Modeling TEL-AML1 Childhood Acute Lymphoblastic Leukaemia using Human Pluripotent Stem Cells

Simon Edward Richardson

UCL Cancer Institute
University College London

A thesis submitted for the degree of Doctor of Philosophy
DECLARATION

I, Simon Edward Richardson, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.
ABSTRACT

Childhood acute lymphoblastic leukaemia (cALL) is distinct from that in adults with higher incidence, better prognosis and a distinct mutational spectrum. One hypothesis for this difference is that cALL arises in transient cells unique to early human development. I explored this in ETV6-RUNX1 cALL where evidence from twins and neonatal heel prick testing has shown that this mutation arises in utero and is an initiating event. I hypothesised that human pluripotent stem cells (hPSCs) could model the developmental features of in utero B cell development. In vitro B cell differentiation of hPSCs produced both pre and proB cells, and a CD19\*IL7R\+ progenitor that switched from myeloid to lympho-myeloid priming during culture, akin to that identified in parallel studies of human fetal liver (FL). At the global transcriptional level the hPSC lymphoid hierarchy mapped closely with FL, with both separating from adult suggesting that hPSCs provide a developmentally relevant model of early FL B lymphopoiesis. I used CRISPR-directed homologous recombination to engineer the expression of ETV6-RUNX1 under the endogenous ETV6 promoter. ETV6-RUNX1 hPSCs displayed a partial block in B cell differentiation at the level of the CD19\*IL7R\+ progenitor. ETV6-RUNX1 expressing proB cells co-expressed an abnormal B-myeloid gene expression signature similar to that seen in the CD19\*IL7R\+ progenitor. These data support a model where expression of ETV6-RUNX1 inhibits lymphoid specification in an early FL CD19\*IL7R\+ lymphomyeloid progenitor, arresting B lineage differentiation and resulting in the production of myeloid-primed B cells. This may explain the propensity for aberrant myeloid antigen expression seen in cALL. ETV6-RUNX1 hPSCs provide a platform for the systematic evaluation of the contribution of additional mutations seen in cALL and may offer a tractable platform for drug screening. In conclusion I propose the human fetal CD19\*IL7R\+ lymphomyeloid progenitor as a candidate target cell for in utero pre-leukemic initiation in cALL.
## TABLE OF CONTENTS

DECLARATION.................................................................................................................. ii
ABSTRACT .......................................................................................................................... iii
TABLE OF CONTENTS ........................................................................................................ iv
TABLE OF FIGURES ........................................................................................................... vii
ABBREVIATIONS .............................................................................................................. xi
ACKNOWLEDGEMENTS ...................................................................................................... xiv

1. CHAPTER 1: Introduction ......................................................................................... 2
   1.1. Haematopoiesis ........................................................................................................ 2
   1.1.1. Haematopoietic Stem Cells ................................................................................ 2
   1.1.2. Haematopoietic Differentiation Hierarchy ........................................................... 5
   1.1.3. Molecular Basis of Haematopoiesis ................................................................. 9
   1.1.4. Fetal haematopoiesis ........................................................................................ 12
   1.1.4.1. ‘Definitive’ Haematopoiesis ............................................................................. 13
   1.1.4.2. Sites of Fetal Haematopoiesis ....................................................................... 14
   1.1.4.3. Regulation of Haematopoietic Specification ............................................... 18
   1.1.5. B cell development ............................................................................................ 20
   1.1.5.1. Stages of B cell Development ....................................................................... 20
   1.1.5.2. Control of B cell Commitment and differentiation ...................................... 22
   1.1.5.3. Cytokine and PreB cell receptor signaling during B cell Differentiation ... 24
   1.1.5.4. Fetal B Cell Development ............................................................................. 26
   1.1.5.5. Lineage Plasticity and the Ontogeny of B cells ............................................. 28
   1.2. Childhood Acute Lymphoblastic Leukaemia ....................................................... 29
   1.2.1. Evidence for the Prenatal Origins of Childhood ALL ....................................... 30
   1.2.2. First Hit Mutations in ALL ............................................................................. 32
   1.2.3. Second Hit Mutations in ALL .......................................................................... 34
   1.2.4. Cancer Stem Cell Hypothesis ......................................................................... 36
   1.2.5. ETV6-RUNX1 Acute Lymphoblastic Leukaemia .............................................. 37
   1.2.5.1. Current Models of ETV6-RUNX1 ALL ......................................................... 40
   1.2.5.2. Cellular Origins of ETV6-RUNX1 Pre-leukaemia ......................................... 46
   1.3. Pluripotency .......................................................................................................... 50
   1.3.1. Murine Embryonic Stem Cells .......................................................................... 50
   1.3.1.1. Murine Epiblast Stem Cells .......................................................................... 51
   1.3.1.2. Ground State Pluripotency in the Mouse ..................................................... 52
   1.3.2. Human Embryonic Stem Cells ........................................................................ 54
   1.3.2.1. Ground State Human Pluripotent Stem Cells .............................................. 55
   1.3.3. Induced Pluripotent Stem Cells ....................................................................... 56
   1.3.4. In Vitro Differentiation of Human Pluripotent Stem Cells ................................. 57
   1.3.4.1. Mesodermal Development and Early Haematopoiesis ...................... 58
   1.3.4.2. Derivation of Haemogenic Endothelium and Definitive Haematopoiesis 59
   1.3.4.3. Generation of Engraftable dHSCs ................................................................. 61
   1.4. Genome Engineering ........................................................................................... 63
   1.4.1. Homologous Recombination ............................................................................. 64
   1.4.2. Site Specific Recombinases ............................................................................. 65
   1.4.3. Site Specific Nucleases .................................................................................... 65
   1.4.3.1. TALENs ...................................................................................................... 66
2.1.1. Transformation of bacteria .................................................. 73
2.1.2. Isolation of plasmid DNA .................................................. 74
2.1.3. Preparation of BAC DNA .................................................. 76
2.1.4. Restriction enzyme digests .................................................. 76
2.1.5. DNA Sequencing ............................................................. 77
2.1.6. Gel extraction ............................................................... 77
2.1.7. DNA Ligation ............................................................... 78
2.1.8. TALEN Assembly by Golden Gate ..................................... 78
2.1.9. BAC Recombineering ...................................................... 81
2.1.10. Southern Blot Hybridisation ............................................ 84
2.1.11. Bulk total RNA isolation ................................................ 88
2.1.12. cDNA preparation ....................................................... 89
2.1.13. Quantitative Real-Time PCR ............................................ 89
2.1.14. Statistics ...................................................................... 89
2.1.15. DJ rearrangement ........................................................... 90
2.1.16. Single Cell Real-Time PCR (Fluidigm) ............................... 90
2.1.17. Low cell number RNA Sequencing .................................. 93
2.1.18. Bioinformatic Analysis ................................................... 94

2.2. Tissue Culture Protocols .............................................................. 95
2.2.1. Human Pluripotent Stem Cell Culture .................................. 95
  2.2.1.1. Mouse Embryonic Fibroblasts ....................................... 95
  2.2.1.2. Mitotic Inactivation and Freezing MEF Aliquots ................. 95
  2.2.1.3. Preparation of feeder Layers for hPSCs ............................... 96
  2.2.1.4. Human Pluripotent Stem Cells ....................................... 96
  2.2.1.5. Thawing hPSCs ........................................................ 97
  2.2.1.6. Routine Culture and Passage of hPSCs on MEFs .................... 97
  2.2.1.7. Maintenance of hPSCs in Matrigel/mTeSR1 or StemFit® ........ 98
  2.2.1.8. Passaging with Gentle Dissociation Reagent ...................... 99
  2.2.1.9. Nuclear staining of fixed and permeabilised cells .................. 99
  2.2.1.10. Karyotyping ........................................................... 100

2.2.2. In vitro B cell Differentiation from hPSCs .............................. 100
  2.2.2.1. OP9 Stroma Maintenance ............................................. 100
  2.2.2.2. Harvesting hPSCs for OP9 Co-culture ................................ 101
  2.2.2.3. Harvesting CD34+ cells ............................................... 101
  2.2.2.4. Magnetic Separation of CD34+ Cells ................................. 102
  2.2.2.5. MSS Maintenance and Co-culture .................................. 103

2.2.3. In vitro differentiation assays .............................................. 103
  2.2.3.1. MSS subculture ........................................................ 103
  2.2.3.2. Semi-solid colony forming assays by methylcellulose ............ 104
  2.2.3.3. Liquid culture on Terasaki plates .................................... 104
  2.2.3.4. Fixing Cells by cytospin for MGG Staining ......................... 105

2.2.4. Flow Cytometric Analysis .................................................. 105

2.2.5. Generation of Knock-in hPSCs ............................................. 106
  2.2.5.1. Transfection of hPSCs by Nucleofection ............................ 106
  2.2.5.2. hPSC colony picking, replica plating and freezing ................ 107
  2.2.5.3. Thawing Candidate clones onto HS27 stroma ...................... 107
3. CHAPTER 3: In Vitro B cell Differentiation of Human Pluripotent Stem Cells
Recapitulates Early Embryonic Lymphopoiesis .................................................. 110
3.1. Introduction ........................................................................................................... 110
3.2. Results .................................................................................................................... 112
   3.2.1. Establishment of in vitro B cell Differentiation from hPSCs ......................... 112
   3.2.2. Molecular and Functional Characterisation of the hPSC-derived IL7R progenitor 117
   3.2.3. Transcriptional comparison of B lymphopoiesis in Human Fetal Liver and from hPSCs. .................................................................................................................. 122
3.3. Discussion ............................................................................................................. 132

4. CHAPTER 4: Generation and Validation of Genome Engineered ETV6-RUNX1 Human Pluripotent Stem Cells .................................................................................................................. 136
4.1. Introduction ............................................................................................................. 136
4.2. Results ..................................................................................................................... 139
   4.2.1. Choice of hPSC cell line .................................................................................. 139
   4.2.2. Validation of Southern Blot Screening Strategy ............................................... 140
   4.2.3. Construction of TALENs ................................................................................ 146
   4.2.4. CRISPR Design ............................................................................................. 149
   4.2.5. Construction of ETV6 Targeting Vector by BAC Recombineering ................ 150
   4.2.6. Transfection of hPSCs by Nucleofection ....................................................... 156
   4.2.7. Assessment of Homologous Recombination Efficiency .................................. 158
   4.2.8. Construction of RUNX1 Knock-in cassette ................................................... 159
   4.2.9. Generation of ETV6-RUNX1 knock-in clones .............................................. 165
   4.2.10. Removal of Knock-in Cassette by Cre Recombinase .................................... 171
4.3. Discussion .............................................................................................................. 172

5. CHAPTER 5: Characterisation of B lymphoid output from ETV6-RUNX1 hPSCs ...... 176
5.1. Introduction ............................................................................................................. 176
5.2. Results .................................................................................................................... 177
   5.2.1. Analysis of ETV6-RUNX1-hiPSC-derived B lymphopoiesis by flow cytometry 177
   5.2.2. Transcriptomic analysis of ETV6-RUNX1-hiPSC-derived B lymphopoiesis .. 182
   5.2.3. Comparison to RUNX1 floxed control hiPSCs .............................................. 188
5.3. Discussion .............................................................................................................. 190

6. CHAPTER 6: Final Remarks and Future Directions ................................................. 195

APPENDIX A: Analysis of B cell development in first trimester Human Fetal Liver ....... 201
REFERENCES .................................................................................................................. 206
TABLE OF FIGURES

Figure 1.1: Classical model of haematopoietic differentiation annotated with human surface markers......................................................... 6
Figure 1.2: Revised murine haematopoietic hierarchy .................................................. 8
Figure 1.3: Model of bistable transcription factor switch ............................................... 12
Figure 1.4: Major haematopoietic sites in the developing human embryo ....................... 17
Figure 1.5: Stages of B cell development .................................................................. 22
Figure 1.6: Transcriptional control of B cell commitment ............................................. 24
Figure 1.7: Incidence of first hit driver mutations in childhood and adult ALL .............. 34
Figure 1.8: Schematic representation of ETV6, RUNX1 and ETV6-RUNX1 fusion ........ 38
Table 1.1: Subsets of haematopoietic progenitors identified during differentiation of hPSCs on OP9 co-culture ........................................ 62
Table 2.1: Antibiotics used in molecular biology protocols ......................................... 74
Table 2.2: Primers used for analysis of D10 rearrangement in IL7R progenitor cells .... 90
Table 2.3: List of qPCR probes used for single cell analysis ....................................... 92
Table 2.4: Cytokines used in in vitro cultures ............................................................ 105
Table 2.5: List of FACS antibodies and reagents used on both hPSC and primary cells .... 106
Figure 3.1: hPSC in vitro B cell differentiation protocol ............................................... 113
Figure 3.2: Analysis of H1 hES cells differentiated on OP9 stroma to D10 for surface expression of the haematopoietic markers CD34 and CD43 .................. 114
Figure 3.3: Left: Differentiated MIFF3 hPSCs were analysed for CD19+ B cells (top) and CD19- IL7R+ progenitors (bottom) at D10 (left) and D31 (right). Right: In vitro image of CD34- MS5 co-cultures at D31 ................................................................. 115
Figure 3.4: H1 hESCs were analysed for CD19+ B cells (top) and IL7R+ progenitor (bottom) at D10 (left) and D31 (right) of differentiation .................................................. 116
Figure 3.5: qPCR of D31 MIFF3 hPSC co-cultures ...................................................... 118
Figure 3.6: B cell potential of the IL7R+ progenitor on MS5 stroma ......................... 119
Figure 3.7: Left: Myeloid potential of the IL7R+ progenitor in semi-solid media .......... 120
Figure 3.8: Single cell qPCR analysis of IL7R+ progenitors from D10 and D31 of differentiation and proB cells from D31 (MIFF3) ................................................ 121
Figure 3.9: Single cell qPCR analysis of IL7R+ progenitor and proB cells from H1 hESCs at D31 of differentiation ......................................................... 122
Figure 3.10: PCA from single cell qPCR data of IL7R+KIT+ progenitor (triangles) and proB cells (squares) from FL (CS17 (progenitor only) and CS20), CB and adult BM ............... 123
Figure 3.11: The direction and magnitude of eigenvectors contributing to separation for lineage-associated (left) and most differentially expressed (right) genes respectively 124
Figure 3.12: PCA calculated from single cell qPCR data from the human primary cells and differentiated hPSCs (MIFF3 and H1) ................................................. 125
Figure 3.13: Agilent Bioanalyzer RNA 6000 pico chip of RNA prepped from replicates of 200 and 500 cell equivalents of TRizol® compared to a dilution of the bulk RNA equivalent to 520 cells ................................................................................ 126
Table 3.1: Optimisation of SMARTer HV cDNA preparation showing cDNA yields and size distribution when titrating amplification cycles to input RNA concentration .......... 128
Figure 3.14: Agilent Bioanalyser high sensitivity DNA chip of cDNA produced from duplicates of 100pg RNA using SMARTer v3.0 cDNA kit varying the number of amplification cycles ................................................................. 128

Table 3.2: Summary of yields and size ranges of cDNA products assayed by Agilent Bioanalyser high sensitivity DNA chip of cDNA produced from duplicates of 100pg RNA using SMARTer v3.0 cDNA kit varying the number of amplification cycles 15, 16, or 17. .................................................................................................................. 129

Figure 3.15: Number of RNASeq reads mapped over the length of each transcript for a representative RNASeq sample showing low levels of 3’ bias ........................................ 130

Figure 3.16: PCA of RNASeq analysis of adult BM, FL and hPSC calculated using primary cell data only .................................................................................................................................. 131

Figure 3.17: PCA of RNASeq analysis of adult BM, FL and hPSC calculated using all data .... 132

Figure 4.1: Genome engineering strategy to generate constitutive RUNX1 knocked into ETV6 locus .......................................................................................................................... 138

Figure 4.2: Overview of Southern blotting strategy used to identify targeting events at the 5’ end of ETV6 intron V ...................................................................................... 141

Figure 4.3: Southern blot screening strategies for ETV6 intron V ........................................ 142

Figure 4.4: Left: Discrepancy in restriction fragment length when probing MIFF3 gDNA with D2 probe compared to the fragment predicted from the human reference genome. Right: Mapping of restriction digest patterns of ETV6 using different sources of DNA 143

Figure 4.5: Mapping of the 2:12026416:12027153 esv223109 structural variant .................. 144

Figure 4.6: Long range PCRs of the ETV6 intron V structural variant ................................. 145

Figure 4.7: In silico TALEN design .................................................................................... 146

Figure 4.8: Summary of TALEN construction by Golden Gate assembly ......................... 147

Figure 4.9: Representative examples of quality control of TALENs constructed by Golden Gate cloning .................................................................................................................. 148

Figure 4.10: Location of ETV6 CRISPR1 gRNA seed sequence and protospacer adjacent motif (PAM) compared to TALEN sites ........................................................................... 149

Figure 4.11: Overview of ETV6 targeting vector construction using BAC recombineering.... 151

Figure 4.12: A test of homologous recombination in hPSCs was performed using a simple NeoR selection cassette ......................................................................................... 152

Figure 4.13: Construction of the PGK NeoR targeting test vector ........................................ 153

Figure 4.14: Check of non-clonal DNA of intermediate vectors produced during recombineering ....................................................................................................................... 155

Figure 4.15: Four clones of p15a_ETV6_NeoR checked by restriction digest ....................... 156

Figure 4.16: Optimisation of nucleofection conditions for MIFF1 and MIFF3 hPSCs ......... 157

Figure 4.17: Summary of homologous recombination frequencies using the p15aETV6NeoR knock-in cassette into MIFF1 and MIFF3 hPSCs ............................................ 158

Figure 4.18: Construction of the RUNX1 knock-in cassette ............................................ 160

Figure 4.19: Ligation of the RUNX1_NeoR cassette into pR6K plasmid .............................. 161

Figure 4.20: Five pR6K_RUNX1 clones triple digested ..................................................... 162

Figure 4.21: Recombineering RUNX1 cassette into p15a_ETV6 vector plasmid............... 163

Figure 4.22: BglII digest of 4 clones of p15a_ETV6_RUNX1 .............................................. 164

Figure 4.23: Final checks of p15a_ETV6_RUNX1 vector .................................................. 165

Figure 4.24: Screening Southern blot of MIFF3 hPSCs nucleofected with linearised p15a_ETV6_RUNX1 targeted using CRISPR Cas9 and Cas9D10A ........................................ 166

Figure 4.25: Confirmatory Southern blot of successfully targeted ETV6-RUNX1 hPSCs targeted using CRISPR Cas9D10A .......................................................... 167

Figure 4.26: Confirmatory Southern blot using internal probe on successfully targeted ETV6-RUNX1 hPSCs targeted using CRISPR Cas9D10A ............................................. 167
Figure 4.27: Quantitative gene-expression analysis of unmodified MIFF3 and 4 successful ETV6-RUNX1 knock-in clones for the pluripotency genes OCT3/4 and NANOG ..........168
Figure 4.28: Immunostaining for nuclear OCT4 on fixed hPSC colonies .........................................................168
Figure 4.29: Flow cytometry analysis of the pluripotency surface markers TRA1-60 and SSEA3 on knock-in cells grown in StemFit media .................................................................169
Figure 4.30: Karyotype of four middle passage ETV6-RUNX1 knock-in hPSC clones grown in mTeSR1 ........................................................................................................................................170
Figure 4.31: Southern blot of Cre reverted ETV6-RUNX1 knock-in hPSC clone 2.8 showing two successfully reverted sub-clones ..........................................................................................................................172
Figure 5.1: Representative flow cytometric analysis of control MIFF3 (top panel) and ETV6-RUNX1 hPSC clone 2.8 (bottom panel) differentiated in vitro and analysed for Venus reporter (left), proB (CD34⁺CD19⁺) and preB (CD34⁻CD19⁺) cells (middle) and CD19 negative IL7R⁺ progenitors (right) ....................................................................................................................177
Figure 5.2: Frequencies of IL7R⁺ progenitor, proB and preB cells in MIFF3 cells compared to ETV6-RUNX1 iPSCs analysed at D31 .............................................................................................................................178
Figure 5.3: Occasional ETV6-RUNX1 iPSC clones with Venus positive (red) and Venus negative (blue) cells identified in the same differentiation culture ..........................................................................................................179
Figure 5.4: qPCR of ETV6-RUNX1 in MIFF3 (grey) compared to Venus⁺ (green) and Venus⁻ (brown) ETV6-RUNX1 IPS-derived cells ........................................................................................................................................180
Figure 5.5: qPCR of ETV6 (TEL) in MIFF3 (grey) compared to Venus⁺ (green) and Venus⁻ (brown) ETV6-RUNX1 IPS-derived cells ........................................................................................................................................181
Figure 5.6: RNASeq data from ETV6-RUNX1-expressing hIPSCs ........................................................................183
Figure 5.7: PCA of single cell qPCR from ETV6-RUNX1 hPSCs ........................................................................184
Figure 5.8: Single cell qPCR data of ETV6-RUNX1⁺ IL7R⁺ progenitors and proB cells from D31 ..............................................185
Figure 5.9: Gene-set enrichment analysis of ETV6-RUNX1 hiPS (red) and control MIFF3 (blue)-derived IL7R⁺ progenitors (top) and proB cells (bottom) at D31 of differentiation .................................................186
Figure 5.10: RNASeq data from LIN⁻ CD34⁺CD45RA⁻, IL7R⁺ progenitor and proB cells derived from FL, adult BM, hPSC and ETV6-RUNX1 hPSCs plotted on a 3D scatter plot, where gene sets for myeloid (log), lymphoid (log) or HSC (linear) programs are plotted on the x, y and z-axis respectively ..................................................................................................................187
Figure 5.11: Flow cytometry analysis at D31 of culture shows re-emergence of Venus negative proB and preB cells in the reverted clone (D6) ...............................................................................................................188
Figure 5.12: Single cell qPCR analysis of proB cells (CD34⁺CD19⁺) from ETV6-RUNX1 reverted clone ........................................................................................................................................................189
Figure 5.13: Co-expression of lineage associated genes in proB cells based on single cell qPCR data from MIFF3, ETV6-RUNX1 knock in and reverted hPSCs ................................................................................190
Figure 6.1: MGEG stained cytospin of macrophage produced from ETV6-RUNX1 expressing proB cells .................................................................197
Figure A1.1: CD19⁺ B cells from human fetal livers (FL) at CS17 and CS20, cord blood (CB) and adult bone marrow (BM) were analysed for surface expression of IL7R ..........201
Figure A1.2: Kinetics of IL7R progenitor emergence during development as percentage of total CD34⁺CD45⁺ cells ..........................................................................................................................202
Figure A1.3: Single cell qPCR analysis of IL7R⁺Kit⁺ progenitors from different developmental stages ........................................................................................................................................203
Figure A1.4: Co-expression of lineage-associated genes in IL7R⁺Kit⁺ progenitors based on single cell qPCR ..................................................................................................................204
ABBREVIATIONS

2-ME 2-mercaptoethanol
ABL Abelson murine leukaemia viral oncogene homologue 1
AGM Aorto-gonado-mesonephros
AHP Angiohaematopoietic progenitors
AID Activation induced deaminase
ALL Acute lymphoblastic leukaemia
AML Acute myeloid leukaemia
AML1 Acute myeloid leukaemia protein 1
AmpR Ampicillin resistance
APLNR Apelin receptor
Ara Arabinose
BAC Bacterial artificial chromosome
BCR Breakpoint cluster region
BLNK B cell linker
BM Bone marrow
BMP Bone morphogenic protein
BTK Bruton’s tyrosine kinase
CB Cord blood
CBF Core binding factor
cDNA Copy DNA
CDR Complementarity determining region
CFU Colony forming unit
ChIP Chromatin immunoprecipitation
CLP Common lymphoid progenitor
CMMP Chronic myelo-monocytic leukaemia
CMP Common myeloid progenitor
CNA Copy number alteration
CNS Central nervous system
CRISPR Clustered regularly interspaced short palindromic repeats
CRLF2 Cytokine receptor like factor 2
CS Carnegie stage
CSC Cancer stem cell
CSF Colony stimulating factor
dHSC Definitive Haematopoietic Stem Cell
DMEM Dulbecco’s modified Eagle’s medium
DMSO Dimethyl sulphoxide
DNA Deoxyribonucleic acid
Dnmt1 DNA methyl transferase 1
DNTT DNA nucleotidylexotransferase
DSB Double strand break
EBF-1 Early B cell factor 1
EDTA Ethylenediaminetetraacetic acid
EGF Epidermal Growth Factor
EPO Erythropoietin
ERG ETS-related gene
ETS E-twenty six
ETV6 ETS translocation variant 6
FACS Fluorescence-activated cell sorting
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>FCS</td>
<td>Fetal calf serum</td>
</tr>
<tr>
<td>FGF</td>
<td>Fibroblast growth factor</td>
</tr>
<tr>
<td>FISH</td>
<td>Fluorescence in situ hybridisation</td>
</tr>
<tr>
<td>FL</td>
<td>Fetal Liver</td>
</tr>
<tr>
<td>FLEX</td>
<td>Flip excision</td>
</tr>
<tr>
<td>Flk1</td>
<td>Fetal liver kinase 1</td>
</tr>
<tr>
<td>FLT-3</td>
<td>Fms-like tyrosine kinase 3</td>
</tr>
<tr>
<td>GAPDH</td>
<td>Glyceraldehyde-3-phosphate dehydrogenase</td>
</tr>
<tr>
<td>gDNA</td>
<td>Genomic DNA</td>
</tr>
<tr>
<td>GentR</td>
<td>Gentamicin resistance</td>
</tr>
<tr>
<td>GM</td>
<td>Granulocyte Monocyte</td>
</tr>
<tr>
<td>GMP</td>
<td>Granulocyte monocyte progenitor</td>
</tr>
<tr>
<td>gRNA</td>
<td>Guide RNA</td>
</tr>
<tr>
<td>GSEA</td>
<td>Gene set enrichment analysis</td>
</tr>
<tr>
<td>GSK</td>
<td>Glycogen synthase kinase</td>
</tr>
<tr>
<td>HA</td>
<td>Homology Arm</td>
</tr>
<tr>
<td>HB</td>
<td>Haemangioblast</td>
</tr>
<tr>
<td>HE</td>
<td>Haemogenic endothelium</td>
</tr>
<tr>
<td>Hh</td>
<td>Hedgehog</td>
</tr>
<tr>
<td>HSC</td>
<td>Haematopoietic stem cell</td>
</tr>
<tr>
<td>HDAC</td>
<td>Histone deacetylase</td>
</tr>
<tr>
<td>hESCs</td>
<td>Human embryonic stem cells</td>
</tr>
<tr>
<td>hiPSCs</td>
<td>Human induced pluripotent stem cells</td>
</tr>
<tr>
<td>HLH</td>
<td>Helix loop helix</td>
</tr>
<tr>
<td>HOX</td>
<td>Homeobox</td>
</tr>
<tr>
<td>hPSCs</td>
<td>Human pluripotent stem cells</td>
</tr>
<tr>
<td>HVMP</td>
<td>Haematovascular mesodermal precursors</td>
</tr>
<tr>
<td>IFN</td>
<td>Interferon</td>
</tr>
<tr>
<td>IgH</td>
<td>Immunoglobulin heavy chain</td>
</tr>
<tr>
<td>IKZF1</td>
<td>Ikaros</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>IL7R</td>
<td>Interleukin 7 receptor</td>
</tr>
<tr>
<td>IVF</td>
<td>In vitro fertilisation</td>
</tr>
<tr>
<td>JAK</td>
<td>Janus kinase</td>
</tr>
<tr>
<td>LIC</td>
<td>Leukaemia initiating cell</td>
</tr>
<tr>
<td>LIF</td>
<td>Leukaemia inhibitory factor</td>
</tr>
<tr>
<td>LIN</td>
<td>Lineage</td>
</tr>
<tr>
<td>LMO2</td>
<td>LIM domain only 2</td>
</tr>
<tr>
<td>LMPP</td>
<td>Lymphoid primed multipotent progenitor</td>
</tr>
<tr>
<td>LT-HSC</td>
<td>Long term repopulating haematopoietic stem cell</td>
</tr>
<tr>
<td>MACS</td>
<td>Magnetic cell sorting</td>
</tr>
<tr>
<td>MEF</td>
<td>Murine embryonic fibroblast</td>
</tr>
<tr>
<td>MEP</td>
<td>Megakaryocytic erythroid progenitor</td>
</tr>
<tr>
<td>MGG</td>
<td>May Grunwald Giemsa</td>
</tr>
<tr>
<td>MHP</td>
<td>Multipotent haematopoietic progenitors</td>
</tr>
<tr>
<td>MIFF</td>
<td>mRNA induced foreskin fibroblast</td>
</tr>
<tr>
<td>MKE</td>
<td>Megakaryocytic erythroid</td>
</tr>
<tr>
<td>MLL</td>
<td>Mixed lineage leukaemia</td>
</tr>
<tr>
<td>MPO</td>
<td>Myeloperoxidase</td>
</tr>
<tr>
<td>MLP</td>
<td>Multipotent lymphoid progenitor</td>
</tr>
<tr>
<td>MPP</td>
<td>Multipotent progenitor</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>MTG</td>
<td>Monothioglycerol</td>
</tr>
<tr>
<td>mTOR</td>
<td>Mammalian target of rapamycin</td>
</tr>
<tr>
<td>NeoR</td>
<td>Neomycin (G418) resistance</td>
</tr>
<tr>
<td>NK</td>
<td>Natural killer</td>
</tr>
<tr>
<td>NSG</td>
<td>NOD SCID gamma mice</td>
</tr>
<tr>
<td>PAM</td>
<td>Protospacer adjacent motif</td>
</tr>
<tr>
<td>PAX5</td>
<td>Paired box 5</td>
</tr>
<tr>
<td>PBS</td>
<td>Phosphate buffered saline</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>PD</td>
<td>Pointed domain</td>
</tr>
<tr>
<td>PDGFR</td>
<td>Platelet derived growth factor receptor</td>
</tr>
<tr>
<td>PGK</td>
<td>Phosphoglycerate kinase promoter</td>
</tr>
<tr>
<td>Ph-like</td>
<td>Philadelphia like</td>
</tr>
<tr>
<td>PI3K</td>
<td>Phosphatidylinositol-3-kinase</td>
</tr>
<tr>
<td>PNA</td>
<td>Peptide nucleic acid</td>
</tr>
<tr>
<td>preBCR</td>
<td>Pre B cell receptor</td>
</tr>
<tr>
<td>pSp</td>
<td>Para-aortic-splanchnopleura</td>
</tr>
<tr>
<td>PU1</td>
<td>Purine box factor 1</td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>Quantitative reverse transcriptase polymerase chain reaction</td>
</tr>
<tr>
<td>RAG</td>
<td>Recombination activating gene</td>
</tr>
<tr>
<td>RIN</td>
<td>RNA integrity number</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RNASeq</td>
<td>RNA sequencing</td>
</tr>
<tr>
<td>ROCK</td>
<td>Rho kinase</td>
</tr>
<tr>
<td>RpsL</td>
<td>Streptomycin resistance gene</td>
</tr>
<tr>
<td>RUNX1</td>
<td>Runt-related transcription factor 1</td>
</tr>
<tr>
<td>SB</td>
<td>Sleeping beauty</td>
</tr>
<tr>
<td>SCF</td>
<td>Stem cell factor</td>
</tr>
<tr>
<td>SCID</td>
<td>Severe combined immunodeficiency</td>
</tr>
<tr>
<td>SCNT</td>
<td>Somatic cell nuclear transfer</td>
</tr>
<tr>
<td>shRNA</td>
<td>Short hairpin RNA</td>
</tr>
<tr>
<td>SLAM</td>
<td>Signaling lymphocyte activation molecule</td>
</tr>
<tr>
<td>SMART</td>
<td>Switching mechanism at 5’ end of RNA template</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>SPIB</td>
<td>SpiB transcription factor</td>
</tr>
<tr>
<td>SSR</td>
<td>Site specific recombinase</td>
</tr>
<tr>
<td>STAT</td>
<td>Signal transducer and activator of transcription</td>
</tr>
<tr>
<td>TALEN</td>
<td>Transcription activator like effector nuclease</td>
</tr>
<tr>
<td>TdT</td>
<td>Terminal deoxynucleotidyl transferase</td>
</tr>
<tr>
<td>TEL</td>
<td>Translocation Ets leukaemia</td>
</tr>
<tr>
<td>TetR</td>
<td>Tetracycline resistance</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>TGF</td>
<td>Transforming growth factor</td>
</tr>
<tr>
<td>TK</td>
<td>Tyrosine kinase</td>
</tr>
<tr>
<td>TPO</td>
<td>Thrombopoietin</td>
</tr>
<tr>
<td>tracrRNA</td>
<td>transactivating RNA</td>
</tr>
<tr>
<td>TSLP</td>
<td>Thymic stromal lymphopoietin</td>
</tr>
<tr>
<td>VEGF</td>
<td>Vascular endothelial growth factor</td>
</tr>
<tr>
<td>VPREB1</td>
<td>Immunoglobulin iota chain</td>
</tr>
<tr>
<td>YS</td>
<td>Yolk sac</td>
</tr>
<tr>
<td>ZFN</td>
<td>Zinc finger nuclease</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

My first thanks go to Prof Tariq Enver for suggesting such a challenging and exciting project, whilst giving me the time and space to develop as a scientist. I would like to express my gratitude to Prof David Linch for his help in establishing this project and crucial assistance with funding. I am also grateful to my secondary supervisor Prof Claudio Stern for useful advice at key stages of the project.

A number of people from the UCL Cancer Institute have provided critical technical support during this project. I would particularly like to thank Charlotta Boiers for agreeing to a hugely productive collaboration on the characterisation of hPSC-derived B cells and for her technical help with FACS sorting and single cell qPCR. My thanks go to John Brown for his support and rigorous training in molecular biology techniques, without which I would not have been able to generate my knock-in constructs. A special thanks goes to Virginia Turati for her help with FACS sorts and assistance with tissue culture, in particular when I was abroad at a key point in the project. Thanks to Javier Herrero and Dapeng Wang for their assistance with the bioinformatic analysis, to Yanping Guo for help with NSG transplantation experiments and to Dr Gill May for advice in establishing the low cell number RNASeq protocol. I am also very grateful to Gill May, Rachael Nimmo and Jason Wray for their critical comments of this thesis and other manuscripts. I would like to thank the whole lab, including Duncan, Eli, Emma, Fiona Florence and Sara for their encouragement and great company.

Outside of the UCL cancer institute I would like to thank Prof Peter Andrews and his group for his help in setting up hPSC cultures and providing MIFF hiPSC lines; Dr Lee Carpenter and Anna French for support and advice on establishing in vitro B cell differentiation; and
Andrew Smith for his advice on genome engineering. I would like to thank the NIHR for funding my academic clinical fellowship and the Wellcome Trust, both for my research training fellowship and also for selecting me to attend two residential courses at the Hinxton genome campus.

Lastly, I would like to thank my wonderful wife May for her endless patience and understanding and my beautiful children, for always giving me perspective.
CHAPTER 1

Introduction
1. **CHAPTER 1: Introduction**

1.1. **Haematopoiesis**

Blood consists of more than ten different effector cell types with diverse functions and lifespans. The major classes of mature blood cell consist of haemoglobin-containing red blood cells (erythrocytes) responsible for the carriage of oxygen and carbon dioxide, platelets that play a role in haemostasis and white blood cells (leukocytes) that form the basis of the immune system. These leukocytes are themselves sub-characterised into those of the myeloid lineage (chiefly granulocytes and monocyte/macrophages) that function as mediators of the innate immune response and lymphoid cells (predominantly B, T and natural killer (NK) cells) that effect the adaptive immune response. The majority of these cells are short-lived, requiring the ongoing replenishment of over $10^{12}$ haematopoietic cells per day (Rieger and Schroeder, 2012). Remarkably, despite the enormous demands in terms of numbers and phenotypes, the entire haematopoietic system is produced and maintained from the progeny of a small number of haematopoietic stem cells (HSCs) located in the bone marrow (BM) (Doulatov et al., 2012).

1.1.1. **Haematopoietic Stem Cells**

Early histopathological studies of the BM identified the coexistence of immature blood cells of multiple lineages, prompting the concept of the HSC by Maximov in 1909 (Maximow, 1909). Work on proving the existence of HSCs accelerated during the Cold War, when research focused on understanding the haematopoietic effects of radiation exposure. In 1951 Lorenz demonstrated that spleen or marrow donation could regenerate the haematopoietic system of irradiated recipients (Lorenz et al., 1951). *In vivo* evidence for the
existence of clonal multipotent HSCs was later provided by Till and McCulloch (Till and McCulloch, 1961)

The defining characteristics of HSCs are an indefinite capacity for both self-renewal and differentiation into all the blood lineages. This functional definition means that HSCs can only be truly identified operationally by a repopulation assay. HSCs are rare cells (approximately 1 in a million BM cells in human (Wang et al., 1997a)) therefore surface markers are routinely used to enrich populations of cells for their presence (Rieger and Schroeder, 2012). Nevertheless such populations remain heterogeneous and ideally analysis should be performed at the single cell or clonal level.

Much of the characterisation of HSCs has been performed in mouse where the cell surface phenotype of HSCs has been highly refined to exist in the Lin⁻Sca1⁻Kit⁺ (LSK) CD34⁻ compartment (Ikuta and Weissman, 1992; Osawa et al., 1996; Spangrude et al., 1988) or alternatively using SLAM markers (LSK CD150⁻CD48⁺CD41⁻) (Kiel et al., 2005). It should be noted that although mouse and human share many similarities, significant haematopoietic differences exist reflecting amongst other things disparities in the lifespan, ecology, size and the balance of blood cell lineage outputs. Thus, markers used for the enrichment of mouse HSCs do not directly translate to human. Most notably human HSCs have generally been considered to be CD34⁺, as demonstrated by successful use of this marker in enriching HSCs for clinical haematopoietic stem cell transplantation (Civin et al., 1984; Kang et al., 2008; Vogel et al., 2000), although evidence has emerged that suggest that more primitive human HSCs may indeed be CD34⁻ (Anjos-Afonso et al., 2013). A series of other markers can refine this human HSC-containing population further including Flt3⁺ (Sitnicka et al., 2003), CD150⁻ (Larochelle et al., 2011), CD90⁻ (Thy1) (Baum et al., 1992), 38 45RA⁻ (Bhatia et al., 1997; Conneally et al., 1997; Lansdorp et al., 1990) culminating in the Lin⁻Rho^low^CD34⁺38
Thy1+45RA CD49f− phenotype, which can enrich human HSCs to 15% purity (Notta et al., 2011).

HSCs have been shown to reside in specific niches within the BM. Bone morphogenic protein (BMP) mutant mice have increased numbers of both HSCs and osteoblasts (Calvi et al., 2003; Zhang et al., 2003) suggesting that osteoblasts contribute to the BM niche for HSCs. Consistent with this intravital imaging has shown HSCs preferentially reside in the periosteal niche of calvarial marrow (Sipkins et al., 2005). In addition to the periosteal niche, HSCs have been shown to migrate to a perivascular endothelial niche, mediated by the chemokine CXCL12 (Kiel et al., 2005). Vertebrates use a diverse range of haematopoietic sites including BM in mammals and birds, kidney in fish and remarkably the frog Rana temporaria switches haematopoietic sites between liver and BM depending on season (Maslova and Tavrovskaia, 1993), thus, whether different niches directly control HSC function is uncertain.

Despite the enormous regenerative demands on the haematopoietic system, HSCs appear predominantly quiescent. Estimates based on telomere length (Shepherd et al., 2004) and X inactivation ratios (Catlin et al., 2011) indicate that human HSCs divide only every 175-350 days. Evidence suggests that HSCs may be maintained in a quiescent state by production of the cytokine thrombopoietin (TPO) from neighbouring osteoblasts (Yoshihara et al., 2007) or by contact with megakaryocytes (Bruns et al., 2014). Such quiescence is believed to protect HSCs from mutagenesis induced by cell division or oxidative stress. Replicative potential is therefore thought to exist in the downstream progenitor populations that lack long-term self-renewal capacity (Sun et al., 2014).
1.1.2. Haematopoietic Differentiation Hierarchy

HSCs produce mature effector blood cells through a series of differentiating progenitor states. As opposed to stem cells, progenitors are defined as cells that retain the capacity to differentiate, but that have lost the capacity for indefinite self-renewal. Classical models organise progenitors into a hierarchy whereby during differentiation they first lose self-renewal before undergoing lineage commitment in a stepwise manner. Studies of mouse haematopoiesis have sought to use surface markers to identify pure progenitor ‘compartments,’ but as with HSCs the lineage fates of progenitors can only be truly understood when examined at the single cell level and recent advances in single cell analysis have challenged some of the fundamental assumptions of these models (Nimmo et al., 2015).

The ‘classical’ human haematopoietic hierarchy is developed from that identified in mouse (Figure 1.1). The first differentiating HSC progeny lose self-renewal capacity and are labeled short term HSCs (ST-HSCs) or multipotent progenitors (MPPs) and in human are thought to be identifiable by loss of Thy1 (Majeti et al., 2007) and CD49f (Notta et al., 2011). Having lost self-renewal capacity the classical model then categorises progenitors as undergoing a series of binary lineage fate decisions. The first of these separates a common myeloid progenitor (CMP) producing erythrocytes, platelets, granulocytes (including neutrophils, eosinophils, mast cells and basophils), macrophages and dendritic cells, from a common lymphoid progenitor (CLP) giving rise to B, T and natural killer (NK) cells. The CMP further differentiates into either a granulocyte monocyte progenitor (GMP) or megakaryocyte erythroid progenitor (MEP). Using the surface markers CD123 (IL3Ra), Flt3 (CD135) and CD45RA these populations can be purified: CMP (LinCD34−CD38−CD135−CD45RA+), GMP (Lin−
CD34⁺CD38⁺CD45RA⁺CD135⁺), MEP (Lin⁻CD34⁺CD38⁻CD45RA⁻CD135⁺) (Doulatov et al., 2010; Manz et al., 2002).

