Identifying Transcription Factor 
Dysregulation in RUNX1-ETO Leukaemia 

A thesis submitted to Cardiff University in candidature for 
the degree of Doctor of Philosophy 

Aleksandra Mazurek Azevedo 

Division of Cancer and Genetics, 
School of Medicine 
Cardiff University 

February 2022
Abstract

Acute myeloid leukaemia (AML) results from clonal expansion of primitive myeloid cells incapable of terminal differentiation, giving rise to an accumulation of ‘blast’ cells at various stages of maturation within the bone marrow niche. AML is a heterogeneous disease with multiple morphological, immunophenotypic and genetic features. This includes the t(8;21) which results in the expression of RUNX1-ETO, and occurs in 12% of AML cases.

To understand the role of RUNX1-ETO in the pathogenesis of AML, our group previously ectopically expressed RUNX1-ETO in normal human haematopoietic stem progenitor cells (HSPC). This resulted in a block in granulocytic differentiation and was associated with increased self-renewal - hallmarks of leukaemia. A subsequent study analysed the transcriptome of these cells and identified 380 differentially expressed genes using an unsupervised approach. This current study has now refined this analysis to determine the most significant changing transcription factors (TFs). Using Pathway Analysis programme (Metacore™), this study identified ZNF217 to be significantly overexpressed compared to control (1.5-fold; p=0.003). ZNF217 is a TF responsible for binding to the promoters of several target genes, such as E-cadherin, as well as cooperating in transcriptional silencing programs by recruiting chromatin modifiers. This study determined that ZNF217 overexpression, as a single abnormality, induced myeloid differentiation of HSPC, particularly within the monocytic population, suggesting that it is unlikely that this TF possesses a role in leukemogenesis on its own. Additionally, ZNF217 was found to be dispensable for myeloid differentiation, as knockdown (KD) of this TF failed to inhibit this process.

Whilst studies have determined the transcriptomic changes observed in cells expressing RUNX1-ETO, there is a paucity of studies quantitating proteomic changes. Therefore, this study also aimed at analysing the proteomic profile of RUNX1-ETO expressing HSPC using quantitative proteomics by SWATH-MS (on different subcellular structures including cytosolic or nuclear fractions). 4,635 proteins were quantified, of which 2,787 were detected in the cytoplasm, and 1,848 in the nucleus. Statistical analysis identified 257 significantly differentially expressed proteins in RUNX1-ETO compared to controls; of which 71% were detected in cytoplasm and 29% in the nucleus. RUNX1-ETO significantly downregulated the expression of C/EBPβ protein and mRNA vs control suggesting transcriptional suppression by RUNX1-ETO. Knocking-down C/EBPβ expression in HSPC, however, failed to induce significant changes in both monocytic and granulocytic development. Interestingly, KD of C/EBPβ in the RUNX1-ETO-expressing cell line, SKNO-1, completely suppressed myeloid cell surface marker expression and gave a concomitant increase in cell proliferation. In non-t(8;21) cells lines (HEL and U937), on the other hand, KD of C/EBPβ ablated cell growth and increased apoptotic frequency, suggesting that the effects of C/EBPβ KD are context dependent.

In conclusion, both transcriptomic and proteomic analysis proved to be useful tools for the identification of potential mediators of the block in terminal differentiation observed in RUNX1-ETO-expressing cells. Subsequently, ZNF217 and C/EBPβ were identified as targets of interest in the context of t(8;21). Whilst it is unlikely that ZNF217 overexpression contributes to leukaemogenic development, additional studies would be necessary to fully determine the role of C/EBPβ in this process.
Acknowledgements

I would like to express my deepest gratitude to my supervisors Prof. Alex Tonks and Prof. Richard Darley for their continuous support and guidance throughout my PhD. Thanks to your encouraging words, your patience, and your knowledge, you have allowed me to grow as a scientist over the last three years, and provided me with invaluable skills for the future. I would also like to thank Dr. Steven Knapper, my mentor, for his advice and for providing me insights into the clinical aspects of AML. I would also like to extend my appreciation to Prof. Tony Whetton, Dr. Andrew Pierce, and Dr. Bethany Geary (University of Manchester), for their contribution to the project, and help and advice on mass spectrometry techniques and analysis.

Within the Haematology Department, I would like to particularly thank Amanda Gilkes for all her hard work during these unprecedented times, which allowed me to safely return to the laboratory to conclude my studies. A special thank you to Dr. (!) Rachael Nicholson, for always being there for me, to help and to listen, for both the good and the bad days that inevitably come with the PhD. I would also like to thank Sara Davies for her training and support during the beginning stages of my PhD. Additionally, I would like to thank Dr. Amanda Tonks, director of PGR, for not only helping me in securing funding for the last months of my PhD, but also for your continuous advice and support throughout this time. Thank you to Dr. Ann Kift-Morgan and Dr. Catherine Naseriyan from the Central Biotechnology Service at Cardiff University for your support and guidance.

I would particularly like to thank my family. Words cannot express how grateful I am for having you in my life, whether you’re near or far. I am endlessly grateful for your constant support and words of encouragement, for advising me when stress got in the way and for believing in me, always. Adam, I want to especially thank you for everything you’ve done for me, for your incredible patience, for your constant support, and for never leaving my side. I couldn’t have done this without you! During my PhD studies, I was lucky enough to have found a Portuguese family in the middle of Cardiff. To Ana, Sara and Tatiana, I would specially like to thank you for your friendship and support during these years, and for constantly encouraging me to keep going. Thank you for always being there, never complaining about hearing my numerous problems.

Lastly, I would like to thank School of Medicine in Cardiff university for providing the funding for a scholarship to undertake this research.

Thank you! Obrigado! Dziękuję!
Publications and Presentations

Publications


Oral Presentations

2018-2021 - Haematology Department Seminar Series, Cardiff University, UK

2019 - Division of Cancer and Genetics PGR Symposium, Cardiff University, UK

2021 - Schools of Medicine and Dentistry PGR symposium, Cardiff University, UK

Poster Presentations

2021 – European Hematology Association; Virtual
# Table of Contents

Abstract  ................................................................. i

Acknowledgements .................................................. ii

Publications and Presentations ................................ iii

Table of Contents ....................................................... iv

Summary of Figures .................................................. xii

Summary of Tables .................................................... xix

List of Abbreviations ................................................ xx

## Chapter 1 ........................................................................... 1

1.1 Haematopoiesis ......................................................... 2

1.1.1 Overview ................................................................. 2

1.1.2 Haematopoietic stem cells ........................................ 3

1.1.2.1 Properties of haematopoietic stem cells ...................... 5

1.1.2.2 Haematopoietic stem cell niche .............................. 5

1.1.3 Hierarchical organisation of haematopoiesis .................. 9

1.1.4 Regulation of normal human haematopoiesis .................. 11

1.1.4.1 Transcription factor regulation of haematopoietic development .............. 11

1.1.4.2 Cytokines and growth factors .............................. 14

1.2 Acute Myeloid Leukaemia .......................................... 16

1.2.1 Overview ................................................................. 16

1.2.2 Pathophysiology of AML ........................................... 16

1.2.2.1 Leukaemic Stem Cells .......................................... 19

1.2.3 Classification of AML ................................................. 21

1.2.4 Epidemiology, diagnosis, and prognosis of AML ............. 24

1.2.5 Treatment of AML .................................................. 26

1.2.5.1 Targeted therapies ............................................ 27

1.3 Pathophysiology of RUNX1-ETO ................................ 29

1.3.1 RUNX1 ................................................................. 30
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.3.1.1</td>
<td>The RUNX family of transcription factors</td>
<td>30</td>
</tr>
<tr>
<td>1.3.1.2</td>
<td>RUNX1 structure and function</td>
<td>30</td>
</tr>
<tr>
<td>1.3.1.3</td>
<td>Transcriptional regulation by RUNX1</td>
<td>33</td>
</tr>
<tr>
<td>1.3.1.4</td>
<td>Post-translational modifications</td>
<td>35</td>
</tr>
<tr>
<td>1.3.1.5</td>
<td>The role of RUNX1 in normal haematopoiesis</td>
<td>36</td>
</tr>
<tr>
<td>1.3.1.6</td>
<td>The role of RUNX1 in AML</td>
<td>37</td>
</tr>
<tr>
<td>1.3.2</td>
<td>ETO</td>
<td>38</td>
</tr>
<tr>
<td>1.3.3</td>
<td>RUNX1-ETO</td>
<td>41</td>
</tr>
<tr>
<td>1.3.3.1</td>
<td>RUNX1-ETO structure</td>
<td>41</td>
</tr>
<tr>
<td>1.3.3.2</td>
<td>Cooperation between RUNX1 and RUNX1-ETO</td>
<td>43</td>
</tr>
<tr>
<td>1.3.3.3</td>
<td>Identification of RUNX1-ETO-mediated transcriptional de-regulation</td>
<td>46</td>
</tr>
<tr>
<td>1.3.3.4</td>
<td>Molecular mechanisms of RUNX1-ETO-induced leukaemia</td>
<td>48</td>
</tr>
<tr>
<td>1.4</td>
<td>ZNF217</td>
<td>50</td>
</tr>
<tr>
<td>1.4.1</td>
<td>ZNF217 in haematopoietic development</td>
<td>53</td>
</tr>
<tr>
<td>1.4.2</td>
<td>ZNF217 as an oncogene</td>
<td>54</td>
</tr>
<tr>
<td>1.4.2.1</td>
<td>Proliferation</td>
<td>54</td>
</tr>
<tr>
<td>1.4.2.2</td>
<td>Replicative immortality</td>
<td>55</td>
</tr>
<tr>
<td>1.4.2.3</td>
<td>Drug resistance</td>
<td>56</td>
</tr>
<tr>
<td>1.4.2.4</td>
<td>Cell differentiation</td>
<td>56</td>
</tr>
<tr>
<td>1.5</td>
<td>C/EBPβ</td>
<td>57</td>
</tr>
<tr>
<td>1.5.1</td>
<td>Structure and function</td>
<td>59</td>
</tr>
<tr>
<td>1.5.2</td>
<td>C/EBPβ in haematopoietic development</td>
<td>62</td>
</tr>
<tr>
<td>1.5.3</td>
<td>C/EBPβ as an oncogene</td>
<td>64</td>
</tr>
<tr>
<td>1.5.3.1</td>
<td>Cell survival</td>
<td>64</td>
</tr>
<tr>
<td>1.5.3.2</td>
<td>Tumour aggressiveness</td>
<td>65</td>
</tr>
<tr>
<td>1.5.3.3</td>
<td>Drug resistance</td>
<td>66</td>
</tr>
<tr>
<td>1.5.4</td>
<td>C/EBPβ in leukaemogenesis</td>
<td>66</td>
</tr>
<tr>
<td>1.6</td>
<td>Aims and objectives</td>
<td>67</td>
</tr>
</tbody>
</table>
Chapter 2 ........................................................................................................................................71

2.1 Molecular Biology ......................................................................................................................72
  2.1.1 Plasmids used in the study ....................................................................................................72
  2.1.2 Bacterial growth medium and selective agar plates .............................................................72
  2.1.3 Isolation and quantification of recombinant plasmid DNA from bacteria ........76
  2.1.4 DNA Sequencing ..................................................................................................................76

2.2 Cell culture ..................................................................................................................................77
  2.2.1 Culture of cell lines ................................................................................................................77
    2.2.1.1 Cell culture of non-adherent cell lines ...........................................................................77
    2.2.1.2 Cell culture of adherent cell lines .................................................................................77
  2.2.2 Determination of cell density .................................................................................................80
  2.2.3 Determination of cell viability ................................................................................................81
  2.2.4 Cryopreservation and thawing of cell lines ........................................................................... 81

2.3 Primary cell culture .....................................................................................................................82
  2.3.1 Isolation of mononuclear cells from human neonatal cord blood ...........................................82
  2.3.2 Isolation of CD34+ HSPC from mononuclear cells ...............................................................82
  2.3.3 Culture of CD34+ HSPC .......................................................................................................83
  2.3.4 Colony forming efficiency .....................................................................................................85
  2.3.5 Assessment of cell morphology .............................................................................................85

2.4 Virus transfection .......................................................................................................................85
  2.4.1 Transfection using Lipofection .............................................................................................87
  2.4.2 Transfection using calcium-phosphate precipitation .............................................................88
  2.4.3 Virus titration .........................................................................................................................88

2.5 Transduction of cell lines and human CD34+ HSPC .................................................................89
  2.5.1 Transduction of cell lines .....................................................................................................89
  2.5.2 Transduction of human CD34+ HSPCs ...............................................................................89

2.6 Determination of protein expression by western blot ...............................................................90
  2.6.1 Total protein extraction .......................................................................................................90
2.6.2 Fractionated protein extraction ................................................................. 90
2.6.3 Protein quantification using Bradford assay .............................................. 91
2.6.4 Protein electrophoresis and electroblotting ............................................. 92
2.6.5 Immunoblotting and protein detection ..................................................... 93
2.6.6 Membrane and gel protein staining ......................................................... 93
2.6.7 Densitometry analysis .............................................................................. 95

2.7 Flow cytometry ......................................................................................... 95

2.7.1 Equipment and Analysis ........................................................................... 95
2.7.2 Estimation of CD34⁺ cell count ................................................................ 95
2.7.3 Fluorescence Activated Cell Sorting ....................................................... 98
2.7.4 Immunophenotypic analysis of transduced CD34⁺ HSPCs ....................... 98
2.7.5 Cell cycle analysis .................................................................................. 102
2.7.6 Apoptosis ................................................................................................ 102
2.7.7 Growth and proliferation assay .............................................................. 105

2.8 Proteomic analysis using SWATH® ............................................................ 105

2.8.1 Sample preparation .................................................................................. 105
2.8.2 Tryptic digestion .................................................................................... 105
2.8.3 Strong Cation Exchange Clean-up ......................................................... 106
2.8.4 Relative quantification by SWATH acquisition ....................................... 106
2.8.5 SWATH-MS statistical analysis .............................................................. 107

2.9 Analysis of existing mRNA expression data ............................................. 107

2.9.1 Analysis of microarray data of RUNX1-ETO induced changes in mRNA expression in CD34⁺ HSPC ................................................................. 107
2.9.2 Publicly available databases ................................................................... 108
  2.9.2.1 Bloodspot .......................................................................................... 108
  2.9.2.2 cBioPortal ....................................................................................... 108

2.10 Meta-analysis of mRNA and proteomic data ......................................... 108

2.10.1 Pathway analysis using Metacore ........................................................ 109
2.10.2 Interactome analysis using Metacore .......................................................... 111
2.10.3 Pathway analysis and identification of transcription factors using Ingenuity Pathway Analysis 112

2.11 Statistical analysis ........................................................................................................ 112

Chapter 3 .............................................................................................................................. 113
3.1 Introduction ....................................................................................................................... 114
3.2 Hypothesis and Aims ........................................................................................................ 116
3.3 Results ............................................................................................................................... 119
  3.3.1 Transcriptional dysregulation mediated by RUNX1-ETO ........................................ 119
    3.3.1.1 RUNX1-ETO expression in HSPC alters TF gene expression and is associated with GO processes that regulate transcription .............................................. 119

  3.3.1.2 Identification of aberrantly expressed TFs driving transcriptional change in AML 127

3.3.2 Quantitative SWATH-MS analysis reveals proteomic changes arising from RUNX1-ETO expression, in CD34+ HSPC ................................................................. 136
    3.3.2.1 Generation of human CD34+ HSPC expressing RUNX1-ETO for SWATH-MS analysis 136

    3.3.2.2 Assessment of CD34+ HSPC nuclear and cytosolic extracts shows sample purity and integrity .................................................................................................................. 136

    3.3.2.3 SWATH-MS analysis detected over 4000 proteins in CD34+ HSPC ...................... 137

    3.3.2.4 Exploratory data analysis ...................................................................................... 142

    3.3.2.5 Statistical model used for data analysis ............................................................... 143

    3.3.2.6 Expression of RUNX1-ETO disrupts normal protein expression in CD34+ cord blood derived HSPC ........................................................................................................ 147

    3.3.2.7 Identification of aberrantly expressed TF in CD34+ HSPC expressing RUNX1-ETO 150

    3.3.2.8 RUNX1-ETO leads to the aberrant expression of 53 proteins ............................ 153

    3.3.2.9 RUNX1-ETO leads to a shift in protein sub-cellular localisation in CD34+ HSPC 157

3.4 Discussion ......................................................................................................................... 162
3.4.1 Transcriptional dysregulation imposed by RUNX1-ETO expression ..........162
3.4.2 Dysregulation of the proteome arising from RUNX1-ETO expression ..........167
3.4.2.1 Alternative analysis of RUNX1-ETO-induced proteomic changes in CD34+ HSPC 171

Chapter 4 ..............................................................................................................174

4.1 Introduction .....................................................................................................175
4.2 Hypothesis and Aims.......................................................................................176
4.3 Results .............................................................................................................177
  4.3.1 Expression of ZNF217 mRNA increases throughout myeloid cell development. 177
  4.3.2 ZNF217 is variably expressed across AML subtypes ................................ 177
  4.3.3 Overexpression of ZNF217 promotes human myeloid development ............184
    4.3.3.1 Generation of ZNF217 overexpressing CD34+ HSPC ...............................184
    4.3.3.2 ZNF217 overexpression supresses myeloid colony formation and self-renewal 186
    4.3.3.3 ZNF217 overexpression supresses the growth of myeloid cells ...............186
    4.3.3.4 ZNF217 overexpression promotes myeloid differentiation .....................190
    4.3.3.5 ZNF217 overexpression in HSPC induces changes in cell cycle ...............194
  4.3.4 Knockdown of ZNF217 does not impact myeloid differentiation .................199
    4.3.4.1 Selection and validation of ZNF217 shRNA constructs .........................199
    4.3.4.2 Knockdown of ZNF217 leads to a decrease in CD34+ HSPC colony forming ability 199
    4.3.4.3 Knockdown of ZNF217 disrupts myeloid cell growth and development ......203
    4.3.4.4 Knockdown of ZNF217 does not affect cell cycle and apoptosis ............206
  4.3.5 The role of ZNF217 in AML cell lines .......................................................210
    4.3.5.1 ZNF217 expression is variable across AML cell lines ............................210
    4.3.5.2 Generation of ZNF217-knockdown AML cell lines ...............................210
    4.3.5.3 ZNF217 knockdown in AML cell lines reduces cell growth by promoting changes in cell cycle and apoptosis .................................................................214
  4.4 Discussion ..................................................................................................217
4.4.1  High ZNF217 in AML is associated with increased risk of relapse .......................... 218
4.4.2  ZNF217 overexpression of promotes myeloid development in HSPC ..................... 219
4.4.3  ZNF217 knockdown in HSPC does not impact myeloid development ..................... 222

Chapter 5 .................................................................................................................. 223

5.1  Introduction ........................................................................................................ 224
5.2  Hypothesis and Aims ........................................................................................ 226
5.3  Results ................................................................................................................ 227

5.3.1  Expression of CEBPB mRNA increases throughout myeloid cell development .. 227
5.3.2  CEBPB is variably expressed across AML subtypes ......................................... 228
5.3.3  Knockdown of C/EBPβ influences normal myeloid development ................. 237

5.3.3.1  Generation of C/EBPβ-knockdown CD34+ HSPC ............................................ 237
5.3.3.2  Knockdown of C/EBPβ leads to a decrease in HSPC colony forming ability .. 237
5.3.3.3  Knockdown of C/EBPβ supresses myeloid cell growth ................................. 240
5.3.3.4  Knockdown of C/EBPβ influences myeloid differentiation ......................... 240
5.3.3.5  Knockdown of C/EBPβ does not result in significant cell cycle or apoptotic changes 247

5.3.4  Overexpression of C/EBPβ influences normal myeloid development .......... 247

5.3.4.1  Generation of C/EBPβ-overexpressing HSPC ............................................... 247
5.3.4.2  C/EBPβ overexpression promotes myeloid colony formation and self-renewal ability 250
5.3.4.3  C/EBPβ promotes myeloid cell growth ....................................................... 250
5.3.4.4  C/EBPβ expression promotes myeloid differentiation .................................... 253
5.3.4.5  Overexpression of C/EBPβ induces change in the cell cycle of HSPC .......... 255

5.3.5  The effect of C/EBPβ mis-regulation in AML cell lines .................................. 260

5.3.5.1  Expression of C/EBPβ is variable across AML cell lines .............................. 260
5.3.5.2  Generation of C/EBPβ-knockdown cell lines ............................................. 260
5.3.5.3  The impact of C/EBPβ knockdown on the AML t(8;21) cell line SKNO-1 ... 265
5.3.5.4  The impact of C/EBPβ knockdown on AML non-t(8;21) cell lines .......... 267
5.4 Discussion.................................................................................................................. 277

5.4.1 CEBPB expression increases during haematopoietic development................. 278
5.4.2 Increased CEBPB expression is associated with poor prognosis in AML ....... 279
5.4.3 Knockdown of C/EBPβ in HSPC does not influence normal myeloid development 279
5.4.4 C/EBPβ overexpression in HSPC promotes myeloid cell growth and development 281
5.4.5 The consequences of C/EBPβ KD in AML cell lines is context-dependant ...... 283

Chapter 6 .......................................................................................................................... 285

6.1 Conclusions............................................................................................................... 286

6.1.1 RUNX1-ETO-induced transcriptomic and proteomic changes in HSPC......... 287
6.1.2 ZNF217 does not contribute to the pathogenesis of AML ......................... 288
6.1.3 KD of C/EBPβ does not contribute to the pathogenesis of AML............. 289
6.1.4 Limitations of the study.................................................................................... 291
6.1.5 Future directions............................................................................................... 293

Supplementary Tables.................................................................................................... 297

Supplementary Figures............................................................................................... 306

References....................................................................................................................... 366
Summary of Figures

Figure 1.1 – Hierarchical representation of human haematopoietic development ........................................... 4
Figure 1.2 – HSC regulation by the BM microenvironment ................................................................. 7
Figure 1.3 – Transcriptional regulation of myeloid development ......................................................... 12
Figure 1.4 – Structural representation of RUNX1 isoforms genes and proteins ........................................... 32
Figure 1.5 – Transcriptional regulation mediated by RUNX1 ........................................................................ 34
Figure 1.6 – Structural representation of the ETO gene and protein ...................................................... 40
Figure 1.7 – RUNX1-ETO structure and isoforms .................................................................................. 42
Figure 1.8 – Mechanisms of transcriptional dysregulation induced by RUNX1-ETO ............................. 45
Figure 1.9 – Structural representation of the ZNF217 gene and protein ............................................. 51
Figure 1.10 – ZNF217 mediated transcriptional repression ................................................................. 52
Figure 1.11 – Structural representation of the C/EBPβ gene and protein isoforms ............................... 60
Figure 1.12 – Hypothesis of the study ................................................................................................... 70
Figure 2.1 – Plasmid DNA constructs used in this study ............................................................................. 75
Figure 2.2 – Morphological analysis of HSPC ....................................................................................... 86
Figure 2.3 – Estimation of CD34+ HSPC in cord blood by flow cytometry .............................................. 97
Figure 2.4 – Gating strategy used to sort GFP-expressing CD34+ HSPC .............................................. 99
Figure 2.5 – Immunophenotypic analysis of CD34+ HSPC ................................................................. 101
Figure 2.6 – Gating strategy used to perform cell cycle analysis ............................................................ 103
Figure 2.7 – Gating strategy used to analyse apoptosis .......................................................................... 104
Figure 2.8 - Strategy for the analysis of differentially expressed genes/protein to study the effects of RUNX1-ETO overexpression in human CD34+ HSPC ......................................................... 110
Figure 3.1 – Strategy for the identification of changes in gene/protein expression arising from the expression of the fusion protein RUNX1-ETO ................................................................. 118
Figure 3.2 – Functional Enrichment Analysis of mRNA changes observed in human CD34+ HSPC expressing RUNX1-ETO ........................................................................................................ 120
Figure 3.3 – TF were identified as the most significantly represented functional class due to RUNX1-ETO expression in CD34+ HSPC .................................................................................................... 122
Figure 3.4 – RUNX1-ETO does not influence the expression of IRF7 in CD34+ HSPC ....................... 130
Figure 3.5 – RUNX1-ETO leads to the upregulation of IRF9 in CD34+ HSPC ........................................... 131
Figure 3.6 – RUNX1-ETO leads to the upregulation of ARID5B in CD34+ HSPC ................................. 132
Figure 3.7 – RUNX1-ETO leads to the upregulation of ZNF217 in CD34+ HSPC ............................... 133
Figure 3.8 – RUNX1-ETO leads to the upregulation of protein targets, in CD34+ HSPC........134
Figure 3.9 – Viral infection of CD34+ cells. CD34+ HSPC cells were analysed for GFP expression on day 3 of cell culture ...................................................................................................................138
Figure 3.10 – Quality control and characterisation of CD34+ HSPC overexpressing RUNX1-ETO ........................................................................................................................................139
Figure 3.11 – GAPDH and Histone H1 protein quantitation by SWATH-MS supports acceptable subcellular fraction efficiencies for cytosol and nuclear protein isolation ..................140
Figure 3.12 – RUNX1 is significantly overexpressed in RUNX1-ETO transfected CD34+ HSPC ........................................................................................................................................141
Figure 3.13 – Principal Component Analysis (PCA) of the protein expression data ........144
Figure 3.14 – Hierarchical clustering of proteins expressed in the cytoplasmic and nuclear compartment of CD34+ HSPC ...................................................................................................................145
Figure 3.15 – Differential protein expression analysis between CD34+ HSPC expressing RUNX1-ETO and control........................................................................................................................................146
Figure 3.16 – Functional Enrichment Analysis of protein changes observed in normal haematopoietic development vs. RUNX1-ETO expressing CD34+ HSPC.................................148
Figure 3.17 – Classification of differentially expressed proteins due to the expression of RUNX1-ETO in CD34+ HSPC ...................................................................................................................149
Figure 3.18 – C/EBPβ expression profile in CD34+ HSPC expressing RUNX1-ETO........154
Figure 3.19 – Target protein and mRNA expression profiles of TF found exclusively in RUNX1-ETO CD34+ HSPC ...................................................................................................................156
Figure 3.20 – Proteins are mis-localised to the nucleus or cytoplasm as a consequence of RUNX1-ETO expression, in CD34+ HSPC ...................................................................................................................159
Figure 3.21 – RUNX1-ETO leads to a shift in localisation of 9 proteins.............................160
Figure 4.1 – ZNF217 mRNA expression in normal haematopoiesis ................................178
Figure 4.2 – ZNF217 mRNA expression levels are variable across AML subtypes ..........179
Figure 4.3 – High levels of ZNF217 is associated with poorer disease-free survival of AML patients ........................................................................................................................................181
Figure 4.4 – High ZNF217 expression is associated with decreased BM blasts and lower mutation count ........................................................................................................................................182
Figure 4.5 – Dysregulated ZNF217 expression is not associated with patient characteristics......183
Figure 4.6 – Transduction of ZNF217 in HEK 293T cells and human HSPC..........................185
Figure 4.7 – Expression and enrichment of GFP+ populations in human CD34+ HSPC ........187
Figure 4.8 – ZNF217 overexpression inhibits myeloid colony formation and decreases self-renewal capacity ................................................................. 188
Figure 4.9 – ZNF217 overexpression inhibits monocytic and granulocytic growth during myeloid development .................................................................................. 189
Figure 4.10 – ZNF217 overexpression disrupts the balance between granulocytic and erythroid populations during myeloid cell development ........................................... 191
Figure 4.11 – ZNF217 overexpression promotes monocytic differentiation ........................................ 193
Figure 4.12 – ZNF217 overexpression promotes granulocytic differentiation by upregulating monocytic markers ......................................................................................... 195
Figure 4.13 – ZNF217 overexpression alters cell morphology during myeloid development ...... 196
Figure 4.14 – ZNF217 overexpression in CD34+ HSPC significantly disrupts normal cell cycle. 197
Figure 4.15 – ZNF217 shRNA constructs successfully reduced the levels of endogenous ZNF217 in K562 cells ......................................................................................... 200
Figure 4.16 – Infection efficiency of transduced CD34+ HSPC using shRNA lentiviral constructs ................................................................................................................... 201
Figure 4.17 – ZNF217 knockdown disrupts myeloid colony formation of HSPC ......................... 202
Figure 4.18 – Knockdown of ZNF217 promotes monocytic cell growth in myeloid development .................................................................................................................. 204
Figure 4.19 – ZNF217 knockdown disrupts the balance between the monocytic and granulocytic populations in CD34+ HSPC during myeloid cell development.......................... 205
Figure 4.20 – Knockdown of ZNF217 does not significantly alter monocytic development ...... 207
Figure 4.21 – Knockdown of ZNF217 does not significantly alter granulocytic development..... 208
Figure 4.22 – Knockdown of ZNF217 has no effect on normal cell cycle or apoptosis ............ 209
Figure 4.23 – ZNF217 expression in leukaemic cell lines .............................................................. 211
Figure 4.24 – ZNF217 is knocked-down in the erythroleukaemia cell line HEL ..................... 212
Figure 4.25 – ZNF217 is knocked-down in the t(8;21) cell line SKNO-1 ................................. 213
Figure 4.26 – ZNF217 knockdown significantly decreases HEL cell proliferation ................ 215
Figure 4.27 – ZNF217 knockdown significantly decreases SKNO-1 cell proliferation ............. 216
Figure 5.1 – CEBPB mRNA expression in normal haematopoiesis ............................................. 229
Figure 5.2 – C/EBPβ endogenous expression throughout haematopoiesis ................................. 230
Figure 5.3 – CEBPB mRNA expression levels are variable across AML subtypes .................. 232
Figure 5.4 – High levels of CEBPB is associated with poorer disease-free survival of AML patients .................................................................................................................. 233
Figure 5.5 – High CEBPB expression is associated with increased WBC and lower mutation count .................................................................................................................................................................................. 234
Figure 5.6 – Dysregulated CEBPB expression is associated with patient characteristics ........... 235
Figure 5.7 – CEBPB expression correlates with cytogenetic and molecular abnormalities ...... 236
Figure 5.8 – C/EBPβ is knocked-down in human CD34+ HSPC .......................................................... 238
Figure 5.9 – C/EBPβ KD disrupts myeloid colony formation of HSPC .................................................. 239
Figure 5.11 – C/EBPβ KD disrupts the balance between the monocytic and granulocytic populations in CD34+ HSPC during myeloid cell development .................................................................................. 242
Figure 5.12 – KD of C/EBPβ has little impact on normal monocytic differentiation .................. 243
Figure 5.13 – KD of C/EBPβ promotes granulocytic development ...................................................... 245
Figure 5.14 – C/EBPβ KD affects cell morphology during myeloid development ......................... 246
Figure 5.15 – KD of C/EBPβ has no significant effect on normal cell cycle or apoptosis .......... 248
Figure 5.16 – C/EBPβ overexpression in human HSPC .............................................................................. 249
Figure 5.17 – C/EBPβ overexpression promotes myeloid colony formation of HSPC .............. 251
Figure 5.18 – Overexpression of C/EBPβ promotes monocytic and granulocytic growth in myeloid development .......................................................................................................................... 252
Figure 5.19 – C/EBPβ overexpression disrupts the balance between myeloid population during haematopoietic development .................................................................................. 254
Figure 5.20 – C/EBPβ overexpression promotes the upregulation of monocytic markers in monocytic progenitors .......................................................................................................................... 256
Figure 5.21 – C/EBPβ overexpression promotes the upregulation of granulocytic markers in granulocytic progenitors .................................................................................................................. 257
Figure 5.22 – C/EBPβ overexpression alters cell morphology during myeloid development ...... 258
Figure 5.23 – Overexpression of C/EBPβ influences normal cell cycle ............................................. 259
Figure 5.24 – C/EBPβ expression in leukaemic cell lines ....................................................................... 261
Figure 5.25 – C/EBPβ KD in the t(8;21) cell line SKNO-1 ................................................................. 262
Figure 5.26 – C/EBPβ KD in the erythroleukaemia cell line HEL ......................................................... 263
Figure 5.27 – C/EBPβ KD in the leukaemic cell line U937 ................................................................. 264
Figure 5.28 – C/EBPβ KD promotes cell growth in SKNO-1 cells, associated with changes in normal cell cycle and apoptotic rate ........................................................................................................... 266
Figure 5.29 – The effect of C/EBPβ KD on cell surface marker expression in SKNO-1 cells .... 268
Figure 5.30 – Morphological effect of C/EBPβ KD in SKNO-1 cells ................................................. 269
Figure 5.31 – C/EBPβ KD significantly reduces cell growth in HEL cells, associated with changes in apoptotic frequency ................................................................. 271
Figure 5.32 – The effect of C/EBPβ KD on cell surface marker expression in HEL cells........ 272
Figure 5.33 – Morphological effect of C/EBPβ KD in HEL cells........................................ 273
Figure 5.34 – C/EBPβ KD significantly reduces cell growth in U937 cells, associated with changes in normal cell cycle and apoptotic rate ................................................................. 274
Figure 5.35 – The effect of C/EBPβ KD on cell surface marker expression in U937 cells...... 275
Figure 5.36 – Morphological effect of C/EBPβ KD in U937 cells........................................ 276
Figure 6.1 – Summary of the main findings described on Chapters 3 and 4 ......................... 295
Figure 6.2 – Summary of the main findings described on Chapter 5 ................................ 296

Supplementary Figure 1 - Representative gating strategy used to immunophenotype HSPC 306
Supplementary Figure 2 – Expression of control GFP and ZNF217 in human HSPC .......... 307
Supplementary Figure 3 – Expression of the stem cell marker CD34 in monocytes during differentiation, following ZNF217 overexpression ................................................................. 308
Supplementary Figure 4 – Expression of the myeloid marker CD11b in monocytes during differentiation, following ZNF217 overexpression ................................................................. 309
Supplementary Figure 5 – Expression of the monocytic marker CD14 in monocytes during differentiation, following ZNF217 overexpression ................................................................. 310
Supplementary Figure 6 – Expression of the stem cell marker CD34 in granulocytes during differentiation, following ZNF217 overexpression ................................................................. 311
Supplementary Figure 7 – Expression of the myeloid marker CD11b in granulocytes during differentiation, following ZNF217 overexpression ................................................................. 312
Supplementary Figure 8 – Expression of the granulocytic marker CD15 in granulocytes during differentiation, following ZNF217 overexpression ................................................................. 313
Supplementary Figure 9 – Expression of control GFP and shZNF217 KD in human HSPC ...... 315
Supplementary Figure 10 – Expression of the stem cell marker CD34 in monocytes during differentiation, following ZNF217 KD ................................................................. 317
Supplementary Figure 11 – Expression of the myeloid marker CD11b in monocytes during differentiation, following ZNF217 KD ................................................................. 319
Supplementary Figure 12 – Expression of the monocytic marker CD14 in monocytes during differentiation, following ZNF217 KD ................................................................. 321
Supplementary Figure 13 – Expression of the stem cell marker CD34 in granulocytes during differentiation, following ZNF217 KD ................................................................. 323
Supplementary Figure 14 – Expression of the myeloid marker CD11b in granulocytes during differentiation, following ZNF217 KD ................................................................. 325
Supplementary Figure 15 – Expression of the granulocytic marker CD15 in granulocytes during differentiation, following ZNF217 KD ................................................................. 327
Supplementary Figure 16 – Expression of control GFP and shC/EBPβ KD in human HSPC ...... 329
Supplementary Figure 17 – Expression of the stem cell marker CD34 in monocytes during differentiation, following C/EBPβ KD ........................................................................... 331
Supplementary Figure 18 – Expression of the myeloid marker CD11b in monocytes during differentiation, following C/EBPβ KD ........................................................................... 333
Supplementary Figure 18 – Expression of the monocytic marker CD14 in monocytes during differentiation, following C/EBPβ KD ........................................................................... 335
Supplementary Figure 20 – Expression of the stem cell marker CD34 in granulocytes during differentiation, following C/EBPβ KD ........................................................................... 337
Supplementary Figure 21 – Expression of the myeloid marker CD11b in granulocytes during differentiation, following C/EBPβ KD ........................................................................... 339
Supplementary Figure 22 – Expression of the granulocytic marker CD15 in granulocytes during differentiation, following C/EBPβ KD ........................................................................... 341
Supplementary Figure 23 – Expression of control GFP and C/EBPβ in human HSPC ............ 342
Supplementary Figure 24 – Expression of the stem cell marker CD34 in monocytes during differentiation, following C/EBPβ overexpression ....................................................................... 343
Supplementary Figure 25 – Expression of the myeloid marker CD11b in monocytes during differentiation, following C/EBPβ overexpression ....................................................................... 344
Supplementary Figure 26 – Expression of the myeloid marker CD14 in monocytes during differentiation, following C/EBPβ overexpression ....................................................................... 345
Supplementary Figure 27 – Expression of the stem cell marker CD34 in granulocytes during differentiation, following C/EBPβ overexpression ....................................................................... 346
Supplementary Figure 28 – Expression of the myeloid marker CD11b in granulocytes during differentiation, following C/EBPβ overexpression ....................................................................... 347
Supplementary Figure 29 – Expression of the granulocytic marker CD15 in granulocytes during differentiation, following C/EBPβ overexpression ....................................................................... 348
Supplementary Figure 30 – ZNF217 overexpression inhibits monocytic and granulocytic growth during myeloid development ........................................................................... 349
Supplementary Figure 31 – ZNF217 overexpression disrupts the balance between granulocytic and erythroid populations during myeloid cell development ................................................................. 350

Supplementary Figure 32 – ZNF217 overexpression promotes monocytic differentiation......... 351

Supplementary Figure 33 – ZNF217 overexpression promotes granulocytic differentiation by upregulating monocytic markers ........................................................................................................ 352

Supplementary Figure 34 – Knockdown of ZNF217 promotes monocytic cell growth in myeloid development ...................................................................................................................................... 353

Supplementary Figure 35 – ZNF217 knockdown disrupts the balance between the monocytic and granulocytic populations in CD34+ HSPC during myeloid cell development .............................................. 354

Supplementary Figure 36 – Knockdown of ZNF217 does not significantly alter monocytic development .................................................................................................................................. 355

Supplementary Figure 37 – Knockdown of ZNF217 does not significantly alter granulocytic development .................................................................................................................................. 356

Supplementary Figure 38 – KD of C/EBPβ decreases granulocytic growth in myeloid development ........................................................................................................................................ 357

Supplementary Figure 39 – C/EBPβ KD disrupts the balance between the monocytic and granulocytic populations in CD34+ HSPC during myeloid cell development .............................................. 358

Supplementary Figure 40 – KD of C/EBPβ has little impact on normal monocytic differentiation ........................................................................................................................................ 359

Supplementary Figure 41 – KD of C/EBPβ promotes granulocytic development .................... 360

Supplementary Figure 42 – Overexpression of C/EBPβ promotes monocytic and granulocytic growth in myeloid development ........................................................................................................ 361

Supplementary Figure 43 – C/EBPβ overexpression disrupts the balance between myeloid population during haematopoietic development ...................................................................................... 362

Supplementary Figure 44 – C/EBPβ overexpression promotes the upregulation of monocytic markers in monocytic progenitors ........................................................................................................ 363

Supplementary Figure 45 – C/EBPβ overexpression promotes the upregulation of granulocytic markers in granulocytic progenitors .................................................................................................... 364

