. & Evans, G.A. (1992) "A *trithorax*-like gene is interrupted by chromosome 11q23 translocations in acute leukaemias." *Nat Genet*, 2(2): 113-118.

Dobbins, S.E., Sherborne, A.L., Ma, Y.P., Bardini, M., Biondi, A., Cazzaniga, G., Lloyd, A., Chubb, D., Greaves, M.F. & Houlston, R.S. (2013) "The silent mutational landscape of infant *MLL-AF4* pro-B acute lymphoblastic leukemia." *Genes Chromosomes Cancer*, 52(10): 954-960.

Dou, Y., Milne, T.A., Tackett, A.J., Smith, E.R., Fukuda, A., Wysocka, J., Allis, C.D., Chait, B.T., Hess, J.L. & Roeder, R.G. (2005) "Physical association and coordinate function of the H3 K4 methyltransferase MLL1 and the H4 K16 acetyltransferase MOF." *Cell*, 121(6): 873-885.

Downing, J.R., Head, D.R., Raimondi, S.C., Carroll, A.J., Curcio-Brint, A.M., Motroni, T.A., Hulshof, M.G., Pullen, D.J. & Domer, P.H. (1994) "The der(11)-encoded *MLL/AF-4* fusion transcript is consistently detected in t(4;11)(q21;q23)-containing acute lymphoblastic leukemia." *Blood*, 83(2): 330-335.

Dreyer, Z.E., Hilden, J.M., Jones, T.L., Devidas, M., Winick, N.J., Willman, C.L., Harvey, R.C., Chen, I.M., Behm, F.G., Pullen, J., Wood, B.L., Carroll, A.J., Heerema, N.A., Felix, C.A., Robinson, B., Reaman, G.H., Salzer, W.L., Hunger, S.P., Carroll, W.L. & Camitta, B.M. (2015) "Intensified chemotherapy without

SCT in infant ALL: results from COG P9407 (Cohort 3)." *Pediatr Blood Cancer*, 62(3): 419-426.

Drynan, L.F., Pannell, R., Forster, A., Chan, N.M., Cano, F., Daser, A. & Rabbitts, T.H. (2005) "*Mll* fusions generated by *Cre-loxP*-mediated de novo translocations can induce lineage reassignment in tumorigenesis." *EMBO J*, 24(17): 3136-3146.

Dzierzak, E. & Bigas, A. (2018) "Blood Development: Hematopoietic Stem Cell Dependence and Independence." *Cell Stem Cell*, 22(5): 639-651.

Eaves, C.J. (2015) "Hematopoietic stem cells: concepts, definitions, and the new reality." *Blood*, 125(17): 2605-2613.

Ernst, P., Wang, J., Huang, M., Goodman, R.H. & Korsmeyer, S.J. (2001) "MLL and CREB bind cooperatively to the nuclear coactivator CREB-binding protein." *Mol Cell Biol*, 21(7): 2249-2258.

Erfurth, F., Hemenway, C.S., de Erkenez, A.C. & Domer, P.H. (2004) "MLL fusion partners AF4 and AF9 interact at subnuclear foci." *Leukemia*, 18(1): 92-102.

Feng, Y., Yang, Y., Ortega, M.M., Copeland, J.N., Zhang, M., Jacob, J.B., Fields, T.A., Vivian, J.L. & Fields, P.E. (2010) "Early mammalian erythropoiesis requires the Dot1L methyltransferase." *Blood*, 116(22): 4483-4491.

Ford, A.M., Ridge, S.A., Cabrera, M.E., Mahmoud, H., Steel, C.M., Chan, L.C. & Greaves, M. (1993) "*In utero* rearrangements in the *trithorax*-related oncogene in infant leukaemias." *Nature*, 363(6427): 358-360.

Frederiks, F., Tzouros, M., Oudgenoeg, G., van Welsem, T., Fornerod, M., Krijgsveld, J. & van Leeuwen, F. (2008) "Nonprocessive methylation by Dot1 leads to functional redundancy of histone H3K79 methylation states." *Nat Struct Mol Biol*, 15(6): 550-557.

Gale, K.B., Ford, A.M., Repp, R., Borkhardt, A., Keller, C., Eden, O.B. & Greaves, M.F. (1997) "Backtracking leukemia to birth: identification of clonotypic gene fusion sequences in neonatal blood spots." *Proc Natl Acad Sci USA*, 94(25): 13950-13954.

Gardner, R., Wu, D., Cherian, S., Fang, M., Hanafi, L.A., Finney, O., Smithers, H., Jensen, M.C., Riddell, S.R., Maloney, D.G. & Turtle, C.J. (2016) "Acquisition of a CD19-negative myeloid phenotype allows immune escape of *MLL*-rearranged B-ALL from CD19 CAR-T-cell therapy." *Blood*, 127(20): 2406-2410.

Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. (2012) "Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria." *Proc Natl Acad Sci USA*, 109(39): 2579-2586.

Gaussmann, A., Wenger, T., Eberle, I., Bursen, A., Bracharz, S., Herr, I., Dingermann, T. & Marschalek, R. (2007) "Combined effects of the two reciprocal t(4;11) fusion proteins MLL.AF4 and AF4.MLL confer resistance to apoptosis, cell cycling capacity and growth transformation." *Oncogene*, 26(23): 3352-3363.

Goardon, N., Marchi, E., Atzberger, A., Quek, L., Schuh, A., Soneji, S., Woll, P., Mead, A., Alford, K.A., Rout, R., Chaudhury, S., Gilkes, A., Knapper, S., Beldjord, K., Begum, S., Rose, S., Geddes, N., Griffiths, M., Standen, G., Sternberg, A., Cavenagh, J., Hunter, H., Bowen, D., Killick, S., Robinson, L., Price, A., Macintyre, E., Virgo, P., Burnett, A., Craddock, C., Enver, T., Jacobsen, S.E., Porcher, C. & Vyas, P. (2011) "Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia." *Cancer Cell*, 19(1): 138-152.

Godfrey, L., Kerry, J., Thorne, R., Repapi, E., Davies, J.O., Tapia, M., Ballabio, E., Hughes, J.R., Geng, H., Konopleva, M. & Milne, T.A. (2017) "MLL-AF4 binds directly to a *BCL-2* specific enhancer and modulates H3K27 acetylation." *Exp Hematol*, 47: 64-75.

Godfrey, L., Crump, N.T., Thorne, R., Lau, I.J., Repapi, E., Dimou, D., Smith, A.L., Harman, J.R., Telenius, J.M., Oudelaar, A.M., Downes, D.J., Vyas, P., Hughes, J.R. & Milne, T.A. (2019) "DOT1L inhibition reveals a distinct subset of enhancers dependent on H3K79 methylation." *Nat Commun*, 10(1): 2803.

Godfrey, L., Crump, N.T., O'Byrne, S., Lau, I.J., Rice, S., Harman, J.R., Jackson, T., Elliott, N., Buck, G., Connor, C., Thorne, R., Knapp, Djhf, Heidenreich, O., Vyas, P., Menendez, P., Inglott, S., Ancliff, P., Geng, H., Roberts, I., Roy, A. & Milne, T.A. (2020) "H3K79me2/3 controls enhancer-promoter interactions and activation of the pan-cancer stem cell marker *PROM1*/CD133 in *MLL-AF4* leukemia cells." *Leukemia*.

Greaves, M.F., Maia, A.T., Wiemels, J.L. & Ford, A.M. (2003) "Leukemia in twins: lessons in natural history." *Blood*, 102(7): 2321-2333.

Greaves, M. (2005) "*In utero* origins of childhood leukaemia." *Early Hum Dev*, 81(1): 123-129.

Greaves, M. (2018) "A causal mechanism for childhood acute lymphoblastic leukaemia." *Nat Rev Cancer*, 18(8): 471-484.

Guenther, M.G., Lawton, L.N., Rozovskaia, T., Frampton, G.M., Levine, S.S., Volkert, T.L., Croce, C.M., Nakamura, T., Canaani, E. & Young, R.A. (2008) "Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia." *Genes Dev*, 22(24): 3403-3408.