Figure 1.1: Classical model of haematopoietic differentiation annotated with human surface markers. HSC, haematopoietic stem cell; MPP, multipotent progenitor; CMP, common myeloid progenitor; CLP, common lymphoid progenitor; MEP, megakaryocytic-erythroid progenitor; GMP, granulocytic monocytic progenitor. (Adapted from Nimmo et al., 2015)

This stepwise, bifurcating model has been challenged, principally by the detailed study of how lymphoid potential is acquired, and by the molecular and functional analysis of progenitors at the single cell level. A lymphoid-primed multipotent progenitor (LMPP) was identified in mouse by the surface phenotype LSKCD34⁺Flt3⁺ (Adolfsson et al., 2005) (Figure 1.2). Contrary to the classical model, this progenitor population has lymphoid bias, myeloid
potential, but negligible MKE capacity. However, in vivo fate mapping using Rag1, IL7R and Flt3 suggests that adult LMPPs contributed almost exclusively to lymphopoiesis raising the important caveat that lineage capacity assayed by either in vitro assays or transplantation does not necessarily mirror lineage commitment under physiological conditions (Boyer et al., 2011; Schlenner et al., 2010; Welner et al., 2009). A major caveat to these lineage-tracing studies, however, is that none of these fate-mapping genes precisely mark the entire LMPP population, for example Flt3 is expressed on multipotent progenitors. Other groups have also described multipotent lymphoid progenitors (MLPs) in human (Doulatov et al., 2010; Laurenti et al., 2013). Single cell gene expression analysis of LMPPs showed that these are the earliest progenitors to show lymphoid priming and that more primitive HSCs and MPPs showed evidence of myeloid gene priming consistent with the earlier phylogenetic origins of the myeloid series (Ciau-Uitz et al., 2014; Mansson et al., 2007). A mechanistic explanation for this is suggested by experiments showing that loss of the epigenetic modifier Dnmt1 impairs B and T cell output due to aberrant activation of myeloid genes (Broske et al., 2009); therefore lymphoid commitment appears not well accounted for by a defined binary lineage choice, but rather the steady silencing of intrinsic myeloid priming.
**Figure 1.2: Revised murine haematopoietic hierarchy taking into account of lymphomyeloid-restricted progenitors (lymphoid primed multipotent progenitors) and MkE biased HSCs. (Adapted from Nimmo et al., 2015)**

Further evidence against the classical model of haematopoiesis has come from the prospective isolation of platelet biased LT-HSCs (Sanjuan-Pla et al., 2013), single cell transcriptomic analysis showing that 50% of the LMPP compartment is already dendritic cell biased (a lineage normally ascribed to have myeloid origins) (Naik et al., 2013) and from single cell clonal assays demonstrating that in adult the human CMP population consists of predominantly unipotent progenitors (Notta et al., 2016). Overall these data indicate that: i) at the single cell level lineage commitment seems to happen much earlier than previously thought; ii) that HSCs have an intrinsic myeloid bias and iii) that the lymphoid system develops progressively from myeloid progenitors whilst being mutually exclusive with MkE potential. Looking ahead novel single cell technologies such as single cell RNASeq and
multiplex mass cytometry coupled with powerful new informatics tools (Qiu et al., 2011) look set to further redefine the classical model of haematopoiesis (Nimmo et al., 2015).

1.1.3. Molecular Basis of Haematopoiesis

The flow of cells through the haematopoietic hierarchy is tightly regulated whilst remaining highly responsive to the requirements of the organism. Waddington proposed a conceptual landscape model of how stem and progenitor cells differentiate in 1957 (Waddington, 1957). In this model a cell (represented as a ball), rolls down a hill (representing differentiation), along ‘canals’ or ‘attractors’ that bifurcate into different lineages. Commitment to a lineage is the point at which a cell crosses a gene expression threshold, becoming irreversibly restricted to one cell fate and losing the potential to differentiate into other cell types. Such a landscape can be considered a three dimensional representation of the global gene expression signature of the different cell states and is defined from the combined output or ‘solution’ of a network of epigenetic modifiers (Enver et al., 2009). Metastable intermediate progenitor states sit in shallow attractor basins, poised to differentiate further or make lineage commitment decisions. The programming of haematopoietic transcriptional networks is such that they will only resolve to a more differentiated cell state providing the directionality of the system, represented by ‘gravity’ or ‘entropy’. Thus under physiological conditions a mature cell will not re-program itself back up the slope into a stem or progenitor cell.

Waddington’s model provides a conceptual framework of how gene regulatory networks might operate at the global level, but does not provide insight into what forces act on stem and progenitor cells to leave their attractor basins and enter differentiation and/or commit to a particular lineage. At the population level haematopoietic differentiation is responsive
to extrinsic signals such as cytokines and other niche signals. Two potential mechanisms could lead to a change in the balance of lineage outputs in response to external stimuli; in the instructive model multipotent progenitor cells are driven to commit to a lineage in response to the signal, whereas in the selective model pre-committed progenitor cells expand or preferentially survive (Enver et al., 1998). Cytokine signaling is characterised by pleiotropic and redundant effects complicating loss of function or stimulation studies. Evidence for instructive signaling has come from continuous imaging of the response of single GMPs to the cytokines MCSF or granulocyte colony stimulating factor (GCSF) (Rieger et al., 2009). Although demonstrating that lineage instruction can occur, it remains unclear as to whether this is the predominant mechanism of lineage commitment in vivo (Enver and Jacobsen, 2009).

Intrinsic to the cell, gene expression is controlled by transcription factors (TFs). These DNA binding proteins recognise sequence specific DNA motifs resulting in the recruitment of complexes of epigenetic modifiers to specific combinations of genetic loci. These complexes result in the activation or repression of target genes either local to DNA binding sites (cis-acting) or more distantly (trans-acting). The overall effect is dynamic and context dependent resulting in a complex interconnected gene regulatory network.

Analysis of TF expression in populations of lineage-committed cells has identified a host of TFs that act as ‘master regulators’ of committed cell states (e.g. GATA1 for MkE, C/EBPα in GMPs, PAX5 in B cells). However, earlier parts of the haematopoietic hierarchy are characterised by heterogeneous low-level expression of multiple lineage-associated genes, termed ‘multilineage priming’ (Hu et al., 1997). This priming is thought to represent the underlying presence of open chromatin at specific loci associated with lineage commitment. Whether this low-level gene expression represents functional or leaky transcription, is
translated into protein expression or represents just a snapshot of complex cyclical or
dynamic gene expression is uncertain. The mechanisms underlying how incoherent lineage
priming is resolved into a committed uni-lineage gene expression programme remain to be
fully resolved. A single cell study of erythroid/myeloid lineage commitment in vitro showed
that erythroid commitment occurs when cells co-express a more coherent early erythroid
gene expression signature, but that this happens in the absence of master regulators such as
GATA1 that effect terminal erythroid differentiation (Pina et al., 2012). Therefore the
process of lineage commitment may start with a global shift in gene expression, which
induces the self-maintaining up-regulation of lineage specific master regulators.

The complexity of TF circuitry means that a single TF can often be associated with
commitment decisions to different lineages (e.g. PU1 is involved with commitment to both
monocytes and B lineages (Carotta et al., 2010)). How expression of the same protein can
result in commitment to different lineages can be explained by variations in TF expression
level (e.g. PU1^hi directs to monocyte whereas PU1^lo directs to B cells (DeKoter and Singh,
2000)) or context specific competition / cooperation with other TFs (Stampfel et al., 2015).
This complexity provides a mechanism by which stochastic changes or ‘noise’ in lineage-
affiliated gene expression levels may result in crossing the threshold to lineage commitment
in a cell autonomous manner (Enver et al., 2009).

Once committed, expression of master-regulators is usually associated with stable
expression of lineage associated genes and repression of alternative lineage genes. Such
behavior can be explained by a simple ‘bistable switch’ model, whereby expression of a TF
auto-regulates itself in a positive feed-forward loop, whilst cross-antagonising master
regulators of alternative cell fates. This is classically demonstrated by the GATA1-PU1 switch
regulating lineage commitment between megakaryocyte/erythroid and myeloid lineages
respectively (Figure 1.3) (Galloway et al., 2005; Rhodes et al., 2005) and has proven a recurrent design feature of haematopoietic lineage commitment networks (Graf and Enver, 2009). The power of TFs to instruct and maintain cell fate decisions is exemplified by their recurrent mutation in leukaemia and by their ability to re-program committed cells to a different lineage upon forced expression (Orkin and Zon, 2008).

![Figure 1.3: Model of bistable transcription factor switch. Left: Transcription factors A and B control commitment to cell states A and B respectively, auto-regulating their own expression whilst inhibiting each other (left). Right: When expressed at low levels, this results in a metastable bipotent progenitor state (a/b) that only commits to a particular lineage attractor basin (A or B) when one TF gains dominance over the other. (Adapted from Enver et al., 2009)](image)

1.1.4. Fetal haematopoiesis

Haematopoiesis in the fetus differs to that in adult reflecting both the different physiological demands of in utero life, and the ontogenic and phylogenetic origins of the haematopoietic system. Despite considerable research, significant aspects of how fetal haematopoiesis is structured remain controversial (Medvinsky et al., 2011). These controversies arise from a number of complexities of blood development: sites of haematopoiesis are separated spatially and developmentally; circulation moves stem cells and progenitors between sites
confounding identification of sites of origin and residence; and blood stem cells are rare and lack specific surface markers to separate them from the tissues from which they arise. Hence there are ongoing controversies over where HSCs are generated, the nature of the cell from which they arise and the functionality of the cells produced.

1.1.4.1. ‘Definitive’ Haematopoiesis

Embryonic blood development is unusual in that blood cells arise before stem cells during ontogeny; namely early yolk sac (YS) haematopoiesis occurs prior to the emergence of definitive HSCs (dHSCs) as defined by engraftment capacity.

The term definitive was first used by Dieterlen-Lievre in 1975 to describe the results of quail chick chimera experiments that showed the intra-embryonic origins of the adult blood system; “Haematopoietic stem cells of the definitive blood cell series originate from a source other than YS, and that source must be intra-embryonic” (Dieterlen-Lievre, 1975). Definitive HSCs are defined operationally as any cells that can provide long-term haematopoietic reconstitution upon transplantation and cannot be prospectively identified by specific surface markers. ‘preHSCs’ may become dHSCs upon further maturation in an embryonic environment such as newborn liver or in utero transplantation (Yoder et al., 1997b). More recently the term definitive has come to mean any fetal or pluripotent stem cell-derived cell/progenitor that can give rise to myeloid and adult-like enucleated erythroid cells, but this does not imply derivation from dHSC (Kennedy et al., 2012).

Two models can account for the specification of dHSCs after emergence of primitive blood cells i) dual origin with independent embryonic/primitive and definitive hierarchies and ii) common origin where primitive haematopoietic cells such as preHSCs develop ‘stemness’
and mature into dHSCs (Medvinsky et al., 2011). Testing these models requires an understanding of the sites of haematopoiesis in the developing embryo.

### 1.1.4.2. Sites of Fetal Haematopoiesis

Haematopoiesis is present at multiple sites during development, both intra-embryonic and extra-embryonic. At day 7 of mouse embryonic development (E7) extra-embryonic mesoderm ingresses through the posterior primitive streak into the YS resulting in the transient production of primitive nucleated red blood cells expressing embryonic haemoglobin. By E8.5 some colony forming unit-spleen (CFU-S) myeloid cells are identifiable along with neonatal-repopulating cKit+CD34+ ‘preHSCs’ at both the YS and aorto- gonado-mesonephros (AGM) region (Ferkowicz et al., 2003; Palis et al., 1999; Yoder et al., 1997a). Tissue resident macrophages (e.g. brain microglia, liver Kupffer cells and Langerhans cells) develop at E8.5-9.5 independently of dHSCs (Schulz et al., 2012).

YS cells are thought to arise from a common blood and vascular progenitor called the haemangioblast (HB). Sabin initially postulated such a bipotent cell in 1920 based on observations of the close physical association of primitive erythrocytes and endothelium in YS ‘blood islands’. Such a bipotent cell has been observed during the differentiation of murine embryonic stem cells (ESCs) (Choi et al., 1998), however, there is a lack of direct evidence of a single cell asymmetrically dividing into blood and endothelium in vivo. Equivalent cells have been found in the posterior primitive streak of the mouse embryo where they also have the capacity to form smooth muscle, however, by the time these cells migrate to the YS they have committed to either blood or endothelium (Huber et al., 2004).
The first long-term repopulating dHSCs appear at the mouse AGM region at E10.5 as tight clusters of haematopoietic cells adherent to the ventral wall of the dorsal aorta, vitelline and umbilical arteries (Medvinsky and Dzierzak, 1996). Evidence from lineage-tracing studies has shown that these cells have endothelial origins (Jaffredo et al., 1998). The phenotype of these cells in mouse is CD45⁺Sca1⁺c-Kit⁺CD34⁺VE-cadherin⁺ and they express the key transcription factors RUNX1, SCL and GATA2 (Dzierzak and Speck, 2008). Apart from CD45 these surface markers are shared with endothelial cells. AGM derived cells are thought to arise directly from endothelial cells termed ‘haemogenic endothelium’ (HE), in contrast to a bipotent haemangioblast. Compared to YS haematopoiesis, AGM derived dHSCs express the transcription factor RUNX1, which appears to mediate an endothelial to haematopoietic transition (North et al., 1999; North et al., 2002). Hematopoietic differentiation directly from endothelial cells has been tracked in live in vitro cultures of ES cells (Eilken et al., 2009) and in vivo live imaging of slices of the mouse AGM (Boisset et al., 2010).

The ontogenic origin of these AGM HE and dHSCs has been highly controversial. Some evidence supports their origins from YS. YS has the capacity to support dHSCs; E9-10 YS has a CD34⁺Kit⁺ preHSC that can mature to dHSC upon transplantation into newborn mice (Yoder et al., 1997b) and by E11.5 YS has its own dHSC population (Cumano et al., 2001). However, the fetal heart begins to beat at E8 and circulation is established by E10 precluding any conclusions regarding the origins of dHSCs based on their presence at a particular site after this time. Strong support for separate YS and AGM hierarchies in amphibians has come from fate mapping in pre-gastrula Xenopus embryo. Fluorescent dye injected into the blastomeres of 32-cell embryo show that different blastomeres contribute to primitive and definitive HSC production (Ciau-Uitz et al., 2000). Similarly in mouse, RUNX1-CreERT2 fate mapping showed that tamoxifen treatment at E7.5 marked haematopoietic cells in the FL, suggesting that the
YS contributes to definitive blood in vivo, although this could relate to the imprecise timing of tamoxifen activity.

Alternatively, in support of an AGM origin for dHSCs, E8 explant cultures from the para-aortic splanchnopleura (pSp – the area that contributes to AGM formation) can generate long term lympho-myeloid output, whereas YS at that time can only produce short term myeloid output (Cumano et al., 2001). Slc8a1/- mice lack a heartbeat preventing circulation of dHSCs until embryonic death. In these mice at E8.5 all myeloid cells reside in the YS with none found at the pSp (Lux et al., 2008). One explanation for this result, however, is that circulation itself is required for dHSC generation at the AGM. In support of this dHSCs only arise in the arterial endothelium and their production can be inhibited by a lack of shear-related nitric oxide signaling at the AGM (Adamo et al., 2009). One hypothesis that tries to unify the origins of YS haemangioblasts and AGM haemogeneic endothelium suggests that early Flk1+ haemangioblasts generate the haemogenic endothelium in a RUNX1-independent manner (Lancrin et al., 2009).

Once dHSCs arise at the AGM at E10.5, dHSCs are seen in arrange of other sites. Placental cells expressing RUNX1 with dHSC potential appear concurrently with dHSCs at the AGM, but the pool declines by E15.5 suggesting they migrate to the fetal liver (FL) (Ottersbach and Dzierzak, 2005). Evidence suggests that the placenta is a site of both de novo dHSC production as well as seeding from the AGM (Rhodes et al., 2008). RUNX1+ clusters and dHSCs are also present in the umbilical cord at E10.5 (de Bruijn et al., 2000; Medvinsky et al., 2011).

The predominant site of haematopoiesis during gestation is the FL. The FL rudiment is colonised from E9 onwards and this later seeds the spleen and thymus (Bertrand et al.,
Colonisation of the FL by dHSCs at E12.5 coincides with their appearance in circulation. It is thought that FL is a site of dHSC expansion mediated by Angiopoietin like factors and Sox17 (Kim et al., 2007; Zhang et al., 2006). The fetal BM is colonised from the FL, but takes only over later in development (Figure 1.4). Of note HSCs at different sites of haematopoiesis exhibit different functional properties, cell surface markers and developmental potentials and transplantation experiments in mice suggest some specific characteristics of FL are intrinsically regulated (Bowie et al., 2007). FL HSCs are predominantly in cycle compared to quiescent in BM. Of particular relevance, FL haematopoiesis exhibits more lineage plasticity than that in adult BM, with single cell clonal analysis of sorted human fractions indicating that progenitors from the CMP compartment are oligopotent, in contrast to adult BM where such cells were unipotent when assayed at the single cell level (Notta et al., 2016). Overall this suggests important functional differences between haematopoiesis in the FL and adult BM, the underlying programming of which has the potential to alter the characteristics of leukaemias that are initiated in utero.

*Figure 1.4: Major haematopoietic sites in the developing human embryo. Time scale shows days of gestation and embryonic development by Carnegie stage (CS). (Adapted from Tavian and Peault, 2005).*
1.1.4.3. Regulation of Haematopoietic Specification

That blood can form in different places at different times suggest that more than one pathway can coordinate the specification of haematopoietic tissues in the embryo. Such pathways combine external cues from niche-derived morphogens with a cell state defined by cell intrinsic TF circuits.

At the anatomical level both primitive and definitive blood form from mesoderm that has been in close contact with endoderm, suggesting that endoderm secretes morphogens that programme mesoderm to acquire haematopoietic potential. YS haematopoiesis requires exposure to the visceral endoderm, which can be substituted for by hedgehog (Hh) signaling (Dyer et al., 2001). AGM haematopoiesis is dependent on the ventro-dorsal patterning of the aorta. The dorsal wall is formed from somitic mesoderm, whereas the haematopoietic-proficient ventral aortic wall arises from lateral mesoderm. In chick, somitic mesoderm can be re-programmed to HE by exposure to ventralising factors (e.g. bFGF, VEGF, TGFβ, BMP4) and this in turn can be antagonised by dorsalising factors (e.g. TGFα, EGF) (Pardanaud and Dieterlen-Lievre, 1999). Hh signaling from the developing gut can also contribute to the development of AGM haematopoiesis in zebrafish (Gering and Patient, 2005; Peeters et al., 2009).

Notch1 signaling is a key pathway involved in both arterial identity and HSC production at the AGM (Kumano et al., 2003). Notch -/- chimeric mice do not contribute to dHSCs (Hadland et al., 2004), but notch deficiency has no effect on YS haematopoiesis. Alternatively arterialisation of veins by deletion of CoupTFII results in the emergence of haematopoietic clusters in the cardinal vein (You et al., 2005). VEGF signaling through its receptor Flk1 and its intermediate Hey2 is thought to contribute to Notch expression.
Notch signaling directly regulates GATA2 (Robert-Moreno et al., 2005), which in turn regulates RUNX1 (Nottingham et al., 2007). Consistent with this GATA2 and RUNX1 over-expression can overcome a lack of notch signaling (Nakagawa et al., 2006; Robert-Moreno et al., 2008). Together these pathways suggest a mechanism of haematopoietic specification at the AGM that is independent to that found in YS.

A number of cell intrinsic haematopoietic TFs are known to be essential for blood specification. SCL/Tal-1 and its partner LMO2 are individually essential for development of both primitive and definitive blood. ETV6 (TEL) is also essential for blood formation in the embryo and probably acts through VEGF to control SCL and GATA2 (Ciau-Uitz et al., 2010; Wang et al., 1997b). GATA2 is a key haematopoietic TF, whose knock out in mouse ablates AGM haematopoiesis and reduces YS erythropoiesis (Tsai et al., 1994). MLL and HOX genes are also essential for AGM haematopoiesis. Interestingly RUNX1 knock out has a highly selective effect on AGM haematopoiesis (Cai et al., 2000; Dzierzak and Speck, 2008; Okuda et al., 1996; Wang et al., 1996). RUNX1 (AML1) is expressed in the endothelial and sub-endothelial mesenchyme at the AGM suggesting that RUNX1 is a master regulator of an endothelial to haematopoietic transition in the dorsal aorta (North et al., 1999). VAV-Cre conditional deletion of RUNX1 in committed hematopoietic cells does not affect dHSC function or numbers, suggesting that RUNX1 has a role in the development, but not maintenance of HSCs.

How these morphogens and TFs act as a network in human in uncertain, but work in *Xenopus* haemangioblast has elucidated some design principles of such a network. Fli1/GATA2 initiates Flk1/Etv2 expression, which is self-maintaining due to positive feedback of Etv2 back onto GATA2. These factors then stimulate VEGFA/Flk1 signaling to activate Scl, which in turn activates a battery of haemangiogenic effector genes (Ciau-Uitz et al., 2013).
Further work has shown that morpholino knockdown of the microRNA mir-142-3p down-regulated all of these factors and this could partially reversed by inhibiting TGFβ. The authors suggest this mir142-3p sits at the apex of the regulatory network and acts as a double negative gate by inhibiting negative TGFβ signaling (Nimmo et al., 2013).

Understanding the mechanisms underlying haematopoietic specification \textit{in vivo} is highly relevant to the development of techniques to produce dHSCs from pluripotent stem cells for use in regenerative medicine and disease modeling.

1.1.5. B cell development

1.1.5.1. Stages of B cell Development

B cells are specialised cells of the adaptive immune system responsible for the production of immunoglobulin proteins that mediate the humoral immune response. The extraordinary diversity of immunoglobulin complementarity determining region (CDR) structures is generated through the controlled somatic mutation of the immunoglobulin heavy and light chain loci. This happens in two phases, first in the bone marrow and subsequently in the lymph node. In summary, during B cell development in the bone marrow the enzyme recombination-activating gene (RAG) mediates genomic recombination of genomically encoded segments (V, D and J) first at the IgH and at then the light chain loci. Additional diversity is added to the junctional regions by terminal deoxynucleotidyl transferase (TdT). If these rearrangements produce a structurally coherent antibody, naive B cells survive to leave the bone marrow. The majority of B cells enter the lymph nodes to become follicular B cells that require tonic signaling from their antibody-containing B cell receptor (BCR). A smaller subset of B cells become marginal B cells in spleen, which are characterised by T cell independence and a limited immunoglobulin repertoire that is biased to cell wall
constituents and senescent self components. On contact with antigen and co-stimulation from T cells, follicular B cells expand and begin further somatic mutation of the Ig locus by the enzyme activation-induced cytidine deaminase (AID). By this mechanism a polyclonal antibody repertoire of increasing affinity is generated and selected by T helper cells. Selected B cells then mature further, first by class switching antibody common chains before further differentiation either into long-lived memory B cells or antibody-secreting plasma cells.

Using cell surface markers Hardy has characterised murine BM B cell development into a number of distinct stages of maturation, which has some degree of overlaps with the stages of human B cell development (Hardy et al., 1991) (Figure 1.5). CLPs express RAG and, in the adult, DNTT (encoding TdT) resulting in the beginning of IgH rearrangement by recombination of $D_H$ and $J_H$ segments. The first committed pre-pro-B cells (Stage A) are characterised by expression of surface $B220^+$ in the mouse (Li et al., 1996). Although this is less well defined in human, Muschen has suggested that as in mouse pre-pro-B cells can be identified by the surface phenotype CD43+CD24loBP-1lo (Buchner et al., 2015), whereas others identify human preproB cells as CD19-CD10+ (Sanz et al., 2010). The next stage of development is the proB stage (Stage B/C), which in the human is marked by the emergence of CD19 expression in concert with CD34 and to some degree CD10 (Hystad et al., 2007).

Survival and development through both these stages of development is dependent on cytokine driven activation of the STAT5 pathway, predominantly by interleukin 7 (IL7). Upon successful rearrangement of $VD_J_H$ segments the cells lose CD34 expression, becoming preB cells and resulting in gain of surface CD20 expression. The preB cells then test the structural integrity of the IgH protein product by transcribing the Igμ heavy chain. This pairs with the surrogate light chain (consisting of λ5 and VpreB proteins) to make the preB cell receptor. Successful pairing results in a proliferative burst (large preB, stage C') followed by IL7
independent rearrangement of the κ (or if unsuccessful λ) light chain loci (small preB, stage D). If successful immature B cells are produced that can express surface IgM (Stage D/E).

This pathway in human has recently been confirmed and elaborated by mass cytometry identifying four key ‘coordination’ points: i) highly proliferative proB cells driven by ligand-dependent STAT5 signaling; ii) low-proliferating preB cells with ligand independent STAT5 signaling undergoing light chain rearrangement; iii) Kappa chain and CD20 expression; iv) naïve B cells (Bendall et al., 2014).

Figure 1.5: Stages of B cell development. (Adapted from Buchner et al., 2015)

1.1.5.2. Control of B cell Commitment and differentiation

Early lymphoid commitment is critically dependent on both Flt3 and stem cell factor (SCF) signaling. Mice lacking Kit, the receptor for SCF, or Flt3 ligand have severely depleted CLP and B lymphopoiesis (Sitnicka et al., 2002; Waskow et al., 2002). IL7 and possibly TSLP signaling are required for survival and proliferation of early B committed progenitors (see below).
A number of transcription factors co-ordinate and maintain B cell commitment and are frequent targets of mutation in B cell malignancies. Ikaros is a member of a family of zinc finger TFs that also includes Helios and Aiolos is expressed in all haematopoietic lineages. Ikaros regulation is complex with eight slice variants, some of which are dominant negatives, and an ability to form both homo and heterodimers. Knock out mice fail to develop the lymphoid lineage, with relative sparing of the myeloid and erythroid lineages indicating a role in early lymphoid commitment (Nichogiannopoulou et al., 1999). Another early lymphoid TF is PU1, encoded by the Spi1 gene. PU1 is expressed in all haematopoietic cells, except erythroblasts, megakaryocytes and T cells. Knockout mice are deficient in myeloid and lymphoid cells, notably macrophages, B cells, osteoclasts and mast cells (McKercher et al., 1996). In vitro retroviral transduction has demonstrated that expression level is highly important with low level expression associated with B cell differentiation, possibly by direct regulation IL7R expression (Peschon et al., 1994), whereas high-level expression promoted macrophage development (DeKoter and Singh, 2000).

B cell commitment is highly dependent on a circuit of TFs including E2A, EBF1 and PAX5 (Figure 1.6). E2A is a member of a group of E proteins containing a basic helix loop helix (bHLH) motif and operates as a homodimer in B cells. Knockout mice do not develop a lymphoid system and lack evidence of DJ rearrangement (Bain et al., 1994). Id family proteins that have a critical role in T cell development negatively regulate E2A. Early B cell Factor 1 (EBF1) is a homodimeric TF specifically expressed in lymphoid progenitors early during B cell commitment. Knockout mice have a few pre-proB cells and lack RAG expression or DJ rearrangement, but have normal E2A levels. EBF1 appears activated by E2A and has an E2A binding site at its promoter. Both E2A and EBF co-operate to activate the key B cell TF PAX5 (Bain et al., 1994), which in turn positively regulates EBF1 in a stable feed forward loop. PAX5 knock out mice develop pre-proB cells that can be maintained in IL7, but that can
switch lineage on cytokine stimulation or transplantation (Nutt et al., 1997). PAX5 expression represses more than 100 genes involved in alternative lineages (e.g. CSF1R, Flt3, Ccl3, Notch1) and activates 170 B cell affiliated genes (e.g. EBF, LEF1, TCF4, Aiolos, Id3, Bach2, IRF4 and IRF8) (Delogu et al., 2006). Conditional deletion of PAX5 in late follicular B cells reactivates alterative lineage programmes. PAX5 is required in B cells until terminal differentiation into plasma cells, which is mediated by the TF BLIMP1. These results suggest that ongoing PAX5 expression is a master regulator of B cell identity in committed B cells, but that in its absence B cells can activate alternative lineage programmes. The power of PAX5 to maintain B cell identity is exemplified by the requirement for PAX5 silencing during induced pluripotent stem cell (iPSC) reprogramming of B cells (Hanna et al., 2008).

![Diagram of transcriptional control of B cell commitment](image)

*Figure 1.6: Transcriptional control of B cell commitment. (Adapted from Rieger and Schroeder, 2012).*

1.1.5.3. Cytokine and PreB cell receptor signaling during B cell Differentiation

The generation of B cells from CLPs is dependent on STAT5 signaling mediated by IL7 and possibly TSLP. This requirement continues until successful recombination of the IgH locus
results in functional preB cell receptor signaling. IL7 signals by dimerising the IL7Rα receptor subunit with the common gamma chain receptor. IL7R signals via JAK 1 and 3 to activate STAT5a/b and also the PI3K pathway. This signaling serves to induce EBF1 and PAX5 expression (Hirokawa et al., 2003), facilitating recombination of the IgH loci (Corcoran et al., 1998) and perhaps most importantly provides a strong survival and proliferation signal. IL7R knock out mice completely lack B and T lymphocytes (Carvalho et al., 2001). However humans with IL7R mutations have an autosomal recessive severe combined immunodeficiency (SCID) disorder characterised by complete loss of T cells, but with normal B cell numbers (Puel et al., 1998). Normal B cell numbers in SCID cases have been attributed to compensatory expansion of B cells in the periphery, or the ability of fetal liver to make B cells in the absence of IL7. Patients are hypogammaglobulinemic, suggesting a functional B cell defect that may represent either an intrinsic B cell defect or a failure of co-stimulation by T cells.

Thymic Stromal Lymphopoietin (TSLP) is a cytokine produced by lung, skin, intestine and thymus with a role in the immune response to helminth infection and allergy. TSLP can also induce proliferation of human fetal liver MPPs to the B lineage (Scheeren et al., 2010). TSLP signals through the CRLF2 receptor, which dimerises with IL7Rα to signal via JAK1/2-STAT5 and PI3K pathways. TSLP therefore has strong similarities to IL7, but CRLF2 knockout mice have normal numbers of B and T cells making its physiological role in lymphoid development uncertain.

The formation of the preBCR is a key checkpoint in B cell development. The preBCR consists of the μ heavy chain paired to the surrogate light chain components λ5 and VpreB. Signaling can be activated either by binding to Galectin1 ligands in the BM (Gauthier et al., 2002) or intrinsically through the N terminal of λ5 (Ubelhart et al., 2010). The function of the preBCR
is threefold. Firstly, preBCR signaling results in allelic exclusion of the remaining IgH allele, repositioning it to centromeric chromatin (Roldan et al., 2005) and temporarily down-regulating RAG enzymes such that the cell only recombines one heavy chain (Grawunder et al., 1995). Secondly, the preBCR positively selects B cells expressing a functional heavy chain, both by providing a proliferative signal that takes over from IL7R signaling (Rolink et al., 2000), and by activating BCL6 (Duy et al., 2011), which counteracts PAX5 mediated BACH2 expression that would otherwise prime B cells for apoptosis (Swaminathan et al., 2013). Thirdly, there is evidence that mice lacking a preBCR have higher levels of autoantibodies, suggesting a role in negative selection (Keenan et al., 2008), although 75% of surviving preBCRs have cross reactivity to self-antigen (Wardemann et al., 2003). After selection, subsequent light chain rearrangement is mediated by the re-expression of RAG enzymes. This process is aided by release of IL7R-STAT5-EZH2 mediated repressive H3K27me3 marks on the immunoglobulin light chain locus and by gradual down-regulation of the surrogate light chain components by BLNK-mediated activation of IRF4, IRF8 and Ikaros.

1.1.5.4. Fetal B Cell Development

FL lymphocytes exhibit a number of differences to adult including lack of TDT expression and IL7R independence. These differences indicate that fetal haematopoiesis may be structured differently to adult, providing a unique cellular context for the in utero initiation of childhood leukaemia. Whether these differences relate to B cell differentiation within a fetal context or rather their origins in a separate wave of lymphopoiesis is uncertain.

In mouse, progenitors with myelo-lymphoid potential can be seen before the emergence of dHSCs (Kieusseian et al., 2012; Rybtsov et al., 2011; Yoshimoto et al., 2012) and RAG1 and IL7R expression has been identified at E10.5 suggesting the presence of lymphoid restricted
progenitors (Kawamoto et al., 2000; Yokota et al., 2006). Work by Boiers et al has identified a YS-derived wave of fetal lympho-myeloid restricted cells (Boiers et al., 2013). The authors first identified a significant population of IL7R⁺ lympho-myeloid progenitors in the E11.5 FL, prior to seeding by dHSCs. Examination of FL from common gamma chain or Flt3 knock out mice showed that these progenitors emerged independently of IL7 or Flt3 signaling. A RAG-GFP reporter identified RAG expressing cells in the E9.5 YS, but not at the pSp, characterised by surface expression of VECadherin, CD45 and CD41. RAG1-cre fate mapping showed that these progenitors contributed to both lymphoid and myeloid lineages, although not tissue resident macrophages. Overall this study supports a model whereby a core lymphoid programme including expression of IL7, Flt3 and RAG emerges in YS myeloid progenitors prior to dHSCs and that these cells contribute to both myeloid and lymphoid lineages in the fetus. Whether such a cell exists in human is uncertain.

One hypothesis is that fetal B cells contribute to a distinct class of B cells termed B1 cells. Identification of B cell malignancies aberrantly expressing CD5 in the 1980s led to the search for a physiological counterpart cell. The results of this work sub-categorised murine B cells into B1a (B220lo CD5⁺), B1b (B220lo CD5⁻ 11b⁺) and B2 B220hi CD5⁻. B2 cells differentiate to form the classical follicular and marginal B cells discussed above. B1 cells are characterised by a restricted antigen repertoire including affinity for self-antigens and repetitive epitopes such as carbohydrates, survive independently of IL7 and respond to TSLP (Montecino-Rodriguez et al., 2006). Functionally B1a cells may mediate defence to specific infections such a Streptococcus pneumoniae (Haas et al., 2005). Clonal studies have shown that single CLPs generate only B1 or B2 outputs (Haas et al., 2005) and that B1 cells are preferentially produced from transplantation of FL HSCs compared to adult BM (Hayakawa et al., 1985). B1 progenitors have been identified in both the YS and pSp leading Dorshkind to propose 3 waves of B1 cell formation (preHSC-derived wave, FL and adult BM) that reduce in
proportion through development (Montecino-Rodriguez and Dorshkind, 2012). FL haematopoiesis is characterised by the expression of the fetal genes Lin28/Lin28b and HSCs transduced with Lin28 preferentially produce B1a and MZ B cells, even in IL7R deficient backgrounds (Yuan et al., 2012). Arid3a has been recently identified as a likely key downstream regulator of Lin28b activity (Zhou et al., 2015). Whether B1 cells exist in human is controversial. Patients with Rheumatoid arthritis have expanded CD5+ B cells and these produce more rheumatoid factor, and human neonatal CB is enriched for CD5+ B cells. Griffin has proposed that the human equivalent is characterised by the surface phenotype CD20+27+43+70-CD11b+- (Griffin et al., 2011), but consensus as to whether such a cell exists in human has not been reached.

1.1.5.5. Lineage Plasticity and the Ontogeny of B cells

Cells can be forced to change fate by lineage diversion, direct transdifferentiation or dedifferentiation to a more primitive state. Understanding the mechanisms by which these intriguing phenomena occur has relevance to regenerative medicine technologies and understanding the natural history of cancer formation. Classical experiments showed that B cells transduced with MCSFR could be forced to transdifferentiate to macrophages after a change in media conditions, possibly related to redox-mediated down-regulation of PAX5 (Borzillo et al., 1990). More recently CEBP/A transduction of B cells coupled with application of myeloid cytokines transdifferentiated B cells to the myeloid lineage, again associated with down-regulation of PAX5 (Heavey et al., 2003; Xie et al., 2004). The importance of PAX5 in maintaining B lineage identity and the tendency of B cells to revert to a myeloid state on PAX5 inhibition is consistent with the myeloid-based model of differentiation whereby B or T lineage potential is acquired in myeloid progenitors (Kawamoto et al., 2010). Further supportive evidence comes from the finding that i) CLPs retain myeloid potential in vitro; ii)
the identification of My-B and My-T progenitors; and iii) that mixed lineage leukaemias usually co-express My-B or My-T antigens.

Perhaps most compellingly the myeloid-based model recapitulates the proposed phylogenetic origins of the adaptive immune system from the innate. B cells have similar basal functions to macrophages, namely antigen incorporation and presentation. Phylogenetically, jawless vertebrate lamprey eels segregate an erythroid and primitive innate immune system (Hagerstrand et al., 1999). The acquisition of a transposon containing RAG like enzymes allows the emergence of the adaptive immune system in subsequent gnathostome vertebrates (Agrawal et al., 1998). B cells in frogs and fish continue to have phagocytic activity (Li et al., 2006). Thus, the evolutionary design of B cell programming might underpin a tendency to de-differentiate revealing an underlying myeloid programme during leukaemogenesis. The greater lineage plasticity seen in fetal blood cells may further explain the tendency for childhood leukaemias to aberrantly express myeloid antigens or switch lineages (Abdelhaleem, 2007; Gerr et al., 2010).

1.2. Childhood Acute Lymphoblastic Leukaemia

The incidence of cancer in children is remarkably high, with a cumulative risk of 1:484 (Böiers, 2015). This high rate of malignancy is thought to result from the acquisition of genetic mutations that specifically affect developmental processes (Li et al., 2005). Acute lymphoblastic leukaemia (ALL) is a malignant proliferation of immature lymphoid progenitor cells that if untreated results in rapid bone marrow failure and death. Although relatively rare in adults, ALL is the commonest malignancy in children accounting for 35% of childhood cancers. Overall 50% of ALL cases present in children and despite cure rates of 80-90%, it remains the commonest cause of death before the age of 20 years old (Hunger and
Mullighan, 2015; Pui et al., 2011). The incidence of ALL peaks between 2-3 years of age, reaching a nadir at age 9, before then slowly increasing with age. ALL is 20% more common in boys. Despite considerable research no environmental risk factors have yet been proven to be associated with ALL. Similarly, the majority of cases have no known inherited predisposition, with the notable exceptions of Down’s syndrome and rare familial mutations in PAX5 and ETV6 (Shah et al., 2013; Zhang et al., 2015). However, some inherited polymorphic variants have been shown to confer increased risk in a minority of cases (e.g. ARID5B, CEBPE, GATA3, IKZF1) (Trevino et al., 2009).

ALL is characterised by the proliferation and expansion of immature lymphoid progenitors that efface the bone marrow resulting in death from the consequences of bone marrow failure. In addition leukaemic cells or ‘blasts’ can infiltrate other haematopoietic and non-haematopoietic organs including the spleen, liver, lymph nodes, CNS, eye and testis. The majority (85%) of cases express antigens consistent with B lineage commitment. 20% of cases aberrantly co-express myeloid antigens, which appears to be more prominent in childhood ALL compared to adult (Abdelhaleem, 2007; Gerr et al., 2010).

1.2.1. Evidence for the Prenatal Origins of Childhood ALL

Compelling evidence from multiple sources has shown that may cases of childhood ALL are initiated in utero (Greaves and Wiemels, 2003). The childhood ALL driven by the ETV6-RUNX1 fusion transcription factor has been particularly informative regarding the prenatal origins of paediatric ALL. ETV6-RUNX1 is formed by the t(12;21) translocation (Fears et al., 1996). Careful analysis by long range PCR has shown that although there is some element of micro-clustering, the fusion break-points are spread through ETV6 intron V and RUNX1
introns I and II (Wiemels et al., 2000). Thus, individual leukaemias contain a unique clonal breakpoint sequence that can be used as a molecular signature to track disease.

Retrospective screening of neonatal blood spots of children with ALL has identified the clonal fusion gene sequence to be present at birth (Hjalgrim et al., 2002; Maia et al., 2001; Wiemels et al., 1999a). Furthermore, banked cord bloods can be shown to contain ETV6-RUNX1 containing pre-leukaemic B cells, as visualised by fluorescence in situ hybridisation (FISH) (Mori et al., 2002). The most persuasive data, however, comes from monochorionic twins, which share a pre-natal blood system through placental vascular anastomoses.

Studies of twins concordant for ALL show their leukaemias share a clonal breakpoint marker, implying a single cell of origin arising in utero and whose progeny were shared between the twins prior to birth (Greaves et al., 2003). Furthermore, in a case where only one twin developed overt ETV6-RUNX1 ALL, a pre-leukaemic B cell clone sharing the same clonal IgH rearrangement was identifiable in the well twin indicating that the twins had shared a pre-leukaemic clone at birth (Hong et al., 2008).

These studies have also demonstrated the surprisingly high incidence of ETV6-RUNX1 pre-leukaemia, but with a low penetrance of subsequent transformation to overt ALL. The concordance rates for two monochorionic twins to be affected by ALL is only 5-10% (Greaves et al., 2003). Furthermore, the detection rate of ETV6-RUNX1 in screened neonatal cord blood (approximately 1%) is approximately one hundred fold greater than the population incidence of ETV6-RUNX1 ALL (1:10000) (Mori et al., 2002). Such observations indicate that ETV6-RUNX1 has relatively weak oncogenic transforming capacity. This is in concordance with the protracted postnatal latency of ETV6-RUNX1, which can be over a decade in some cases (Maia et al., 2004; Wiemels et al., 1999b). Thus, ETV6-RUNX1 is commonly acquired in
uteri and is likely to be an initiating event, but in isolation is insufficient to transform B cells to ALL requiring the acquisition of second hit oncogenic mutations.

1.2.2. First Hit Mutations in ALL

B lineage ALL is predominantly categorised according to the initiating mutation driving the disease. Leukaemias such as ALL are characterised by recurrent chromosomal abnormalities that have been amenable to evaluation by fluorescent in situ hybridization (FISH) for many decades. Two broad groups of chromosomal abnormality account for the majority of cases: aneuploidy (i.e. changes in the chromosome number) and chromosomal translocations. 20-30% of childhood ALL exhibit high hyperdiploidy (>50 chromosomes), which confers a very good prognosis. Other significant aneuploidies include hypodiploid ALL (<44 chromosomes, 2-3%), which confers a bad prognosis (Nachman et al., 2007), low hypodiploidy (30-39 chromosomes), which is associated with P53 or Li Fraumeni syndrome (Holmfeldt et al., 2013) and trisomy 21 (Down’s syndrome).

Chromosomal translocations usually result in dysregulation of a normal gene or the formation of a novel chimeric protein product. The commonest chromosomal translocation in childhood ALL is t(12;21) resulting in the formation of the ETV6-RUNX1 (TEL-AML1) fusion transcription factor. This mutation accounts for 25% of cases of childhood ALL making it the commonest single genetic mutation that initiates childhood ALL. Other fusion proteins include TCF3-PBX1, BCR-ABL1 (resulting from a t(9;22) translocation forming the Philadelphia chromosome) and a group of translocations involving the MLL gene on chromosome 11 with one of over 60 translocation partners. Translocations that dysregulate oncogene expression often results in genes being inserted into constitutively active loci such as the immunoglobulin or T cell receptor (TCR) loci (e.g. IGH-MYC/CRLF2/EpoR, TCR-TLX1/3).
In a significant minority of cases an initiating chromosomal abnormality cannot be detected. New genomic technologies are beginning to dissect this heterogeneous group identifying novel subgroups such as those with an internal amplification of chromosome 21 (iAMP21) that share characteristics with trisomy 21 ALL (Heerema et al., 2013) and recurrent mutations in ERG (Clappier et al., 2014). Furthermore, analysis of gene expression profiles has demonstrated the shared transcriptional features of leukaemias driven by seemingly disparate oncogenes. 15% of childhood ALL has a Philadelphia-like (Ph-ALL) phenotype, sharing a similar gene expression profile with that found in t(9;22) ALL and are characterised by recurrent abnormalities that activate receptor tyrosine kinase (TK) signaling (Roberts et al., 2014). Such TKs have been subcategorised into those of the ABL class (ABL1, ABL2, CSF1R (MCSFR/v-fms), PDGFRβ) and those signaling through the JAK/STAT pathway (IL7R, CRLF2 (TSLPR), EPOR).