Supplementary Figure 46 – C/EBPβ KD in the Kasumi-1 cell line........................................................................................................................................ 365
Summary of Tables

Table 1.1 – Recurrent mutations in AML ................................................................. 18
Table 1.2 – FAB classification of AML ................................................................. 22
Table 1.3 – WHO (2016) classification of AML .............................................. 23
Table 1.4 – ELN 2017 risk stratification of AML ............................................. 25
Table 2.1 – Plasmid vectors used in this study ..................................................... 73
Table 2.2 – Primers used for direct sequencing ................................................... 77
Table 2.3 – Culture conditions for non-adherent cells ....................................... 78
Table 2.4 – Culture conditions for adherent cells .............................................. 80
Table 2.5 – Cell culture media used for the growth and development of CD34+ HSPC .......... 84
Table 2.6 – List of primary antibodies used in this study .................................... 94
Table 2.7 – Technical specifications of the Accuri™ C6 Plus used in this study .......... 96
Table 2.8 – List of antibodies used for cord blood immunophenotyping ............... 96
Table 2.9 – Antibodies used for immunophenotypic analysis of transduced CD34+ HSPC ...... 100
Table 3.1 – Identification of RUNX1-ETO induced TF dysregulation in CD34+ HSPC .......... 123
Table 3.2 – Summary showing the effects of RUNX1-ETO expression in CD34+ HSPC on gene and protein expression in prioritised targets ........................................ 135
Table 3.3 – Identification of RUNX1-ETO induced TF protein dysregulation in CD34+ HSPC .. 151
Table 3.4 – TF detected exclusively in RUNX1-ETO expressing CD34+ HSPC .......... 155
Table 3.5 – Identification of ‘shifted’ proteins as a result of RUNX1-ETO expression ....... 155
Table 4.1 - Summary of the main findings regarding the role of ZNF217 in normal human haematopoiesis ......................................................................................... 221
Table 5.1 – Summary of the main findings regarding the role of C/EBPβ in normal human haematopoiesis ......................................................................................... 282

Supplementary Table 1 – Compensation matrix .................................................. 297
Supplementary Table 2 – Identification of significantly dysregulated proteins, as a result of RUNX1-ETO expression ........................................................................ 298
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,25(OH)₂D₃</td>
<td>1,25-Dihydroxyvitamin D3</td>
</tr>
<tr>
<td>ALL</td>
<td>Acute Lymphoblastic Leukaemia</td>
</tr>
<tr>
<td>AmBic</td>
<td>Ammonium Bicarbonate</td>
</tr>
<tr>
<td>αMEM</td>
<td>Minimum Essential Medium Eagle - Alpha Modification</td>
</tr>
<tr>
<td>AML</td>
<td>Acute Myeloid Leukaemia</td>
</tr>
<tr>
<td>ANG1</td>
<td>Angiopoietin 1</td>
</tr>
<tr>
<td>ANXA2</td>
<td>Annexin 2</td>
</tr>
<tr>
<td>APC</td>
<td>Allophycocyanin</td>
</tr>
<tr>
<td>APL</td>
<td>Acute Promyelocytic Leukaemia</td>
</tr>
<tr>
<td>ATO</td>
<td>Arsenic Trioxide</td>
</tr>
<tr>
<td>ATRA</td>
<td>All-Trans Retinoic Acid</td>
</tr>
<tr>
<td>BM</td>
<td>Bone Marrow</td>
</tr>
<tr>
<td>β-ME</td>
<td>β-Mercaptoethanol</td>
</tr>
<tr>
<td>BSA</td>
<td>Bovine Serum Albumin</td>
</tr>
<tr>
<td>bZIP</td>
<td>C-terminal Leucine-Zipper Dimerization Domain</td>
</tr>
<tr>
<td>C/EBP</td>
<td>CCAAT/Enhancer Binding Protein</td>
</tr>
<tr>
<td>CAR</td>
<td>CXCL12-abundant reticular cell</td>
</tr>
<tr>
<td>CB</td>
<td>Cord Blood</td>
</tr>
<tr>
<td>CBF</td>
<td>Core Binding Factor</td>
</tr>
<tr>
<td>CBW</td>
<td>Cord Blood Wash</td>
</tr>
<tr>
<td>CD</td>
<td>Cluster of Differentiation</td>
</tr>
<tr>
<td>CDK</td>
<td>Cyclin Dependent Kinase</td>
</tr>
<tr>
<td>CDS</td>
<td>Coding Sequence</td>
</tr>
<tr>
<td>CFU</td>
<td>Colony-Forming Unit</td>
</tr>
<tr>
<td>ChIP-Seq</td>
<td>Chromatin Immunoprecipitation Sequencing</td>
</tr>
<tr>
<td>CHOP</td>
<td>C/EBP-Homologous Protein</td>
</tr>
<tr>
<td>c-KIT</td>
<td>Tyrosine-Protein Kinase KIT</td>
</tr>
<tr>
<td>CLP</td>
<td>Common Lymphoid Progenitor Cell</td>
</tr>
<tr>
<td>CML</td>
<td>Chronic Myeloid Leukaemia</td>
</tr>
<tr>
<td>CMP</td>
<td>Common Myeloid Progenitor Cell</td>
</tr>
<tr>
<td>CR</td>
<td>Complete remission</td>
</tr>
<tr>
<td>CSC</td>
<td>Cancer Stem Cell</td>
</tr>
<tr>
<td>CSF</td>
<td>Colony Stimulating Factor</td>
</tr>
<tr>
<td>CtBP</td>
<td>C-Terminal Binding Protein</td>
</tr>
<tr>
<td>CXCL12</td>
<td>CXC-Chemokine ligand 12</td>
</tr>
<tr>
<td>DC</td>
<td>Dendritic Cell</td>
</tr>
<tr>
<td>DEG</td>
<td>Differentially Expressed Genes</td>
</tr>
<tr>
<td>DEP</td>
<td>Differentially Expressed Proteins</td>
</tr>
<tr>
<td>dH₂O</td>
<td>Deionised Water</td>
</tr>
<tr>
<td>DIA</td>
<td>Data-Independent Acquisition</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco’s Modified Eagle Medium</td>
</tr>
<tr>
<td>DMNT</td>
<td>DNA Methyltransferase</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>DMSO</td>
<td>Dimethyl Sulfoxide</td>
</tr>
<tr>
<td>EGR</td>
<td>Early Growth Response</td>
</tr>
<tr>
<td>ELN</td>
<td>European LeukemiaNet</td>
</tr>
<tr>
<td>EMP</td>
<td>Erythro-myeloid Progenitor Cell</td>
</tr>
<tr>
<td>EMT</td>
<td>Epithelial-Mesenchymal Transition</td>
</tr>
<tr>
<td>EPO</td>
<td>Erythropoietin</td>
</tr>
<tr>
<td>ERK</td>
<td>Extracellular Signal-Regulated Kinases</td>
</tr>
<tr>
<td>ETO</td>
<td>Eight-Twenty-One</td>
</tr>
<tr>
<td>ETS</td>
<td>E26 Transformation-Specific</td>
</tr>
<tr>
<td>FAB</td>
<td>French-American-British</td>
</tr>
<tr>
<td>FACS</td>
<td>Fluorescently Activated Cell Sorting</td>
</tr>
<tr>
<td>FBS</td>
<td>Foetal Bovine Serum</td>
</tr>
<tr>
<td>FC</td>
<td>Fold Change</td>
</tr>
<tr>
<td>FDA</td>
<td>US Food and Drug Administration</td>
</tr>
<tr>
<td>FDR</td>
<td>False Discovery Rate</td>
</tr>
<tr>
<td>FGF</td>
<td>Fibroblast Growth Factor</td>
</tr>
<tr>
<td>FITC</td>
<td>Fluorescein isothiocyanate</td>
</tr>
<tr>
<td>FLI1</td>
<td>Friend Leukaemia Integration 1</td>
</tr>
<tr>
<td>FLT3</td>
<td>Fms-like Tyrosine Kinase 3</td>
</tr>
<tr>
<td>FSC</td>
<td>Forward Scatter</td>
</tr>
<tr>
<td>GAPDH</td>
<td>Glyceraldehyde 3-Phosphate Dehydrogenase</td>
</tr>
<tr>
<td>GATA</td>
<td>Globin Transcription Factor</td>
</tr>
<tr>
<td>G-CSF</td>
<td>Granulocyte Colony Stimulating Factor</td>
</tr>
<tr>
<td>GFI1</td>
<td>Growth-Factor Independent 1</td>
</tr>
<tr>
<td>GFP</td>
<td>Green Fluorescent Protein</td>
</tr>
<tr>
<td>GlyA</td>
<td>Glycophorin A (CD235a)</td>
</tr>
<tr>
<td>GM-CSF</td>
<td>Granulocyte/Macrophage Colony Stimulating Factor</td>
</tr>
<tr>
<td>GMP</td>
<td>Granulocyte-Monocyte Progenitor Cell</td>
</tr>
<tr>
<td>GO</td>
<td>Gene Ontology</td>
</tr>
<tr>
<td>GSC</td>
<td>Glioma Stem Cell</td>
</tr>
<tr>
<td>HBSS</td>
<td>Hank’s Balanced Salt Solution</td>
</tr>
<tr>
<td>HDAC</td>
<td>Histone Deacetylase</td>
</tr>
<tr>
<td>HMA</td>
<td>Hypomethylating Agents</td>
</tr>
<tr>
<td>HMEC</td>
<td>Human Mammary Epithelial Cells</td>
</tr>
<tr>
<td>HR</td>
<td>Hazard Ratio</td>
</tr>
<tr>
<td>HRP</td>
<td>Horseradish Peroxidase</td>
</tr>
<tr>
<td>HSC</td>
<td>Haematopoietic Stem Cell</td>
</tr>
<tr>
<td>HSPC</td>
<td>Haematopoietic Stem and Progenitor Cell</td>
</tr>
<tr>
<td>IDH</td>
<td>Isocitrate Dehydrogenase (NADP(+))</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>IMDM</td>
<td>Iscove Modified Dulbecco’s Medium</td>
</tr>
<tr>
<td>Inv</td>
<td>Inversion</td>
</tr>
<tr>
<td>IOSE</td>
<td>Human Ovarian Surface Epithelial Cell Line</td>
</tr>
<tr>
<td>IPA</td>
<td>Ingenuity Pathway Analysis</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>IRF</td>
<td>Interferon-Regulatory Factor</td>
</tr>
<tr>
<td>IT-HSC</td>
<td>Intermediate-Term Haematopoietic Stem Cell</td>
</tr>
<tr>
<td>ITD</td>
<td>Internal Tandem Duplication</td>
</tr>
<tr>
<td>KD</td>
<td>Knockdown</td>
</tr>
<tr>
<td>KO</td>
<td>Knockout</td>
</tr>
<tr>
<td>KRAS</td>
<td>Kirsten Rat Sarcoma Virus</td>
</tr>
<tr>
<td>LAP</td>
<td>Liver-Enriched Activating Protein</td>
</tr>
<tr>
<td>LAP*</td>
<td>Liver-Enriched Activating Protein*</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani</td>
</tr>
<tr>
<td>LDS</td>
<td>Lithium Dodecyl Sulphate Sample Buffer</td>
</tr>
<tr>
<td>LepR</td>
<td>Leptin Receptor Perivascular-Expressing Cell</td>
</tr>
<tr>
<td>Lin</td>
<td>Lineage</td>
</tr>
<tr>
<td>LIP</td>
<td>Liver-Enriched Inhibitory Protein</td>
</tr>
<tr>
<td>LMPP</td>
<td>Lymphoid-Primed Multipotent Progenitor Cell</td>
</tr>
<tr>
<td>LSC</td>
<td>Leukaemic Stem Cell</td>
</tr>
<tr>
<td>LSD</td>
<td>Lysine-Specific Histone Demethylase</td>
</tr>
<tr>
<td>LT-HSC</td>
<td>Long-Term Haematopoietic Stem Cell</td>
</tr>
<tr>
<td>MACS</td>
<td>Magnetic-Activated Cell Sorting</td>
</tr>
<tr>
<td>M-CSF</td>
<td>Macrophage Colony Stimulating Factor</td>
</tr>
<tr>
<td>MDP</td>
<td>Monocyte-Dendritic Progenitor Cell</td>
</tr>
<tr>
<td>MDS</td>
<td>Myelodysplastic syndrome</td>
</tr>
<tr>
<td>MEP</td>
<td>Megakaryocytic-Erythroid Progenitor Cell</td>
</tr>
<tr>
<td>MFI</td>
<td>Mean Fluorescent Intensity</td>
</tr>
<tr>
<td>MgCl2</td>
<td>Magnesium Chloride</td>
</tr>
<tr>
<td>miRNA</td>
<td>microRNAs</td>
</tr>
<tr>
<td>MLL</td>
<td>Mixed Lineage Leukaemia</td>
</tr>
<tr>
<td>MPP</td>
<td>Multipotent Progenitor Cell</td>
</tr>
<tr>
<td>MS</td>
<td>Mass Spectrometry</td>
</tr>
<tr>
<td>MSC</td>
<td>Mesenchymal Stem Cell</td>
</tr>
<tr>
<td>mSin3a</td>
<td>Mammalian Sin3a</td>
</tr>
<tr>
<td>N-CoR</td>
<td>Human Nuclear Receptor Co-Repressor</td>
</tr>
<tr>
<td>Nes</td>
<td>Nestin-positive Cell</td>
</tr>
<tr>
<td>NF-κB</td>
<td>Nuclear Factor-κB</td>
</tr>
<tr>
<td>NHR</td>
<td>Nervy Homology Region</td>
</tr>
<tr>
<td>NK</td>
<td>Natural Killer Cell</td>
</tr>
<tr>
<td>NLS</td>
<td>Nuclear Localisation Signal</td>
</tr>
<tr>
<td>NPM1</td>
<td>Nucleophosmin-1</td>
</tr>
<tr>
<td>NRAS</td>
<td>Neuroblastoma Rat Sarcoma Virus</td>
</tr>
<tr>
<td>OE</td>
<td>Overexpression</td>
</tr>
<tr>
<td>OIS</td>
<td>Oncogene-Induced Senescence</td>
</tr>
<tr>
<td>OPN</td>
<td>Osteopontin</td>
</tr>
<tr>
<td>OS</td>
<td>Overall Survival</td>
</tr>
<tr>
<td>P300</td>
<td>Histone Acetyltransferase p300</td>
</tr>
<tr>
<td>PB</td>
<td>Peripheral Blood</td>
</tr>
</tbody>
</table>
PBS
Phosphate Buffer Saline
PCA
Principal Component Analysis
PE
Phycoerythrin
PerCP-Cy5.5
Peridinin chlorophyll protein – Cyanine5.5
PI
Propidium iodide
PLZF
Promyelocytic Leukaemia Zinc-Finger
PRMT
Protein Arginine N-Methyltransferase
PVDF
Polyvinylidene Difluoride
RD
Regulatory Domain
RHD
Runt Homology Domain
RNASeq
RNA Sequencing
ROI
Region of Interest
RPMI
Roswell Park Memorial Institute-1640 medium
RUNX
Runt-Related Transcription Factor
SA-PerCP
Streptavidin-Peridinin Chlorophyll Protein
SB
Staining Buffer
SCF
Stem Cell Factor
SCL
Stem-Cell Leukaemia Factor
SCX
Strong Cation Exchange
shRNA
Short Hairpin RNA
SMRT
Silencing Mediator of Retinoic Acid and Thyroid Hormone Receptors
SSC
Side Scatter
ST-HSC
Short-Term Haematopoietic Stem Cell
STAT3
Signal Transducer and Activator of Transcription 3
SWATH-MS
Sequential Window Acquisition of all Theoretical Mass Spectra
T
Translocation
TAD
Transactivation Domain
TBS
Tris-Buffered Saline
TBS-T
Tris-Buffered Saline with Tween-20
TCGA
The Cancer Genome Atlas
TET2
Tet Methylcytosine Dioxygenase 2
TF
Transcription Factor
TGF-β
Transforming Growth Factor Beta
TKD
Tyrosine Kinase Domain Mutations
TP53
Tumour Protein p53
TPO
Thrombopoietin
TRCN
RNAi Consortium Number
TSP1
Thrombospondin-1
UC
Universal Container
WBC
White Blood Count
WHO
World Health Organization
WT1
Wilms Tumour 1
ZNF
Zinc Finger Protein
Chapter 1

Introduction
1.1 Haematopoiesis

1.1.1 Overview

Haematopoiesis is the process through which cellular constituents of the blood are continually replenished throughout an individual’s lifetime. It is composed of various highly specialised cells with numerous functions, including oxygen transport and immune defence (Jagannathan-Bogdan and Zon, 2013). This process is initiated by a common precursor, a haematopoietic stem cell (HSC), as found in the early 1900s. These cells are characterised by their ability to self-renew or give rise to different progenitor cells that can proliferate and differentiate into mature cells (Laurenti and Göttgens, 2018). Early studies showed that bone marrow (BM) failure in radiation exposed recipients could be rescued by injecting spleen or BM cells from non-exposed donors (Lorenz et al., 1951). Research on this topic intensified when Till and McCulloch demonstrated that the regenerative potential of HSCs could be assayed by performing in vivo repopulation assays, thus supporting the existence of multipotential HSCs (Till and McCulloch, 1961). These studies contributed to an understanding of a development hierarchy, in which multipotent HSCs are found at the top, whilst terminally differentiated cells sit on the bottom (1.1.3).

Throughout embryonic development, haematopoiesis occurs in successive waves, each one of them temporarily and spatially restricted, resulting in the development of haematopoietic progenitor cells. This process can be further divided into two main stages. In early embryonic development, blood cells are produced in the yolk sac, in a process termed ‘primitive’ haematopoiesis. During this time, there is an increased production of red blood cells, to promote tissue oxygenation as the embryo undergoes rapid growth. As well as primitive nucleated erythrocytes, there is also a low frequency of primitive macrophages and megakaryocytes (Fukuda, 1973; Luckett, 1978). The next stage of blood cell development is termed ‘definitive’ haematopoiesis, with the development of erythro-myeloid progenitors (EMP), which will give rise to definitive erythrocytes and most myeloid lineages, as well as early B and T progenitors. This process occurs initially in the aorta-gonad-mesonephros (AGM), later occurring in the placenta, foetal liver, spleen and BM (Ivanovs et al., 2017). Postnatally, haematopoiesis occurs primarily in the BM; however, in times of haematopoietic stress or injury, this process can occur in the liver or spleen (Butler et al., 2010).
1.1.2 Haematopoietic stem cells

HSCs possess two main characteristics that make them unique in the haematopoietic system. Firstly, these cells are multipotential and are therefore able to differentiate into all functional blood cells. Secondly, they have the ability to self-renew, giving rise to identical daughter cells without differentiating. Loss and/or gain of function studies in mice have identified basic developmental principles that control the emergence of haemogenic tissues during ontogeny, haematopoiesis in the adult. However, due to differences observed in basic biology and haematology, as well as therapy development, there was a constant need to complement these mouse studies with human primary cells. Subsequently, based on Till and McCulloch experiments, several groups focused on investigating human haematopoiesis using human colony-forming progenitors scored in in vitro colony-forming unit (CFU) assays (Moore et al., 1973; Pike and Robinson, 1970). Subsequently, the ability to engraft human haematopoietic cells into immune-deficient mice, allowed the development of several humanised mouse models to study HSC development in vivo (Bosma et al., 1983; McCune et al., 1988; Shultz et al., 2005; Shultz et al., 1995; Rongvaux et al., 2011).

A major obstacle in studying HSCs relies in the rarity of these cells. In general, only 1 in $10^6$ cells in the human BM is defined as a transplantable HSC (Wang et al., 1997) based on the simultaneous detection of several independent cell surface markers. The first marker to be identified in these cells was CD34, expressed in less than 5% of all blood cells, and in > 99% of human HSC (Civin et al., 1984). However, as CD34 is expressed both in HSCs, and in progenitor cells, there was a need to search for additional markers of stemness. Subsequent studies identified CD90 as a stem cell marker (Baum et al., 1992), whilst CD45RA and CD38 were identified as markers of progenitor cells (Bhatia et al., 1997; Conneally et al., 1997; Lansdorp et al., 1990). Furthermore, human HSC were characterised by the absence of lineage markers (Lin$^-$), resulting in their classification as Lin$^-\text{CD34}^+\text{CD38}^\text{-CD90}^\text{-CD45RA}^-$ (Figure 1.1).

At the cellular level, the inhibition of self-renewal occurs as lineage programming is initiated. Hence, it would be anticipated that this would also happen at the molecular level, resulting in the concept of multilineage priming, proposed as a mechanism through which HSCs are able to maintain their multipotency potential (Miyamoto et al., 2002).
Figure 1.1 – Hierarchical representation of human haematopoietic development

HSC possess an increased self-renewal capacity (curved arrow), gradually giving rise to mature progeny with reduced capacity for self-renewal and increased differentiation potential. The phenotypic cell surface marker of each population is shown. Adapted from (Weiskopf et al., 2016; Guilliams et al., 2018).

**HSC** – Haematopoietic Stem Cell; **MPP** – Multipotent Progenitor Cell; **CMP** – Common Myeloid Progenitor Cell; **MEP** – Megakaryocytic-Erythroid Progenitor Cell; **GMP** – Granulocyte-Monocyte Progenitor Cell; **MDP** – Monocyte-Dendritic cell Progenitor Cell; **CLP** – Common Lymphoid Progenitor Cell; **NK cell** – Natural Killer Cell; **GlyA** – Glycophorin A (CD235a); **CD** – Cluster of Differentiation; **Lin** – Lineage markers.
However, the HSC transcriptional programme is characterised by several unique metabolic and cellular properties, not intuitively linked with multipotency (Laurenti and Göttgens, 2018). In fact, approximately 70% of all transcriptional changes observed between HSCs and early progenitor cells occur independently of lineage fate (Laurenti et al., 2013). HSCs typically reside in a quiescent (Wilson et al., 2008; Foudi et al., 2009), autophagy-dependent (Warr et al., 2013; Ho et al., 2017) and glycolytic state (Simsek et al., 2010; Takubo et al., 2013), with low mitochondrial activity (Vannini et al., 2016; Ito et al., 2016) and tight regulation of protein synthesis (Signer et al., 2014). Progenitor cells, on the other hand, are highly proliferative and metabolically active cells (Laurenti and Göttgens, 2018). However, it is important to acknowledge that these characteristics are not absolute. Whilst the initiation of transcriptional programs specific to lineage determination may occur regardless of loss of stem cell characteristics, several regulators can be involved in both, like runt-related transcription factor 1 (RUNX1, aka AML1) (Cai et al., 2015). These observations suggest that the processes through which lineage determination coordinates with changes in the cellular state remains to be fully understood.

1.1.2.1 Properties of haematopoietic stem cells

It is currently recognised that the HSC population is heterogeneous, comprising, at least, two subsets, differing in their repopulation abilities and cycling properties (Foudi et al., 2009; Qiu et al., 2014; Takizawa et al., 2011). These are long-term HSCs (LT-HSC), capable of long-term engraftment, and short-term HSCs (ST-HSC), derived from LT-HSCs. Whilst LT-HSCs have a lifelong self-renewing potential (months-years), the ST-HSCs have limited ability (days-weeks), as they firstly give rise to multipotent progenitors (MPPs), further branching into common-myeloid progenitors (CMPs) and common-lymphoid progenitors (CLPs), with the aim of replenishing the haematopoietic system (Morrison and Weissman, 1994; Guenechea et al., 2001; Christensen and Weissman, 2001). Furthermore, phenotypically, LT-HSC are defined as Lin\(^{+}\)CD34\(^{+}\)CD45RA\(^{-}\)CD90\(^{-}\)CD49f\(^{-}\)CD38\(^{-}\), lacking CD38 or any lineage-restricted antigen. Upon loss of CD49f and CD90 expression, these give rise to transiently engrafting multipotent progenitors, the ST-HSCs, characterised as Lin\(^{+}\)CD34\(^{+}\)CD45RA\(^{-}\)CD49f\(^{-}\)CD90\(^{-}\)CD38\(^{-}\) cells (Notta et al., 2011).

1.1.2.2 Haematopoietic stem cell niche

The concept of an HSC niche was first suggested by Schofield in 1978, who proposed that a physical niche of cells resides within the BM (Schofield, 1978). HSCs largely rely on their
microenvironment, made up of a complex network of cells and secreted factors, for the regulation of quiescence, proliferation, self-renewal and differentiation (Morrison and Scadden, 2014). Normally, most of the HSCs are in a quiescent state, but can become active as a response to infectious stress, such as interferon-mediated signalling, and increase their proliferative rate or promote differentiation (Essers et al., 2009; Wilson et al., 2008; Baldridge et al., 2010). Currently, it is considered that the BM microenvironment is composed of two niches, able to maintain and regulate HSCs: the endosteal and the vascular niches (Figure 1.2).

In the outer BM, HSCs reside closely to the endosteal bone surface, as several studies showed that HSCs isolated from this region presented a higher degree of proliferation and an increased long-term hematopoietic reconstitution potential (Haylock et al., 2007; Grassinger et al., 2010). More differentiated cells, on the other hand, are generally found in the central BM region, within the perivascular niche.

The endosteum comprises the region between bone and BM, and it’s lined with a heterogeneous group of osteoblastic cells at various stages of differentiation, from which a fraction of are fully mature osteoblasts able to synthesise bone. Moreover, osteoclasts, bone-absorbing cells, also line the endosteum and together regulate bone formation. Osteoblasts synthesize several cytokines that are suggested to contribute to the maintenance and regulation of HSCs, such as such as CXC-chemokine ligand 12 (CXCL12 aka stromal-derived factor 1 [SDF-1]), stem-cell factor (SCF), osteopontin (OPN), granulocyte colony-stimulating factor (G-CSF), annexin 2 (ANXA2), angiopoietin 1 (ANG1), and thrombopoietin (TPO) (Taichman and Emerson, 1994; Ponomaryov et al., 2000; Calvi et al., 2003; Arai et al., 2004; Stier et al., 2005; Jung et al., 2007; Yoshihara et al., 2007).

CXCL12 is mainly produced by immature osteoblasts and endothelial cells, and controls HSC homing, retention, and repopulation (Ponomaryov et al., 2000). Even though SCF is mainly produced by perivascular cells (Ding et al., 2012), this cytokine is also produced by osteoblasts, and plays an important role in HSC maintenance. Equally, OPN, secreted by osteoblasts and other cells, plays a critical role in the retention, migration, and control of HSC proliferation and differentiation in the endosteal surface (Nilsson et al., 2005; Stier et al., 2005; Grassinger et al., 2009). G-CSF is necessary to support myelopoiesis (Lord et al., 1975; Morad et al., 2008), whilst ANXA2 regulates HSC homing and engraftment (Jung et al., 2007).
Figure 1.2 – HSC regulation by the BM microenvironment

HSC reside in a specialised environment within the BM, divided into two main niches: endosteal and perivascular. Adapted from (Galán-Diez et al., 2018).

**HSC** – Haematopoietic Stem Cell; **MEP** – Megakaryocytic-Erythroid Progenitor Cell; **GMP** – Granulocyte-Monocyte Progenitor Cell; **CAR cell** - CXCL12-abundant reticular cell.
In addition to secreting cytokines, osteoblastic cells have also been shown to express several membrane-bound ligands and adhesion receptors which contribute to HSC maintenance, including Jagged, a ligand for Notch receptors expressed in HSCs, and N-cadherin. However, both of these factors have proven to be controversial, with studies conflicting in regard to their specific role in HSC maintenance (Varnum-Finney et al., 2000; Butler et al., 2010; Maillard et al., 2008; Mancini et al., 2005; Li and Zon, 2010; Hosokawa et al., 2010). Several other factors are thought to play a role in HSC maintenance, including bone-degrading osteoclasts (Batard et al., 2000; Bhatia et al., 1999; Silver et al., 1988) and macrophages (Chang et al., 2008; Winkler et al., 2010).

In addition to the endosteal region, several studies have suggested that vascular environments are also involved in the maintenance of HSC, comprising the perivascular niche, essential for gas exchange, delivery of nutrients to cells and waste removal in the BM (Kiel et al., 2005) (Figure 1). The perivascular niche is highly heterogenous and contains distinct cell types, including endothelial cells, blood vessels that deliver both oxygen and nutrients, and can be found in the lining of sinusoids, specialised blood vessels that from an extensive network throughout the BM (Kiel et al., 2005). These cells regulate HSC fate both by direct cell-to-cell contact, as well as by the secretion of angiocrine factors (Butler et al., 2010; Kobayashi et al., 2010). These include CXCL12, vascular endothelial growth factor A (VEGF-A), fibroblast growth factor 2 (FGF2), ANG1, thrombospondin-1 (TSP1), and Notch ligands, which have all been described as playing important roles in maintaining the stem-cell pool and regulate HSC self-renewal capacity (Butler et al., 2010; Kobayashi et al., 2010; Poulos et al., 2013; Rafii et al., 2016). Other cell constituents present in the perivascular niche include Leptin Receptor Perivascular-Expressing Cells (LepR⁺), a major source of several growth factors, including CXCL12 and SCF (Ding et al., 2012; Ding and Morrison, 2013). Moreover, these cells are responsible for the differentiation of mesenchymal stem cells (MSCs) into adipocytes, osteoblasts, and chondrocytes (Zhou et al., 2014). Equally, Nestin-positive (Nes⁺) cells produce soluble factors involved in HSC maintenance, including CXCL12 and SCF, suggesting that MSCs are directly involved with HSCs (Méndez-Ferrer et al., 2010). CXCL12 abundant reticular (CAR) cells have been shown to be important in the homing and localisation of HSC within the BM. As the name suggests, these cells secrete extremely high levels of CXCL12, as well as SCF. Other molecules found in the perivascular niche include megakaryocytes and platelets (Banu and Williams, 1995; Alexander et al., 1996).
1.1.3 Hierarchical organisation of haematopoiesis

In order to define the relationship between HSCs and their progenies, a model for the haematopoietic hierarchy was initially proposed as a tree-like roadmap detailing the differentiation process starting from an HSC, into cells with declining multilineage potential leading to unilineage commitment (Kondo et al., 1997; Akashi et al., 2000; Adolfsson et al., 2005; Wilson et al., 2008) (Figure 1.1). Such distinction was made possible due to the use of phenotypical analysis of the different lineages, based on flow cytometry-based cell sorting (Rieger and Schroeder, 2012). LT-HSC sit at the apex of the hierarchical tree and, upon certain stimuli, can enter cell cycle (Schoedel et al., 2016), and progress into ST-HSC, which in turn differentiate into MPP. These cells present a higher frequency of cell-cycle progression and differentiation activity, but are incapable of self-renewing (Morrison et al., 1997). MPPs, in turn, give rise to CLPs, that only possess lymphoid-restricted differentiation abilities (Kondo et al., 1997), and CMPs, that can differentiate into megakaryocytic/erythroid progenitors (MEPs) and granulocyte/macrophage progenitors (GMPs) (Akashi et al., 2000; Na Nakorn et al., 2002). GMPs ultimately give rise to monocytes, macrophages and granulocytes, including basophils, neutrophils and eosinophils, whilst MEP cells differentiate into erythrocytes, megakaryocytes and platelets (Rieger and Schroeder, 2012). Cells derived from the lymphoid lineage are responsible for mediating immune response, whilst cells from the myeloid lineage are involved in several physiological processes, including blood clotting, oxygen transport and innate immunity.

Additional studies have shown that this classical model of haematopoiesis is more complex than initially thought. For instance, intermediate-term HSCs (IT-HSC) represent a transitory population between LT-SCS and ST-HSC (Yamamoto et al., 2013; Benveniste et al., 2010; Zhang et al., 2018). Furthermore, the MPP population can be further characterised into MPP1, MPP2, MPP3 and MPP4 (Wilson et al., 2008; Pietras et al., 2015). In the revised model, MPP1 resembles ST-HSC, whilst MPP2-4 are parallel subpopulations at a downstream level, responsible for sustaining blood production at steady-state (Pietras et al., 2015). In addition, another subpopulation has been identified, expressing high levels of Fms-like tyrosine kinase 3 (FLT3), defined as lymphoid-primed multipotent progenitors (LMPPs). Functionally, these generate cells within the lymphoid lineages (Adolfsson et al., 2001; Adolfsson et al., 2005;
Forsberg et al., 2006; Boyer et al., 2011), and phenotypically resemble MPP4 (Pietras et al., 2015).

The traditional hierarchical model suggests that HSC differentiation is a stepwise process, however, several studies have shown that megakaryocyte differentiation is able to bypass most of these steps, and its lineage outcome is mainly derived from HSCs (Woolthuis and Park, 2016). In mice, HSCs were found to highly express von Willebrand factor (Vwf), a megakaryocyte-specific gene (Månsson et al., 2007). These cells were shown to possess robust short- and long-term reconstitution ability for megakaryocytes post-transplantation, and their maintenance was modulated by TPO, essential for megakaryocyte development and platelet production (Sanjuan-Pla et al., 2013; de Sauvage et al., 1996). Subsequent studies equally showed that megakaryocytes are predominately and directly derived from HSCs, further contributing to the hypothesis that these cells can bypass the differentiation stages between MPPs and MEPs (Notta et al., 2016).

More recent studies using single-cell technology, combined with computational analysis, have shown that haematopoiesis is a continuous process, as opposed to a stepwise progress to lineage commitment (Velten et al., 2017). Moreover, unilineage-restricted cells have been suggested to originate directly from a ‘continuum of low-primed undifferentiated HSPCs’ (CLOUD-HSPC), without any major transition through the multi- and bi-potent stages (Velten et al., 2017). However, even though these new techniques have further allowed the understating of normal human haematopoiesis, it has led to the generation of conflicting observations. Instead, some authors defend the existence of a structured hierarchy with a heterogenous haematopoietic landscape (Buenrostro et al., 2018; Karamitros et al., 2018; Pellin et al., 2019).

Altogether, with the use of high-throughput and single-cell methodologies, such as RNA sequencing, mass cytometry and immunophenotypic analysis, it was possible to not only identify new populations, but also enlighten the specific processes cell undergo, until reaching a terminal differentiated state, suggesting a need to revise the classical hematopoietic hierarchy roadmap.
1.1.4 Regulation of normal human haematopoiesis

1.1.4.1 Transcription factor regulation of haematopoietic development

Transcription factors (TFs) play an essential role during haematopoiesis, being responsible for several processes, such as stem cell maintenance, lineage commitment and differentiation. These are often cell-type restricted, which means they can drive the expression of characteristic lineage-specific target genes, leading to the development of a specific cell subtype (Figure 1.3). The differentiation of haematopoietic precursors is associated with two main processes: the reduction on the cells’ self-renewal potential and the stepwise acquisition of specific lineage identity. However, the interpretation of the roles of these TFs was framed on the classical model, and the current understanding shows this to be oversimplistic (1.1.3).

Myeloid cell differentiation is orchestrated by a small number of TFs, which include PU.1 (aka SPI1) (Klemsz et al., 1990), CCAAT/enhancer binding proteins (including CEBPA, CEBPB and CEBPE) (Zhang et al., 1997; Tanaka et al., 1995a; Yamanaka et al., 1997), growth-factor independent 1 (GFI1) (Hock et al., 2003) and interferon-regulatory factor 8 (IRF8) (Holtschke et al., 1996). These also include, at the stem-cell level, RUNX1 (Okuda et al., 1996) and stem-cell leukaemia factor (SCL or TAL1) (Shivdasani et al., 1995). These TFs are often classified as master regulators of haematopoiesis due to their ability to promote the expression of several myeloid genes, such as those encoding receptors for macrophage-CSF (M-CSF), G-CSF or granulocyte/macrophage-CSF (GM-CSF).

The first TF to be expressed are those that orchestrate the formation of the HSC pool, which include SCL and RUNX1. Mice lacking either one of these factors are embryonically lethal and have no detectable haematopoiesis, indicating that these are essential in the generation of foetal HSCs (Goode et al., 2016; Suter et al., 2011). In adult haematopoietic cells, conditional deletion of these genes showed that they are not required for the maintenance of HSCs in the BM; instead, the effects observed are more lineage-specific. Regarding RUNX1, its conditional ablation inhibited CLP production, blocked B- and T- cell maturation and decreased the formation of platelets (Ichikawa et al., 2004). A subsequent study showed that this ablation contributed to the development of mild myeloproliferative syndrome (Growney et al., 2005).
Chapter 1

Figure 1.3 – Transcriptional regulation of myeloid development

Developmental stages and transcriptional regulation of monocytic and granulocytic differentiation. RUNX1 and stem-cell leukaemia factor (SCL) are required for the generation HSC, whilst growth-factor independent 1 (GFI1) and CCAAT/enhancer binding protein-α (C/EBPα) regulate HSC self-renewal potential. Moreover, C/EBPα is essential for the differentiation of CMP into GMPs. For granulocytic development, both GFI1 and C/EBPε are crucial mediators. Monocytic differentiation relies on the expression of PU.1 and IRF8. Bars represent the controlled expression of different TF during myeloid cell development. Adapted from (Rosenbauer and Tenen, 2007).
Whereas expression of PU.1 is essential for the differentiation of HSCs into CMPs, the next step of development into GMPs requires the expression of C/EBPα. C/EBPα is a basic-region leucine zipper TF expressed by HSCs, myeloid progenitors and granulocytes, but absent in macrophages (Akashi et al., 2000; Radomska et al., 1998). In mice, C/EBPα deficiency led to the absence of GMPs and granulocytic-committed cells, but normal numbers of CMPs (Zhang et al., 1997; Zhang et al., 2004b). Interestingly, C/EBPα is not required for granulocytic differentiation following GMP commitment (Zhang et al., 2004b). In addition to having a role in myeloid differentiation, C/EBPα is also responsible for controlling the cells’ self-renewal potential, as ablation of C/EBPα in HSCs results in an increased repopulation activity in mouse transplant models (Zhang et al., 2004b). Additionally, this TF has also been shown to be a strong promoter of cell-growth arrest by regulating cell cycle exit (Johnson, 2005).

Following development into GMPs, cells undergo further differentiation into macrophages or granulocytes, following the expression of both PU.1 and IRF8. IRF8 is expressed in several haematopoietic lineages, including HSCs, B cells, dendritic cells and in resting T cells (Driggers et al., 1990). Furthermore, expression of IRF8 is higher in macrophages, but not in granulocytes (Tamura et al., 2000). In mice, loss of IRF8 led to the development of myeloproliferative syndrome resembling CML (Holtschke et al., 1996), characterised by an increased number of granulocytic-derived cells at the expense of macrophages (Scheller et al., 1999).

Granulocytic differentiation requires the expression of two additional TF, GFI1 and C/EBPε. GFI1 is a transcriptional repressor detected in HSCs, neutrophils and B- and T-precursor cells (Hock et al., 2003; Karsunky et al., 2002a; Hock and Orkin, 2006). Mice deficient for GFI1B were seen to lack lymphocytic progenitor differentiation, as well as neutrophilic granulocytes (Hock et al., 2003; Karsunky et al., 2002b), even though the development of early myeloid progenitors remained unchanged. Similarly, ablation of C/EBPε led to abnormal granulocytic development beyond the promyelocyte stage (Yamanaka et al., 1997). In addition to having a role in granulocytic differentiation, deletion of GFI1 also impacts normal HSC development. Even though, phenotypically, GFI1-deficient cells resembled normal HSCs, these were functionally impaired, as they displayed reduced self-renewal capacity due to cell cycle changes (Hock et al., 2004).