Gundry, M.C., Brunetti, L., Lin, A., Mayle, A.E., Kitano, A., Wagner, D., Hsu, J.I., Hoegenauer, K.A., Rooney, C.M., Goodell, M.A. & Nakada, D. (2016) "Highly Efficient Genome Editing of Murine and Human Hematopoietic Progenitor Cells by CRISPR/Cas9." *Cell Rep*, 17(5): 1453-1461.

Hardy, R.R., Carmack, C.E., Shinton, S.A., Kemp, J.D. & Hayakawa, K. (1991) "Resolution and characterization of pro-B and pre-pro-B cell stages in normal mouse bone marrow." *J Exp Med*, 173(5): 1213-1225.

Harman, J.R., Thorne, R., Jamilly, M., Tapia, M., Crump, N.T., Rice, S., Beveridge, R., Morrissey, E., de Bruijn, Mftr, Roberts, I., Roy, A., Fulga, T.A. & Milne, T.A. (2021) "A *KMT2A-AFF1* gene regulatory network highlights the role of core transcription factors and reveals the regulatory logic of key downstream target genes." *Genome Res.*

Harrison, C.J., Hills, R.K., Moorman, A.V., Grimwade, D.J., Hann, I., Webb, D.K., Wheatley, K., de Graaf, S.S., van den Berg, E., Burnett, A.K. & Gibson, B.E. (2010) "Cytogenetics of childhood acute myeloid leukemia: United Kingdom Medical Research Council Treatment trials AML 10 and 12." *J Clin Oncol*, 28(16): 2674-2681.

He, N., Liu, M., Hsu, J., Xue, Y., Chou, S., Burlingame, A., Krogan, N.J., Alber, T. & Zhou, Q. (2010). "HIV-1 Tat and host AFF4 recruit two transcription elongation factors into a bifunctional complex for coordinated activation of *HIV-1* transcription." *Mol Cell*, 38(3): 428-438.

Helsmoortel, H.H., Bresolin, S., Lammens, T., Cave, H., Noellke, P., Caye, A., Ghazavi, F., de Vries, A., Hasle, H., Labarque, V., Masetti, R., Stary, J., van den Heuvel-Eibrink, M.M., Philippe, J., Van Roy, N., Benoit, Y., Speleman, F., Niemeyer, C., Flotho, C., Basso, G., Te Kronnie, G., Van Vlierberghe, P. & De Moerloose, B. (2016) "*LIN28B* overexpression defines a novel fetal-like subgroup of juvenile myelomonocytic leukemia." *Blood*, 127(9): 1163-1172.

Hern, W.M. (1984) "Correlation of fetal age and measurements between 10 and 26 weeks of gestation." *Obstet Gynecol*, 63(1): 26-32.

Hilden, J.M., Dinndorf, P.A., Meerbaum, S.O., Sather, H., Villaluna, D., Heerema, N.A., McGlennen, R., Smith, F.O., Woods, W.G., Salzer, W.L., Johnstone, H.S., Dreyer, Z., Reaman, G.H. & Children's Oncology Group (2006) "Analysis of prognostic factors of acute lymphoblastic leukemia in infants: report on CCG 1953 from the Children's Oncology Group." *Blood*, 108(2): 441-451.

Hofer, T. & Rodewald, H.R. (2018) "Differentiation-based model of hematopoietic stem cell functions and lineage pathways." *Blood*, 132(11): 1106-1113.

Holyoake, T.L., Nicolini, F.E. & Eaves, C.J. (1999) "Functional differences between transplantable human hematopoietic stem cells from fetal liver, cord blood, and adult marrow." *Exp Hematol*, 27(9): 1418-1427.

Hom, R.A., Chang, P.Y., Roy, S., Musselman, C.A., Glass, K.C., Selezneva, A.I., Gozani, O., Ismagilov, R.F., Cleary, M.L. & Kutateladze, T.G. (2014) "Molecular mechanism of MLL PHD3 and RNA recognition by the Cyp33 RRM domain." *J Mol Biol*, 400(2): 145-154.

Horton, S.J., Jaques, J., Woolthuis, C., van Dijk, J., Mesuraca, M., Huls, G., Morrone, G., Vellenga, E. & Schuringa, J.J. (2013) "*MLL-AF9*-mediated immortalization of human hematopoietic cells along different lineages changes during ontogeny." *Leukemia*, 27(5): 1116-1126.

Hunger, S.P., McGavran, L., Meltesen, L., Parker, N.B., Kassenbrock, C.K. & Bitter, M.A. (1998) "Oncogenesis *in utero*: fetal death due to acute myelogenous leukaemia with an *MLL* translocation." *Br J Haematol*, 103(2): 539-542.

Iacobucci, I. & Mullighan, C.G. (2017) "Genetic Basis of Acute Lymphoblastic Leukemia." *J Clin Oncol*, 35(9): 975-983.

Ichii, M., Oritani, K. & Kanakura, Y. (2014) "Early B lymphocyte development: Similarities and differences in human and mouse." *World J Stem Cells*, 6(4): 421-431.

Ivanovs, A., Rybtsov, S., Welch, L., Anderson, R.A., Turner, M.L. & Medvinsky, A. (2011) "Highly potent human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros region." *J Exp Med*, 208(12): 2417-2427.

Jackson, T.R., Ling, R.E. & Roy, A. (2021) "The Origin of B-cells: Human Fetal B Cell Development and Implications for the Pathogenesis of Childhood Acute Lymphoblastic Leukemia." *Front Immunol*, 12: 637975.

Jacoby, E., Nguyen, S.M., Fountaine, T.J., Welp, K., Gryder, B., Qin, H., Yang, Y., Chien, C.D., Seif, A.E., Lei, H., Song, Y.K., Khan, J., Lee, D.W., Mackall, C.L., Gardner, R.A., Jensen, M.C., Shern, J.F. & Fry, T.J. (2016) "CD19 CAR immune pressure induces B-precursor acute lymphoblastic leukaemia lineage switch exposing inherent leukaemic plasticity." *Nat Commun*, 7: 12320.

Jacobsen, S.E.W. & Nerlov, C. (2019) "Haematopoiesis in the era of advanced single-cell technologies." *Nat Cell Biol*, 21(1): 2-8.

Jansen, M.W., Corral, L., van der Velden, V.H., Panzer-Grumayer, R., Schrappe, M., Schrauder, A., Marschalek, R., Meyer, C., den Boer, M.L., Hop, W.J., Valsecchi, M.G., Basso, G., Biondi, A., Pieters, R. & van Dongen, J.J. (2007) "Immunobiological diversity in infant acute lymphoblastic leukemia is related to the occurrence and type of *MLL* gene rearrangement." *Leukemia*, 21(4): 633-641.

Jardine, L., Webb, S., Goh, I., Quiroga Londono, M., Reynolds, G., Mather, M., Olabi, B., Stephenson, E., Botting, R.A., Horsfall, D., Engelbert, J., Maunder, D., Mende, N., Murnane, C., Dann, E., McGrath, J., King, H., Kucinski, I., Queen, R., Carey, C.D., Shrubsole, C., Poyner, E., Acres, M., Jones, C., Ness, T., Coulthard, R., Elliott, N., O'Byrne, S., Haltalli, M.L.R., Lawrence, J.E., Lisgo, S., Balogh, P., Meyer, K.B., Prigmore, E., Ambridge, K., Jain, M.S., Efremova, M., Pickard, K., Creasey, T., Bacardit, J., Henderson, D., Coxhead, J., Filby, A., Hussain, R., Dixon, D., McDonald, D., Popescu, D.M., Kowalczyk, M.S., Li, B., Ashenberg, O., Tabaka, M., Dionne, D., Tickle, T.L., Slyper, M., Rozenblatt-Rosen, O., Regev, A., Behjati, S., Laurenti, E., Wilson, N.K., Roy, A., Göttgens, B., Roberts, I., Teichmann, S.A. & Haniffa, M. (2021) "Blood and immune development in human fetal bone marrow and Down syndrome." *Nature*, 598(7880): 327-331.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A. & Charpentier, E. (2012) "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity." *Science*, 337(6096): 816-821.

Jude, C.D., Climer, L., Xu, D., Artinger, E, Fisher, J.K. & Ernst, P. (2007) "Unique and independent roles for MLL in adult hematopoietic stem cells and progenitors." *Cell Stem Cell*, 1(3): 324-337.