Overall the incidence of these initiating mutations is significantly different between adult and childhood ALL, with ETV6-RUNX1 accounting for 25% of childhood ALL, yet <2% of adult ALL (Figure 1.7).
Figure 1.7: Incidence of first hit driver mutations in childhood (left) and adult (right) ALL.

(Adapted from Böiers, 2015)

1.2.3. Second Hit Mutations in ALL

The majority of these first hit mutations are insufficient in themselves to transform lymphoid progenitors into ALL. As such they must either arise in, or confer self-renewal potential to, long-lived HSCs or lymphoid progenitor cells generating a pre-leukaemic clone with acquisition of further co-operating mutations resulting in transformation to overt leukaemia. This is consistent with a minimal ‘two-hit’ hypothesis proposed by Knudson (Knudson, 1971). Gilliland and Speck extended this concept by observing that co-operating mutations in leukaemia are often drawn from two classes of oncogene: class I mutations that typically activate receptor TK pathways providing growth and survival signals; and class II mutations, often affecting transcription factors, which result in differentiation arrest (Speck and Gilliland, 2002).
Compared to adult, childhood ALL has relatively few secondary mutations at diagnosis (on average 10-20 non-silent coding mutations) with double the number at relapse (Ma et al., 2015; Papaemmanuil et al., 2014). This maybe due to the relative paucity of age related acquired ‘passenger’ mutations that do not contribute to leukaemic transformation. These second-hit mutations recurrently affect key pathways including the transcriptional regulation of lymphoid development/differentiation, cell cycle, the p53/RB pathway, growth factors (e.g. Ras, PI3K, JAK/STAT signaling) and, at relapse, chemotherapy resistance pathways such as mutations in nucleoside metabolism and epigenetic modifiers.

Loss of the B cell master regulator PAX5 is the commonest recurrent somatic mutation in B ALL, implying a tumour suppressor role (Mullighan et al., 2007). PAX5 mutations act in a class II manner arresting B cell differentiation, but the mutations are almost invariably haploinsufficient suggesting that complete loss of PAX5 might be inconsistent with a B ALL transcriptional programme.

Mutations in cytokine signaling are common in ALL including activation or overexpression of cytokine receptors (e.g. CRLF2, IL7R), activation of second messenger signaling (e.g. JAK, Lnk) and the generation of novel oncogenic TKs (e.g. BCR-ABL1). Ph-like ALLs are characterised by aberrant STAT5 signaling, including through non-lymphoid receptors such as those generated by IgH-EPOR, EBF1-PDGFRβ, ATF7IP-PDGFRβ translocations or Flt3+ expression (Roberts et al., 2012). Ph-like ALL is almost always associated with deletion of IKAROS suggesting a strong cooperative role (Mullighan et al., 2009). Muschen has proposed that STAT5 signaling is required in ALLs lacking a functional preBCR. According to this functional classification, the majority of ALLs (85%) are preBCR negative (including ETV6-RUNX1, BCR-ABL1, MLL and hyperdiploid and Ph-like ALL) (Geng et al., 2015; Trageser et al., 2009). Despite the absence of a preBCR, BCR-ABL1 ALLs can activate BTK signaling by expression of
a truncated splice variant of BTK that recruits full length BTK for transactivation by BCR-ABL1 kinase (Feldhahn et al., 2005). PreBCR signaling appears mutually exclusive to a Ph-like ALL gene expression signature, suggesting that the preBCR may act as a tumour suppressor in these cases (Trageser et al., 2009). This may be due to preBCR inhibition of cytokine receptor STAT5 signaling via BLNK or hyper-activation of preB cell receptor signaling resulting in apoptosis. Such observations demonstrate the degree to which the outcome of co-operation between first and second hit mutations is dependent on the differentiation stage of the cell in which they act. Such a model also predicts that in classes of preBCR negative ALL where class I activating mutations are relatively rare (e.g. ETV6-RUNX1), differentiation arrest might trap B cell progenitor compartments that constitutively activate signaling pathways that are permissive for leukaemic development.

1.2.4. Cancer Stem Cell Hypothesis

That leukaemia forms within differentiating haematopoietic hierarchies led John Dick to consider whether leukaemias themselves might be arranged into hierarchies of functionally distinct cells. Using cell surface makers of HSCs, he identified a sub-population of cells within acute myeloid leukaemia (AML) that were enriched for repopulating activity by transplantation and that could regenerate the bulk of the leukaemia (Bonnet and Dick, 1997). The major implication of this theory was that only a subset of tumour cells retains indefinite self-renewal potential (cancer stem cells, CSCs or leukaemia initiating cells, LICs), with the majority of the tumour being relatively short-lived bulk differentiated cells. If CSCs retained this capacity in patients, then they would act to both maintain the tumour and be the source of post-treatment relapse, implying that they were the key therapeutic targets for long-term cure.
The concept has proven highly controversial, not least because the development of more permissive immunosuppressed mice have increased the frequency of CSCs in some cancers to a level which challenges their relevance (Quintana et al., 2008). Furthermore the plasticity of immune phenotypes in leukaemia makes extrapolation of supposed hierarchies onto leukaemia difficult (le Viseur et al., 2008). What remains important from the CSC concept however is that tumours consist of functionally heterogeneous cells. This heterogeneity is increasingly seen to arise from multiple mechanisms, including underlying sub-clonal genetic variegation (Anderson et al., 2011), cell cycle status, in particular quiescence (Lutz et al., 2013) and the role of niche and microenvironment in modifying cell behavior (Duan et al., 2014). Nevertheless, in the absence of a complete block in haematopoietic differentiation in the first hit cell, it remains likely that remnants of normal haematopoietic differentiation hierarchies may persist, at least during the early stages of pre-leukaemic or leukaemic development, and that these compartments may have distinct functional properties (Hong et al., 2008).

### 1.2.5. ETV6-RUNX1 Acute Lymphoblastic Leukaemia

The chimeric transcription factor ETV6-RUNX1 (TEL-AML1) results from the translocation t(12;21) and accounts for 25% of childhood ALL, but <2% of adult ALL (Fears et al., 1996; Mullighan, 2012; Romana et al., 1995b). Evidence from both neonatal blood spots and monochorionic twins of ETV6-RUNX1 ALL cases, supported by deep sequencing have demonstrated that this mutation frequently occurs in utero and is unequivocally an initiating event (Greaves, 2003; Papaemmanuil et al., 2014). Cases are invariably of B lineage ALL, usually differentiated to the pre (CD34+CD19+) or proB cell (CD34+CD19+) phenotype. Despite a propensity for late relapse it is a cytogenetic subgroup that is associated with a good prognosis (Attarbaschi et al., 2004).
The t(12;21) translocation breakpoint can be anywhere along the 12kb length of ETV6 intron V into either intron I or II of RUNX1, fusing the N terminal of ETV6 (amino acids 1-336) to almost all of RUNX1 (amino acids 22-480) (Golub et al., 1995; Romana et al., 1995a; Romana et al., 1995b) (Figure 1.8). RUNX1 is one of a family of three TFs containing a DNA binding domain related to Drosophila Runt along with a transcriptional activation and repression domain. RUNX1 normally binds to DNA in a complex with core binding factor β (CBFβ) and is also named CBFα. RUNX1 is the most commonly recurrently mutated gene in acute leukaemia including multiple translocation partners as a first hit event, frequent deletions in acute myeloid leukaemia (AML) and recurrent amplification in ALL (Speck and Gilliland, 2002). In addition, RUNX1 is mutated in a rare familial platelet disorder resulting in predisposition to AML (Song et al., 1999). The precise regulation and expression of differentially spliced isoforms of RUNX1 is highly complex (Speck and Gilliland, 2002; Tsuzuki et al., 2007). As discussed above RUNX1 is critical for the development of blood from haemogenic endothelium, but is also expressed in most mature blood cells including B cells. Unlike constitutive knock-out mice, conditional knockout mice retain haematopoiesis, but exhibit impaired lymphopoiesis and increased multipotent progenitor numbers (Ichikawa et al., 2004). RUNX1 binds to the enhancers of genes involved with the pre-B cell transition suggesting a role in B cell differentiation (Niebuhr et al., 2013).
Figure 1.8: Schematic representation of ETV6, RUNX1 and ETV6-RUNX1 fusion. Arrows indicate fusion points in t(12;21). ETV6 contains an oligomerisation pointed domain (PD), repression domain and ETS family DNA binding domain. RUNX1 contains the DNA binding RUNT domain, an mSin3A interaction domain (SID), a region reported to interact with the p300 histone acetyl transferase (p300ID) and a transcriptional activation domain.

(Adapted from Zelent et al., 2004).

ETV6 is one of the ETS family of TFs containing an 88 amino acid DNA binding and protein-protein interaction domain (Wasylyk et al., 1993). In addition a helix loop helix (HLH) containing pointed domain (PD) at the N terminus mediates homo or hetero-dimerisation. ETV6 recruits multiple transcriptional repressors including SMRT, mSin3A, NCor and HDAC3. ETV6 was first identified through fusion to PDGFRβ in CMML (Golub et al., 1994) and is recurrently rearranged with multiple translocation partners in a number of haematological malignancies. Knockout mice die in gestation at E10.5 demonstrating vascular defects in the YS (Wang et al., 1997b). Chimeric mice generated from ETV6 knockout and wild-type ES cells show normal definitive haematopoiesis in the YS and fetal liver, but an absence of all haematopoietic lineages in the bone marrow (Wang et al., 1997b; Wang et al., 1998).

The precise molecular action of ETV6-RUNX1 remains uncertain. In contrast to other gene fusions involving RUNX1, ETV6-RUNX1 contains all the significant domains of RUNX1 including the C terminal regulatory domains. ETV6 loses its DNA binding ETS domain, contributing its repression and N terminal pointed domains. Thus, it is hypothesised that ETV6-RUNX1 will aberrantly repress RUNX1 target loci, but may also result in complex abnormal protein-protein interactions (Zelent et al., 2004). Consistent with this hypothesis studies have shown repression of haematopoietic reporters in the presence of ETV6-RUNX1.
(Fears et al., 1997; Hiebert et al., 1996; Uchida et al., 1999) although a significant minority of direct ETV6-RUNX1 binding targets are in fact activated (J Wray, personal communication).

A number of studies have suggested a concerted activity by ETV6-RUNX1 on key oncogenic or haematopoietic pathways. shRNA knock-down of ETV6-RUNX1 identified down-regulation of the PI3K-AKT-mTOR pathway and HSC signature genes and the PI3K/AKT/mTOR pathway was subsequently shown to be required for the survival of ETV6-RUNX1 cells (Fuka et al., 2012; Fuka et al., 2011). Another study has indicated a difference in response to cytokine signaling with ETV6-RUNX1 expressing B cells being less responsive to growth inhibition by TGFβ, which was postulated to be mediated via direct binding of ETV6-RUNX1 to the TGFβ signaling target Smad3 (Ford et al., 2009). Overall the molecular activity of ETV6-RUNX1 appears complex and locus specific, with activity being mediated both at the transcriptional and direct protein level. Understanding the specific functional impact of ETV6-RUNX1 on early B cell development is therefore key to identifying the most relevant target pathways.

1.2.5.1. Current Models of ETV6-RUNX1 ALL

Understanding the pre-leukaemic activity of ETV6-RUNX1 first requires establishing its expression as a sole genetic abnormality in an appropriate B cell progenitor. The availability of the pre-leukaemic initiated cells from primary patient samples is limiting. Although 1% of children are born with circulating ETV6-RUNX1 initiated cells, such cells are rare and there are few or no surface phenotypes that can enrich for the pre-leukaemic cells (Hong et al., 2008). Immortalised cell lines from ETV6-RUNX1 ALL cells such as REH and AT2 have been in use for many decades (Greaves and Janossy, 1978; Zhang et al., 1993), however, the presence of multiple mutations, the transforming process and decades in culture confound the study of the pre-leukaemic initiating activity of ETV6-RUNX1. Furthermore these
homogeneous cell lines do not allow study of the functional cellular compartments and hierarchies of the pre-leukaemic state. Therefore a number of ETV6-RUNX1 model systems exist and have provided useful insights into the contribution of ETV6-RUNX1 in the initiation and transformation of ALL.

At the simplest level, the DNA binding and transcriptional impact of ETV6-RUNX1 can be assayed by transgenically expressing ETV6-RUNX1 in B cell lines. A key method of examining its DNA binding targets is by chromatin immunoprecipitation (ChIP), but ETV6-RUNX1 exposes few novel epitopes for IP antibodies (Linka et al., 2013). The only model system suitable for ETV6-RUNX1 ChIP in the presence of wild type ETV6 is in the mouse lymphoid cell line BaF3 transgenically expressing mifepristone-inducible V5-tagged ETV6-RUNX1 (Ford et al., 2009). Although tractable, this system fails to model many of the important characteristics of ETV6-RUNX1 pre-leukaemia in human including: i) the effect of the in utero developmental context; ii) the impact of ETV6-RUNX1 during early B cell differentiation/commitment; and iii) failure to model the impact of haploinsufficiency of wild type ETV6 and RUNX1, which are key tumour suppressor and oncogeneic second hit target genes respectively.

Murine transplantation studies have provided inconsistent phenotypes that appear to depend both on the developmental and differentiation stage of the cell targeted. In vivo expression of ETV6-RUNX1 under the B cell IGH enhancer failed to cause leukaemia, most likely due to the requirement for expression in a developmentally more primitive compartment and the absence of co-operating second hits mutations (Andreasson et al., 2001). Syngenic transplantation of murine BM HSCs retrovirally transduced with ETV6-RUNX1 generated leukaemia in 2 of 9 mice (T-ALL and B-ALL) (Bernardin et al., 2002). When HSCs from mice deficient in p16/p19 were transfected 6/8 mice developed leukaemia, but of
unspecified lineages. To investigate the effect of ETV6-RUNX1 on B cell differentiation, Tsuzuki et al transplanted ETV6-RUNX1 transduced syngeneic BALB/C mouse HSCs and demonstrated a partial arrest of early B cell differentiation, resulting in expansion of the pro-B cell compartment and increased numbers of HSCs (Tsuzuki et al., 2004). Of note, there was also some increase in myeloid progenitor activity, but with no myeloid differentiation block. These findings correlated with the work of Fischer et al 2005 in C57BL/6J-Ly5.2 mice (Fischer et al., 2005). Together these studies indicate that ETV6-RUNX1 is acquired in more primitive stem or progenitor cells and results in a lineage selective partial differentiation block at the proB cell stage.

To model the effect of ETV6-RUNX1 on prenatal B cell development, Morrow et al used a model of forced ETV6-RUNX1 expression in murine fetal liver (FL) HSCs. In this context ETV6-RUNX1 expression resulted in the enhancement of B cell progenitor self-renewal capacity as measured by increased in vitro re-plating efficiency, colony forming assays and in vivo reconstitution (Morrow et al., 2004). In addition, this study also found a B cell differentiation arrest, but at the pre rather than pro-B cell stage. These findings indicate that the developmental stage in which ETV6-RUNX1 is acquired is also important.

Hong et al used a combination of primary patient material and lentiviral transduction of cord blood (CB) to study the hierarchical relationships between the immunophenotypic compartments in ETV6-RUNX1 pre-leukaemia (Hong et al., 2008). An aberrant CD34⁺CD38⁻CD19⁺ immunophenotype had previously been shown to select for leukaemia propagating ‘stem-B’ cells that were transplantable into NOD/SCID mice (Castor et al., 2005). Hong et al identified and fluorescence-activated cell sorting (FACS) sorted a population of immunophenotypic stem-B cells from the peripheral blood of a monochorionic twin of an ETV6-RUNX1 ALL patient. Such twins share a blood system prenatally due to placental
vascular anastomoses, thus the well twin carries a pre-leukaemic clone ancestrally related to that in the sibling’s ALL. As predicted, these rare stem-B cells shared the same partial \( IGH \) \( DJ \) rearrangement as the twin’s leukaemia and appeared to be hierarchically related to a more mature VDJ recombined and CD10⁺ bulk ALL cells. As the pre-leukaemic stem-B cells in the well twin were rare, human cord blood (CB) lentivirally transduced with ETV6-RUNX1 was xenotransplanted into NOD/SCID mice. This recapitulated the immune phenotype and \( DJ_{H} \) rearrangement status of the pre-leukaemic stem-B fraction and was demonstrated to have self-renewal by serial transplantation. Although subsequent transplantation studies in NSG mice have shown that compartments in addition to the ‘stem-B’ compartment can have repopulating capacity, this study highlights how pre-leukaemia itself may have its own differentiation hierarchies, and indicates that different compartments may have different functional properties.

De Laurentiis investigated the impact of ETV6-RUNX1 during B lineage commitment by lentivirally transducing the murine multipotent (erythroid / myeloid / lymphoid) EML1 cell line (de Laurentiis et al., 2015). In the presence of the B cell cytokines IL7 and FLT3L, transduced cells showed a block in B cell development resulting in high levels of cell death after 9 days. Transcriptional analysis identified the interferon (IFN) \( \alpha/\beta \) pathway to be down regulated by ETV6-RUNX1, potentially mediated by a reduction in mTOR-mediated activation of IRF3. Treatment of the cell line REH with IFNβ induced differentiation as judged by an increase in surface IgM expression.

To test whether the level of ETV6-RUNX1 expression affected the observed phenotype Tsuzuki et al used a retrovirus expressing ETV6-RUNX1 up- or down-stream of an internal ribosomal entry site (IRES) to provide high or low ETV6-RUNX1 expression respectively (Tsuzuki and Seto, 2013). The low-level virus gave expression levels one sixth lower than the
high producing ETV6-RUNX1 levels consistent those in the cell line REH when compared by western blot. High-level expression in murine FL HSCs inhibited in vitro differentiation of B cells and expanded the myeloid compartment. Conversely low-level expression resulted in higher B cell re-plating efficiency, higher competitive and serial engraftment and an incomplete B cell differentiation block at the pro-B cell stage. ETV6-RUNX1 transgenically expressed under the CD19 promoter gave similar results suggesting an action on an early B cell compartment. ETV6-RUNX1 co-operated with N-RAS mutation in pro-B cells to increase the B-ALL leukaemogenic properties of this mutation. Intriguingly gene set enrichment analysis (GSEA) suggested that ETV6-RUNX1 up-regulated embryonic stem cell gene sets.

A number of transgenic animal models have been developed to assess the effect of ETV6-RUNX1 in vivo. A transgenic zebrafish model of ETV6-RUNX1 under a ubiquitous promoter exhibited fatal hyperplasia of lymphoid blasts in 6% of cases, but the precise lineage of the blasts was not clear and whether the hyperplasia was cell autonomous was not tested (Sabaawy et al., 2006). A series of conditional mouse knock-in (KI) models indicated that ETV6-RUNX1 may operate at the level of primitive HSCs (Schindler et al., 2009). The mice were designed to conditionally knock-in the cDNA of RUNX1 at the endogenous ETV6 locus in response to lineage-restricted expression of cre recombinase. Expression in the early mouse embryo in response to GATA1-cre resulted in increased self-renewal of FL cells in vitro, which was lost in the newborn. Mx-cre KI in HSCs increased HSC and myeloid numbers, but resulted in absent B cell development, whereas expression in early lymphoid cells using CD2-cre showed no differences to wild type. Using ENU mutagenesis to induce transforming second hits in HSC or lymphoid cells expressing ETV6-RUNX1 resulted in T lineage malignancies.

A constitutive KI mouse model inserted the entire cDNA for human RUNX1 into ETV6 exon V.
along with a sleeping beauty (SB) transposase to facilitate the inducible introduction of random second-hit mutations (van der Weyden et al., 2011). When crossed with transposon array mice, the mice developed leukaemias, but predominantly of the T and myeloid lineage rather than B-ALL. A follow up study crossed this model with Pax5 heterozygous mice showing that although they continued to see myeloid and T cell leukaemias, the proportion of B ALL cases increased in combination with Pax5 heterozygosity. As anticipated the contribution of the Pax5 mutation was to enforce differentiation arrest at a preB cell stage, as opposed to the effect of back crossing with Ink4a deficient mice, which resulted in a shorter latency to leukaemia development. Recurrent SB mediated second hit mutations centred on B cell maturation pathways, including activating mutations of the JAK-STAT5 pathway (van der Weyden et al., 2015).

In summary, the existing models of ETV6-RUNX1 pre-leukaemia have produced some conflicting results, but a number of themes have emerged from multiple systems. The predominant phenotype observed is that of a block in B cell differentiation, with some studies showing enhanced self-renewal at the HSC or early B cell stage. Where reported the myeloid lineage appears more proliferative, but with no differentiation block. Why the addition of second hits to these models does not recapitulate the B lineage leukaemias seen in children remains unresolved, but may be due to the induction of non-relevant second hits by random mutagenesis approaches and/or the failure to target these mutations to the appropriate B lineage progenitor compartment impacted by ETV6-RUNX1. Given the lessons learnt from these models, an ideal system would express physiologically relevant levels of ETV6-RUNX1 in the human fetal B lymphoid progenitor targeted by ETV6-RUNX1 in utero.
1.2.5.2. **Cellular Origins of ETV6-RUNX1 Pre-leukaemia**

The models above indicate that the impact of ETV6-RUNX1 is highly dependent on the developmental and differentiation stage of the cell in which it is expressed. Therefore in order to model its mechanism of action requires a detailed understanding both of the ontogenic origin of the cell in which it is acquired and at what level of the differentiation hierarchy it targets.

ETV6-RUNX1 ALL is a paradigmatic disease for the study of the early leukaemic initiation and tumour evolution. Early studies demonstrated that ETV6-RUNX1 is unequivocally an initiating event and is frequently acquired in utero resulting in a covert pre-leukaemic clone that persists into childhood (Gale et al., 1997). The prenatal origins of ETV6-RUNX1 have been corroborated by several studies including the identification of pre-leukaemic clones in the monochorionic twins of ALL cases (Greaves, 2003), the retrospective identification of ETV6-RUNX1 in the neonatal heel pricks of ETV6-RUNX1 ALL cases (Greaves and Wiemels, 2003) and the identification of ETV6-RUNX1 in 1% of healthy newborns (Mori et al., 2002). Single cell targeted genomic sequencing of monochorionic twins has shown that the earliest shared clonal rearrangements involve both the IGH and TCR loci including DJ/ rearrangements (Alpar et al., 2015). Such findings are consistent with ETV6-RUNX1 targeting a fetal RAG-expressing lymphoid progenitor cell of both T and B lineage potential. These studies also demonstrate that ETV6-RUNX1 is insufficient to transform to ALL alone, correlating with the pre-leukaemic models described above. This is also consistent with the long latency to transformation to ALL (Wiemels et al., 1999b), the low concordance of ALL between monochorionic twins (10%) (Greaves, 2003) and 100 fold higher rate of ETV6-RUNX1 identifiable at birth compared to cases of ETV6-RUNX1 ALL (Mori et al., 2002).
The development of second hits is a crucial bottleneck in the development of ALL. Of all second hits, the commonest is loss or silencing of the remaining allele of ETV6 (Greaves and Wiemels, 2003; Mullighan et al., 2007) suggesting a role as a tumour suppressor (Fenrick et al., 2000) potentially by direct dimerisation to ETV6-RUNX1 (Lopez et al., 1999). In addition, multiple allelic gains of RUNX1 are highly recurrent further complicating how to produce a genetically precise pre-leukaemic model. Other genes identified by large-scale array and deep sequencing studies include PAX5, Ikaros, CDKN2A, BTG1, TBL1XR1, nr3c2, BTLA, RAG1, RAG2, ATF7IP and MGA (Bateman et al., 2010; Hunger and Mullighan, 2015; Papaemmanuil et al., 2014).

The sequence in which these second hit mutations are acquired in ETV6-RUNX1 ALL has proven to be highly informative as to how tumours develop genetic heterogeneity more generally. Traditionally it was predicted that mutations were acquired in a linear sequential manner with loss of minor clones at each stage of evolution. Studies of single cells using multiparametric FISH, however, have demonstrated a much more complex picture reminiscent of Darwinian branching polytypic evolution with the same genes being independently mutated in different sub-clones (Anderson et al., 2011). Importantly, recombination-activating genes (RAG) appear to mediate these recurrent mutations indicating that ETV6-RUNX1 targets a RAG-expressing lymphoid progenitor (Papaemmanuil et al., 2014).

RAG1 and 2 are recombinase enzymes involved in the somatic recombination of the IGH and TCR loci. The enzymes are targeted to VDJ sites by recombination signal sequences (RSS) consisting of highly conserved heptamer (CACAGTG) and less conserved nonamer (ACAAAAAACC) sequences separated by a 12-23bp spacer (Fugmann et al., 2000). In addition TDT (encoded by DNTT) adds non-template sequence (NTS) to the recombination ends,
further diversifying the immune repertoire. RAG enzymes can also recognise other sequences including non-RSS sequences, open chromatin, H3K4me3, non B DNA and deaminated methylCpGs, allowing off target recombinase activity. A large deep sequencing study of ETV6-RUNX1 ALL showed high levels of enrichment of acquired mutations at RSS sites associated with the addition of NTS compared to mutations in solid malignancies and hypodiploid ALL controls, suggesting a dominant role of RAG and TDT in the development of second hit mutations in ETV6-RUNX1 ALL (Papaemmanuil et al., 2014). Consistent with this ETV6-RUNX1 cases had relatively few point mutations (average 14, range 1-95). Of note the ETV6-RUNX1 translocation itself did not share these RAG motifs, suggesting that whereas the cell of impact is almost certainly a RAG expressing lymphoid progenitor, the cell in which ETV6-RUNX1 is acquired could be more primitive. RAG appeared to be preferentially directed to active genes through binding to H3K4me3 active promoter and enhancer sites. Although RAG was itself recurrently targeted, it was usually only heterozygously mutated. The authors also noted that compared to other malignancies there was a relative lack of activating kinase mutations.

The Muschen group investigated whether activation induced cytidine deaminase (AID), a further enzyme involved with somatic hypermutation of the IGH locus in the lymph node germinal centre reaction, was also contributing to second hit mutagenesis in ETV6-RUNX1 ALL (Swaminathan et al., 2015). The authors noted a close correlation of genes deleted in ALL with AID DNA binding sites identified by ChIP, although this contrasted with the failure to associate AID motifs with mutations in ETV6-RUNX1 ALL by deep sequencing (Papaemmanuil et al., 2014). Previous studies have shown that 70% of chromosomal translocations in B cells involve a break at a CpG dinucleotide (Tsai et al., 2008) and these authors suggested a mechanism where AID mediated deamination from pC to T results in a T:G mismatch 2500x more stable than a U:G mismatch forming a stable DNA bubble
amenable to cleavage by RAG. As such concomitant AID and RAG expression is potentially highly mutagenic at CpG sites, but importantly RAG and AID expression are normally strictly segregated, with RAG active in differentiating bone marrow lymphoid progenitors and AID active later in lymph node B cells. Muschen and colleagues continued to show that RAG and AID were in fact concurrently active in small preB cells and that IL7R signaling normally down-regulated AID activity prior to this stage of differentiation. The authors concluded that such a mechanism could explain the observed association between the transformation of ETV6-RUNX1 ALL and childhood infections.

Overall, the clear prenatal origins of ETV6-RUNX1 as a first hit and rarity of ETV6-RUNX1 ALL in adults indicate that ETV6-RUNX1 arises in a transient cell unique to early human fetal development. The identification of common ancestral clones containing partial DJ \( IGH \) or even TCR rearrangements in the monochorionic twins of ETV6-RUNX1 childhood ALL cases indicate that transformation from pre-leukaemia to leukaemia occurs within an early RAG-expressing multipotent lymphoid progenitor compartment that has either been trapped by differentiation arrest and/or expanded by ETV6-RUNX1 (Alpar et al., 2015). Ongoing RAG and potentially AID activity further genetically diversifies this transformed clone (Papaemmanuil et al., 2014; Swaminathan et al., 2015). The clinical features of aberrant co-expression of myeloid antigens and lineage switching commonly seen in many childhood B-ALLs indicate that the first hit mutation may arise in a target cell that might additionally harbor some myeloid as well as lymphoid lineage programming (Abdelhaleem, 2007; Gerr et al., 2010; Suggs et al., 2007). This is supported by the observation that fetal haematopoietic hierarchies are more lineage permissive than adult (Boiers et al., 2013; Notta et al., 2016). We hypothesised that the B lymphoid output from human pluripotent stem cells (hPSCs) may provide a source of fetal-like B or B/myeloid progenitor cells in which to explore the first hit function of ETV6-RUNX1.
1.3. Pluripotency

Stem cells are characterised by the ability to both indefinitely self-renew and to differentiate into more restricted progeny. Stem cells can be classified depending on the range of cell types they are capable of differentiating into. Totipotent cells are present in the developing zygote and blastomere and have the capacity to differentiate into all intra-embryonic and extra-embryonic tissues; however, their capacity for indefinite self-renewal has not been demonstrated. Pluripotent stem cells (PSCs) are capable of differentiating into all the tissues of the body, but with essentially no contribution to the extra-embryonic membranes or placenta. It has, however, been reported that ground state murine PSCs grown in 2i culture conditions (see below) can become totipotent (Morgani et al., 2013). Tissue resident stem cells such as HSCs are multipotent as they are restricted to generating mature cell types solely of their tissue of origin. In some cases stem cells can be unipotent, for example spermatogonial stem cells only give rise to a single cell type. Protocols are not established for the long-term maintenance and manipulation of HSCs in vitro, but such techniques do exist for hPSCs from which haematopoietic progenitor cells can be derived. Given their ability to generate cells, and to some degree tissues, from any of the three germ layers (endoderm, mesoderm and ectoderm) hPSCs are emerging as a powerful tool for regenerative medicine and disease modeling.

1.3.1. Murine Embryonic Stem Cells

Murine embryonic stem cells (ESCs) were first derived by explanting cells from blastocysts or inner cell masses onto a layer of mitotically inactivated fibroblast “feeder cells” in fetal calf serum (FCS)-containing medium (Evans and Kaufman, 1981; Martin, 1981). These culture conditions had been previously defined to maintain embryonal carcinoma cells (Martin and
Evans, 1975). Characteristics of pluripotent stem cells include unlimited growth in vitro, specific cell surface marker and transcription factor expression, the ability to form tissues from all three germ layers in a teratoma assay and most stringently the functional capacity to contribute to germline chimeras (Bradley et al., 1984; De Los Angeles et al., 2015).

Pluripotency relies on extrinsic stimuli from tissue culture conditions stabilising a metastable cell state that would not persist during normal development. Central to this cell state are the TFs Oct4 and Sox2, which function together as a heterodimer (Avilion et al., 2003). A further TF Nanog stabilises the PSC state and is required to generate the inner cell mass in vivo, although it is not strictly required for the maintenance of pluripotency in mouse ESCs and is only transcribed at low levels in mouse epiblast stem cells (see below) (Chambers et al., 2007; Silva et al., 2009). Overall, these TFs work co-operatively and to some extent redundantly, with Oct4 being the most stringently required for ongoing pluripotency. Together they auto-regulate their own expression (Boyer et al., 2005) whilst maintaining bivalent chromatin marks (H3K4me3 with H3K27me3) on the promoters of lineage regulating genes (Bernstein et al., 2006). Despite the expression of these core pluripotency factors mouse ES cells grown in standard conditions are functionally and molecularly heterogeneous and are termed ‘primed’.

### 1.3.1.1. Murine Epiblast Stem Cells

Pluripotent stem cells can be derived from different stages of mouse development, indicating that the property of pluripotency is present in a number of distinct cell states. After implantation the rodent embryo develops into an egg cylinder with a cup-shaped pluripotent epiblast. Attempts to derive ESCs from such post-implantation epiblasts were initially met with failure, but pluripotent cell lines have been derived and maintained in
distinct culture conditions (Brons et al., 2007; Tesar et al., 2007). These post-implantation derived pluripotent stem cells have been termed epiblast-derived stem cells (Epi-SCs). Although Epi-SCs appear functionally pluripotent and generate teratomas containing cells derived from all three germ layers, they exhibit significant differences to murine ESCs: they do not effectively contribute to chimeras; their gene expression profiles show that although they express Oct4, they have low or no expression of naïve pluripotency markers such as Rex1 and express higher levels of post-implantation factors such as Fgf5 and brachyury; their morphology is more two-dimensional; they are not maintained in standard murine ESC conditions, but require stimulation through the FGF/activin axis; they do not clone easily from single cells; and they exhibit striking heterogeneity in their differentiation biases (Bernemann et al., 2011). Whether this cellular heterogeneity is an artifact of culture conditions or reflects the lineage priming biases of the cells of the egg-cylinder is not known.

1.3.1.2. Ground State Pluripotency in the Mouse

Biases in lineage priming such as that seen in conventional mESCs and EpiSCs is a technical barrier to the directed differentiation of PSCs in defined conditions. This has led to the search for conditions in which ESCs can be maintained in a naïve or ground state, free from lineage bias. Rodents exhibit a reproductive behavior called diapause, whereby development of a pre-implantation fertilised blastocyst can be arrested until conditions for gestation are more favourable. It is likely that diapause has facilitated derivation of ESCs in rodents (Brook and Gardner, 1997; Evans and Kaufman, 1981) and is probably mediated by a relatively stable transcription factor circuit that defines a pluripotent state at the naïve epiblast stage. Work pioneered by the laboratory of Austin Smith has shown that such a primitive pluripotent state can be stabilised by modifying external signaling allowing refinements to murine ESC culture conditions. In summary, feeder cells can be replaced by
leukaemia inhibitory factor (LIF) (Koopman and Cotton, 1984; Smith et al., 1988; Smith and Hooper, 1987; Smith and Hooper, 1983; Williams et al., 1988). This itself can be replaced by the combination of a glycogen synthase kinase-3 (GSK3) inhibitor (CH99021) and the potent MEK inhibitor (PD0325901), labeled 2i conditions (Wray et al., 2011; Wray et al., 2010). Ground state pluripotency is characterised by X chromosome reactivation in female cells, low global DNA methylation, expression of a host of distinct naïve TFs (e.g. Klf4, Klf2, Esrrb, Tfcp2l1, Tbx3 and Gbx2) and by Oct4 expression driven from its distal, rather than proximal enhancer.

A perfect ground state PSC should completely lose epigenetic lineage priming, becoming a blank slate or ‘tabula rasa’, however, even in current ground state conditions naïve mES exhibit a variable expression of pluripotency TFs (Morgani et al., 2013). Nevertheless, these conditions maintain a more homogenous population of Nanog-expressing naïve ESCs which has facilitated important technical advances such as the derivation of ESCs from difficult murine strains (Batlle-Morera et al., 2008; Nichols et al., 2009) and even other rodents such as rat (Buehr et al., 2008).

As the technology to control PSCs has matured, more stringent functional assays of pluripotency have been proposed (De Los Angeles et al., 2015). Simple in vitro differentiation assays or in vivo teratoma formation can demonstrate capacity to differentiate into cells from the three germ layers. The ability to contribute to a developing embryo can be assayed by non-gestational chimeras, gestational complementation, tetraploid gestational complementation (whereby the tetraploid-derived cells contribute only the extra-embryonic tissues), single cell gestational complementation and ultimately gestation from a single input cell. However, all but the first of these chimeric assays are ethically unacceptable assays of human PSCs.
1.3.2. Human Embryonic Stem Cells

Human ESCs are significantly different to those of mouse, resulting in a considerable delay before their successful derivation. The first primate pluripotent cell line was derived in 1995 from Rhesus macaque blastocysts (Thomson et al., 1995). Human ESCs were subsequently obtained in 1998 using donated supernumerary embryos from in vitro fertilisation (IVF) treatment (Thomson et al., 1998). Despite their derivation from blastocysts and their capacity to produce teratomas the properties of hESCs have proven closer to those of rodent Epi-SCs (Chia et al., 2010; Smith, 2001). In particular hESCs do not respond to LIF or 2i, but require signaling through the FGF and activin pathways for self-renewal and like murine EpiSCs are markedly heterogeneous between and within cell lines with significant biases in differentiation potential (Osafune et al., 2008). Molecularly, hPSCs utilise the Oct4 proximal enhancer, express high levels of DNA methylation and inactivate the X chromosome (Hanna et al., 2010). This difference to mESCs may arise from the process of derivation from primate embryos. In non-rodent mammals such as primates, the epiblast develops as a flat epithelial disk, which may develop more readily in explant culture than rodent egg cylinder morphogenesis, thus developing to the equivalent of a post-implantation EpiSC state. Furthermore, primates do not exhibit diapause, so the ability to extrinsically stabilise ground-state pluripotency may be easier in, or potentially specific to, rodents.

This similarity to murine EpiSCs under current culture conditions presents a number of technical limitations to the use of hPSCs. Firstly, hPSCs exhibit marked variation in differentiation potential both between and within cell lines. Secondly, hPSCs survive single cell dissociation poorly, limiting the ability to sub-clone cells of interest, for example during genome engineering. hPSCs have traditionally been maintained on MEF feeders in high
concentrations of exogenous FGF. Although commercial feeder free platforms exist they are expensive and the majority are not suitable for GMP use in humans. Thus there is considerable interest in deriving and maintaining ground state hPSCs in defined conditions that are both clonogenic and lack significant differentiation bias.

1.3.2.1. **Ground State Human Pluripotent Stem Cells**

It is unclear what characteristics naïve human PSCs might share with ground state mESCs, however, evidence that such a state may exist in human has been provided by single cell RNASeq. This showed that the naïve pluripotency markers defined in mouse were highly up-regulated in pre-implantation human morula and epiblasts compared to hESCs (Yan et al., 2013). Two recent papers have suggested that pharmacological control of signaling pathways may be able to stabilise ground state hPSCs. Jaenisch’s group used a genome-engineered reporter of the Oct4 distal enhancer to identify a number of inhibitors of signaling kinases (including BRAF, LYC/SRC, GSK3, LIF, Activin, ROCKi) that in combination could induce and maintain cells that morphologically and transcriptionally resembled naïve mESCs. However a number of issues remained including residual X inactivation, slow proliferation, karyotypic instability and failure to contribute to non-gestational chimeras (Theunissen et al., 2014).

Smith’s group used doxycycline inducible KLF2 and NANOG to reset hPSCs to a ground state that was capable of growing in the presence of 2i/LIF (Takashima et al., 2014). The cells cloned without the use of rock inhibitor and died in standard FGF/knockout-serum replacement conditions. This reset state could be maintained pharmacologically with the addition of a protein kinase C (PKC) inhibitor and cells could be grown at least temporarily feeder-free on matrigel or laminin. Naïve hPSCs shared features with ground state mouse
including teratoma formation, expression of naïve factors (except Essrb), increased oxidative phosphorylation, globally reduced DNA methylation, a global transcriptional state that clustered near mouse naïve ES and some degree of contribution to mouse blastocysts. The lack of requirement for Essrb may explain failure of 2i to work in hPSCs and the fragility of naïve pluripotent state in human compared to rodent. Although a promising advance, technical issues remain, not least the requirement of transgenes to re-programme existing hPSC lines and the ongoing requirement for feeders (A Smith, personal communication).

1.3.3. Induced Pluripotent Stem Cells

Almost all cells in the body retain a full complement of genomic DNA opening the possibility that any cell could be epigenetically reprogrammed back to pluripotency (Gurdon 1962). By retrovirally transducing candidate ESC regulators into murine fibroblasts and then exposing them to ESC culture conditions Takahashi and Yamanaka derived small numbers of ESC-like colonies (Takahashi and Yamanaka, 2006). Subsequent testing identified a minimal group of four factors (OSKM - Oct4, Sox2, Klf4 and c-myc) that were sufficient to reprogram fibroblasts to a pluripotent-like state labeled induced pluripotent stem cells (iPSCs). This work was soon followed up in human (Takahashi et al., 2007) and subsequently a range of factors, gene delivery systems and small molecule adjuncts have been described to facilitate reprogramming of multiple cell sources. Human iPSCs allow the ethical derivation of hPSC lines from theoretically any genetic background or diseased cell type. The potential power of iPSCs has been demonstrated by the successful treatment of a mouse model of sickle cell anaemia using genetically corrected autologous iPSCs (Hanna et al., 2007).

Systems analysis of iPSC reprogramming has demonstrated that iPSCs closely resemble ESCs, although they do transit through intermediate states during reprogramming and some iPSCs
can retain an epigenetic memory of their original cell state (Kim et al., 2010; Mikkelsen et al., 2008). Some groups have suggested that such ‘memory’ limits iPSC differentiation potential compared to cells derived by somatic cell nuclear transfer (SCNT) (Boland et al., 2009), although experience in haematopoietic differentiation from hPSCs have suggested that iPSCs may be more permissive to in vitro differentiation than hESCs (Carpenter et al., 2011). The combination and concentration of re-programming factors used are important, with high Oct4/Klf4 and low myc/Sox2 found to be favourable (Carey et al., 2011). Using more stringent tetraploid complementation assays of functional pluripotency Buganim et al have shown that many murine iPSCs derived using traditional OSKM factors in fact generate ‘poor quality’ chimeras, whereas a novel group of factors (SNEL - Sall4, Nanog, Esrrb and Lin28) generated smaller numbers of higher quality iPSC colonies (Buganim et al., 2014). Interestingly myc expression was associated with higher levels of DNA damage including trisomy 8 and abnormal histone H2A.X deposition, possibly providing a novel mechanism for why iPSCs exhibit gene expression biases. This notwithstanding, overall human iPSCs can be considered to be largely equivalent to hESCs, with the key caveat that expression of reprogramming factors should be transient and ideally avoid any integration into the genome.