Taken together, haematopoietic development is a complex process, relying on the hierarchical expression of key TFs, responsible for promoting HSC differentiation into mature
myeloid cells. The dysregulation of these TFs has been previously shown to play an important role in the leukaemogenic process, and often present themselves as therapeutic targets (Tenen, 2003). Several studies have shown that absence of a lineage specific TF results in a perturbed differentiation process or complete absence of the respective lineage. For instance, knocking-out of GATA-1, an erythroid TF, results in a complete absence of erythroid and megakaryocytic differentiation in mice, whilst other lineages remained unperturbed (Pevny et al., 1991). Moreover, whilst removal of RUNX1 or TAL1 influences HSC formation and the development of the entire haematopoietic system, absence of PU.1 largely affects the myeloid and B-cell lineages.

1.1.4.2 Cytokines and growth factors

The production of haematopoietic-derived cells is under the control of several haematopoietic cytokines (Kondo et al., 2000; Mossadegh-Keller et al., 2013; Brown et al., 2018). Cytokines are extracellular ligands that can stimulate biological responses through the binding and activation of cytokine receptors. Important cytokines within the haematopoietic system include interleukins (ILs), CSFs, interferons, erythropoietin (EPO) and TPO. Similarly to haematopoietic development, cytokines have been arranged in a hierarchical system, as sub-populations of HSC express certain cytokine receptors associated with specific cell lineages (Figure 1.1). Moreover, certain cell types require the action of multiple cytokines simultaneously, observed in HSCs and megakaryocyte progenitors. For instance, whilst IL-3 stimulates the growth of most lineages, EPO exclusively regulates erythroid development (Ihle, 1992).

EPO supports erythroid progenitor cells by promoting the expression of the erythroid-affiliated TF GATA-1 (Koury and Bondurant, 1990; Grover et al., 2014). Ablation of EPO and its respective receptor results in in vivo death mid-gestation due to severe anaemia, even though these mice displayed erythroid progenitor cells within the BM (Wu et al., 1995; Kieran et al., 1996; Lin et al., 1996). Moreover, these subjects showed signs of early erythropoiesis, suggesting that another cytokine, possibly TPO, is responsible for supporting early erythropoiesis (Kieran et al., 1996). Indeed, absence of TPO or its receptor resulted not only in loss of normal number of platelets and megakaryocytes, but also reduced numbers of HSCs, associated with a significant reduction in stem cell expansion following BM transplantation,
suggesting a non-redundant role in stem cell activity (Gurney et al., 1994; Alexander et al., 1996; Carver-Moore et al., 1996; Solar et al., 1998; Fox et al., 2002).

In high concentrations, FLT3 has been shown to drive HSC development towards a myeloid-lymphoid fate, whilst supressing megakaryocyte and erythroid development, by promoting to the upregulation of PU.1 (Tsapogas et al., 2014; Onai et al., 2006). Targeted disruption of FLT3 resulted in a significant reduction in the BM haematopoietic progenitor pool size and subsequent decrease in the number of mature myeloid, B-, natural killer (NK) and dendritic cells (DC) (Mackarehtschian et al., 1995).

SCF is another cytokine that has been shown to be essential for HSC survival, proliferation and differentiation (Broudy, 1997; Hartman et al., 2001). Loss of SCF or its receptor c-kit results in severe macrocytic anaemia (Ogawa et al., 1991). Moreover, following the administration of a neutralising antibody against c-kit, adult mice show complete depletion of progenitor cells and, eventually, absence of all mature myeloid and erythroid cells within the BM (Ogawa et al., 1991).

By growing daughter cells of granulocyte-macrophage colony forming cells in the presence of either G- or GM-CSF, Metcalf and Burgess observed the differentiation of these cells into granulocytes and macrophages, respectively (Metcalf and Burgess, 1982). Unexpectedly, null mutations of GM-CSF or its corresponding receptor had no consequence in the normal numbers of myeloid progenitor cells (Stanley et al., 1994; Nishinakamura et al., 1995; Robb et al., 1995). On the other hand, mice with a null mutation in both G-CSF and its receptor were shown to present incomplete granulopoiesis, characterised by chronic neutropenia, a decrease in mature myeloid cells in the BM and decreased neutrophil release from the BM (Lieschke et al., 1994; Liu et al., 1996; Semerad et al., 2002). However, mice with ablated G-CSF receptor only presented a mild reduction in the proportion of committed myeloid progenitors. These observations promoted the discussion that cytokines have the ability to act compensatively upon the complete absence of a certain molecule, as deletion of a single cytokine or receptor never results in the complete absence of a specific haematopoietic lineage (Brown et al., 2018).
1.2 Acute Myeloid Leukaemia

1.2.1 Overview

Acute myeloid leukaemia (AML) is a malignant blood cancer characterised by the clonal expansion of abnormal myeloid cells, firstly within the BM, with subsequent exfiltration of these cells through the bloodstream into other organs, such as spleen and liver. As the name suggests, this is an extremely fast-growing disease, with symptoms appearing following a few weeks of disease development. These can include fatigue, fever, recurrent infections, persisting bruising and bleeding, amongst others. AML is the most common acute leukaemia in adults (Khwaja et al., 2016); though rare in comparison with the most prevalent cancer types. For instance, whilst approximately 40,000 people are diagnosed every year in the UK with colorectal cancer, only about 3,000 are diagnosed with AML (Cancer Research UK, 2021). AML is diagnosed based on the accumulation of myeloid blasts within the BM, which are subjected to immunophenotyping, cytogenetic and molecular characterisation to distinguish between AML subtypes (1.2.3). Currently, the presence of abnormal molecular and cytogenetic features at diagnosis are considered the most important prognostic factors and are highly predictive of complete remission (CR) rates, disease-free survival, risk of relapse and overall survival (OS) (Döhner et al., 2017). However, the fact that these cells often exhibit several molecular abnormalities, contributes to disease heterogeneity (Ley et al., 2013; Lindsley et al., 2015). Even though significant improvements have been made in the treatment of this disease, prognosis in elder patients, who account for the majority of the new cases, remains poor (Shah et al., 2013a).

1.2.2 Pathophysiology of AML

AML develops as a result of genetic and epigenetic changes in myeloid progenitor cells. In most cases, there is no predisposing factor that makes individuals more susceptible to developing AML. Several models have shown that more than one cooperating abnormality is necessary to develop AML (Grisolano et al., 2003; Schessl et al., 2005). Originally, the developmental process of AML was thought of as a ‘2-hit’ model, in which it was suggested that the clonal expansion of malignant cells required the acquisition and cooperation of, at least, two mutations from different protein/gene classes (Conway O'Brien et al., 2014; Gilliland and Griffin, 2002). According to this model, Class I mutations, which result in activated signalling
pathways that regulate cell proliferation and survival, must occur in combination with Class II mutations, involving TFs implicated in the regulation of cell differentiation and self-renewal (Takahashi, 2011; Kihara et al., 2014). Common Class I mutations include FLT3, kirsten- or neuroblastoma-rat sarcoma virus gene (K/NRAS, respectively), tumour protein p53 (TP53) and tyrosine-protein kinase KIT (c-KIT). Interestingly, these mutations often occur in sub-clonal cellular fractions, occurring late in pathogenesis of disease, suggesting that these are a result of late clonal events; and whilst they may impact treatment response, the evidence that they initiate disease is scarce (Papaemmanuil et al., 2016). Further studies have also highlighted the role of signal transducer and activator of transcription 3 (STAT3) in the stimulation of cellular proliferation and survival (Cook et al., 2014; Ghoshal Gupta et al., 2008; Yamada and Kawauchi, 2013). Enhanced tyrosine phosphorylation of STAT3 occurs as a result of increased secretion of cytokines, such as IL-6, or due to mutations in receptor tyrosine kinases, seen in up to 50% of q AML cases and associated with a worse prognosis (Schuringa et al., 2000; Steensma et al., 2006).

Significant Class II mutations include nucleophosmin-1 (NPM1) and CEBPA, which are found in ~ 27% and 6% of cases, respectively (Ley et al., 2013) (Table 1.1), as well as the expression of the fusion proteins RUNX1-ETO (aka AML1-ETO or RUNX1-RUNX1T1) and PLM-RARA. However, in recent years, this has been shown to be an over-simplification of the leukaemogenic process, with the identification of mutations within epigenetic regulators, classified as Class III mutations (Sun et al., 2018). These have been shown to possess downstream effects on both cellular differentiation and proliferation, and include mutations in DNA-methylation and post-translational histone modification genes, such as DNA methyltransferase 3 alpha (DMNT3A), Tet methylcytosine dioxygenase 2 (TET2), Wilms tumour 1 (WT1), and isocitrate dehydrogenase (NADP(+)) 1 / 2 (IDH1 / IDH2) (Ley et al., 2013; Patel et al., 2012).

A low number of coding sequence mutations are found in AML, compared to most solid epithelial tumours (Ley et al., 2013). As a result, mutations associated with the development of AML are well defined, and their presence is important when classifying disease subtype (1.2.3), infer prognosis (1.2.4) and respective treatment (1.2.5).
Table 1.1 – Recurrent mutations in AML

Table outlining recurrent molecular and cytogenetic aberrations are frequently observed in AML. Adapted from (Grove and Vassiliou, 2014).

<table>
<thead>
<tr>
<th>Type of mutation</th>
<th>Gene</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Signal transduction Genes</strong></td>
<td>FLT3, NRAS, c-KIT, PTPN11</td>
<td>59%</td>
</tr>
<tr>
<td><strong>DNA modification Genes</strong></td>
<td>DNMT3A, TET2, IDH1/2</td>
<td>44%</td>
</tr>
<tr>
<td><strong>Chromatin Modifiers</strong></td>
<td>MLL-fusions, ASXL1, EZH2</td>
<td>30%</td>
</tr>
<tr>
<td></td>
<td>NPM1</td>
<td>27%</td>
</tr>
<tr>
<td><strong>Fusion Genes</strong></td>
<td>PML-RARA, MYH11-CBFB, RUNX1-ETO</td>
<td>25%</td>
</tr>
<tr>
<td><strong>Myeloid transcription factors</strong></td>
<td>CEBPA, RUNX1</td>
<td>22%</td>
</tr>
<tr>
<td><strong>Tumour Suppressor Genes</strong></td>
<td>TP53, WT1, PHF6</td>
<td>16%</td>
</tr>
<tr>
<td><strong>Spliceosome Genes</strong></td>
<td>SF3B1, SRSF2, U2AF1</td>
<td>14%</td>
</tr>
<tr>
<td><strong>Cohesins</strong></td>
<td>SMC1A, SMC3, RAD1, STAG2</td>
<td>13%</td>
</tr>
</tbody>
</table>
Collectively, the most frequent mutations in AML can be found in genes related to signalling pathways, such as FLT3, presented as internal tandem duplications (ITD), or tyrosine kinase domain mutations (TKD); c-KIT and RAS, present in over half of AML patients (Table 1.1). Additionally, mutations in epigenetic regulators occur in 40% of patients (Table 1.1). For instance, DMNT3A mutations are observed in 20-25% of AML patients (Ley et al., 2010), and mouse knockout (KO) studies have shown that this gene plays a crucial role in limiting the self-renewal potential of haematopoietic stem and progenitor cells (HSPC) and in regulating myeloid differentiation (Challen et al., 2011). However, despite their prevalence in AML, these mutations alone are not sufficient to promote disease development. Progression to leukaemia often requires the acquisition of driver mutations, such as the expression of the fusion protein PML-RARA, not detected in healthy individuals (Abelson et al., 2018; Welch et al., 2012). Chromosomal aberrations have in fact been widely described in AML, and can include inversions (inv), such as inv(16), responsible for the expression of CBFβ-MYH11, or reciprocal translocations (t), including t(8;21) and t(15;17), resulting in the formation of the chimeric proteins RUNX1-ETO and PML-RARA, respectively, which can alter the cells differentiation process.

Even though several studies have attempted to understand the leukaemogenic process that leads to the development of AML, much remains to be understood. As suggested by the model described above, the pathogenesis and behaviour of AML relies on the interaction between different somatic mutations and chromosomal rearrangements, as well as epigenetic alterations. For instance, the c-KIT mutation has been associated with t(8;21) and inv(16), and its presence/absence significantly influences disease prognosis (De Kouchkovsky and Abdul-Hay, 2016). Similarly, mutations in NPM1 frequently occur with FLT3-ITD, or with mutations in the epigenetic genes DNMT3A and IDH-1 or IDH-2 (Patel et al., 2012).

1.2.2.1 Leukaemic Stem Cells

The existence and involvement of leukaemic stem cells (LSCs) in the development of AML has been discussed for several decades (Fialkow, 1974; Griffin and Löwenberg, 1986; Lapidot et al., 1994). This is based on the hypothesis that AML populations may mimic the hierarchical developmental process observed in normal haematopoiesis; just as normal HSC can give rise to progressively more mature cells, malignant cells can similarly generate a bulk population of AML cell blasts. However, whilst the normal stem cell has the ability to differentiate into the haematopoietic lineage, a mutated stem cell, even though it possesses similar properties to the
normal stem cell, can either divide carrying defects; or remain undifferentiated, accumulating as immature progenitor cells, or blasts (Jordan, 2007). These cells can undergo self-renewal, are multipotent, and highly proliferative. However, the chemotherapeutic agents used to effectively eradicate blast cells, have little, if any effect in the LSC population (Jordan, 2007; Hanekamp et al., 2017).

Phenotypically, LSC are described as having an HSC-like phenotype, expressing the CD34 marker, with the ability to initiate leukaemia when xenografted into immunosuppressed mice (Bonnet and Dick, 1997). This was examined by Bonnet and Dick by transplanting patient samples with an immature (CD34+CD38−) and more mature phenotypic cells (CD34+CD38+) into immunodeficient mice (Bonnet and Dick, 1997). Mice transplanted with more immature cells developed symptoms of overt disease, not observed in recipients of the more mature cellular population (Bonnet and Dick, 1997). Moreover, CD34+CD38− cells were able maintain their leukaemia initiation effect upon serial transplantations (Bonnet and Dick, 1997). These studies demonstrated the importance of LSC in the development and progression of AML. Subsequent studies have corroborated these finding by showing that LSCs can initiate leukaemia when transplanted into non-obese diabetic/severe combined immunodeficiency (NOD/SCID) mice through xenotransplantation assays (Feuring-Buske et al., 2003). Moreover, these cells were shown to possess unlimited self-renewal capacity, in a similar way to normal HSCs, associated with a higher proliferative potential compared to normal haematopoietic cells (Reinisch et al., 2015). Altogether, these findings indicate that LSCs can contribute to disease progression. However, due to their heterogeneity, there is no unique phenotype that defines this specific cellular subtype. For instance, even though the CD34 is a well-established LSC surface marker, c30% of patients have no CD34 expression in leukaemic blasts, possibly due to their origin in CD34− haematopoietic progenitors (Ng et al., 2016; Sarry et al., 2011; Quek et al., 2016; Taussig et al., 2010). Several additional cell surface markers have been identified as upregulated in LSCs, including CD44 (Jin et al., 2006), CD123 (Jin et al., 2009), and CD47 (Jaiswal et al., 2009); a more detailed phenotypical analysis of these cells will undoubtfully contribute to the development of new therapies targeted against this specific cell subtype, essential for achieving long-term remission.
1.2.3 Classification of AML

Several classification systems have been developed and employed to improve prognosis and treatment of AML, based on aetiology, morphology, immunophenotyping and genetics. In the 1970s, AML was categorized according to the French-American-British (FAB) classification system, mainly based on morphology and immunophenotypic criteria (Neame et al., 1986) (Table 1.2). This classified AML into eight major subtypes, ranging from FAB-M0, corresponding to immature AML cells, up to FAB-M7, described as acute megakaryoblastic leukaemia, displaying the most differentiated blasts (Bennett et al., 1976). However, to incorporate further AML characteristics, such as clinical features, morphology, immunophenotyping, cytogenetics and molecular genetics, the World Health Organization (WHO) developed a new classification system in 2008, with new revised versions released in 2016 (Vardiman et al., 2009; Arber et al., 2016). According to this model, AML can be subclassified into six categories (Table 1.3). AML with recurrent genetic abnormalities, includes AML with balanced translocations/inversions, as well as AML with gene mutations, accounting for approximately 20-30% of patients. Core binding factor (CBF) AML, which include the t(8;21)(q22;q22.1) and inv(16)(p13.1q22), are present in this category.

In 2017, a new European LeukemiaNet (ELN) classification system for the diagnosis and management of adult patients diagnosed with AML was developed, which stratified patients into three different outcome groups, based on the cytogenetics and mutation status of the ASXL1, CEBPA, FLT3, NPM1, RUNX1, and TP53 genes: favourable, intermediate or adverse (Döhner et al., 2017) (Table 1.4). For instance, the CBF leukaemias described above have relatively favourable outcomes. Similarly, patients with biallelic CEBPA mutations have been classified in the favourable risk group. On the other hand, poorer outcomes are associated with complex chromosomal alterations, including those involving the deletion of chromosome arms 5q, 7q and/or 17p, frequently occurring in combination with TP53 mutations.
**Table 1.2 – FAB classification of AML**

AML subtypes based on FAB classification. Adapted from (Kabel *et al.*, 2017).

<table>
<thead>
<tr>
<th>FAB subtype</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>M0</td>
<td>Undifferentiated acute myeloblast leukaemia</td>
</tr>
<tr>
<td>M1</td>
<td>Acute myeloblast leukaemia with minimal maturation</td>
</tr>
<tr>
<td>M2</td>
<td>Acute myeloblastic leukaemia with maturation</td>
</tr>
<tr>
<td>M3</td>
<td>Acute promyelocytic leukaemia</td>
</tr>
<tr>
<td>M4</td>
<td>Acute myelomonocytic leukaemia</td>
</tr>
<tr>
<td>M4&lt;sub&gt;EO&lt;/sub&gt;</td>
<td>Acute myelomonocytic leukaemia with BM eosinophilia</td>
</tr>
<tr>
<td>M5</td>
<td>Acute monocytic leukaemia</td>
</tr>
<tr>
<td>M6</td>
<td>Acute erythroid leukaemia</td>
</tr>
<tr>
<td>M7</td>
<td>Acute megakaryoblastic leukaemia</td>
</tr>
</tbody>
</table>
Table 1.3 – WHO (2016) classification of AML

AML subtypes based on the WHO (2016) stratification. Adapted from (Döhner et al., 2017).

**AML and related neoplasms**

**AML with recurrent genetic abnormalities**

- AML with t(8;21) (q22;q22.1); *RUNX1-RUNX1T1*
- AML with inv(16) (p13.1q22) or t(16;16) (p13.1;q22); *CBFβ-MYH11*
- APL with *PML-RARA*
- AML with t(9;11) (p21.3;q23.3); *MLLT3-KMT2A*
- AML with t(6;9) (p23;q34.1); *DEK-NUP214*
- AML with inv(3) (q21.3q26.2) or t(3;3)(q21.3;q26.2); *GATA2, MECOM*
- AML (megakaryoblastic) with t(1;22) (p13.3;q13.3); *RBM15-MKL1*
- Provisional entity: AML with *BCR-ABL1*
- AML with mutated *NPM1*
- AML with biallelic mutations of *CEBPA*
- Provisional entity: AML with mutated *RUNX1*

**AML with myelodysplasia-related changes**

**Therapy-related myeloid neoplasms**

**AML, not otherwise specified (NOS)**

- AML with minimal differentiation
- AML without maturation
- AML with maturation
- Acute monoblastic/monocytic leukaemia
- Pure erythroid leukaemia
- Acute megakaryoblastic leukaemia
- Acute basophilic leukaemia
- Acute panmyelosis with myelofibrosis

**Myeloid sarcoma**

**Myeloid proliferations related to Down syndrome**

- Transient abnormal myelopoiesis (TAM)
- Myeloid leukaemia associated with Down Syndrome

*APL* – Acute Promyelocytic Leukaemia
Even though cytogenetic analysis remains a critical tool for the diagnosis and stratification of AML patients, around 45% of all diagnosed cases present a normal karyotype (Grimwade et al., 2001). In general, normal karyotype AML is considered intermediate risk type; however, this depends on the degree of heterogeneity observed in these patients, hence why further molecular analysis is necessary for prognostic outlook and patient stratification. FLT3-ITD mutations are classified as a high risk AML, associated with increased risk of relapse (Yanada et al., 2005). NPM1 mutations, on the other hand, are associated with a favourable outcome (Döhner et al., 2005). However, patients diagnosed with a combined NPM1, and FLT3-ITD mutations correlate with intermediate factor risk.

1.2.4 Epidemiology, diagnosis, and prognosis of AML

AML is one of the most common types of leukaemia in adults, with a slightly higher incidence in men, as compared to women (56% vs. 44%; (Cancer Research UK, 2021)). Even though AML can occur in any age group, it is predominant in older adults, with an average age at diagnosis of 68 years old (Short et al., 2018). Since the early 1990s, AML incidence has increased by 29% in the UK (Cancer Research UK, 2021), partially due to the increase of therapy-related AML cases (McNerney et al., 2017). This incidence has remained stable in patients between the ages of 0-59 years; however, these have progressively increased in the age groups 60-69, 70-79 and >80 years old, by 17%, 36% and 72%, respectively (Cancer Research UK, 2021).

Furthermore, incidence of AML is higher in Caucasian people, as compared to Hispanics, African and Asian/Pacific Islanders (Kirtane and Lee, 2017). Several environmental risk factors have the ability to predispose individuals to the development of AML (Short et al., 2018), although no genetic factors have been identified to date. The risk of developing AML increases upon exposure to DNA damaging-agents, including benzene, cigarette smoke, ionising radiation and cytotoxic chemotherapy (Khwaja et al., 2016). An increased risk of developing AML can also be observed in first-degree relatives of patients diagnosed with several types of haematological disorders. Moreover, specific inherited disorders carry a particularly high risk of AML development, including Down syndrome, Fanconi anaemia and Bloom syndrome (Seif, 2011).
Table 1.4 – ELN 2017 risk stratification of AML

AML classification system based on the ELN stratification, from 2017. Adapted from (Döhner et al., 2017).

<table>
<thead>
<tr>
<th>Risk Category</th>
<th>Genetic Abnormality</th>
</tr>
</thead>
</table>
| **Favourable** | t(8;21) (q22;q22.1); *RUNX1-RUNX1T1*  
                | inv(16) (p13.1q22) or t(16;16)(p13.1;q22); *CBFB-MYH11*  
                | Mutated *NPM1* without *FLT3-ITD* or with *FLT3-ITD*low (allelic ratio < 0.5)  
                | Biallelic mutated *C/EBPα* |
| **Intermediate** | Mutated *NPM1* and *FLT3-ITD*high (allelic ratio ≥ 0.5)  
                  | Wild-type *NPM1* without *FLT3-ITD* or with *FLT3-ITD*low (allelic ratio < 0.5)*  
                  | t(9;11) (p21.3;q23.3); *MLLT3-KMT2A*  
                  | Cytogenetic abnormalities not classified as favourable or adverse |
| **Adverse** | t(6;9) (p23;q34.1); *DEK-NUP214*  
               | t(v;11q23.3); *KMT2A* rearranged  
               | t(9;22) (q34.1;q11.2); *BCR-ABL1*  
               | inv(3) (q21.3q26.2) or t(3;3) (q21.3;q26.2); *GATA2, MECOM*  
               | −5 or del(5q); −7; −17/abn(17p)  
               | Complex karyotype (3 or more chromosomal abnormalities), monosomal karyotype  
               | Wild-type *NPM1* and *FLT3-ITD*high (allelic ratio ≥ 0.5)  
               | Mutated *RUNX1* (if not co-occurring with favourable AML subtypes)  
               | Mutated *ASXL1* (if not co-occurring with favourable AML subtypes)  
               | Mutated *TP53* (associated with AML complex and monosomal karyotype)  

* Without adverse-risk lesions
The presence of >20% myeloblasts in the peripheral blood, including myeloblasts, monoblasts or megakaryoblasts, undoubtedly indicates a leukaemia. Immunophenotypic analysis of the patient samples can further facilitate the subclassification into the different AML subtypes (Bennett et al., 1976; Bene et al., 1995). Confirmation of AML arises following the expression of two of the following markers, detected in myeloblasts: MPO, CD13, CD33, CDw65 and CD117. Lymphoid antigens are detectable in ~25% of individuals with AML. The T cell antigen CD7 has been reported in 10–30% of patients (Jha et al., 2013), whilst the B cell antigen CD19 is abnormally expressed in 5-34% of patients (Shorbagy et al., 2016). Cytogenetic analysis and screening for common gene mutations and rearrangements are an essential part of not only the diagnostic process, but also for prognosis and treatment options. In fact, for three cytogenic groups, AML diagnosis can be based on cytogenetic analysis alone, regardless of blood count - t(8;21)(q22;q22), inv(16)(p13.1q22) and t(15;17)(q22;q12). Patients diagnosed with t(15;17) are associated with a good prognosis, along with those with t(8;21) and inv(16), whilst cytogenetically normal patients usually present an intermediate prognosis. Patients with a complex karyotype (three or more chromosomal abnormalities), inv(3) or t(6;9), on the other hand, are associated with an extremely poor prognosis (Döhner et al., 2017).

1.2.5 Treatment of AML

Even though conventional treatment for AML has remained similar over the last few decades, more targeted and specific therapies have emerged based on the patient’s disease subtype. An initial assessment is made based on the patients’ characteristics, which include age, performance status and pre-treatment comorbidities. The treatment strategy used for AML is often divided into induction, consolidation, and maintenance therapy. The main aim of induction therapy is to achieve CR, by using cytotoxic chemotherapeutic agents. In patients suitable for intensive treatment, standard chemotherapy agents include the use of a combination of infused cytarabine for 7 days, followed by 3 days of anthracycline, commonly daunorubicin (Short et al., 2018). In younger patients (<60 years), CR is observed in approximately 60-85% of patients; however, only 40-60% of older patients (>60 years) achieve remission; these represent almost 50% of all patients diagnosed and are, in general, unsuitable for intensive chemotherapy, due to comorbidities and poor performance status. Alternatively, these patients might be suited for less intensive regimes, including low dose cytarabine or hypomethylating agents (HMA), such as Azacitidine (Dombret et al., 2015; Seymour et al., 2017). CR is
achieved once the tumour burden has been reduced, with <5% blasts in the BM, the absence of Auer rods and extramedullary leukaemia, a neutrophil count of >1,000/μL and a platelet count >100,000/μL (Döhner et al., 2015).

Having achieved CR, consolidation chemotherapy is necessary to reduce the risk of relapse (Burnett et al., 2011a), generally centred on a cytarabine-based regimen. Alternatively, patients can be eligible for an allogenic HSC transplant. Even though this procedure usually improves the outcome of intermediate and poor-risk AML patients, it is highly dependent on patient selection, donor and optimal regimen (Short et al., 2018). Relapsed disease and leukaemia-associated complications are the most common causes of death. AML relapse is often associated with a substantial increase in molecular complexity, with multiple new subclones and mutations identified at the time of relapse, leading to increased resistance to cytotoxic chemotherapy (Short et al., 2018). For this reason, there is a constant need to develop new and more effective therapies, especially for older AML patients.

1.2.5.1 Targeted therapies

Treatment options for AML have been completely revolutionised by the development of a targeted therapy against acute promyelocytic leukaemia (APL). Currently, these patients are treated with a combination of all-trans retinoic acid (ATRA) and arsenic trioxide (ATO), which induce PML-RARA degradation and cell differentiation (reviewed in (Tomita et al., 2013)). Approximately 80-90% of patients diagnosed with APL are expected to achieve CR, associated with a long-term survival of over 98% (Khwaja et al., 2016; Sanz et al., 2019).

Since the discovery and approval of ATRA, several efforts have been made into developing new targeted therapies against specific mutations, namely FLT3. In patients, mutant FLT3 is constitutively active, leading to the induction of cell proliferation, whilst suppressing differentiation (Meshinchi and Appelbaum, 2009). Inhibition of tyrosine kinase receptors has been successfully used for the treatment of other blood cancers, including Philadelphia-chromosome positive leukaemias (De Kouchkovsky and Abdul-Hay, 2016). For the treatment of AML, first-generation targeted therapies were initially developed, including Sorafenib (Zhang et al., 2008b; Borthakur et al., 2011; Crump et al., 2010) and Midostaurin (Stone et al., 2005; Fischer et al., 2010). However, these are associated with a high incidence of off-target effects (De Kouchkovsky and Abdul-Hay, 2016; Short et al., 2018), which led to the
development of second-generation drugs, such as Quizartinib (Zarrinkar et al., 2009), Crenolanib (Galanis et al., 2014) and Gilteritinib (Ueno et al., 2014). A recurrent concern with FLT3-targeted therapies is the development of secondary mutations, mostly associated with its kinase domain (Daver et al., 2015). To overcome this, conventional FLT3 inhibitors are often used with other treatments, including HMA (Chang et al., 2016; Williams et al., 2013), Venetoclax (Ma et al., 2019; Singh Mali et al., 2021) or proteosome inhibitors (Larrue et al., 2016; Walker et al., 2016; Saliba et al., 2017). Several novel FLT3 inhibitors are currently being developed, such as multikinase inhibitors, including Ponatinib (Shah et al., 2013b), Cabozantinib (Lu et al., 2016; Fathi et al., 2018), Pexidartinib (Smith et al., 2020) and Ibrutinib (Wu et al., 2016); and next-generation inhibitors, for instance FF-10101 (Yamaura et al., 2018).

AML poses as the ideal target for monoclonal antibody-based immunotherapy, as cells are largely present in the blood and BM, tissues readily accessible to antibodies. However, applying this has been challenging as not many leukaemia-specific antigens have been identified to date. AML cells typically express antigens found on normal myeloid progenitor and differentiated cells, such as macrophages and monocytes, particularly CD45 (expressed in 97.2% cells), CD33 (95.3%), and CD13 (94.3%) (Khalidi et al., 1998). Additionally, CD56 expression in t(8;21) AML has been previously associated with a higher rate of relapse (Iriyama et al., 2013), making this marker a potential therapeutic target. The first clinically viable monoclonal antibody to be approved in haematological malignancies was Gemtuzumab ozogamicin (Mylotarg®), which targets CD33. By combining Mylotarg® with conventional induction chemotherapy on the first day of therapy of previously untreated AML patients, the MRC AML15 trial reported a significant survival benefit without increased toxicity in younger patients with favourable cytogenetics, particularly CBF leukaemias (Burnett et al., 2011b). Currently, this drug is used for the treatment of newly diagnosed CD33-positive AML in adults and for the treatment of relapsed or refractory CD33-positive AML in adults and paediatric patients above the age of 2 (Williams et al., 2019).

Patients unable to tolerate standard chemotherapy have profited from Venetoclax, a Bcl-2 inhibitor, often used in combination with the HMA decitabine or azacytidine (Williams et al., 2019). In 2018, the US food and drug administration (FDA) approved the combination of Venetoclax with either low dose cytarabine or demethylation therapy for older or unfit AML patients following successful clinical trials (Wei et al., 2019; DiNardo et al., 2019). Even
though there were improvements in remission rates and OS, resistance to targeted therapies remains a challenge in the treatment of AML (DiNardo et al., 2019). FLT3-ITD and TP53 defects were shown to correlate with adaptive resistance to Venetoclax alone as well as in combination with other therapies (DiNardo et al., 2019).

IDH inhibitors are used in patients with AML associated with either IDH1 or IDH2 mutations, including Ivosidenib (Popovici-Muller et al., 2018), an IDH1 inhibitor, and Enasidenib (Stein et al., 2017; Yen et al., 2017) targeting IDH2. Recently, a new therapeutic agent, CPX-351, has been approved for the treatment of AML with myelodysplasia-related changes and therapy-related AML, based on a liposomal formulation of cytarabine and daunorubicin, showing improved OS as compared to traditional ‘3+7’ treatment regimens (Lancet et al., 2014; Lancet et al., 2018).

Alternatively, treatment of AML has the potential to benefit from a less harmful approach, using chimeric antigen receptor (CAR) T cells, based on genetically engineered T cells able to recognise specific antigen epitopes. Following activation, CAR T cells are able to secrete several anti-tumour cytokines, resulting in the recruitment of other immune cells, tumour cell elimination and in the inhibition of tumour relapse (Marofi et al., 2021; Zhang and Xu, 2017). However, due to the challenge in identifying suitable cell surface markers in AML blasts, this therapy has encountered many challenges. Nevertheless, several markers have been identified as of interest in these cells, including CD33 (Kenderian et al., 2015; Wang et al., 2015b), CD38 (Yoshida et al., 2016) and CD123 (Gill et al., 2014; Thokala et al., 2016).

All these studies show that genomic profiling is of the upmost importance upon treatment design in AML. Currently, there are several treatment options based on different strategies, including monoclonal antibodies, as well as checkpoint inhibitors and cellular therapies, in a clinical trial stage (reviewed in (Short et al., 2018)). However, due to the heterogeneity observed in AML patients, these options are only available for a reduced cohort of patients.

1.3 Pathophysiology of RUNX1-ETO

The t(8;21)(q22;q22) reciprocal translocation is the most frequent chromosomal abnormality observed in AML and is frequently associated with the FAB-M2 subtype of AML (Rowley, 1984) (1.2.3). Cells with this translocation are usually associated with impaired
granulocytic maturation and expression of the cell surface markers CD13, CD19, CD34 and CD36 (Bitter et al., 1987). Molecular studies identified the rearrangement of the RUNX1 gene, located at chromosome 21q22, with the ETO gene, at chromosome 8q22, thus generating the fusion protein RUNX1-ETO (aka AML1-ETO, RUNX1-RUNXIT1) (Miyoshi et al., 1993; Miyoshi et al., 1991). AML patients with t(8;21) are often associated with good prognosis (Table 1.4), as compared to other subtypes of AML. However, a significant proportion of these patients eventually relapse, which can be attributed to a high heterogeneity within RUNX1-ETO leukaemia (Qin et al., 2017).

1.3.1 RUNX1

1.3.1.1 The RUNX family of transcription factors

Runt-related TF (RUNX) proteins belong to a family of TFs described as master regulators. Within the last decades, members of this family have been implicated in several processes, including proliferation, cell differentiation, apoptosis, and lineage determination. Furthermore, RUNX family members have been linked to the development of oncogenic processes and signalling pathways associated with cancer development (Otálora-Otálora et al., 2019).

RUNX genes were firstly identified in Drosophila melanogaster, where the runt gene was found to be essential for early embryonic segmentation (Gergen and Butler, 1988; Nüsslein-Volhard and Wieschaus, 1980). In mammals, there are three RUNX genes, each with distinct expression patterns subject to tissue-specification: RUNX1, RUNX2 and RUNX3. RUNX1 is localised on human chromosome 21q22 and is functionally important for haematopoietic cell differentiation (Yamagata et al., 2005; Dowdy et al., 2010). RUNX2 (or AML3) is located on chromosome 6p21 and has been described to be essential in osteogenesis (Lian and Stein, 2003; Lian et al., 2004; Komori, 2003). Lastly, RUNX3 (or AML2) is responsible for regulating gastric epithelial growth, and its located on chromosome 1p36 (Fukamachi, 2006). In cancer, RUNX genes have been associated with both oncogenic and tumour suppressor roles (Blyth et al., 2005; Blyth et al., 2010; Kilbey et al., 2008).

1.3.1.2 RUNX1 structure and function

All RUNX genes share a similar genomic structure (Levanon and Groner, 2004), and their expression is regulated by two promoters, one distal (P1) and one proximal (P2). Depending on the cells stage of differentiation, the promotors are triggered and generate distinct RUNX
isoforms (Ito, 2008; Levanon and Groner, 2004). Moreover, RUNX transcript alternative splicing can also lead to the expression of isoforms with altered properties (Bae et al., 1994; Miyoshi et al., 1995; Tanaka et al., 1995b; Stewart et al., 1997; Bangsow et al., 2001). This leads to distinct developmental expression patterns observed for all RUNX isoforms. At the transcriptional level, RUNX1 is the largest gene with nine exons and three extensively described isoforms: RUNX1a, RUNX1b and RUNX1c (Figure 1.4A). The proximal promoter P2 is responsible for the transcription of isoforms RUNX1a and RUNX1b, whilst RUNX1c is under the control of the distal promoter P1 (Miyoshi et al., 1995). Structurally, all RUNX1 protein isoforms comprise a highly conserved DNA binding domain (also known as Runt domain), a 128-amino acid sequence located near the N-terminus region, responsible for DNA binding at the consensus ‘PyGPyGGTPy’ RUNX motif (Kamachi et al., 1990; Wang and Speck, 1992), protein interactions (Lilly et al., 2016; Nagata et al., 1999), as well as for the nuclear localization of RUNX factors (Michaud et al., 2002; Telfer et al., 2004) (Figure 1.4B). RUNX1a lacks the transcriptional regulatory domains found in the C-terminal domain of the other RUNX1 isoforms (Tsuzuki et al., 2007). Moreover, the RUNX1c isoform possesses an additional five amino acid motif (VWRPY), essential for the recruitment of the Groucho/TLE family of co-repressors (Levanon et al., 1998; Seo et al., 2012b; Yarmus et al., 2006). All RUNX1 proteins further possess a conserved nuclear matrix-targeting signal sequence, important for the regulation of its activity and nuclear localization (Zaidi et al., 2001; Zeng et al., 1998). This region represents the basis for the functional diversity observed in RUNX proteins, including their ability to function as regulators of transcription (Chuang et al., 2013).

The multiple RUNX1 isoforms have been shown to play specific roles in HSC development and in the regulation of embryonic haematopoiesis. The RUNX1a isoform, the shorter of the three, lacks the transactivation domain and is thought to act as a doming negative (Levanon et al., 2001), promoting haematopoietic commitment (Ran et al., 2013) and increasing HSC renewal (Tsuzuki et al., 2007; Tsuzuki and Seto, 2012). In adult haematopoiesis, RUNX1c is the dominant isoform, whilst RUNX1b is exclusive to progenitor sub-populations of granulocytes/macrophages, megakaryocytes and lymphoid lineages (Bee et al., 2009; Draper et al., 2016; Telfer and Rothenberg, 2001). In fact, terminal differentiation of granulocytes/macrophages and lymphoid cells relies on the downregulation of RUNX1b, and its expression correlates with increased proliferation and colony-forming unit-culture activity in MEPs (Draper et al., 2016).
Figure 1.4 – Structural representation of RUNX1 isoforms genes and proteins

(A) Genomic organisation of the human RUNX1 locus. The RUNX1a and RUNX1b isoforms are transcribed from the proximal promoter P2, whilst RUNX1c is under the control of the distal promoter P1 and contains a unique N-terminal sequence encoded by exons 1 and 2. Black boxes correspond to untranslated regions, whilst blue boxes represent coding regions. Adapted from (van der Kouwe and Staber, 2019); (B) Schematic representation of the RUNX1a, RUNX1b and RUNX1c proteins and their functional domains. RUNX1a solely contains the RHD domain, whereas RUNX1b and RUNX1c contain the RHD, TAD and VWRPY motif. Below are represented TF and regulators that interact with RUNX1, and the corresponding regions they interact with. Adapted from (van der Kouwe and Staber, 2019; Lam and Zhang, 2012).