Karamitros, D., Stoilova, B., Aboukhalil, Z., Hamey, F., Reinisch, A., Samitsch, M., Quek, L., Otto, G., Repapi, E., Doondeea, J., Usukhbayar, B., Calvo, J., Taylor, S., Goardon, N., Six, E., Pflumio, F., Porcher, C., Majeti, R., Gottgens, B. & Vyas, P. (2018) "Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells." *Nat Immunol*, 19(1): 85-97.

Kerry, J., Godfrey, L., Repapi, E., Tapia, M., Blackledge, N.P., Ma, H., Ballabio, E., O'Byrne, S., Ponthan, F., Heidenreich, O., Roy, A., Roberts, I., Konopleva, M., Klose, R.J., Geng, H. & Milne, T.A. (2017) "MLL-AF4 Spreading Identifies Binding Sites that Are Distinct from Super-Enhancers and that Govern Sensitivity to DOT1L Inhibition in Leukemia." *Cell Rep*, 18(2): 482-495.

Korotkevich, G., Sukhov, V. & Sergushichev A. (2018) "Fast gene set enrichment analysis." *bioRxiv*.

Krivtsov, A.V., Twomey, D., Feng, Z., Stubbs, M.C., Wang, Y., Faber, J., Levine, J.E., Wang, J., Hahn, W.C., Gilliland, D.G., Golub, T.R. & Armstrong, S.A. (2006) "Transformation from committed progenitor to leukaemia stem cell initiated by *MLL-AF9*." *Nature*, 442(7104): 818-822.

Krivtsov, A.V., Feng, Z., Lemieux, M.E., Faber, J., Vempati, S., Sinha, A.U., Xia, X., Jesneck, J., Bracken, A.P., Silverman, L.B., Kutok, J.L., Kung, A.L. & Armstrong, S.A. (2008) "H3K79 methylation profiles define murine and human *MLL-AF4* leukemias." *Cancer Cell*, 14(5): 355-368.

Krivtsov, A.V., Figueroa, M.E., Sinha, A.U., Stubbs, M.C., Feng, Z., Valk, P.J., Delwel, R., Dohner, K., Bullinger, L., Kung, A.L., Melnick, A.M. & Armstrong, S.A. (2013) "Cell of origin determines clinically relevant subtypes of *MLL*-rearranged AML." *Leukemia*, 27(4): 852-860.

Krivtsov, A.V., Evans, K., Gadrey, J.Y., Eschle, B.K., Hatton, C., Uckelmann, H.J., Ross, K.N., Perner, F., Olsen, S.N., Pritchard, T., McDermott, L., Jones, C.D., Jing, D., Braytee, A., Chacon, D., Earley, E., McKeever, B.M., Claremon, D., Gifford, A.J., Lee, H.J., Teicher, B.A., Pimanda, J.E., Beck, D., Perry, J.A., Smith, M.A., McGeehan, G.M., Lock, R.B. & Armstrong, S.A. (2019) "A Menin-MLL Inhibitor Induces Specific Chromatin Changes and Eradicates Disease in Models of *MLL*-Rearranged Leukemia." *Cancer Cell*, 36(6): 660-673.

Kumar, A.R., Yao, Q., Li, Q., Sam, T.A. & Kersey, J.H. (2011) "t(4;11) leukemias display addiction to *MLL-AF4* but not to *AF4-MLL*." *Leuk Res*, 35(3): 305-309.

Lansdorp, P.M., Dragowska, W. & Mayani, H. (1993) "Ontogeny-related changes in proliferative potential of human hematopoietic cells." *J Exp Med*, 178(3): 787-791.

Laurenti, E. & Göttgens B. (2018) "From haematopoietic stem cells to complex differentiation landscapes." *Nature*, 553(7689): 418-426.

Lavau, C., Szilvassy, S.J., Slany, R. & Cleary, M.L. (1997) "Immortalization and leukemic transformation of a myelomonocytic precursor by retrovirally transduced *HRX-ENL*." *EMBO J*, 16(14): 4226-4237.

Leach, B.I., Kuntimaddi, A., Schmidt, C.R., Cierpicki, T., Johnson, S.A. & Bushweller, J.H. (2013) "Leukemia fusion target AF9 is an intrinsically disordered transcriptional regulator that recruits multiple partners via coupled folding and binding." *Structure*, 21(1): 176-183.

Lepus, C.M., Gibson, T.F., Gerber, S.A., Kawikova, I., Szczepanik, M., Hossain, J., Ablamunits, V., Kirkiles-Smith, N., Herold, K.C., Donis, R.O., Bothwell, A.L., Pober, J.S. & Harding, M.J. (2009) "Comparison of human fetal liver, umbilical cord blood, and adult blood hematopoietic stem cell engraftment in NOD-*scid*/γc-/-, Balb/c-*Rag1*-/-γc-/-, and C.B-17-*scid*/bg immunodeficient mice." *Hum Immunol*, 70(10): 790-802.

le Viseur, C., Hotfilder, M., Bomken, S., Wilson, K., Rottgers, S., Schrauder, A., Rosemann, A., Irving, J., Stam, R.W., Shultz, L.D., Harbott, J., Jurgens, H., Schrappe, M., Pieters, R. & Vormoor, J. (2008) "In childhood acute lymphoblastic leukemia, blasts at different stages of immunophenotypic maturation have stem cell properties." *Cancer Cell*, 14(1): 47-58.

Li, Y., Wen, H., Xi, Y., Tanaka, K., Wang, H., Peng, D., Ren, Y., Jin, Q., Dent, S.Y., Li, W., Li, H. & Shi, X. (2014) "AF9 YEATS domain links histone acetylation to DOT1L-mediated H3K79 methylation." *Cell*, 159(3): 558-571.

Liang, D.C., Shih, L.Y., Fu, J.F., Li, H.Y., Wang, H.I., Hung, I.J., Yang, C.P., Jaing, T.H., Chen, S.H. & Liu, H.C. (2006) "*K-Ras* mutations and *N-Ras* mutations in childhood acute leukemias with or without *mixed-lineage leukemia* gene rearrangements." *Cancer*, 106(4): 950-956.

Lin, C., Smith, E.R., Takahashi, H., Lai, K.C., Martin-Brown, S., Florens, L., Washburn, M.P., Conaway, J.W., Conaway, R.C. & Shilatifard, A. (2010) "AFF4, a component of the ELL/P-TEFb elongation complex and a shared subunit of *MLL* chimeras, can link transcription elongation to leukemia." *Mol Cell*, 37(3): 429-437.

Lin, S., Luo, R.T., Ptasinska, A., Kerry, J., Assi, S. A., Wunderlich, M., Imamura, T., Kaberlein, J.J., Rayes, A., Althoff, M.J., Anastasi, J., O'Brien, M.M., Meetei, A.R., Milne, T.A., Bonifer, C., Mulloy, J.C. & Thirman, M.J. (2016) "Instructive Role of MLL-Fusion Proteins Revealed by a Model of t(4;11) Pro-B Acute Lymphoblastic Leukemia." *Cancer Cell*, 30(5): 737-749.

Lin, S., Luo, R.T., Shrestha, M., Thirman, M.J. & Mulloy, J.C. (2017) "The full transforming capacity of *MLL-Af4* is interlinked with lymphoid lineage commitment." *Blood*, 130(7): 903-907.

Lobato, M.N., Metzler, M., Drynan, L., Forster, A., Pannell, R. & Rabbitts, T.H. (2008) "Modeling chromosomal translocations using conditional alleles to

recapitulate initiating events in human leukemias." *J Natl Cancer Inst Monogr*, 39: 58-63.

Long, H. K., Rose, N. R., Blackledge, N. P. & Klose, R. J. (2018) "Biochemical Identification of Nonmethylated DNA by BioCAP-Seq." *Methods Mol Biol*, 1766: 15-29.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E. & Church, G.M. (2013) "RNA-guided human genome engineering via Cas9." *Science*, 339(6121): 823-826.

Malouf, C. & Ottersbach, K. (2018) "The fetal liver lymphoid-primed multipotent progenitor provides the prerequisites for the initiation of t(4;11) *MLL-AF4* infant leukemia." *Haematologica*, 93: 38-43 (e5).