1.3.4. In Vitro Differentiation of Human Pluripotent Stem Cells

The true power of pluripotent stem cells relies on their ability to generate specific cells of interest. In the case of haematopoietic differentiation, a number of methodologies have been developed for the directed in vitro differentiation of hPSCs to haematopoietic progenitors and B cells. Kauffman demonstrated the first successful differentiation of hESCs using S17 mouse stromal co-culture (Kaufman et al., 2001). Subsequently, several protocols for the haematopoietic differentiation of hESCs have been published based on either embryoid body or 2D co-cultures, including serum and feeder free adaptations (Slukvin,
Where tested, differentiation of hPSCs has provided largely similar results to that from hESCs (Choi et al., 2009). In general, all differentiation protocols involve two stages: firstly haemato-endothelial specification, and secondly amplification and lineage commitment from established multipotent haematopoietic progenitors. Techniques for the first stage were based on work in mouse using established mesodermal morphogens such as BMP4, WNT, FGF2 and VEGF. The second stage uses standard haematopoietic in vitro culture conditions, but optimised for cell number and developmental stage. A peculiar issue affecting the in vitro differentiation of hPSCs is that unlike in the developing fetus, different ‘waves’ of haematopoiesis spatially co-exist. This has led to the need to identify novel surface markers that can distinguish mesodermal and haematoendothelial developmental intermediates. The Slukvin group who pioneered the use of the OP9 coculture system has performed much of this characterisation (summarised in Table 1.1) and the descriptions below are based on this protocol.

**1.3.4.1. Mesodermal Development and Early Haematopoiesis**

Early mesodermal development in the mouse is identified by the surface markers KDR (FLK1)⁺ and PDGFRα⁺, but in human PSCs some KDR expression is present prior to differentiation making it a non-specific marker. By using a GFP reporter for the early mesodermal factor MIXL1, apelin receptor (APLNR) was identified as a novel marker of the early primitive streak in human (Davis et al., 2008). KDR⁺, APLNR⁺, PDGFRα⁺ nascent mesoderm is enriched for a small population of cells with blast CFC (BL-CFC) potential when cultured in the presence of FGF2 and VEGF, but not haematopoietic cytokines (Kennedy et al., 2007). These early emerging haemangioblast-like cells give rise to ‘epitheliod’ cores with surrounding vascular and haematopoietic derivatives with MkE and monocytic potential.
Emergence of later lateral plate mesoderm is characterised by loss of PDGFRα in the mouse (Sakurai et al., 2006). As such, Choi et al examined the expression of KDR (FLK1)+, PDGFRα and APLNR during hPSC differentiation (Choi et al., 2012). KDR (FLK1), PDGFRα and APLNR are also found on more mature human cells so they first excluded more mature haematopo-endothelial progenitors (EMH-Lin−: endothelial (CD31/VE-cadherin, endothelial-mesenchymal CD73/CD105 and haematopoietic CD43/45). Using this strategy they identified a EMH-Lin− /APLNR+/KDR+/PDGFRα−/lo population which they called haematovascular mesodermal precursors (HVMP). These cells down-regulate primitive streak genes whilst co-expressing lateral plate and extraembryonic mesoderm genes with genes associated with angiohaematopoietic development (e.g. TAL1, HHES, LMO2, GATA2, ETV2). Functionally this population is enriched in myeloid-potent haemato-endothelial progenitors.

**1.3.4.2. Derivation of Haemogenic Endothelium and Definitive Haematopoiesis**

The derivation of definitive haematopoietic stem and progenitor cells from hPSCs requires the identification of haemogenic endothelial (HE) progenitors in hPSC cultures (Sturgeon et al., 2013). However, classical surface markers are unable to specifically distinguish HE cells. CD34 is expressed on all haematopoietic, endothelial and mesenchymal stem cells. VE-Cadherin is expressed on endothelial cells (including non-HE) and in mouse is usually combined with CD41a (embryonic haematopoietic progenitors) and CD45 (pan-haematopoietic). A significant advance in the field was the identification that leukosialin (CD43) marked the entire population of emerging haematopoietic cells (Vodyanik et al., 2006). CD34+CD43− cells had no CFC potential, but could generate blood and endothelium upon further OP9 co-culture. Time-lapse imaging suggests CD43 may function by forming a repulsive barrier around haematopoietic cells as they bud from endothelium (Manjunath et
al., 1995). The first CD43lo cells can be detected at day 4 of OP9 co-culture and co-express the erythroid marker glycophorin A (CD235a) (Choi et al., 2012). These angiohaematopoietic progenitors (AHPs) can generate blood and endothelium including cytokine independent myeloid colonies, but in contrast to the early haemangioblastic colonies (above) these did not contain endothelial cores. Exclusion of cells expressing the marker CD73 (5’nucleotidase) could define HVMP-derived HE (VECadherin’CD43 CD73 ) from non-haemogenic endothelium. The transcription factor SOX17, which is required for specification of haemogenic endothelium, is down-regulated during haematopoietic development from hPSC-derived HE (Nakajima-Takagi et al., 2013).

By day 8 of OP9 co-culture, Lin’CD34+CD43+CD45lo multipotent haematopoietic progenitors (MHP) with broad lymphomyeloid potential are present. These cells have CFC capacity similar or higher to that from equivalent fractions of cord blood (Salvagiotto et al., 2008) and share the surface phenotype of adult HSCs (including CD90+, Kit+, 38, CD45RA, Rho). These MHPs have lymphoid potential, including NK (Woll et al., 2005) and T cell potential on OP9-DLL1/4, which was found exclusively in the CD34+43lo fraction (Timmermans et al., 2009). Generation of B lineage output by M55 co-culture supplemented with cytokines (IL3, IL7, SCF, Flt3L) has proven challenging for many laboratories and, until recently had failed to produce IgM-expressing naïve B cells (Carpenter et al., 2011). A recent update from the Carpenter group has shown that re-plating CD45 cells into a further 3 weeks of M55 co-culture and removing exogenous rhIL7 generated a small population of IgM+ naïve B cells (French et al., 2015). This suggests that ongoing IL7 exposure can block hPSC-B differentiation beyond the preB stage, possibly by inhibiting light chain recombination.

Consistent with studies of T cell output (Kennedy et al., 2012), the HE enriched compartment (VE-Cadherin’CD43 CD235a CD73 ) exhibited >80 fold greater B potential compared to the haematopoietic specified fraction (CD43+/CD235a+) derived from day 10 of
OP9 co-culture, suggesting that haematopoietic progenitors derived later in co-culture have enhanced lymphoid potential and potentially providing a new strategy for sorting B-lymphoid potent HE. This notwithstanding, the best characterised system for the directed differentiation of multiple haematopoietic lineages including B lineage from hESCs (Vodyanik et al., 2005) and hiPSCs (Carpenter et al., 2011) utilises OP9 stromal co-culture to produce Lin CD34⁻CD43⁻CD45⁺/- multipotent haematopoietic progenitors (MHPs), with subsequent re-plating onto MS5 stroma to produce CD19⁺ B-lineage output.

1.3.4.3. Generation of Engraftable dHSCs

T, and where achieved B, lymphoid output is considered a marker of generation of definitive haematopoiesis, although recent studies in mouse demonstrating the emergence of lymphoid progenitors prior to dHSCs questions the validity of this assumption (Boiers et al., 2013). Of note engraftable HSC-like cells have never been robustly derived from hPSCs. Xenotransplantation of hESCs or hESC-derived CD34⁺ cells into mouse and sheep produced detectable haematopoietic cells in the bone marrow months after injection, but at very low levels (0.1-2%) and with mainly myeloid reconstitution (Ledran et al., 2008; Narayan et al., 2006). Two studies have shown haematopoietic differentiation in vivo from hiPSC teratomas grown in the presence of OP9-DLL1/WNT3a stroma or hSCF/hTPO via a pump, but both methods also gave low levels of engraftment (Amabile et al., 2013; Suzuki et al., 2013). A number of hypotheses have been suggested to account for the failure to generate engraftable HSCs from hPSCs including limited migratory potential, failure to complete HSC maturation or a failure to capture or expand rare HSCs. In addition hPSC-derived HSC-like cells have differences in gene expression compared to FL including failure to up-regulate HoxA9 and Notch pathways (McKinney-Freeman et al., 2012) and the homing chemokine receptor CXCR4 (Salvagiotto et al., 2008). Overall these suggest that haematopoietic
differentiation from hPSCs models waves of haematopoiesis derived from early
developmental 'pre-HSCs,' which may be advantageous when investigating the prenatal
origins of childhood leukaemia.

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Phenotype</th>
<th>Day on OP9 co-culture</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Posterior Mesoderm (PM)</td>
<td>EML Managers Lin&lt;sup&gt;+&lt;/sup&gt; APLNR&lt;sup&gt;+&lt;/sup&gt; PDGRFa&lt;sup&gt;+&lt;/sup&gt;</td>
<td>3</td>
<td>Express typical lateral plate/extra-embryonic mesoderm genes. Potential to form FGF2 dependent haemangioblast colonies in serum free media.</td>
</tr>
<tr>
<td>Haematovascular mesodermal precursors (HVMP)</td>
<td>EML Managers Lin&lt;sup&gt;+&lt;/sup&gt; KDR&lt;sup&gt;hi&lt;/sup&gt; APLNR&lt;sup&gt;+&lt;/sup&gt; PDGFR&lt;sup&gt;low/&lt;/sup&gt;</td>
<td>4</td>
<td>Express lateral plate/extraembryonic mesoderm genes with angiohaematopoietic, but not primitive streak genes. Potential to form haematoendothelial clusters on OP9.</td>
</tr>
<tr>
<td>Angiogenic Haematopoietic Progenitors (AHP)</td>
<td>VE-cadherin&lt;sup&gt;+&lt;/sup&gt; CD73 CD43&lt;sub&gt;low&lt;/sub&gt; CD235α&lt;sup&gt;+&lt;/sup&gt; CD41α&lt;sup&gt;+&lt;/sup&gt; CD117&lt;sup&gt;+&lt;/sup&gt;</td>
<td>5</td>
<td>Primarily haematopoietic with haematoendothelial potential.</td>
</tr>
<tr>
<td>Erythromegakaryocytic progenitors (EMkP)</td>
<td>VE-cadherin&lt;sup&gt;+&lt;/sup&gt; CD73&lt;sup&gt;+&lt;/sup&gt; CD43&lt;sup&gt;+&lt;/sup&gt;/CD235α&lt;sup&gt;+&lt;/sup&gt; CD41α&lt;sup&gt;+&lt;/sup&gt;</td>
<td>5</td>
<td>Early EMk progenitors</td>
</tr>
<tr>
<td>Non-haemogenic endothelial progenitors (NonHEP)</td>
<td>VE-cadherin&lt;sup&gt;+&lt;/sup&gt; CD73&lt;sup&gt;+&lt;/sup&gt; CD235α/CD43&lt;sup&gt;+&lt;/sup&gt; CD117&lt;sub&gt;hi&lt;/sub&gt;</td>
<td>5</td>
<td>Endothelial progenitors</td>
</tr>
<tr>
<td>Haemogenic Endothelial Progenitors (HEP)</td>
<td>VE-cadherin&lt;sup&gt;+&lt;/sup&gt; CD73&lt;sup&gt;+&lt;/sup&gt; CD235α/CD43&lt;sup&gt;+&lt;/sup&gt; CD117&lt;sub&gt;int&lt;/sub&gt;</td>
<td>5</td>
<td>Endothelial charateristics, lack haematopoietic CFC potential, but can mature into blood and endothelium on OP9 co-culture. Recently shown to have high lymphoid potential on further co-culture.</td>
</tr>
<tr>
<td>Multipotent Haematopoietic Progenitors (MHP)</td>
<td>Lin&lt;sup&gt;+&lt;/sup&gt; CD34&lt;sup&gt;+&lt;/sup&gt; CD43&lt;sup&gt;+&lt;/sup&gt; CD45&lt;sup&gt;+&lt;/sup&gt; CD38&lt;sup&gt;+&lt;/sup&gt;</td>
<td>8</td>
<td>Lineage negative haematopoietic progenitors with the potential to form the full spectrum of myeloid colonies. Used as the basis of B lymphoid differentiation.</td>
</tr>
</tbody>
</table>
Table 1.1: Subsets of haematoendothelial progenitors identified during differentiation of hPSCs on OP9 co-culture (Adapted from Choi et al., 2012)

1.4. Genome Engineering

Cancer is a disease characterised by the accumulation of multiple genetic and epigenetic changes occurring within a specific cell type. Genes that are activated during carcinogenesis are termed oncogenes and typically include the overexpression of physiological cellular regulators or the creation of novel oncogenic proteins such as fusion proteins formed from chromosomal translocations. Conversely tumour suppressor genes act to inhibit cancer formation and require inactivation of one or both alleles.

Understanding how mutations contribute to carcinogenesis ideally requires a model that accurately recapitulates the genetic lesions in the appropriate human cellular context in which they arise. The field of cancer genetics uses many of the approaches developed by functional geneticists. The gold standard for deciphering gene function is to examine a loss of function phenotype. In so-called ‘forward genetics’ this is achieved by inducing semi-random mutations in model organisms or cell lines and then screening for observable phenotypes and identifying the affected genes. Alternatively ‘reverse genetics’ targets a specific gene and then analyses the phenotypic consequences, but this requires precise targeting of a specific gene or gene product. To analyse the abnormal expression of a gene, or the expression of a novel gene product requires over-expression or targeted knock-in technology. Retroviral and lentiviral transfection of cells with cDNA is the main methodology used to introduce genes of interest into cells and is relatively cheap, quick and technically straightforward. Viral transfection has a number of technical limitations, however, including size limits on the packaging of cDNA, supra-physiological overexpression of the transgene,
off target effects from random integration into the genome and the potential for epigenetic silencing of the transgene over time. A number of other transgenic approaches exist that allow the random integration of larger complex DNA constructs including non-coding regulatory elements such as BAC transgenesis and PiggyBac transposon-mediated integration (Cadinanos and Bradley, 2007; Poser et al., 2008). In general these methods exhibit less efficient transduction rates, particularly in primary cells, limiting their application to cell lines that can be sub-cloned and re-expanded.

1.4.1. Homologous Recombination

Targeted insertion of transgenes by gene targeting or genome engineering has the ability to more faithfully recreate oncogeneic mutations. Work in the 1980s by Capecchi and Smithies showed that precise genomic targeting could be by achieved in mouse ES cells by providing constructs on a targeting vector sharing homology to the targeted locus (Smithies et al., 1985; Thomas et al., 1986). Large homology arms (e.g. 10kb) were required to achieve successful targeted homologous recombination with rare targeting events selected and candidate knock-in clones screened by Southern blot hybridisation (Rudolph et al., 1994). A major breakthrough in the field was the ability to genetically engineer these long homology arms directly from BAC DNA in live cultures of E. coli (Muyrers et al., 1999). Recombineering employs λ phage proteins driven from an inducible plasmid to promote homologous recombination (HR) between DNA molecules in liquid cultures of Escherichia coli. Linear donor DNA containing terminal 50bp regions of homology (usually as a PCR product) is transformed by electroporation and specifically recombines by HR with cognate plasmid or BAC DNA sequences on induction of the recombineering enzymes. Although recombination is a rare event successful HR can be selected by the exchange of positive/negative selection cassettes (Bird et al., 2012). Recombineering has a number of advantages over traditional
cloning or polymerase chain reaction (PCR) approaches, including its independence from restriction sites, lack of DNA size limits, high speed and low rates of mutagenesis making it well suited to the generation of large gene targeting vectors (Testa et al., 2003).

1.4.2. Site Specific Recombinases

Recombineering can be combined with a toolkit of plasmids encoding generic DNA cassettes to make complex conditional alleles (Schnutgen et al., 2006). These alleles make use of site-specific recombination (SSR) sequences such as LoxP, Frt or Rox that are recognised by the highly efficient phage recombinase enzymes Cre, Flp and Dre respectively. When such sequences are aligned in the same direction the related enzyme will delete or ‘flox’ the intervening DNA, leaving a single recombination site; whereas if aligned in opposite directions the recombinase will invert the intervening sequence in a reversible manner. Of the above enzymes Cre recombinase is the most active and commonly used (Sauer and Henderson, 1988). LoxP has been further engineered into multiple heterotypic variants that recombine only with their homotypic counterparts, allowing the design of complex conditional alleles such as the FLEX switch (Schnutgen et al., 2003). Using recombineering to target standardised SSR-containing cassettes to critical exons in mouse ES cells the international knockout mouse consortium has successfully produced knockout and conditional knockout mice for almost every coding gene (Fu et al., 2010; Skarnes et al., 2011).

1.4.3. Site Specific Nucleases

Although precise and to some degree scalable, gene targeting by HR is slow and some alleles have proven challenging to target. In addition, whereas spontaneous HR occurs frequently in
mouse ES cells, it is very infrequent in hPSCs meaning that the majority of clones surviving selection will have randomly integrated the targeting vector. Methods to improve the efficiency and accuracy of gene targeting have focused on creating site specific double strand breaks (DSB) in genomic DNA to activate non homologous end joining (NHEJ) or precise homologous recombination (HR) DNA repair pathways (Rouet et al., 1994). Early approaches at inducing targeted DSBs used DNA binding oligonucleotides or peptide nucleic acids (PNAs) linked to DNA crosslinking agents, but these large molecules were limited by their poor cellular penetration (Faruqi et al., 1998). Peter Dervan conjugated naturally occurring DNA binding agents into a small molecule hairpin polyamide structure that bound DNA in a sequence specific manner though the minor-groove sterically hindering transcription factor binding (Gottesfeld et al., 1997). Elucidation of the modular DNA-reading structure of eukaryotic zinc finger DNA binding proteins led to the development of modular DNA recognition proteins that could be constructed from a library of plasmid intermediates. By conjugating two adjacent DNA binding domains to the dimeric endonuclease FokI, Kim et al generated zinc finger nucleases (ZFN) that could be administered to cells by transient transgenesis of mRNA or plasmid DNA (Kim et al., 1996). Although active in hPSCs (Hockemeyer et al., 2009), zinc finger nuclease libraries were complex and expensive to design, construct and validate limiting their widespread use.

1.4.3.1. **TALENs**

The first truly tractable modular DNA binding proteins were transcription activator-like effectors (TALEs). TALEs were identified in the plant pathogenic bacteria *Xanthomonas*, which secrete them to regulate specific genes in their plant host (Bogdanove et al., 2010). TALE DNA binding domains consist of 7-34 highly homologous direct repeats. Each repeat contains two critical amino acids at positions 12 and 13 termed repeat variable domains
(RVDs) that read the DNA base pair sequence by a simple one to one code (Boch et al., 2009). Arrays of RVDs can be constructed in a plasmid to target 15-20bp, which when transcribed in vivo produces a large protein of 500-700 amino acids. When fused to the FokI endonuclease pairs of TALE-Nucleases (TALENs) targeting a total of 40-50bp can be designed to produce highly specific DNA DSBs. In the absence of a HR donor, a DSB will usually result in error prone NHEJ, likely resulting in a heritable knock out mutation. DNA binding TALE arrays can also be fused to other effectors such as transcriptional repressors (e.g. KRAB-TALE) producing a knockdown phenotype similar to RNAi (Cong et al., 2012). Cloning and sequencing of TALEN-encoding plasmids is complicated by the highly repetitive nature of TALE repeats. Several techniques have been developed to facilitate TALEN construction, usually involving the creation of ‘libraries’ of plasmids encoding array motifs, with the most widely used method being golden gate cloning (Cermak et al., 2011). Whilst TALENs solved many of the limitations of traditional gene targeting, including genome engineering hPSCs (Hockemeyer et al., 2011), the technology was soon largely superseded by CRISPR:Cas9 (Doudna and Charpentier, 2014).

1.4.3.2. CRISPR:Cas9

In 1987 Japanese researchers identified an unusual region of Escherichia coli chromosomal DNA they termed clustered regulatory interspaced palindromic repeats (CRISPR) (Ishino et al., 1987). In 2005 it was demonstrated that the spacer sequences between these arrays of repeats were of plasmid or viral origin. The CRISPR region was transcribed as a single precursor crRNA and then cleaved into individual crRNAs (Pourcel et al., 2005) whilst a series of CRISPR associated genes (Cas) situated 5’ of the CRISPR region were shown to represent an operon of nucleases. Overall these findings suggested the presence of a bacterial adaptive immune system whereby plasmid or phage DNA is incorporated between the
CRISPR repeats and subsequently transcribed as antisense RNA to direct degradation of the invading DNA by Cas nucleases.

Three types of CRISPR system have been identified. Types I and II read a short DNA sequence 3’ of an obligate binding site or protospacer adjacent motif (PAM). Types I and III need a large complex of Cas proteins, whereas type II just uses a single cas nuclease. An additional trans activating RNA (tracrRNA) encoded upstream of the type II cas locus in *Streptococcus pyogenes* is essential for the maturation of the crRNA (Deltcheva et al., 2011). Thus a single Cas9 nuclease binds the tracrRNA:crRNA duplex, directing a sequence specific DSB to a 20nt sequence upstream of a PAM. Combining the tracrRNA and crRNA into one strand makes a single guide RNA (gRNA) (Jinek et al., 2012). X ray crystallography has shown that gRNA binding causes a large conformational change in Cas9, which further changes on binding to DNA. The process of DNA binding starts by PAM recognition though conserved Cas9 arginine residues, which possibly initiates DNA melting, strand invasion and RNA:DNA hybrid formation (Sternberg et al., 2014). *Streptococcus* Cas9 cleaves each strand of DNA through separate catalytic domains (termed ‘HNH’ and ‘RuvC’); mutating either of these makes a single strand ‘nickase,’ whereas mutating both makes a DNA binding protein (dCas9).

Expressing class II Cas9 and a gRNA from separate plasmids has shown rates of targeted mutagenesis as high as 50% in eukaryotic cells (Mali et al., 2013). Although at the population level such mutagenesis would be heterogeneous, this level of activity permits the cloning of both hetero and homozygous knockouts. In addition to high levels of cutting, the other major advantage of CRISPR-mediated genome engineering is the ease and low cost in which new gRNAs can be programmed. Compared to other directed nucleases, CRISPR has a small number of design principles. The most stringent is the obligate need for a PAM sequence at the 3’ end of the DNA binding domain (NGG for Cas9). The sequence specificity of the gRNA
is most rigorously mediated by the most 3’ 8-12bp seed sequence and this relatively short
sequence needs to be checked against other potential binding sites in the genome. Binding
in the sense or antisense orientation does not affect cutting, but homopolymeric DNA
targets are best avoided and chromatin structure is known to affect binding kinetics (Doench
et al., 2014; Wu et al., 2014). gRNAs are usually driven from an RNAPolIII promoter (such as
U6, H1, 7SK, tRNA) as these transcribe small RNAs at high levels. Tissue specific gRNA
expression would require expression through RNAPolIII and techniques to avoid mRNA
capping machinery by flanking the gRNA with ribozyme or CsY4 RNA cleavage sites have
been developed (Nissim et al., 2014).

The major concern with CRISPR-directed nucleases is the potential for off target
mutagenesis. After recognising the PAM gRNA:Cas9 interrogates the 5’ flanking DNA (Cong
et al., 2013). Cas9 ChIPSeq has shown significant off target binding, usually to sites with
sequence homology to the 5bp upstream of the PAM, but cleavage is infrequent at these
sites (Wu et al., 2014). Live imaging has shown that off target binding is usually short lived
(<1 second) and that Cas9 bound heterochromatin less frequently and with slower kinetics
(Knight et al., 2015). GUIDESeq has shown that off target cutting events are highly variable
between gRNAs ranging from 0-150 sites per gRNA (Tsai et al., 2015). Modifications to
reduce off target mutagenesis include reducing the expression level of Cas9 (Hsu et al.,
2013), shortening the gRNA to 17/18bp to reduce the annealing entropy (Fu et al., 2014),
dCas9-FokI dimerisation akin to that used in TALEN and ZFNs (Tsai et al., 2014) and reducing
Cas9 helicase activity thus increasing the stringency binding afforded by sequence specific
RNA:DNA strand invasion (Slaymaker et al., 2016). A highly useful adaptation for CRISPR-
directed HR is the use of single or double Cas9 nickases. ssDNA breaks are repaired with
extremely high fidelity such that even using two adjacent nickases to create a DSB reduces
off target cutting 50-1500 fold (Ran et al., 2013). Transient expression of a single nickase in conjunction with a HR donor is highly unlikely to result in off target mutagenesis.

The technical advantages of CRISPR:Cas9 have produced a paradigm shift in both forward genetics and genome engineering. Large-scale knockout screening libraries have been constructed in lentivirus (Wang et al., 2014). The Jaenisch group has used multiplex CIRPSR to disrupt 8 alleles simultaneously in mouse ES and make double KO animals (Wang et al., 2013) and this work has been extended to conditional and reporter alleles (Yang et al., 2013). The fusion of dCas9 to the transcriptional effectors KRAB or VP64 produces inhibitory (CRISPRi) or activating (CRISPRa) respectively (Cheng et al., 2013; Gilbert et al., 2013). Live imaging of specific loci has been established using GFP tagged dCas9 (Chen et al., 2013).

Recently mice have been engineered with Dox / Cre inducible Cas9 alleles further facilitating the ease of gene editing in this key model system. Excitingly, the potential clinical utility of CIRPSR gene editing has been shown by the successful correction of the cystic fibrosis CFTR mutation in adult somatic intestinal stem cells, which could subsequently be grown into intestinal organoids (Schwank et al., 2013).

The applicability of CRISPR in cancer modeling has also been pioneered by a number of groups: CRISPR inactivation of P53 in a lymphoma model conferred resistance to doxorubicin (Malina et al., 2013); a single allele CRISPR knockout of MLL3 in mouse haematopoietic stem and progenitor cells confirmed MLL3 as a haploinsufficient tumour suppressor in acute myeloid leukaemia (AML) (Heckl et al., 2014); and multiplex targeting of tumour suppressor genes in mouse haematopoietic stem and progenitor cells resulted in de novo AML (Chen et al., 2014). Although Cas9:gRNA delivery techniques remain a limitation, CRISPR knock-out has also been used to induce somatic mutations in vivo: hydrodynamic gene transfer into hepatocytes generated hepatocellular carcinoma (Xue et al., 2014) and large scale
chromosomal re-arrangement could be induced to recreate Eml4-Alk lung tumours (Maddalo et al., 2014). These early *in vitro* and *in vivo* studies demonstrate the potential of CRISPR-directed genome engineering to dissect the contribution of initiating and collaborating mutations in the origins and clonal evolution of cancers such as leukaemia.

1.5. Aim of the study: A human pluripotent stem cell model of ETV6-RUNX1 pre-leukaemia

Despite accounting for 25% of childhood ALL, the molecular and functional impact of the first hit mutation ETV6-RUNX1 is uncertain. Current models of ETV6-RUNX1 have produced variable results depending on: i) the level of ETV6-RUNX1 expression; ii) the stage of B cell differentiation that is targeted; and iii) the developmental hierarchy studied. I hypothesised that the *in utero* origins and childhood tropism of ETV6-RUNX1 results from its initiation in transient stem cell or progenitor compartments unique to early human development, and that B lymphoid output from human pluripotent stem cells would recapitulate the cellular hierarchies seen in the developing human fetus. If so, by genome engineering hPSCs we could study the effect of ETV6-RUNX1 under the control of the endogenous *ETV6* promoter in the appropriate developmental context.

The study consisted of 3 broad objectives:

1) The establishment of B lymphoid differentiation from hPSCs and its comparison to B lymphopoiesis in human fetal liver

2) Generation of validated hPSC clones genome engineered to harbour ETV6-RUNX1.

3) Molecular and functional characterisation of the B lymphoid output from ETV6-RUNX1 expressing hPSCs.
CHAPTER 2

Materials and Methods
2. CHAPTER 2: Materials and Methods

2.1. Molecular Biology

2.1.1. Transformation of bacteria

New plasmids (0.1-1ng) or ligation mixtures (10-50% mixture) were transformed into chemically competent sub-cloning efficiency DH5α (Invitrogen) or high efficiency NEB5α Escherichia coli (New England BioLabs - NEB) respectively. Plasmid DNA was added to 50μl of E. coli in a 1.5ml sterile microcentrifuge tube and incubated on ice for 30 minutes. The plasmids were transformed by heat shock in a water bath at 42°C for 40 seconds and subsequently incubated on ice for an additional 5 minutes. Cultures were recovered in 950μL of SOC outgrowth medium (NEB) at 37°C for 1 hour in a bacterial shaker (250 rpm). 1, 10 and 90% of this mix was then inoculated onto pre-warmed selection plates prepared with LB agar (1.5g Bacto Agar (BD Bioscience) per 100ml LB broth 1% w/v Bacto Tryptone (BD Bioscience), 0.5% w/v Bacto Yeast Extract (BD Bioscience), 1% w/v Sodium Chloride (NaCl), [pH 7.0]) and 100μg/ml Ampicillin (Sigma-Aldrich) and incubated at 37°C overnight.

Antibiotic concentrations for plate and liquid cultures are shown in Table 2.1.
<table>
<thead>
<tr>
<th>Antibiotic</th>
<th>Solvent</th>
<th>Stock (mg/ml)</th>
<th>Working (μg/ml)</th>
<th>Dilution times</th>
<th>Medium (ml)</th>
<th>Add (μl)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ampicillin</td>
<td>H2O</td>
<td>100</td>
<td>50 Liquid</td>
<td>2000</td>
<td>100</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>100 Plate</td>
<td>1000</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Chloramphenicol</td>
<td>100% EtOH</td>
<td>30</td>
<td>10 Liquid</td>
<td>3000</td>
<td>100</td>
<td>33.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>30 Plate</td>
<td>1000</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Kanamycin</td>
<td>H2O</td>
<td>30</td>
<td>5 Liquid</td>
<td>100</td>
<td>100</td>
<td>16.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>15 Plate</td>
<td>100</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Tetracycline</td>
<td>75% EtOH</td>
<td>10</td>
<td>3 Liquid</td>
<td>100</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5 Plate</td>
<td>100</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>Gentamicin</td>
<td>H2O</td>
<td>50</td>
<td>1 Liquid</td>
<td>100</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>2 Plate</td>
<td>100</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Streptomycin</td>
<td>H2O</td>
<td>50</td>
<td>100 Liquid</td>
<td>100</td>
<td>200</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>200 Plate</td>
<td>100</td>
<td>400</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.1: Antibiotics used in molecular biology protocols including recombineering (unless otherwise stated in text). EtOH, ethanol.

2.1.2. Isolation of plasmid DNA

For routine analytical purposes small amounts of plasmid DNA were ‘mini-prepped’ by modified alkaline lysis. Individual bacterial clones were inoculated into 6ml LB broth containing appropriate selection and incubated in a shaker at 37°C overnight in a 50ml Falcon tube (BD). 850μl culture was thoroughly mixed with 150μl (15%) sterile glycerol, transferred to a cryovial (Nunc) and flash frozen on dry ice for storage at -80°C. The
remainder of the bacterial culture was used to extract DNA using the QIAprep® Miniprep Plasmid kit (Qiagen) according to the manufacturer’s instructions. In brief, the bacterial culture was thoroughly re-suspended in 250μl of buffer P1 containing ‘lyseblue’ indicator and transferred to a microcentrifuge tube. 250μl of buffer P2 (checked for SDS precipitation) was added and mixed thoroughly by inverting the tube 4–6 times to lyse the cells until the solution turned uniformly blue. The tube was briefly centrifuged to remove material from the lid and after 5mins, 350μl of buffer N3 was added and mixed thoroughly by inverting the tube 4–6 times until no areas of blue remained. The lysed mixture was centrifuged for 10 min at 13,000 rpm in a microcentrifuge. The liquid lysate was then transferred by pipetting to a QIAprep spin column and centrifuged at maximum speed for 60 seconds and the flow-through discarded. The column was washed first by adding 0.5ml buffer PB and centrifuging for 60s; then by adding 0.75ml buffer PE and centrifuging for 60s. The flow-through was discarded at each step before the column was finally centrifuged at full speed for an additional 1 min to remove residual ethanol-containing PE buffer. The QIAprep column was placed in a clean 1.5 ml microcentrifuge tube and 20-50μl IDTE Buffer (10 mM Tris·Cl, pH 8.0) or water were added to the centre of each QIAprep spin column to elute DNA. The solution was incubated for 1 min, followed by a 1 min max speed centrifugation of the tube. The final DNA concentration was measured using a spectrophotometer (NanoDrop.ND-1000, Lebtech International) or by Qubit BR fluriometry (Molecular Probes).

Larger quantities of plasmid was produced by processing larger bacterial cultures in a similar way using plasmid plus Midi or Maxi Kits (Qiagen) according to manufacturer’s instructions.
2.1.3. Preparation of BAC DNA

Bacterial artificial chromosome (BAC) clones were purchased from Source Biosciences as an agar stab and inoculated onto agar plates with chloramphenicol selection (15μg/ml). Clones were grown out in 5ml LB containing 10μg/ml chloramphenicol and DNA prepared using a modified alkaline lysis protocol and spin column (ZR BAC DNA miniprep kit, Zymo research) according to manufacturer’s instructions. Bacteria were re-suspended in 200μl P1 buffer and transferred to a microcentrifuge tube. Cells were lysed in 200μl buffer P2 for 3 minutes, checking that the solution had turned purple. This lysis was neutralised by addition of 400μl P3 making the solution turn yellow. The mixture was incubated at room temperature for 2 minutes before centrifugation for 3 minutes at maximum speed on a bench top microcentrifuge. The supernatant was transferred to a Zymo-spin™ IC-XL column and centrifuged again for 30 seconds. The flow through was discarded and washed with 200μl endo wash buffer and then 400μl plasmid wash buffer, centrifuging and discarding the flow through after each step. BAC DNA was eluted by adding at least 10μl of the provided DNA elution buffer to the column and centrifuged to a clean microcentrifuge tube.

2.1.4. Restriction enzyme digests

Restriction enzyme digests were performed according to manufacturer’s instructions. The required amount of DNA (calculated to give greater than 10ng of the shortest fragment required for visualisation on a ultraviolet (UV)-transilluminated agarose gel) was digested in a mix prepared with approximately 10U/μl of restriction enzyme per 1μg of DNA, 10x restriction buffer (supplemented with additional 100x BSA 10μg/μl if required) to a final volume of 5-10μl adjusted with water. DNA was digested from between 1 hour to overnight (ON) usually at 37°C, depending on the characteristics and efficiency of the restriction
enzyme adopted (https://www.neb.com/tools-and-resources/interactive-tools/enzyme-finder). The digested products were checked by gel electrophoresis on 0.7-2% w/v agarose gels depending on the size of the generated fragment/s (agarose (Invitrogen), 1x TAE buffer, 0.5% ethidium bromide (Sigma)) and visualised on a UV transilluminator. Fragments required for use in downstream ligation reactions were visualised using a blue light box prior to cutting out of relevant bands.

2.1.5. DNA Sequencing

DNA sequencing was performed by Source Biosciences. Clean DNA in 5μl water was provided at the following concentrations: plasmid 100ng/μl; PCR product 1ng/μl/100bp; primers 3.2pmol/μl.

2.1.6. Gel extraction

Digested DNA products were cut from the agarose gel on a blue box and the gel fragment weighed. DNA was purified using GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) according to the manufacturer’s instructions. 1μl/mg of capture buffer type 3 was added to the excised gel and incubated at 60°C for 15–30 minutes in a benchtop microcentrifuge tube shaker until the agarose was completely dissolved. The sample mix was then transferred onto a GFX MicroSpin column and collection tube, incubated at room temperature for 1 minute and spun at 16000g for 30 seconds. The flow through was discarded by emptying the collection tube and the GFX MicroSpin column washed with 500μl of wash buffer type 1. DNA was eluted by centrifugation in 30-50μl of IDTE (pH8.0).
2.1.7. DNA Ligation

1-2μg vector and the appropriate molar quantity of insert were digested by appropriate restriction digest in a 25μl reaction volume. The completion of the digest was checked by running 0.5-1μl of digested and undigested DNA on an analytical agarose gel and either increasing incubation time or adding more enzyme added if restriction digest was incomplete. Once digested samples were loaded onto an 8 lane preparative mini-gel leaving single lane gaps between samples. This gel was run overnight at 15-20V using a pump to circulate the TAE buffer. After gel extraction of both vector and insert, the vector was treated with Antarctic Phosphatase (NEB) for 15mins and cleaned over a minelute column (Quiagen). Fragments were checked on an analytical agarose gel and quantified by Qubit HS DNA fluriometry. Ligations were set up in 10μl final volumes using a ratio of 1:1 inserts (if multiple inserts) and 1:3 vector:insert for 2-18 hours at room temperature. Negative (vector self) and positive (0.5μg 10kb+ ladder (Invitrogen)) controls were also ligated. Successful ligations were checked by running 5μl on an analytical gel and compared to un-ligated vector. If a significant shift in migration pattern was seen then 1-2μl of ligation mix was transformed into NEB5α high efficiency E coli (NEB).

2.1.8. TALEN Assembly by Golden Gate

TAL Effector Nucleases (TALENs) were constructed as a pair of plasmids encoding DNA binding proteins attached to the obligate dimeric restriction endonuclease FokI. The DNA binding region of the TALEN contains a series of repeat variable domains (RVDs) joined together into an array to generate a repetitive protein structure that enters into the major groove of DNA. Two critical amino acids interact with the base pairs in a sequence specific manner (NH-guanine, NI - adenine, NG-thymine, HD-cytosine). Construction of a specific
TALEN array is complicated by the repetitive sequences of the RVDs. Golden gate assembly was chosen as a relatively robust and scalable means of assembling the RVDs with the sequence specific array is then cloned into a generic expression vector containing a promoter and FokI endonuclease (Voytas golden gate kit version 2.0, Addgene) (Cermak et al., 2011).

Firstly a library of all required repeat variable domains (RVD), intermediate backbones, last repeats and promoter/FokI backbones were stored as glycerol stocks, midi-prepped (Qiagen), verified by both restriction digest & sequencing and 100ng/μl working stocks made. TALENs were designed in silico using the Cornell University online tool (https://tale-nt.cac.cornell.edu/node/add/talen). Binding or cutting sites avoided known methylation marks as annotated in Ensembl for H1 hESC. Genomic DNA of the target hPSCs cells was sequenced and heterozygous SNPs checked and avoided. TALENs were designed to minimise the potential for off-target cutting; each monomer was made sufficiently long to be genome specific and the FokI nucleases used were obligate heterodimers.

RVDs were assembled following cleavage of the component parts by the type 2S restriction enzyme BsaI (NEB) leaving sequence-specific cohesive ends which were re-ligated into half arrays and integrated into an intermediate fusion vector of the appropriate length (“pFus”) in a single 20μl reaction containing: 100ng (1μl) each plasmid, 1μl BsaI-HF, 1μl T4 ligase, 2μl fresh T4 ligase buffer, 1μl BSA (@2mg/ml) and cycled for 10x 37°C (5mins)/16°C (10mins) followed by 50°C (5mins) and 80°C (5mins). Unligated linear DNA was then digested by adding 1μl 10U/μl plasmid safe ATP dependent DNAase (Epibio) plus 1μl 10mM dATP for 1 hour at 37°C. 5μl of ligation reaction was used to transform chemically competent DH5α E. coli (Invitrogen), which after outgrowth were selected overnight on agar plates containing spectinomycin 50μg/ml, 40μg/ml X-gal and 10μM IPTG. Three white intermediate clones
were checked by directly inoculating colonies into the following screening PCR reaction (water 11.5μl, 10x Buffer 2μl, 5xQ solution 4μl, 10mM dNTPs 0.4μl, 2μl 0.25μM primer mix pCR8_F1 (TTGATGCTGGCAGTTCCCT) pCR8_R1 (CGAACCGACAGGCTATGT), 0.1μl HotStarTaq (Quiagen): Conditions 95°C 15mins, 35 cycles of 94°C 45s / 55°C 45s / 72°C 1min 45s, final extension 72°C 10min). Correct clones are identified by the presence of a band of the correct RVD array size on the background of a smear and laddering effect. A successful colony was picked and grown overnight for glycerol stocks and mini-prepped plasmid DNA was checked by restriction digest with AflII and XbaI to release the array of fused repeats and by sequencing using the primers pCR8_F1 and pCR8_R1.

Two additional backbone plasmid kits were purchased from Addgene containing different promoter/heterodimeric Fokl combinations. Backbone “A” was driven by a CMV promoter and was linked to a DD/RR heterodimeric Fokl. Backbone “B” was driven by the powerful pCAGS promoter and was linked to the KKR/ELD heterodimeric Fokl additionally engineered to harbour the “Sharkey” mutation to increase cutting efficacy. The intermediate half arrays were assembled into the final TALEN through a second golden gate reaction by mixing the two half arrays (150ng each), a C-terminal last-repeat (150ng) and a promoter/Fokl destination plasmid (75ng) with 1μl fresh Esp3I (ThermoFisher), 1μl T4 ligase (NEB), 2μl fresh T4 ligase buffer in a 20μl total reaction volume. A single cycle reaction could be used for this step: 37°C 10mins, 16°C 15mins, 37°C 15mins, 80°C 5mins. 5μl of reaction transformed into DH5α chemically competent Escherichia coli and following outgrowth were plated on agar plates supplemented with ampicillin 50μg/ml and X-gal/IPTG as above.

Three white colonies were checked by inoculation into the following PCR: water 11.5μl, 10x Buffer 2μl, 5xQ solution 4μl, 10mM dNTPs 0.4μl, 2μl 0.25μM primer mix TAL_F1 (TTGGCGTCGGCAACAGTGG) TAL_R2 (GGCGACGAGGTGGTCGTTGG), 0.1μl HotStarTaq
(Quiagen): Conditions 95°C 15mins, 35 cycles of 94°C 45s / 55°C 45s / 72°C 3m 45s, final extension 72°C 10min. A smear with or without faint bands around 2-3kb indicated the formation of the complete TALE array. A successful colony was picked and grown in liquid culture overnight for preparation glycerol stocks and mini-prepped plasmid DNA was checked by restriction digest with StuI and AatII to release the array of fused repeats and by sequencing using the primers seqTALEN 5-1 (CATCGCGCAATGCACTGAC) and TAL_R2. Of note, the final RVD arrays were too repetitive and long to sequence in their entirety, thus sequencing of the intermediate half arrays was essential.

2.1.9. BAC Recombineering

A homologous recombination vector targeting ETV6 intron V was constructed by BAC recombineering (RP11_418C2 Source Biosciences) (Fu et al., 2010). BAC-containing DH10B E. coli were grown overnight in 1.4ml of LB/chloramphenicol (10μg/ml). All cultures were conducted in microcentrifuge tubes with a single piecing in the lid to allow aeration, using a bench-top Eppendorf® microcentrifuge tube shaker (950rpm). On day 2, 40μl of overnight culture were outgrown by inoculation into 1.4ml LB/ chloramphenicol (10μg/ml) at 37°C. These cells were made recombineering proficient by electroporation with up to 100ng of the inducible pSC101-BAD-gbaA-tet plasmid (recombineering plasmids provided by Francis Stewart, University of Dresden). All transformations were prepared by washing the cells washing 3 times with 1ml ice-cold ddH₂O taking care not to lose cells during each wash. The final pellet was gently re-suspended in 40μl ice-cold water, mixed with DNA and electroporated using the Ec1 setting (BioRad). Cells were recovered at 30°C (the psc101 plasmid is lost at higher temperatures) in 1ml SOC for 2 hours without selection and after
recovery 100μl of the culture was inoculated into a new tube containing 900μl LB chloramphenicol (10 μg/ml) and tetracycline (4 μg/ml) and incubated at 30°C for 20 hours.