NLS – Nuclear localisation signal; RHD – Runt homology domain; TAD – Transactivation domain.
1.3.1.3 Transcriptional regulation by RUNX1

The RUNX family of genes represent the α subunit of a heterodimeric complex formed by dimerization with the core binding factor subunit β (CBFβ), ubiquitously expressed and encoded by a single gene in mammals, which in itself is unable to bind to DNA (Kamachi et al., 1990; Ogawa et al., 1993; Wang et al., 1993). However, upon binding to members of the RUNX family, CBFβ increases the RUNX DNA-binding affinity and promotes complex stabilization, by triggering flexible DNA-recognition loops (Bravo et al., 2001; Huang et al., 2001; Tang et al., 2000; Yan et al., 2004). The cooperative process between RUNX1 and CBFβ is responsible for not only regulating ubiquitin-mediated degradation of RUNX1 (Huang et al., 2001), but also enhancing RUNX1 phosphorylation/acetylation responsible for a decreased interaction with transcriptional repressors (Wee et al., 2008). The RUNX1-CBFβ complex has been shown to interact with PU.1, C/EBPα, histone acetyltransferase p300 (p300), mammalian Sin3a (mSin3a) and Friend leukaemia integration 1 (FLI1) (Zhang et al., 1996; Petrovick et al., 1998; Kitabayashi et al., 1998b; Imai et al., 2004; Huang et al., 2009b). Furthermore, it also possesses the ability to regulate several molecules, including the growth factors GM-CSF, MPO and IL-1; the surface receptors T Cell Receptor Alpha and Beta Locus (TCRA and TCRB, respectively), M-CSF receptor and FLT3; the signalling molecule Cyclin Dependent Kinase Inhibitor 1A (CDKN1A); transcriptional activators STAT3 and MYC; and, lastly, proliferation and survival regulators BLK and BCL-2 (Michaud et al., 2003; Ito, 2004). In addition to its interaction with CBFβ, RUNX1 has been shown to interact with other TFs and transcriptional co-regulators, including ETS Proto-Oncogene 1 (ETS1), with which it coordinates transcriptional activity through the Runt domain within RUNX1, thus eliminating the requirement for CBFβ and resulting in enhanced DNA-binding ability of the two proteins and the synergistic activation of a promoter (Ito, 2008; Kim et al., 1999). Even though, on their own, RUNX TFs are often characterised as weak, their interaction with other TF help regulate its target genes in a tissue-specific manner by promoting chromatin de-condensation, allowing the recruitment of other transcriptional regulators (Zaret and Carroll, 2011) (Figure 1.5).
RUNX1 can act both as a transcriptional activator or repressor, depending on the presence of coactivators/corepressor at a specific time point. (A) CBFβ stabilizes RUNX1 and protects it from degradation. Histone acetyltransferases (HATs) are recruited via p300 and CBP proteins, activating gene transcription through histone acetylation. Adapted from (Duque-Afonso et al., 2014; Brettingham-Moore et al., 2015). (B) RUNX1 can also recruit corepressors and other epigenetic modifiers, such as PRMT6 and PRMT4, as well as histone deacetylases (HDAC), which result in the repression of RUNX1 activity. Open circles represent methylation sites [unmethylated]. Adapted from (Brettingham-Moore et al., 2015)
1.3.1.4 Post-translational modifications

In addition to being transcriptionally regulated, RUNX1 is also under the control of several post-translational mechanisms, which influence its protein activity, subcellular localisation and stability, correlated with genes essential for myeloid and lymphoid differentiation (Zhao et al., 2008; Seo et al., 2012a). These processes are generally associated with chromatic modifiers, co-factors and other TFs targeting certain regulatory regions.

Transcriptional activation relies in the recruitment of co-activator proteins to certain promoters, which often possess histone acetyltransferase activity. RUNX proteins, due to their abundance in lysine residues, are often modified by lysine acetyltransferases (KAT), which stimulate their transcriptional activity (Blumenthal et al., 2017). P300 is a member of this family, and its mediated acetylation of RUNX1 increases the latter’s binding to DNA and its transcriptional activation (Yamaguchi et al., 2004). Lysine acetyltransferase 6A (MOZ) is another RUNX1 co-activator that, upon binding to the RUNX1 C-terminal transactivation domain promotes the expression of genes involved in monocyte/macrophage differentiation (Kitabayashi et al., 2001).

The transcriptional activity of RUNX1 is further regulated by methyltransferases. The mixed lineage leukaemia (MLL) lysine methyltransferase physically interacts with the N-terminal region of RUNX1, stabilising it by inhibiting its poly-ubiquitination (Huang et al., 2011). The protein arginine N-methyltransferase-1 (PRMT1) disrupts the association between RUNX1 and the co-repressor SIN3A, thus enhancing RUNX1 transcriptional activity and promoting its binding to target gene promoters (Zhao et al., 2008).

The most studied post-translational mechanism to control RUNX1 activity is phosphorylation. This is mediated by kinases activated by haematopoietic cytokines and growth factors, as well as cell cycle regulatory proteins. Extracellular signal-regulated kinases (ERK) phosphorylate the C-terminus domain of RUNX1, thus preventing RUNX1 from interacting with SIN3A and enhancing RUNX1-mediated transcription (Imai et al., 2004). Additionally, RUNX1 is phosphorylated by the homeodomain interacting protein kinase 2 (HIPK2), thus inducing p300 phosphorylation and subsequent transcriptional activation (Wee et al., 2008; Aikawa et al., 2006). During cell cycle, the G1 to S transition is directly regulated by RUNX1, following which the RUNX1 protein is ubiquitously degraded during the G2/M phase, a process triggered by cyclin-dependent kinase-1 (CDK1) and CDK6 (Biggs et al., 2006). Furthermore, CDK1/2/6
also phosphorylate RUNX1 (Zhang et al., 2008a), reducing its interaction with histone deacetylase-1 (HDAC1) and HDAC3, further promoting transcriptional activation (Guo and Friedman, 2011). Moreover, upon cytokine stimulation, the PIM1 kinase possesses the ability to interact with RUNX1, increasing its transactivation activity (Aho et al., 2006).

In addition to increasing RUNX1 transcriptional abilities, post-translational modifications can also negatively impact transcription, through deacetylation, methylation and phosphorylation processes. Physiologically, HDAC complexes participate in several processes, including chromatin remodelling and gene expression, and have been shown to interact with RUNX1, in particular HDAC1, SIN3A and Gro/TLE (Levanon et al., 1998; Imai et al., 2004). Interestingly, an increased recruitment of co-repressors is observed in chromosomal translocations involving RUNX1, as compared to wildtype RUNX1, suggesting the repression of its target genes (Guidez et al., 2000).

Additionally, the methyltransferase PRMT4, highly expressed in HSCs, leads to the assembly of a repressive complex that influences myeloid differentiation (Vu et al., 2013). RUNX1 has the ability to form a repressor complex with PRMT6, thus mediating the repression of genes preceding megakaryocytic differentiation (Herglotz et al., 2013). Similarly, RUNX1 phosphorylation can impact its gene expression. Megakaryocytic differentiation has been shown to be repressed by RUNX1 tyrosine phosphorylation (Huang et al., 2012). RUNX1’s interaction with c-SRC and the tyrosine phosphatase non-receptor type 11 (SHP2) leads to the inhibition of the interaction between RUNX1 and CBFβ and, consequently, GATA-1 and FLI1 (Huang et al., 2012).

1.3.1.5 The role of RUNX1 in normal haematopoiesis

RUNX1 is expressed in all haematopoietic cells except for mature erythrocytes (North et al., 1999; North et al., 2002). Absence of RUNX1 was shown to result in foetal death due to severe haemorrhaging of the central nervous system, due to defective angiogenesis, as well as lack of definite haematopoiesis (Okuda et al., 1996; Wang et al., 1996). However, continuous expression of RUNX1 is not necessary for survival in adult mice, indicated by absence of lethality (Cai et al., 2011; Growney et al., 2005; Ichikawa et al., 2004; Jacob et al., 2010; Putz et al., 2006). In this model, RUNX1 deletion resulted in either the expansion (Growney et al., 2005; Ichikawa et al., 2004) or exhaustion (Jacob et al., 2010) of phenotypic HSCs. However,
studies suggest that deletion of RUNX1 results in a significant expansion of the HSPC compartment, as well as reduced apoptosis and ribosome biogenesis, possible contributing to a pre-leukaemic condition (Cai et al., 2011; Growney et al., 2005; Ichikawa et al., 2004; Jacob et al., 2010; Cai et al., 2015). Moreover, absence of RUNX1 in adult mice also resulted in decreased numbers of B and T cells, as well as lower platelet counts (Growney et al., 2005; Ichikawa et al., 2004). Additionally, a role for RUNX1 in balancing megakaryocytic differentiation has been proposed, through its interaction with AP-1, p300, GATA and ETS TFs (Elagib et al., 2003; Pencovich et al., 2013; Pencovich et al., 2011). RUNX1 has further been found to act as a transcriptional co-repressor, suggesting that its role is highly context dependant, relying on the complexes in which it functions (Figure 1.5).

In addition to its role in haematopoiesis, RUNX1 has also been associated with several immune cell processes, including T lymphocyte development in the thymus, in a RUNX1-dependent manner (Ikawa et al., 2004; Kawamoto et al., 1999; Krueger and von Boehmer, 2007; Perry et al., 2004; Petrie, 2007). RUNX1 induces BCL11b expression which, consequently, promotes the expression of T cell-lineage specific genes, including Thpok (Zbtb7b) and RUNX3 (Kojo et al., 2017; Liu et al., 2010). Conversely, loss of RUNX1 results in a block in normal T cell development (Egawa et al., 2007; Sato et al., 2003). Furthermore, RUNX1 expression is essential for the development of natural killer T cells (Tachibana et al., 2011). In the BM, loss of RUNX1 results in defects in early B cell development, as this is required, together with CBFβ, to cooperate with the TF EBF for progression into the pro-B cell stage (Maier et al., 2004; Seo et al., 2012a).

1.3.1.6 The role of RUNX1 in AML

Considering the role RUNX1 plays in haematopoietic development (1.3.1.5), it is not surprising that this gene has been found to be a recurrent target for mutations and chromosomal abnormalities in haematological disorders. RUNX1 has been shown to be implicated in more than 50 chromosomal translocations reported in paediatric acute lymphoblastic leukaemia (ALL), AML and myelodysplastic syndrome (MDS) (De Braekeleer et al., 2011). The most common chromosomal translocations involving RUNX1 are the t(8;21)(q22;q22), resulting in the expression of the fusion protein RUNX1-ETO (1.3.3); the t(12;21)(p13;q22), leading to the expression of ETV6-RUNX1, present in about 25% of patients with pre-B cell ALL (Golub et al., 1995); and t(3;21)(q26;q22), detected in 3% of therapy-related MDS and AML (Rubin et
al., 1990; Rubin et al., 1987), in which the N-terminal portion of RUNX1 is fused with one of three genes present on chromosome 3, including EVI, MDS1 or EAP (Nucifora et al., 1994). However, even though these translocations represent a major risk factor, additional cooperating mutations are necessary for leukaemogenesis to occur.

Another phenomenon related to cancer development involving the RUNX1 gene is the occurrence of somatic point mutations, identified in approximately 15% of de novo AML, especially in the FAB-M0 subtype, and in 3% of paediatric AML (Tang et al., 2009; Greif et al., 2012; Mendler et al., 2012; Skokowa et al., 2014). Even though mutations are also common in therapy-related MDS and AML (16-40%) (Harada et al., 2003; Christiansen et al., 2004), these are considered a rare event in leukaemia (Preudhomme et al., 2000; Harada et al., 2004). Somatic mutations can either be mono-allelic or bi-allelic, with the latter being predominant in undifferentiated AML (Osato, 2004). Mono-allelic mutations are usually localised in the Runt domain and influence the DNA-binding ability of RUNX1 (Mangan and Speck, 2011; Song et al., 1999; Matheny et al., 2007). Traditionally, these are defined as loss-of-function mutations; however, some mutants have shown gain of function. For instance, expression of DNA-binding RUNX1 mutants in BM progenitors led to an increased replating efficiency, even as compared to RUNX1-deficient cells (Cammenga et al., 2007). Furthermore, other mutations in the RUNX1 N-terminal domain have been observed, that do not result in clear transactivation defects, and are particularly associated with epigenetic modifications in RUNX1 target genes (Huang et al., 2011). Mutations are further associated with older age, male gender, and poor disease prognosis, as compared to wild-type RUNX1 (Tang et al., 2009; Greif et al., 2012; Mendler et al., 2012). RUNX1 mutations are often observed in combination with FLT3-ITD, FLT3-TKD or mutations in driver genes, such as CEBPA, DNMT3A, NRAS, KIT, IDH1, IDH2, and WT1 (Gaidzik et al., 2011; Schnittger et al., 2011; Greif et al., 2012; Mendler et al., 2012). Interestingly, mutations in the RUNX1 and NPM1 genes are suggested to be mutually exclusive (Greif et al., 2012; Mendler et al., 2012; Gaidzik et al., 2011).

1.3.2 ETO

The ETO family of genes comprises three members: the first member to be identified, eight twenty-one (ETO, or MTG8); the myeloid transforming gene related protein-1 (MTGRI); and the myeloid transforming gene chromosome 16 (MTG16) (Miyoshi et al., 1993; Fracchiolla et al., 1998; Gamou et al., 1998; Kitabayashi et al., 1998a). Until its identification as RUNX1’s
fusion partner and member of the RUNX1-ETO fusion protein, the ETO protein was virtually unknown (Erickson et al., 1992; Miyoshi et al., 1993; Nisson et al., 1992). Its gene is located on chromosome 8q22 and consists of 11 exons (Wolford and Prochazka, 1998) (Figure 1.6A). The ETO protein consists of four evolutionary conserved domains, termed nervy homology regions (NHR) 1-4, or zinc-finger motifs (Figure 1.6B). Whilst NHR1 and NHR2 are responsible for the interaction with other protein motifs (Lutterbach et al., 1998; Davis et al., 1999; McGhee et al., 2003), NHR4, consisting of two non-classical zinc fingers, mediates protein interactions, but its unable to bind DNA (Lutterbach et al., 1998; Wang et al., 1998). Additionally, NHR3 is suggested to be involved with specific co-repressors (Hildebrand et al., 2001). Under physiological conditions, ETO is found in the nucleus, due to the presence of a non-canonical nuclear localisation signal (NLS), found between NHR1 and NHR2 (Davis et al., 1999; Odaka et al., 2000; Sacchi et al., 1998). ETO is expressed in several tissues, but its most abundant in the heart, brain, lungs and testis (Erickson et al., 1994; Wolford and Prochazka, 1998). Moreover, similarly to RUNX1, ETO is expressed in HSPC, but absent from differentiated leukocytes (Erickson et al., 1996).

ETO generally functions as a transcriptional co-repressor (Hiebert et al., 2001). In support of this, earlier studies aiming at isolating the human nuclear receptor co-repressor (N-CoR) were performed using ETO as bait (Wang et al., 1998). This complex had been previously shown to interact with DNA-bound nuclear receptors, thus repressing normal gene transcription though the recruitment of HDAC complexes (Hörlein et al., 1995). NH4 has been shown to be sufficient to promote the interaction of ETO with the N-CoR complex, as deletion of NHR 1-3 motifs abolished any binding between ETO and N-CoR, whilst a portion of ETO containing the NHR4 domain was enough to promote binding (Wang et al., 1998). Subsequent studies by Lutterbach et al. confirmed these observations and added that ETO can bind to the central portion of the N-CoR complex (Lutterbach et al., 1998).

Moreover, follow-up studies have established a relationship between ETO and the silencing mediator of retinoic acid and thyroid hormone receptors (SMRT) complex (Gelmetti et al., 1998; Chen and Evans, 1995), as well with HDACs (Wang et al., 1996; Rao et al., 1996; Wang et al., 1998; Lutterbach et al., 1998), the latter in combination with mSin3A, leading to the establishment of a N-CoR/mSin3A/HDAC1 complex (Wang et al., 1998).
Figure 1.6 – Structural representation of the ETO gene and protein

(A) Genomic organisation of the human ETO gene. Black boxes correspond to untranslated regions, whilst yellow boxes represent coding regions. Adapted from (Lin et al., 2017). (B) Schematic representation of the ETO protein and its functional domains. The ETO protein comprises 4 evolutionary conserved nervy homology domains (NHR), as well as an NLS. Below are represented regulators and complexes that interact with ETO, and the corresponding regions they interact with. Adapted from (Davis et al., 2003; Lam and Zhang, 2012).

NHR – Nervy Homology Region; NLS – Nuclear Localisation Signal
These observations suggest that ETO is a component of several co-repressor complexes \textit{in vivo}. However, several authors have demonstrated that, even though NHR4 is necessary to promote the interactions between ETO and both the SMRT and the N-CoR complex \textit{in vitro}, it is not sufficient, and requires the presence of other NHR domains (Zhang \textit{et al.}, 2001; Hildebrand \textit{et al.}, 2001).

ETO has been suggested to interact with sequence-specific DNA binding proteins, including the transcriptional repressors promyelocytic leukaemia zinc-finger (PLZF) and GFI1 (Melnick \textit{et al.}, 2000b; Costoya and Pandolfi, 2001; Grimes \textit{et al.}, 1996a; McGhee \textit{et al.}, 2003). PLZF is usually found in haematopoietic cells and has been shown to be downregulated as cells differentiate into the myeloid lineage (Costoya and Pandolfi, 2001). ETO binds PLZF, promoting the activity of histone deacetylases and enhancing the transcriptional repression induced by this protein (Melnick \textit{et al.}, 2000b). GFI1 expression is elevated in the thymus, spleen, BM and testis, and has been shown to play a role in haematopoietic cell differentiation and survival (Grimes \textit{et al.}, 1996a; Grimes \textit{et al.}, 1996b; Tong \textit{et al.}, 1998). Similarly to ETO, GFI1 is nuclear matrix-attached, and generally interacts with HDAC-containing complexes (McGhee \textit{et al.}, 2003). Altogether, these observations support the notion that ETO can interact with several co-repressor proteins and complexes, suggesting a role in haematopoietic development, specifically cell differentiation and proliferation.

ETO is involved in a non-random chromosomal translocation that binds it to the RUNX1 gene on chromosome 21 (Erickson \textit{et al.}, 1992; Gamou \textit{et al.}, 1998). Briefly, ETO contributes to the ability of RUNX1-ETO to bind to histone-deacetylase complexes, playing an essential role in the development of t(8;21) leukaemia (Hiebert \textit{et al.}, 2001) (1.3.3).

1.3.3 \textbf{RUNX1-ETO}

1.3.3.1 \textit{RUNX1-ETO structure}

The chromosomal breakpoints that generate RUNX1-ETO occur within intron 5 of the \textit{RUNX1} gene and in intron 1 of the \textit{ETO} gene (Tighe and Calabi, 1995; Zhang \textit{et al.}, 2002) (Figure 1.7A-B). The exact mechanism through which this translocation occurs remains unclear; however, Wnt/β-catenin signalling has been suggested to be an instigator of the genomic proximity between the \textit{RUNX1} and \textit{ETO} genes, facilitating translocation events (Ugarte \textit{et al.}, 2015).
Figure 1.7 – RUNX1-ETO structure and isoforms

(A) Genomic structure of RUNX1 on chromosome 21 and ETO on chromosome 8. The translocation occurs between exons 5 and 6 of the RUNX1 gene, and exons 1 and 3 of ETO. Black boxes indicate untranslated regions, whilst blue/yellow boxes indicate coding sequences. Adapted from (Lin et al., 2017). (B) Structure of the fusion gene RUNX1-ETO, containing exons 1-5 from the RUNX1 gene, and exons 2-11 derived from ETO. Adapted from (Swart and Heidenreich, 2021). (C) Protein structure of the full-length RUNX1-ETO and derived isoforms. RUNX1-ETO possesses the RHD domain (blue box) from RUNX1, responsible for DNA binding and heterodimerisation with CBFβ, as well as four NHR domain (green boxes), derived from the ETO domain, involved in the binding of co-repressor complexes. RUNX1-ETO also contains an NLS, represented in red. RUNX1-ETO9a lacks NHR3 and 4 domains, whilst RUNX1-ETO lacks solely the latter. Adapted from (Lin et al., 2017).

NHR – Nervy homology region; NLS – Nuclear localisation signal; RHD - Runt homology domain.
RUNX1-ETO encompasses the N-terminal region of the RUNX1 protein, containing the Runt domain, and the majority of the ETO protein, including the four NHR regions, resulting in a 752 amino-acid fusion protein (Miyoshi et al., 1993) (Figure 1.7C). The Runt domain present in the RUNX1 portion is responsible for mediating DNA binding, as well as the heterodimerisation with the co-factor CBFβ (1.3.1.5). Within the ETO portion, NHR2 has been shown to be important for leukaemic development, as it’s responsible for the recruitment on the NCor-SIN3A co-repressor complex, together with HDAC (Liu et al., 2006; Kwok et al., 2009), and its disruption results in a decrease in the self-renewal ability of RUNX1-ETO-expressing HSCs (Byrd et al., 2002). NHR4 is responsible for the recruitment of SMRT, as well as SIN3 and HDACs, via NCoR (Hug and Lazar, 2004), assisted by NHR3. Furthermore, between NHR1 and 2, the ETO domain contains an NLS, essential for the RUNX1-ETO nuclear localisation (Odaka et al., 2000; Barseguian et al., 2002).

In addition to the full-length RUNX1-ETO, containing exons 1-5 from RUNX1, and exons 2-11 from ETO, several isoforms of the RUNX1-ETO transcript have been identified. These include the alternative exons 9a and 11a, providing a stop codon following the amino acid encoded by exon 9 and 11, respectively (Wolford and Prochazka, 1998). From the 9a exon, a C-terminal truncated protein is generated, termed RUNX1-ETO9a, lacking both NHR3 and 4 (Figure 1.7C). This protein has a decreased ability to repress RUNX1-mediated gene activation (DeKelver et al., 2013b); however, expression of RUNX1-ETO9a promotes faster leukaemic development as compared to the unspliced form in mice (Yan et al., 2006), a reason why it is often used for RUNX1-ETO studies, in detriment of the full-length protein. RUNX1-ETO11a equally leads to the generation of a truncated RUNX1-ETO isoform, lacking the NHR4 domain, with reduced transcriptional activity (Kozu et al., 2005) (Figure 1.7C). However, the exact mechanism through which this isoform contributes to leukemogenesis has not been determined yet.

1.3.3.2 Cooperation between RUNX1 and RUNX1-ETO

A central mechanism for RUNX1-ETO-induced leukaemogenesis relies on the dominant inhibition of native RUNX1. Studies examining the consequences of RUNX1-ETO expression in mice have been essential to further the understanding of the leukaemogenic process observed in AML t(8;21). Similarly to previous observations in RUNX1 KO mice (1.3.1.5), RUNX1-ETO heterozygous mice display early embryonic lethality and identical haematopoietic defects (Yergeau et al., 1997; Okuda et al., 1998).
This suggests that RUNX1-ETO is able to block RUNX1 function, in a dominant-negative approach, thus influencing normal haematopoietic processes and leading to the development of a preleukaemic condition. Several studies have shown that the mechanisms through which RUNX1-ETO represses RUNX1-induced gene activation is through the recruitment of co-repressors by the ETO moiety (Gelmetti et al., 1998; Lutterbach et al., 1998; Wang et al., 1998; Amann et al., 2001; Zhang et al., 2001; Hiebert et al., 2001) (Figure 1.8A). To bypass embryonic death, a conditional RUNX1-ETO knock-in model was developed, which allowed the expansion of myeloid progenitor cells (Higuchi et al., 2002). However, even though these cells presented aberrant growth and differentiation phenotypes, RUNX1-ETO alone did not lead to the development of AML, suggesting that additional molecular abnormalities are necessary for leukaemogenesis (1.2.2).

In human cord blood (CB) derived CD34+ HSPC, ectopic expression of RUNX1-ETO resulted in the suppression of both erythroid and granulocytic growth, with an increased self-renewal capacity (Mulloy et al., 2002; Tonks et al., 2003; Tonks et al., 2004). Subsequent studies have further validated these observations, in which RUNX1-ETO was responsible for blocking normal myeloid development in embryonic stem cells, as a result of cell cycle arrest and through the interference with RUNX1 chromatin binding (Nafria et al., 2020). Similarly, these studies demonstrated that RUNX1-ETO alone is insufficient to induce leukaemia (Rhoades et al., 2000; Yuan et al., 2001; Higuchi et al., 2002; Mulloy et al., 2003; Tonks et al., 2004) and additional genetic abnormalities are necessary for this process to occur (Schessl et al., 2005; Wang et al., 2011b; Zhao et al., 2014; Goyama et al., 2016).

Subsequent studies have shown that there is more complexity to this process than initially thought. RUNX1 has previously been shown to be mutated in several types of AML (1.3.1.6); however, no inactivating mutations are observed in t(8;21) AML, suggesting that functional RUNX1 is necessary to promote the leukaemogenic process induced by RUNX1-ETO (Goyama and Mulloy, 2011). Later studies have supported this hypothesis, in which knockdown (KD) of RUNX1 resulted in growth inhibition and decreased survival of RUNX1-ETO leukaemic cells (Ben-Ami et al., 2013; Goyama et al., 2013).

Mechanistically, chromatin immunoprecipitation sequencing (ChIP-Seq) and RNA-sequencing (RNASeq) studies have identified RUNX1 as a member of the TF complex containing RUNX1-ETO, controlling gene activation or repression based on chromatin binding (Li et al., 2016).
Figure 1.8 – Mechanisms of transcriptional dysregulation induced by RUNX1-ETO

(A) For transcriptional repression, RUNX1-ETO recruits corepressors, including N-Cor, HDACs, mSin3A, further interacting with DNMT proteins no promote DNA methylation, resulting in the repression of target genes. (B) For transcriptional activation, RUNX1-ETO and native RUNX1, coupled with CBFβ, bind to RUNX1-motifs, promoting the recruitment of AP-1. RUNX1-ETO also interacts with the coactivators p300, PRMT1 and JMJD1C to activate target gene transcription. Representative target genes/miRNAs regulated by RUNX1-ETO are also shown. Open/full circles represent methylation sites (unmethylated and methylated, respectively). Adapted from (Lin et al., 2017; Duque-Afonso et al., 2014).
Additionally, KD of RUNX1 or RUNX1-ETO results in ‘inverse’ gene expression profiles, suggesting opposite roles in gene regulatory processes (Ben-Ami et al., 2013). Further studies using siRNA have further elucidated the transcriptomic program driven by RUNX1-ETO (Ptasinska et al., 2014; Ptasinska et al., 2012). This fusion protein has been shown to significantly reprogramme the transcriptional network and maintain leukaemia by impairing the expression of approximately 1,400 genes linked to myeloid and granulocytic differentiation (Ptasinska et al., 2014; Ptasinska et al., 2012). Approximately 50% of all genes affected were shown to either maintain normal expression or be upregulated by RUNX1-ETO, and were associated with several processes, including cell cycle progression, glycolysis, oxidative phosphorylation, MTOR signalling, RNA processing, and ribosome biogenesis (Ptasinska et al., 2014; Ptasinska et al., 2012).

1.3.3.3 Identification of RUNX1-ETO-mediated transcriptional de-regulation

RUNX1-ETO regulates gene expression through interactions with several different proteins via its Runt domain, through binding to CBFβ, a feature that is maintained in RUNX1-ETO expressing cells. In normal haematopoiesis, the RUNX1/CBFβ complex is essential for the emergence of haematopoietic cells and previous studies have shown that loss of either one of the factors results in embryonic death with lack of definitive haematopoiesis (1.3.1.5). Similarly, CBFβ promotes the binding of RUNX1-ETO to DNA, (Roudaia et al., 2009), although the exact nature of this mechanism remains to be elucidated (Gu et al., 2000; Kanno et al., 1998).

RUNX1-ETO requires the recruitment of complexes comprised of multiple proteins to specific target genes. Earlier studies have described RUNX1-ETO as a dominant negative regulator, acting upon RUNX1 transcriptional target genes, thus inhibiting normal myeloid differentiation (Meyers et al., 1995). Through its interaction with NCoR and mSin3A, ETO recruits HDACs, thus leading to changes in normal chromosome structure to allow a closer conformation on RUNX1-ETO target promoters (Lutterbach et al., 1998; Wang et al., 1998; Gelmetti et al., 1998; Amann et al., 2001; Hildebrand et al., 2001). In addition to HDACs, RUNX1-ETO has also been shown to recruit DNA methyltransferases (DNMTs), thus promoting DNA methylation (Liu et al., 2005). Subsequent studies performed in the human t(8;21) cell line Kasumi-1 have identified an endogenous complex containing RUNX1-ETO, as well as several other haematopoietic TF, including CBFβ, E proteins (HEB and E1A), LYL1, LMO22 and its binding partner LDB1 (Sun et al., 2013). KD of any of the members of this
complex resulted in a decrease in the protein level of other components, suggesting a mutual stabilisation mechanism.

In addition to acting as a transcriptional repressor, RUNX1-ETO can induce the expression of several target genes through the interaction with different transcriptional activators (Figure 1.8B). The p300 acetyltransferase increases RUNX1-ETO transcriptional activation ability by promoting its acetylation and recruiting transcriptional pre-initiation complexes (Wang et al., 2011a). The methyltransferase PRMT1, on the other hand, promotes residue methylation and contributes to gene activation (Shia et al., 2012). Similarly, RUNX1-ETO directly recruits the histone demethylase JMJD1C, enhancing transcriptional upregulation (Chen et al., 2015). Moreover, several haematopoietic TF have been shown to interact with the Runt domain of this fusion protein, including GATA1, CEBPA, and PU.1 (Pabst et al., 2001).

Transcriptomic analysis has revealed significantly altered gene expression profiles as a result of the expression of RUNX1-ETO in t(8;21) AML (Ross et al., 2004; Valk et al., 2004). Genome-wide studies performed in the t(8;21) cell lines Kasumi-1 and SKNO-1, as well as t(8;21) patient samples, revealed that 60-80% of RUNX1-ETO binding sites are shared with RUNX1 (Gardini et al., 2008; Ptasinska et al., 2012; Li et al., 2016; Ben-Ami et al., 2013). As described above (1.3.3.2), a basal level of RUNX1 is necessary for maintaining the cell growth induced by RUNX1-ETO, as the complete absence of its regulated differentiation genes is not feasible (Gardini et al., 2008). This further results in the recruitment of protein complexes with opposing roles. For instance, whilst RUNX1-ETO recruits HDACs, RUNX1 cooperates with p300 on genes repressed by this fusion protein (Ptasinska et al., 2014). Another example is the recruitment of DNMT1 and DNMT3A by RUNX1-ETO, whilst RUNX1 recruits TET histone demethylases (Gu et al., 2014) (Figure 1.8A). However, not all RUNX1-ETO binding sites contain RUNX1 motifs, suggesting that there are other factors that contribute to the binding of this fusion protein to DNA (Ptasinska et al., 2012; Maiques-Diaz et al., 2012). This binding can occur through E proteins, including HEB and E2A, as previous ChIP-Seq analyses identified the overrepresentation of E-boxes in RUNX1-ETO binding regions (Gardini et al., 2008; Zhang et al., 2004a; Sun et al., 2013). Moreover, genome-wide analysis identified ERG and FLI1, members of the ETS family of TFs, as facilitators in the RUNX1-ETO binding process towards DNA, further occupying similar genomic regions as to this fusion protein (Martens et al., 2012). Similarly, SP1 has been shown to interact with RUNX1-ETO and
promote DNA binding through SP1-binding sites; in fact, SP1 inhibition has been demonstrated to attenuate RUNX1-ETO-mediated transcriptional repression of target genes (Maiques-Diaz et al., 2012). Further studies have associated RUNX1-ETO binding to chromatic accessibility (Saeed et al., 2012; Maiques-Diaz et al., 2012). However, most of these studies have primarily focused on transformed AML cell lines. To address this, Tonks et al. performed microarray analysis on normal human haematopoietic progenitor cells transduced with RUNX1-ETO. This allowed the identification of dysregulated genes when RUNX1-ETO was expressed as a single abnormality in human CD34+ cells (Tonks et al., 2007). This study further promoted the identification of several genes dysregulated in RUNX1-ETO HSPC, and their potential involvement in the leukaemogenic process, including CD200 and γ-catenin.

1.3.3.4 Molecular mechanisms of RUNX1-ETO-induced leukaemia

The process through which RUNX1-ETO can induce cell proliferation involves several factors. HIF1α is a TF known to be overexpressed in t(8;21) patients, and the interaction between this factor and RUNX1-ETO have been shown to be necessary to drive proliferation in both in vivo and in vitro models (Peng et al., 2008). Mechanistically, HIF1α interacts with DNMT3a, leading to DNA hypermethylation (Gao et al., 2015). RUNX1-ETO further leads to the upregulation of Pontin, an ATPase involved in several cellular functions, including cell cycle progression and proliferation (Breig et al., 2014). Dysregulated signalling pathways, such as Wnt and RAS, have been shown to drive cell proliferation through its constitutive activation in t(8;21) patients (Kuchenbauer et al., 2006; Bacher et al., 2006). In HSC, expression of RUNX1-ETO results in an enhanced self-renewal potential, coupled with a block in normal cell differentiation.

RUNX1-ETO directly alters gene expression by interfering with the normal function of myeloid TF. By interacting with PU.1 and C/EBPa, RUNX1-ETO reduces their DNA binding activity, thus influencing their normal expression (Pabst et al., 2001; Vangala et al., 2003). Furthermore, RUNX1-ETO influences GATA1 expression, a major erythroid TF, by preventing its acetylation and normal transcriptional activity (Choi et al., 2006).

Additionally, RUNX1-ETO has been proposed to possess a role in the regulation of apoptosis by inhibiting normal apoptotic processes, through the upregulation of the anti-apoptotic proteins BCL-1 and BCL-XL (Klampfer et al., 1996; Chou et al., 2012). Subsequent
depletion of RUNX1-ETO results in an increase in apoptosis and cell cycle arrest. Further, RUNX1-ETO has been demonstrated to downregulate the expression of genes involved in DNA repair, including POLE and OGG1, leading to an increase in DNA damage (Alcalay et al., 2003; Krejci et al., 2008).

Dysregulation of tumour suppressor genes has also been observed in RUNX1-ETO induced leukaemia. RUNX1-ETO has been shown to transcriptionally repress p14ARF (aka CDKN2A) and NF1 (Linggi et al., 2002; Yang et al., 2005). RUNX3, another member of the RUNX family and a tumour-suppressor, is also suppressed by RUNX1-ETO (Cheng et al., 2008). Conversely, RUNX1-ETO also has the ability to upregulate the expression of cyclin-dependent kinase inhibitor 1A (p21WAF1 or CDKN1A) (Peterson et al., 2007). Transcriptional regulation of genes involved in signalling pathways has been shown to be modulated by RUNX1-ETO, including through the expression of γ-catenin (JUP) and Cox-2, which in turn activate Wnt signalling (Müller-Tidow et al., 2004; Zheng et al., 2004b; Zhang et al., 2013; Yeh et al., 2009; Tonks et al., 2007). Suppression of the leukaemogenic potential in RUNX1-ETO was achieved upon the use of Cox inhibitors, and by deleting β- or γ-catenin, indicating a critical role of the Wnt pathway in the maintenance of LSC in RUNX1-ETO cells (Zhang et al., 2013; Yeh et al., 2009). The TPO/MPL signalling pathway has been identified as a key pathway involved in the survival and leukemogenesis process induced by RUNX1-ETO through the upregulation of the anti-apoptotic protein Bcl-xL, as well as promotion of the PI3K/AKT and JAK/STAT pathways (Pulikkan et al., 2012; Chou et al., 2012). Further pathways that have been associated with RUNX1-ETO include NF-kB signalling (Nakagawa et al., 2011) and UBASH3B/CBL (Goyama et al., 2015).

Recently, through the integration of several techniques, including RNASeq, ChIP-Seq and DNase I hypersensitive site-se1 experiments, it was possible to identify CCND2 as a target of interest in RUNX1-ETO-driven leukaemogenesis, a member of the CDK6/CCND2 kinase complex (Martinez-Soria et al., 2018). Depletion of either RUNX1-ETO or CCND2, as well as pharmacological inhibition of CDK6, resulted in cell senescence, suggesting that RUNX1-ETO is responsible for driving cell cycle progression through the G1 phase by promoting the expression of both CDK6 and CCND2 (Byrd et al., 2002; Martinez-Soria et al., 2018).

Altogether, these observations indicate that RUNX1-ETO leads to significant changes in normal transcriptional processes, likely mediated through changes in TF expression. Whilst
several transcriptomic studies have been performed, these usually rely on an unsupervised analysis, without focusing on a specific gene class. Additionally, whilst transcriptomic analysis can be used as a strategy for target identification, it is not a powerful predictor of protein expression, reason why these techniques are often combined with alternative technologies, focused on analysing the cells’ proteomic profile. In line with this, Singh et al. aimed at identifying the RUNX1-ETO target proteins that could lead to novel insights into the pathogenesis of RUNX1-ETO-induced leukaemia on a post-genomic functional level. This was achieved by inducing the expression of RUNX1-ETO in a Tet-off-inducible U937 cell line coupled with mass spectrometry analysis (Singh et al., 2010). The authors showed that in these cells, the cells’ protein profile is drastically changed due to the expression of RUNX1-ETO, and were able to identify several changing proteins, including NM23 and HSP27. However, several limitations are observed, as these studies have primarily focused on transformed AML cell lines, and in total transcript/protein expression profiles.

1.4 ZNF217

In Chapter 3, I identified zinc finger protein 217 (ZNF217) as a significantly upregulated gene at the transcriptional level in RUNX1-ETO expressing CD34+ HSPC (Chapter 3). ZNF217 belongs to the Kruppel-like family of TFs, and contains eight C2H2 zinc finger motifs, as well as a proline-rich region (Kallioniemi et al., 1994) (Figure 1.9). The ZNF217 gene is located on chromosome 20q13, seen to be frequently amplified in human tumours (Collins et al., 1998; Tabach et al., 2011). ZNF217 binds to DNA sequences that regulate gene expression (Cowger et al., 2007; Krig et al., 2007), is a component of the HDAC complex (CoREST-HDAC) (Cowger et al., 2007; Thillainadesan et al., 2012; Thillainadesan et al., 2008), and can be found in complexes with the transcriptional co-repressor C-terminal binding protein 1 (CtBP1) (Shi et al., 2003), lysine-specific histone demethylase 1A (LSD1) (Hakimi et al., 2003; Lee et al., 2005; Shi et al., 2004) and KDM5B/JARID1B/ PLU-1, and the methyltransferases G9a and EZH2 (Quinlan et al., 2007; Cowger et al., 2007; Thillainadesan et al., 2012; Thillainadesan et al., 2008; Quinlan et al., 2006) (Figure 1.10A).
Figure 1.9 – Structural representation of the ZNF217 gene and protein

(A) Genomic organisation of the human ZNF217 gene. Black boxes correspond to untranslated regions, whilst green boxes represent coding regions. (B) Schematic representation of the ZNF217 protein and its functional domains. The ZNF217 protein comprises eight C2H2 zinc finger regions and a proline-rich domain, including an NLS motif. Below are represented the main functional domains of ZNF217. Adapted from (Quinlan et al., 2007).
Figure 1.10 – ZNF217 mediated transcriptional repression

(A) Schematic figure representing ZNF217-containing transcriptional repressor complexes identified through Co-IP studies. (B) Four models have been proposed for ZNF217 recruitment and function as a component of transcriptional repressor complexes: (i) ZNF217 binds DNA directly and orients an associated repressor complex; (ii) Another TF recruits ZNF217 and its partners, resulting in transcriptional repression; (iii) ZNF217 cooperates with other TF in the binding and recruitment of corepressors; (iv) ZNF217 binds directly to DNA and represses transcription through intrinsic enzymatic activity. Adapted from (Quinlan et al., 2007).
Even though ZNF217 was initially thought to act as a transcriptional co-repressor (Quinlan et al., 2007), several studies have shown that it can also positively regulate the expression of specific genes (Krig et al., 2007; Thillainadesan et al., 2008; Krig et al., 2010; Vendrell et al., 2012).