Malouf, C., Antunes, E.T.B., O'Dwyer, M., Jakobczyk, H., Sahm, F., Landua, S.L., Anderson, R.A., Soufi, A., Halsey, C. & Ottersbach, K. (2021) "miR-130b and miR-128a are essential lineage-specific codrivers of t(4;11) *MLL-AF4* acute leukemia." *Blood*, 138(21): 2066-2092.

Mann, G., Attarbaschi, A., Schrappe, M., De Lorenzo, P., Peters, C., Hann, I., De Rossi, G., Felice, M., Lausen, B., Leblanc, T., Szczepanski, T., Ferster, A., Janka-Schaub, G., Rubnitz, J., Silverman, L.B., Stary, J., Campbell, M., Li, C.K., Suppiah, R., Biondi, A., Vora, A., Valsecchi, M.G., Pieters, R. & Interfant-99 Study Group (2010) "Improved outcome with hematopoietic stem cell transplantation in a poor prognostic subgroup of infants with *mixed-lineage-leukemia (MLL)*-rearranged acute lymphoblastic leukemia: results from the Interfant-99 Study." *Blood*, 116(15): 2644-2650.

Marschalek, R. (2020) "The reciprocal world of *MLL* fusions: A personal view." *Biochim Biophys Acta Gene Regul Mech*, 1863(7): 194547.

Marshall, N.F. & Price, D.H. (1995) "Purification of P-TEFb, a transcription factor required for the transition into productive elongation." *J Biol Chem*, 270(21): 12335-12338.

Marshall, N.F., Peng, J., Xie, Z. & Price, D.H. (1996) "Control of RNA polymerase II elongation potential by a novel carboxyl-terminal domain kinase." *J Biol Chem*, 271(43): 27176-27183.

McMahon, K.A., Hiew, S.Y., Hadjur, S., Veiga-Fernandes, H., Menzel, U., Price, A.J., Kioussis, D., Williams, O. & Brady, H.J. (2007) "*Mll* has a critical role in fetal and adult hematopoietic stem cell self-renewal." *Cell Stem Cell*, 1(3): 338-345.

Meder, L., Konig, K., Dietlein, F., Macheleidt, I., Florin, A., Ercanoglu, M.S., Rommerscheidt-Fuss, U., Koker, M., Schon, G., Odenthal, M., Klein, F., Buttner, R., Schulte, J.H., Heukamp, L.C. & Ullrich, R.T. (2018) "*LIN28B* enhanced tumorigenesis in an autochthonous *KRAS*(G12V)-driven lung carcinoma mouse model." *Oncogene*, 37(20): 2746-2756.

Medvinsky, A. & Dzierzak, E. (1996) "Definitive hematopoiesis is autonomously initiated by the AGM region." *Cell*, 86(6): 897-906.

Metzler, M., Forster, A., Pannell, R., Arends, M.J., Daser, A., Lobato, M.N. & Rabbitts, T.H. (2006) "A conditional model of *MLL-AF4* B-cell tumourigenesis using invertor technology." *Oncogene*, 25(22): 3093-3103.

Meyer, C., Hofmann, J., Burmeister, T., Groger, D., Park, T.S., Emerenciano, M., Pombo de Oliveira, M., Renneville, A., Villarese, P., Macintyre, E., Cave, H., Clappier, E., Mass-Malo, K., Zuna, J., Trka, J., De Braekeleer, E., De Braekeleer, M., Oh, S.H., Tsaur, G., Fechina, L., van der Velden, V.H., van Dongen, J.J., Delabesse, E., Binato, R., Silva, M.L., Kustanovich, A., Aleinikova, O., Harris, M.H., Lund-Aho, T., Juvonen, V., Heidenreich, O., Vormoor, J., Choi, W.W., Jarosova, M., Kolenova, A., Bueno, C., Menendez, P., Wehner, S., Eckert, C., Talmant, P., Tondeur, S., Lippert, E., Launay, E., Henry, C., Ballerini, P., Lapillone,

H., Callanan, M.B., Cayuela, J.M., Herbaux, C., Cazzaniga, G., Kakadiya, P.M., Bohlander, S., Ahlmann, M., Choi, J.R., Gameiro, P., Lee, D.S., Krauter, J., Cornillet-Lefebvre, P., Te Kronnie, G., Schafer, B.W., Kubetzko, S., Alonso, C.N., zur Stadt, U., Sutton, R., Venn, N.C., Izraeli, S., Trakhtenbrot, L., Madsen, H.O., Archer, P., Hancock, J., Cerveira, N., Teixeira, M.R., Lo Nigro, L., Moricke, A., Stanulla, M., Schrappe, M., Sedek, L., Szczepanski, T., Zwaan, C.M., Coenen, E.A., van den Heuvel-Eibrink, M.M., Strehl, S., Dworzak, M., Panzer-Grumayer, R., Dingermann, T., Klingebiel, T. & Marschalek, R. (2013) "The *MLL* recombinome of acute leukemias in 2013." *Leukemia*, 27(11): 2165-2176.

Meyer, C., Burmeister, T., Groger, D., Tsaur, G., Fechina, L., Renneville, A., Sutton, R., Venn, N.C., Emerenciano, M., Pombo-de-Oliveira, M.S., Barbieri Blunck, C., Almeida Lopes, B., Zuna, J., Trka, J., Ballerini, P., Lapillonne, H., De Braekeleer, M., Cazzaniga, G., Corral Abascal, L., van der Velden, V.H.J., Delabesse, E., Park, T.S., Oh, S.H., Silva, M.L.M., Lund-Aho, T., Juvonen, V., Moore, A.S., Heidenreich, O., Vormoor, J., Zerkalenkova, E., Olshanskaya, Y., Bueno, C., Menendez, P., Teigler-Schlegel, A., Zur Stadt, U., Lentes, J., Gohring, G., Kustanovich, A., Aleinikova, O., Schafer, B.W., Kubetzko, S., Madsen, H.O., Gruhn, B., Duarte, X., Gameiro, P., Lippert, E., Bidet, A., Cayuela, J.M., Clappier, E., Alonso, C.N., Zwaan, C.M., van den Heuvel-Eibrink, M.M., Izraeli, S., Trakhtenbrot, L., Archer, P., Hancock, J., Moricke, A., Alten, J., Schrappe, M., Stanulla, M., Strehl, S., Attarbaschi, A., Dworzak, M., Haas, O.A., Panzer-Grumayer, R., Sedek, L., Szczepanski, T., Caye, A., Suarez, L., Cave, H. & Marschalek, R. (2018) "The *MLL* recombinome of acute leukemias in 2017." *Leukemia*, 32(2): 273-284.

Milne, T.A., Briggs, S.D., Brock, H.W., Martin, M.E., Gibbs, D., Allis, C.D. & Hess, J.L. (2002) "MLL targets SET domain methyltransferase activity to *Hox* gene promoters." *Mol Cell*, 10(5): 1107-1117.

Milne, T.A., Martin, M.E., Brock, H.W., Slany, R.K. & Hess, J.L. (2005) "Leukemogenic MLL fusion proteins bind across a broad region of the *Hoxa9*

locus, promoting transcription and multiple histone modifications." *Cancer Res*, 65(24): 11367-11374.

Milne, T.A., Kim, J., Wang, G.G., Stadler, S.C., Basrur, V., Whitcomb, S.J., Wang, Z., Ruthenburg, A.J., Elenitoba-Johnson, K.S., Roeder, R.G. & Allis, C.D. (2010) "Multiple interactions recruit MLL1 and MLL1 fusion proteins to the *HOXA9* locus in leukemogenesis." *Mol Cell*, 38(6): 853-863.

Milne, T.A., (2017) "Mouse models of *MLL* leukemia: recapitulating the human disease." *Blood*, 129(16): 2217-2223.

Molenaar, J.J., Domingo-Fernandez, R., Ebus, M.E., Lindner, S., Koster, J., Drabek, K., Mestdagh, P., van Sluis, P., Valentijn, L.J., van Nes, J., Broekmans, M., Haneveld, F., Volckmann, R., Bray, I., Heukamp, L., Sprussel, A., Thor, T., Kieckbusch, K., Klein-Hitpass, L., Fischer, M., Vandesompele, J., Schramm, A., van Noesel, M.M., Varesio, L., Speleman, F., Eggert, A., Stallings, R.L., Caron, H.N., Versteeg, R. & Schulte, J.H. (2012) "LIN28B induces neuroblastoma and enhances *MYCN* levels via *let-7* suppression." *Nat Genet*, 44(11): 1199-1206.