The homology arms for the ETV6 targeting vector were then sub-cloned from the BAC by gap-repair into a generic targeting backbone containing a diphtheria toxin (DTA) eukaryotic negative selection cassette (p15a-DTA-AmpR). Cells were prepared by inoculating 40-100μl of overnight culture into 1.4ml LB/ chloramphenicol (10 μg/ml) and tetracycline (4 μg/ml) cultures and outgrowing for 2-4 hours at 30°C until the LB started to turn turbid. Growing for long enough to provide a sufficient number of *Escherichia coli* is essential. XmaI linearised p15a-DTA-AmpR backbone plasmid was amplified by Phusion® high-fidelity PCR (NEB) using PAGE-purified primers (Integrated DNA Technologies - IDT) designed with 50bp homology arms aligned in an inverted format to facilitate gap-repair from the BAC (5’:

GCTATATATATAATATTATATACGAATATAATTATGGCTGACTTAATAAGATGATCTTCT

TGAGATCG; 3’:

CGTAAAAAGTTCTGTCATTAACGTGAATAGCCATAAAAAAGCCCTTTAATCTCTTACCAATGCTTAAATCA

GTGAGG) (Bold = homology arms). PCR conditions; 25μl PCR mix: 5x Phusion® buffer (5μl), 100% DMSO 1.25μl, 10mM dNTP 0.5μl, Primers (50pmol/μl) 0.5μl each, Phusion® polymerase 0.25μl, linearised template (3ng/μl) 0.5μl; cycling conditions: 98°C 30s, 35 cycles 98°C 10s / 54°C 30s / 72°C 1min 30s, final extension 72°C 10mins. The PCR product was then treated with DpnI to digest methylated plasmid template further reducing the risk of background transformation and cleaned over a Minelute® column (Qiagen). Following induction by the addition of 20μl 10% arabinose the *E. coli* were grown at 37°C for 40mins prior to electroporation with 400ng of cleaned ETV6-specific p15a-DTA-AmpR PCR product re-suspended in water. DNA and arabinose negative controls were also included and are essential. The cells were recovered in SOC for 2 hours at 30°C then overnight cultures in 1.4ml LB tetracycline (4μg/ml), ampicillin (50μg/ml) and streptomycin (200μg/ml) at 30°C to
counter-select for transformation with p15a plasmid PCR template (which contains an rpsL streptomycin sensitivity cassette).

The following day the E. coli were arabinose induced and electroporated as above with 200ng of a column-purified PCR product of the rpsL-GentaR positive/negative selection cassette containing 50bp homology arms targeting the cleavage point of ETV6 CRISPR gRNA (Primers F: CACCATCTCAATTAGGGCCTCTCAAGGCTCTTGAGGGCAATTGGAGGCATTAATCTAATTTTGTTTGAC; R: AGCGGCTCATAAACCAAGGCTATTAGGACTCTCATCAGACAAGTAGGTGGCCGTACTTGAGGG) (Bold = homology arms). PCR conditions; 25μl PCR mix: 5x Q5® buffer (5μl), 10mM dNTP 0.5μl, Primers (25μmol) 0.5μl each, Q5® polymerase polymerase 0.25μl (NEB), template plasmid DNA (50ng/μl) 1μl; cycling conditions: 98°C 30s, 35 cycles 98°C 10s / 72°C 30s / 72°C 45s, final extension 72°C 2mins. Product 1102bp. The rpsL/gentaR cassette is on a pR6K plasmid backbone that can only be grown in Pir116+E. coli (kindly provided by Dr Andrew Smith) preventing background transformation of PCR template plasmid DNA. The cells were recovered in SOC for 2 hours at 30°C then cultured overnight in 1.4ml LB with tetracycline (4μg/ml), ampicillin (50μg/ml) and gentamicin (4μg/ml). Note gentamicin must be fresh and activity checked with a negative control.

The following day the rpsL/GentaR positive/negative selection cassette was replaced with the RUNX1 knock-in cassette. The cassette was designed in silico and commercially synthesised as in two sequence-verified fragments (GeneArt™, Invitrogen). These fragments were released from plasmid DNA by XhoI SbfI double digest and ligated together into a pBluescript plasmid, prior to XhoI ligation into a pR6K-AmpR plasmid backbone in pir116+E coli (to prevent background transformation during recombineering). The knock-in cassette
flanked by 50bp homology arms (included in the synthesis) targeting the knock-in site was released from a pR6K plasmid backbone by Ascl NotI double digest. Following arabinose induction the *E. coli* were electroporated with 600ng of linearised knock-in construct. DNA and arabinose negative controls were also included. The cells were recovered in SOC for 1 hour at 37°C then cultured overnight in 1.4ml LB with ampicillin (50μg/ml) and kanamycin (5μg/ml) at 37°C to remove the psc101 recombineering plasmid.

The next day plasmid DNA was prepared from the cultures and retransformed into chemically competent DH5α *E. coli*, plated on Kan/Amp agar and clones verified by restriction digest and sequencing across all recombineering junctions and the knock-in cassette. A verified clone was used to prepare tissue culture (TC)-grade plasmid DNA by Maxiprep (Qiagen), linearised with *Xho*I (NEB) and cleaned over a Minelute® (Qiagen) column into TC-grade water for use in hPSC nucleofection (Lonza)

### 2.1.10. Southern Blot Hybridisation

Successful knock-in (KI) clones were checked for successful targeting by Southern blot hybridisation looking for a change in the restriction digest pattern at the *ETV6* intron V locus. Two methods of genomic DNA (gDNA) preparation and digestion were used for i) bulk hPSC cells grown in 6 well plates or ii) those screened after growth in 96 well plates.

Genomic DNA was prepared from hPSCs grown on 6 well plates was prepared by phenol chloroform extraction. A confluent well of hPSCs was mechanically scraped into 750μl of media and transferred into a screw-cap 1.5ml microcentrifuge tube. After centrifugation at 400g for 3mins, the supernatant was carefully removed and the pellet washed in 500μl PBS. Following further centrifugation the pellet was re-suspended in 677.2μl isolyis buffer
(Trizma HCl pH7.5 20mM, EDTA 50mM, KCl 50mM, NaCl 100mM, SDS 0.5%, Igepal 0.1%, NP-40 0.1%) supplemented with 2.8μl RNAaseA (100mg/ml) (Sigma) and 7μl Proteinase K (20mg/ml) (Invitrogen) and incubated at 65°C overnight. gDNA was phase extracted by the addition of 750μl phenol chloroform (Invitrogen) and mixed fully, but gently by inversion. The tube was briefly spun and the entire emulsion transferred to a pre-spun phaselock® (Eppendorf) tube. After centrifugation at 16000g for 5mins the aqueous phase was removed to a new Safelock® (Eppendorf) microcentrifuge tube and a 2.5x volume of 100% ethanol added. The solution was mixed carefully to reveal a visible web of gDNA. This was pelleted by centrifugation at 16000g for 10mins, followed by a wash in fresh 70% ethanol. After a final spin all the remaining ethanol was carefully removed with a 20μl pipette and the pellet dried at 37°C for 5mins. The pellet was re-suspended overnight at room temperature in 50-100μl IDTE (pH8.0) spiked with RNAaseA. gDNA was stored at 4°C and handled gently to avoid shearing, particularly when probing for large DNA fragments. 2-4μg of gDNA was digested with 60 units (3μl) of the appropriate restriction enzyme/buffer in a 50-100μl total volume. Depending on the half-life of the enzyme an additional 40U (2μl) was added after 6-8 hours and the digest continued overnight. DNA integrity was analysed by running 2.5-5μl of the digest on a 0.8% agarose gel. Once the digest had been confirmed to work the digested gDNA was cleaned by addition of 10% 3M sodium acetate and precipitated in 2.5x volume of 100% ethanol and washed as above. The pellet was re-suspended in 10-20μl IDTE overnight. The digest was quantified by Qubit BR and 1μg loaded into a well of a 0.8% agarose gel. A λHindIII ladder (NEB), melted at 65°C was included as a control. The gel was run overnight at 15-20V using a circulating pump to prevent buffer exhaustion. The next day the gel was briefly visualised on a UV transilluminator next to a fluorescent ruler prior to DNA transfer (see below).
Candidate clones grown in 96 well TC plates were grown to confluence, washed once with 100μl PBS and lysed in 50μl cell lysis buffer (NaCl 100mM, TrisHCl pH 7.5 50mM, EDTA 10mM, SDS 0.5%, proteinase K 0.8mg/ml) overnight at 65°C in a sealed humid box (Rudolph et al., 1994). DNA was precipitated by direct addition of 50-60μl isopropanol taking care to carefully vortex and flick-spin the plate at every step. After 30mins the plate was centrifuged at 2000g (faster speeds were found to crack the TC plastic) for 20mins at room temperature. The plate was carefully inverted to remove the supernatant, before 3 careful washes with 100μl 70% ethanol administered with a multichannel pipette. The plates were dried between each inversion on paper towels to avoid cross contamination between wells. The pellets were dried in a hybridisation oven at 37°C for 5-10mins until all the ethanol had evaporated, but taking care not to over-dry the pellets. Pellets were directly re-suspended in 30μl restriction enzyme mixture (10x buffer, 0.2mg/ml BSA, spermidine 1mM, RNAaseA 20μg/ml, 10U enzyme/30μl) and digested in an incubator in a sealed humid box overnight. The next day 5μl of concentrated restriction enzyme mix was added to each well (10x buffer, 0.2mg/ml BSA, spermidine 1mM, RNAaseA 20μg/ml, 10U enzyme/5μl) and further digested in an incubator in a sealed humid box for 4-5 hours. After this time the plate was removed and incubated in a hybridisation oven for 1-2 hours to force evaporation of restriction enzyme mix bringing the total volume to approximately 20-30μl per well. 3.5μl of 10x loading buffer was added and the samples are then ready for loading onto a 200 well 0.8% agarose gel alongside a melted λHindIII ladder (0.5μg/lane) (NEB). 200 well agarose gels were run at 110V for 2-3 hours before being briefly visualised on a UV transilluminator next to a fluorescent ruler.

After visualisation the gel was depurinated in 250mM HCl for 10mins, rinsed in ddH20, denatured in 2x 15min 0.5M NaOH/1.5M NaCl, rinsed in ddH20, neutralised in 1.5M NaCl/TrisHCl 0.5M pH7.5, rinsed in ddH20 and equilibrated in 20xSSC (Severn Biotech) for
10mins. The gel was carefully transferred to a capillary transfer apparatus onto a wick of 3MM paper over a reservoir of 20x SSC. The top right of the gel was nicked and the edges of the gel sealed with parafilm®. A positively charged nylon membrane (Roche) was cut to be 1cm wider and longer than the gel, wetted in ddH2O and then 20xSSC and placed over the gel taking care to remove any bubbles. The nylon was nicked in the top right corner, covered with wetted 3MM paper and a large stack of weighted paper towels added to the top. The transfer apparatus was covered with Saran® film for protection and left to transfer overnight.

The following morning the nylon was removed with forceps, UV crosslinked (120mJ Stratalinker™) and rinsed in ddH2O. The gel was washed in TAE/Ethidium bromide and visualised to ensure transfer of DNA was complete. The nylon membrane was pre-hybridised at 42°C in 15-25ml DIG Easyhyb® buffer (Roche) for a minimum of 60mins in a rotating hybridisation bottle. Large or multiple blots were separated by a layer of nylon mesh. A Southern probe (SP2) was produced using HotStar® Taq polymerase (Qiagen) from 500pg ETV6 BAC. Primers SP2-F: CCCATCTGAGGGACTGTG (Tm 56.9) SP2-R: GTATCCGCAGCTAAAGGACTTG (Tm 56.4). PCR mix: 10x buffer 5μl, dNTP (10mM) 1μl, primers (2.5μM) 2.5μl, HotStar Taq 0.25μl and template DNA (0.5ng/μl) 1μl. Cycling conditions: 95°C 15min, 94°C 30s, 56°C 30s, 72°C 2min, repeat 35 cycles, 72°C 10min. The SP2 probe was labeled using 32P-dATP to maximise sensitivity (Agilent). 200 well blots required 200ng probe, whereas a 30 lane gel needed no more than 100ng. 1-200ng SP2 clean PCR product was mixed with 10-20μl random nonamers made up to 34-68μl with water and the DNA melted at 95°C for 5mins. Then 10-20μl ATP deficient buffer, 5-10ul 32P-dATP (Hartmann SRP-203) and 1-2μl Klenow fragment were added and the mixture incubated at 37°C for 45mins. The preparation was column purified (GE microspin) and the activity flow through containing the probe compared to that of the column. Good labeling was typically a 10-fold
enrichment of radioactivity in the flow though. 1.5μg λHindIII ladder (NEB) was labeled in the same way and a 1:1000 dilution made in IDTE pH8.0. 5-20μl of this dilution (sufficient to produce 10-50 counts per second when held close to a Geiger Muller tube) was added to the labeled Southern probe. The combined probes were denatured for 5min at 95°C, before being carefully added to the hybridising buffer. The membrane was left to hybridise overnight at 42°C in a rotating bottle. The next day the membrane was washed twice in low stringency buffer (0.5xSSC/0.1%SDS) for 5min (warmed to re-dissolve any SDS crystals) and twice in high stringency buffer (0.1XSSC/0.1%SDS) at 68°C for 15min. If significant counts were still detectable on the nylon then the membrane was given a third high stringency wash in a large square box. The membrane was then wrapped in polythene, excess buffer removed and the edges sealed. The nylon was exposed to a phosphor screen in a cassette for 12-36 hours prior to imaging on a Typhoon phosphorimager.

2.1.11. Bulk total RNA isolation

Cells were harvested into a screw cap microcentrifuge tube with 0.8-1ml TRIZol® (Invitrogen) mixed and frozen at -80°C. On thawing 200μl chloroform was added per 1ml TRIZol® and the sample mixed vigorously. After centrifugation at 4°C for 10mins the aqueous phase was removed to a fresh Safelock® tube and RNA precipitated by the addition of an equal volume of isopropanol. After mixing, the RNA was pelleted by centrifugation at 16kg for 15mins, washed in 70% ethanol and the resulting pellet re-suspended in 10μl ddH2O at 55°C for 10mins. The final RNA concentration was measured using a spectrophotometer (NanoDrop.ND-1000, Lebtech International).
2.1.12. cDNA preparation

cDNA was prepared using SuperScript® III (Invitrogen). 1μg RNA in 8μl water was first treated with DNAase (Promega) to remove contaminating gDNA (1μl DNAase buffer, 1μl DNAase). After 30 mins at 37°C the reaction was stopped by the addition of 1μl stop solution. This reaction proceeded directly to cDNA synthesis, first with the addition of 1μl random primers (100ng/μl) (Invitrogen) and 1μl dNTP (10mM), which was denatured at 65°C for 5mins. The mixture was cooled on ice for 1 minute, then 4μl 5x first strand buffer, 1μl dithiothreitol (DTT 0.1M), 1μl RNaseOUT™ (Invitrogen 40U/μl) and 1μl SuperScript® III added. The mixture was incubated at 25°C for 5mins, 50°C for 30-60mins and 70°C for 15mins. A 1:200 dilution was used for downstream applications.

2.1.13. Quantitative Real-Time PCR

Quantitative PCR was performed in a 20μl reaction using 10μl SYBR® Green master mix (Applied Biosystems) on 2μl cDNA using a 200nM primer mix (OCT3/4 F: GAAGGAGAGCTGGAGCAAA, R: CTTCTGCTTCAGGAGCTTG; NANOG F: GATTTGTGGGCTGAGAAA, R: CAGGGCTGTCCTGAATAAGC) and normalized to 8-ACTIN using TaqMan® probe (1μl in 20μl reaction) Hs99999903_M1 (ThermoFisher).

2.1.14. Statistics

Data are shown as mean values with standard deviation (SD) if not stated otherwise. Statistics were performed using a non-parametric Mann-Whitney test or Kruskal-Wallis/Dunns test for multiple comparisons.
### 2.1.15. DJ rearrangement

Analysis of DJ\(_\mu\) status was performed by Dr Charlotta Boiers. Cells were FACS sorted together with control populations directly into Kapa Express DNA extraction Kit (KAPABiosystems) and DNA extracted according to manufacture’s recommendation. RT-PCR was performed mixing DNA with Maxima SYBR Green (ThermoFisher) and the primers listed in table 2.2 (van Dongen et al., 2003). Two reactions were performed – one with DJ\(_\mu\)1-6 and the J\(_\mu\) consensus primers and one with DJ\(_\mu\)7 and the J\(_\mu\) consensus primer, the latter also rise gives to germline bands. Samples were run with the following program: 15 min hold at 95°C followed by 50 cycles of 30 sec at 95°C, 30 sec at 60°C and 45 sec at 72°C and ending with one cycle of 15 sec at 95°C, 20 sec at 60°C and 15 sec at 95°C. The PCR product was run on a gel (2% Agarose, ethidium bromide) with TBE buffer (ThermoScientific).

<table>
<thead>
<tr>
<th>Primers - DJ rearrangement</th>
</tr>
</thead>
<tbody>
<tr>
<td>DJ(<em>\mu)1-6 - J(</em>\mu) reaction</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>DJ(<em>\mu)7 - J(</em>\mu) reaction</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

*Table 2.2: Primers used for analysis of DJ\(_\mu\) rearrangement in IL7R progenitor cells.*

### 2.1.16. Single Cell Real-Time PCR (Fluidigm)

Cells were directly sorted (single cells or 25 cells/population) into 4\(\mu\)l of lysis buffer (65\(\mu\)M dNTP (TaKaRa), 0.4% NP40 (Sigma), 2.4mM DTT (Invitrogen), 0.5U/\(\mu\)l RNaseOUT (Invitrogen) in nuclease-free water). Dr Charlotta Boiers kindly performed the single cell qPCR reactions. cDNA synthesis and target specific pre-amplification was done using Cell Direct one-step qRT-PCR kit (Life Technologies). Pre-amplification mastermix was added to
each well consisting of 6.25µL of 2X reaction buffer, 1µl of SuperScript III RT/Platinum Taq mix and 1.5µl of TaqMan assay mix. TaqMan assay mix was prepared by mixing equal volume of all target specific primers (Table 2.3) (Life Technologies). No-RT controls were treated in parallel except that no SuperScript III RT enzyme was included with Taq polymerase (Invitrogen) included instead. The PCR conditions were 60 min at 50°C, 2 min at 95°C and 25 cycles of 15 sec at 95°C and 4 min at 60°C. For samples with 25 cells, only 22 cycles were run. Pre-amplified product was diluted 1 to 5 and loaded onto a 48.48 Fluidigm® chip together with Taqman universal MasterMix (Life Technologies) and the Taqman assays listed in table below with the appropriate loading reagents according to manufacturer’s instructions (BioMark 48.48 Dynamic array platform (Fluidigm)). Bulk data were normalised to GAPDH to calculate relative expression and all values were displayed as CT values (not normalised) in a heatmap. In some single cell sorts, index sorting was used and populations separated based on the index analysis. Cells not expressing GAPDH were not analysed further. In the analysis of ETV6-RUNX1⁺ cells, only cells expressing ETV6-RUNX1 were considered. Only cells expressing IL7R were analysed for the IL7R progenitor and only cells expressing CD19 considered when analysing proB cells.
<table>
<thead>
<tr>
<th>Assay</th>
<th>Assay Id</th>
<th>Used for PCA</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCR9</td>
<td>Hs00258535 m1</td>
<td>x</td>
</tr>
<tr>
<td>CD10</td>
<td>Hs00153510 m1</td>
<td>x</td>
</tr>
<tr>
<td>CD19</td>
<td>Hs00174333 m1</td>
<td>x</td>
</tr>
<tr>
<td>CD33</td>
<td>Hs01076281 m1</td>
<td>x</td>
</tr>
<tr>
<td>CD34</td>
<td>Hs00990732 m1</td>
<td>x</td>
</tr>
<tr>
<td>CD38</td>
<td>Hs01120071 m1</td>
<td>x</td>
</tr>
<tr>
<td>CD62L (SELL)</td>
<td>Hs00174151 m1</td>
<td>x</td>
</tr>
<tr>
<td>CD7</td>
<td>Hs00196191 m1</td>
<td>x</td>
</tr>
<tr>
<td>CRLF2</td>
<td>Hs00845692 m1</td>
<td>x</td>
</tr>
<tr>
<td>CSF1R (M-CSFR)</td>
<td>Hs00911250 m1</td>
<td>x</td>
</tr>
<tr>
<td>CSF2RA (GM-CSFR)</td>
<td>Hs00538900 m1</td>
<td>x</td>
</tr>
<tr>
<td>CSF3R (G-CSFR)</td>
<td>Hs0114427 m1</td>
<td>x</td>
</tr>
<tr>
<td>DNTT</td>
<td>Hs00172743 m1</td>
<td>x</td>
</tr>
<tr>
<td>TCF3 (E2A)</td>
<td>Hs00413032 m1</td>
<td>x</td>
</tr>
<tr>
<td>EBF1</td>
<td>Hs00395513 m1</td>
<td>x</td>
</tr>
<tr>
<td>EPOR</td>
<td>Hs00181092 m1</td>
<td>x</td>
</tr>
<tr>
<td>ETV6</td>
<td>HS00231101 m1</td>
<td>x</td>
</tr>
<tr>
<td>ETV6-RUNX1*</td>
<td>F TEL: CTCTGTCTCCCCGCCTGAA</td>
<td>P TEL: TCCCAATGGGCATGGCGTGC</td>
</tr>
<tr>
<td>FLT3</td>
<td>Hs00174900 m1</td>
<td>x</td>
</tr>
<tr>
<td>GATA1</td>
<td>Hs01085823 m1</td>
<td>x</td>
</tr>
<tr>
<td>GATA3</td>
<td>Hs00231122 m1</td>
<td>x</td>
</tr>
<tr>
<td>GAPDH</td>
<td>Hs09999905 m1</td>
<td>x</td>
</tr>
<tr>
<td>IGHM</td>
<td>Hs00378435 m1</td>
<td>x</td>
</tr>
<tr>
<td>IKZF1 (IKAROS)</td>
<td>Hs00172991 m1</td>
<td>x</td>
</tr>
<tr>
<td>IL2RG</td>
<td>Hs00953624 m1</td>
<td>x</td>
</tr>
<tr>
<td>IL3Rα</td>
<td>Hs00608141 m1</td>
<td>x</td>
</tr>
<tr>
<td>IL7R</td>
<td>Hs00902334 m1</td>
<td>x</td>
</tr>
<tr>
<td>KIT</td>
<td>Hs00174029 m1</td>
<td>x</td>
</tr>
<tr>
<td>LIN28B</td>
<td>Hs01013729 m1</td>
<td>x</td>
</tr>
<tr>
<td>MPL</td>
<td>Hs00180489 m1</td>
<td>x</td>
</tr>
<tr>
<td>MPO</td>
<td>Hs00165162 m1</td>
<td>x</td>
</tr>
<tr>
<td>NOTCH1</td>
<td>Hs01062011 m1</td>
<td>x</td>
</tr>
<tr>
<td>PAX5</td>
<td>Hs00277134 m1</td>
<td>x</td>
</tr>
<tr>
<td>PRETA</td>
<td>Hs00300125 m1</td>
<td>x</td>
</tr>
<tr>
<td>SPI1 (PU.1)</td>
<td>Hs02786711 m1</td>
<td>x</td>
</tr>
<tr>
<td>RAG2</td>
<td>Hs00379177 m1</td>
<td>x</td>
</tr>
<tr>
<td>TCF7</td>
<td>Hs00175273 m1</td>
<td>x</td>
</tr>
<tr>
<td>VWF</td>
<td>Hs00169795 m1</td>
<td>x</td>
</tr>
</tbody>
</table>

Table 2.3: List of qPCR probes used for single cell analysis. Those used in the principal component analysis (PCA) are marked. * (Gabert et al., 2003)
2.1.17. Low cell number RNA Sequencing

22-400 cells were directly sorted into 800μl TRizol® and frozen at -80°C. All pre-PCR steps were performed in a clean PCR hood after UV treatment. RNA was extracted from the aqueous phase after addition of 160μl chloroform, precipitated with an equal volume of isopropanol supplemented with 20μg linear polyacrylamide and washed once in fresh 80% ethanol. The RNA pellet was re-suspended in 3μl DNAase/RNAase free water at 50°C for 10 minutes and then kept on ice or stored at -80°C. 0.5-1μl of RNA was quantitated using an Agilent Bioanalyser RNA 6000 pico chip.

100pg (or 1μl of stock if unquantifiable) of RNA used for first strand cDNA synthesis using SMARTER v3 (Clontech) before 16 cycles of amplification according to manufacturer’s instructions. cDNA was purified on Agencourt *AMPure*XP magnetic beads, washed twice with fresh 80% ethanol and eluted in 17μl elution buffer. 1μl cDNA was checked and quantified on an Agilent Bioanalyser high sensitivity DNA chip.

Sequencing libraries were produced using Illumina® Nextera XT tagmentation according to manufacturer’s instructions except using 150pg input cDNA, 5mins tagmentation and amplification 12 cycles using Illumina® XT 24 index primer kit. Libraries were cleaned using an equal volume (50μl) Agencourt *AMPure*XP magnetic beads and re-suspended in 20μl elution buffer. Libraries were checked and quantified on an Agilent Bioanalyser high sensitivity DNA chip (using quantification size range 150-2000bp) and by Qubit® dsDNA BR (Molecular Probes). Libraries were pooled to a normalised concentration of 1.5nM and sequenced on an Illumina® NextSeq 500 using 150bp paired end kit as per manufacturer’s instructions by Tony Brookes, Institute of Child Health, UCL.
2.1.18. Bioinformatic Analysis

Sequencing data was mapped and analysed by Dr Dapeng Wang and Javier Herrero, Bill Lyons Informatics Centre, UCL Cancer Institute. Raw reads were initially processed with PRINSEQ-lite v0.20.4 (Schmieder and Edwards, 2011) to trim low-quality reads. They were further aligned with TopHat2 v2.0.11 (Kim et al., 2013) on the GRCh37 human assembly and FPKM values for protein-coding genes were calculated with Cufflinks v2.2.0 (Trapnell et al., 2010) using uniquely mapping reads only. Genes with an FPKM < 5 in every sample were filtered out. The remaining ones were log-transformed using the formula log2 (FPKM+1).

Finally, expression values were normalised upon quantiles (Bolstad et al., 2003) prior to the PCA analysis. When building the PCA map based on the expression in primary samples only (FL CS17, FL CS21-22 and Adult BM), the FL CS17 IL7R Progenitor sample was detected as a clear outlier. A new map based on all other primary cells was used instead and the FL CS17 IL7R Progenitor sample was projected on the new map. When building the PCA map based on the expression of all wild-type samples (hPSC D10, hPSC D31 and the aforementioned primary samples), the FL CS17 IL7R Progenitor sample rested on the correct region of the PCA map. Gene Set Enrichment Analysis (GSEA) was performed with permutation type=gene set and the gene sets used were obtained from (Laurenti et al., 2013). In the 3D scatter plot (Figure 5.10) signature expression values correspond to the average FPKM for the genes that define each signature. GMP and CMP gene sets were joined to define the myeloid signature. Similarly, the EarlyB and ProB gene sets were joined to define the lymphoid signature.
2.2. Tissue Culture Protocols

2.2.1. Human Pluripotent Stem Cell Culture

2.2.1.1. Mouse Embryonic Fibroblasts

Laboratory stocks of passage 0 (P0) MEFs were thawed rapidly from liquid nitrogen storage in a 37°C water bath and washed in 10 ml MEF medium (DMEM (Invitrogen 41965-039), 10% fetal calf serum (FCS), 2mM L-glutamine) by centrifugation at 200g for 5 minutes. These cells were seeded into a T25 culture flask in MEF media and designated P1. MEFs were grown to 90% confluence and then passaged by incubation with 0.25% trypsin (w/v) /5mM EDTA (Sigma) for 5 minutes at 37°C. MEFs were split 1:3-1:5 until P4.

2.2.1.2. Mitotic Inactivation and Freezing MEF Aliquots

P4 MEFs were harvested from approximately fifteen T175 flasks by trypsinisation and re-suspended in 20ml MEF media. The re-suspended cells were mitotically inactivated using 30Gy X-irradiation. The P4 irradiated (P4i) MEFs were centrifuged at 200g, re-suspended in 10ml MEF media, live cells counted in a haemocytometer in the presence of 50% trypan blue and frozen aliquots prepared. Each well of a 6 well plate required 1.6 x 10^5 viable irradiated MEFs; thus, taking into account an empirically defined 10% loss of viability during freezing, vials of 1.08x10^6 P4i MEFs were prepared to cover 6 wells of a 6 well plate. Cells were frozen slowly at -80°C in 90% FCS/10%DMSO.
2.2.1.3. Preparation of feeder Layers for hPSCs

6-well tissue culture plates were treated overnight with 0.1% bovine gelatin (Sigma) diluted in PBS. Overnight gelatinisation is recommended to maximise viability of irradiated MEFs (R. Malladi, Oxford University D.Phil. Thesis). One vial of P4i MEFs were thawed, washed and reconstituted in 12ml MEF media as above and 2ml distributed to each of the 6-wells of the pre-gelatinised 6-well plate. The 6-well plates were gently agitated to achieve an even distribution of cells and transferred to a tissue culture incubator overnight for use the following day.

2.2.1.4. Human Pluripotent Stem Cells

Human pluripotent stem cells used in this project were the IPS lines mRNA induced foreskin fibroblast 1 and 3 (MIFF1 & 3) (University of Sheffield, http://hpscreg.eu/cell-line/UOSi001-B) and H1 ES cells (WiCell, European hESC registry (http://hpscreg.eu/cell-line/WAe001-A)). The Steering Committee for the UK Stem Cell approved all work on human ES cells (project number SCSCO3-1). MIFF1 and 3 hIPS lines were derived by Dr Christian Unger and kindly provided by Dr Christian Unger and Prof Peter Andrews (Centre for Stem Cell Biology, Sheffield University). Protocols for routine tissue culture of hPSCs were acquired from the DPhil. thesis of R. Malladi (Enver lab) and the University of Sheffield Centre for Stem Cell Biology (P. Andrews, personal communication). The Andrews group also kindly provided two days practical training in tissue culture techniques.

The viability of hPSCs is significantly reduced by single cell dissociation, although this can be improved in the presence of a ROCK inhibitor such as Y27632 (Cambridge Biotechnologies). Thus where possible hPSCs are handled gently and maintained as clumps of 20-100 cells.
Procedures such as freezing/thawing, passaging and obligate single-cell requirements such as FACS or nucleofection were conducted in the presence of 10μM Y27632.

2.2.1.5. Thawing hPSCs

A vial containing hPSCs was removed from long-term storage in liquid nitrogen and thawed rapidly in a 37°C water bath. The freshly thawed cells were gently transferred from the vial into a 15ml conical tube containing and the DMSO slowly diluted by the addition of 10ml of hES medium (DMEM/F12 (Invitrogen), 20% knock-out serum replacement (Invitrogen), 1mM L-glutamine (Invitrogen), 1% non-essential amino acids (Invitrogen), 8ng/ml recombinant human basic fibroblast growth factor (hbFGF) (Peprotech) and 0.1mM β-mercaptoethanol (Sigma)). 10μM Y27632 was added to the wash media during thawing and for the first 24 hours of culture. The hPSCs cells were centrifuged for 5 minutes no faster than 200g, the supernatant removed and the cell pellet gently re-suspended in 2.5ml of hES medium and transferred to one well of a 6-well plate of freshly prepared P4i MEFs. This plate was placed in a tissue culture incubator overnight without being disturbed. Due to the thermolability of hbFGF complete media changes are required on a daily basis until the colonies of hES cells were large enough to require passaging, which typically take seven to fourteen days after thawing.

2.2.1.6. Routine Culture and Passage of hPSCs on MEFs

Once the colonies of hPSCs cells were large enough, they were passaged by an enzymatic/mechanical method. Prior to passaging any grossly differentiated hES colonies were identified under a microscope and marked for removal using a sterile pipette tip. hES medium was then removed from the 6-well plates, the cells washed once with 2ml warm
DMEM/F12 and then incubated with 1ml warmed collagenase IV solution (1mg/ml Invitrogen in DMEM/F12). Collagenase IV was reconstituted from powder in DMEM/F12 to produce 10x concentrated aliquots and sterile filtered. Of note it rapidly inactivates at 4°C and required incubation at 37°C for good colony lifting. The plates were incubated at 37°C for 10 minutes before the collagenase was removed and the well gently washed with twice with 2ml DMEM/F12. 1ml hES media supplemented with 10μM Y27632 was added to the cells, which were then mechanically dissociated by cell scraping, gently triturated to produce clumps of 50-100 hPSCs and the desired fraction transferred into a 15ml falcon tube containing the appropriate amount of hES media supplemented with Y27632 (e.g. 12ml for 1x 6 well plate). hPSCs were typically passaged at a ratio of 1:6-8.

2.2.1.7. Maintenance of hPSCs in Matrigel/mTeSR1 or StemFit®

Matrigel (BD/Corning) and mTeSR1 (Stem cell technologies - SCT) were prepared according to manufacturer’s instructions. In brief, hPSC-certified matrigel was thawed overnight on ice at 4°C and then aliquoted into chilled 15ml falcons, using chilled pipette tips, on ice according to the advised dilution factor. Each 15ml falcon received enough matrigel to produce 12.5ml of 1x solution and was stored at -20°C. Matrigel aliquots were thawed on ice and re-suspended in 12.5ml of chilled DMEM/F12 using a chilled serological pipette. 1ml was then dispensed to each well of 2x6-well plates. These could be stored at 4°C for up to 7 days and used after 1 hour of warming at 37°C. Plates were not allowed to desiccate and matrigel was removed just prior to cell seeding. mTeSR1 was reconstituted by thawing the 5x supplement at 4°C overnight, adding to the supplied basal media. StemFit® (Aginomoto) media was used later in the project in place of mTeSR1 and was similarly made by thawing frozen supplements B and C and adding to basal media A. Media aliquots were warmed to
37°C prior to use to avoid thermo-depletion of hbFGF and were discarded after 7 days at 4°C.

2.2.1.8. **Passaging with Gentle Dissociation Reagent**

Enzymatic/mechanical passaging using dispase (SCT) was initially used, using manufacturer’s protocol. This involved 2x2ml washes with DMEM/F12 before and after a 7-minute incubation with dispase 1mg/ml at 37°C.

A simpler, cheaper and more effective means of passaging was later found to be the non-enzymatic EDTA-based gentle dissociation reagent (SCT). This reagent can be stored at room temperature and requires no wash steps. Cells are incubated at room temperature for 5-7mins, the reagent is replaced with mTeSR1/StemFit supplemented with Y27632 and then the cells are mechanically dissociated into clumps. This reagent was also found to produce reliable single cell suspensions when incubated for 13mins at 37°C following pre-incubation with 10μM Y27632 for 2 hours.

2.2.1.9. **Nuclear staining of fixed and permeabilised cells**

hPSC colonies were fixed in their wells using 4% paraformaldehyde and permeabilised for 2 hours (PBS w/o Ca/Mg, 1% BSA, 3%FCS, 0.1% TritonX100) before staining overnight at 4°C with 1:200 dilution OCT3/4 primary antibody (SantaCruz 9801). The following day the well was washed with PBS 3 times quickly and then for 2 15 minute washes. Secondary stain was with Cy™5 anti-Rabbit IgG (Jackson Labs) at 1:1000 dilution in permeabilisation buffer. After one hour this antibody was washed as above and the cells counterstained with DAPI prior to visualisation on a fluorescence microscope.
2.2.1.10. Karyotyping

Confirmation of normal karyotype by G banding was kindly performed by Duncan Baker (Sheffield Diagnostic Genetics Service) and Dr Christian Unger (Centre for Stem Cell Biology, University of Sheffield).

2.2.2. In vitro B cell Differentiation from hPSCs

2.2.2.1. OP9 Stroma Maintenance

Anna French and Lee Carpenter (University of Oxford, NHSBT) provided technical advice and training and helped assay reagents for the in vitro B cell differentiation protocol. The quality of the OP9 murine bone marrow stromal layer is critical to generation of CD34+/43+/45+ multipotent haematopoietic progenitors. The stroma were grown on 10cm dishes (BD) that had been treated with 0.1% bovine gelatin (Sigma) overnight. Cells were grown in 10ml of OP9 maintenance media (OP9-M: αMEM from powder (Invitrogen), 20% batch tested defined FCS (Hyclone SH30070.03), 100μM monothioglycerol (Sigma), 1x penicillin/streptomycin (Invitrogen)) until confluent. Serum batch is critical. Later in the project serum was sourced from Invitrogen (qualified FBS, 26140079). Cells were passaged every 4 days and failure to adequately trypsinise the stromal layer is a common cause of increased adiposity (unacceptable if >15%) resulting in stromal layer failure. Passaging involves two thorough washes with 5ml DPBS and incubation at 37°C with 5ml freshly thawed trypsin (diluted in DPBS to final concentration 0.05% with 0.5mM EDTA) for 5mins or until the stromal layer begins to visibly dissociate on swirling. At that point the trypsin is quenched with 5ml OP9-M media, the cells triturated, centrifuged at 300g for 5mins, resuspended in fresh maintenance media and split 1:4 onto gelatinised 10cm dishes. hPSCs
were seeded onto OP9 stroma that had been overgrown for 8-12 days post passage to produce a thick matrix. 5ml half media exchanges were conducted on the stroma on days 4 (and day 8 if necessary).

2.2.2.2. **Harvesting hPSCs for OP9 Co-culture**

hPSCs were grown to large colonies for 5-6 days post passage. One well of a 6 well dish was trypsinised and counted to estimate cell numbers. If colonies become over-confluent they can be roughly scored with a sterile pipette tip prior to collagenase treatment to create large sheets of cells. hPSCs were incubated at 37°C with fresh collagenase IV 1mg/ml for 10-90mins with occasional swirling to detach large whole colonies or sheets of cells. Once colonies had lifted they were gently washed by gravity 3 times in 10ml OP9 differentiation media (OP9-D: αMEM from powder, 10% batch-tested defined FCS, 100µM monothioglycerol, 1x penicillin/streptomycin). 1.5-2x10^6 hPSCs were seeded in 10ml of OP9-D media per 10cm dish of OP9 maintaining the hPSCs as large clumps or whole colonies. A complete media change was conducted on day 1 to remove dead cells and replaced with 20ml OP9-D media, half media changes (10ml) were conducted on days 4, 6 and 8 with CD34+ cells harvested on day 10. Timing of feeding is critical.

2.2.2.3. **Harvesting CD34+ cells**

The 20ml OP9-D media was removed and saved in a 50ml falcon to capture non-adherent cells. If more than one 10cm plate are combined the media is centrifuged at 300g for 5mins with the majority of supernatant removed to provide adequate space for the subsequent harvest. The OP9 matrix is first digested with 7ml fresh collagenase IV 1mg/ml in DEMEM/F12 at 37°C, which also swirled gently to wash off residual serum that can inhibit
subsequent trypsinisation. The collagenase was removed and discarded after 25mins and 7ml fresh 0.05% trypsin/0.5mM EDTA was added and incubated at 37°C. In order to prevent the layer from self-adhering into an indigestible ball the stromal matrix was scored at the start of trypsinisation. After 7mins the plates were gently swirled and examined; the reaction was quenched with OP9-D media when the stromal layer begins to separate. Breaking up the stromal layers and hPSC colonies can require aggressive trituration particularly if the enzymes are not fresh, the trypsin is quenched too early or the matrix coalesces into a ball. Trituration is performed first with a 10ml serological pipette and then with a 1ml pipette. Once triturated the media is strained through a pre-wetted 40μm nylon filter (BD Falcon) into the original 50ml falcon. The plate is then washed with 5ml MACS buffer (DPBS, 2% batch-tested defined FCS (Hyclone SH30070.03), 2mM EDTA) and additionally the filter washed twice with 2x1ml MACS buffer to maximise yield from the remaining stromal matrix.

2.2.2.4. Magnetic Separation of CD34+ Cells

After centrifugation at 300g for 5 minutes the supernatant is discarded and the cellular pellet re-suspended in 300-320μl MACS buffer (20μl of cells can then be removed for use as an unstained control if required). Cells are incubated at 4°C with 100μl Fc receptor blocking reagent (Miltenyi) for 5mins and then 100μl of CD34+ MACS beads (Miltenyi) for 30mins. A Miltenyi LS MACS column is loaded into an appropriate magnet, a 30μm sterile nylon cell strainer (CellTrics) added and both primed with 3ml MACS buffer. After incubation the cells are washed with 5ml MACS buffer, centrifuged at 300g at 4°C for 5mins and the supernatant completely removed. The pellet is re-suspended in 500μl MACS buffer and added to the centre of the cell strainer. The column is washed with 3x3ml MACS buffer before the column is removed from the magnet and the cells extracted by a rapid plunge using 5ml MACS
buffer. After centrifugation as above the cells are re-suspended in 0.5-1ml OP9-D media and counted.

**2.2.2.5. **MS5 Maintenance and Co-culture

MS5 stroma are maintained in pre-gelatinised T75 flasks (2 hours, 0.1% w/v porcine gelatin made from powder in DPBS and double autoclaved (Sigma)) in MS5 maintenance media (MS5-M: αMEM from powder, 10% batch-tested defined FCS, 1x penicillin/streptomycin). The cells grow rapidly and are split 1:4-1:12 when 80% confluent and not permitted to overgrow. Up to 2x10^5 MS5 cells are seeded in MS5-M media onto a pre-gelatinised 6 well plate 1-3 days prior to use for differentiation. 12.5-50x10^4 CD34+ cells (defined as small bright live cells counted by haemocytometer post-MACS separation) are added to each well in 2ml OP9-D media supplemented with IL3 10ng/ml, IL7 20ng/ml, Flt3L 50ng/ml, SCF 50ng/ml. 2ml of the above media (excluding IL3) is added at day 7. A further feed is performed at day 14 by carefully removing 2ml media from the side of each well (where there are fewer non adherent cells) and replacing with 2ml media (constituted as per day 7). Cells are harvested at day 21 by mechanical trituration with a 1ml pipette. Enzymatic treatment with both collagenase and trypsin can affect antigen detection by FACS and is best avoided at this analytical stage.