The exact mechanism through which ZNF217 is able to repress gene expression is not fully understood; however, it has been speculated that this occurs through multiple mechanisms (Krig et al., 2007) (Figure 1.10B). The first hypothesis is based on the direct binding of ZNF217 to promoters through its zinc finger domains 6 and 7, thus recruiting the CtBP co-repressor complex through direct protein-protein interactions (Krig et al., 2007). The second model relies on the recruitment of other DNA binding factors responsible for mediating the interaction between the ZNF217/CtBP complex and other promotor regions (Krig et al., 2007). It has also been hypothesised that ZNF217 and other TF are recruited to the same promoters and later interact with CtBP to inhibit gene expression (Krig et al., 2007). The last theory relies on ZNF217 repressing gene expression by itself, through an unknown mechanism (Krig et al., 2007). Furthermore, several studies have shown that ZNF217 gene expression levels do not often correlate with the corresponding protein expression, due to the interference of microRNAs (miRNA) (Li et al., 2015; Bai et al., 2014; Szczyrba et al., 2013) and promoter methylation (Renner et al., 2013; Etcheverry et al., 2010).

1.4.1 ZNF217 in haematopoietic development

The exact role of ZNF217 in the haematopoietic process has not been described; however, some studies have examined the role of ZNF217 in the epithelial-mesenchymal transition (EMT). This occurs during normal embryonic development and is important for lineage determination. During this process, the cell undergoes multiple biochemical changes that allow it to assume a mesenchymal phenotype, with increased migratory ability, invasiveness and increased apoptotic resistance (Kalluri and Neilson, 2003). However, this process has been shown to be involved in tumour progression with metastatic expansion, through the generation of tumour cells with stem cell properties, which play a significant role in drug resistance associated with cancer treatment (Nieto et al., 2016; Lambert et al., 2017; Moustakas and de Herreros, 2017).

ZNF217 has been reported as a promoter of EMT in human mammary epithelial cells (HMEC) and was found to negatively correlate with the expression of several epithelial
markers, such as E-cadherin, a direct target for ZNF217 (Krig et al., 2007; Vendrell et al., 2012). Furthermore, this process was followed by an increase in the expression of mesenchymal proteins and increase on the mRNA levels of TFs identified as promoters of EMT, such as Snail1/2, Twist1/2 and ZEB1/2 (Vendrell et al., 2012). Additionally, in human breast cancer cells and tumours, the promoters of Snail1 and Snail2 genes were found to be enriched by ZNF217 (Krig et al., 2007). Furthermore, overexpression of ZNF217 in the same model was shown to result in the sustained activation of the Transforming growth factor β (TGF-β) pathway through its binding to the TGFB2 and TGFB3 promoters (Vendrell et al., 2012). Conversely, inhibition of the TGF-β pathway resulted in the reversal of the ZNF217-induced EMT, suggesting that this process incorporates the transcriptional repression of E-cadherin coupled with the constitutive activation of the TGF-β-activated signalling pathway (Vendrell et al., 2012). In late-stage tumours, several studies have described a redirection of TGF-β signalling from suppressing cell proliferation to EMT activation (Pirozzi et al., 2011).

1.4.2 ZNF217 as an oncogene

Due to its localisation, ZNF217 has been extensively studied as a potential oncogene and biomarker of disease. In fact, several studies have shown that ZNF217 has the ability to interfere with multiple processes associated with the ‘hallmarks of cancer’. These include sustained proliferative signals, evasion from growth suppressors, replicative immortality, resistance to cell death, cancer stem cell enrichment, and activation of invasion and metastasis (Hanahan and Weinberg, 2011; Hanahan and Weinberg, 2000). Moreover, several tumours have shown to possess increased copies of the ZNF217 gene (Plevova et al., 2010; Fang et al., 2010; Rahman et al., 2012), in variable frequency according to the type of tumour, and have been linked to a poorer outcome in some studies (Quinlan et al., 2007; Peiró et al., 2002; Rooney et al., 2004; Ginzinger et al., 2000). Even though several roles have been proposed for ZNF217 in the development of different cancers, including breast and ovarian, no relationship has been established between the role of this TF and leukaemogenesis.

1.4.2.1 Proliferation

One of the most fundamental characteristics of cancer cells relates to the cells ability to sustain chronic proliferation. ZNF217 has been described as having a major role in this process through the dysregulation of signals which allow progression through cell cycle and growth,
as well as through the disruption of anti-proliferative signalling. In breast and ovarian cancer cells, overexpression of ZNF217 was shown to promote cell proliferation (Li et al., 2014; Thollet et al., 2010), whilst silencing it led to a reduction in cell growth in prostate, colorectal, ovarian and breast cancer cells (Szczyrba et al., 2013; Li et al., 2015; Rahman et al., 2012; Sun et al., 2008a; Thollet et al., 2010). In vivo studies agreed with these observations, as constitutive expression of ZNF217 in breast and ovarian cancer cells promoted the growth and rate of tumour formation in mice (Vendrell et al., 2012; Li et al., 2014; Thollet et al., 2010; Littlepage et al., 2012). This was associated with a substantial increase in the proportion of cells in S phase (Li et al., 2014), coupled with the abnormal expression of several cyclins and, post-translationally, through the upregulation of Aurora kinase A (Aurora-A) (Thollet et al., 2010). One of the genes thought to play a role in ZNF217-induced proliferation is ErbB3, positively regulated by the TF, both in vivo and in human breast cancer cells (Krig et al., 2010). Furthermore, in normal HMEC and in breast cancer cell lines, ectopic expression of ZNF217 induced ErbB3 protein overexpression (Krig et al., 2010; Vendrell et al., 2012), as well as its heterodimer ErbB2. Both ErbB proteins function as a ‘oncogenic unit’ and result in the activation of the PI3K/Akt and MAPK survival pathways (Krig et al., 2010; Littlepage et al., 2012).

An additional mechanism for the sustained proliferative ability of ZNF217-overexpressing cells has been proposed, involving the p15\(^{ink4b}\) tumour suppressor, which inhibits cell cycle progression at the G\(_1\)/S transition (Kim and Sharpless, 2006). The ZNF217/CoREST complex is able to inhibit p15\(^{ink4b}\) activity through the recruitment of the methyltransferase DNMT3A, which in turn leads to promoter hypermethylation (Thillainadesan et al., 2012). By preventing the recruitment of co-factors responsible for the demethylation of p15\(^{ink4b}\), overexpression of ZNF217 results in the block of the anti-proliferative TGF-β signalling pathway (Thillainadesan et al., 2012), shown to promote malignancy (Massagué, 2008; Ikushima and Miyazono, 2010).

1.4.2.2 Replicative immortality

One of the first studies that supported the hypothesis that ZNF217 may function as an oncogene relied on the transduction of finite life-span HMEC with the ZNF217 gene, giving rise to an immortalised cell line, with increased telomerase activity and resistance to TGF-β-induced anti-proliferative action (Nonet et al., 2001). Subsequent studies demonstrated that, in an SV40 Tag/tag expressing, p53/pRB-deficient, human ovarian surface epithelial cell line
(IOSE), ZNF217 overexpression promoted neoplastic progression, associated with high telomerase activity and genomic changes linked to ovarian carcinomas (Li et al., 2007). The oncogenic translation elongation factor EEF1A2 has been shown to be upregulated in ZNF217-overexpressing IOSE cells (Li et al., 2007; Abbas et al., 2015), and its overexpression in the same cells was shown to promote neoplastic progression (Sun et al., 2008b). Additionally, both immortalised HMEC and IOSE cells showed a stabilization of telomere length, with increased activity, features associated with bypassing senescence (Nonet et al., 2001; Li et al., 2007). By overexpressing ZNF217, cells exhibited a decrease in the normal apoptotic rate, as a consequence of functionally compromised telomeres (Huang et al., 2005). These observations suggest that ZNF217 acts by promoting the de-repression of telomerases, thus allowing premalignant cells to overcome senescence due to telomere dysfunction (Nonet et al., 2001).

### 1.4.2.3 Drug resistance

Common therapeutic treatment options often include the combination of different drugs for enhanced effect. These can include doxorubicin, which inhibits topoisomerase II and induces double-strand DNA breaks, leading to ATM-dependent p53-mediated apoptosis, or taxanes, such as paclitaxel, which lead to cell cycle arrest and apoptosis through microtubule-stabilising processes. Ectopic expression of ZNF217 has been shown to contribute to both doxorubicin and paclitaxel resistance in breast cancer cells, whilst silencing it resulted in the opposite effect, with increased drug sensitivity (Thollet et al., 2010; Huang et al., 2005). This is thought to occur through the interference of apoptotic signals induced by the drugs, due to ZNF217 expression (Thollet et al., 2010; Huang et al., 2005). In fact, ZNF217-mediated resistance to paclitaxel in breast cancer cells resulted in significant changes in intrinsic mitochondrial apoptotic pathways, associated with the dysregulation of the anti-apoptotic proteins BCL-2 and BCL-xL, and the pro-apoptotic proteins Bad, Bak and Bax (Thollet et al., 2010). Moreover, the anti-apoptotic potential of the ZNF217 protein has also been found to be mediated through the PI3K/Akt pathway, as ectopic expression of this protein led to the activation of the pathway, whilst silencing ZNF217 resulted in decreased Akt phosphorylation (Krig et al., 2010; Huang et al., 2005).

### 1.4.2.4 Cell differentiation

Identically to LSCs, cancer stem cells (CSCs) possess multilineage differentiation ability, as well as an increased self-renewal potential, and are thought to be the promoters of
tumorigeneses and metastases (Nguyen et al., 2012). Several studies have shown that ZNF217 binds to the promoter region of several genes involved in the differentiation process and organ development, resulting in a block in cell differentiation, as well as CSC maintenance (Krig et al., 2007; Vendrell et al., 2012; Littlepage et al., 2012). In primary mammary epithelial cells, as well as in breast cancer cells, overexpression of ZNF217 led to an increase in the formation of mammospheres with self-renewal potential (Vendrell et al., 2012; Littlepage et al., 2012; Nguyen et al., 2014). Moreover, this was associated with a repression of adult stem cell expression signature, frequently found downregulated in cancer (Littlepage et al., 2012). Lastly, following treatment with retinoic acid, a pro-differentiation agent, ZNF217 expression was repressed in embryonal cells, suggesting that, in differentiated adult cells, aberrant expression of ZNF217 might suppress differentiation, thus leading to tumorigenesis (Krig et al., 2007).

Glioma stem cells (GSCs) are thought to be the driving population leading to the development of glioblastoma multiforme, the most common and aggressive type of primary brain tumour (Singh et al., 2004). These cells have been shown to significantly upregulate ZNF217 expression, as compared to non-GSCs (Mao et al., 2011). Moreover, downregulation of ZNF217 was observed upon GSCs forced differentiation, whilst knocking it down in these cells led to a decrease in cell proliferation and reduction in the stem-like cell population (Mao et al., 2011). A possible mechanism for this is through the involvement of Aurora-A, suggested to regulate self-renewal potential through the stabilization and activation of Wnt/β-catenin signalling (Xia et al., 2013). Additionally, the GSC niche requires hypoxic conditions, and ZNF217 has been demonstrated to be regulated by several HIF both through direct and indirect mechanisms (Mao et al., 2011). These observations suggest that ZNF217 may promote the hypoxia-indices stemness observed in GSCs through its role as a downstream target of HIF.

1.5 C/EBPβ

In Chapter 3, I identified the C/EBPβ protein as a significantly downregulated protein in the nuclear compartment of RUNX1-ETO expressing CD34+ HSPC (Chapter 3). The C/EBP family comprises six structurally and functionally homologous TFs (C/EBPα, C/EBPβ, C/EBPδ, C/EBPγ, C/EBPε, and CHOP) responsible for regulating several physiological processes, including cell proliferation, differentiation, inflammation and metabolism, as well
as oncogene-induced senescence and tumorigenesis (Cao et al., 1991; Diehl, 1998; Poli, 1998; Zahnow, 2002; Ramji and Foka, 2002; Sebastian and Johnson, 2006). All members of this family share a highly conserved C-terminal leucine-zipper dimerization domain (bZIP) involved in DNA binding (Ramji and Foka, 2002), whilst the N-terminal region is less conserved, except for three activation domains, common to all C/EBP family members (Kowenz-Leutz et al., 1994; Williams et al., 1995; Angerer et al., 1999; Williamson et al., 1998), responsible for promoting interactions with transcriptional co-activators (Mink et al., 1997).

The first protein to be identified was C/EBPα, isolated from rat liver and suggested to play an essential role in adipocyte differentiation (Johnson et al., 1987). With time, other C/EBP proteins were identified and attributed to multiple processes. C/EBPβ was first identified in 1990 and was initially described as a bZIP structured factor binding to the IL-1-responsive element within the IL-6 promoter (Akira et al., 1990). C/EBPβ is highly expressed across several tissues, including liver, lung, spleen and kidneys, as well as in myelomonocytic cells and macrophages (Williams et al., 1991; Katz et al., 1993; Haas et al., 2010; Gutsch et al., 2011). Several external factors can contribute to the activation or inhibition of C/EBPβ, including differentiation- or proliferation-inducing agents, hormones, cytokines and inflammatory molecules, through the activation of specific pathways (Ramji and Foka, 2002). Moreover, several additional mechanisms can contribute to C/EBPβ’s expression, including transcriptional processes, mTOR-mediated alternative translation, post-translational modifications, and protein–protein interactions (Zahnow, 2009; Nerlov, 2007; Tsukada et al., 2011). The expression of cytokines or chemokines (and their respective receptors), pro-inflammatory genes, pro-proliferative or differentiation-related genes, as well as metabolic enzymes are regulated through the activation of C/EBPβ (Ramji and Foka, 2002). Subsequently, several cellular functions are influenced by C/EBPβ expression, including proliferation (Haas et al., 2010; Gutsch et al., 2011), differentiation (Katz et al., 1993; Gutsch et al., 2011; Pham et al., 2007), metabolic regulation (Liu et al., 1999; Croniger et al., 2001), and orchestration of the immune response (Poli, 1998; Zwergal et al., 2006; Cappello et al., 2009). Additionally, C/EBPβ is also involved in the pathogenesis of several diseases, such as cancer hyper- or hypo-inflammation, and bacterial/viral infections (Ramji and Foka, 2002; Liu et al., 2009).
1.5.1 Structure and function

*CEBPB* mRNA expression is regulated by several TF, through its TATA box and different binding sites present in its promoter region, including C/EBPβ itself (Chang *et al.*, 1995; Niehof *et al.*, 2001a; Foka *et al.*, 2001), STAT3 (Niehof *et al.*, 2001b), SP1 (Berrier *et al.*, 1998), members of the cAMP responsive element binding protein (CREB)/activating TF family (Berrier *et al.*, 1998; Niehof *et al.*, 1997), early growth response 2 (EGR2) (Chen *et al.*, 2005), Fos related antigen 2 (Fra-2) (Chang *et al.*, 2004), sterol-regulatory element-binding protein 1c (SREBP1c) (Le Lay *et al.*, 2002), myoblastosis TF (Myb) (Mink *et al.*, 1999), and RARα (Duprez *et al.*, 2003). *CEBPB* is an introneless gene, consisting of one single exon and located in chromosome 20q13.13, coding for the production of a single mRNA molecule (Calkhoven *et al.*, 2000; Xiong *et al.*, 2001) (**Figure 1.11A**). However, due to the presence of three initiation codons on different sites (Calkhoven *et al.*, 2000; Duprez *et al.*, 2003), it can generate three protein isoforms: liver-enriched activating protein* (LAP* or full-length [FL]), liver-enriched activating protein (LAP) and liver-enriched inhibitory protein (LIP) (Zahnow, 2009; Sears and Sealy, 1994; Condamine *et al.*, 2015) (**Figure 1.11B**). LAP* is a 44 kDa protein consisting of 345 amino acids; LAP is slightly shorter, with 42 kDa and 322 amino acids, whilst LIP represents a truncated version of C/EBPβ, with 20 kDa and 145 amino acids (Zahnow, 2009; Sears and Sealy, 1994).

As described above (1.5), the C/EBPβ protein structure is characterised by an N-terminal region containing up to three transactivation domains (TAD) and two regulatory domains (RD1 and RD2) (Ramji and Foka, 2002; Williams *et al.*, 1995). On the C-terminal region, C/EBPβ contains a basic DNA-binding domain adjacent to the bZIP dimerization domain (Ramji and Foka, 2002; Zahnow, 2009; Lekstrom-Himes and Xanthopoulos, 1998; Wedel and Ziegler-Heitbrock, 1995; Tsukada *et al.*, 2011). Both LAP* and LAP contain all three TAD and both regulatory domains, whilst LIP lacks all TAD regions, possesses a full-length RD2 and a truncated form of RD1 (Ramji and Foka, 2002; Williams *et al.*, 1995). Several factors influence dimerization, cellular localization, DNA binding, and the transactivation activity of C/EBPβ, such as post-translational modifications, phosphorylation, acetylation, methylation, sumoylation, and proteolysis (Ramji and Foka, 2002; Zahnow, 2009; Tsukada *et al.*, 2011; Hattori *et al.*, 2003).
Figure 1.11 – Structural representation of the C/EBPβ gene and protein isoforms

(A) Genomic organisation of the human CEBPB gene. A cis-regulatory uORF in the single exon CEBPB transcript regulates initiation of translation at conserved in-frame start sites to generate three C/EBPβ isoforms. (B) Schematic representation of the three C/EBPβ protein isoforms and their functional domains. LAP* and LAP isoforms contain three transactivation domains, as well as two regulatory regions RD1 and RD2. LIP, on the other hand, lacks the transactivation domain, and only contains the regulatory region RD2. All three isoforms possess a DNA binding domain, couples with a leucine zipper region, making up the bZIP domain. Adapted from (Zahnow, 2002).
Moreover, the different C/EBPβ isoforms themselves can regulate the protein’s transactivation ability, through the presence (LAP* and LAP) or absence (LIP) of intrinsic transactivation potential or through the appropriate dimerization/interaction partners (Ramji and Foka, 2002; Zahnow, 2009; Tsukada et al., 2011). Furthermore, the ratio of LIP to LAP*/LAP is an important factor in C/EBPβ function (Descombes and Schibler, 1991; Ossipow et al., 1993; Li et al., 2008) (1.5.3). Additionally, C/EBPβ target gene activation also relies on the composition of target gene promoters and localisation (Ramji and Foka, 2002; Zahnow, 2009; Tsukada et al., 2011).

LAP*, LAP and LIP possess distinctive functions. Regarding ‘pure’ transactivation activity, LAP* is a weaker activator of gene transcription, as compared to LAP, due to the formation of an additional disulphide bond (Williams et al., 1995; Su et al., 2003). However, upon the recruitment of the switch/sucrose nonfermentable’ (SWI/SNF) nucleosome remodelling complex by its N-terminal domain, LAP* promotes not only the activation of genes otherwise silenced, but also the interaction with the myeloid-specific TF c-Myb (Kowenz-Leutz and Leutz, 1999). Through the interaction with other TF and regulators, LAP* represents a more effective (co)regulator of myeloid genes than LAP (Kowenz-Leutz and Leutz, 1999). LIP, on the other hand, is considered a dominant inhibitor of transcriptionally active C/EBPs, including LAP* and LAP (Descombes and Schibler, 1991). However, a recent study has demonstrated the ability of LIP to act compensatively in mice lacking the CEBPB gene, suggesting that the classification of LIP as a dominant inhibitor relies on temporal and cellular contexts, in the adipogenic differentiation process (Bégay et al., 2018). However, the exact mechanism underlying this process remains to be understood.

C/EBPβ further possesses the ability to interact with proteins outside the C/EBP family, thus increasing the accessible DNA binding motifs and changing C/EBP protein functions and specificities (Tsukada et al., 2011). These include other bZIP-structured TF, such as members of the Jun/Fos (Hsu et al., 1994) and the CREB/ATF families (Vallejo et al., 1993), as well as non-bZIP proteins, including nuclear factor-κB (NF-κB) subunits p65 (Zwergal et al., 2006) and p50 (LeClair et al., 1992), glucocorticoid receptor (Savoldi et al., 1997), E2F proteins (Sebastian et al., 2005; Johnson, 2005), c-Myb (Kowenz-Leutz and Leutz, 1999), PRMT4 (Kowenz-Leutz et al., 2010), CBP (Guo et al., 2001), PU.1 (Pham et al., 2007; Tissières et al., 2009), death-associated protein 6 (Daxx) (Wethkamp and Klempnauer, 2009), in addition to CBF and RUNX proteins (Tahirov et al., 2001).
1.5.2 C/EBPβ in haematopoietic development

C/EBPβ is highly expressed in myelomonocytic cells, as well as monocytes and macrophages (1.5). In these cells, C/EBPβ controls several processes such as proliferation and differentiation through the activation or repression of specific target genes. In myelomonocytic cells, C/EBPβ is most importantly responsible for regulating proliferation and cell cycle progression (Friedman, 2007; Sato et al., 2020). In HSCs transduced with a FLT3-wild type receptor, FLT3 ligand induced the expression of the truncated C/EBPβ protein LIP, thus leading to a decrease in the LAP/LIP ratio, associated with proliferation (Haas et al., 2010). Furthermore, LIP-deficient murine embryonic fibroblasts show a significant reduction in cell duplication and a decrease in the expression of the proliferative marker’s cyclins A1, A2, B1, E1 and E2 (Wethmar et al., 2010). The consequences arising from the absence of C/EBPβ on the proliferative process, however, are time-dependent, as cells in early developmental stages respond differently to more differentiated cells. For instance, absence of C/EBPβ in murine BM-derived progenitor cells, led to a significant decrease in the number of myeloid colonies, grown in the presence of G-CSF, GM-CSF and IL-3, and in the number of generated cells, as compared to control cells (Hirai et al., 2006). In more differentiated cells from the same model, on the other hand, C/EBPβ KO increased cell proliferation; in vivo, C/EBPβ-KO mice presented hyperplastic haematopoiesis and hypermyeloproliferation (Screpanti et al., 1995) whilst in vitro studies showed that macrophages lacking the CEBPB gene display an increased proliferative rate, as well as enhanced proportion of cells exhibiting S or G2/M cell cycle markers (Gutsch et al., 2011). Although usually regarded as transcriptional activators, LAP* and LAP are thought to be the mediators of this process by promoting the repression of c-Myc in monocytes (Gutsch et al., 2011; Zhang et al., 2011a). By doing this, C/EBPβ represses the transcription of its pro-proliferative target genes, such as cyclin D, whilst promoting the expression of cell cycle inhibitors, such as p27, usually repressed by c-Myc (Gutsch et al., 2011). Subsequent studies have suggested that C/EBPβ interacts with the RB-E2F protein complex, leading to RB-dependent cell cycle arrest and repression of E2F target genes (Sebastian et al., 2005; Johnson, 2005). In summary, in differentiated monocytic cells, the predominant expression of the LAP* and/or LAP results in an arrest in cell cycle, specifically in the G0/G1 phase, and a reduced proliferative ability (Gutsch et al., 2011).
C/EBPβ isoforms have previously been shown to contribute to monocytic differentiation (Friedman, 2007). This process is characterised by both a significant upregulation of the LAP* and LAP isoforms (Katz et al., 1993; Gutsch et al., 2011; Pham et al., 2007; Natsuka et al., 1992; Pan et al., 1999; Ji and Studzinski, 2004; Zhang et al., 2011a) and an increase in the LAP/LIP ratio (Gutsch et al., 2011) is also observed in the differentiation process of other cells, including hepatocytes (Descombes and Schibler, 1991; Buck et al., 1994) and adipocytes (Zhu et al., 2007; Vigilanza et al., 2011). Additionally, C/EBPβ is involved in the regulation of the expression of several differentiation-associated genes, including CD14 (Pan et al., 1999; Ji and Studzinski, 2004; Xu et al., 2008), Fcγ receptor II (FcγRII) (Gorgoni et al., 2002), monocyte-specific esterase (MSE) (Ji and Studzinski, 2004), 1α-hydroxylase (Stoffels et al., 2006), and the cytoplasmic proline-rich tyrosine kinase 2 (Pyk2) (Park et al., 2008). Moreover, C/EBPβ has been shown to interact with PU.1 as a transcriptional cofactor (Pham et al., 2007; Tissières et al., 2009). At specific time-points, LAP* and LAP have the potential to promote c-Myc-mediated transcription by facilitating chromatic opening, through the recruitment of histone acetylating cofactors of the p300/CBP family (by LAP* and LAP) and/or the SWI/SNF chromatin remodelling complex (exclusively by LAP*) (Kowenz-Leutz et al., 2010; Plachetka et al., 2008). As a consequence of the activation of differentiation-associated genes, there is a decrease in the cells’ proliferative rate, coupled with morphological changes (Gutsch et al., 2011; Zhang et al., 2011a) and increased antimicrobial ability (Zhang et al., 2011a). Even though C/EBPβ-KO mice have the ability to give rise to macrophage-like cells, these showed a reduced functional potential, since their response to external stimuli is impaired (Screpanti et al., 1995; Tanaka et al., 1995a). Subsequently, macrophages derived from C/EBPβ-KO mice shown a reduction in the expression of differentiation markers (Koschmieder et al., 2009), as well as atypical morphological characteristics (Gutsch et al., 2011). Moreover, C/EBPβ has been suggested to be required for the survival of peripheral blood monocytes, whilst it’s not necessary for the development and maintenance of mature cells of other lineages at steady state (Tamura et al., 2015a).

In addition to monocytic development, LAP* and LAP have been shown to promote granulocytic differentiation (Hirai et al., 2006; Popernack et al., 2001). In murine primary BM cells, C/EBPβ overexpression resulted in a reduction in the proportion of myeloid progenitor cells and promoted granulocytic differentiation (Popernack et al., 2001). Moreover, in response to external stimuli, C/EBPβ has the ability to promote both proliferation and differentiation of
HSPC to supply granulocytes in demand (Satake et al., 2012). In the absence of C/EBPβ, however, mice infected with C. albicans showed a reduced response in terms of emergency granulopoiesis (Hirai et al., 2006). Moreover, in patients with congenital neutropenia, addition of G-CSF led to a significant upregulation of C/EBPβ in BM-derived myeloid cells, resulting in urgent granulocytic development (Skokowa and Welte, 2009). Osteoclast formation, on the other hand, is regulated by LIP, but inhibited by LAP* or LAP, through the regulation of the osteoclastogenesis inhibitor MafB (Smink et al., 2009; Smink et al., 2012). Similarly, uORF-deficient mice incapable of inducing LIP expression display increased levels of MafB, resulting in impaired osteoclastogenesis (Wethmar et al., 2010). Recent studies have shown that C/EBPβ is upregulated in HSPC as a response to stress (Sato et al., 2020), and this is essential for the cells to regenerate, thus meeting the increasing myeloid cell demand. Furthermore, the expression of the different isoforms is dependent on cell state: first, upregulation of LIP in an early regenerative stage allows the proliferation of LT-HSC and their differentiation into MPP, whilst LAP*/LAP induce subsequent myeloid differentiation at a later stage of regeneration (Sato et al., 2020).

1.5.3 C/EBPβ as an oncogene

Due to their ability to block cell growth and respond to DNA damage, C/EBP proteins are generally identified as tumour suppressor factors. However, upon certain stimuli and depending on cell type and isoform present, these can elicit opposite effects. C/EBPβ function highly resembles that of C/EBPα, since it has been shown to promote cell differentiation, suppress tumorigenesis and block proliferation. When not required, C/EBPβ is maintained in a latent state through the action of auto-inhibitory factors aimed at suppressing its DNA binding and transactivation functions (Williams et al., 1995; Kowenz-Leutz et al., 1994).

1.5.3.1 Cell survival

The role of C/EBPβ in cancer is thought to be mediated through its role in regulating cell survival and apoptosis. For instance, upon induced DNA damage, C/EBPβ plays an essential role in the survival process of hepatic stellate cells (Buck et al., 2001) and in macrophages, C/EBPβ is required for survival in response to Myc-Raf transformation (Wessells et al., 2004). Moreover, C/EBPβ-induced cell survival in response to DNA damage has been shown to be
mediated through a reduction in p53 expression and activity (Yoon et al., 2007; Ewing et al., 2008).

In addition to its involvement in apoptosis, several studies have showed that C/EBPβ has a role in oncogene-induced senescence (OIS). Under normal circumstances, senescence in characterised by a state of irreversible growth arrest that can act as a barrier to malignant transformation. Forced expression of C/EBPβ LAP has been shown to result in cell cycle arrest in hepatocarcinoma cells (Buck et al., 1994), keratinocytes (Zhu et al., 1999) and fibroblasts (Johnson, 2005). Moreover, cooperation between C/EBPβ and RB-E2F proteins has been shown to result in an irreversible cell cycle arrest at the G1/S phase, through Ras-induced cellular senescence (Sebastian et al., 2005). Lastly, several C/EBPβ-induced pro-inflammatory cytokines and chemokines have been shown to contribute to OIS (Acosta et al., 2008; Kuilman et al., 2008). An additional association between Ras and C/EBPβ was described as studies showed that in the context of C/EBPβ deficiency, v-HA-Ras transgenic mice showed reduced skin-tumorigenic potential, suggesting that C/EBPβ plays an oncogenic role in the downstream Ras signalling pathway (Poli, 1998). Moreover, gene expression analysis in human cancers suggested that C/EBPβ is involved in cyclin D1-induced oncogenic signature (Zahnow et al., 1997). Increased expression of C/EBPβ, and its association with oncogenic development has been reported in several cancer types, including breast, ovarian, colorectal renal and gastric tumours. However, the exact role of C/EBPβ in the regulation of survival, apoptosis and senescence is highly context specific.

1.5.3.2 Tumour aggressiveness

In general, few mutations have been identified in C/EBPβ, and those identified are not thought to contribute to the development of epithelial cancers (Vegesna et al., 2002). However, the region in which the C/EBPβ gene is found has been shown to be amplified in a small proportion of human cancers, and it has been associated with lobular carcinoma in situ of the breast (Mastracci et al., 2006). An increase in CEBPB mRNA levels results in an increase in protein translation, associated with a rise in C/EBPβ isoform expression and the imbalance of the LAP/LIP ratio, with the upregulation of the LIP isoform, observed in oestrogen-receptor-negative, aneuploid, highly proliferative breast tumours, with poor prognosis (Zahnow et al., 1997; Milde-Langosch et al., 2003). Dysregulation in the normal LAP/LIP balance has also been related to a defective TGF-β-dependent cytostatic response in metastatic breast cancer
However, forced expression of LAP*/LAP, thus normalising the LAP/LIP ratio was able to restore TGF-β cytostatic response and reduce the proliferative ability of metastatic cells (Gomis et al., 2006).

The increase in CEBPB levels has been observed in a more-aggressive subset of tumours, associated with metastatic breast cancer (van de Vijver et al., 2002), a high tumour grade (van ’t Veer et al., 2002; Ma et al., 2004; Finak et al., 2008) and overall poorer prognosis (van de Vijver et al., 2002), as compared to less-aggressive tumours. Additionally, the expression of C/EBPβ was found to be heightened in malignant ovarian tumours, as compared to borderline or benign tumours (Sundfeldt et al., 1999). These observations suggest that the transcriptional control or regulation of mRNA stability may influence CEBPB expression in more aggressive tumour types.

1.5.3.3 Drug resistance

C/EBPβ has been suggested to contribute to multi-drug resistance in breast cancer (Combates et al., 1994; Conze et al., 2001; Chen et al., 2004), through the P-glycoprotein transporter, encoded by MDR1, further associated with poor prognosis (Leonessa and Clarke, 2003). Further, the MDR1 gene is regulated by C/EBPβ in HepG2 hepatoma cells (Combates et al., 1994) and in MCF7 breast cancer cells (Conze et al., 2001). Similar observations were made in ovarian cancer, in which C/EBPβ was shown to promote the expression of drug resistance genes (Liu et al., 2018), through its interaction with the telomeric silencing 1-like (DOT1L) protein. Inhibition or KD of DOT1L led to a decrease in the promotion of the expression of drug resistance genes by C/EBPβ, and actually reversed cisplatin resistance (Liu et al., 2018). These observations suggest that expression of C/EBPβ influences treatment outcome and the development of drug-resistance events.

1.5.4 C/EBPβ in leukaemogenesis

Of all the members of the C/EBP family, the role of C/EBPα in the leukaemogenic process has by far been the most extensively studied (Nerlov, 2004; Koschmieder et al., 2009). Nonetheless, C/EBPβ has been shown to be related to the development of several myelomonocytic leukaemias (Wetmar et al., 2010; Sebastian and Johnson, 2006; Duprez, 2004). C/EBPβ has been shown to be dysregulated in FLT3-ITD-mutated AML patients, with an increased concentration of LIP observed in these patients (Haas et al., 2010). Accordingly,
an increased LIP expression has been shown in several ITD-positive monocytic cell lines, including MV4;11, MOLM-13 and PL21 (Haas et al., 2010). Furthermore, LIP is also highly expressed in the erythroleukaemia-like cell line, K562 (Wall et al., 1996), as well as in the anaplastic large cell lymphoma cell lines SUDHL-1, Ki-JK, and Karpas 299 (Quintanilla-Martinez et al., 2006). LAP* and LAP, on the other hand, are generally reduced in rapidly proliferating cells, such as in BM-derived cells from CML patients in blast crisis (Guerzoni et al., 2006).

Upregulation of LAP* or LAP has been shown to reduce cell proliferation and induce the differentiation of leukaemic cell lines, and supressing leukaemogenesis in in vivo models (Guerzoni et al., 2006). In the NB4 AML cell line, upregulation of LAP*/LAP following ATRA treatment resulted in the differentiation of less developed myelomonocytic cells into granulocytes (Duprez et al., 2003). However, this was not observed in ATRA-resistant cells, in which it was not possible induce the expression of any C/EBPβ isoform (Duprez et al., 2003). Moreover, treatment of primary AML cells with deltanoids (Studzinski et al., 2005), U937 cells with vitamin D3 (Koschmieder et al., 2007) or ATRA (Chen et al., 2009), and HL-60 cells with 1,25-Dihydroxyvitamin D3 [1,25(OH)2D3] (Studzinski et al., 2005) also resulted in an increase in C/EBPβ protein levels associated with the differentiation of these cells into monocytes. Similarly, expression of C/EBPβ in the BCR-ABL cell line 32D was found to be repressed, as this fusion protein negatively regulates C/EBPβ (Guerzoni et al., 2006). However, this was restored upon treatment with imatinib, following which C/EBPβ was able to induce differentiation and inhibit the proliferation of these cells (Guerzoni et al., 2006). Subsequently, treatment of AML patients with vitamin D3 or derivates might represent an alternative therapeutical approach for patients that do not respond to classical chemotherapeutic agents, through the induction of C/EBPβ-supported cell differentiation (Hughes et al., 2010). However, the role in C/EBPβ in the development of t(8;21) leukaemia has not been determined.

1.6 Aims and objectives

The t(8;21) translocation, resulting in the expression of the fusion protein RUNX1-ETO, is a common chromosomal translocation in AML. Even though it is associated with a favourable prognosis, the exact mechanism underlying the leukaemogenic process remains unknown.
Identifying new therapeutic targets is therefore important if we are to progress the treatment of patients with AML.

Previous studies performed within our group have shown that RUNX1-ETO is able to inhibit the development of haematopoietic cells, promote the growth of immature cells and increase the cells’ self-renewal ability (Tonks et al., 2003; Tonks et al., 2004). Transcriptomic analysis of these cells identified hundreds of genes impacted by RUNX1-ETO expression (Tonks et al., 2007). However, this analysis was performed using an unsupervised approach, focusing on the highest changes observed in RUNX1-ETO expressing cells. Given that transcriptional regulation is likely mediated by changes in TF expression, this study will re-analyse transcriptomic data obtained from Tonks et. al., and focus specifically on changes observed in this gene class (Tonks et al., 2007). Further, even though transcriptomic analysis has proved to be an essential strategy for target identification, it is not a powerful predictor of protein expression. These approaches are often combined with alternative technologies, aiming at analysing the cells’ proteomic profile. In addition, no study has so far quantified the cytosolic and nuclear protein expression profile of CD34+ HSPC expressing RUNX1-ETO. I hypothesize that the early differentiation block observed upon the expression of RUNX1-ETO is mediated by changes in TF expression, and that by identifying these, it will be possible to reverse the phenotype and promote normal cell development (Figure 1.12). This hypothesis will be investigated through the following objectives:

**To analyse existing transcriptomic data to identify novel changes in TF mRNA expression arising from RUNX1-ETO expression**

Pre-existing transcriptomic data relating to RUNX1-ETO-induced changes in gene expression in human CD34+ HSPC will be analysed to identify changes in TF mRNA expression (Chapter 3).

**To quantify changes in protein expression in human CD34+ cord-blood derived HSPC expressing RUNX1-ETO**

Cytosolic and nuclear protein fractions will be extracted from human RUNX1-ETO-expressing and control CD34+ HSPC and quantitative proteomic analysis will be performed,
using Sequential Window Acquisition of all Theoretical Mass Spectra (SWATH-MS), further allowing the identification of a target of interest for subsequent studies (Chapter 3).

**Determine the effects of ZNF217 overexpression or KD on normal human myeloid development**

Functional analysis will be performed using overexpression and KD (shRNA) vectors to determine the role of ZNF217 in normal human haematopoiesis. Specifically, I will determine whether ZNF217 disrupts myeloid cell proliferation and differentiation. Furthermore, the effect of ZNF217 on proliferation and apoptosis in AML cell lines will be assessed using flow cytometry (Chapter 4).