Moorman, A.V., Ensor, H.M., Richards, S.M., Chilton, L., Schwab, C., Kinsey, S.E., Vora, A., Mitchell, C.D. & Harrison, C.J. (2010) "Prognostic effect of chromosomal abnormalities in childhood B-cell precursor acute lymphoblastic leukaemia: results from the UK Medical Research Council ALL97/99 randomised trial." *The Lancet Oncology*, 11(5): 429-438.

Mueller, D., Bach, C., Zeisig, D., Garcia-Cuellar, M.P., Monroe, S., Sreekumar, A., Zhou, R., Nesvizhskii, A., Chinnaiyan, A., Hess, J.L. & Slany, R.K. (2007) "A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification." *Blood*, 110(13): 4445-4454.

Muench, M.O., Cupp, J., Polakoff, J., Roncarolo, M.G. (1994) "Expression of CD33, CD38, and HLA-DR on CD34+ human fetal liver progenitors with a high proliferative potential." *Blood*, 83(11): 3170-3181.

Mullighan, C.G., Goorha, S., Radtke, I., Miller, C.B., Coustan-Smith, E., Dalton, J.D., Girtman, K., Mathew, S., Ma, J., Pounds, S.B., Su, X., Pui, C.H., Relling, M.V., Evans, W.E., Shurtleff, S.A. & Downing, J.R. (2007) "Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia." *Nature*, 446(7137): 758-764.

Muntean, A.G., Giannola, D., Udager, A.M. & Hess, J.L. (2008) "The PHD fingers of MLL block MLL fusion protein-mediated transformation." *Blood*, 112(12): 4690-4693.

Muntean, A.G., Tan, J., Sitwala, K., Huang, Y., Bronstein, J., Connelly, J.A., Basrur, V., Elenitoba-Johnson, K.S. & Hess, J.L. (2010) "The PAF complex synergizes with MLL fusion proteins at *HOX* loci to promote leukemogenesis." *Cancer Cell*, 17(6): 609-621.

Nakamura, T., Mori, T., Tada, S., Krajewski, W., Rozovskaia, T., Wassell, R., Dubois, G., Mazo, A., Croce, C.M. & Canaani, E. (2002) "ALL-1 is a histone methyltransferase that assembles a supercomplex of proteins involved in transcriptional regulation." *Mol Cell*, 10(5): 1119-1128.

Notta, F., Zandi, S., Takayama, N., Dobson, S., Gan, O.I., Wilson, G., Kaufmann, K.B., McLeod, J., Laurenti, E., Dunant, C.F., McPherson, J.D., Stein, L.D., Dror, Y. & Dick, J.E. (2016) "Distinct routes of lineage development reshape the human blood hierarchy across ontogeny." *Science*, 351(6269): aab2116.

O'Byrne, S., Elliott, N., Rice, S., Buck, G., Fordham, N., Garnett, C., Godfrey, L., Crump, N.T., Wright, G., Inglott, S., Hua, P., Psaila, B., Povinelli, B., Knapp, Djhf, Agraz-Doblas, A., Bueno, C., Varela, I., Bennett, P., Koohy, H., Watt, S. M., Karadimitris, A., Mead, A.J., Ancliff, P., Vyas, P., Menendez, P., Milne, T.A., Roberts, I. & Roy, A. (2019) "Discovery of a CD10 negative B-progenitor in human fetal life identifies unique ontogeny-related developmental programs." *Blood*.

Okuda, H., Kawaguchi, M., Kanai, A., Matsui, H., Kawamura, T., Inaba, T., Kitabayashi, I. & Yokoyama, A. (2014) "MLL fusion proteins link transcriptional coactivators to previously active CpG-rich promoters." *Nucleic Acids Res*, 42(7): 4241-4256.

Pal, D., Blair, H.J., Elder, A., Dormon, K., Rennie, K.J., Coleman, D.J., Weiland, J., Rankin, K.S., Filby, A., Heidenreich, O. & Vormoor, J. (2016) "Long-term *in vitro* maintenance of clonal abundance and leukaemia-initiating potential in acute lymphoblastic leukaemia." *Leukemia*, 30(8): 1691-1700.

Pieters, R., Loonen, A.H., Huismans, D.R., Broekema, G.J., Dirven, M.W., Heyenbrok, M.W., Hahlen, K. & Veerman, A.J. (1990) "*In vitro* drug sensitivity of cells from children with leukemia using the MTT assay with improved culture conditions." *Blood*, 76(11): 2327-2336.

Pieters, R., Huismans, D.R., Loonen, A.H., Hahlen, K., van der Does-van den Berg, A., van Wering, E.R. & Veerman, A.J. (1991) "Relation of cellular drug resistance to long-term clinical outcome in childhood acute lymphoblastic leukaemia." *Lancet*, 338(8764): 399-403.

Pieters, R. ,Kaspers, G.J., Klumper, E. & Veerman, A.J. (1994) "Clinical relevance of *in vitro* drug resistance testing in childhood acute lymphoblastic leukemia: the state of the art." *Med Pediatr Oncol*, 22(5): 299-308.

Pieters, R., Schrappe, M., De Lorenzo, P., Hann, I., De Rossi, G., Felice, M., Hovi, L., LeBlanc, T., Szczepanski, T., Ferster, A., Janka, G., Rubnitz, J., Silverman, L., Stary, J., Campbell, M., C. Li, K., Mann, G., Suppiah, R., Biondi, A., Vora, A. & Valsecchi, M.G. (2007) "A treatment protocol for infants younger than 1 year with acute lymphoblastic leukaemia (Interfant-99): an observational study and a multicentre randomised trial." *Lancet*, 370(9583): 240-250.

Pieters, R., De Lorenzo, P., Ancliffe, P., Aversa, L.A., Brethon, B., Biondi, A., Campbell, M., Escherich, Ferster, G.A., Gardner, R.A., Kotecha, R.S., Lausen,

B., Li, C.K., Locatelli, F., Attarbaschi, A., Peters, C., Rubnitz, J.E., Silverman, L.B., Stary, J., Szczepanski, T., Vora, A., Schrappe, M. & Valsecchi, M.G. (2019) "Outcome of Infants Younger Than 1 Year With Acute Lymphoblastic Leukemia Treated With the Interfant-06 Protocol: Results From an International Phase III Randomized Study." *J Clin Oncol* 37(25): 2246-2256.

Popescu, D.M., Botting, R.A., Stephenson, E., Green, K., Webb, S., Jardine, L., Calderbank, E.F., Polanski, K., Goh, I., Efremova, M., Acres, M., Maunder, D., Vegh, P., Gitton, Y., Park, J.E., Vento-Tormo, R., Miao, Z., Dixon, D., Rowell, R., McDonald, D., Fletcher, J., Poyner, E., Reynolds, G., Mather, M., Moldovan, C., Mamanova, L., Greig, F., Young, M.D., Meyer, K.B., Lisgo, S., Bacardit, J., Fuller, A., Millar, B., Innes, B., Lindsay, S., Stubbington, M.J.T., Kowalczyk, M.S., Li, B., Ashenberg, O., Tabaka, M., Dionne, D., Tickle, T.L., Slyper, M., Rozenblatt-Rosen, O., Filby, A., Carey, P., Villani, A.C., Roy, A., Regev, A., Chedotal, A., Roberts, I., Göttgens, B., Behjati, S., Laurenti, E., Teichmann, S.A. & Haniffa, M. (2019) "Decoding human fetal liver haematopoiesis." *Nature*, 574(7778): 365-371.

Pradeepa, M.M., Sutherland, H.G., Ule, J., Grimes, G.R. & Bickmore, W.A. (2012) "Psip1/Ledgf p52 binds methylated histone H3K36 and splicing factors and contributes to the regulation of alternative splicing." *PLoS Genet*, 8(5): e1002717.

Pradeepa, M.M., Grimes, G.R., Taylor, G.C., Sutherland, H.G. & Bickmore, W.A. (2014) "Psip1/Ledgf p75 restrains *Hox* gene expression by recruiting both trithorax and polycomb group proteins." *Nucleic Acids Res*, 42(14): 9012-9032.