**2.2.3. In vitro differentiation assays**

**2.2.3.1. **MS5 subculture

To evaluate B lymphoid potential indicated populations were plated on MS5 stroma for B cell lineage. Media (αMEM (Invitrogen), with 1% Penicillin/Streptomycin, 1% L-glutamine
(2mM, Invitrogen), 1% 2-Mercaptoethanol and 10% FBS (HyClone) supplemented with cytokines were half changed weekly (e.g. 100μl of a 200μl media in a single well of a 96 well plate). Cytokines were added according to table 2.4. After 2 weeks cells were harvested and stained and analysed for expression of lineage markers.

**2.2.3.2. Semi-solid colony forming assays by methylcellulose**

Cell populations were plated in complete methylcellulose (MethoCult H4435, SCT) to assay myelo-erythroid output. Colonies were scored at day 12-14. Colonies could be picked and pooled and morphology evaluated on May-Grünwald (VWR International) and Giemsa (VWR International) stained cytospin slides.

**2.2.3.3. Liquid culture on Terasaki plates**

Myeloid output could be tested in non-adherent liquid culture. Bulk sorted populations were plated at a density of 20-40 cells/well in IMDM (Invitrogen) supplemented with 1% Penicillin/Streptomycin, 1% L-glutamine (2mM), 1% 2-Mercaptoethanol (0.1mM) and 20% FBS supplemented with cytokines (TPO, SCF, FLT3L, IL3, MCSF, GMCSF) according to table 2.4. Cultures were scored after 12-14 days. The cultures were picked and morphology and lineage evaluated on May-Grünwald (Merck) and Giemsa (Merck) stained cytospin slides.
<table>
<thead>
<tr>
<th>Growth factor</th>
<th>Supplier</th>
<th>Final concentration</th>
<th>Assay</th>
</tr>
</thead>
<tbody>
<tr>
<td>human Granulocyte Monocyte-Colony Stimulating Factor (hGM-CSF)</td>
<td>PeProTech</td>
<td>25ng/ml</td>
<td>Terasaki</td>
</tr>
<tr>
<td>human interleukin 3 (hIL-3)</td>
<td>PeProTech</td>
<td>25ng/ml</td>
<td>Terasaki</td>
</tr>
<tr>
<td>human Monocyte-Colony Stimulating Factor (hM-CSF)</td>
<td>PeProTech</td>
<td>25ng/ml</td>
<td>Terasaki</td>
</tr>
<tr>
<td>human Flt3 Ligand (hFlt3L)</td>
<td>PeProTech</td>
<td>25-50ng/ml</td>
<td>Terasaki/B/T</td>
</tr>
<tr>
<td>Human Stem Cell Factor (hSCF)</td>
<td>PeProTech</td>
<td>25-50ng/ml</td>
<td>Terasaki/B/T</td>
</tr>
<tr>
<td>human Thrombopoietin</td>
<td>PeProTech</td>
<td>25-50ng/ml</td>
<td>Terasaki/B/T</td>
</tr>
<tr>
<td>human interleukin 7 (hIL-7)</td>
<td>PeProTech</td>
<td>5-20ng/ml</td>
<td>T</td>
</tr>
</tbody>
</table>

**Table 2.4: Cytokines used in in vitro cultures.**

### 2.2.3.4. Fixing Cells by cytospin for MGG Staining

Slides were labeled with a pencil and adapted to slide holder, filter and funnel. The filter was moistened with 100μl PBS and spun at 350g for 2mins. Cells were re-suspended in 100μl PBS added to the funnel and spun at 1000g for 10mins (minimum acceleration). The slides were air-dried overnight prior to staining. Slides were immersed in May-Grünwald (VWR International) for 5 minutes, washed in PBS twice and then stained with fresh Giemsa (VWR International, made up from 10x stock) for 10-20mins. Slides were washed in water, air-dried and a cover slip mounted with pertex in the fume hood.

### 2.2.4. Flow Cytometric Analysis

Single cell suspensions were blocked with Fc receptor binding myeloma antibodies and then stained with specific monoclonal antibodies listed (Table 2.5). Before analysis and cell sorting the sample was dissolved in viability dye. For cell sorting, samples were sometimes enriched with CD34 isolation beads before staining was performed. Progenitors were identified according to the following markers (KIT and CD38 were only used on primary cells, not on *in vitro* differentiated hPSCs): LIN^CD34^-CD45RA^: LIN^CD19^CD34^-(CD38^-)CD45RA^-
(KIT⁺), IL7R⁺ (KIT⁺) progenitor: LIN CD19⁺CD34⁺(CD38⁺)CD45RA⁻IL7R⁺ (KIT⁺); ProB: LIN⁻CD34⁻CD19⁻; CD34⁺: LIN⁻CD19⁻CD34⁺.

<table>
<thead>
<tr>
<th>Antibody-conjugate</th>
<th>Clone</th>
<th>Supplier</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>CD45RA FITC</td>
<td>MEM-56</td>
<td>Invitrogen</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD34-FITC</td>
<td>561</td>
<td>Biolegend</td>
<td>IPS samples</td>
</tr>
<tr>
<td>CD43 FITC</td>
<td>IG10</td>
<td>BD</td>
<td>IPS samples D10</td>
</tr>
<tr>
<td>KIT (CD17)-PE</td>
<td>YB5.B8</td>
<td>BD</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD19-PE</td>
<td>HIB19</td>
<td>Biolegend</td>
<td>IPS samples</td>
</tr>
<tr>
<td>CD61 PE CY5 (LIN)</td>
<td>B159</td>
<td>BD</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD35 PE CY5 (LIN)</td>
<td>HIT3a</td>
<td>Biolegend</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD8a PE CY5 (LIN)</td>
<td>RPA-T8</td>
<td>Biolegend</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD235-PE CY5 (LIN)</td>
<td>GA-B2 (HR2)</td>
<td>BD</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD45RA PE CY7</td>
<td>H101</td>
<td>Biolegend</td>
<td>IPS samples</td>
</tr>
<tr>
<td>CD38-PE CY7</td>
<td>HIT3</td>
<td>eBioscience</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>IL-7R (CD127)-BV421</td>
<td>HIL-7R-M21</td>
<td>BD</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>CD19-BV605</td>
<td>SA-DA4</td>
<td>eBioscience</td>
<td>IPS samples D31</td>
</tr>
<tr>
<td>CD19 APC</td>
<td>SA-DA4</td>
<td>eBioscience</td>
<td>IPS samples D31</td>
</tr>
<tr>
<td>CD34-A700</td>
<td>S25C1</td>
<td>BD</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD58-APC</td>
<td>A700</td>
<td>BD</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>IL7R (CD127)-BV421</td>
<td>S25C1</td>
<td>BD</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>CD19-APC</td>
<td>SA-DA4</td>
<td>eBioscience</td>
<td>IPS samples D31</td>
</tr>
<tr>
<td>CD58-APC</td>
<td>A700</td>
<td>BD</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD34-A700</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>IL7R (CD127)-BV421</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>CD19-APC</td>
<td>SA-DA4</td>
<td>eBioscience</td>
<td>IPS samples D31</td>
</tr>
<tr>
<td>CD58-APC</td>
<td>A700</td>
<td>BD</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD34-A700</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>IL7R (CD127)-BV421</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>CD19-APC</td>
<td>SA-DA4</td>
<td>eBioscience</td>
<td>IPS samples D31</td>
</tr>
<tr>
<td>CD58-APC</td>
<td>A700</td>
<td>BD</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD34-A700</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>IL7R (CD127)-BV421</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>CD19-APC</td>
<td>SA-DA4</td>
<td>eBioscience</td>
<td>IPS samples D31</td>
</tr>
<tr>
<td>CD58-APC</td>
<td>A700</td>
<td>BD</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD34-A700</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>IL7R (CD127)-BV421</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>CD19-APC</td>
<td>SA-DA4</td>
<td>eBioscience</td>
<td>IPS samples D31</td>
</tr>
<tr>
<td>CD58-APC</td>
<td>A700</td>
<td>BD</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD34-A700</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>IL7R (CD127)-BV421</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>CD19-APC</td>
<td>SA-DA4</td>
<td>eBioscience</td>
<td>IPS samples D31</td>
</tr>
<tr>
<td>CD58-APC</td>
<td>A700</td>
<td>BD</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD34-A700</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>IL7R (CD127)-BV421</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>CD19-APC</td>
<td>SA-DA4</td>
<td>eBioscience</td>
<td>IPS samples D31</td>
</tr>
<tr>
<td>CD58-APC</td>
<td>A700</td>
<td>BD</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD34-A700</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>IL7R (CD127)-BV421</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>CD19-APC</td>
<td>SA-DA4</td>
<td>eBioscience</td>
<td>IPS samples D31</td>
</tr>
<tr>
<td>CD58-APC</td>
<td>A700</td>
<td>BD</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD34-A700</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>IL7R (CD127)-BV421</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>CD19-APC</td>
<td>SA-DA4</td>
<td>eBioscience</td>
<td>IPS samples D31</td>
</tr>
<tr>
<td>CD58-APC</td>
<td>A700</td>
<td>BD</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD34-A700</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>IL7R (CD127)-BV421</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>CD19-APC</td>
<td>SA-DA4</td>
<td>eBioscience</td>
<td>IPS samples D31</td>
</tr>
<tr>
<td>CD58-APC</td>
<td>A700</td>
<td>BD</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD34-A700</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>IL7R (CD127)-BV421</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>CD19-APC</td>
<td>SA-DA4</td>
<td>eBioscience</td>
<td>IPS samples D31</td>
</tr>
<tr>
<td>CD58-APC</td>
<td>A700</td>
<td>BD</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD34-A700</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>IL7R (CD127)-BV421</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>CD19-APC</td>
<td>SA-DA4</td>
<td>eBioscience</td>
<td>IPS samples D31</td>
</tr>
<tr>
<td>CD58-APC</td>
<td>A700</td>
<td>BD</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD34-A700</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>IL7R (CD127)-BV421</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>CD19-APC</td>
<td>SA-DA4</td>
<td>eBioscience</td>
<td>IPS samples D31</td>
</tr>
<tr>
<td>CD58-APC</td>
<td>A700</td>
<td>BD</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD34-A700</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>IL7R (CD127)-BV421</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>CD19-APC</td>
<td>SA-DA4</td>
<td>eBioscience</td>
<td>IPS samples D31</td>
</tr>
<tr>
<td>CD58-APC</td>
<td>A700</td>
<td>BD</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD34-A700</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>IL7R (CD127)-BV421</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>CD19-APC</td>
<td>SA-DA4</td>
<td>eBioscience</td>
<td>IPS samples D31</td>
</tr>
<tr>
<td>CD58-APC</td>
<td>A700</td>
<td>BD</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD34-A700</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>IL7R (CD127)-BV421</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>CD19-APC</td>
<td>SA-DA4</td>
<td>eBioscience</td>
<td>IPS samples D31</td>
</tr>
<tr>
<td>CD58-APC</td>
<td>A700</td>
<td>BD</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD34-A700</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>IL7R (CD127)-BV421</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>CD19-APC</td>
<td>SA-DA4</td>
<td>eBioscience</td>
<td>IPS samples D31</td>
</tr>
<tr>
<td>CD58-APC</td>
<td>A700</td>
<td>BD</td>
<td>Human primary samples</td>
</tr>
<tr>
<td>CD34-A700</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
<tr>
<td>IL7R (CD127)-BV421</td>
<td>H101</td>
<td>Biolegend</td>
<td>Human primary and IPS samples</td>
</tr>
</tbody>
</table>

Table 2.5: List of FACS antibodies and reagents used on both hPSC and primary cells.

2.2.5. Generation of Knock-in hPSCs

2.2.5.1. Transfection of hPSCs by Nucleofection

hPSCs were harvested as single cell using gentle dissociation reagent following 2 hours pre-incubation with rock inhibitor. 2x10⁶ hPSCs were transfected with 2µg Xhol linearised RUNX1 knock-in construct and 4µg each of Cas9D10A (pCDNA3.3topo hNLS-hCas9D10ANLS, kindly provided by Dr Beth Payne, UCL CI) and a single custom ETV6 guide RNA (pBS-ETV6gRNA: target GGCAATTGGAGGCTTCTGCT) or 4µg of each TALEN pair using human stem cell nucleofector kit 2 using programme B016, Nucleofector IIA (Lonza). Cells were recovered into 10ml mTeSR1/Y27632 on a 10cm matrigel-coated plate. Media was changed daily and at 48 hours 50µg/ml G418 selection was commenced.
2.2.5.2. **hPSC colony picking, replica plating and freezing**

At day 8-10 single colonies were picked under microscopic guidance into a 96 well plate. Once confluent, clones were replica plated: one plate was gowned for freezing at -80°C in 90% knockout serum replacement (Invitrogen)/10% DMSO (Sigma) supplemented with Y27632; the other was grown and lysed for screening by southern blot. Lysis conditions: 65°C overnight in 50μl cell lysis buffer (NaCl 100mM, Tris-HCl pH 7.5 50mM, EDTA 10mM, SDS 0.5%, Proteinase K 0.8mg/ml) as described above (section 2.1.10).

2.2.5.3. **Thawing Candidate clones onto HS27 stroma**

Candidate knock-in clones identified by screening Southern blot hybridisation were thawed onto matrigel-coated 24 plates prepared with confluent irradiated HS27 feeder layers (HS27 media and cells kindly provided by Andrew Smith, University of Oxford). A 96 well plate was thawed in the airflow of the TC hood. Cells were not centrifuged and DMSO was diluted-out by direct transfer into the 2ml mTeSR1 media (supplemented with rock inhibitor for the first 24 hours). The clones were fed daily with mTeSR1 and later passaged into 1-3 matrigel coated 6 wells for freezing aliquots and confirmatory Southern blot.

2.2.5.4. **Reversion of RUNX1 Cassette by TAT-Cre recombinase**

10 000 hPSCs from ETV6-RUNX1 knock-in clone 2.8 were seeded into a matrigel coated 12 well plate in StemFit® media. The following day the cells were incubated for 3 hours in 500μl StemFit® media containing 5μM TAT-Cre (Millipore) (filter sterilised). hPSCs were passaged after 48 hours as single cells and plated at limiting dilutions in a 12 well plate. Emerging
colonies were picked under microscopic guidance into 96 well plates and processed as above for screening by Southern blot hybridisation using KpnI-HF digest and internal Neo probe.
CHAPTER 3

In Vitro B cell Differentiation of Human Pluripotent Stem Cells Recapitulates Early Embryonic Lymphopoiesis
3. CHAPTER 3: In Vitro B cell Differentiation of Human Pluripotent Stem Cells

Recapitulates Early Embryonic Lymphopoiesis

3.1. Introduction

ETV6-RUNX1 is the commonest single initiating mutation driving childhood ALL and is commonly acquired \textit{in utero}, yet its impact on human fetal lymphopoiesis has not been fully elucidated. Current models of ETV6-RUNX1 have shown variable phenotypes, which appear to be affected by the level of ETV6-RUNX1 expression and the developmental stage and differentiation status of B cell progenitor studied. Overall these studies indicate that accurately modeling ETV6-RUNX1 pre-leukaemia would require the development of a human system expressing physiologically relevant levels of ETV6-RUNX1 in the appropriate B lineage stem or progenitor cell targeted by ETV6-RUNX1 \textit{in utero}. The aim of this project was to generate a genetically and developmentally accurate model of ETV6-RUNX1 pre-leukaemia using hPSC-derived B cell differentiation as a model of early human fetal lymphopoiesis.

Studies of ETV6-RUNX1 cases have indicated some of the characteristics anticipated of the ETV6-RUNX1 target cell. The clear prenatal origins of ETV6-RUNX1 pre-leukaemia and its rarity as a cause of adult ALL led us to hypothesise that ETV6-RUNX1 arises in a transient cell state unique to early human fetal development. Although ETV6-RUNX1 is a leukaemia of pro or preB cells, the identification of common ancestral clones containing partial \textit{DJ IGH} or even \textit{TCR} rearrangements in the monochorionic twins of ETV6-RUNX1 childhood ALL cases indicates that leukemic transformation occurs within an early \textit{RAG}-expressing lymphoid progenitor potentially of B or T cell potential (Alpar et al., 2015). Furthermore, the clinical features of aberrant co-expression of myeloid antigens and lineage switching commonly
seen in many childhood B-ALLs indicate that the disease may arise in a target cell that might additionally harbor some myeloid as well as lymphoid lineage programming (Abdelhaleem, 2007; Gerr et al., 2010; Suggs et al., 2007). This is further supported by the observation that fetal myeloid haematopoietic hierarchies exhibit greater levels of multilineage-priming than those in adult (Boiers et al., 2013; Notta et al., 2016).

In principle *in vitro* differentiation of human pluripotent stem cells (hPSCs) offers a good system to model the impact of oncogenes such as ETV6-RUNX1 on early embryonic hematopoiesis (Slukvin, 2013). hPSCs are known to produce cells with characteristics of primitive hematopoiesis, such as expression of embryonic hemoglobin isoforms. As elaborated in the introduction, reports of the production of transplantable definitive hematopoietic stem or progenitor cells from hPSCs have proven contentious indicating that differentiation of hPSCs may model ‘preHSC’-derived haematopoiesis (Kyba and Daley, 2003). However the protocols for *in vitro* B cell differentiation have proved particularly challenging and so there is a lack of clarity as to what stage of development hPSC-derived B lymphopoiesis models. If hPSC-derived B cell precursors recapitulate the important developmental characteristics of the earliest B lymphoid progenitor cells in the human embryo, then hPSCs could provide a tractable model to explore the impact of cALL oncogenes on this currently inaccessible arena of human development.

I therefore began by establishing a protocol for *in vitro* B cell differentiation from hPSCs. Once established I compared the B lymphoid hierarchies produced from hPSC with those seen in adult bone marrow (BM) and fetal liver (FL). In this regard I was helped by the results of parallel work performed in this laboratory by Dr Charlotta Boiers, which characterised the earliest emerging B cells in human fetal liver (results summarised in appendix A). I was therefore able to explore how closely B cell precursors derived from hPSCs shared the key developmental characteristics of the earliest emerging human B lymphoid progenitor cells.
3.2. Results

3.2.1. Establishment of in vitro B cell Differentiation from hPSCs

For these studies I used the human embryonic stem cell (hES) line, H1 and an induced pluripotent stem cell line (hiPS), MIFF3 (kindly provided by Prof Peter Andrews and Dr Christian Unger, University of Sheffield). The MIFF3 line was chosen as a fully characterised pluripotent hiPS line derived by a laboratory specialising in hPSCs. Importantly the IPS re-programming factors were transiently delivered as mRNA and do not integrate into the genome thus cannot affect the differentiation potential of the cells or confound the analysis of the impact of ETV6-RUNX1 expression.

An OP9/MS5 sequential co-culture system was used to differentiate hPSCs to the B cell lineage (Figure 3.1). This system was based on that described using H1 hES cells by Vodjanik and colleagues in 2005 and validated with hiPSCs by Lee Carpenter in 2011 (Carpenter et al., 2011; Vodyanik et al., 2005). OP9 stroma was kindly provided by Igor Slukvin and MS5 stroma were later kindly provided by Lee Carpenter (discussed below). Co-cultured haematopoietic cells were analysed at two time-points; after 10 days of OP9 co-culture (D10) to assay the formation of CD34+ haematoendothelial and CD34+/43+ haematopoietic progenitors, and after further co-culture in lymphoid conditions on MS5 stroma for the formation of B cells at D31 (±2 days).
Figure 3.1: hPSC in vitro B cell differentiation protocol. Whole colonies of hPSCs are harvested onto a layer of overgrown OP9 and co-cultured for 10 days. After CD34 enrichment cells were either analysed (D10) or co-cultured for an additional 21 days on MS-5 in the presence of lymphoid cytokines as indicated (D31) (Carpenter et al., 2011).

The long timescales involved and use of multiple non-defined reagents including stroma and batch dependent fetal calf serum (FCS) complicated the establishment of the protocol. According to published literature the handing of the OP9 co-culture is considered to be the most challenging aspect of the protocol (Choi et al., 2011). Nevertheless after a small number of pilot attempts I was able to specify CD34+/43+ early haematopoietic progenitors by D10 (Figure 3.2). Consistent with published data, a typical yield of CD34⁺/43⁺ haematopoietic progenitors was $10^5$ cells, from an input of $10^6$ hPSCs. This success was probably attributable to the use of verified OP9 stroma from the Slukvin lab kept strictly according to protocol. Subsequent experiments were performed using serum tested against this batch of serum validated in these experiments.
Figure 3.2: Analysis of H1 hES cells differentiated on OP9 stroma to D10 for surface expression of the haematoendothelial markers CD34 and CD43. Cells are CD34 MACS enriched and gated as viable.

Having achieved haemato-endothelial specification from hPSCs, I then attempted to differentiate these CD34⁺ cells to B cells on MS5 stroma supplemented with cytokines (IL3, IL7, SCF and Flt3-L). B cell output proved unsuccessful when using MS5 stroma from within the lab, however, changing to MS5 stroma validated to produce hiPSC-derived B cells in the Carpenter laboratory resulted in the expansion of haematopoietic progenitors with the formation of cobblestone colonies (Figure 3.3 right). FACS analysis of these D31 cultures showed the presence of CD34⁺CD19⁺ proB and CD34 CD19⁺ pre B cells from MIFF3 hiPSCs (Figure 3.3 left).

Work by Charlotta Boiers in human FL had recently identified a lineage negative IL7R progenitor cell population (referred to herein as IL7R progenitor) with lymphoid and myeloid potential that is abundant in early FL (Figures A1.1 and A1.2). An IL7R progenitor of similar phenotype (CD34⁺CD45RA⁻IL7R⁻CD19⁻) to the IL7R progenitor observed in FL was also identifiable at D31 of hPSC co-culture. This progenitor was also seen as early as D10,
significantly before the emergence of CD19⁺ B cells and prior to the addition of lymphoid cytokines. Although CD38 and Kit were used in the initial identification of the FL IL7R progenitor population, the surface expression of these activation markers was found to be affected by culture conditions (e.g. presence of exogenous SCF) and they were not used for further analysis of hPSC-derived cells.

**Figure 3.3:** Left: Differentiated MIFF3 hPSCs were analysed for CD19⁺ B cells (top) and CD19⁻IL7R⁺ progenitors (bottom) at D10 (left) and D31 (right). Cells were gated CD45⁺, further gating as indicated. All cells at D10 are CD19 negative when compared to fluorescence-minus-one (FMO) cord blood controls. Mean percentages of CD45⁺ cells, n=5-10. Right: In vitro image of CD34-MS5 co-cultures at D31.

I was also able to reproducibly produce pro and preB cells and IL7R progenitors from H1 hESCs (Figure 3.4). This result was in contrast with reports from the Carpenter laboratory
where they had failed to produce B cells from H1 hESCs using an equivalent protocol (Carpenter et al., 2011).

**Figure 3.4:** H1 hESCs were analysed for CD19^+ B cells (top) and IL7R^+ progenitor (bottom) at D10 (left) and D31 (right) of differentiation. Cells were gated CD45^+ or CD45^-IGM^-.

Further gating is indicated in the figure. Numbers show mean percentages of CD45^+ cells.

Representative reanalysis of sorted MIFF3 hPSC-derived cells is shown to the right of each plot, IL7R^+ progenitor (red) and LIN^- CD34^+ CD45RA^- control cells (blue), proB cells right top panel. Cells were typically sorted with \( \geq 94\% \) purity.

In summary, these experiments established a reliable B cell differentiation protocol from both hES and hiPSCs (collectively referred to as hPSCs) that produced immunophenotypically defined pro and preB cells along with a candidate IL7R progenitor population similar to that first identified in FL.
3.2.2. Molecular and Functional Characterisation of the hPSC-derived IL7R progenitor

Having demonstrated the ability to differentiate hPSCs into cells with an IL7R progenitor phenotype at both day 10 and day 31 of co-culture I next examined the lineage potentials of the hPSC-derived CD19 negative IL7R+ progenitor compartment to see if it correlated with that identified in the FL (see figure 3.4 for sorting strategy). As hypothesised, 25 cell RT-qPCR (reverse transcriptase quantitative polymerase chain reaction) gene expression analysis of IL7R progenitor and preB cells sorted from D31 MIFF3 co-cultures confirmed the progressive up-regulation of key lymphoid genes including IL7R, RAG2, EBF1 and PAX5 in both compartments. EBF1 was expressed in the IL7R progenitor compartment consistent with early lymphoid commitment along with low levels of the B lineage committed gene PAX5. The IL7R progenitor did not express the key erythroid TF GATA1 consistent with absence of MkE potential. hPSC-derived cells also expressed LIN28B suggesting a similarity to fetal tissues (Figure 3.5) (Yuan et al., 2012).
Figure 3.5: qPCR of D31 MIFF3 hiPSC co-cultures. LIN'CD34+CD45RA−, IL7R+ progenitors and CD34+CD19+ proB cells were analysed for expression of lymphoid (IL7R, RAG2, EBF1, PAX5), megakaryocytic/erythroid (GATA1) and fetal specific (LIN28B) genes. Data shown normalised to GAPDH, n=2-3.

The IL7R population identified in the FL is characterised by both lymphoid and myelomonocytic potential when cultured in vitro. The above qPCR results suggested that this population was highly B lineage primed. I tested if the D31 hPSC-derived IL7R progenitor from MIFF3 hiPSC could give rise to B cells during in vitro co-culture with MS5 stroma. 100-500 progenitor cells as available were sorted directly into media and added to a single MS5-containing well of a 96 well plate and cultured for 2 weeks. As expected the hiPSC-derived IL7R progenitor from MIFF3 showed robust CD19+B cell output (>70% of CD45+ cells) (Figure 3.6).
Figure 3.6: B cell potential of the IL7R<sup>+</sup> progenitor was investigated on MS5 stroma. The progenitor was sorted from MIFF3 D31 and cultured for an additional 12-14 days and analysed by flow cytometry for CD19<sup>+</sup> cells. LIN<sup>−</sup>CD34<sup>+</sup>CD45RA<sup>−</sup> cells were cultured in parallel as control. Graph shows percent CD19<sup>+</sup> cells of total CD45<sup>+</sup> cells in the culture, Mean+SD, n=2. To the right is a representative FACS analysis of the IL7R<sup>+</sup> progenitor readout.

The FL IL7R progenitor compartment has myelo-monocytic potential, but lacks erythroid capacity when assayed by semi-solid culture. To test the erythroid potential of the hPSC-derived IL7R progenitors, 500 cells were sorted from D31 of B cell differentiation culture and cultured with Methocult H4435 (Stem Cell Technologies) for 2 weeks. IL7R progenitors produced small numbers of macrophage colonies, confirmed on MGG staining of pooled cytospun colonies (Figure 3.7). No erythroid potential was seen from the IL7R compartment, but BFU-E and mixed GEMM colonies were readily formed from a CD34<sup>+</sup>CD45RA<sup>−</sup> control fraction even after 3 weeks of lymphoid co-culture. IL7R progenitors from the early D10 time point of OP9 co-culture failed to grow when directly sorted either on MS5 or in semi-solid erythro-myeloid culture (data not shown).
Figure 3.7: Left: Myeloid potential of the IL7R⁺ progenitor was investigated in semi-solid media. The progenitor was sorted from MIFF3 and H1 D31 and cultured for an additional 12-14 days. Lin⁻CD34⁺CD45RA⁻ hPSCs were cultured in parallel as control. Graph shows percent colonies per plated number of cells. Mean±SD, n=2. Right: Macrophages from MIFF3 hPSC-derived IL7R⁺ progenitor culture.

Work by Charlotta Boiers has shown that the IL7R progenitor population switches from myeloid to lympho-myeloid lineage priming during fetal development (Appendix A and Figures A1.3 and A1.4). Having demonstrated the myelomonocytic and B lineage potentials of the hPSC-derived IL7R progenitor, I next examined whether this compartment mirrored the myeloid to lymphomyeloid ‘switch’ in lineage priming seen in the FL IL7R progenitors. Single cells from both D10 (IL7R progenitors only) and D31 (IL7R and proB cells) were directly sorted into 4μl lysis buffer for analysis by multiplex RT-qPCR (performed by Charlotta Boiers). Gene sets were defined as either myeloid (red), early lymphoid (blue) or B committed (green) consistent with those used by Dr Boiers for analysis of human primary cells (Appendix A). MIFF3 D31 proB cells expressed a coherent B-lymphoid gene expression signature, with some minor incoherent activation of myeloid affiliated genes (Figure 3.8 right). As with the FL IL7R expression was high in all the proB cells despite this being less
obvious at the surface protein level, likely due to receptor internalisation in response to high levels of IL7 in culture. Single cell RT-qPCR showed that the early D10 IL7R progenitor exhibited a strong and relatively uniform myeloid signature similar to the early fetal CS17 IL7R progenitor (Figure 3.8 left), but by D31 there appeared to be a mix of myeloid, lympho-myeloid and B lineage primed cells (Figure 3.8 middle).

**Figure 3.8:** Single cell qPCR analysis of IL7R⁺ progenitors from D10 and D31 of differentiation and proB cells from D31 (MIFF3). Each column represents a single cell. Coloured by CT value. Gene sets: red myeloid; blue lymphoid and green B lineage related genes. 35-58 cells per time-point, n=2 per population.

To ensure that this myelo-lymphoid signature was not specific to the MIFF3 hiPSC line, we also analysed IL7R progenitor and proB cells derived from H1 hESCs at D31. The pattern of gene expression was similar to that seen in MIFF3 hiPSCs with evidence of lympho-myeloid priming in the IL7R progenitor population (Figure 3.9).
Figure 3.9: Single cell qPCR analysis of IL7R⁺ progenitor and proB cells from H1 hESCs at D31 of differentiation. Each column represents a single cell. Colour code based on CT values. Genes labeled in red are myeloid, blue lymphoid and green B cell genes, n=1, 20 cells investigated.

In summary, these data show that both hES and hIPS-derived B cells highly express IL7R, and that hPSC differentiation produces an IL7R progenitor population that shares a number of key features with their FL counterparts: i) B lineage and myelomonocytic functional outputs in vitro, ii) absence of erythroid potential and iii) a switch from myeloid to lympho-myeloid priming during development.

3.2.3. Transcriptional comparison of B lymphopoiesis in Human Fetal Liver and from hPSCs.

Having demonstrated immunophenotypic and functional similarities between FL and hPSC-derived B lymphoid hierarchies, in collaboration with Charlotta Boiers I next sought to assess the extent of this similarity by comparing the expression of lineage affiliated genes with that seen in adult BM and neonatal CB.
Firstly, to determine whether the overall pattern of gene expression assayed by qPCR could objectively separate the IL7R progenitor and proB cell populations derived from primary FL, BM and CB samples we analysed the single cell qPCR data by principal component analysis (PCA) (Figure 3.10). FL proB cells (blue squares) segregated from adult (red squares), with neonatal CB proB (pink squares) overlapping both populations reflecting the relatively small differences seen between the proB cells sorted from different stages of development. The IL7R progenitors clearly separated from proB cells, with those from FL clustering away from both CB and adult BM and the early CS17 (light blue triangles) also segregating from later CS20 FL IL7R progenitors (dark blue triangles).

Figure 3.10: PCA from single cell qPCR data of IL7R⁺KIT⁺ progenitor (triangles) and proB cells (squares) from FL (CS17 (progenitor only) and CS20), CB and adult BM. Each dot represents a single cell. Single cell qPCR performed by Charlotte Boiers. PCA constructed by Javier Herrero.
This analysis suggested that the biggest differences in the expression of lineage-affiliated genes were between adult and fetal IL7R progenitors, with additional large changes occurring during early fetal development. Among the genes contributing most strongly to the signature of the early FL IL7R progenitor we found the myeloid tyrosine kinase receptor CSF1R (M-CSFR, v-fms) and LIN28B (Figure 3.11).

Figure 3.11: Left and right graphs show the direction and magnitude of eigenvectors contributing to separation for lineage-associated (left) and most differentially expressed (right) genes respectively. Red: myeloid; Blue: lymphoid; Green: B lineage. 36 genes were used for the PCA, n=2-3 per developmental stage. Dashed ellipses represent IL7R progenitor and proB cell clusters from all sources as indicated in the key. Single cell qPCR performed by Charlotta Boiers. PCA constructed by Javier Herrero.

Reconstruction of this PCA now including data from hPSC cultures (Figure 3.12) showed that the early D10 hPSC-derived IL7R progenitors (light green triangles) overlapped closely with early CS17 FL (light blue ellipse), whereas by D31 IL7R progenitors (dark green triangles) overlapped both the CS17 and CS20 FL progenitor (light and dark blue ellipses respectively) reflecting the mix of myeloid and lympho-myeloid primed cells seen in Figure 3.8. The hPSC proB cells (green squares) clustered away from the IL7R progenitors overlapping in part with the territory occupied by the CS20 FL proB cells and distinct from adult BM proB cells.
Figure 3.12: PCA calculated from single cell qPCR data from the human primary cells and differentiated hPSCs (MIF3 and H1). hPSCs are from D10 (IL7R⁺ progenitor only) and D31 (IL7R⁺ progenitor and proB cells). Each dot represents a single cell (only hPSCs shown).

Highlighted ellipses represent primary human cell clusters, 36 genes were used for the PCA, n=2-3. Single cell qPCR performed by Charlotta Boiers. PCA constructed by Javier Herrero.

Therefore the hPSC-derived IL7R progenitor and proB cells appear to resemble the FL cells both functionally and transcriptionally. In particular, at the time-points studied, the IPS cultures also appear to recapitulate the myeloid to lympho-myeloid transition seen in the IL7R progenitor during embryonic development including the co-expression of both myeloid and B committed lineage programmes seen in FL CS20 IL7R progenitors.

I next used RNA sequencing as a global measure of the similarity between hPSC and FL derived B cell hierarchies. The number of IL7R progenitors that could be sorted from either
primary samples or hPSC co-cultures was in the low hundreds requiring optimisation of a low cell number protocol for cDNA generation. Cells were sorted directly into TRIZOL® and RNA extracted using a standard phase separation (see methods). With assistance from Dr Gill May I first tested whether this method of RNA extraction could produce intact RNA from low cell numbers as assayed by an Agilent Bioanalyser RNA 6000 pico chip. I split a sample of 4000 MIFF3-derived proB cells sorted into TRIZOL to test whether it was possible to prepare RNA from the equivalent of 200 or 500 cells fractions (prepared in duplicate). RNA integrity and quantitation was assayed by loading 1μl of the 3μl stock onto a Bioanalyser chip and compared to bulk-prepared RNA diluted 1:5 to be equivalent to that from 520 cells (Figure 3.13).

![Figure 3.13: Agilent Bioanalyser RNA 6000 pico chip of RNA prepped from replicates of 200 (n=2) and 500 (n=2) cell equivalents of TRIZOL® compared to a dilution of the bulk RNA equivalent to 520 cells (n=1). (1μl of 3μl RNA analysed).](image-url)
All samples produced visible ribosomal RNA (rRNA) bands indicating that low amounts of RNA prepared in this way were not grossly degraded. One of the 500 cell replicates produced quantifiable levels of RNA (579pg/μl, 3.4pg/cell), showing that the yields were reduced somewhat from that measured in the 520 cell equivalent diluted bulk RNA (983pg/μl, 5.6pg/cell). The RNA integrity number (RIN) is calculated from the ratio of the rRNA peaks and provides a crude measure of RNA degradation, with numbers closer to 10 indicative of higher quality RNA. RIN averages were 5.0 for 200 cell, 9.0 for 500 cells and 9.4 for the 520 cell equivalent population. This indicated that preparing RNA from 500 cells resulted in good quality RNA by this measure. Although the RIN numbers for the 200 cell samples was low, the appearance of the 18S and 20S rRNA bands indicates that this likely reflects the noise in signal generated by being below the limits of RNA quantification. Thus, although it is impossible to accurately quality control RNA using the Bioanalyser from the 200 cell samples, visual inspection of the Bioanalyser trace suggested that the RNA was intact and should be amenable to amplification during cDNA preparation.

CDNA preparation for low RNA input is best achieved using a template switching reverse transcriptase technology followed by PCR amplification of the first strand. Although amplification is required to produce sufficient cDNA for sequencing library preparation, over-amplification can result in unwanted biases in transcript representation. cDNA quality can be assessed by measuring both the total cDNA yield and the size distribution of cDNA transcripts using an Agilent Bioanalyser high sensitivity DNA chip. Previous studies performed by Gill May using the SMARTer HV cDNA kit (Clontech) had shown that 500pg of input RNA required 17 cycles of amplification to produce the recommended cDNA yields (target range 2-7ng). I first used the SMARTer HV kit to assess whether the 200 cell RNA samples could be amplified to produce similar quality cDNA to that from the 500 cell and bulk fractions (Table 3.1). Assuming a doubling of cDNA yield per amplification cycle, I
titrated the number of amplification cycles to the estimated input RNA. This result suggested that good quality cDNA at the required yields could be retrieved from as low as 200 cell equivalents, even when this RNA was not accurately quantifiable by Bioanalyser RNA Pico chip.

<table>
<thead>
<tr>
<th>Input Cell Equivalents</th>
<th>Input RNA (pg)</th>
<th>Number of Amplification cycles</th>
<th>Output Yield (ng) (Target range 2-7ng)</th>
<th>Average Size (kb) (Target peak 2kb)</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>100*</td>
<td>19</td>
<td>3.1</td>
<td>1.88</td>
</tr>
<tr>
<td>500</td>
<td>500</td>
<td>17</td>
<td>2.5</td>
<td>1.99</td>
</tr>
<tr>
<td>2600</td>
<td>2000</td>
<td>15</td>
<td>1.8</td>
<td>1.90</td>
</tr>
</tbody>
</table>

Table 3.1: Optimisation of SMARTer HV cDNA preparation showing cDNA yields and size distribution when titrating amplification cycles to input RNA concentration.

Thus, based on a minimum predicted yield of 1pg of RNA per cell and a minimum sort of 200 cells I optimised cDNA synthesis for a standardised input of 100pg RNA for all samples. For these studies I used the updated SMARTer v3.0 cDNA kit (Clontech) for ultra low RNA input, which has a target cDNA yield of 2-10ng. Given that this updated protocol is known to be more efficient at cDNA generation I tested amplifying 100pg of input RNA (diluted from the bulk fraction used above) using 15, 16 and 17 cycles of amplification (Figure 3.14).
16 cycles produced an average of 7ng of cDNA with no discernable impact on the size distribution of cDNA fragments when visualized on an Agilent Bioanalyser high sensitivity DNA chip (Table 3.2). 17 cycles resulted in some over-amplification based on a cDNA yield of 12ng and some increase in the average size of cDNA product, whereas 15 cycles only produced 3.6ng yield, which was near the lower limit of the expected range. 16 cycles of amplification from 100pg input RNA using the SMARTer V3.0 kit was therefore chosen for cDNA library preparation from small cell numbers as it produced optimal cDNA yield for library preparation in the expected size range.

<table>
<thead>
<tr>
<th>Number of Amplification cycles</th>
<th>Output Concentration (pg/μl) in 17μl</th>
<th>Average Yield (ng) (Target range 2-10ng)</th>
<th>Average Size (bp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>202</td>
<td>3.6</td>
<td>2624</td>
</tr>
<tr>
<td>15</td>
<td>233</td>
<td></td>
<td>2614</td>
</tr>
<tr>
<td>16</td>
<td>417</td>
<td>7</td>
<td>2695</td>
</tr>
<tr>
<td>16</td>
<td>367</td>
<td></td>
<td>2667</td>
</tr>
<tr>
<td>17</td>
<td>733</td>
<td>12</td>
<td>2702</td>
</tr>
<tr>
<td>17</td>
<td>680</td>
<td></td>
<td>2751</td>
</tr>
</tbody>
</table>

Table 3.2: Summary of yields and size ranges of cDNA products assayed by Agilent Bioanalyser high sensitivity DNA chip of cDNA produced from duplicates of 100pg RNA using SMARTer v3.0 cDNA kit varying the number of amplification cycles 15, 16, or 17.

Sequencing libraries were then produced from these cDNA libraries using Illumina® Nextera XT tagmentation according to manufacturer’s instructions (see methods). Sequencing reads were mapped and quality controlled *in silico* by Dr Dapeng Wang (Bill Lyons Informatics)
Unit). Importantly given the polyA-based priming of the SMARTer reverse transcriptase we did not observe significant levels of 3’ bias in the mapped reads. A representative map of read number across transcript length from one RNASeq replicate is shown in Figure 3.15 demonstrating only minor 3’ bias.

![Graph showing RNASeq reads mapped over length of transcripts](image)

Figure 3.15: Number of RNASeq reads mapped over the length of each transcript for a representative RNASeq sample showing low levels of 3’ bias. Courtesy of Dapeng Wang.

Using this new RNASeq pipeline, I first constructed a PCA ‘map’ based on primary BM and FL samples comparing IL7R progenitor and proB populations from CS20 fetal livers and adult BM (Figure 3.16). A primitive Lin’CD34’CD45RA’ population was used as a control (Lin’CD34’CD38’CD45RA’). The PCA separated the samples based on both differentiation stage (PC1, 28.7% explained variance) and developmental origin (PC2, 16.2% explained variance) with the adult BM hierarchy (red) clearly separating from FL (blue). Based on this map we interpolated the hPSC lymphoid hierarchy onto this calculation, asking with which of these two developmental pathways it would cluster. hPSC differentiation closely mapped that of the primary cells from human FL capturing the majority of the developmental signature that distinguishes FL from adult BM.
Figure 3.16: PCA of RNASeq analysis using primary cells only: LIN CD34+CD45RA- (circles), IL7R+ progenitor (triangles) and proB cells (squares) from adult BM (red, n=3) and FL (blue, CS21-22, n=2). hPSC-derived cells (green) and the early CS17 FL IL7R+ progenitor (pale blue) were placed on the PCA calculation (haloed). hPSCs (MIFF3 and H1) are from D10 (IL7R+ progenitor only) and D31 (LIN CD34+CD45RA-, IL7R+ progenitor and proB cells) of differentiation. Ellipses represent the CD34, IL7R progenitor and proB clusters from all samples. PCA1 and PCA2 are shown and each dot represents one sample. PCA constructed by Javier Herrero.

Whilst this interpolation strategy was used to mitigate any disproportionate impact of culture related signatures distorting the PCA, it risked masking important additional differences between hPSC and primary cells. Reassuringly, however, a PCA calculated simultaneously using primary and hPSC sample data produced remarkably similar results although as expected with additional differences between FL and hPSC samples (Figure
3.17). Collectively these global gene expression profiles support the hypothesis that in vitro hPSC differentiation provides a developmentally relevant model of early embryonic human lymphopoiesis.

**Figure 3.17:** PCA calculated using RNA sequencing data from all human primary samples and hPSCs from D10 (IL7R\(^+\) progenitor only) and D31 (LIN\(^-\) CD34\(^+\) CD45RA\(^-\), IL7R\(^+\) progenitor and proB cells) of differentiation. CS17 IL7R progenitor is put onto the PCA, since it is an outlier, shown with a grey halo. PCA1 and PCA2 are shown and each dot represents one individual sample. Legend is shown to the right. PCA constructed by Javier Herrero.