**Determine the effects of C/EBPβ overexpression or KD on normal human myeloid development**

Functional analysis will be performed using overexpression and KD (shRNA) vectors to determine the role of C/EBPβ in normal human haematopoiesis. Specifically, I will determine whether C/EBPβ disrupts myeloid cell proliferation and differentiation. Furthermore, the effect of C/EBPβ on proliferation and apoptosis in AML cell lines will be assessed using flow cytometry (Chapter 5).
This study hypothesizes that \textit{RUNX}-\textit{ETO} induces changes in TF expression, which subsequently lead to a block in normal cell differentiation. By identifying these TF, it will be possible to modulate their expression and promote normal haematopoietic development. These changes will be evaluated by analysing previously generated transcriptomic data, as well as newly performed proteomic analysis.
Chapter 2

Materials and Methods
2.1 Molecular Biology

2.1.1 Plasmids used in the study

Retroviruses, along with lentiviruses, are RNA viruses capable of infecting eukaryotic cells and are widely used to deliver nucleic acids to many cell types in a variety of experimental systems. They do this by using the reverse transcriptase enzyme to produce DNA within the host. This DNA is incorporated into the host’s genome using an integrase enzyme and is then replicated with the host’s DNA. A key feature of both viral vectors is that they produce replication-defective, or self-inactivating, particles, allowing the delivery of the desired sequence, without continued viral replication in the target cells (Cooray et al., 2012). Whereas transfection of nucleic acids results only in transient transgene expression, the activity of the viral integrase in retroviral and lentiviral-based systems allows for stable integration of the transgene which is then inherited and continuously expressed over repeated cell divisions. Lentiviruses, unlike retroviruses, are the only retroviral family member that can integrate in non-dividing cells, making them one of the most efficient gene delivery systems (Dufait et al., 2012). The constructs used in the project are outlined in Table 2.1 and Figure 2.1.

2.1.2 Bacterial growth medium and selective agar plates

Selective Luria-Bertani (LB) broth media was prepared by dissolving 20g of LB broth (ThermoFisher Scientific, Loughborough, UK) in 1 L of deionised water (dH₂O) and autoclaved. LB Agar plates were prepared by adding 35g of LB agar (Sigma-Aldrich, Poole, UK) to 1L of dH₂O, and autoclaved. Medium were supplemented with 100 μg/mL ampicillin (Sigma-Aldrich) for selective growth.

Previously stored glycerol stocks of transformed E.coli cells were spread over ampicillin selective LB agar plates and incubated overnight at 37°C. A single colony was aseptically picked and grown in 5 mL LB broth supplemented with ampicillin (100 μg/mL) at 37°C on a rotary shaker at 225 revolutions per minute (rpm) for 5-6 h. The culture was subsequently diluted by transferring 150 μL to 150 mL of LB broth supplemented with ampicillin (100 μg/mL) and cultured overnight at 37°C, on a rotary shaker (225 rpm). Prior to plasmid DNA isolation and quantification (2.1.3), 850 μL of inoculated broth were added to 150 μL of glycerol and stored at -80 °C.
<table>
<thead>
<tr>
<th>Target</th>
<th>#</th>
<th>Plasmid</th>
<th>Studies</th>
<th>Vector Name</th>
<th>Target sequence</th>
<th>Match region</th>
<th>Marker(s)</th>
<th>Host bacteria</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>99</td>
<td>Pinco</td>
<td>Overexpression</td>
<td>Control</td>
<td>n/a</td>
<td>n/a</td>
<td>GFP</td>
<td>Stbl3</td>
<td>(ThermoFisher) (Tonks et al., 2003; Grignani et al., 1998)</td>
</tr>
<tr>
<td></td>
<td>407</td>
<td>pLV</td>
<td>Overexpression</td>
<td>Control</td>
<td>n/a</td>
<td>n/a</td>
<td>Puromycin;</td>
<td>Stbl3</td>
<td>VectorBuilder</td>
</tr>
<tr>
<td></td>
<td>497</td>
<td>pLV</td>
<td>Knockdown</td>
<td>shRNA Control</td>
<td>Scrambled</td>
<td>n/a</td>
<td>Puromycin;</td>
<td>DH5α</td>
<td>VectorBuilder</td>
</tr>
<tr>
<td></td>
<td>530</td>
<td>pLV</td>
<td>Overexpression</td>
<td>Control</td>
<td>n/a</td>
<td>n/a</td>
<td>Neomycin;</td>
<td>Stbl3</td>
<td>VectorBuilder</td>
</tr>
<tr>
<td>RUNX1-ETO</td>
<td>1</td>
<td>Pinco</td>
<td>Overexpression</td>
<td>RUNX1-ETO</td>
<td>RUNX1-ETO</td>
<td>n/a</td>
<td>GFP</td>
<td>Stbl3</td>
<td>(Tonks et al., 2003)</td>
</tr>
<tr>
<td></td>
<td>543</td>
<td>pLV</td>
<td>Overexpression</td>
<td>ZNF217 OE</td>
<td>ZNF217 (NM_006526.2)</td>
<td>n/a</td>
<td>GFP</td>
<td>VB UltraStable</td>
<td>VectorBuilder</td>
</tr>
<tr>
<td>ZNF217</td>
<td>552</td>
<td>pLV</td>
<td>Knockdown</td>
<td>shZNF217 #1</td>
<td>TRCN0000364558</td>
<td>CDS</td>
<td>Puromycin;</td>
<td>VB UltraStable</td>
<td>VectorBuilder</td>
</tr>
<tr>
<td></td>
<td>554</td>
<td>pLV</td>
<td>Knockdown</td>
<td>shZNF217 #2</td>
<td>TRCN000013060</td>
<td>CDS</td>
<td>Puromycin;</td>
<td>VB UltraStable</td>
<td>VectorBuilder</td>
</tr>
<tr>
<td></td>
<td>556</td>
<td>pLV</td>
<td>Knockdown</td>
<td>shZNF217 #3</td>
<td>TRCN0000364635</td>
<td>3'UTR</td>
<td>Puromycin;</td>
<td>VB UltraStable</td>
<td>VectorBuilder</td>
</tr>
<tr>
<td></td>
<td>558</td>
<td>pLV</td>
<td>Knockdown</td>
<td>shZNF217 #4</td>
<td>TRCN0000364634</td>
<td>CDS</td>
<td>Puromycin;</td>
<td>VB UltraStable</td>
<td>VectorBuilder</td>
</tr>
</tbody>
</table>
### C/EBPβ

<table>
<thead>
<tr>
<th>#</th>
<th>Vector Type</th>
<th>Overexpression</th>
<th>Gene</th>
<th>Control</th>
<th>Delivery</th>
<th>Selectant</th>
<th>Concentration</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>570</td>
<td>pLV</td>
<td>Overexpression</td>
<td>C/EBPβ OE</td>
<td>n/a</td>
<td>VB UltraStable</td>
<td>Puromycin; GFP</td>
<td>VectorBuilder</td>
<td></td>
</tr>
<tr>
<td>565</td>
<td>pLV</td>
<td>Knockdown</td>
<td>shC/EBPβ #1</td>
<td>TRCN0000007440</td>
<td>3′UTR</td>
<td>VB UltraStable</td>
<td>Puromycin; GFP</td>
<td>VectorBuilder</td>
</tr>
<tr>
<td>565</td>
<td>pLV</td>
<td>Knockdown</td>
<td>shC/EBPβ #2</td>
<td>TRCN0000364533</td>
<td>CDS</td>
<td>VB UltraStable</td>
<td>Puromycin; GFP</td>
<td>VectorBuilder</td>
</tr>
<tr>
<td>567</td>
<td>pLV</td>
<td>Knockdown</td>
<td>shC/EBPβ #3</td>
<td>TRCN0000364495</td>
<td>CDS</td>
<td>VB UltraStable</td>
<td>Puromycin; GFP</td>
<td>VectorBuilder</td>
</tr>
</tbody>
</table>

### Other

<table>
<thead>
<tr>
<th>#</th>
<th>Vector Type</th>
<th>Overexpression</th>
<th>Gene</th>
<th>Control</th>
<th>Delivery</th>
<th>Selectant</th>
<th>Concentration</th>
<th>Origin</th>
</tr>
</thead>
<tbody>
<tr>
<td>158</td>
<td>pMD2</td>
<td>n/a</td>
<td>pMD2</td>
<td>Addgene_12259</td>
<td>n/a</td>
<td>n/a</td>
<td>Sibl3</td>
<td>Addgene</td>
</tr>
<tr>
<td>189</td>
<td>pMD2</td>
<td>n/a</td>
<td>pMD2</td>
<td>Addgene_12260</td>
<td>n/a</td>
<td>n/a</td>
<td>DH5α</td>
<td>Addgene</td>
</tr>
</tbody>
</table>

# indicates laboratory plasmid reference number; 1 DNA sequence with no mammalian target; 2 all plasmids contain an ampicillin resistant gene; n/a - not applicable; shRNA – short hairpin RNA; TRCN – The RNAi Consortium Number; GFP – Green Fluorescent Protein; Plasmids were purchased from VectorBuilder (Guangzhou, China) or Addgene (Massachusetts, USA).
Figure 2.1 – Plasmid DNA constructs used in this study

Representative plasmid maps of (A) PINCO control vector, with a EGFP selectable marker; (B) shRNA with EGFP:T2A:Puro selectable marker and U6 promoter; (C) lentiviral control vector with a EF-1α promotor and EGFP/Neomycin selectable markers; (D) lentiviral control with EGFP:T2A:Puro selectable makers and EF-1α promotor. All vectors were created using SnapGene Software (GSL Biotech LLC, USA).

ΔU3/3' LTR - truncated HIV-1 3’long terminal repeat; Δ5' LTR/5'UTR - truncated HIV-1 5’ long terminal repeat; AmpR - ampicillin resistance gene; CMV enhancer – human cytomegalovirus (CMV) immediate early enhancer; CMV promotor - CMV immediate early promotor; cPPT/CTS - central polypurine tract; EF-1α core promotor - core promotor for human elongation factor EF-1α; EGFP – Green fluorescent protein; EGFP:T2A:Puro - EGFP and Puromycin linked by T2A; HIV-1 Ψ - HIV-1 packaging signal; MMLV Ψ - packaging signal of Moloney murine leukaemia virus (MMLV); NeoR – neomycin resistance gene; PGK promotor - mouse phosphoglycerate kinase 1 promotor; PuroR - puromycin resistance gene; RRE - Rev response element; RSV promotor - Rous sarcoma virus promotor; SV40 poly(A) signal - SV40 polyadenylation signal; SV40 ori - SV40 origin of replication; U6 - RNA polymerase III U6 promotor; WPRE - woodchuck hepatitis virus post-transcriptional regulatory element.
2.1.3 Isolation and quantification of recombinant plasmid DNA from bacteria

Following overnight bacterial growth, plasmid DNA was isolated using the QIAGEN HiSpeed® Plasmid Maxi Kit (QIAGEN, Manchester, UK) according to the manufacturer’s instructions. Briefly, grown bacterial culture was centrifuged at 6,000 xg for 15 min, at 4°C. Following supernatant discard, pellet was resuspended in 10 mL of P1 buffer, containing RNase I (resuspension buffer), followed by the addition of 10 mL of P2 buffer (lysis buffer). The solution was mixed by slowly inverting the tube and incubated at room temperature (RT) for 5 min, before adding 10 mL of P3 buffer (neutralization buffer). Lysate was transferred onto a QIA filter cartridge and incubated at RT for 10 min. In the meantime, a HiSpeed Maxi tip was equilibrated by adding 10 mL of QBT buffer (equilibration buffer). Lysate was added to the column and allowed to pass through the resin by gravity flow, before being washed with 60 mL of QC buffer (wash buffer). Plasmid DNA was eluted by adding 15 mL of QF buffer and precipitated by adding 10.5 mL of isopropanol (Sigma-Aldrich). Following a 5 min incubation, the eluted DNA was passed through a QIA precipitator using a 30 mL syringe and washed with 2 mL of 70% v/v ethanol (Sigma-Aldrich). The plasmid DNA, retained in the precipitator, was air-dried and collected by passing 1 mL of dH2O, through the syringe. The elute was then passed through the precipitator once more to ensure maximum DNA recovery.

Isolated plasmid DNA was quantified using a NanoDrop OneC Spectrophotometer (ThermoFisher Scientific), according to the manufacturer’s instructions. To access DNA purity, the ratio of absorbance at 260/280 nm was used. A ratio between 1.8-2 is considered good quality; proteins absorb at 280 nm, hence lower ratios indicate the presence of high concentration of proteins. This was combined with the analysis of the A260/230 ratio, that accesses sample purity in terms of salts and other contaminants, which absorb at 230 nm. A sample with a ratio within 2-2.2 is considered “pure”. Pure DNA samples were used to generate recombinant virus, as described in 2.4.

2.1.4 DNA Sequencing

To validate DNA sequences, Sanger sequencing was performed by Eurofins Genomics, using the Eurofins Genomic’s SmartSeq kit sequencing service. For this, 50-100 ng of plasmid DNA, eluted in water, was made up to a volume of 15 μL and combined with 10 μM of the desired primer (Table 2.2). Sequences were verified by performing a pairwise sequence
alignment using ClustalW2 (https://www.ebi.ac.uk/Tools/msa/clustalw2/). Vectors purchased from VectorBuilder were not used to perform sequencing, as the company provides this data as a quality control check.

Table 2.2 – Primers used for direct sequencing

Table outlining primer sequence (5’-3’) used for sequencing PINCO vectors above (2.1.4).

<table>
<thead>
<tr>
<th>Primer</th>
<th>Vector</th>
<th>Strand</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>PINCO</td>
<td>Reverse</td>
<td>5’ TTA TGT AAC GCG GAA CTC CA 3’</td>
</tr>
<tr>
<td>P2</td>
<td>PINCO</td>
<td>Forward</td>
<td>5’ TAG AAC CTC GCT GGA AAG GA 3’</td>
</tr>
</tbody>
</table>

2.2 Cell culture

All tissue culture work was performed in a Class II biosafety cabinet and the waste disinfected with bleach and/or autoclaved. Cells were cultured under aseptic conditions in a humidified incubator at 37°C and 5% CO₂, unless otherwise specified.

2.2.1 Culture of cell lines

2.2.1.1 Cell culture of non-adherent cell lines

The origin and specific culture conditions of non-adherent cells used in this study are outlined in Table 2.3. All lines were maintained in the appropriate culture media supplemented with 10% (v/v) heat-inactivated foetal bovine serum (FBS; 77133, Labtech International Ltd, Sussex, UK), 1% (v/v) L-glutamine (Invitrogen, California, USA) and 20 µg/mL Gentamycin (Life Technologies, California, USA), unless otherwise specified. Cells were maintained in log phase growth, at a density of 1-8x10⁵ cells/mL, in the corresponding culture media.

2.2.1.2 Cell culture of adherent cell lines

The characteristics of adherent cells used in this study have been described in Table 2.4. Cell lines were maintained with Dulbecco’s Modified Eagle Medium (DMEM; Merck Life Science, Gillingham, UK) supplemented with 10% (v/v) FBS, 1% (v/v) L-glutamine and 20 µg/mL Gentamycin. Adherent cells were passaged when reaching 90% confluence.
Table 2.3 – Culture conditions for non-adherent cells

All cell lines were cultured in the presence of 10% v/v FBS, 2mM L-Glutamine and 20 µg/mL gentamicin, unless otherwise stated; all cell lines are derived from human patients, unless otherwise stated.

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Description</th>
<th>FAB</th>
<th>Molecular characteristics</th>
<th>Media</th>
<th>Seeding density (cells/mL)</th>
<th>Doubling time</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEL</td>
<td>Erythroleukaemia</td>
<td>M6</td>
<td>JAK2 mutation (V617F)</td>
<td>RPMI 1640</td>
<td>1x10^5 - 1x10^6</td>
<td>36h</td>
<td>ECACC</td>
</tr>
<tr>
<td>HL-60</td>
<td>APL</td>
<td>M2</td>
<td>t(5;17)(p11;q11)</td>
<td>RPMI 1640</td>
<td>5x10^5 – 1.5x10^6</td>
<td>40h</td>
<td>ATCC</td>
</tr>
<tr>
<td>K562</td>
<td>CML</td>
<td>-</td>
<td>t(9;22)(q34;q11) - BCR-ABL1</td>
<td>RPMI 1640</td>
<td>1x10^6</td>
<td>30-40h</td>
<td>ECACC</td>
</tr>
<tr>
<td>Kasumi-1</td>
<td>AML</td>
<td>M2</td>
<td>t(8;21) (q22;q22) – RUNX1-ETO</td>
<td>RPMI 1640; 20% v/v FBS</td>
<td>3x10^5</td>
<td>48-72h</td>
<td>DSMZ</td>
</tr>
<tr>
<td>KG1</td>
<td>Erythroleukaemia</td>
<td>M1</td>
<td>-</td>
<td>RPMI 1640</td>
<td>2x10^5 – 1x10^6</td>
<td>35-40h</td>
<td>ATCC</td>
</tr>
<tr>
<td>MV-4-11</td>
<td>Myelomonocytic Leukaemia</td>
<td>M5</td>
<td>t(4;11)(q21;q23) – KMT2A-AFF1</td>
<td>IMDM</td>
<td>1x10^5 – 1x10^6</td>
<td>50h</td>
<td>ATCC</td>
</tr>
<tr>
<td>NB4</td>
<td>APL</td>
<td>M3</td>
<td>t(15;17)(q22;q11-12.1) – PML-RARA</td>
<td>RPMI 1640</td>
<td>5x10^5 - 1x10^6</td>
<td>35-45h</td>
<td>ATCC</td>
</tr>
<tr>
<td>NOMO-1</td>
<td>AML</td>
<td>M5</td>
<td>t(9;11)(p22;q23) – KMT2A-MLLT3</td>
<td>RPMI 1640</td>
<td>5x10^5</td>
<td>35h</td>
<td>DSMZ</td>
</tr>
<tr>
<td>OCI-AML2</td>
<td>AML</td>
<td>M4</td>
<td>DNMT3A mutation (R635W)</td>
<td>aMEM; 20% v/v FBS</td>
<td>7x10^5 - 1x10^6</td>
<td>40-50h</td>
<td>DSMZ</td>
</tr>
<tr>
<td>OCI-AML5</td>
<td>AML</td>
<td>M4</td>
<td>t(1;19)(p13;p13)</td>
<td>aMEM; 20% v/v FBS; 10 ng/mL GM-CSF</td>
<td>5x10^5 - 1.5x10^6</td>
<td>30-40h</td>
<td>DSMZ</td>
</tr>
<tr>
<td>PLB985</td>
<td>Clone of HL-60</td>
<td></td>
<td></td>
<td>RPMI 1640</td>
<td>5x10^5 – 1x10^6</td>
<td>30h</td>
<td>ATCC</td>
</tr>
<tr>
<td>SKNO-1</td>
<td>AML</td>
<td>M2</td>
<td>t(8;21) (q22;q22) – RUNX1-ETO</td>
<td>RPMI 1640; 10 ng/mL GM-CSF</td>
<td>2x10^5 - 2x10^6</td>
<td>35-50h</td>
<td>DSMZ</td>
</tr>
<tr>
<td>TF1</td>
<td>Erythroleukaemia</td>
<td>M6</td>
<td>-</td>
<td>RPMI 1640</td>
<td>5x10^5 – 1x10^6</td>
<td>70h</td>
<td>ATCC</td>
</tr>
<tr>
<td>THP-1</td>
<td>Acute Monocytic Leukaemia</td>
<td>M5</td>
<td>(9;11)(p21;q23) leading to KMT2A-MLLT3 (MLL-MLLT3; MLL-AF9)</td>
<td>RPMI 1640</td>
<td>2x10^5 - 4x10^5</td>
<td>40-50h</td>
<td>ECACC</td>
</tr>
<tr>
<td>TK6</td>
<td>CML</td>
<td>-</td>
<td>t(9;22)(q34;q11) - BCR-ABL1</td>
<td>RPMI 1640</td>
<td>2x10^5 - 1x10^6</td>
<td>12h</td>
<td>Prof Allan, Newcastle Uni.</td>
</tr>
<tr>
<td>U937</td>
<td>Histiocytic Lymphoma</td>
<td>M5</td>
<td>t(10;11)(p12;q14) – MLLT10-PICALM</td>
<td>RPMI 1640</td>
<td>1x10^5 - 2x10^6</td>
<td>30-40h</td>
<td>ATCC</td>
</tr>
<tr>
<td>------</td>
<td>---------------------</td>
<td>----</td>
<td>--------------------------------</td>
<td>-----------</td>
<td>-----------------</td>
<td>--------</td>
<td>------</td>
</tr>
</tbody>
</table>

αMEM - Minimum Essential Medium Eagle- Alpha Modification; APL – Acute Promyelocytic Leukaemia; ATCC – American Type Culture Collection (USA); CML – Chronic Myeloid Leukaemia; DMEM - Dulbecco’s Modified Eagle’s Medium; DSMZ - German Collection of Microorganisms and Cell Cultures (Germany); ECACC - European Collection of Authenticated Cell Cultures (UK); FBS – Heat-Inactivated Foetal Bovine Serum (Labtech, East Sussex, UK); Gent – Gentamycin (Gibco, ThermoFisher Scientific); IMDM - Iscove Modified Dulbecco’s Medium; L-Glu – L-Glutamine (Gibco, ThermoFisher Scientific); RPMI - Roswell Park Memorial Institute-1640 Medium.
Confluent cell monolayers grown were detached from the flask by incubation with 3 mL (F25) or 5 mL (F75) of pre-warmed trypsin (500 μg/mL; Fisher Scientific) at RT for 3 min. Trypsin was neutralized by adding an equal volume of media, following which cells were disaggregated from the wall of the flask and transferred to a universal container (UC). Following centrifugation at 270 xg for 10 min, supernatant was discarded, and cell pellets resuspended in 20 mL of culture media. Cells were counted, as described in 2.2.2, and seeded at the density required for subsequent assays.

Table 2.4 – Culture conditions for adherent cells

All cell lines were cultured in DMEM, in the presence of 10% (v/v) FBS, 1% (v/v) L-Glutamine and 20 μg/mL gentamicin.

<table>
<thead>
<tr>
<th>Cell line</th>
<th>Description</th>
<th>Seeding Density (cells/mL)</th>
<th>Doubling time</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEK 293T</td>
<td>Human embryonic kidney cells</td>
<td>Ratio 1:6 as necessary or every 72h</td>
<td>24-30h</td>
<td>ECACC</td>
</tr>
<tr>
<td>HeLa</td>
<td>Human epithelial adenocarcinoma cells</td>
<td>Ratio: 1:4 as necessary or every 3-5 days</td>
<td>40-48h</td>
<td>ATCC</td>
</tr>
<tr>
<td>Phoenix</td>
<td>Human epithelial kidney cells</td>
<td>Ratio 1:3 as necessary or every 36h</td>
<td>24-30h</td>
<td>Prof Nolan, Stanford University</td>
</tr>
</tbody>
</table>

2.2.2 Determination of cell density

Cell density was verified before each passage and experiment using a Neubauer chamber (Hawksley, Brighton, UK). To do this, 10 µL of cell suspension were added to the chamber, and cell count was performed under the microscope. Density was determined using the following equation:

\[ \text{Cell density} = \frac{\text{Cells per quadrant}}{4} \times 10^4 / \text{mL} \]

Alternatively, counting by flow cytometry was performed by adding 50 µL of cell suspension to a mini flow tube and acquiring 10 µL of cells through an Accuri C6 Plus flow cytometer (2.7). After excluding cell debris, based on forward scattered light (FSC) and side scattered light (SSC), the number of events was multiplied by 100 to obtain the correct cell density.
(cells/mL). To ensure proper counting, beads seeded at a known density were acquired, in triplicate, before determining the density of each cell culture.

2.2.3 Determination of cell viability

TO-PRO™-3 Iodide (ThermoFisher) is a viability stain that is impermeant to live cells but can pierce through the membrane of dead cells, allowing their identification within a cell culture. This stain exhibits far-red fluorescence with excitation at 642 nm and emission at 661 nm. Cell viability was assessed by adding 1 µL of TO-PRO™-3 (5 µM) to 50 µL of cell suspension and analysed using an Accuri Plus C6 cytometer, as described in 2.7.

2.2.4 Cryopreservation and thawing of cell lines

For cryopreservation, cell suspensions were transferred to a UC and centrifuged at 270 x g for 10 min. Supernatants were discarded, and the pellets resuspended in 500 µL/vial of appropriate culture media (Table 2.3, Table 2.4). An equal volume of freezing media (Iscove Modified Dulbecco’s Medium (IMDM, Gibco) supplemented with 30% (v/v) FBS and 20% (v/v) Dimethyl Sulfoxide (DMSO; Sigma-Aldrich)), was added to the cells, following which samples were immediately transferred to 1.8 mL cryovials and placed in controlled refrigerated containers (Cool Cell™ Cell Freezing Container, BioCision, California, USA). These were placed in a -80°C freezer overnight to allow the cells to equilibrate. For long-term storage, cells were transferred to liquid nitrogen containers.

Cryopreserved cells were thawed by adding 1 mL of pre-warmed FBS, followed by rapid thawing at 37°C in a water bath. Cells were transferred to a UC containing 5 mL of fresh media, followed by a 5 min centrifugation at 270 x g. Supernatant was discarded, pellet was resuspended in the appropriate media and transferred to a new F25 tissue culture flask, to allow recovery overnight. The following day, cells were transferred to a new F75 flask and supplemented with fresh media (Table 2.3; Table 2.4). HSPC were recovered as described in 2.3.2.
2.3 Primary cell culture

2.3.1 Isolation of mononuclear cells from human neonatal cord blood

Human neonatal CB was obtained from healthy full-term pregnancies at the Maternity Unit of the University Hospital of Wales in Cardiff, with the informed consent of mothers undertaking elective caesareans. The permission for the use of CB was authorized through South East Wales local research ethics committee (licence number 06/WSE03/6). In addition, fresh CB was purchased from the NHS Bone Marrow and Transplant Unit in London. Mononuclear cells were isolated from cord blood by density gradient centrifugation using Ficoll-Paque™ (GE Healthcare, Little Chalfont, UK). Briefly, cord blood was diluted by adding an equal volume of Hank’s Balanced Salt Solution (HBSS; Gibco) supplemented with 10 mg/mL heparin, 20 µg/mL gentamycin and 25 mM HEPES solution (Gibco). Approximately 7 to 8 mL of diluted cord blood were gently layered on 5 mL of Ficoll-Paque™ and centrifuged at 400 xg for 40 min, with slow acceleration and brake off. The mononuclear cells found at the interface between the plasma and the Ficoll-Paque™ solution were carefully aspirated and transferred to a UC containing CB Wash (CBW; RPMI 1640, 5% v/v FBS, 20 µg/mL gentamycin and 10 mg/mL heparin), followed by a 10 min centrifugation at 270 xg. The mononuclear cells were washed again until supernatant was clear ensuring no platelet contamination. Cells were resuspended in 20 mL of CBW, from which 10 µL were removed and added to 190 µL of CBW and 1.5 µL of zaponin (Beckman-Coulter, High Wycombe, UK) for cell counting (2.2.2). Following centrifugation cells were frozen as described in 2.2.4 at a density of 5 x 10^7 mononuclear cells/vial. Once thawed, CD34+ cells were isolated as described in 2.3.2. Estimation of CD34 positivity within each sample was performed as described in 2.7.2.

2.3.2 Isolation of CD34+ HSPC from mononuclear cells

Following mononuclear cell isolation, CD34+ HSPC were isolated using the MiniMACS® Indirect CD34 MicroBead Kit (Miltenyi Biotec, Woking, UK) according to the manufacturer’s instructions. Briefly, to each vial of frozen cells, 1 mL of FBS and 20 µL of DNase (10 µg/mL; Sigma-Aldrich) were added and rapidly thawed in a 37°C water bath. Following viral content transfer to a labelled UC, an equal volume of a 1x Phosphate Buffer Saline (PBS) solution supplemented with 5 mM magnesium chloride (MgCl₂; Sigma-Aldrich) at RT was added
dropwise to the culture, over 3 minutes, resulting in the doubling of the volume. Two subsequent doubling dilution steps were performed. Cells were gently centrifuged and an additional 10 µL of DNase (10 µg/mL) were added. Subsequent volumes were determined per 1x10^8 cells, based on previous counts. Centrifuged pellets were resuspended in 400 µL of cold column buffer (MACS buffer; 1x PBS; 1% v/v Bovine Serum Albumin [BSA; Biosera]; 5 mM MgCl₂) and incubated at 4°C for 15 min with 100 µL hapten-conjugated monoclonal CD34 antibody (reagent A1) in the presence of FcR blocking agent (reagent A2). Cells were washed with 5 mL of MACS buffer to stop the reaction and centrifuged at 270 xg for 5 min. Washed cells were resuspended in 400 µL MACS buffer and incubated at 4°C with 100 µL of anti-hapten microbeads (reagent B) for 15 min. Subsequently, 5 mL of MACS buffer were added, and the cell suspension flowed through a cell 40 µm cell strainer, following which cells were centrifuged at 270 xg for 5 min. The cell pellet was later resuspended in 500 µL of MACS buffer and transferred to a pre-equilibrated MS/LS column (depending on desired cell number) (Miltenyi Biotec). Once column flow stopped, three wash steps were performed by adding 500 µL of MACS buffer to the column, following which CD34⁺ HSPC enriched population was eluted in 1 mL MACS buffer. The eluted cells were transferred to a second column and previously described wash steps performed to maximise enrichment. Following culture elution, cells were counted (2.2.2) and subcultured at 2x10^5 cells/mL in appropriate medium (2.3.3). Isolation of CD34⁺ HSCPC is defined as ‘day 0’ in growth and differentiation experiments (2.5, 2.7, 2.8).

2.3.3 Culture of CD34⁺ HSPC

Human CD34⁺ HSPC were cultured in IMDM containing 20% v/v FBS, 1% v/v BSA, 45 mM β-mercaptoethanol (β-ME; Sigma-Aldrich), 360 µg/mL of 30% iron saturated human transferrin (Roche Diagnostics; Switzerland) and 20 µg/mL gentamycin. Basal media was subsequently supplemented with specific combinations of cytokines, based on cell requirements (Table 2.5). After isolation, CD34⁺ HSPC were cultured in cytokine rich media, defined as 36S^high, to promote recovery and cell cycle progression. On day 3 of culture, and following retro- or lentiviral transduction (2.5), cells were maintained in 3S^lowG/GM (Table 2.5) to support myeloid cell differentiation. Media was replenished every 2-3 days to guarantee appropriate cytokine concentrations. Cells were seeded at different densities, according to developmental stages.
Table 2.5 – Cell culture media used for the growth and development of CD34+ HSPC

Human CD34+ HSPC were cultured in IMDM supplemented with 20% v/v FBS, 1% v/v BSA, human transferrin (30 mg/mL), β-ME (9 mM) and gentamicin (20 µg/mL). Different combinations of cytokines were added to the growth according to developmental stage.

<table>
<thead>
<tr>
<th>Days</th>
<th>Application</th>
<th>Cytokine*Concentration</th>
<th>Media</th>
</tr>
</thead>
<tbody>
<tr>
<td>0-3 Cell recovery</td>
<td></td>
<td>IL-3 50 ng/mL</td>
<td>36S&lt;sup&gt;high&lt;/sup&gt;</td>
</tr>
<tr>
<td></td>
<td></td>
<td>IL-6 25 ng/mL</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SCF 50 ng/mL</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G-CSF 25 ng/mL</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GM-CSF 25 ng/mL</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Flt3 50 ng/mL</td>
<td></td>
</tr>
<tr>
<td>3-13 Long-term culture</td>
<td></td>
<td>IL-3 5 ng/mL</td>
<td>38&lt;sup&gt;low&lt;/sup&gt;G/GM</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SCF 5 ng/mL</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>G-CSF 5 ng/mL</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>GM-CSF 5 ng/mL</td>
<td></td>
</tr>
</tbody>
</table>

* All cytokines were purchased from Biolegend (London, UK); **IL-3** – human interleukin-3; **IL-6** - human interleukin-6; **SCF** – human Stem Cell Factor; **G-CSF** - human Granulocyte Colony-Stimulating Factor; **GM-CSF** - human Granulocyte-Macrophage Colony Stimulating Factor; **Flt3** - Human Fms-like tyrosine kinase-3 ligand.
Within the 3rd and 6th days of culture, cells were seeded at a 0.5x10^5 cells/mL, to allow maximum proliferative rate. From days 6 through 10, cells were seeded at 2x10^5 cells/mL, and 3x10^5 cells/mL from 10 onwards.

### 2.3.4 Colony forming efficiency

To determine cloning-forming cell frequency, a limiting dilution colony assay was performed on Fluorescently Activated Cell Sorted (FACS) CD34+ HSPC (2.7.3) transduced with different viral plasmids (2.5.2), in 96-U plates, at a density of 0.3 cells/well, to ensure the presence of a cell within every third well, and incubated at 37°C with 5% CO₂. Individual colonies (>50 cells) and clusters (>5 cells) were scored by counting via light microscopy following 7 days of growth. To assess self-renewal efficiency, identified colonies were harvested, counted, and re-plated at a density of 1 cell/well, and cultured for an additional week before re-scoring, as mentioned above.

### 2.3.5 Assessment of cell morphology

To analyse cell morphology, 3 x 10^4 cells were transferred to a pre-assembled cytopspin slide chamber with a glass slide and centrifuged 60 xg for 5 min, using a Cytospin 3 (Fisher Scientific). Slides were subsequently stained with May-Grünwald-Giemsa for morphology examination and scanned using Zeiss AxioScan Z1 Slide Scanner (Carl Zeiss, New York, USA) at 20x magnification. Differential counts were performed using QuPath (Edinburgh, UK). Cells were defined as being in an early, intermediate, or late phase of development if they presented as myeloblasts/promyelocytes, myelocytes/metamyelocytes and band/segmented cells, respectively, as described in Figure 2.2.

### 2.4 Virus transfection

To optimize viral transduction of cell lines and CD34+ HSPC, thus ensuring high titre virus, two transfection methods were tested, as described below. Lipofection (2.4.1) was performed to generate lentivirus, using HEK 293T cells, whilst the calcium-phosphate method proved to be more efficient when producing retroviral particles using the Phoenix cell line (2.4.2).
Assessment of cell morphology was performed on sorted primare cells, on day 17 of culture, using May-Grünwald-Giemsa staining. Cells were scored as granulocytes, in different stages of differentiation, and identified in early, intermediate, or late developmental stages. Additionally, cells were scored as monocytes and macrophages.

**Figure 2.2 – Morphological analysis of HSPC**
2.4.1 Transfection using Lipofection

Lipofection was achieved the Lipofectamine® 3000 Kit (Thermo Fisher Scientific) according to the manufacturer’s instructions. Briefly, cells were counted and seeded the day before transfection at $5 \times 10^6$ (T25) or $15 \times 10^6$ cells (T75) per flask in a previously Poly-D-Lysine (Trevigen®, Abingdon, UK) coated flask. Flasks were incubated overnight at 37°C, in 5% CO$_2$ incubator to allow monolayer formation. The following day, cell confluence was assessed, and two tubes were prepared as follows.

For retroviral production, tube A was prepared by adding 8 µg (T25) or 24 µg (T75) of the desired vector to 667 µL (T25) or 2 mL (T75) of Opti-MEM Media (Gibco) and 16µL (T25) or 48 µL (T75) of the Lipofectamine 3000 Reagent. Tube B was prepared by adding 667 µL (T25) or 2 mL (T75) of Opti-MEM Media and 10 µL (T25) or 57 µL (T75) of the P3000 Enhancer Reagent. To prepare lipid-DNA complexes, contents of Tube A were transferred to Tube B and incubated for 15-20 min at RT. Following incubation, media in each flask was reduced to 2 mL (T25) or 6 mL (T75), and the lipid-DNA complex was carefully added to the cells. Flasks were incubated for 6 h at 37°C, following which time all media was removed, and replaced with fresh. Cells were incubated for 24 h, following which time supernatant was removed and discarded, and fresh media was added. Retrovirus harvesting was performed 48- and 72h post-transfection, with media being changed between each harvest. Viral supernatants were centrifuged at 270 xg for 10 min to remove cell debris, and aliquoted into 1.8mL cryovials, following with they were snap frozen in liquid nitrogen. For long-term storage, cryovials were kept at -80°C and thawed when necessary.

For lentiviral transfection, two additional plasmids were added to Tube A: 2.2 µg (T25) or 6.6 µg (T75) of pMD2, an envelope plasmid, and 4 µg (T25) or 12 µg (T75) of psPAX2, a packaging plasmid, as well as 2.2 µg (T25) or 6.6 µg (T75) of the desired vector; tube B was prepared as described for above. For viral harvest, lentiviral supernatant was collected 24- and 48h post-transfection and pooled (stored overnight on each collection at 4°C). Once both viral harvests were performed and combined, viral supernatants were centrifuged at 270 xg for 10 min and aliquoted/stored as described above.
To assess target protein expression, cells were collected, and proteins extracted, according to 2.6. Furthermore, where cells were transfected with virus encoding for GFP expression, visual inspection by fluorescent microscope was performed.

### 2.4.2 Transfection using calcium-phosphate precipitation

To improve transduction efficiencies in human CD34+ HSPC, an alternative method of virus transfection was used. For this, Phoenix virus-packaging cells were transfected using the Calcium phosphate transfection kit (Invitrogen™) according to the manufacturer’s instructions (this method was only used for retroviral production). Briefly, cells were counted and initially seeded at 8x10^6 cells per T75 flask. The following morning, media was replaced with 14 mL of fresh pre-warmed media. By the afternoon, tissue grade sterile water was mixed with 45 μg of plasmid DNA and 45 μL of CaCl₂ (2.5M), followed by the dropwise addition of 450 μL of 2x HEPES-buffered saline (HBS; 50mM) whilst the solution was bubbled with a pipette. Following vortexing, the solution was left to precipitate for 20 min at RT. Prior to the transfection, 15 μL of chloroquine (25 μM) was added to the cultures, followed by the dropwise addition of DNA to the medium. Cells were incubated 37°C and 5% CO₂ until the following day, by which growth media was changed and cells were incubated at 33°C to maximise retroviral stability. Viral supernatant was harvested at 48 and 72 h post-transfection as previously described (2.4.1).

### 2.4.3 Virus titration

Virus titration was performed using K562 cells. 96-well untreated plates were coated with Retronectin® (30 μg/mL; Takara Biotech Inc., Shiga, Japan), a recombinant human fibronectin peptide fragment that promotes binding between cells and viruses (Tonks et al., 2005), and incubated at 4°C overnight. The following day, Retronectin® was removed and washed with blocking buffer (PBS with 1% (v/v) BSA) for 30 min at RT. Following blocking, 100 μL of K562 cells at a density of 1 x 10^5 cells/mL were added to each well, before adding 50 μL of viral supernatant (2.3.4) to each well; retroviral and lentiviral aliquots of known titre (standards) were also included to standardise the data. Cells were cultured for two days at 37°C and 5% CO₂. Following incubation, media was carefully aspirated and washed to remove residual virus, by gently adding PBS without disturbing the cells attached to the wells. Cells
were resuspended in Staining Buffer (SB; 1x PBS supplemented with 0.02% (v/v) sodium azide and 1% (v/v) BSA) and analysed by flow cytometry for GFP expression (2.7).

2.5 Transduction of cell lines and human CD34+ HSPC

2.5.1 Transduction of cell lines

For overexpression studies, using either retro- or lentiviral vectors, 12-well non-tissue culture plates were coated Retronectin® (30 μg/mL) and incubated at 4°C overnight. The following day, Retronectin® was aspirated and replaced with 1% (v/v) BSA as previously described (2.4.3). The BSA was aspirated and replaced with 1 mL of retro- or lentiviral supernatant. The culture plate was centrifuged at 3000 xg for 120 minutes at RT, following which the supernatant was removed from the wells and replaced with the appropriate number of cells (4x10^5 cells/mL), subsequently incubated at 37°C overnight. Cells expressing GFP were assessed alongside parental cell cultures to measure infection rates, the day following infection, by flow cytometry (2.7).