Prieto, C., Marschalek, R., Kuhn, A., Bursen, A., Bueno, C. & Menendez, P. (2017) "The *AF4-MLL* fusion transiently augments multilineage hematopoietic engraftment but is not sufficient to initiate leukemia in cord blood CD34(+) cells." *Oncotarget*, 8(47): 81936-81941.

Pui, C.H. (2000) "Acute lymphoblastic leukemia in children." *Curr Opin Oncol*, 12(1): 3-12.

Pui, C.H. & Campana, D. (2007) "Age-related differences in leukemia biology and prognosis: the paradigm of *MLL-AF4*-positive acute lymphoblastic leukemia." *Leukemia*, 21(4): 593-594.

Ramakers-van Woerden, N.L., Beverloo, H.B., Veerman, A.J., Camitta, B.M., Loonen, A.H., van Wering, E.R., Slater, R.M., Harbott, J., den Boer, M.L., Ludwig, W.D., Haas, O.A., Janka-Schaub, G.E. & Pieters, R. (2004) "*In vitro* drug-resistance profile in infant acute lymphoblastic leukemia in relation to age, *MLL* rearrangements and immunophenotype." *Leukemia*, 18(3): 521-529.

Rayes, A., McMasters, R.L. & O'Brien, M.M. (2016) "Lineage Switch in *MLL*-Rearranged Infant Leukemia Following CD19-Directed Therapy." *Pediatr Blood Cancer*, 63(6): 1113-1115.

Rebel, V.I., Miller, C.L., Eaves, C.J. & Lansdorp, P.M. (1996) "The repopulation potential of fetal liver hematopoietic stem cells in mice exceeds that of their liver adult bone marrow counterparts." *Blood*, 87(8): 3500-3507.

Rehe, K., Wilson, K., Bomken, S., Williamson, D., Irving, J., den Boer, M. L., Stanulla, M., Schrappe, M., Hall, A. G., Heidenreich, O. & Vormoor, J. (2013) "Acute B lymphoblastic leukaemia-propagating cells are present at high frequency in diverse lymphoblast populations." *EMBO Mol Med*, 5(1): 38-51.

Reichel, M., Gillert, E., Angermuller, S., Hensel, J.P., Heidel, F., Lode, M., Leis, T., Biondi, A., Haas, O.A., Strehl, S., Panzer-Grumayer, E.R., Griesinger, F., Beck, J.D., Greil, J., Fey, G.H., Uckun, F.M. & Marschalek, R. (2001) "Biased distribution of chromosomal breakpoints involving the *MLL* gene in infants versus children and adults with t(4;11) ALL." *Oncogene*, 20(23): 2900-2907.

Reimer, J., Knoss, S., Labuhn, M., Charpentier, E.M., Gohring, G., Schlegelberger, B., Klusmann, J.H. & Heckl, D. (2017) "CRISPR-Cas9-induced t(11;19)/*MLL-ENL* translocations initiate leukemia in human hematopoietic progenitor cells *in vivo*." *Haematologica*, 102(9): 1558-1566.

Rice, S., Jackson, T., Crump, N.T., Fordham, N., Elliott, N., O'Byrne, S., Fanego, Mdml, Addy, D., Crabb, T., Dryden, C., Inglott, S., Ladon, D., Wright, G., Bartram, J., Ancliff, P., Mead, A.J., Halsey, C., Roberts, I., Milne, T.A. & Roy, A. (2021) "A human fetal liver-derived infant *MLL-AF4* acute lymphoblastic leukemia model reveals a distinct fetal gene expression program." *Nat Commun*, 12(1): 6905.

Richardson, S.E., Ghazanfari, R., Chhetri, J., Enver, T. & Boiers, C. (2021) "*In vitro* differentiation of human pluripotent stem cells into the B lineage using OP9-MS5 co-culture." *STAR Protoc*, 2(2): 100420.

Roberts, I., Fordham, N.J., Rao, A. & Bain, B.J. (2018). "Neonatal leukaemia." *Br J Haematol*, 182(2): 170-184.

Rossi, M.I., Yokota, T., Medina, K.L., Garrett, K.P., Comp, P.C., Schipul, A.H., Jr. & Kincade, P.W. (2003) "B lymphopoiesis is active throughout human life, but there are developmental age-related changes." *Blood*, 101(2): 576-584.

Rossi, J.G., Bernasconi, A.R., Alonso, C.N., Rubio, P.L., Gallego, M.S., Carrara, C.A., Guitter, M.R., Eberle, S.E., Cocce, M., Zubizarreta, P.A. & Felice, M.S. (2012) "Lineage switch in childhood acute leukemia: an unusual event with poor outcome." *Am J Hematol*, 87(9): 890-897.

Rowe, R.G., Lummertz da Rocha, E., Sousa, P., Missios, P., Morse, M., Marion, W., Yermalovich, A., Barragan, J., Mathieu, R., Jha, D.K., Fleming, M.D., North, T.E. & Daley, G.Q. (2019) "The developmental stage of the hematopoietic niche regulates lineage in *MLL*-rearranged leukemia." *J Exp Med*.

Roy, A., Cowan, G., Mead, A.J., Filippi, S., Bohn, G., Chaidos, A., Tunstall, O., Chan, J.K., Choolani, M., Bennett, P., Kumar, S., Atkinson, D., Wyatt-Ashmead, J., Hu, M., Stumpf, M.P., Goudevenou, K., O'Connor, D., Chou, S.T., Weiss, M.J., Karadimitris, A., Jacobsen, S.E., Vyas, P. & Roberts, I. (2012) "Perturbation of fetal liver hematopoietic stem and progenitor cell development by trisomy 21." *Proc Natl Acad Sci USA*, 109(43): 17579-17584.

Roy, A., Wang, G., Iskander, D., O'Byrne, S., Elliott, N., O'Sullivan, J., Buck, G., Heuston, E.F., Wen, W.X., Meira, A.R., Hua, P., Karadimitris, A., Mead, A.J., Bodine, D.M., Roberts, I., Psaila, B. & Thongjuea, S. (2021) "Transitions in lineage specification and gene regulatory networks in hematopoietic stem/progenitor cells over human development." *Cell Rep*, 36(11): 109698.

Sanchez, M.J., Holmes, A., Miles, C. & Dzierzak, E. (1996) "Characterization of the first definitive hematopoietic stem cells in the AGM and liver of the mouse embryo." *Immunity*, 5(6): 513-525.

Sanders, D.S., Muntean, A.G. & Hess, J.L. (2011) "Significance of *AF4-MLL* reciprocal fusion in t(4;11) leukemias?" *Leuk Res*, 35(3): 299-300.

Secker, K.A., Keppeler, H., Duerr-Stoerzer, S., Schmid, H., Schneidawind, D., Hentrich, T., Schulze, Hentrich, J.M., Mankel, B., Fend, F. & Schneidawind, C. (2019) "Inhibition of DOT1L and PRMT5 promote synergistic anti-tumor activity in a human *MLL* leukemia model induced by CRISPR/Cas9." *Oncogene*, 38(46): 7181-7195.

Shi, Y., Beckett, M.C., Blair, H.J., Tirtakusuma, R., Nakjang, S., Enshaei, A., Halsey, C., Vormoor, J., Heidenreich, O., Krippner-Heidenreich & A., van Delft, F.W. (2020) "Phase II-like murine trial identifies synergy between dexamethasone and dasatinib in T-cell acute lymphoblastic leukemia." *Haematologica*.

Silverman, L. B. (2007) "Acute lymphoblastic leukemia in infancy." *Pediatr Blood Cancer*, 49(7 Suppl): 1070-1073.

Somervaille, T.C. & Cleary, M.L. (2006) "Identification and characterization of leukemia stem cells in murine *MLL-AF9* acute myeloid leukemia." *Cancer Cell*, 10(4): 257-268.

Stam, R.W., Schneider, P., de Lorenzo, P., Valsecchi, M.G., den Boer, M.L. & Pieters, R. (2007) "Prognostic significance of high-level *FLT3* expression in *MLL*-rearranged infant acute lymphoblastic leukemia." *Blood*, 110(7): 2775-2775.