3.3. Discussion

ETV6-RUNX1 initiates leukaemia in utero suggesting a specific impact on human fetal B lymphopoiesis. However, early first trimester B lymphoid development in the human remains poorly characterised and those cells that are available provide a limited resource for disease modeling. I tested whether hPSC-derived B lymphopoiesis could model the
distinctive features of FL B lymphopoiesis identified by Charlotta Boiers in parallel experiments on primary human FL.

We found that in vitro differentiation of hPSCs recreated the B cell hierarchy seen in FL, including a lympho-myeloid IL7R progenitor. Although IL7R progenitors harvested at D31 had been grown in B cell conditions for a month, we were able to demonstrate both B and myelomonocytic potential on in vitro culture. We were also able to show that undifferentiated CD34^+CD45RA^-^IL7R^-^ bulk cells were still able to produce erythroid output at this time point, but this was never seen from the IL7R progenitor. Surprisingly, an immunophenotypically-defined IL7R progenitor was also present at D10 in small numbers. This cell was not amenable to in vitro culture, but by single cell RT-qPCR appeared to harbour an entirely myeloid gene expression programme. By D31 the IL7R progenitor exhibited both myeloid and lymphoid gene expression programmes, including the highly unusual co-expression of myeloid and B committed genes (including the B cell master regulator PAX5), mirroring that seen in the FL. Thus the hPSC-derived IL7R progenitors share the key functional and molecular lineage attributes of their FL counterparts.

I used RNASeq to test how similar hPSC lymphopoiesis was compared to that in FL and adult BM at the global transcriptional level. An initial analysis of all samples by PCA suggested that hPSCs model an earlier stage of human lymphopoiesis than that represented from the samples analysed in FL (Figure 3.17). While this may be the case, this separation may in part reflect difficulties in comparing the global transcriptional profiles of primary samples with hPSC-derived cells due to the specific impact of culture in vitro. In support of this, interpolating the analysis of the global transcriptional level profile of the hPSC-derived hierarchy onto a PCA calculated by only using primary samples (Figure 3.16) demonstrated a remarkable degree of overlap with the developmental profile of FL.
The developmental correlates in vivo of hematopoietic progenitors derived from hPSCs in vitro is an area of considerable controversy (Slukvin, 2013). T-cell and where achieved B-cell lymphoid output is considered evidence of production of ‘definitive’ hematopoiesis, but as yet reproducible derivation of engraftable dHSCs from hPSCs has been elusive (Kennedy et al., 2012). However the identification of lymphoid progenitors in the murine YS prior to the emergence of dHSCs suggests that human B lymphopoiesis may be possible from preHSC-like cells (Boiers et al., 2013). Whereas the stages of FL development analysed here are subsequent to the emergence of dHSCs at the AGM, it remains possible that hPSCs are modeling an earlier lymphoid hierarchy and that this, rather than culture artifact, is contributing to the additional differences seen with FL in figure 3.17.

Whether or not hPSCs represent an earlier B lymphoid hierarchy, this careful comparative analysis show that at the time-points studied hPSC-derived lymphopoiesis capture significant components of both the lineage and global transcriptional landscape seen in the earliest emerging FL B cells and IL7R progenitors. As such we conclude that hPSCs offer an experimentally tractable system in which to model pre-leukemic phase of ALL initiated by ETV6-RUNX1 in early human ontogeny.
CHAPTER 4

Generation and Validation of Genome Engineered

ETV6-RUNX1 Human Pluripotent Stem Cells
4. CHAPTER 4: Generation and Validation of Genome Engineered ETV6-RUNX1 Human Pluripotent Stem Cells

4.1. Introduction

I aimed to assess the impact of ETV6-RUNX1 expression during the in vitro B cell differentiation of hPSCs to model its effect on early embryonic human lymphopoiesis. Evidence from other model systems had suggested that the expression level of ETV6-RUNX1 affects the phenotype observed (Tsuzuki and Seto, 2013). I therefore decided to utilise the genetic tractability of hPSCs to genome engineer a more accurate model of ETV6-RUNX1 expression than is achievable by viral transgenesis.

Three approaches were considered. The simplest of these involved an inducible knock-in approach. Here the cDNA of ETV6-RUNX1 under the control of a doxycycline-inducible promoter would be introduced into hPSCs either via a PiggyBac transposon or genome engineered into the AAVS1 safe harbor locus using commercially available zinc finger nucleases (Sigma). The advantage of this approach was its relative technical feasibility and the control over the timing of ETV6-RUNX1 expression. However, this approach does not put ETV6-RUNX1 under the control of the endogenous ETV6 promoter and was likely to result in high levels of ETV6-RUNX1 in those cells induced to express it. Furthermore, the cells would retain two functional alleles of ETV6, which is likely to act as a tumour suppressor in ALL.

The second approach I considered was engineering the t(12;21) translocation. This could be achieved either by genome engineering site-specific recombination sites (e.g. LoxP) into each breakpoint locus with translocation mediated by Cre expression, or by the induction of double strand breaks (DSBs) at each locus using TALENS. This approach would provide the
most authentic genetic background, including physiological levels of ETV6-RUNX1 expression and disruption of one allele each of ETV6 and RUNX1. Although there was precedent for chromosomal engineering in mouse (Forster et al., 2003), at the time of starting this project chromosomal engineering had only once been reported in hPSCs (Brunet et al., 2009). A particular difficulty anticipated was how to identify and select cells that had successfully translocated, as ETV6-RUNX1 was not anticipated to result in an identifiable phenotype that could be distinguished in mixed cultures.

The final genome engineering strategy considered was a constitutive knock-in approach, similar to that used in murine models of ETV6-RUNX1 (van der Weyden et al., 2011). In this approach a custom targeted nuclease (e.g. TALEN or later CRISPR) would direct a homologous recombination vector to insert a RUNX1 knock-in cassette into ETV6 intron V (Figure 4.1). The cassette would contain the cDNA for RUNX1 exons II-VIII linked to the bright mVenus fluorescent reporter by a self-cleaving 2A peptide. RUNX1 would be spliced into the ETV6 transcript through the engrailed 2 splice acceptor. ETV6-RUNX1 is tagged at the C terminus using a V5 epitope to assist with specific immunoprecipitation of ETV6-RUNX1 protein. The V5 tag is protected from the N terminal residues of the 2A peptide by the presence of an additional furin cleavage site. The whole RUNX1 cassette is flanked by head-to-tail LoxP sites, allowing its removal by Cre recombinase. A 3’ Flp floxable NeoR cassette, driven by the eukaryotic PGK promoter, allows selection of targeted hPSCs with G418. The NeoR gene is also driven by the prokaryotic Em7 promoter allowing kanamycin selection of RUNX1 knock-in cassette-containing plasmids during subcloning.
Figure 4.1: Genome engineering strategy to generate constitutive RUNX1 knocked into ETV6 locus. A single ETV6 gRNA directs Cas9D10A to induce a single strand nick towards the 5’ end of ETV6 intron V. A constitutive knock-in cassette encoding a splice acceptor (SA), cDNA for RUNX1 exons II-VIII V5 tagged at the C terminus and linked to the mVenus fluorescent reporter by a self cleaving furin/T2A peptide is inserted by CRISPR-directed homologous recombination. The cassette is flanked by LoxP sites and includes a 3’ triple stop/polyA tail followed by a FLP floxable NeoR positive selection cassette driven by the eukaryotic PGK and prokaryotic Em7 promoters. Not to scale.

This targeted constitutive knock-in approach has the advantage that ETV6-RUNX1 is placed under the control of the endogenous ETV6 promoter, providing physiological control of ETV6-RUNX1 expression levels. Of note, ETV6 expression in hPSCs is low (Peter Andrews, personal communication) and transgenic ETV6-RUNX1 expression does not obviously affect hPSC cultures (Ram Malladi, personal communication). An additional advantage of this approach is that it results in haploinsufficiency of ETV6, further recreating the genetic defects in ETV6-RUNX1 pre-leukaemia, although both alleles of RUNX1 remain intact.

Although still technically challenging, such knock-in strategies are commonly used in mouse ESCs and TALEN-directed genome engineering had been reported in hPSCs at the time of designing this project (Hockemeyer et al., 2011). The main conceptual caveats to the design were the potential for down-regulation of ETV6-RUNX1 by alternative splicing or allelic silencing and the potential for ETV6-RUNX1 expression affecting B lineage commitment too.
early during in vitro differentiation. This notwithstanding it was decided to proceed with the constutive knock-in approach as detailed in figure 4.1.

The ETV6-RUNX1 constitutive knock-in approach required the validation of a number of components including: i) a strategy for screening candidate knock-in clones by Southern blot hybridization; ii) construction of targeted nucleases (TALENs and/or CIRPSR); iii) construction of an ETV6 homologous recombination targeting vector by BAC recombineering; iv) testing the efficiency of homologous recombination; v) construction of the RUNX1 knock-in cassette and assembly into the validated ETV6 targeting vector; vi) the nucleofection, selection and screening of targeted clones; vii) validation of candidate clones by Southern blot genotyping, karyotyping, assessing markers of pluripotency and differentiation potential; and finally viii) reversion of the RUNX1 cassette by Cre recombinase to ensure that any phenotypes seen are directly attributable to ETV-RUNX1 expression. The technical details of steps ii-v above are detailed in sections 4.2.2-4.2.8 below.

4.2. Results

4.2.1. Choice of hPSC cell line

hPSC lines are known to exhibit differences in both their handling in hPSC culture conditions and differentiation capacity. At the time of starting this project in vitro B cell differentiation had been successful from multiple hIPSC lines by the Carpenter group, but not from hESCs (Carpenter et al., 2011), although I subsequently achieved this, as detailed in chapter 3. It was therefore decided to use a hIPSC line for the ETV6-RUNX1 knock-in model. After discussion with Peter Andrews (University of Sheffield), I decided to test two new isogenic hIPSC lines derived in the Andrews’ group by mRNA transfection of primary human foreskin
fibroblasts (MIFF1 and 3). The use of mRNA for transient expression of the IPS factors was seen as critical to avoid the possibility of confounding my results by ongoing expression of potentially oncogenic reprogramming factors.

4.2.2. Validation of Southern Blot Screening Strategy

A number of approaches have been developed to screen genome-targeted cells including long range PCR, qPCR assessment of loss of heterozygosity and Southern blot (SB) hybridisation, of which SB is considered the most sensitive and specific. Genomic DNA (gDNA) is prepared and digested with a restriction enzyme (e.g. KpnI) then size separated by agarose gel electrophoresis and transferred onto a nylon membrane. The fragment of interest can be detected using a radiolabelled PCR product of a specific locus contained within the fragment. To avoid background signal genome specific probes are designed against evolutionarily conserved regions free of repetitive sequences and then checked against the human genome for off target binding sites. The knock-in cassette is delivered in a vector containing a large (e.g. 10kb) region of homology to the targeted locus. Homologous recombination is a rare event, thus the majority of clones surviving selection will have the cassette randomly integrated into the genome. Therefore, to be specific for the targeted allele any probe must be external to the extent of the vector’s homology arms. On successful targeting the cassette changes the length of the loci’s restriction fragment, either by increasing its length, or preferably shortening it by the introduction of a novel restriction site. An overview of the Southern screening strategy used in this project is presented in Figure 4.2.
Figure 4.2: Overview of Southern blotting strategy used to identify targeting events at the 5' end of ETV6 intron V. Genomic DNA is digested with KpnI releasing a 25kb wild type (WT) fragment from both ETV6 alleles detectable using a radiolabelled probe to an evolutionary conserved region (red). The probe is located outside the long homology arms of the vector (orange) and is therefore specific to the endogenous ETV6 alleles. Upon knock-in the RUNX1 cassette introduces a new KpnI site resulting in the detection of a shorter 8.7kb allele.

Two restriction digest patterns and probe combinations were tested (Figure 4.3). The first used an Ncol dDNA digest producing a 7021bp wild type gDNA fragment, which can be detected using the radiolabelled 453bp D2 PCR product (so called because it was one of two products that had been designed to sequence verify the Downstream ends of the ETV6 BAC homology arms). I later changed to use a KpnI digest to release a 25kb gDNA fragment, which was probed using a longer 1634bp southern probe 2 (SP2) to maximise signal intensity from screening blots.
Figure 4.3: Southern blot screening strategies. TALEN1 cleavage site is annotated as the site of knock-in with surrounding 5' and 3' targeting homology arms delimited. Wild type restriction fragments for both Ncol (7021) and KpnI (25kb) are shown under sequence track. Probes must lie outside of homology arms to accurately detect correct integration. The D2 probe can be used with either digest, but the longer SP2 probe is useful only when probing the larger KpnI fragment. RFLP: restriction digest length polymorphism.

I tested the digest and probing patterns using control ETV6 BAC DNA compared to commercial control pooled human gDNA (Promega) and gDNA prepared from the isogenic hIPS lines MIFF 1 and 3 (Figure 4.4). An initial blot of MIFF3 DNA using Ncol with the D2 probe revealed a band that was approximately 500bp shorter than anticipated from the reference human genome (Figure 4.4 left). This discrepancy could result from either a structural variant in the MIFF3 gDNA, or due to a running artifact from the preparation of my gDNA. I tested for the latter using BAC DNA (digested with either KpnI or Ncol) run with or without mouse gDNA to recreate the loading conditions of gDNA (Figure 4.4 right). The BAC DNA produced the bands expected showing no effect on migration from the addition of mouse gDNA. In parallel I digested MIFF1 gDNA with multiple enzyme combinations and compared these to pooled hgDNA cut with either Ncol or KpnI. Pooled hgDNA gave two band sizes after Ncol digestion, indicating heterogeneity for a structural variant within in the Ncol restriction fragment. Ncol-digested MIFF1 gDNA produced a single fragment of the same size as the shorter of those from pooled hgDNA. Multiple other restriction digests of
MIFF1 gDNA also produced fragments shorter than those predicted from the reference human genome. These results indicate that MIFF1 and 3 contain an unexpected deletion within the ETV6 locus.

![Figure 4.4: Left: Discrepancy in restriction fragment length when probing MIFF3 gDNA with D2 probe compared to the fragment predicted from the human reference genome (7021bp). Right: Mapping of restriction digest patterns of ETV6 BAC DNA ± mouse genomic DNA (left) compared to commercial pooled human gDNA (Promega) (middle) and MIFF3 hIPS gDNA (right). ETV6 BAC DNA produces the predicted band sizes after digestion with both KpnI (25kb, left) and Ncol (7kb, right). Migration of ETV6 BAC DNA fragments is not altered by co-loading with mouse genomic DNA (mgDNA). Pooled hgDNA gives two band sizes after Ncol digestion, indicating heterogeneity for a deletion in the Ncol restriction fragment. Ncol digested MIFF1 gDNA produces a similar fragment to the smaller of those from pooled hgDNA (6.4kb). Using multiple other restriction enzymes over the locus also produces fragments that are shorter than predicted.](image-url)
The presence of a likely structural variant in the targeting region could have significant implications for the design of the HR donor vector and screening strategy. I next searched the Ensembl genome browser for known structural variants in the region mapped by the restriction digests above (Figure 4.5). Ensembl included a number of known deletions affecting an area including the end of the planned 3’ homology arm.

Figure 4.5: Mapping of the 2:12026416:12027153 esv223109 structural variant (green). Interrogation of structural variants in Ensembl genome browser showed a known 738bp deletion in the fragment mapped in figure 4.4 (2:12026416:12027153 esv223109). A long range PCR was designed to cover the 3’ targeting homology arm (LR1) and two shorter PCRs were designed to map the homology arm (T5) and the area surrounding the deletion (T6).

I designed a long range PCR approach to map this region further (Figure 4.5). The LR1 product (Fig 4.6 left) showed a clear difference between BAC (long), pooled hgDNA (mix) and MIFF1/3 (short). The T5 primers were designed to amplify the region 5’ of the deletion including the 3’ targeting homology arm (Fig 4.6 middle), whereas the T6 primers flank the deletion (Figure 4.6 right). As predicted the T5 product was identical between MIFF1 and BAC, whereas the T6 product was approximately 750bp shorter in the MIFF gDNA compared to BAC.
Figure 4.6: Long range PCRs of the ETV6 intron V structural variant. Left: Long range PCR (LR1) comparing ETV6 BAC DNA, pooled hgDNA and MIFF1 and 3 gDNA. Predicted band: 8510bp. BAC DNA produces predicted band, pooled gDNA appears heterogeneous for at least 3 structural variants, whereas MIFF1 and 3 are homozygous for a shorter structural variant. T5 PCR (middle) shows no difference between MIFF1 gDNA and BAC, whereas T6 PCR (right) shows the presence of a deletion in MIFF 1 and 3 gDNA compared to BAC. This fragment was sequence verified.

The T6 PCR product from MIFF hIPSC gDNA was sequenced confirming the presence of the known esv223109 structural variant shown in figure 4.5. As a result the design of the 3’ homology arm was shortened to avoid the deletion. In addition, the now validated KpnI/SP2 Southern strategy was used in subsequent blots to maximise sensitivity of the probe (compared to the shorter D2 probe).
4.2.3. Construction of TALENs

TALE Effector Nucleases (TALENs) are targeted nucleases used in genome engineering to increase gene-targeting efficiency by inducing homologous recombination in response to DNA damage. TALENs consist of a pair of DNA binding proteins linked to the DNA endonuclease FokI, which on dimerisation induces a targeted double strand break (DSB). TALEN binding and cutting of DNA is affected by DNA methylation and sequence polymorphisms (Bultmann et al., 2012). I therefore sequence verified a 150bp region of interest at the 5’ end of ETV6 intron V in MIFF3 gDNA confirming the presence of 2 single nucleotide polymorphisms (SNPs) (Figure 4.7). I further annotated the sequence to include known methylation marks in H1 human ES cells (Ensembl). Using this information, I designed three TALEN pairs to cut in the 150bp region of interest avoiding these features at sites of binding or cutting.

Figure 4.7: In silico TALEN design. Three TALEN pairs were designed to cut at the 5’ end of ETV6 intron V. Known sites of methylation and SNPs were avoided at sites of binding or cutting.

TALENs can be introduced into cells as mRNA or plasmid DNA. De novo assembly of TALEN-encoding plasmids is complicated by the highly repetitive nature of the repeat variable domain (RVD)-containing arrays. Golden gate cloning (Cermak et al., 2011) is a relatively
An efficient method for building TALEN plasmids is summarised in Figure 4.8 and detailed in methods.

**Figure 4.8: Summary of TALEN construction by Golden Gate assembly.** Stage 1 (top) involves the assembly of arrays of up to 10 RVDs into intermediate plasmids (pFus).

Plasmids are chosen on the basis of the nucleotide that they bind (NI:A, NG:T, NH:G, HD:C) and their position in the array (1-10). Repeats are released from library plasmids by the type 2S restriction enzyme BsaI and ligated into the intermediate pFus vector in a thermocycling Golden Gate reaction. Sequence specific sticky ends maintain the correct ordering of inserts. Clones of the intermediate vectors are checked by restriction digest and sequencing, before these are combined with a half last repeat and destination vector in a second golden gate reaction using the enzyme Esp3I. The destination vector contains a specific promoter and FokI endonuclease. (Spec: Spectinomycin resistance, AmpR Ampicillin resistance).

The DNA sequence is read by 2 critical RVD amino acids, but the surrounding sequence depends on their position within the array. Thus, first a library of validated plasmids was produced each containing inserts for one of four specific RVDs at a particular position in the array from 1-10. In the first golden gate reaction RVDs are pooled and released in a single
reaction by the type 2S restriction enzyme BsaI, which also opens an intermediate pFus vector. Type 2S enzymes cut away from their restriction site allowing the release of multiple fragments with sequence-specific ends that can direct re-ligation. The reaction also contains T4 DNA ligase, so during thermocycling the mixture of cut RVD inserts and pFus vector re-ligate into an intermediate plasmid containing up to 10 RVDs, in so doing losing the BsaI restriction sites. These intermediate plasmids (usually 4 per TALEN pair) are cloned into E coli and selected on plates containing spectinomycin and XGal/IPTG. White colonies are picked and screened by colony PCR to produce a band of the correct size and a smear/laddering pattern (Figure 4.9 left).

Figure 4.9: Representative examples of quality control of TALENs constructed by Golden Gate cloning. Left: Triplicate colony PCRs of intermediate vectors with different lengths of RVD arrays compared to empty vector. A band of the correct size with a laddering and smearing pattern is characteristic of success. Right: Digest of completed ETV6 TALEN3 using StuI and AatII releasing the common RVD array of expected length (2.3kb). Left and right TALEN arrays are ligated into one of two promoter-heterodimeric FokI destination vectors (DD/RR or KKR/ELD).

Clones passing screening are checked by restriction digest and sequencing. After validation the (usually two) pFus intermediate plasmids are ligated with a last repeat plasmid and
destination vector by a second golden gate reaction. After ampicillin and blue/white selection, candidate clones are again screened by colony PCR, restriction digest (Figure 4.9 right) and sequencing (not shown).

Using the above technique, three pairs of RVD arrays were constructed (Figure 4.7) and combined into two different pairs of destination vectors purchased from Addgene. ‘A’ destination plasmids were driven by the CMV promoter and contained the DD/RR obligate heterodimeric FokI endonuclease, whereas ‘B’ destination plasmids contain the highly active pCAGS promoter and the KKR/ELD obligate heterodimeric FokI nuclease made hyperactive by the ‘Sharkey’ mutation.

4.2.4. CRISPR Design

Whilst constructing the ETV6 targeting TALENs an alternative technique for the generation of site-specific double strand breaks was published using CRISPR-Cas9 (Mali et al., 2013). Although there were early concerns regarding the potential for off target effects of CRISPR, a key advantage is that CRISPR gRNAs are synthesised as gBlocks or cloned from oligonucleotides making them easy to generate. I designed a single gRNA targeting the same ETV6 break point as TALEN1 above, aiming to i) compare the efficiency of CRISPR versus TALEN and ii) as a back up approach in case the ETV6 locus was not targetable using TALENs.
**Figure 4.10: Location of ETV6 CRISPR1 gRNA seed sequence and protospacer adjacent motif (PAM) compared to TALEN sites.**

ETV6 TALEN1 includes a GG dinucleotide, required for the obligate CRISPR protospacer adjacent motif (PAM) (Figure 4.10). The 5' gRNA seed sequence to this PAM was tested *in silico* as being unique in the genome (Target: GGCAATGGAGGCTTCTGCT). The gRNA sequence, driven by the U6 promoter, was ordered as a gBlock (Integrated DNA Technologies) and cloned into pBluescript by HindIII/EcoRI digestion and re-ligation (pBS_ETV6gRNA1). Human codon-optimised Cas9 and the Cas9D10A single strand ‘nickase’ mutant (a gift from George Church (Addgene plasmid #41816), modified to include a 5’NLS by Dr Elspeth Payne) were kindly provided by Dr Elspeth Payne, UCL Cancer Institute. Both Cas9 and Cas9Nickase plasmids were verified by restriction digestion and sequencing (not shown).

### 4.2.5. Construction of ETV6 Targeting Vector by BAC Recombineering

Gene targeting vectors containing large regions (e.g. 10kb) of homology to the targeting site are constructed from bacterial artificial chromosomes (BACs) by recombineering. The complete recombineering strategy used to make the RUNX1 knock-in cassette is outlined in figure 4.11 and detailed in the methods section.
Figure 4.11: Overview of ETV6 targeting vector construction using BAC recombineering.

Step 1 – making bacteria recombineering proficient: ETV6 (TEL) BAC-containing E. coli are electroporated with the pSC101gbaATetR inducible recombineering plasmid.

Step 2 – sub-cloning homology arms: The E. coli are transformed again with linear p15aDTAmpR targeting plasmid with 50bp homology arms to the ends of the ETV6 homology arms. On arabinose induction the ETV6 homology arms are sub-cloned from the BAC into the p15a plasmid. Successful recombination is selected for with Ampicillin. The BAC is counter-selected by its large size.

Step 3 – insertion of a positive-negative selection cassette: The E. coli are transformed with linear pR6K_rpsL/gentaR positive-negative selection plasmid with 50bp homology arms to the targeted locus in ETV6. On arabinose induction the cassette is inserted into the sub-cloned ETV6 p15a plasmid. Successful recombination is selected for with Gentamicin.

Step 4: The positive-negative selection cassette is replaced with the linearised RUNX1 knock-in cassette released from a pR6K donor plasmid. Successful recombination is selected for with kanamycin. Non-recombineered clones are counter-selected with streptomycin. The psc101 plasmid is excluded by growing the bacteria at 37°C.
**Step 5:** The final targeting vector is cloned, checked and linearised prior to nucleofection into hPSCs.

To test the ability to target the *ETV6* allele using the above nucleases I first generated a simple *ETV6* targeting vector containing a PGK_NeoR positive selection cassette. This cassette was produced by PCR amplification from a pR6K plasmid (provided by Francis Stewart, University of Dresden), using primers with 50bp of homology to the *ETV6* TALEN1 cut site (Figure 4.12).

![Figure 4.12](image)

*Figure 4.12: A test of homologous recombination in hPSCs was performed using a simple NeoR selection cassette. The cassette was PCR amplified from a pR6K donor using primers containing 50bp homology arms and recombineered directly in place of step 3 above.*

Although this PCR reaction produced a band of the expect size, it provided insufficient DNA for use in recombineering. I therefore purified the PCR product by gel extraction and successfully re-amplified the cassette using shorter primers designed to the ends of the new 50bp homology arms (Figure 4.13).
Figure 4.13: Construction of the PGK NeoR targeting test vector. **Left:** PGKNeoR was amplified by PCR to include 50bp homology arms to ETV6 and the band gel extracted (1641bp). **Middle:** PGKNeoR band checked after gel extraction. **Right:** PGKNeoR was further amplified by PCR using short primers to the ends of the ETV6 homology arms to provide sufficient yield for recombineering.

DH10B E coli were purchased containing the ETV6 BAC clone used for the reference human genome (Source Biosciences). E coli were grown in liquid culture overnight and made recombineering proficient by the electroporation of the psc101_BAD_gbaA_tetR plasmid (Step 1, Figure 4.11). This plasmid inducibly expresses λ phage recombination proteins (gba) from the arabinose inducible BAD promoter. The psc101 origin of replication is heat sensitive, such that it is maintained at 30°C, but lost at 37°C. Expression of the recombination proteins in response to arabinose allows the introduction of donor single strand DNA containing homologous ends into replicating plasmid DNA by invasion of the lagging strand.
The transformed E coli were grown overnight at 30°C under selection with chloramphenicol and tetracycline. The next day the 10kb homology arms were sub-cloned from the BAC into the generic gene targeting backbone plasmid p15a_DTA_AmpR (Step 2, Figure 4.11). The plasmid expresses the diphtheria toxin A (DTA) gene as a eukaryotic negative selection cassette to select against non-targeted integration into hPSCs. The E coli were electroporated with a PCR product from the p15a plasmid containing 50bp homology arms aligned so as to sub-clone the ends of the targeting vector homology arms. After arabinose induction of the recombineering proteins, the E coli were selected overnight in ampicillin in liquid culture.

The next day DNA was prepared from a bulk culture (grown at 37°C to eliminate the psc101 plasmid) and digested for evidence of successful recombination (Figure 4.14 left). Having confirmed successful sub-cloning of ETV6 into p15a, E coli grown overnight at 30°C to maintain the psc101 plasmid were stored as glycerol stocks at -80°C for use in the final RUNX1 targeting vector. A further outgrowth of these cells was electroporated with the PGK_NeoR PCR product produced above (Figure 4.13 right). After overnight selection in kanamycin and ampicillin at 37°C (to eliminate the psc101 recombineering plasmid) the bulk E coli were checked for successful recombination by restriction digest of plasmid DNA (Figure 4.14 right).
Figure 4.14: Check of non-clonal DNA of intermediate vectors produced during recombineering. Left: p15a_ETV6 - predicted bands: Ncol (11183), SmaI (9062, 2121), XbaI (6075, 2761, 1845, 502), HindIII (5760, 3496, 2241, 286). Right p15a_ETV6_NeoR – predicted bands KpnI (12724), XbaI (6075, 4302, 1845, 502), HindIII (5760, 5037, 2241, 286).

After confirmation of successful insertion of the PGK_NeoR cassette into the p15_ETV6 plasmid the DNA was used to transform DH5α E coli, cells selected on Kan/Amp plates and candidate clones picked for validation. Four clones were analysed by restriction digest (Figure 4.15), which showed that clones 2 and 3 consisted of concatamers of the p15a_ETV6 plasmid with and without the NeoR insert.
Figure 4.15: Four clones of p15a_ETV6_NeoR were checked by restriction digest. Predicted bands - KpnI 12724, XbaI 6075, 4302 (vs. 2761 in p15aETV6), 1845, 502; HindIII 5160, 5037 (vs. 3496 in p15aETV6), 2241, 286. Clones 2 & 3 show extra bands consistent with concatamers of p15a ETV6 and p15aETV6NeoR. Clone 1 was rejected due to a 47bp deletion in the PGK promoter found on sequencing.

The two remaining clones were sequenced across all recombineering junctions, identifying a 47bp deletion in the PGK promoter of clone 1. Thus, p15a_ETV6_NeoR clone 4 was validated and used to test the efficiency of gene targeting ETV6 in MIFF1 and 3 hiPSCs.

4.2.6. Transfection of hPSCs by Nucleofection

Successful gene targeting requires the efficient transfection of both the TALEN/CRISPR nuclease-encoding plasmids and also the large linearised gene-targeting vector, the latter ideally directly into the nucleus. As opposed to transient transfection, which can be readily achieved through lipofection, transfection with gene targeting constructs is most efficiently achieved by electroporation-based techniques (Andrew Smith, personal communication).
Nucleofection® (Lonza) is a commercial electroporation system that offers higher efficiency transfection compared to traditional electroporation; although the device and reagents are expensive, each nucleofection requires five fold less of both DNA and hPSCs than electroporation, mitigating the additional cost (William Skarnes, personal communication).

I first optimised nucleofection conditions for the hPSC lines MIFF1 and 3 according to the manufacturer’s guidelines. hPSCs were transfected using a Nucleofector® II device (Lonza). Two manufacturer-supplied solutions were tested against the 5 nucleofection programmes suggested for hPSCs using the manufacturer-supplied GFP-expressing plasmid (pmaxGFP®).

![GFP](image)

**Figure 4.16:** Optimisation of nucleofection conditions for MIFF1 and MIFF3 hiPSCs. Five nucleofector II programmes (A012, A013, A023, A027 & B016) were tested using two nucleofection solutions (1 left and 2 right) from the Lonza human stem cell kit.

Transfection of MIFF3 was more efficient than MIFF1 in all conditions tested (median difference 40%, range 16-65%). Overall Solution 2 and programme B016 gave the highest percentage (24%) GFP positive cells in MIFF3 hiPSCs (Figure 4.16).
4.2.7. Assessment of Homologous Recombination Efficiency

I next tested the ability of my TALEN/CRISPR nucleases and ETV6 targeting vector to insert the PGK_NeoR cassette into MIFF1 and 3 hiPSCs. 2x10^6 hIhPSCs were transfected with 2μg ScaI linearised p15a_ETV6_PGK_NeoR construct and 4μg each of Cas9D10A/Cas9 and ETV6 gRNA or 4μg, of each of the 6 TALEN pairs produced above. After 48 hours hIhPSCs were selected in G418 and individual colonies picked at 8-10 days into 96 well plates prior to screening by Southern blot hybridisation using the KpnI/SP2 probe strategy. A summary of targeting efficiencies is presented in Figure 4.17.

![Figure 4.17: Summary of homologous recombination frequencies using the p15aETV6NeoR knock-in cassette into MIFF1 and MIFF3 hIhPSCs. Percentage frequency of knock-in bands was calculated as a proportion of the number of clones producing a visible wild type band on screening by Southern blot.](image)

As expected in all conditions tested the majority of resistant colonies were not targeted at the ETV6 locus consistent with random integration of the NeoR cassette. No targeted integrations were observed in the absence of a targeting nuclease. TALEN pairs 1 and 3
produced a 2-5% knock-in rate, but the cells exposed to TALEN2 mostly died and no successful knock-in clones were isolated. The most striking result was the very high efficiency of targeted knock-in with both Cas9 and the Cas9D10A nickase, even though the latter was guided by a single gRNA. The high targeting efficiency and low chance of off target mutagenesis using the Cas9D10A nickase indicated that this was the best nuclease for targeting the ETV6 locus. Although MIFF1 hIPSCs had a slightly higher knock-in frequency using CRISPR nucleases, the colony morphology of MIFF3 was found to be more amenable to picking clonal colonies and therefore MIFF3 was used in subsequent targeting experiments.

4.2.8. Construction of RUNX1 Knock-in cassette

Having validated the ability of the p15a_ETV6 vector to target a selection cassette to the ETV6 locus, I next constructed the RUNX1 targeting cassette and recombineered it into the p15a_ETV6 vector. The sequence for the cassette design described in figure 4.1 was constructed in silico. The cassette was flanked by 50bp homology arms to the ETV6 locus to allow insertion into the p15a_ETV6 vector by recombineering. The cassette was submitted for commercial synthesis (GeneArt) and due to its large size and complexity was delivered as two sequence-verified fragments: a 4114bp 5’ fragment containing the RUNX1 knock-in cassette flanked by XhoI and SbfI restriction sites; and a 1818bp 3’ NeoR selection cassette, flanked by SbfI and XhoI sites. These fragments were released from their plasmid backbones by restriction digest and purified by gel extraction (Figure 4.18 left).

Linear DNA successfully recombineers into the desired plasmid relatively rarely, whereas circular plasmid DNA (e.g. PCR template or left after release of the cassette) will transform bacteria very efficiently, resulting in the same antibiotic sensitivity pattern. To avoid this problem of ‘background’ transformation of plasmid DNA, recombineering cassettes are
released from pR6K donor plasmids that can only replicate in \( \text{pir}^{116^+} \) E coli. Therefore the synthesised 5’ and 3’ RUNX1 cassette fragments needed to be both ligated together and inserted into a pR6K vector. In case the pR6K vector was not amenable to standard cloning, I also designed the complete RUNX1 cassette to be flanked by both attL1/L2 gateway sites and 50bp recombineering arms homologous to the pR6K_AmpR vector, although I did not resort to either of these approaches. As ligation of multiple inserts can reduce cloning efficiency, and I had not used the pR6K vector before, I ligated the 5’ and 3’ fragments into both pR6K_AmpR and also pBluescript (pBS) vectors (Figure 4.18 right).

Figure 4.18: Construction of the RUNX1 knock-in cassette. Left: Gel extraction of pR6K vector linearised by Xhol (1818bp), 3’ RUNX1 insert cut with SbfI/Xhol (1818bp), 5’ RUNX1 insert cut with Xhol/SbfI (4114bp) and the pBS vector linearised by Xhol (2958bp). Right: Check of ligation of RUNX1 5’ and 3’ fragments into pR6K (left) and pBS (right). Negative control: Vector self-ligations. Positive control: ladder self-ligation.

Although both ligation appeared to work well compared to vector self-ligation controls, no successful clones were produced from ligating both inserts into the pR6K vector, however, the co-ligated insert was retrieved at high efficiency from the pBS vector. Thus, the complete RUNX1 cassette was released from a sequence verified clone of pBS_RUNX1 by Xhol.
digestion and gel purified (Figure 4.19 left). In addition a new batch of pR6K_Amp vector was XhoI digested and dephosphorylated (Figure 4.19 middle). The results of this ligation are shown in Figure 4.19 (right), and produced multiple successful pR6K_ETV6RUNX1 clones.

![Image](image_url)

*Figure 4.19: Ligation of the RUNX1_NeoR cassette into pR6K plasmid. Left: Gel extraction of the combined RUNX1 cassette from pBS vector by XhoI (5932bp). Middle: pR6K vector checked post XhoI linearisation and phosphatase treatment (1818bp). Right: pR6K ligation with single RUNX1 XhoI knock-in insert. Negative control: Vector self-ligations. Positive control: ladder self-ligation.*

These clones were tested by restriction digest (Figure 4.20) and the entire RUNX1 cassette was successfully sequence verified for clone 2. This pR6K_ETV6RUNX1 plasmid was then used for recombineering into the p15a_ETV6 targeting vector produced above.
Figure 4.20: Five pR6K_RUNX1 clones triple digested. Predicted bands - Scal (7736), XhoI (1818, 5918), HindIII (1774, 2817, 3145). Clone 2 was sequence verified and used as recombineering donor. Sizes of the 1kb+ ladder are indicated.

Recombineering works most efficiently when under both positive selection for the desired event, and negative selection against the intermediate construct being engineered. I therefore PCR amplified the generic rpsL/gentaR negative/positive selection cassette from a pR6K vector using primers containing 50bp homology arms to the ETV6 locus (Step 3, Figure 4.11). This cassette was introduced into the p15a_ETV6 vector by recombineering and positively selected with gentamicin. Gentamicin is unstable and appropriate controls were essential to ensure adequate selection. The addition of the rpsL cassette made the DH10B E coli (intrinsically resistant to streptomycin) sensitive to streptomycin allowing negative selection for the p15a_ETV6_rpsL/GentaR intermediate plasmid in subsequent recombineering steps.
The RUNX1 knock-in cassette flanked by ETV6 specific 50bp homology arms was released from the pR6K vector by Ascl NotI double digest (Figure 4.21 left). This was recombineered into the p15a_ETV6_rpsL/GentaR under positive selection with ampicillin and kanamycin and negative selection using streptomycin (Step 4, Figure 4.11). Successful recombineering was checked by restriction digest of the bulk culture (Figure 4.21 right) and the plasmid was cloned in DH5α E coli.

**Figure 4.21:** Recombineering RUNX1 cassette into p15a_ETV6 vector plasmid. Left: RUNX1 knock-in cassette (5932bp) released from pR6K vector (1818bp) by Ascl NotI digestion. Right: Right: *BglII* digest of bulk recombineered p15a_ETV6_RUNX1. Predicted bands – 5698, 4917, 4442, 809, 690.

Four p15a_ETV6_RUNX1 clones were isolated and validated by restriction digest (Figure 4.22). Clone 1 was sequence verified over all recombineering junctions and through the entire knock-in cassette.
Figure 4.22: BglII digest of 4 clones of p15a_ETV6_RUNX1. Top short run. Bottom – longer run. Predicted bands – 5698, 4917, 4442, 809, 690. Clone 1 sequenced correctly and was used for gene targeting.

I next tested the functionality of the site-specific recombination sequences by transient expression of Cre and Flp recombinase in p15a_ETV6_RUNX1 clone 1 containing E coli. Analysis of bulk DNA by restriction digest showed induction of recombination on exposure to both Cre and Flp (Figure 4.23 left). On testing the LoxP sites successful recombination occurred even in the absence of arabinose induction consistent with some background expression from the BAD promoter and the high efficiency of Cre mediated recombination. The validated p15a_ETV6_RUNX1 vector was prepared by Maxi Prep (Qiagen) and linearised with XhoI for use in hiPSCs (Figure 4.23 right)
Figure 4.23: Final checks of p15a_ETV6_RUNX1 vector. Left: Check of site-specific recombination sites in E. coli containing DNA from p15a_ETV6_RUNX1 clone 1. E. coli were transformed with arabinose-inducible Cre or Flp recombinase. DNA from bulk cultures was prepared with or without Arabinose (Ara) induction of recombination. BglII digest predicted bands – Unmodified plasmid: 5698, 4917, 4442, 809, 690; Cre recombined 6412, 4917, 809, 690; Flp recombined 5898, 4917, 3017, 809, 690. Both enzymes worked as expected in the Ara-induced controls. Cre recombinase resulted in recombination even in the absence of arabinose consistent with its high efficiency of recombination. Right: XhoI linearisation of p15a_ETV6_RUNX1 clone 1 prior to nucleofection into hPSCs (band 16686).

4.2.9. Generation of ETV6-RUNX1 knock-in clones

Using the reagents and techniques validated above I next targeted the RUNX1 knock-in cassette into ETV6 intron V of MIFF3 hPSCs. In order to maximise the chances of successful targeting I used both Cas9 and Cas9D10A nickase approaches. Overall these approaches
produced 189 clones for screening by Southern blot hybridisation (38 from Cas9, 151 from Cas9D10A) (Figure 4.24).

**Figure 4.24: Screening Southern blot of MIFF3 hIPSCs nucleofected with linearised p15a_ETV6_RUNX1 targeted using CRISPR Cas9 and Cas9D10A. WT 25kb, KI 8.7kb.**

The screening Southern blot identified 17 candidate knock-in clones, of which three 1.2, 1.3 and 1.8 were generated by Cas9. In general Cas9 targeted clones failed to grow well after picking and clone 1.3 did not survive thawing. In addition clone 1.2 did not maintain appropriate pluripotent morphology in vitro and was found to have an inappropriate size band on internal Southern probing. Furthermore the clone targeted with Cas9 in lane 9, top row, figure 4.24 had a knock-in band of the wrong size, consistent with the experience of other groups (Konstantinos Anastassiadis, personal communication). Together these findings
suggest that Cas9 may be having functionally significant off target effects in MIFF3 hPSCs and therefore Cas9 targeted clones were not taken forward for validation.

Four knock-in clones generated using the Cas9D10A nickase were validated and used in this study (clone numbers 1.6, 1.9, 2.1, 2.8). Targeting was checked by Southern blot using the SP2 external probe and clonality was checked visually by equal stoichiometry of the knock-in and wild type allele bands (Figure 4.25).

![Figure 4.25: Confirmatory Southern blot of successfully targeted ETV6-RUNX1 hIPSCs targeted using CRISPR Cas9D10A. KpnI digest, external SP2 probe. WT 25kb, KI 8.7kb.]

To ensure that the ETV6 allele was the only site of integration the knock-in cassette was internally probed using a NeoR specific PCR product on BamH1 digested DNA, revealing a single band in all four clones indicating a single integration event (Figure 4.26).

![Figure 4.26: Confirmatory Southern blot using internal probe on successfully targeted ETV6-RUNX1 hIPSCs targeted using CRISPR Cas9D10A. BamHI digest, internal Neo probe looking for single site of genomic integration. KI 18.5kb.]

167
The candidate clones were checked for ongoing expression of pluripotency markers, including qPCR for OCT3/4 and NANOG (Figure 4.27), nuclear staining of OCT4 (Figure 4.28) and surface expression of the pluripotency associated antigens SSEA3 and Tra 1-60 (Figure 4.29).

![Expression vs MIFF3 graph]

**Figure 4.27:** Quantitative gene-expression analysis of unmodified MIFF3 and 4 successful ETV6-RUNX1 knock-in clones for the pluripotency genes OCT3/4 and NANOG. Expression is normalised to β-ACTIN and presented relative to unmodified MIFF3. Mean ± SD, n=2.