For KD studies using shRNA constructs, pre-coating of cells was unnecessary given the higher titre virus generated from these constructs. Instead, cells were resuspended at 4x10^5 cells/mL and seeded into a 12-well plate, and 1 mL of culture was added along with 0.5-1 mL of viral supernatant. Cells were assessed for GFP expression as previously described. Following successful cell transduction, cultures were subjected to antibiotic selection to enrich the culture for transduced cells and ensure a pure population. To this end, transduced cells were incubated with 10 µg/mL puromycin (Sigma-Aldrich; P8833), until parental non-transduced cultures were no longer viable. Following selection, transduced cell lines were maintained in normal growth media until confluency was achieved, by which point cells were cryopreserved (2.2.4) and total protein lysates were collected to validate protein expression by western blot analysis (2.6). KD studies were performed within 3 weeks of generating the cell lines.

2.5.2 Transduction of human CD34+ HSPCs

For overexpression and KD studies, plates were coated Retronectin® (30 μg/mL) and incubated at 4°C overnight, as previously described (2.5.1). Following BSA washing, viral supernatants were added to the corresponding wells and plates centrifuged at 3000 xg for 120 minutes at RT. Following centrifugation, the viral supernatant was removed and CD34+ cells,
previously seeded at 2 x 10^5 cells/mL for overnight recovery following isolation (2.3.2), were added to the wells. For overexpression studies, the following day, cells were temporarily transferred to a UC and placed at 37°C in the incubator, whilst another round of infection was performed as above. For KD studies, however, a second viral coating step was not performed due to high viral titre. Instead, cells were incubated with the lentiviral particles for 2 consecutive days, before being harvested and analysed on day 3 of culture. Cells were subsequently analysed through flow cytometry and used to perform the assays described in 2.7.2-2.7.6, 2.3.4 and 2.8.

2.6 Determination of protein expression by western blot

2.6.1 Total protein extraction

To generate total protein lysates, the appropriate number of cells to be fractionated (1x10^6 AML cells; 2x10^6 HSPC) was centrifuged at 270 x g for 10 min. Supernatant was discarded, and cell pellets were washed using 10 mL of ice-cold tris-buffered saline (TBS; 50 mM Tris [Sigma-Aldrich] dissolved in dH_2O and adjusted to pH 7.6 with HCl; 150 mM NaCl [Sigma - Aldrich] dissolved in dH_2O). Supernatant was discarded and pellets snap frozen in liquid nitrogen before protein extraction. When required, frozen cell pellets were thawed on ice in the presence of 1 mg/mL DNase for 5 min and resuspended in 50 µL of homogenisation buffer (0.25 M sucrose [Sigma-Aldrich], 10 mM HEPES-KOH pH7.2 [Gibco], 1mM magnesium acetate [Sigma-Aldrich], 0.5 mM EDTA [Sigma-Aldrich], 0.5 mM EGTA [Sigma-Aldrich], 12.6 M BME [Sigma-Aldrich], 1 tablet of EDTA free-protease inhibitor [Roche Diagnostics; Basel, Switzerland], 1% v/v x100-Triton [Sigma-Aldrich]). Cells were incubated on ice for 30 min and subjected to occasional vortexing. Subsequently, cell suspension was transferred to ice cold Eppendorf tubes (Eppendorf, Stevenage, UK) and centrifuged for at 16,000 x g for 5 min at 4°C, using a Biofuge fresco Heraeus centrifuge (ThermoFisher Scientific). Supernatant was collected transferred to a new pre-chilled Eppendorf and stored at -80°C until required. Protein concentration was determined by Bradford assay (2.6.3).

2.6.2 Fractionated protein extraction

To generate fractionated protein cell lysates, extractions were performed using the Biovision Nuclear/Cytosol Fractionation Kit (Cambridge Bioscience, Cambridge, UK), following manufacturer’s instructions. Briefly, the appropriate number of cells (2x10^6 AML cells; 5x10^6
HSPC) was centrifuged at 270 xg for 10 min, following which supernatant was discarded and pellets washed with 20 mL of ice-cold TBS. Subsequently, cell pellets were resuspended in 200 µL of Cytosol Extraction Buffer A, supplemented with Protein Inhibitor Cocktail (PIC) and Dithiothreitol (DTT) and vortexed, followed by a 10-min incubation on ice. Additionally, 11 µL of Cytosolic Extraction Buffer B were added, cells were vortexed again and incubated on ice for a further minute. Cell pellets were vortexed and centrifuged at 10,000 xg for 8 min, at 4˚C. The supernatant, containing the cytosolic fraction was collected and stored at -80˚C. The remaining cell pellet was washed with 500 µL of PBS supplemented with 5 mM MgCl₂ and centrifuged for 3 min at 10,000 xg. Supernatant was discarded and the pellet snap frozen in liquid nitrogen. Following 2-3 cycles of rapid freezing and thawing in the vapour phase on liquid nitrogen, thus allowing the rupture of the nuclear membrane, 2 µL/million cells of benzonase (25 U/µL; Merck Millipore, USA) was added, and cells were incubated on ice for 30 min, subjected to occasional vortexing. Following incubation, 50-100 µL of TEAB lysis buffer (0.5 M Triethylammonium bicarbonate buffer; 0.05% SDS [ThermoFisher Scientific]; 0.1% Protease Inhibitor Cocktail [Sigma-Aldrich]; 0.1% Phosphatase Inhibitor Cocktail [Sigma-Aldrich], dH₂O) was added, and cells were kept on ice for another 30 min, with occasional vortexing. Following incubation, cell lysates were centrifuged at 10,000 xg for 10 min, at 4˚C and supernatant collected to a fresh tube. Protein quantification was carried out as described below (2.6.3).

2.6.3 Protein quantification using Bradford assay

Protein quantification was performed using Bradford Assay (Ernst and Zor, 2010). Total protein extract samples were diluted 1:100 using ddH₂O, whilst cytosolic/nuclear samples were diluted 1:50. Each sample was assayed in duplicate by adding 10 µL of diluted lysates alongside 10 µL protein standards, ranging between 0-100 µg/mL (Sigma-Aldrich), which were also added in duplicate. Bradford’s reagent (Sigma-Aldrich; B6916) was diluted 1:1 in dH₂O, and 190 µL of diluted reagent was added to each well. The plate was then incubated for 5 at RT, protected from light. Absorbances were measured at 595 nm using a plate reader. The protein standards were used to generate a standard curve from which the concentration of protein in each sample was determined.
2.6.4 Protein electrophoresis and electroblotting

All reagents and equipment were purchased from Invitrogen and were prepared according to the manufacturer’s instructions, unless otherwise stated. Prior to SDS-polyacrylamide gel electrophoresis, samples were denatured in a 70°C water bath in the presence of NuPAGE™ 4x Lithium Dodecyl Sulphate Sample Buffer (LDS), NuPAGE™ 10x Sample Reducing Agent and dH2O. Protein gel electrophoresis was performed using a NuPAGE™ XCell SureLock™ Mini-cell and XCell II blot module system. NuPAGE™ 4-12% Bis-Tris gels (12- or 17-well) were placed in a module to which 1x NuPAGE™ MOPS SDS Running Buffer, supplemented with 500 µL of NuPAGE™ antioxidant was added. Between 10 – 20 µg of protein were added to each well, to a maximum volume of 10 µL (in a 17-well gel) to 20 µL (in a 12-well gel). Samples were loaded alongside MagicMarker XP protein ladder (1/10 dilution in 4x LDS sample buffer and dH2O), to allow visualisation of protein molecular weight. Following sample loading, the outer chamber of the gel tank was filled with approximately 600 mL of 1x MOPS buffer. Proteins were separated through electrophoresis at 200V for 50 min.

During this period, 1x NuPAGE™ Transfer buffer supplemented with 1 mL NuPAGE™ Antioxidant and 10% v/v methanol (20% v/v if blotting two gels) was prepared and used to pre-soak blotting pads and filter paper. Polyvinylidene difluoride (PVDF) membranes were soaked for 10 min in methanol prior to electroblotting. Following electrophoresis, the gels were removed from their cassettes and assembled between the pre-soaked PVDF membrane and filter paper, followed by the pre-soaked blotting pads. Following module assembly, 1x transfer buffer was used to fill the transfer cell until the pads were completely submerged, and 600 mL dH2O was used to fill the outer chamber. Membranes were electroblotted for 1 h at 30V.

Following transfer, electroblotted membranes were washed twice with dH2O using a plate shaker (1 revolution / sec) for 5 min. Proteins were visualised for both equal loading and to facilitate cutting of membrane using 20 mL of Ponceau S solution (Sigma-Aldrich), for 30 sec. The Ponceau S stain was removed using dH2O before membranes were blocked with 2.5% (w/v) powdered milk diluted in TBS supplemented with 0.1% (v/v) Tween-20 (TBS-T; Sigma-Aldrich) for 30 min at RT on a rotary shaker. Membranes were subsequently washed for 15 min in TBS-T, followed by 3 x 5-min cycles of further TBS-T washes. Membranes were used for immunoblotting (2.6.5), as described below.
2.6.5 Immunoblotting and protein detection

Following membrane blocking and subsequent washes, membranes were incubated with primary antibodies as described in Table 2.6, in 2.5% (w/v) powdered milk diluted TBS-T overnight at 4°C. The following day, membranes were washed with TBS-T, as described above (2.6.4), and incubated with the appropriate anti-rabbit or anti-mouse horseradish peroxidase (HRP) conjugated secondary antibody, diluted in 1% (w/v) powdered milk diluted in TBS-T for 1 h at RT. Following incubation, membranes were washed with TBS-T as previously described (2.6.4). HRP-conjugated antibodies were exposed through a chemiluminescence reaction using Amersham™ ECL™ Prime (GE HealthCare Life Sciences), according to the manufacturer’s instructions. Briefly, equal volumes of Amersham™ ECL™ Luminol Enhancer solution (solution A) and Amersham™ ECL™ peroxide solution (solution B) were added to the membranes, to a combined volume of 4 mL. Following a 5 min incubation at RT, protected from light, excess substrate was removed using filter paper. Membranes were visualized using a LAS-3000 digital imager (Fujifilm UK Ltd, Bedfordshire, UK), using an exposure time of up to 30 min.

Equal loading was further assessed using Glyceraldehyde 3-Phosphate Dehydrogenase (GAPDH) or Histone H1 antibodies for total/cytosolic and nuclear protein extracts, respectively, using fluorescent antibodies. For this, following chemiluminescent imaging, membranes were washed with TBS-T, as described above (2.6.4). Subsequently, antibodies were diluted in 2.5% (w/v) powdered milk diluted TBS-T and incubated for 1h at RT, protected from light, following which membranes were washed for 5 min using TBS-T. Membranes were imaged directly after probing using an Odyssey® FC Imaging System (LI-COR Biosciences, Nebraska, USA) for 30s. Image files were saved as .TIFF Images and analysed using Adobe Photoshop CS4. Densiometric analysis was performed as described in 2.6.7.

2.6.6 Membrane and gel protein staining

Alternatively, following protein transfer, to confirm protein loading and sample integrity, membranes were stained with SYPRO® Ruby protein blot stain (ThermoFisher Scientific) for 15 min in a rotary shaker. Following incubation, these were washed 4-6 times with dH2O for 1 min and visualized in an Odyssey® Fc Imaging System (LI-COR Biosciences, Cambridge, UK). These membranes were not suitable for subsequent antibody incubation.
Table 2.6 – List of primary antibodies used in this study

Table outlining the source, clone, conjugate, and dilution of the antibodies used in western blot analysis (2.6.5). Primary antibodies were stored at -20°C and secondary antibodies at 4°C.

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Cat. #</th>
<th>Species</th>
<th>Conjugate</th>
<th>Conditions</th>
<th>Dilution</th>
<th>Source</th>
<th>RRID</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary antibodies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AML1 (RUNX1)</td>
<td>#4334</td>
<td>Rabbit</td>
<td>-</td>
<td>4°C · O/N · Chemi</td>
<td>1:5000</td>
<td>CST</td>
<td>AB_2184099</td>
</tr>
<tr>
<td>ARID5B</td>
<td>PA5-78597</td>
<td>Rabbit</td>
<td>-</td>
<td>4°C · O/N · Chemi</td>
<td>1:1000</td>
<td>Fisher Scientific</td>
<td>AB_2735478</td>
</tr>
<tr>
<td>C/EBPβ</td>
<td>#90081</td>
<td>Rabbit</td>
<td>-</td>
<td>4°C · O/N · Chemi</td>
<td>1:1000</td>
<td>CST</td>
<td>Not defined</td>
</tr>
<tr>
<td>IRF7</td>
<td>PA5-79519</td>
<td>Rabbit</td>
<td>-</td>
<td>4°C · O/N · Chemi</td>
<td>1:1000</td>
<td>Fisher Scientific</td>
<td>AB_2746635</td>
</tr>
<tr>
<td>IRF9</td>
<td>D2T8M</td>
<td>Rabbit</td>
<td>-</td>
<td>4°C · O/N · Chemi</td>
<td>1:1000</td>
<td>CST</td>
<td>AB_2799885</td>
</tr>
<tr>
<td>ZNF217</td>
<td>#720352</td>
<td>Rabbit</td>
<td>-</td>
<td>4°C · O/N · Chemi</td>
<td>1:1000</td>
<td>Fisher Scientific</td>
<td>AB_2716919</td>
</tr>
<tr>
<td>Secondary Antibodies</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anti-mouse</td>
<td>NA931</td>
<td>Donkey</td>
<td>HRP</td>
<td>RT · 1h</td>
<td>1:5000</td>
<td>GE Healthcare</td>
<td>AB_772210</td>
</tr>
<tr>
<td>Anti-rabbit</td>
<td>NA934</td>
<td>Donkey</td>
<td>HRP</td>
<td>RT · 1h</td>
<td>1:5000</td>
<td>GE Healthcare</td>
<td>AB_772206</td>
</tr>
<tr>
<td>Loading Controls</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GAPDH</td>
<td>GA1R</td>
<td>Mouse</td>
<td>DyLight® 680</td>
<td>RT · Fluor 1h ·</td>
<td>1:10000</td>
<td>Fisher Scientific</td>
<td>AB_2537657</td>
</tr>
<tr>
<td>Histone H1</td>
<td>AE-4</td>
<td>Mouse</td>
<td>AF790</td>
<td>RT · Fluor 1h ·</td>
<td>1:5000</td>
<td>SCBT</td>
<td>AB_675641</td>
</tr>
</tbody>
</table>

**AF90** – Alexa Fluor 790; **Chemi** – Chemiluminescence; **CST** – Cell Signalling Technologies (London, UK); **Fluor** – Fluorescence; **HRP** – Horseradish peroxidase; **O/N** – Overnight; **RRID** – Research Resource Identifier; **SCBT** – Santa Cruz Biotechnology (Heidelberg, Germany)
2.6.7 Densitometry analysis

Relative changes in protein expression were semi-quantitatively measured using Image J software (Fiji; v.2.0.0.71; U. S. National Institutes of Health, Maryland, USA). Briefly, a region of interest (ROI) was drawn around a band of interest. From this, a histogram of peak intensity was generated, and a baseline of background intensity was set from the area surrounding the band within the ROI. Subsequently, the area around the curve was calculated to give an arbitrary intensity value. The band intensities of the protein of interest were normalised to the band intensity of the loading control for each sample. Furthermore, to compare protein expression between a test sample and the corresponding control, the first was normalised to the latter.

Unless otherwise stated, due to the limited availability of cord blood, and the high number of cells necessary for proteomic analysis, western blot analysis was only performed once (n=1).

2.7 Flow cytometry

2.7.1 Equipment and Analysis

For this study, a benchtop Accuri C6 Plus Flow Cytometer™ (Accuri Cytometers, BD) was used. The technical specifications of this instrument are outlined in Table 2.7. Data analysis was performed using FCS Express® ver.6 software (De Novo Software, Pasadena, USA).

2.7.2 Estimation of CD34+ cell count

The percentage of CD34+ in human cord blood samples was determined by performing an immunophenotyping assay, in parallel with the isolation protocol (2.3.2). Briefly, 100 μL of undiluted blood were stained with 10 μL of CD34- phycoerythrin (PE) and 5 μL of CD45-allophycocyanin (APC), in parallel with a control sample stained with 10 μL of IgG1-PE and 5 μL of CD45-APC (Table 2.8). Following an incubation at 4°C for 30 min, 5 mL of FACSTM Lysis Buffer (BD, Berkshire, UK) was added to each tube and incubated at RT for 10 min. The cells were washed in 5 mL 1x PBS, followed by a centrifugation at 200 x g for 5 min. and supernatant discarded. Cells were resuspended in residual supernatant and analysed by flow cytometry according to the gating strategy described in Figure 2.3.
Table 2.7 – Technical specifications of the Accuri™ C6 Plus used in this study

<table>
<thead>
<tr>
<th>Fluorophore</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laser excitation</td>
<td>Red – 640 nm</td>
</tr>
<tr>
<td></td>
<td>Blue – 488 nm</td>
</tr>
<tr>
<td>Emission Detection</td>
<td>FITC/GFP – 533/30 nm (FL1)</td>
</tr>
<tr>
<td></td>
<td>PE/PI – 585/40 nm (FL2)</td>
</tr>
<tr>
<td></td>
<td>PerCP-Cy™5.5 – &gt; 670 nm (FL3)</td>
</tr>
<tr>
<td></td>
<td>APC – 675/25 nm (FL4)</td>
</tr>
<tr>
<td>Flow rate</td>
<td>10-100 µL/min</td>
</tr>
</tbody>
</table>

APC - Allophycocyanin; FITC - Fluorescein isothiocyanate; GFP – Green fluorescent protein; PE – Phycoerythrin; PI – Propidium iodide; PerCP-Cy5.5 – Peridinin chlorophyll protein – Cyanine5.5

Table 2.8 – List of antibodies used for cord blood immunophenotyping

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Fluorescent conjugate</th>
<th>Clone</th>
<th>Dilution</th>
<th>Source</th>
<th>RRID</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgG1</td>
<td>PE</td>
<td>MOPC-21</td>
<td>1/10</td>
<td>Biolegend</td>
<td>AB_2847829</td>
</tr>
<tr>
<td>CD34</td>
<td>PE</td>
<td>581</td>
<td>1/10</td>
<td>Biolegend</td>
<td>AB_1731862</td>
</tr>
<tr>
<td>CD45</td>
<td>APC</td>
<td>HI30</td>
<td>1/20</td>
<td>Biolegend</td>
<td>AB_314400</td>
</tr>
</tbody>
</table>

APC - Allophycocyanin; PE – Phycoerythrin; RRID – Research Resource Identifier.
Figure 2.3 – Estimation of CD34+ HSPC in cord blood by flow cytometry

Representative bivariate dot plots showing the gating strategy used for the estimation of CD34+ HSPC in human cord blood samples. (A) Example gate used to exclude cell debris, based on FSC and SSC; (B) Using the “Non-debris” gate, the expression of CD45 was used to identify mononuclear cells (CD45+); (C) The percentage of CD34+ cells were estimated from within the mononuclear cell population. Background auto fluorescence was established using CD34+ cells stained with the isotype control IgG1-PE (Table 2.8).
2.7.3 Fluorescence Activated Cell Sorting

Following transduction (2.5.2), CD34+ HSPC were subjected to FACS for GFP positivity to ensure an enriched population (Figure 2.4), using a BD FACSARiaIII™, within the Central Biotechnology Services (CBS) facilities, Cardiff University. Cells were counted and centrifuged at 270 xg for 10 min. Supernatant was discarded and cells were resuspended at a final density of 1 x 10^6 cells/mL in PBS containing 1% BSA and 5 ng/mL IL-3, to minimize background. Resuspended cells were passed through a 40 µM cell strainer into 15 mL falcon tubes. Uninfected (mock) cells were used to set background autofluorescence. Cells were sorted at RT, using the 100 µm nozzle, at a low flow rate. Sorted cells were collected in 15 mL falcon tubes containing 8 mL of IMDM supplemented with 10% (v/v) FBS and 20 mg/mL gentamycin. Following sorting, cells were transferred to UCs and centrifuged for 5 min at 270 xg, at RT. Following centrifugation, supernatant was discarded, and cells resuspended in 1 mL 5^lowG/GM (Table 2.5) for cell enumeration by flow cytometry (2.2.2). Cell density was adjusted as appropriate, for a final concentration of 2 x 10^5 cells/mL. Sorted cells were subsequently used in colony assays (2.3.4) and morphology slides (2.3.4).

2.7.4 Immunophenotypic analysis of transduced CD34+ HSPCs

To assess myeloid cell differentiation, immunostaining was carried at different time-points using a panel of cell surface markers (Table 2.9). CD13-APC, in combination with CD36-biotin was used for lineage discrimination, along with streptavidin PerCP-Cy™5.5, in a two-step antibody labelling. Furthermore, cells were incubated with one on the following PE-labelled markers: IgG1, CD11b, CD14, CD15 and CD34. Briefly, 2.5 - 5 x 10^5 cells were washed and resuspended in 75 µL SB. Antibodies were added to a U-bottom 96-well plate, followed by the addition of 15 µL of cell suspension. Plates were briefly centrifuged to bring down the contents of each well, and carefully vortexed prior to incubation at 4°C for 30 min. Subsequently, 150 µL of SB was added to each well and its contents transferred to mini flow tubes. Cells were centrifuged at 200 x g for 5 min and supernatant discarded, following which samples were labelled with streptavidin-peridinin chlorophyll protein (SA-PerCP-Cy™5.5; BD Pharmingen, USA) for 30 min at 4°C. Stained cells were washed and resuspended in 100 µL of SB for analysis using the Accuri™C6 Plus cytometer (2.7.1), using the gating strategy outlined in Figure 2.5 and the appropriate compensation settings (Supplementary Table 1).
Figure 2.4 – Gating strategy used to sort GFP-expressing CD34+ HSPC

Representative bivariate dot plots showing FACS-enriched CD34+ HSPC cultures for transduced cells, according to GFP positivity, using FACSariaIII™. (A) Example gate used to exclude cell debris, based on FSC and SSC; (B) Example gating used to exclude doublets, based on the previously established “Non-debris”. (C) The percentage of GFP+ cells within a bulk culture was established based on the “Non-debris” and “Doublet exclusion” gates shown in A and B. Background auto fluorescence was established using cells subjected to the equivalent viral infection procedure but in the absence of virus (for CD34+ cultures – mock culture).
Table 2.9 – Antibodies used for immunophenotypic analysis of transduced CD34+ HSPC

<table>
<thead>
<tr>
<th>Antibody</th>
<th>Fluorescent conjugate</th>
<th>Clone</th>
<th>Dilution</th>
<th>Source</th>
<th>RRID</th>
</tr>
</thead>
<tbody>
<tr>
<td>IgG1</td>
<td>PE</td>
<td>MOPC-21</td>
<td>1/17</td>
<td>Biolegend</td>
<td>AB_2847829</td>
</tr>
<tr>
<td>CD11b</td>
<td>PE</td>
<td>ICRF44</td>
<td>1/17</td>
<td>Biolegend</td>
<td>AB_314158</td>
</tr>
<tr>
<td>CD13</td>
<td>APC</td>
<td>WM15</td>
<td>1/17</td>
<td>Biolegend</td>
<td>AB_314182</td>
</tr>
<tr>
<td>CD14</td>
<td>PE</td>
<td>HCD14</td>
<td>1/17</td>
<td>Biolegend</td>
<td>AB_830679</td>
</tr>
<tr>
<td>CD15</td>
<td>PE</td>
<td>W6D3</td>
<td>1/17</td>
<td>Biolegend</td>
<td>AB_756012</td>
</tr>
<tr>
<td>CD34</td>
<td>PE</td>
<td>581</td>
<td>1/17</td>
<td>Biolegend</td>
<td>AB_1731862</td>
</tr>
<tr>
<td>CD36</td>
<td>Biotin</td>
<td>SMO</td>
<td>1/100</td>
<td>Ancell Corporation</td>
<td>Not defined</td>
</tr>
<tr>
<td>PerCP-Cy™5.5 Streptavidin</td>
<td>PerCP-Cy™5.5</td>
<td>-</td>
<td>1/100</td>
<td>BD Biosciences</td>
<td>AB_2868934</td>
</tr>
</tbody>
</table>

APC - Allophycocyanin; PE – Phycoerythrin; PerCP-Cy5.5 – Peridin chlorophyll protein – Cyanine5.5; RRID – Research Resource Identifier.
Figure 2.5 – Immunophenotypic analysis of CD34+ HSPC

Representative bivariate plots and histograms outlining the gating strategy used for immunophenotypic analysis of CD34+ HSPC throughout haematopoietic development. Cells were gated according to the “Non-debris” gate showed in Figure 2.4A. Appropriate compensation was applied before the analysis. (A) The CD13 and CD36 markers were used to discriminate between the granulocytic, monocytic and erythroid lineages within GFP+ population (Figure 2.4C). (B-C) Each subpopulation was further examined for differentiation markers, such as the granulocytic marker CD15 or monocytic marker CD14 (Table 2.9) (Isotype - filled histogram; CD14/CD15 – open histogram).
2.7.5 Cell cycle analysis

To determine changes in CD34+ HSPC cell cycle, propidium iodide (PI, Sigma-Aldrich) was used to label intracellular DNA. For this analysis, 5 x 10^4 cells were transferred to a UC and washed with 20 mL of 1x PBS. Following centrifugation, supernatant was discarded, and cultures resuspended in 300 µL of PBS, following which cells were fixed for 30 min on ice by adding 700 µL absolute ethanol. After fixation, samples were stored at -20°C for an O/N incubation. The following day, cells were washed with 10 mL of 1x PBS and centrifuged at 270 xg for 10 min. Supernatant was discarded and cells were resuspended in 75 µL 1x PBS. Subsequently, cells were incubated with 25 µL staining solution, containing 40 µg/µL PI and 0.1 mg/mL RNase (Sigma-Aldrich) diluted in 1x PBS, at 37°C for 30 min. Samples were acquired within 20 min following incubation using the Accuri™C6 Plus cytometer (2.7.1). Cell cycle analysis was performed using FCS Express’s Multicycle AV DNA analysis tool plug-in, according to the gating strategy outlined in Figure 2.6.

2.7.6 Apoptosis

To determine the proportion of apoptotic cells, an apoptosis assay using Annexin V (ThermoFisher) was performed. Annexin V is a calcium-dependent phospholipid-binding protein, that binds to phosphatidylserine (PS), usually located in the inner leaflet of the plasma membrane. When a cell becomes apoptotic, PS is translocated to the extracellular membrane leaflet, and is detected by the fluorescent labelled Annexin V in a calcium-dependent manner. To perform this assay, 5 x 10^4 cells were transferred onto a labelled UC and washed with 1x PBS, followed by an additional wash with 1x Annexin V Binding Buffer (ThermoFisher). Supernatant was discarded and cells resuspended in 50 µL Binding Buffer. Afterwards, 5 µL of APC-conjugated Annexin V was added and cells incubated for 10-15 min at RT. Following incubation, cells were washed with 1x Binding Buffer and resuspended in 100 µL of the same solution. 5 µL of PI or 7-AAD were added to the UCs and cells analysed by flow cytometry (2.7.1), using the setting described in Figure 2.7.
Figure 2.6 – Gating strategy used to perform cell cycle analysis

Representative bivariate plots and histogram describing the gating strategy used for the analysis of cell cycle status by flow cytometry. (A) Gating strategy used to exclude cell debris based on FSC and SSC; (B) Gating strategy used to exclude doublet and cell aggregates based on Height and Area parameters. (C) Histogram showing DNA content analysis, and determination of the proportion of cells within the G1, S and G2 phases on cell cycle.
Figure 2.7 – Gating strategy used to analyse apoptosis

Representative bivariate dot plots showing (A) gating strategy used to identify cell debris and exclude them for further analysis, based on FSC and SSC; (B) gating strategy used to exclude doublet cells, based on PI-height and PI-area; (C) quadrant gating used to determine viable (Q4), pre-apoptotic (Q3) and late apoptotic cells (Q2).
2.7.7 Growth and proliferation assay

To measure growth and proliferation of cell lines, a proliferation/viability assay was performed. Cells in log phase growth were pelleted in a UC and washed in 10 mL of serum free medium. Cells were seeded at a concentration of 1x10^5 cells/mL in a 24-well plate. Following seeding, 50 µL of cell suspension was recovered and transferred to FACS tubs to confirm cell density. In addition, 1 µL of 5 µM TOPRO-3 was added to each tube to check cell viability, as described in 2.2.3.

The following day, cells in each well were resuspended and 50 µL of cell suspension were transferred to FACS tubes. To each tube, 1 µL of TO-PRO was added, and suspension mixed. The Accuri Plus B6 flow cytometer was used to count the cells and cell’s viability was calculated for each sample (2.2.2). This process was repeated for 5 consecutive days.

2.8 Proteomic analysis using SWATH®

Sample preparation was performed at Cardiff University. Sample processing was performed during a visit to Manchester University, whilst analysis was performed in the Stoller Biomarker Discovery Centre in the University of Manchester by Dr. Andrew Pierce and Dr. Bethany Geary, in Prof. Anthony Whetton’s group.

2.8.1 Sample preparation

Human cord blood-derived CD34+ HSPC were isolated as described in 2.3.2. Subsequently, cells were transduced with either a control (PINCO GFP) or a RUNX1-ETO PINCO vector (n=3), according to 2.5.2. On the third day of cell culture, efficiency of infection was assessed through GFP expression (2.7), following which nuclear and cytosolic proteins were extracted. Following protein quantification (2.6.3), samples were stored at -80°C until processing at the University of Manchester (2.8.2).

2.8.2 Tryptic digestion

Samples were prepared by diluting 10 µg of extract 30 µL of 30 nM Ammonium Bicarbonate (AmBic). Disulphide bonds were reduced by adding 3 µL of 50mM TCEP, followed by an incubation at 60°C for 1 h. Subsequently, proteins were alkylated by adding 3.3 µL of 60mM iodoacetamide and incubated at RT for 30 min in the dark. Lastly, proteins were
digested by adding trypsin at a 10:1 substrate:enzyme ratio (i.e. by adding 10 µg of trypsin per 100 µg of protein). Samples were incubated overnight at 37°C.

### 2.8.3 Strong Cation Exchange Clean-up

To remove contaminating solutions from the digestion process and to purify the sample, a UltraMicroSpin™ (The Nest Group, Massachusetts, USA; [URL: https://www.nestgrp.com]) column based on Strong Cation Exchange (SCX) was used, according to the manufacturers protocol. Columns were conditioned by adding 100 µL of 100% methanol and centrifuged for 1 min at 110 x g. The columns were subsequently washed with 50 µL of water and centrifuged 1 min at 110 x g. After discarding the liquid, 100 µL of a 0.2M monosodium phosphate supplemented with 0.3M sodium acetate solution (pH 3-6.5) was added, and columns centrifuged for 10 sec to allow for the solution to be in contact with the filter. Columns were left for 1 h at RT, following which the solution was eluted by spinning the columns for 1 min at 110 xg. The final step of conditioning consisted of washing the columns with 100 µL of water, followed by a new step of centrifugation. The columns were then equilibrated by adding 100 µL of 25mM ammonium format in 85% acetonitrile and centrifuged for 1 min at 110 xg.

In the meantime, samples were placed in a miVac DNA concentrator for 1 h to allow cell pellets to dry and solvents to evaporate. Tryptic peptides were afterwards resuspended in 100 µL of 25 mM ammonium format in 85% acetonitrile and added to the column, which was centrifuged for 1 min at 110 xg. After discarding the supernatant, columns were washed by adding 50 µL of 25mM ammonium format in 85% acetonitrile. Samples were eluted by placing the columns onto new 2mL microcentrifuge tubes, and by adding 30 µL of 15 mM formic acid, followed by a 1 min centrifugation at 110 xg. Elution was repeated one more time to maximize sample yield.

### 2.8.4 Relative quantification by SWATH acquisition

Samples were analysed using SCIEX (Framingham, Boston, MA, USA) TripleTOF 6600 instrumentation coupled on-line to an Ultimate 3000 HPLC (Dionex), according to the standard operating procedures of the Stoller Biomarker Discovery Centre (Manchester, UK). The analytical column used was an Eksigent LC system with a nanoLC 400 autosampler and nanoLC 425 pump module. Peptides were eluted using a gradient from a 98% water, 2%
acetonitrile and 0.1% formic acid buffer to an 80% (v/v) acetonitrile, 20% (v/v) water, 0.1% (v/v) formic acid buffer over 120 minutes at 0.3uL/min. Nuclear and cytosolic samples from control and RUNX1-ETO cells were analysed using data-independent acquisition (DIA) method. Raw spectral data files were analysed using openSWATH (version 2.0.0) using the Pan Human Library (Rosenberger et al., 2014) followed by scoring and filtering using pyProphet (version 0.18.3). Resultant output files were combined using the MSproteomicstools feature alignment script. Further filtering and protein abundance summarisation and normalisation were performed in R (version 3.4.1) using the packages SWATH2stats and MSstats from Bioconductor (release 3.5). Statistical significance was analysed according to section 2.8.5, and pathway analysis according to section 2.9.2.

2.8.5 SWATH-MS statistical analysis

Partek® Genomics Suite® (v. 7.0, St. Louis, USA) was used to perform statistical analysis on the protein expression dataset. To do this, detected proteins were partitioned according to their sub-cellular localisation (cytoplasmic or nuclear). Subsequently, only proteins that were commonly detected in all 3 biological replicates (for each condition) were used to perform statistical analysis. An ANOVA was used to determine statistically dysregulated protein changes between RUNX1-ETO expressing and control cells, and a cut-off value of p < 0.05 was applied to the analysis. Mean values of protein expression were used for calculation of fold-change (FC). No fold-change cut offs were applied in this analysis. Cytosolic and nuclear proteins were then combined to generate a new list for pathway and network analysis, performed using Metacore™, according to (2.10).

2.9 Analysis of existing mRNA expression data
2.9.1 Analysis of microarray data of RUNX1-ETO induced changes in mRNA expression in CD34⁺ HSPC

Previously, the effects of aberrant transcriptional activity of RUNX1-ETO were analysed by Affymetrix® microarray gene expression in a study performed by Tonks et al (Tonks et al., 2007). Dysregulated genes arising from this analysis were downloaded from https://www.nature.com/articles/2404961#supplementary-information, Supplementary Table 2, accession number E-MEXP-583. Dysregulated probe-sets were converted to Entrez Gene ID using the DAVID tool (Database for Annotation, Visualization and Integrated Discovery ver.
6.7) (Huang et al., 2009a). Pathway and Interactome analysis were performed using Metacore™, as described below (2.10).

2.9.2 Publicly available databases

2.9.2.1 Bloodspot

mRNA expression in human haematopoietic cells for the different targets mentioned in this study was accessed through the online database tool Bloodspot (Bagger et al., 2016) using the ‘Normal haematopoiesis with AML” dataset, and the corresponding probesets (URL: http://servers.binf.ku.dk/bloodspot/) (GSE42519; (Rapin et al., 2014)).

2.9.2.2 cBioPortal

In addition, mRNA expression was also assessed in AML patients through cBioPortal (Cerami et al., 2012; Gao et al., 2013) (URL: https://cbiportal.org/). Analysis was performed based on the Cancer Genome Atlas (TCGA), NEJM 2013 dataset, comprising of whole-genome or whole-exome sequencing of 163 adult de novo AML samples (Ley et al., 2013). For patient survival analyses, patient samples were excluded based on (i) untreated patients and/or (ii) diagnosed with APL. The remaining patients were stratified according to mRNA expression of the gene of interest into the upper (high expression) or lower (low expression) quartiles. Furthermore, patients were analysed based in clinical attributes associated with high and low mRNA target expression, including FAB classification, cytogenetic or molecular abnormalities, and overall-survival. All mRNA expression data within cBioPortal is represented as RNASeq RSEM.

2.10 Meta-analysis of mRNA and proteomic data

Metacore™ (Clarivate™ Analytics, UK), is a web-based bioinformatics tool used for the analysis of multiple omics data, including microarray, RNASeq and proteomics datasets. This analysis is based on published peer-reviewed literature combined into a curated knowledge database, MetaBase, covering over 2700 scientific journals. In all the studies described below, the following elements were selected for the input dataset: for the mRNA analysis, Entrez Gene ID identifiers were used, as well as their corresponding fold-change values; for the protein analysis, the protein name and the corresponding p-value were used.
Ingenuity Pathway Analysis (IPA) is another web-based software application, alike Metacore™. Similarly, it allows the analysis and interpretation of microarray and proteomics data, by performing not only pathway analysis, but also by identifying key regulators of a specific dataset, predicting downstream effects, and providing potentially useful target data. IPA is based on the Ingenuity Knowledge Database, derived from manually curated literature from over 300 scientific journals, and abstracts from more than 3600 peer-reviewed journals. The workflow used to analyse both datasets has been described in Figure 2.8.

2.10.1 Pathway analysis using Metacore

The Metacore™ Pathway Analysis tool was used to perform an Enrichment analysis of the dataset of dysregulated genes/proteins. The background of the analysis of the microarray dataset, defined as the base gene list in which all analyses was performed, was defined as “Affy // U133A”, corresponding to the Microarray GeneChip® previously used for the mRNA analysis. The resulting pathways were ranked according to their p-value, based on a hypergeometric distribution. Metacore™ also uses p-value calculation to evaluate a network’s relevance to Gene Ontology biological processes classification. The p-value is calculated as follows:

\[
pVal(r, n, R, N) = \frac{\min(n, R)}{N!} \sum_{i=\max(r, R-n)}^{\min(n, R)} \frac{R! n! (N-R)! (N-n)!}{i! (R-i)! (n-i)! (N-R-n+i)!}
\]

The set of all nodes in the Metacore™ database of interactions represent the “Global network” of size \( N \). Respectively, a set of nodes corresponding to our dataset is \( R \) and set of nodes in the network under evaluation is \( n \), considering that \( r \) nodes in \( n \) turned out to be “marked” by association with user’s list. \( R \) and \( r \) could also represent nodes marked by their association with some other list, such as “process” category from Gene Ontology. For the evaluation of statistical significance, Metacore™ considers the null hypothesis which states that the subsets \( R \) and \( n \) are independent and, therefore, the size of their intersection follows the hypergeometric distribution. The alternative hypothesis states that there is positive correlation between the subsets. Based on these assumptions, Metacore™ is able to calculate a p-value as the probability that intersection of two randomly selected sub-sets of \( N \) would have having the size of \( n \) or larger.
Figure 2.8 - Strategy for the analysis of differentially expressed genes/protein to study the effects of RUNX1-ETO overexpression in human CD34+ HSPC

Diagram illustrating the strategy employed to analyse the consequences of RUNX1-ETO expression in CD34+ HSPC. Differentially expressed genes had been previously filtered according to Tonks et. al using p-value < 0.05 and a fold-change cut-off of 1.5 (Tonks et al., 2007); differentially expressed proteins were identified by applying a cut-off of p-value < 0.05. Enrichment analysis was performed using Metacore and IPA (for mRNA analysis); TF overconnectivity was performed using Metacore.
The p-value was set with a False Discovery Rate (FDR) of <0.05 (no more than 5% false positives) and used to estimate the probability of a random intersection between the input probe set IDs with ontology gene IDs within the Metacore™ database. The lower p-value, the higher relevance of the gene ontology IDs to the dataset, showing as a higher rating for the given probe sets of the corresponding pathway.