Stam, R.W., Den Boer, M.L., Schneider, P., de Boer, J., Hagelstein, J., Valsecchi, M.G., de Lorenzo, P., Sallan, S.E., Brady, H.J., Armstrong, S.A. & Pieters, R. (2010) "Association of high-level *MCL-1* expression with *in vitro* and *in vivo* prednisone resistance in *MLL*-rearranged infant acute lymphoblastic leukemia." *Blood*, 115(5): 1018-1025.

Stam, R.W., Schneider, P., Hagelstein, J.A., van der Linden, M.H., Stumpel, D.J., de Menezes, R.X., de Lorenzo, P., Valsecchi, M.G. & Pieters, R. (2010) "Gene expression profiling-based dissection of *MLL* translocated and *MLL* germline acute lymphoblastic leukemia in infants." *Blood*, 115(14): 2835-2844.

Stavropoulou, V., Kaspar, S., Brault, L., Sanders, M.A., Juge, S., Morettini, S., Tzankov, A., Iacovino, M., Lau, I.J., Milne, T.A., Royo, H., Kyba, M., Valk, P.J.M. & Peters, Ahfm, Schwaller, J. (2016) "*MLL-AF9* Expression in Hematopoietic Stem Cells Drives a Highly Invasive AML Expressing EMT-Related Genes Linked to Poor Outcome." *Cancer Cell*, 30(1): 43-58.

Steger, D.J., Lefterova, M.I., Ying, L., Stonestrom, A.J., Schupp, M., Zhuo, D., Vakoc, A.L., Kim, J.E., Chen, J., Lazar, M.A., Blobel, G.A. & Vakoc, C.R. (2008) "DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells." *Mol Cell Biol*, 28(8): 2825-2839.

Stein, E.M., Garcia-Manero, G., Rizzieri, D.A., Tibes, R., Berdeja, J.G., Savona, M.R., Jongen-Lavrenic, M., Altman, J.K., Thomson, B., Blakemore, S.J., Daigle, S.R., Waters, N.J., Suttle, A.B., Clawson, A., Pollock, R., Krivtsov, A., Armstrong, S.A., DiMartino, J., Hedrick, E., Lowenberg, B. & Tallman, M.S. (2018) "The DOT1L inhibitor pinometostat reduces H3K79 methylation and has modest clinical activity in adult acute leukemia." *Blood*, 131(24): 2661-2669.

Stein, E.M., Aldoss, I., DiPersio, J.F., Stone, R.M., Arellano, M.L., Rosen, G., Meyers, M.L., Huang, Y., Smith, S., Bagley, R.G. & Thirman, M. (2021) "Safety and Efficacy of Menin Inhibition in Patients (Pts) with *MLL*-Rearranged and *NPM1*

Mutant Acute Leukemia: A Phase (Ph) 1, First-in-Human Study of SNDX-5613 (AUGMENT 101)." *Blood*, 138(Supplement 1): 699-699.

Symeonidou, V. & Ottersbach, K. (2021) "*HOXA9/IRX1* expression pattern defines two subgroups of infant *MLL-AF4*-driven acute lymphoblastic leukemia." *Exp Hematol*, 93: 38-43.

Symeonidou, V., Jakobczyk, H., Bashanfer, S., Malouf, C., Fotopoulou, F., Kotecha, R.S., Anderson, R.A., Finch, A.J. & Ottersbach, K. (2021) "Defining the fetal origin of *MLL-AF4* infant leukemia highlights specific fatty acid requirements." *Cell Rep*, 37(4): 109900.

Takahashi, S. & Yokoyama, A. (2020) "The molecular functions of common and atypical MLL fusion protein complexes." *Biochim Biophys Acta Gene Regul Mech*, 1863(7): 194548.

Tavian, M., Cortes, F., Charbord, P., Labastie, M.C. & Peault, B. (1999) "Emergence of the haematopoietic system in the human embryo and foetus." *Haematologica*, 84 Suppl EHA-4: 1-3.

Tavian, M., Hallais, M.F. & Peault, B. (1999) "Emergence of intraembryonic hematopoietic precursors in the pre-liver human embryo." *Development*, 126(4): 793-803.

Telenius, J. & Hughes, J.R. (2018) "NGseqBasic - a single-command UNIX tool for ATAC-seq, DNaseI-seq, Cut-and-Run, and ChIP-seq data mapping, high-resolution visualisation, and quality control." *bioRxiv*.

Thirman, M.J., Gill, H.J., Burnett, R.C., Mbangkollo, D., McCabe, N.R., Kobayashi, H., Ziemin-van der Poel, S., Kaneko, Y., Morgan, R., Sandberg, A.A. Chaganti, R.S.K., Larson, R.A., Le Beau, M.M., Diaz, M.O., & Rowley, J.D. (1993) "Rearrangement of the *MLL* gene in acute lymphoblastic and acute myeloid leukemias with 11q23 chromosomal translocations." *N Engl J Med*, 329(13): 909-914.

Tirtakusuma, R., Szoltysek, K., Milne, P., Grinev, V.V., Ptasinska, A., Meyer, C., Nakjang, S., Hehir-Kwa, J.Y., Williamson, D., Cauchy, P., Assi, S.A., Imperato, M.R., Vogiatzi, F., Lin, S., Wunderlich, M., Stutterheim, J., Komkov, A., Zerkalenkova, E., Evans, P., McNeill, H., Elder, A., Martinez-Soria, N., Fordham, S.E., Shi, Y., Russell, L.J., Pal, D., Smith, A., Kingsbury, Z., Becq, J., Eckert, C., Haas, O.A., Carey, P., Bailey, S., Skinner, R., Miakova, N., Collin, M., Bigley, V., Haniffa, M., Marschalek, R., Harrison, C.J., Cargo, C.A., Schewe, D., Olshanskaya, Y., Thirman, M.J., Cockerill, P.N., Mulloy, J.C., Blair, H.J., Vormoor, J., Allan, J.M., Bonifer, C., Heidenreich, O. & Bomken, S. (2021) "Epigenetic regulator genes direct lineage switching in *MLL-AF4* leukaemia." *bioRxiv*.

Torres, R., Martin, M.C., Garcia, A., Cigudosa, J.C., Ramirez, J.C. & Rodriguez-Perales, S. (2014) "Engineering human tumour-associated chromosomal translocations with the RNA-guided CRISPR-Cas9 system." *Nat Commun*, 5: 3964.

Tkachuk, D.C., Kohler, S. & Cleary, M.L. (1992) "Involvement of a homolog of *Drosophila trithorax* by 11q23 chromosomal translocations in acute leukemias." *Cell*, 71(4):691-700.

Tomizawa, D., Miyamura, T., Imamura, T., Watanabe, T., Moriya Saito, A., Ogawa, A., Takahashi, Y., Hirayama, M., Taki, T., Deguchi, T., Hori, T., Sanada, M., Ohmori, S., Haba, M., Iguchi, A., Arakawa, Y., Koga, Y., Manabe, A., Horibe, K., Ishii, E. & Koh, K. (2020) "A risk-stratified therapy for infants with acute lymphoblastic leukemia: a report from the JPLSG MLL-10 trial." *Blood*, 136(16): 1813-1823.

Trentin, L., Giordan, M., Dingermann, T., Basso, G., Te Kronnie, G. & Marschalek, R. (2009) "Two independent gene signatures in pediatric t(4;11) acute lymphoblastic leukemia patients." *Eur J Haemotol*, 83(5): 406-419.

Trentin, L., Bresolin, S., Giarin, E., Bardini, M., Serafin, V., Accordi, B., Fais, F., Tenca, C., De Lorenzo, P., Valsecchi, M.G., Cazzaniga, G., Kronnie, G.T. & Basso,

G. (2016) "Deciphering *KRAS* and *NRAS* mutated clone dynamics in *MLL-AF4* paediatric leukaemia by ultra deep sequencing analysis." *Sci Rep*, 6: 34449.

Ugale, A., Norddahl, G.L., Wahlestedt, M., Sawen, P., Jaako, P., Pronk, C.J., Soneji, S., Cammenga, J. & Bryder, D. (2014) "Hematopoietic stem cells are intrinsically protected against *MLL-ENL*-mediated transformation." *Cell Rep*, 9(4): 1246-1255.