![Images of hPSC colonies]

**Figure 4.28:** Immunostaining for nuclear OCT4 (Santa Cruz) on fixed hPSC colonies. Anti-rabbit Cy5 secondary antibody (Jackson labs). 40x magnification.
Figure 4.29: Flow cytometry analysis of the pluripotency surface markers TRA1-60 (PE) and SSEA3 (AF647) on cells grown in StemFit media.

Finally, the intermediate passages of the 4 knock-in clones were confirmed to have a normal karyotype by G banding (kindly performed by Duncan Baker and Christian Unger, University of Sheffield) (Figure 4.30).
Figure 4.30: Karyotype of four middle passage ETV6-RUNX1 knock-in hPSC clones grown in mTeSR1. Courtesy of Duncan Baker.

Overall these results demonstrate successful gene targeting of the RUNX1 knock-in cassette into the ETV6 locus in MIFF3 hPSCs resulting in the generation of four ETV6-RUNX1 knock-in clones that retained a normal karyotype and ongoing expression of pluripotency markers.

Of note, at this time a new source of hPSC media became available for testing in our lab (StemFit, Ajinomoto) and when cultured in these conditions the ETV6-RUNX1 knock-in clones had preferable morphology to when cultured in mTeSR1. After validation that B cell differentiation from hPSCs grown in StemFit was equivalent to that from mTeSR1, I used StemFit for all subsequent experiments.
4.2.10. Removal of Knock-in Cassette by Cre Recombinase

hPSC cultures are heterogeneous and cloning of genome engineered hPSCs risks isolating cells with different functional characteristics confounding the analysis of the effects of the knock-in gene. As an additional control we therefore floxed the RUNX1 cassette from hPSCs using Cre recombinase so as to be able to unequivocally attribute the cellular and molecular phenotypes seen in the knock-in cells to the presence of ETV6-RUNX1. Although Cre is an efficient enzyme, in the absence of positive selection for recombination this proved challenging. Two approaches were used to introduce Cre recombinase: lipofection with a puromycin selectable Cre expressing plasmid and exposure to cell permeable TAT-Cre (Millipore). Cells were then sub-cloned either by direct FACS sorting into 96 well plates or by picking colonies after limiting dilution. Only 1 of 1300 FACS sorted cells survived and was not successfully reverted. Of the clones picked after limiting dilution 2 sub-clones of ETV6-RUNX1 knock-in 2.8 were found to be successfully reverted (D5 and D6, Figure 4.31) and a further two had evidence of a mix of knock-in and floxed cells. These two reverted lines were used as additional controls for the observed phenotypes of knock-in cells compared to unmodified control MIFF3 hPSCs.
Figure 4.31: Southern blot of Cre reverted ETV6-RUNX1 knock-in hPSC clone 2.8 showing two successfully reverted sub-clones (D5 and D6). Kpnl digest, NeoR internal probe. KI 5kb, Cre floxed 17kb.

4.3. Discussion

Genome engineering of hPSCs has traditionally proved challenging due to difficulties in cloning and the low efficiency of homologous recombination compared to that seen in mouse ESCs. The development of targeted nucleases such as TALENS and CRISPR/Cas9 has made gene targeting in hPSCs feasible, increasing both the number of targetable loci and the efficiency of gene targeting. Nevertheless, in my experience selecting and sub-cloning successfully targeted hPSCs remains laborious.

The high efficiency of homologous recombination using CRISPR/Cas9 seen here compared to TALENs is consistent with the published literature (Ding et al., 2013). The gRNA used here targets close to that of the binding site for ETV6 TALEN1, thus it seems unlikely that the differences in knock-in efficiency can be explained by the underlying DNA structure or methylation. Unlike TALENS, which are artificially fused to the FokI endonuclease for use in genome engineering, CRISPR/Cas9 has specifically evolved in bacteria to maximise the efficiency of DNA cutting. Although the rates of off target cutting are debated (Doudna and Charpentier, 2014), in these experiments Cas9 appeared toxic to MIFF3 hPSCs resulting in
changes in morphology, survival and the identification of a knock-in band of the wrong size. Although less mutagenic, published reports indicate that Cas9D10A nickase is markedly less efficient than Cas9 and have suggested alternative approaches such as double nickases (Cong et al., 2013). Here, the Cas9D10A nickase directed by a single gRNA was efficient at inducing targeted homologous recombination and resulted in less detectable toxicity to the hiPSCs than Cas9. Given the low risk of off target mutagenesis from Cas9D10A, the lack of any predicted off-target binding sites in coding exons and the generation of four separate knock-in clones and 2 floxed revertants as controls, I have not invested in targeted sequencing of the predicted CRISPR off target sites.

The use of large 10kb homology arms on the ETV6 targeting vector resulted in the requirement for colony screening by Southern blot hybridisation. Such vectors are based on those used in mouse ESCs without targeting nucleases (Smithies et al., 1985; Thomas et al., 1986) and are best constructed by BAC recombineering. More recent reports using CRISPR nucleases in hPSCs suggest that targeting vectors can use very short homology arms of just a few hundred base pairs, which can be produced by PCR amplification of BAC or genomic DNA. Such a modification to vector design could obviate the need for BAC recombineering and allow colony screening by PCR. Nevertheless, my experience is that once established BAC recombineering and Southern screening are relatively robust and scalable techniques.

Although derived and validated in mTeSR1/matrigel conditions, over time the knock-in hPSCs lost colony morphology and tended to grow as a diffuse monolayer. In contrast, when changed to StemFit/matrigel conditions the knock-in clones generated compact uniform colonies that were stable in long-term culture and easier to harvest for B cell differentiation. Whether this change of morphology in mTeSR1 relates to the destabilisation of the pluripotency circuit by ETV6-RUNX1 is uncertain. This seems unlikely, however, given both
the negligible level of ETV6 expression in normal hPSCs (P Andrews, personal communication) and the ability to grow hPSCs expressing high levels of virally transfected ETV6-RUNX1 (Ram Malladi, personal communication).

Successful sub-cloning of Cre reverted clones was made difficult by the lack of positive selection and the difficulties in isolating pure sub-clones of hPSCs. Despite sorting or picking in excess of 2000 clones, only two pure clones were isolated, both using cell permeable TAT-Cre on ETV6-RUNX1 knock-in clone 2.8. I have not yet succeeded in removing the PGK-NeoR cassette despite two attempts using TAT-Flp, and the commercial availability of this reagent has recently proven limiting. It is reported that PGK_NeoR cassettes can have significant effects on gene expression at distant loci (Strathdee et al., 2006; Zhu et al., 2015) and therefore the removal of just the RUNX1 cassette with the ongoing presence of the NeoR cassette is a more stringent control for the specificity of the impact of ETV6-RUNX1.

Overall this work has derived four ETV6-RUNX1 knock-in clones and two RUNX1 reverted sub-clones on the MIF3 hPSC background. These lines were next used to test the functional and molecular impact of ETV6-RUNX1 expression on human fetal lymphopoiesis, using in vitro B cell differentiation of hPSCs as a model system.
CHAPTER 5

Characterisation of B lymphoid output from

ETV6-RUNX1 hPSCs
5. CHAPTER 5: Characterisation of B lymphoid output from ETV6-RUNX1 hPSCs

5.1. Introduction

Having validated the \textit{in vitro} B cell differentiation of hPSCs as a model of fetal lymphopoiesis and generated ETV6-RUNX1 knock-in and reverted control hiPS lines, I next tested the functional and molecular effects of ETV6-RUNX1 expression during \textit{in vitro} B cell differentiation of knock-in hiPSCs. ETV6-RUNX1 alone is insufficient to generate leukaemia as a single mutation and is generally considered to have ‘weak’ oncogenic activity (Mori et al., 2002). The phenotype seen most consistently in other models of ETV6-RUNX1 pre-leukaemia has been a block in B cell differentiation either at the proB or preB cell stage indicating that ETV6-RUNX1 predominantly acts in the pre-leukaemic setting by arresting B cell development (Morrow et al., 2004; Schindler et al., 2009; Tsuzuki and Seto, 2013; Tsuzuki et al., 2004). In addition a small number of papers have reported enhancement of self-renewal capacity (Hong et al., 2008; Morrow et al., 2004).

Evidence from cord blood models suggests that ETV6-RUNX1 pre-leukaemia is itself organised as a functional hierarchy (Hong et al., 2008). Given that the most significant difference seen between adult and fetal B lymphoid hierarchies was the B-myeloid programming of the IL7R progenitor, I was particularly interested to examine the impact of ETV6-RUNX1 on this population, which constitutes a putative \textit{in utero} target cell. I therefore used flow cytometry to look for evidence of a B cell differentiation arrest with or without associated expansion of upstream compartments. I then FACS sorted the available B lymphoid progenitor compartments to investigate the transcriptional impact of ETV6-RUNX1 expression during fetal B lymphoid commitment.
5.2. Results

5.2.1. Analysis of ETV6-RUNX1-hiPSC-derived B lymphopoiesis by flow cytometry

Consistent with a block in B cell differentiation, ETV6-RUNX1 expressing clones produced either no (clones 1.6 and 1.9) or low (clones 2.1 and 2.8) numbers of immunophenotypically-defined CD19⁺ proB and preB cells relative to control cultures. However, an IL7R progenitor could be detected at numbers similar to that in MIFF3 control cultures in all four knock-in lines (an example is shown in Figure 5.1).

![Flow cytometry analysis](image)

**Figure 5.1: Representative flow cytometric analysis of control MIFF3 (top panel) and ETV6-RUNX1 hPSC clone 2.8 (bottom panel) differentiated in vitro and analysed for venus reporter (left), proB (CD34⁺CD19⁺) and preB (CD34⁻CD19⁺) cells (middle) and CD19 negative IL7R⁺ progenitors (right). Viable cells were gated as indicated.**

Considerable variation in B cell output exists even when differentiating control hiPSCs. To gain an appreciation of the consistency of the B cell differentiation block seen from ETV6-RUNX1 hiPSCs compared to control MIFF3 I analysed the frequencies of IL7R progenitor,
proB and preB cell compartments as a proportion of venus positive or venus negative CD45+ blood cells seen in D31 ETV6-RUNX1 or MIFF3 control differentiations respectively (Figure 5.2). Overall the IL7R progenitor was not significantly reduced in number compared to MIFF3 controls, whereas proB cells were much reduced and preB cells were rarely seen at this time-point.

![Graph showing frequencies of IL7R+ progenitor, proB and preB cells in MIFF3 cells compared to ETV6-RUNX1 iPSCs analysed at D31. Percentage of Venus— (MIFF3) or Venus+ (ETV6-RUNX1 hIPS) CD45+ blood cells. Mean±SD. Each dot represents one replicate. Of note the majority of ETV6-RUNX1 hIPSC differentiations undertaken were CD34 enriched prior to FACS sorting for molecular analysis and therefore did not yield sufficient un-enriched material to be used in this analysis, thus only 6 ETV6-RUNX1 replicates are included here.](image)

Rarely, both Venus— and Venus+ cells were seen together in the culture. This could either represent contamination with wild type cells or silencing of the knock in allele. This phenomenon was most apparent from knock in clone 1.9 (an extreme example is shown in Figure 5.3, left), which was surprising given that clone 1.9 always produced stoichiometric knock-in and wild type bands on Southern blot consistent with a pure knock-in clone rather than contamination by wild type MIFF3 cells (Figure 4.25). Although rarely seen and usually
providing only a minor Venus− population, such mixed clones offered an additional opportunity to control for the effect of ETV6-RUNX1. In mixed cultures the Venus− population had robust B cell production compared to Venus+ cells, which produced the IL7R progenitor, but little or no B cell output (Figure 5.3 middle and right). This demonstrated that loss of B cell production is directly related to ETV6-RUNX1 expression, rather than dependent on any experimental variation in culture conditions and further suggested the impact of ETV6-RUNX1 is cell autonomous.

Figure 5.3: Occasionally ETV6-RUNX1 iPSC clones had Venus positive (red) and Venus negative (blue) cells identified in the same differentiation culture. Frequencies show percentage of CD45+V+ or CD45+V− cells respectively for this sample (ETV6-RUNX1 clone #1.9).

RT-qPCR analysis of sorted compartments from control MIFF3 and Venus− and Venus+ ETV6-RUNX1 cells showed a direct relationship between Venus and ETV6-RUNX1 expression validating the Venus reporter (Figure 5.4).
Figure 5.4: qPCR of ETV6-RUNX1 in MIFF3 (grey) compared to Venus⁺ (green) and Venus⁻ (brown) ETV6-RUNX1 IPS-derived cells. Data normalised to expression of GAPDH. Mean±SD. Number of replicates is indicated in the figure.

Loss of the remaining allele of \textit{ETV6} is a common second hit in ETV6-RUNX1 ALL suggesting \textit{ETV6} exerts a tumour suppressor role (Fenrick et al., 2000). qPCR on sorted compartments showed that \textit{ETV6} expression was comparable between MIFF3 and Venus⁺ and Venus⁻ ETV6-RUNX1 knock-in populations in all compartments tested, indicating that \textit{ETV6} haploinsufficiency does not contribute to the observed phenotype (Figure 5.5).
Figure 5.5: qPCR of ETV6 (TEL) in MIFF3 (grey) compared to Venus+ (green) and Venus- (brown) ETV6-RUNX1 IPS-derived cells. Data normalised to expression of GAPDH. Mean+SD. Number of replicates is indicated in the figure. ETV6 probe used targets the 3’ portion of ETV6, therefore does not detect ETV6-RUNX1 fusion.

I next tested whether ETV6-RUNX1 CD34+ progenitor cells from D31 of culture had enhanced engraftment capacity compared to control MIFF3 (which was anticipated not to engraft as per published reports) (Slukvin, 2013). I MACS-enriched CD34+ cells from D31 cultures and 10^5 CD34+ cells were injected into the livers of newborn NSG mice (mouse experiments kindly performed by Dr Yanping Guo). The CD34+ compartment encompassed a range of differentiated intermediates from multipotent progenitors to proB cells. Transplantation into newborn mouse has been previously been shown to generate definitive haematopoiesis from ‘preHSC’-derived material (Yoder et al., 1997b) and was therefore chosen as the engraftment method most likely to succeed. Mice were sacrificed at 2 and 3 months, but no evidence of any human engraftment was seen from either ETV6-RUNX1 or control cells (MIFF3 n=12 mice, ETV6-RUNX1 1.9 n=8, ETV6-RUNX1 2.1 n=2). This negative result was consistent with the published lack of evidence of haematopoietic engraftment.
from hPSC-derived material and therefore additional attempts at deriving engraftment from ETV6-RUNX1 were not pursued.

5.2.2. Transcriptomic analysis of ETV6-RUNX1-hiPSC-derived B lymphopoiesis

I next analysed the transcriptional programming of the ETV6-RUNX1 progenitor compartments. I performed global RNASeq on ETV6-RUNX1 IL7R progenitors and those small numbers of proB cells that were produced in ETV6-RUNX1 hPSC cultures. Overlaying these data on the PCA of primary human and control hPSC-derived cells (Figure 3.16) showed that ETV6-RUNX1⁺ hPSC-derived CD19⁺ 'proB' cells (purple squares) clustered with CS17 FL (blue triangles) and control hPSC-derived IL7R progenitors (green triangles), rather than control hPSC-derived proB cells (green squares) (Figure 5.6). This suggested that the global gene expression program seen in ‘proB’-like ETV6-RUNX1 expressing cells is very different from that in wild-type cells with a similar immunophenotype. In fact these ETV6-RUNX1 expressing ‘proB’ like cells closely resemble the IL7R progenitor at the global transcriptional level, further indicating that ETV6-RUNX1 arrests B cell differentiation in the vicinity of the IL7R progenitor.
Figure 5.6: RNASeq data from ETV6-RUNX1-expressing hPSCs overlaid (haloed) on the existing PCA map calculated from primary samples and overlaying control hPSC-derived data (haloed) for comparison. PCA1 and PCA2 are shown; each dot represents one sample.

PCA constructed by Javier Herrero.

With assistance from Charlotta Boiers, I analysed these compartments by single cell qPCR to assess the extent of heterogeneity and co-expression of lineage-affiliated genes. A new PCA calculated using the previous primary and hPSC-derived single cell data with the ETV6-RUNX1 expressing cells showed the majority of ETV6-RUNX1 IL7R progenitors (purple triangles) clustered with the early D10 hPSC/CS17 FL progenitors (light green triangles, light blue ellipse), with relatively few overlapping with CS20 FL (dark blue ellipse). More strikingly the ETV6-RUNX1 ‘proB’ cells (purple squares) were markedly heterogeneous, spanning from the CS17 FL IL7R progenitor compartment to the CS20 FL/CB proB cluster (Figure 5.7). The PCA suggested that the differentiation trajectory from IL7R progenitor to ‘proB’ cell is diverted by expression of ETV6-RUNX1.
Figure 5.7: PCA of single cell qPCR data calculated from primary cells, control hPSCs and ETV6-RUNX1 hIPS cells from D10 (control hPSC IL7R+ progenitor only) and D31 (IL7R+ progenitor and proB cells). Each dot represents a single cell. Top: MIFF3 and H1 hPSCs plotted on the new PCA calculation for comparison. Bottom: ETV6-RUNX1 samples. 36 genes were used for the PCA, n=2-3 replicates. Single cell qPCR performed by Charlotta Boiers. PCA constructed by Javier Herrero.

Examination of the myeloid and lymphoid gene expression programs showed that the D31
ETV6-RUNX1 IL7R progenitors exhibited an almost uniform myeloid signature (Figure 5.8 left) similar to that seen in early D10 hPSC/CS17 FL IL7R progenitors (Figure 3.8). In contrast ETV6-RUNX1 ‘proB’ cells showed widespread co-expression of a B-myeloid signature, significantly enhanced relative to that in control hPSC-derived proB cells.

**Figure 5.8: Single cell qPCR data of ETV6-RUNX1+ IL7R progenitors and proB cells from D31. Each column represents a single cell. Coloured by CT value. Genes labeled in red are myeloid, blue lymphoid and green B cell genes. Only cells expressing ETV6-RUNX1 are shown. 56-85 cells investigated per population, n=3-4.**

To gain a broader appreciation of how ETV6-RUNX1 was affecting lineage-associated gene expression I used gene-set enrichment analysis (GSEA) of the RNASeq data interrogating previously reported HSC, Megakaryocyte/Erythroid (MkE), myeloid and B cell associated signatures (Laurenti et al., 2013) (Figure 5.9). The ETV6-RUNX1+ IL7R progenitor and ‘proB’ cells were highly enriched for HSC, MkE and myeloid genes compared to hPSC-derived controls. Both the ETV6-RUNX1 progenitor and ‘proB’ cells displayed components of B cell identity, although the early B cell signature was less well developed in than in controls. This suggested that expression of ETV6-RUNX1 imposes or maintains multilineage gene
expression priming characteristic of primitive cells, which cannot be fully resolved as the cells initiate B lineage commitment.

Figure 5.9: Gene-set enrichment analysis of ETV6-RUNX1 hiPS (red) and control MIFF3 (blue)-derived IL7R⁺ progenitors (top) and proB cells (bottom) at D31 of differentiation. Lineage affiliations of gene sets used are indicated above (Laurenti et al., 2013). NES: normalised enrichment score. Kindly performed by Dr Dapeng Wang.

GSEA provides a relative measure of lineage commitment between samples. To compare the absolute extent of B lineage commitment in the ETV6-RUNX1 hierarchy compared to primary and hPSC-derived cells, the RNASeq data was visualised on a 3D scatter plot, with axes defined by the HSC, myeloid (CMP/GMP) and B lymphoid (early B/proB) gene-sets above (Figure 5.10; left: cell state; right: source of cells) (Laurenti et al., 2013). The adult BM (red), FL (blue) and hPSC-derived (green, Figure 5.10 right) differentiation hierarchies navigated similar differentiation pathways, first down-regulating HSC genes, then switching from lympho-myeloid priming to lymphoid commitment between the IL7R progenitor
(yellow balls) to proB (orange balls) compartments (Figure 5.10 left). The IL7R progenitor samples from the more primitive CS17 FL and D10 hPSC retained a relatively myeloid signature, consistent with the single cell qPCR data (Figure A1.3 and 3.8). Despite the large relative enrichment for HSC and myeloid genes in the ETV6-RUNX1 IL7R progenitor seen in figure 5.9, in absolute terms these cells clustered closely with the control hPSC progenitors. The most striking difference, however, was that ETV6-RUNX1 expressing ‘pro-B’ cells clustered with the IL7R progenitors from early CS17 FL and D10/D31 of hPSC co-culture, demonstrating a marked defect in the extent of their B lymphoid program.

Figure 5.10: RNASeq data from LIN- CD34+CD45RA- (green), IL7R+ progenitor (yellow) and proB cells (orange) derived from FL (blue), adult BM (red), hPSC (green) and ETV6-RUNX1 hIPSCs (purple) plotted on a 3D scatter plot, where gene sets for myeloid (log), lymphoid (log) or HSC (linear) programs are plotted on the x, y and z-axis respectively. Left: points coloured for differentiation hierarchy. Right: same points coloured for comparison between primary, hPSC and ETV6-RUNX1 hIPSC samples. Kindly produced by Dr Javier Herrero.
Together these analyses show that the transcriptional program of the immunophenotypically-defined ETV6-RUNX1 expressing ‘proB’ cells is in fact very different from primary and hPSC-control counterparts and instead more closely resembles the global and lineage-affiliated transcriptional programs of the more primitive lympho-myeloid IL7R progenitor.

5.2.3. Comparison to RUNX1 floxed control hIPSCs

To directly attribute the cellular and molecular phenotypes seen in the ETV6-RUNX1 knock-in cells to ETV6-RUNX1 expression, I analysed the differentiation outputs from the two lines (2.8_D5 and 2.8_D6) that had the RUNX1 cassette removed using TAT-CRE. Both reverted clones differentiated to proB and preB cells at levels not seen from any of the knock-in clones (Figure 5.11).

Figure 5.11: ETV6-RUNX1 was floxed using cell permeable TAT-cre. Flow cytometry analysis at D31 of culture shows re-emergence of Venus negative proB and preB cells in the
reverted clone (D6). Percentage of total CD45\(^+\) Venus\(^+\) and CD45\(^+\) Venus\(^-\) cells respectively, clone #2.8 and 2 corresponding reverted clones derived from #2.8 (D5 and D6 n=2), mean percentage values of the reverted clone is shown.

To test whether the reverted clone’s proB and preB cells exhibited a normal pattern of lineage priming I sorted single cells for qPCR analysis (Figure 5.12). The gene expression signature of the proB cells was restored almost precisely to that seen in unmodified MIFF3, with a marked reduction in the abnormal co-expression of B and myeloid programmes (Figure 5.13).

![ProB reverted clone](image)

**Figure 5.12: Single cell qPCR analysis of proB cells (CD34\(^+\)CD19\(^+\)) from ETV6-RUNX1 reverted clone. Each column represents one single cell. Color code according to CT values. Genes labeled in red are myeloid, blue lymphoid and green B cell genes. 34 cells investigated in 1 experiment.**
Figure 5.13: Co-expression of lineage associated genes in proB cells based on single cell qPCR data from MIFF3, ETV6-RUNX1 knock in and reverted hPSCs. A cell is considered myeloid (GM) if expressing 2 of the myeloid genes, lymphoid (L) if 2 of the genes listed in blue and B primed (B) if expressing 2 of the genes listed in green. Co-expression of GM with L or B genes are labeled GM/L and GM/B respectively. Data from proB D31 are shown (MIFF3, ETV6-RUNX1 IPS and ETV6-RUNX1 IPS floxed clone D6).

Overall these results indicate that both the functional B cell differentiation block and the abnormal B-myeloid programming seen in the ETV6-RUNX1 knock-in cultures is directly attributable to ETV6-RUNX1 expression, rather than any effect of hiPS variability or the PGK-NeoR selection cassette.

5.3. Discussion

Previous studies of ETV6-RUNX1 pre-leukaemia suggested the level of ETV6-RUNX1 expression affects its functional impact (Tsuzuki and Seto, 2013). Consequently I used a targeted knock-in approach placing ETV6-RUNX1 under the control of the endogenous ETV6 promoter in hIPSCs. As in mouse models, early expression of ETV6-RUNX1 during hPSC lymphoid differentiation appears non-problematic (Schindler et al., 2009; van der Weyden et al., 2011). Consistent with several previous ETV6-RUNX1 models the predominant phenotype observed was a partial block in B cell development (Morrow et al., 2004; Tsuzuki
and Seto, 2013; Tsuzuki et al., 2004), but here seen at the level of the IL7R progenitor.

Whether or not the IL7R progenitor is indeed the in utero target cell for ETV6-RUNX1, the partial nature of the differentiation block allowed molecular analysis of the crucial downstream proB cells affording examination of the phenotypic compartments at the same stage of differentiation seen in bulk ALL and pre-leukaemic cord血液. These ETV6-RUNX1 proB cells exhibited markedly abnormal gene expression characterised by a limited B cell signature with co-expression of HSC and myeloid genes. Such co-expression of conflicting signaling and transcriptional pathways may produce a network level lineage conflict, which has been hypothesised to be a major hallmark of first hit oncogenes (Banerji et al., 2013; Enver and Greaves, 1998). Once trapped, such cells may have ongoing IL7-STAT5 signaling and differential responses to niche signals or cytokine environments produced in response to infection, a proposed accelerator of leukemic progression (Ford et al., 2009; Greaves, 2006).

Whether ETV6-RUNX1 is potent enough to re-program a committed B cell or is acting to extend the B/myeloid signature of the upstream IL7R progenitor into proB cells is uncertain and would require testing using a conditional knock-in model. In adult, IPS reprogramming and myeloid trans-differentiation of committed B cells have demonstrated the power of PAX5 to both maintain B lineage identity and extinguish myeloid programming (Hanna et al., 2008; Xie et al., 2004), yet here ETV6-RUNX1 facilitates multilineage priming despite co-expression of PAX5. The increasing evidence for lineage plasticity in fetal haematopoietic tissues (Notta et al., 2016) and the dual B-myeloid signature of the fetal IL7R progenitor outlined in chapter 4 indicate that fetal B cells may be more amenable to lineage reprogramming.

The IL7R progenitor exhibits a number of the predicted characteristics of the target cell for the pre-leukemic activity of ETV6-RUNX1. It is a relatively abundant B cell progenitor present
early in the first trimester, when changes in placental structure may increase fetal vulnerability to first hit mutations (Pijnenborg et al., 2006). Monochorionic twin studies showing shared clonal DJH or TCR rearrangement in single cells indicate that ETV6-RUNX1 arises in or impacts a lymphoid progenitor prior to B cell commitment (Alpar et al., 2015; Hong et al., 2008). Expression of ETV6-RUNX1 impairs the myeloid to lympho-myeloid transition seen in the IL7R progenitor during development, partially arresting B cell differentiation. The partial nature of the B cell differentiation block may explain the requirement for recurrent second hit mutations targeting other lymphoid master regulators such as PAX5. Furthermore RAG recombinase appears to mediate the majority of second hit mutations in ETV6-RUNX1 ALL (Papaemmanuil et al., 2014) and both the arrested IL7R progenitor and myeloid expressing ‘proB’ cells might therefore be exposed to ongoing RAG and AID exposure if their normal maturation is delayed (Swaminathan et al., 2015).

Overall, the presence of such a developmentally restricted B-myeloid population offers an attractive explanation as to why ETV6-RUNX1 ALL is overwhelmingly a disease of childhood and why ETV6-RUNX1 B-ALL frequently co-expresses B and myeloid surface markers (Abdelhaleem, 2007; Gerr et al., 2010). The transient progenitor-like nature of this population, coupled with the relatively weak oncogenic impact of ETV6-RUNX1, may explain the low rate of progression to ALL and the relatively good response of ETV6-RUNX1 ALL to chemotherapy.

Although I was unable to demonstrate enhancement of self-renewal capacity by NSG engraftment, this negative result is most likely consistent with the well-documented difficulties in engrafting hPSC-derived haematopoietic material into NSG mice (Slukvin, 2013). Whether demonstration of enhanced self-renewal can be achieved by changes in transplantation protocol, such as use of larger cell numbers, in utero transplantation, or the use of humanised mice more permissive to myeloid engraftment remains to be tested.
Consistent with the single cell transcriptional analysis, preliminary evidence indicates that ETV6-RUNX1 hPSC-derived proB cells are able to produce macrophages when sorted into liquid culture with myeloid cytokines (clone 2.8, n=1 – data not shown). However, this result has not yet been replicated or rigorously tested against multiple MIFF3 control or reverted hiPSC-derived material. If demonstrated to be both replicable and ETV6-RUNX1 specific this capacity for myeloid transdifferentiation may provide an attractive avenue for myeloid differentiation therapy in ETV6-RUNX1 ALL or pre-leukaemia.

Together these results suggest the hiPS ETV6-RUNX1 knock-in model provides a tractable and developmentally relevant setting for the examination of how ETV6-RUNX1 initiates pre-leukaemia in utero. The model suggests that ETV6-RUNX1 primarily acts to preclude the effective resolution of a myeloid-lymphoid cell fate choice that exists during ontogeny when lymphoid programming develops in myeloid primed progenitors. Whether this cellular hierarchy also offers a venue for the origin of non ETV6-RUNX1 associated ALL is an exciting, but as yet untested hypothesis. Further analysis of this hPSC-based model may identify fetal specific vulnerabilities in ETV6-RUNX1 ALL, informing new and more specific therapeutic targets which could be tested within this tractable platform. In addition, the model will afford systematic introduction and evaluation of the secondary mutations prevalent in ETV6-RUNX1 ALL. Finally the concept that developmentally restricted target cells are the venues for the initiation of childhood diseases (Li et al., 2005) suggests that hPSCs could be more broadly used to model childhood malignancies.
CHAPTER 6

Final Remarks and Future Directions
Together these results indicate that in vitro B cell differentiation from hPSCs recapitulates fetal B lymphopoiesis and that ETV6-RUNX1 blocks fetal B cell differentiation at the level of an IL7R⁺ lympho-myeloid progenitor compartment, resulting in the formation of proB cells aberrantly expressing myeloid antigens. This molecular and functional phenotype in a tractable and scalable model system now provides a platform for the systematic identification of targetable pathways and the screening of therapeutic compounds.

The current analysis of the RNASeq data has focused on comparing primary and hPSC samples at the global transcriptional level, and further investigating the lineage affiliated gene sets implicated by the single cell qPCR analysis of ETV6-RUNX1 expressing cells. GSEA analysis should identify other candidate oncogenic pathways upregulated in the ETV6-RUNX1 proB cells. In addition, more detailed analysis of the RNASeq data from primary samples should characterise the components of the fetal specific gene expression signature. Our hypothesis would indicate that this fetal programme is required for the effect of ETV6-RUNX1, but dispensable for normal ‘adult’ haematopoiesis, making it an attractive drug target.

PCA analysis of the single cell qPCR data indicates that ETV6-RUNX1 proB cells are a relatively heterogeneous population spanning from the IL7R compartment to proB cells (Figure 5.7). Once established in the laboratory, single cell RNASeq analysis will afford the unbiased transcriptional interrogation of these cells, potentially demonstrating sub-compartments during this transition. Similarly low cell number ChIP (e.g. iChIP) against the V5 epitope tag will allow elucidation of direct ETV6-RUNX1 binding sites (Lara-Astiaso et al., 2014). Together these data can be combined with the systems level analysis of ETV6-RUNX1
model systems and primary ALL samples already established in the lab. This analysis should dissect the contributions from the fetal gene expression programme and ETV6-RUNX1 core programme, and define what elements of those components persist into ETV6-RUNX1 ALL.

The primary functional effect of ETV6-RUNX1 is a partial block in B cell differentiation at the IL7R progenitor to proB cell transition. On occasion, however, small numbers of more mature preB cells have been observed. Analysis of these preB cells will be informative, as i) this is the phenotypic compartment most commonly seen in ETV6-RUNX1 cALL and therefore may be a key cell of impact, or alternatively ii) if B cells maturing past the IL7R-proB cell block are molecularly and functionally ‘normal’ in ETV6-RUNX1 pre-leukaemia then this would indicate that B lineage differentiation therapy as a useful therapeutic strategy, at least for the pre-leukaemic phase of the disease. To this end, the recently published extended B cell differentiation protocol should be useful in increasing the numbers of preB cells available for analysis and allowing analysis of differentiation to the IgM expressing naïve B cell stage, which is more mature than that seen in ETV6-RUNX1 ALL (French et al., 2015). Similarly, although the myeloid potential of ETV6-RUNX1 expressing IL7R progenitors or proB cells has not been fully investigated, preliminary data indicates that ETV6-RUNX1 expressing proB cells can be readily differentiated to macrophages (Figure 6.1). Of note ETV6-RUNX1 is not associated with myeloid leukaemias, and previous models of ETV6-RUNX1 have not demonstrated a block in myeloid differentiation. Therefore, if ETV6-RUNX1 expressing cells retain myeloid potential this may be another avenue to explore differentiation therapies.
Figure 6.1: MGG stained cytospin of macrophage produced from ETV6-RUNX1 expressing proB cells grown in liquid culture with myeloid cytokines (SCF, Flt3L, TPO, IL3, MCSF, GCSF) (n=1).

Preliminary attempts to demonstrate the self-renewal potential of control or ETV6-RUNX1 derived CD34+ cells by engraftment into newborn livers of NSG mice failed. Engraftment may be achievable by modifying the transplantation protocol, cell number or strain of mouse used, but given the established difficulties associated with engrafting hPSC-derived haematopoietic cells and the weak oncogenic impact of ETV6-RUNX this is anticipated to be very challenging. In the absence of engraftment, serial re-plating assays such as LT-CIC or extended B cell differentiation may provide an indication of the self-renewal capacity of ETV6-RUNX1 expressing cellular compartments.

My results indicate that the B-myeloid signature seen in ETV6-RUNX1 expressing proB cells arises as a result of the unique co-expression of B and myeloid programmes seen in the fetal IL7R progenitor, providing a plausible explanation as to why ETV6-RUNX1 is a childhood disease. A major caveat to this, however, is the absence of an equivalent control comparison of ETV6-RUNX1 in adult IL7R progenitors. In the absence of such an adult model, the fetal contribution to the effect of ETV6-RUNX1 could be further validated by comparison with the immunophenotypes or transcriptional profiles of ETV6-RUNX1 ALL samples, or ideally from
FACS sorted fractions of cord bloods containing rare ETV6-RUNX1 pre-leukaemic cells, should suitable sorting markers become available.

The constitutive ETV6-RUNX1 model created here demonstrates the earliest point in the haematopoietic hierarchy that ETV6-RUNX1 can impact and implicates the IL7R progenitor as a candidate target cell for ETV6-RUNX1. The requirement for the contribution of the B-myeloid transcriptional context afforded by the IL7R progenitor could be tested in hPSCs using a conditional knock-in approach, such as a cre-FLEX switch system (Schnutgen et al., 2003). By inducing ETV6-RUNX1 knock-in after B lineage commitment, this model would demonstrate the latest point in differentiation that ETV6-RUNX1 is able to act, and assess whether the B-myeloid signature seen in the ETV6-RUNX1 proB cells here is dependent on the programming seen in the IL7R progenitor.

Lastly, this model now affords the systematic evaluation of the contribution of second hit mutations in the transformation of ETV6-RUNX1 pre-leukaemia. These second hits are likely to result in stronger phenotypes, but may themselves impede haematopoietic differentiation from hPSCs. An inducible knockout system could be generated by stably introducing a Tet-inducible Cas9 on a by piggyBac transposon (kindly provided by Konstantinos Anastassiadis) into ETV6-RUNX1 hPSCs. gRNAs specific to single or multiple second hit target genes could then be introduced into the ETV6-RUNX1 hPSCs and Cas9 mutagenesis induced at different stages during differentiation either on bulk cultures or FACS sorted cellular compartments.

In summary, this hPSC model of ETV6-RUNX1 pre-leukaemia has demonstrated the lineage conflicting effect of ETV6-RUNX1 during fetal lymphopoiesis. Further analysis of this system will identify an ETV6-RUNX1 core transcriptional network and its interaction with fetal
specific signatures. The model provides a platform for discovering and validating novel therapeutic targets in ETV6-RUNX1 pre-leukaemia and ALL.
APPENDIX A

Analysis of B cell development in

First trimester Human Fetal Liver
APPENDIX A: Analysis of B cell development in first trimester Human Fetal Liver

Unpublished work by Charlotta Boiers in the Enver lab undertaken parallel my own work has investigated first trimester human fetal livers (FL) looking for early B cell progenitors that might be candidate target cells for pre-leukaemic initiation by ETV6-RUNX1. This work has demonstrated that the first emerging B cells in the human FL arise at Carnegie stage 17 (Figure 1.4) and express high levels of IL7R compared to neonatal cord blood (CB) or adult bone marrow (BM) (60-88% FL total CD19⁺ B cells vs. 9.8% BM total CD19⁻) (Figure A1.1).

Figure A1.1: CD19⁺ B cells from human fetal livers (FL) at CS17 and CS20, cord blood (CB) and adult bone marrow (BM) were analysed for surface expression of IL7R. Viable cells were gated CD45⁺LIN⁻, further gating as indicated. Top panel; mean percentage CD19⁺ B cells of total CD45⁺ cells. Bottom panel; mean percentage of CD19⁺ B cells expressing IL7R. Courtesy of Charlotta Boiers.

Gene expression analysis by bulk and single cell RT-qPCR using a panel of lineage affiliated
genes showed that the FL proB cells exhibited only minor differences to adult proB cells for the genes tested. Notable differences other than high levels of IL7R included high-level expression of LIN28B and low expression of DNNT in the FL, consistent with published literature (Li et al., 1993).

Given that IL7R expression was highly prominent on emerging fetal B cells we hypothesised that these cells arise from a lineage negative IL7R⁺ progenitor. Using CD19 as a marker for B lineage commitment in the adult led to the identification of a CD19 negative lineage negative CD34⁺CD45RA⁻CD19 IL7R⁺ progenitor compartment (referred to as IL7R progenitor). Lineage positive antigens included CD19, CD3, CD8a, CD56, CD235a. The progenitor was also positive for Kit (CD117, the receptor for stem cell factor) and the activation marker CD38. This progenitor was relatively abundant prior to the emergence of FL B cells (Figure A1.2) and was demonstrated to have B, T, NK and myelomonocytic potential when assayed by in vitro culture, but has lost MkE potential (data not shown). Furthermore, the IL7R progenitor was shown to have DJ₄ rearrangements by gDNA PCR (data not shown).

Figure: A1.2: Kinetics of IL7R progenitor emergence during development as percentage of total CD34⁺CD45⁺ cells. Each dot represents one biological sample. Mean±SD. Statistical comparison versus CB control. Courtesy of Charlotta Boiers.
Consistent with these functional outputs, single cell RT-qPCR gene expression analysis of the IL7R progenitor showed enrichment for both myeloid and lymphoid genes including RAG2, but little or no expression of genes affiliated with the MkE lineage (Figure A1.3). Comparing IL7R progenitors at the time of the first emerging B cells (CS17) with those approximately a week later (CS20) revealed a significant switch in gene expression pattern. CS17 IL7R progenitors expressed predominantly myeloid genes with the majority of cells lacking an identifiable or coherent lymphoid program. By CS20 early lymphoid and B cell specific gene expression programs were expressed whilst retaining the myeloid signature, suggesting that the IL7R progenitors had transitioned from a myeloid primed to a lympho-myeloid primed transcriptional state. Qualitatively the lymphoid program appeared different in adult BM compared to CS20 FL, with higher expression of a core lymphoid program, but less consistent expression of B lineage restricted genes.

![Figure A1.3](image)

*Figure A1.3: Single cell qPCR analysis of IL7R⁺KIT⁺ progenitors from different developmental stages; FL CS17 (left), FL CS20 (middle) and adult BM (right). Each column represents a single cell, coloured by CT value. Gene sets: red, myeloid; blue, lymphoid; green, B lineage related. 44-60 cells and n=3 per developmental stage. Courtesy of Charlotta Boiers.*
Using the single cell RT-qPCR data we asked to what extent lymphoid and myeloid programs were co-resident in individual IL7R progenitor cells through development. Cells were considered myeloid primed if expressing two or more of CD33, IL3RA, MPO, CSF1R, CSF2RA and CSF3R; lymphoid primed if expressing 2 or more of CD7, CD10 and RAG2; or B committed if expressing 2 or more of EBF1, PAX5 and CD19. This analysis revealed a transition from a purely myeloid progenitor state at CS17 to a lympho-myeloid state in CB and adult BM, transiting through a highly unusual B-myeloid (GM/B) state at CS20 not present at other (CB and adult BM) developmental stages (Figure A1.4). In summary, as a RAG-expressing B cell progenitor with a transient developmental lympho-myeloid program the fetal liver IL7R progenitor has many of the attributes anticipated of the ETV6-RUNX1 in utero target cell.

Figure A1.4: Co-expression of lineage-associated genes in IL7R⁺KIT⁺ progenitors based on single cell qPCR. Cells are considered myeloid (GM), lymphoid (L) or B-primed (B) if expressing 2 genes representative of that lineage, as defined in figure 3A. Co-expression of 2 or more of each of the GM and L and or B genes are labeled GM/L and GM/B respectively. Only cells expressing IL7R were analysed. Courtesy of Charlotta Boiers.
REFERENCES


de Bruijn, M.F., Speck, N.A., Peeters, M.C., and Dzierzak, E. (2000). Definitive hematopoietic stem cells first develop within the major arterial regions of the mouse embryo. The EMBO journal 19, 2465-2474.


Ma, X., Edmonson, M., Yergeau, D., Muzny, D.M., Hampton, O.A., Rusch, M., Song, G.,
to relapse in pediatric B-acute lymphoblastic leukaemia. Nat Commun 6, 6604.

Maddalo, D., Manchado, E., Concepcion, C.P., Bonetti, C., Vidigal, J.A., Han, Y.C.,
of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. Nature 516, 423-
427.


Protracted postnatal natural histories in childhood leukemia. Genes Chromosomes Cancer
39, 335-340.


Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church,

Malina, A., Mills, J.R., Cencic, R., Yan, Y., Fraser, J., Schippers, L.M., Paquet, M., Dostie, J., and
development 27, 2602-2614.


Maximow, A. (1909). Der Lymphozyt als gemeinsame Stammzelle der ver- schiedenen Blutelemente in der embryonalen Entwicklung und im postfetalen Leben der Sa¨ ugetiere. Folia Haematol (Frankf) 8, 125-134.


Wang, J.C., Doedens, M., and Dick, J.E. (1997a). Primitive human hematopoietic cells are enriched in cord blood compared with adult bone marrow or mobilized peripheral blood as measured by the quantitative in vivo SCID-repopulating cell assay. Blood 89, 3919-3924.