Ugale, A., Sawen, P., Dudenhoffer-Pfeifer, M., Wahlestedt, M., Norddahl, G.L. & Bryder, D. (2017) "*MLL-ENL*-mediated leukemia initiation at the interface of lymphoid commitment." *Oncogene*. 36(22): 3207-3212.

Viswanathan, S.R., Powers, J.T., Einhorn, W., Hoshida, Y., Ng, T.L., Toffanin, S., O'Sullivan, M., Lu, J., Phillips, L.A., Lockhart, V.L., Shah, S.P., Tanwar, P.S., Mermel, C.H., Beroukhim, R., Azam, M., Teixeira, J., Meyerson, M., Hughes, T.P., Llovet, J.M., Radich, J., Mullighan, C.G., Golub, T.R., Sorensen, P.H. & Daley, G.Q. (2009) "*Lin28* promotes transformation and is associated with advanced human malignancies." *Nat Genet*, 41(7): 843-848.

Vora, A., Goulden, N., Wade, R., Mitchell, C., Hancock, J., Hough, R., Rowntree, C. & Richards, S. (2013). "Treatment reduction for children and young adults with low-risk acute lymphoblastic leukaemia defined by minimal residual disease (UKALL 2003): a randomised controlled trial." *Lancet Oncol*, 14(3): 199-209.

Wang, P., Lin, C., Smith, E.R., Guo, H., Sanderson, B.W., Wu, M., Gogol, M., Alexander, T., Seidel, C., Wiedemann, L.M., Ge, K., Krumlauf, R. & Shilatifard, A. (2009) "Global analysis of H3K4 methylation defines MLL family member targets and points to a role for MLL1-mediated H3K4 methylation in the regulation of transcriptional initiation by RNA polymerase II." *Mol Cell Biol*, 29(22): 6074-6085.

Wang, Z., Song, J., Milne, T.A., Wang, G.G., Li, H., Allis, C.D. & Patel, D.J. (2010) "Pro-isomerization in MLL1 PHD3-bromo cassette connects H3K4me readout to CyP33 and HDAC-mediated repression." *Cell*, 141(7): 1183-1194.

Ward, E., DeSantis, C., Robbins, A., Kohler, B. & Jemal, A. (2014) "Childhood and adolescent cancer statistics, 2014." *Cell*, 141(7): 1183-94.

Wei, J., Wunderlich, M., Fox, C., Alvarez, S., Cigudosa, J.C., Wilhelm, J.S., Zheng, Y., Cancelas, J.A., Gu, Y., Jansen, M., Dimartino, J.F. & Mulloy, J.C. (2008) "Microenvironment determines lineage fate in a human model of *MLL-AF9* leukemia." *Cancer Cell*, 13(6): 483-495.

Wiedenheft, B., Sternberg, S.H. & Doudna, J.A. (2012) "RNA-guided genetic silencing systems in bacteria and archaea." *Nature*, 482(7385): 331-338.

Wiemels, J.L., Cazzaniga, G., Daniotti, M., Eden, O.B., Addison, G.M., Masera, G., Saha, V., Biondi, A. & Greaves, M.F. (1999) "Prenatal origin of acute lymphoblastic leukaemia in children." *Lancet*, 354(9189): 1499-1503.

Wilhelm, A. & Marschalek, R. (2021) "The role of reciprocal fusions in *MLL*-r acute leukemia: studying the chromosomal translocation t(4;11)." *Oncogene*, 40(42): 6093-6102.

Wilkinson, A.C., Ballabio, E., Geng, H., North, P., Tapia, M., Kerry, J., Biswas, D., Roeder, R.G., Allis, C.D., Melnick, A., de Bruijn, M.F. & Milne, T.A. (2013) "RUNX1 is a key target in t(4;11) leukemias that contributes to gene activation through an AF4-MLL complex interaction." *Cell Rep*, 3(1): 116-127.

Wright, G., Watt, E., Inglott, S., Brooks, T., Bartram, J. & Adams, S.P. (2019) "Clinical benefit of a high-throughput sequencing approach for minimal residual disease in acute lymphoblastic leukemia." *Pediatr Blood Cancer*, 66(8): 27787.

Wysocka, J., Swigut, T., Milne, T.A., Dou, Y., Zhang, X., Burlingame, A.L., Roeder, R.G., Brivanlou, A.H. & Allis, C.D. (2005) "WDR5 associates with histone H3 methylated at K4 and is essential for H3 K4 methylation and vertebrate development." *Cell*, 121(6): 859-872.

Yamada, T., Yamaguchi, Y., Inukai, N., Okamoto, S., Mura, T. & Handa, H. (2006) "P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation." *Mol Cell*, 21(2): 227-237.

Yamaguchi, Y., Takag, T., Wada, T., Yano, K., Furuya, A., Sugimoto, S., Hasegawa, J. & Handa, H. (1999) "NELF, a multisubunit complex containing RD, cooperates with DSIF to repress RNA polymerase II elongation." *Cell*, 97(1): 41-51.

Yang, Z., Yik, J.H., Chen, R., He, N., Jang, M.K., Ozato, K. & Zhou, Q. (2005) "Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4." *Mol Cell*, 19(4): 535-545.

Yokoyama, A., Somervaille, T.C., Smith, K.S., Rozenblatt-Rosen, O., Meyerson, M. & Cleary, M.L. (2005) "The menin tumor suppressor protein is an essential oncogenic cofactor for *MLL*-associated leukemogenesis." *Cell*, 123(2): 207-218.

Yokoyama, A. & Cleary, M.L. (2008) "Menin critically links MLL proteins with LEDGF on cancer-associated target genes." *Cancer Cell*, 14(1): 36-46.

Yokoyama, A., Lin, M., Naresh, A., Kitabayashi, I. & Cleary, M.L. (2010) "A higher-order complex containing AF4 and ENL family proteins with P-TEFb facilitates oncogenic and physiologic MLL-dependent transcription." *Cancer Cell*, 17(2): 198-212.

Yoshimoto, M., Montecino-Rodriguez, E., Ferkowicz, M.J., Porayette, P., Shelley, W.C., Conway, S.J., Dorshkind, K. & Yoder, M.C. (2011) "Embryonic day 9 yolk sac and intra-embryonic hemogenic endothelium independently generate a B-1 and marginal zone progenitor lacking B-2 potential." *Proc Natl Acad Sci USA*, 108(4): 1468-1473.

Yu, B.D., Hess, J.L., Horning, S.E., Brown, G.A. & Korsmeyer, S.J. (1995) "Altered *Hox* expression and segmental identity in *Mll*-mutant mice." *Nature*, 378(6556): 505-508.

Yu, B.D., Hanson, R.D., Hess, J.L., Horning, S.E. & Korsmeyer, S.J. (1998) "*MLL*, a mammalian *trithorax-group* gene, functions as a transcriptional maintenance factor in morphogenesis." *Proc Natl Acad Sci USA*, 95(18): 10632-10636.

Yuan, J., Nguyen, C.K., Liu, X., Kanellopoulou, C. & Muljo, S.A. (2012) "*Lin28b* reprograms adult bone marrow hematopoietic progenitors to mediate fetal-like lymphopoiesis." *Science*, 335(6073): 1195-1200.

Zeisig, B.B., Garcia-Cuellar, M.P., Winkler, T.H. & Slany, R.K. (2003) "The oncoprotein *MLL-ENL* disturbs hematopoietic lineage determination and transforms a biphenotypic lymphoid/myeloid cell." *Oncogene*, 22(11): 1629-1637.

Zeisig, D.T., Bittner, C.B., Zeisig, B.B., Garcia-Cuellar, M.P., Hess, J.L. & Slany, R.K. (2005) "The eleven-nineteen-leukemia protein ENL connects nuclear MLL fusion partners with chromatin." *Oncogene*, 24(35): 5525-5532.

Zeleznik-Le, N.J., Harden, A.M. & Rowley, J.D. (1994) "11q23 translocations split the "AT-hook" cruciform DNA-binding region and the transcriptional repression domain from the activation domain of the *mixed-lineage leukemia (MLL)* gene." *Proc Natl Acad Sci USA*, 91(22): 10610-10614.

Zhou, J., Ng, S.B. & Chng, W.J. (2013) "*LIN28/LIN28B*: an emerging oncogenic driver in cancer stem cells." *Int J Biochem Cell Biol*, 45(5): 973-978.