2.10.2 Interactome analysis using Metacore

The Metacore™ Interactome tool was used to identify genes/proteins overrepresented within the given datasets. This tool provides an estimation of the level of interconnectivity within the input dataset based on published literature. This algorithm estimates statistically significant interactions in the set, and enrichment of the dataset according to certain genes/proteins.

In this study, to identify statistically significant TF responsible for the regulation of other genes/proteins, the algorithm “Interactome Transcription Factor” was used. The results were ranked by p-value using the basic formula for a hypergeometric distribution, as described in 2.10.1. Metacore™ provides an alternative statistical analysis based on the level of over- and under-connectivity of each TF, designated z-score. The z-score measures the difference between the obtained number of proteins and the expected average number of proteins corresponding to genes expressed in units of standard dispersion, and is calculates as below:

\[
Z - \text{score} = \frac{r - n \frac{R}{N}}{\sqrt{n \left(\frac{R}{N}\right) \left(1 - \frac{R}{N}\right) \left(1 - \frac{n - 1}{N - 1}\right)}}
\]

Where \( r \) represents the number of proteins derived from the current protein list that have interactions with given protein; \( R \) represents the total number of proteins in the GeneGo global network that have interactions with the given protein; \( n \) refers to the total number of proteins in given protein list and \( N \) is based on the total number of proteins in the GeneGo global network. All analyses were performed considering an FDR of < 0.05.
2.10.3 Pathway analysis and identification of transcription factors using Ingenuity Pathway Analysis

Similarly, IPA was used to analyse the microarray dataset. For the microarray analysis, the background was defined before any analysis was made as Affy // U133A. To identify TF, Upstream regulator analysis algorithm was used, allowing to predict the regulators responsible for the changes in gene expression. Once more, genes were ranked according to their p-value, calculated using a Right-Tailed Fisher’s Exact Test, reflecting the likelihood that the association or overlap between a set of significant molecules from our dataset and a given process/pathway/transcription neighbourhood is due to random chance. The smaller the p-value the less likely that the association is random.

2.11 Statistical analysis

Statistical analysis was performed using a paired sample t-test, or one-way ANOVA with Bonferroni’s multiple test correction, unless otherwise stated. Values of p<0.05 were considered statistically significant. All statistical analyses were conducted using GraphPad Prism (ver. 8) (GraphPad Software, California, USA). Details regarding the statistical tests used are labelled in figure legends.
Chapter 3

RUNX1-ETO Induces Changes in mRNA and Protein Expression of Transcription Factors
3.1 Introduction

AML is characterised by a block in terminal myeloid cell differentiation resulting in the accumulation of non-differentiated leukaemic cells in the BM and PB. It is widely accepted that abnormal activity of critical genes, such as TFs, which are involved in haematopoietic cell development, survival, proliferation, and maturation contribute to the development of this disease (Saultz and Garzon, 2016; De Kouchkovsky and Abdul-Hay, 2016). Chromosomal translocations are one of the most common cytogenetic abnormalities in AML and result in the expression of abnormal oncogenic fusion proteins (1.2.2). These translocations almost invariably consist of at least one TF protein (or a shortened truncated form) frequently involved in the regulation of myeloid cell differentiation (Lee et al., 2006). Consequently, AML-associated fusion proteins function as aberrant transcriptional regulators with the potential to interfere with the normal processes of haematopoietic differentiation.

One of the most common translocations in AML is the t(8;21), which results in the expression of the oncogenic fusion protein RUNX1-ETO (Rowley, 1990; Reikvam et al., 2011). Expression of this protein is the product of the fusion of RUNXI (chromosome 21), a critical TF in haematopoietic development, with ETO (chromosome 8) (1.3.3). To investigate the in vivo mechanisms through which RUNX1-ETO contributes to leukaemogenesis, Okuda et al., generated a murine model containing a single allele of RUNX1-ETO whose expression was regulated by the endogenous transcriptional regulatory elements of murine RUNX1 (Okuda et al., 1998). The study showed that RUNX1-ETO not only disrupts normal RUNX1-mediated transcriptional activity, but also contributes to the generation of abnormal haematopoietic progenitor cells. In addition, mice that are heterozygous for the RUNX1-ETO allele are foetally incompatible with life due to haemorrhaging in the central nervous system and exhibit a severe block in foetal liver haematopoiesis (Yergeau et al., 1997). In vitro studies using an experimental model based on normal human primary HSPC (CD34+ cord blood derived) enabled the analysis of the effects of RUNX1-ETO expression as a single abnormality on normal human haematopoiesis. These studies showed that ectopic expression of this fusion protein was able to inhibit the development of haematopoietic cells, promote the growth of immature blood cells and increase self-renewal – hallmarks of AML (Tonks et al., 2003; Tonks et al., 2004; Schnerch et al., 2012; Mulloy et al., 2002).
Increasing data has shown that developmental abnormalities observed in the presence of the RUNX1-ETO fusion protein disturbs one or more transcriptional processes that regulate haematopoietic development (1.3.3.4). To understand the mechanism through which this fusion protein blocks haematopoietic development, transcriptomic analysis was performed using DNA oligonucleotide microarrays (Tonks et al., 2007). Using the human cord blood-derived model, Tonks et al., found 380 genes to be dysregulated in the presence of the RUNX1-ETO fusion protein when compared to control cells. The analysis identified several dysregulated genes known to be implicated in AML and in the arrest of cell development, including PU.1 and CEBPA; both widely described in the literature as master regulators of haematopoiesis (Nerlov and Graf, 1998; Koschmieder et al., 2005). Furthermore, gene changes were also observed throughout myeloid cell development, including monocytic, granulocytic and erythroid lineages. These dysregulated genes included Sox4, CD200 and γ-catenin, suggesting that the effect of RUNX1-ETO on gene expression in human haematopoietic progenitor cells in highly context-dependent (Tonks et al., 2007; Morgan et al., 2013; Damiani et al., 2015; Lu et al., 2017).

Further studies have since been performed to try to understand the molecular mechanism through which this fusion protein contributes to leukaemia development, by analysing the transcriptome or proteome of RUNX1-ETO expressing cells. Ptasinska et al. performed a genome-wide association study in t(8;21) AML patients coupled with RUNX1-ETO expressing cell lines (Kasumi-1 and SKNO-1) to determine the alterations in the epigenetic landscape caused by the expression of this protein (Ptasinska et al., 2012). The authors were able to demonstrate that the expression of genes involved in haematopoietic differentiation and self-renewal is controlled by RUNX1-ETO; these genes also included the well-established TF C/EBPα, PU.1 and NFE2. The same study showed that knockdown of RUNX1-ETO using siRNA, led to an epigenetic reprogramming and a genome-wide redistribution of RUNX1 binding, influencing processes involved in myeloid differentiation, proliferation, and self-renewal. Furthermore, RUNX1-ETO knockdown in primary t(8;21) patient cells and cell lines resulted in the upregulation of genes involved in myeloid differentiation, whilst downregulated genes were shown to be involved in cell proliferation and cell cycle progression, resulting in the inversion of the typical t(8;21) phenotype. An additional study performed by Martinez-Soria et al. using an RNA interference (RNAi) approach identified Cyclin D2 (CCND2) as a crucial transmitter of RUNX1-ETO-driven leukaemic propagation (Martinez-Soria et al.,
Moreover, Singh et al. identified RUNX1-ETO mediated changes in protein expression by inducing the expression of this oncoprotein in a Tet-off-inducible U937 cell line coupled with mass spectrometry analysis (Singh et al., 2010). The authors showed that in these cells, the protein profile is drastically changed due to the expression of RUNX1-ETO, and were able to identify several changing proteins, including NM23 and HSP27. These specific changes were also observed in AML patient samples (Singh et al., 2010).

However, the above studies have primarily focused on transformed AML cell lines, and in general gene/protein expression, using an unsupervised analysis of transcriptomic/proteomic changes. Given that transcriptional regulation is likely mediated by changes in TF expression, this study will re-analyse transcriptomic data obtained from microarray studies performed by Tonks et al., and use a supervised approach to focus specifically on changes observed in TF. Furthermore, whilst mRNA abundance can be used as a strategy for target identification, it is not a powerful predictor of protein expression. For this reason, this study will use both TF transcriptomics and proteomics expression data to identify potential targets of the phenotype observed in t(8;21), by performing quantitative MS. In addition, no study has specifically examined the cytosolic and nuclear protein expression profile of CD34+ HSPC upon the expression of the RUNX1-ETO fusion protein, focusing instead on changes to total protein expression. Therefore, MS analysis will be employed on cytosolic and nuclear fractions of RUNX1-ETO expressing human CD34+ HSPC to identify, firstly, changes in TF expression, and secondly, aberrant subcellular protein localisation that could help explain the phenotypic changes observed in AML t(8;21). Analysis of transcriptomic and proteomic data sets will be used to investigate similarities between differentially expressed genes and proteins, as a result of the expression of the RUNX1-ETO fusion protein. However, it is important to take into consideration the well-recognised delay between mRNA expression and protein production, alongside the short half-life of mRNA against the extended half-life of some proteins. This understanding leads to the expectation that, despite both analysis being performed following three days of HSPC transduction, some measure of discrepancy will inherently remain.

### 3.2 Hypothesis and Aims

I hypothesise that the RUNX1-ETO fusion protein dysregulates TF that mediate the block in haematopoietic cell differentiation. The overall aim of this chapter is to identify changes in
the expression of TFs relevant in the context of t(8;21) AML by analysing the transcriptome and proteome of human CD34\(^+\) HSPC expressing RUNX1-ETO (Figure 3.1). The specific objectives are:

**To determine RUNX1-ETO induced changes in TF mRNA expression using normal human CD34\(^+\) HSPC.**

Pre-existing Affymetrix transcriptome data relating to RUNX1-ETO-induced changes in mRNA gene expression in human CD34\(^+\) HSPC is available (Tonks *et al.*, 2007). This dataset will be analysed to identify changes in TF mRNA abundance in CD34\(^+\) HSPC expressing RUNX1-ETO, compared to normal CD34\(^+\) HSPC controls. This will be achieved using pathway analysis software Metacore™ and IPA®, which are manually curated databases used to analyse omics data. In addition, identification of dysregulated genes will be compared to publicly available transcriptomic data from AML patients to identify clinically relevant changing TF mRNA.

**To determine RUNX1-ETO induced changes in TF protein expression using normal human CD34\(^+\) HSPC.**

To quantify significant changes in TF protein expression as a result of RUNX1-ETO expression, SWATH-MS analysis (2.8) will be performed on cytosolic and nuclear extracts of human CD34\(^+\) HSPC expressing RUNX1-ETO and compared to control. This will allow the identification of TFs inappropriately localised in either subcellular compartment, thus interfering with normal cellular processes. Pathway analysis using Metacore™ will be used to identify mis-regulated proteins arising from the expression of RUNX1-ETO. Furthermore, examination of the cytoplasmic to nuclear protein expression ratio in control and RUNX1-ETO CD34\(^+\) HSPC will allow the identification of differential protein expression with regard to these subcellular compartments.
Chapter 3

Figure 3.1 – Strategy for the identification of changes in gene/protein expression arising from the expression of the fusion protein RUNX1-ETO

Flow diagram showing the experimental strategy used to identify key target genes/proteins in RUNX1-ETO expressing cells, compared to control, for further study. Human CD34+ HSPC were isolated from neonatal cord blood and transduced with retrovirus co-expressing either RUNX1-ETO and GFP or GFP alone (control). Transduced cells were assessed for transduction efficiency and RNA/proteins extracted on day 3 of cell culture. Samples were analysed by performing Affymetrix Microarray (n=4) or SWATH-MS (n=3) analysis to determine mRNA and protein expression, respectively. Differentially expressed genes/proteins were uploaded into pathway analysis software for pathway analysis and target identification (Part 2), with a specific focus on TF. Additionally, aberrantly localised proteins were analysed in the cytoplasmic and nuclear compartments. Following identification of a target(s) of interest, these was used to perform overexpression and/or KD functional studies in CD34+ HSPC and AML cell lines (Part 3; Chapter 4, Chapter 5).
3.3 Results

3.3.1 Transcriptional dysregulation mediated by RUNX1-ETO

3.3.1.1 RUNX1-ETO expression in HSPC alters TF gene expression and is associated with GO processes that regulate transcription

Previously, the effects of RUNX1-ETO expression on the cell’s transcriptome was analysed in cord blood-derived CD34+ HSPC by contrasting control against a RUNX1-ETO overexpression vector (Tonks et al., 2007). However, the analysis did not consider an assessment of changes to cellular pathways or gene expression networks but rather focused on highest fold changing transcripts. A more focused approach based on changes observed in TF coupled with ‘Pathway’ analysis will refine this analysis, leading to the identification of novel changes in transcriptional changes.

To understand the biological implications of RUNX1-ETO expression, functional ‘Enrichment Analysis’ was performed using Metacore™. Based upon a fold change value of ±1.5, significantly changed probe-sets (465) were matched to 380 unique genes and subsequent Gene Ontology (GO) analysis was performed. Several GO Processes were found to be significantly disrupted as a result of the expression of RUNX1-ETO, including ‘positive and negative cell regulation processes’ and ‘cell proliferation’ (Figure 3.2A), consistent with the hallmarks of AML (Niebuhr et al., 2008). This approach is effective in identifying processes that are involved in RUNX1-ETO-induced gene dysregulation; however, it does not consider the directionality of expression changes, which is an inherent limitation of overrepresentation analysis, such as the one presented in this study. Nevertheless, this is a well-recognised and widely utilised technique for identifying the most broad processes which have been perturbed in this setting.

Alternative analysis examining Metacore™ curated Pathways Maps identified several disrupted cell developmental processes, of which ‘transcription regulation of granulocytic development’ was found to be significantly over-represented, supporting previous observations that RUNX1-ETO expression impairs haematopoietic differentiation (Figure 3.2B) (Niebuhr et al., 2008; Tonks et al., 2004). Additionally, ‘immune response’ related pathways were found to be disrupted as a consequence of RUNX1-ETO expression (Vangala et al., 2003; Choi et al., 2006).
Figure 3.2 – Functional Enrichment Analysis of mRNA changes observed in human CD34+ HSPC expressing RUNX1-ETO

Enrichment Analysis of the 380 significantly dysregulated genes (Tonks et al., 2007) in CD34+ HSPC expressing RUNX1-ETO, when compared to control, using Metacore™. Top 10 Significantly dysregulated (A) GO processes and (B) Pathway Maps, ranked by - Log (p-Value). Processes can further be clustered (1) Developmental process, (2) Immune response-related process and (3) Asthma-related process. A False Discovery Rate (FDR) of 0.05 was applied to both analyses.
Changes observed in the transcriptome of RUNX1-ETO expressing CD34+ HSPC, when compared to control, correlate with the phenotypical features observed in vitro and in t(8;21) AML patients, in which cells display an arrest in normal development, as well as corroborating that multiple physiological processes are disrupted as a consequence of the expression of this fusion protein.

To determine the proportion of functional classes of genes dysregulated as a consequence of RUNX1-ETO expression, a classification for all 380 dysregulated genes was performed using Metacore™ (Figure 3.3A). In order to statistically support these observations, an Enrichment Analysis by Protein Function using Metacore™ was performed. TF were identified as the most over-represented group of genes (determined by Z-score) whose expression was significantly disrupted by RUNX1-ETO expression (48 network objects, corresponding to 44 TFs), followed by receptor ligands and general enzymes (Figure 3.3B). Changes in TF expression supports the pathway analysis described previously in which transcriptional and development processes are significantly associated with RUNX1-ETO expression. Table 3.1 shows that 22 TFs were found to be upregulated in RUNX1-ETO cells compared to control, whilst 22 were found to be downregulated under the same conditions.

Since developmental disruption is likely mediated through changes in TF regulation, we attempted to use the 380 dysregulated gene list to determine the TF which regulated the highest number of genes within this list. As different pathway analysis programs can rely on different databases, two parallel approaches were undertaken using Metacore™ and IPA® to maximise the resulting targets. Initially, Metacore™, in combination with its built-in ‘Interactome Transcription Factors’ algorithm was used to identify the upstream TF(s) responsible for the changing mRNA expression of downstream genes. This algorithm achieves this by determining the density and statistical significance of the interactions of each TF with other genes from the dataset. This approach resulted in the identification of 25 TFs (Table 3.1). The most significantly connected TF included PU.1, which is a master regulator of myeloid cells (Nerlov and Graf, 1998; Koschmieder et al., 2005); ERG, a member of the erythroblast transformation-specific family, responsible for regulating haematopoiesis and the differentiation and maturation of megakaryocytic cells (Knudsen et al., 2015); and CEBPA, a coordinator of proliferation arrest and differentiation of myeloid progenitors (Koschmieder et al., 2005).
Figure 3.3 – TF were identified as the most significantly represented functional class due to RUNX1-ETO expression in CD34+ HSPC

(A) Pie chart indicating the classification of the differentially expressed genes using Metacore™. 380 genes were found to be dysregulated in RUNX1-ETO expressing cells when compared to control. Numbers after each class represent the number of genes identified. (B) Table indicating the classification of ‘network objects’ dysregulated upon the expression of RUNX1-ETO. Metacore™ identified TF as the most significantly changed class of genes, followed by receptors. However, it failed to identify over 50% of all ‘network objects’ present in the dataset. (Actual - number of network objects from the dataset(s) for a given protein class; n - number of network objects in the activated dataset; R - number of network objects of a given protein class in the complete database or background list; N - total number of network objects in the complete database or background list.). As z-scores can be negative, the higher z-scores correspond to shorter ‘distance’ and most meaningful results.
Table 3.1 – Identification of RUNX1-ETO induced TF dysregulation in CD34+ HSPC

TF expression had been previously determined by performing microarray analysis on CD34+ HSPC transfected either with a RUNX1-ETO retroviral vector, or a control plasmid, on day 3 of cell culture. To identify significantly overconnected genes within the dataset, the Metacore™, ‘Interactome Transcription Factors’ and the IPA® ‘Upstream Regulators’ algorithms were used. ✓ and X denotes if the gene was identified or not (respectively) by the algorithm used. The results from both approaches were used to generate a list of 35 genes for subsequent analysis. p-values were calculated by both software as described in 2.10.

RNASeq data from the TCGA 2013 dataset (Ley et al., 2013) (GSE13159) was used to determine mRNA expression in AML M2 patients, divided into t(8;21), expressing the fusion protein RUNX1-ETO, and non-t(8;21) patients. mRNA expression in normal human HSC were obtained from (Rapin et al., 2014) (GSE42519). Each gene has a corresponding Affymetrix ID unique to the Hu133A GeneChip® (Affymetrix®).

Fold-change (FC) values represent the regulation of each gene when compared to control cells (green – upregulated; red – repressed, when compared to control cells). (↑) denotes upregulated or (↓) downregulated gene expression according to each analysis.

* Genes are represented by multiple Affymetrix IDs. For genes with multiple Affymetrix IDs, values of fold-change were averaged. (n/a denotes not applicable; CDS denotes coding sequence)

<table>
<thead>
<tr>
<th>Gene</th>
<th>FC</th>
<th>Affymetrix ID</th>
<th></th>
<th>Metacore™</th>
<th>IPA®</th>
<th>Literature review</th>
<th>AML M2 t(8;21) vs non-t(8;21)</th>
<th>AML t(8;21) vs. HSC</th>
<th>CDS (kbp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EGR2</td>
<td>3.69</td>
<td>205249_at</td>
<td>✓</td>
<td>3.65 x 10^-13</td>
<td>✓</td>
<td>2.61 x 10^-4</td>
<td>↓</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>HOXB6</td>
<td>3.05</td>
<td>205366_s_at</td>
<td>X</td>
<td>-</td>
<td>X</td>
<td>-</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>ID1</td>
<td>2.99</td>
<td>208937_s_at</td>
<td>X</td>
<td>-</td>
<td>✓</td>
<td>2.89 x 10^-4</td>
<td>(Wang et al., 2015a)</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>CREM*</td>
<td>2.75</td>
<td>209967_s_at</td>
<td>✓</td>
<td>2.39 x 10^-4</td>
<td>✓</td>
<td>5.53 x 10^-3</td>
<td>-</td>
<td>↓</td>
<td>n/a</td>
</tr>
<tr>
<td>ARID5B</td>
<td>2.59</td>
<td>212614_at</td>
<td>X</td>
<td>-</td>
<td>✓</td>
<td>4.11 x 10^-2</td>
<td>-</td>
<td>↑</td>
<td>↑~2.9</td>
</tr>
<tr>
<td>KLF6</td>
<td>2.58</td>
<td>208961_s_at</td>
<td>✓</td>
<td>4.38 x 10^-4</td>
<td>X</td>
<td>-</td>
<td>(DeKelver et al., 2013a)</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>
**Gene** | **FC** | **Affymetrix ID** | **Metacore™** | **IPA®** | **Literature review** | **AML M2 t(8;21) vs non-t(8;21)** | **AML t(8;21) vs. HSC** | **CDS (kbp)**  
--- | --- | --- | --- | --- | --- | --- | --- | ---  
ZBTB20* | 2.51 | 222357_at | X | - | ✓ | 1.01 x 10^-2 | - |  
AHR | 2.42 | 205749_at | ✓ | 8.84 x 10^-10 | X | - | (Ly et al., 2019) | n/a | n/a | n/a  
NFKB2 | 2.12 | 207535_s_at | ✓ | 2.56 x 10^-7 | ✓ | 3.77 x 10^-6 | (Nakagawa et al., 2011; Zhou et al., 2012) | n/a | n/a | n/a  
RUNX1 | 1.91 | 1440878_at | ✓ | 6.9 x 10^-34 | ✓ | 1.57 x 10^-13 | n/a | n/a | n/a  
HLF | 1.88 | 204753_s_at | X | - | X | - | n/a | n/a | n/a | n/a  
IRF9 | 1.78 | 203882_at | X | - | ✓ | 4.65 x 10^-3 | - |  
MXI1 | 1.73 | 202364_at | X | - | ✓ | 2.73 x 10^-4 | (Weng et al., 2017) | n/a | n/a | n/a  
SOX4 | 1.65 | 201416_at | ✓ | 1.62 x 10^-4 | ✓ | 2.09 x 10^-4 | (Nafria et al., 2020; Fernando et al., 2017) | n/a | n/a | n/a  
IRF7 | 1.64 | 208436_s_at | ✓ | 3.06 x 10^-4 | ✓ | 9.07 x 10^-4 | - | No change |  
E2F5 | 1.62 | 221586_s_at | ✓ | 1.38 x 10^-5 | ✓ | 1.40 x 10^-4 | (Huang et al., 2020) | n/a | n/a | n/a  
TCF7L2 | 1.62 | 216035_s_at | ✓ | 6.26 x 10^-20 | ✓ | 2.46 x 10^-3 | (Saenz et al., 2020; Zhao et al., 2019) | n/a | n/a | n/a  
ERG | 1.61 | 213541_s_at | ✓ | 2.31 x 10^-18 | ✓ | 1.28 x 10^-7 | (Martens et al., 2012; Mandoli et al., 2016) | n/a | n/a | n/a  
SMARCA1 | 1.55 | 203874_s_at | X | - | X | - | n/a | n/a | n/a | n/a  
HBP1 | 1.54 | 209102_s_at | ✓ | 1.10 x 10^-2 | X | - | - |  
HOXA5 | 1.54 | 213844_at | X | - | X | - | n/a | n/a | n/a | n/a  
ZNF217 | 1.5 | 203739_at | X | - | ✓ | 3.79 x 10^-3 | - | No change |  
EGR3 | -1.52 | 206115_at | ✓ | 3.69 x 10^-3 | X | - | (Cheng et al., 2015) | n/a | n/a | n/a  
IRF8 | -1.52 | 204057_at | ✓ | 4.88 x 10^-8 | ✓ | 2.57 x 10^-9 | (Pogosova-Agadjanyan et al., 2013; Liss et al., 2021) | n/a | n/a | n/a
<table>
<thead>
<tr>
<th>Gene</th>
<th>FC</th>
<th>Affymetrix ID</th>
<th>Metacore™ Interactome p-value</th>
<th>Metacore™ Upstream Regulators p-value</th>
<th>IPA® p-value</th>
<th>Literature review</th>
<th>AML M2 (t(8;21) vs non-t(8;21))</th>
<th>AML t(8;21) vs. HSC</th>
<th>CDS (kbp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>JUND</td>
<td>-1.53</td>
<td>203752_s_at</td>
<td>✓</td>
<td>✓</td>
<td>8.5 x 10^{-13}</td>
<td>3.77 x 10^{-6} (Plasinska et al., 2019; Martinez-Soria et al., 2017)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>E2F8</td>
<td>-1.55</td>
<td>219990_at</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>(Steinauer et al., 2019; Jakobczyk et al., 2021)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>TSC22D1</td>
<td>-1.56</td>
<td>215111_s_at</td>
<td>X</td>
<td>X</td>
<td>-</td>
<td>(De Braekeleeer et al., 2014)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>CBFA2T3</td>
<td>-1.57</td>
<td>208056_s_at</td>
<td>✓</td>
<td>✓</td>
<td>1.39 x 10^{-2}</td>
<td>2.88 x 10^{-2} (Zhou et al., 2017; Deb and Somervaille, 2016)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>RARA</td>
<td>-1.57</td>
<td>203749_s_at</td>
<td>✓</td>
<td>X</td>
<td>1.58 x 10^{-5}</td>
<td>(Kawagoe et al., 2007; Liu et al., 2017)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>ID2</td>
<td>-1.58</td>
<td>213931_at</td>
<td>X</td>
<td>-</td>
<td>✓</td>
<td>(Plasinska et al., 2012; Melnick et al., 2000a)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>MYCN</td>
<td>-1.59</td>
<td>209757_s_at</td>
<td>X</td>
<td>-</td>
<td>✓</td>
<td>(Shimizu et al., 2000; Li et al., 2019)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>ZBTB16</td>
<td>-1.61</td>
<td>205883_at</td>
<td>✓</td>
<td>6 x 10^{-7}</td>
<td>✓</td>
<td>(Soria et al., 2002; Shimizu et al., 2000; et al.)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>C/EBPε</td>
<td>-1.64</td>
<td>214523_at</td>
<td>✓</td>
<td>1.06 x 10^{-6}</td>
<td>✓</td>
<td>(Adewale et al., 2003)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>ZFP36L2</td>
<td>-1.65</td>
<td>201367_s_at</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>CIAO1</td>
<td>-1.66</td>
<td>217501_at</td>
<td>X</td>
<td>-</td>
<td>X</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>RUNX3*</td>
<td>-1.69</td>
<td>204197_s_at</td>
<td>✓</td>
<td>9 x 10^{-10}</td>
<td>✓</td>
<td>*</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>GATA2</td>
<td>-1.76</td>
<td>209710_at</td>
<td>✓</td>
<td>2.25 x 10^{-30}</td>
<td>✓</td>
<td>(Sood et al., 2016; Loke et al., 2017)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>PU.1</td>
<td>-1.76</td>
<td>209454_s_at</td>
<td>✓</td>
<td>1.18 x 10^{-35}</td>
<td>✓</td>
<td>(Vangala et al., 2003; Eder et al., 2016)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>GATA1</td>
<td>-1.81</td>
<td>210446_at</td>
<td>X</td>
<td>-</td>
<td>✓</td>
<td>(Elagib et al., 2003; Choi et al., 2006)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>GFI1B</td>
<td>-1.94</td>
<td>208501_at</td>
<td>✓</td>
<td>5.46 x 10^{-5}</td>
<td>✓</td>
<td>(Möröy and Khandanpour, 2019)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>TEAD3</td>
<td>-2.09</td>
<td>209454_s_at</td>
<td>X</td>
<td>-</td>
<td>-</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>C/EBPα</td>
<td>-2.34</td>
<td>204039_at</td>
<td>✓</td>
<td>1.57 x 10^{-22}</td>
<td>✓</td>
<td>(Pabst et al., 2001; Grossmann et al., 2005)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Gene</td>
<td>FC</td>
<td>Affymetrix ID</td>
<td>Metacore™</td>
<td>IPA®</td>
<td>Literature review</td>
<td>AML M2 t(8;21) vs non-t(8;21)</td>
<td>AML t(8;21) vs. HSC</td>
<td>CDS (kbp)</td>
<td></td>
</tr>
<tr>
<td>------</td>
<td>-----</td>
<td>---------------</td>
<td>------------</td>
<td>------</td>
<td>------------------</td>
<td>-------------------------------</td>
<td>---------------------</td>
<td>-----------</td>
<td></td>
</tr>
<tr>
<td>KLF1</td>
<td>-2.59</td>
<td>210504_at</td>
<td>✓</td>
<td>1.05 x 10^-5</td>
<td>✓</td>
<td>8.47 x 10^-3</td>
<td>(Kuvardina et al., 2015; Mansoor et al., 2020)</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>NF-E2</td>
<td>-2.64</td>
<td>209930_s_at</td>
<td>✓</td>
<td>1.68 x 10^-7</td>
<td>✓</td>
<td>3.43 x 10^-5</td>
<td>(Ptasinska et al., 2012; Wang et al., 2010a)</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>
Other overconnected TF included \textit{EGR2} and \textit{JUND}, known for protecting cells from p53-dependent senescence and apoptosis, reflecting the diversity of genes and processes influenced by the RUNX1-ETO fusion protein (Morita \textit{et al.}, 2016; Hernandez \textit{et al.}, 2008).

The second approach used the IPA® algorithm ‘Upstream Regulators’. Similarly, this method aims to identify the cascade of upstream transcriptional regulators based on how many known downstream targets of each transcription regulator are represented in the dataset. This approach identified 35 upstream transcriptional regulators. However, transcriptional regulators refers to both TFs and regulators of transcription; after filtering non-relevant results, 31 TF were identified for further investigation, as shown in \textbf{Table 3.1}. The results obtained from both proprietary software approaches were combined, resulting in the generation of a list containing 35 overconnected TF, of which 22 were identified by both approaches, suggesting concordance between the underlying databases. Subsequent steps of analyses were based on the combined 35-gene list.

\textbf{3.3.1.2 Identification of aberrantly expressed TFs driving transcriptional change in AML}

To identify novel genes that could be involved in the arrest in cell differentiation observed upon the expression on the RUNX1-ETO fusion protein, subsequent analysis filtered the 35-TF based on key literature search criteria, including novelty, significance in terms of its role in AML t(8;21) and model system used. Several genes identified had already been extensively studied in AML t(8;21), including \textit{PU.1 (SPI1), CEBPA, NFKB2} and \textit{NFE2} (Vangala \textit{et al.}, 2003; Pabst \textit{et al.}, 2001; Zhou \textit{et al.}, 2015; Wang \textit{et al.}, 2010a; Jutzi \textit{et al.}, 2019). Genes that lacked the novelty factor were excluded from further analysis.

Subsequently, to determine whether the pattern of expression of the TF identified above correlated with mRNA expression observed in patients, this study examined the transcriptome of AML blast samples from AML M2 patients harbouring the (8;21) translocation, and compared them to non-t(8;21) patients. In order to do this, the publicly available TCGA dataset generated by RNA-sequencing was analysed for the mRNA expression pattern of each gene (Ley \textit{et al.}, 2013). Moreover, an additional analysis examined gene expression patterns in undifferentiated HSC and compared them to that of patients diagnosed with t(8;21) AML (Ley \textit{et al.}, 2013; Rapin \textit{et al.}, 2014). In order to refine the gene list, the dataset was filtered by only examining genes for which expression was concordant between analysed studies (genes with
no change in one of the studies, but same trend in the other as the CD34+ cord blood-derived model were still considered) (Table 3.1).

Furthermore, a feasibility analysis was performed to ensure downstream analysis was possible; for instance, genes with a coding sequence bigger than 3.5 kb were excluded due to an inability to overexpress it in our model system (Table 3.1). Based on all these factors, four gene targets were chosen for further study: ARID5B, IRF7, IRF9 and ZNF217. Given that the expression or function of these genes has not yet been described in AML, this study initially assessed their protein expression in RUNX1-ETO expressing CD34+ HSPC by western blot.

IRF7 is a master regulator of Type I IFN-dependant immune response, and its dysregulated expression has been described in prostate and breast cancer (3.4) (Yang et al., 2016; Tara et al., 2018). Increased expression of IRF7-regulated genes has also been associated with prolonged metastatic-free survival (Savitsky et al., 2010; Takaoka et al., 2008). Figure 3.4 shows IRF7 mRNA was 1.6-fold upregulated in CD34+ HSPC expressing RUNX1-ETO, when compared to control cells; however, no difference was observed at the protein level. Another member of the IRF family, IRF9 has been described as having a role in Type-I IFN-mediated cellular response and has been shown to stimulate the p53 pathway, preventing oncogene-induced malignant cell transformation and in inducing DNA damage-induced apoptosis (Savitsky et al., 2010; Takaoka et al., 2008). Even though other IRF genes have been linked to myeloid cell differentiation (Tamura et al., 2015b; Wang and Morse, 2009), this has not been described in AML. As shown in Figure 3.5, IRF9 mRNA was 2.7-fold upregulated in CD34+ RUNX1-ETO expressing cells compared to control, whilst a 3-fold increase in IRF9 protein expression was observed in the cytoplasm of CD34+ HSPC expressing the fusion protein RUNX1-ETO. A 1.5-fold increase in the expression of this protein in the nucleus of RUNX1-ETO expressing cells was also observed.

ARID5B has been identified for its role in embryonic development, cell type-specific gene expression and cell growth regulation (Patsialou et al., 2005). Its aberrant expression or mutations have been widely described in ALL and APL, however its role in AML remains unidentified. mRNA analysis of CD34+ HSPC expressing RUNX1-ETO showed a 4-fold upregulation ARID5B in these cells when compared to control (Figure 3.6). ARID5B protein was found to be detected in the nucleus of both RUNX1-ETO and control CD34+ HSPC; however, expression level was the same. Moreover, no ARID5B protein was detected in the cytoplasm of either control or RUNX1-ETO expressing CD34+ cells.
Lastly, ZNF217 has been widely described as having a crucial role in the development of multiple types of cancer, including breast, ovarian and prostate, and its increased expression has been linked to a poorer outcome in this disease (see 0) (Rahman et al., 2012; Rooney et al., 2004). Furthermore, ZNF217 has been implicated in the maintenance of an undifferentiated phenotype in glioma and breast primary cultures (Cohen et al., 2015). ZNF217 mRNA and protein is upregulated by 1.7 and >10 fold respectively in cells expressing the RUNX1-ETO fusion protein (Figure 3.7).

Additionally, SWATH-MS analysis was performed to determine the proteomic profile of RUNX1-ETO CD34+ HSPC, as compared to control (see 3.3.2 for a complete description of this analysis). Regarding these 4 targets, SWATH-MS was able to detect 3 of the potentially relevant proteins, but failed to identify IRF7, detectable through western blotting technique. Nevertheless, this analysis showed that the IRF9 protein was found to be exclusively expressed in the nuclear compartment of RUNX1-ETO cells, with no detectable expression in control CD34+ HSPC (Figure 3.8A). SWATH-MS was able to detect the ARID5B protein, previously found in the nucleus of both control and RUNX1-ETO CD34+ HSPC; however, this was identified as being expressed in the cytoplasm of these cells (Figure 3.8B). Lastly, the ZNF217 protein was detected as being overexpressed in the nucleus of RUNX1-ETO cells, as compared to control (Figure 3.8C). Even though both SWATH-MS and western blotting analysis were performed on identical samples, both techniques measured distinct protein expression levels, and were not concordant between each other. The discrepancies between SWATH-MS and western blotting results will be further discussed in section 0.

In summary, this study was able to identify four potentially relevant targets in the development of AML t(8;21) (Table 3.2). Even though several targets were found to be upregulated at the mRNA level, these changes did not equate to the corresponding protein expression. For this reason, additional methods for analysing the proteomic profile of these cells are necessary.
Figure 3.4 – RUNX1-ETO does not influence the expression of IRF7 in CD34+ HSPC

(A) IRF7 mRNA expression following retroviral transfection (day 3 of cell culture), in CD34+ HSPC transduced with a control or RUNX1-ETO vector (Tonks et al., 2007) (n=4) (Paired T-test; ns - not significant). (B) Western blot analysis of IRF7 expression in CD34+ HSPC transduced with either a control or a RUNX1-ETO vector, on day 3 of cell culture (n=1). The OCI-AML2 cell line was used as a positive control (PC). GAPDH was used as a loading control for cytosolic fractions; Histone H1 was used as a loading control for nuclear fractions. Both proteins were used as a verification of subcellular fraction. (C) Western blot quantification by densitometry analysis using ImageJ. Relative protein expression was determined by normalising each cytosolic and nuclear sample to GAPDH or Histone H1, respectively, and to each corresponding ‘control’ (n=1).
Figure 3.5 – RUNX1-ETO leads to the upregulation of IRF9 in CD34+ HSPC

(A) IRF9 mRNA expression following retroviral transfection (day 3 of cell culture), in CD34+ HSPC transduced with a control or RUNX1-ETO vector (Tonks et al., 2007) (n=4). Significant differences between RUNX1-ETO and control cells was analysed using paired t-test; * denotes p<0.05. (B) Western blot analysis of IRF9 expression in CD34+ HSPC transduced with either a control or a RUNX1-ETO vector, on day 3 of cell culture (n=1). The OCI-AML2 cell line was used as a positive control (PC). GAPDH was used as a loading control for cytosolic fractions; Histone H1 was used as a loading control for nuclear fractions. Both proteins were used as a verification of subcellular fraction. (C) Western blot quantification by densitometry analysis using ImageJ. Relative protein expression was determined by normalising each cytosolic and nuclear sample to GAPDH or Histone H1, respectively, and to each corresponding ‘control’ (n=1).
Figure 3.6 – RUNX1-ETO leads to the upregulation of ARID5B in CD34+ HSPC

(A) ARID5B mRNA expression following retroviral transfection (day 3 of cell culture), in CD34+ HSPC transduced with a control or RUNX1-ETO vector (Tonks et al., 2007) (n=4). Significant differences between RUNX1-ETO and control cells was analysed using paired t-test; ** denotes p<0.01. (Appropriate band is indicated by the arrow). (B) Western blot analysis of ARID5B expression in CD34+ HSPC transduced with either a control or a RUNX1-ETO vector, on day 3 of cell culture (n=1). The HeLa cell line was used as a positive control (PC). GAPDH was used as a loading control for cytosolic fractions; Histone H1 was used as a loading control for nuclear fractions. Both proteins were used as a verification of subcellular fraction. (C) Western blot quantification by densitometry analysis using ImageJ. Relative protein expression was determined by normalising each cytosolic and nuclear sample to GAPDH or Histone H1, respectively, and to each corresponding ‘control’ (n=1).
Figure 3.7 – RUNX1-ETO leads to the upregulation of ZNF217 in CD34+ HSPC

(A) ZNF217 mRNA expression following retroviral transfection (day 3 of cell culture), in CD34+ HSPC transduced with a control or RUNX1-ETO vector (Tonks et al., 2007) (n=4). Significant differences between RUNX1-ETO and control cells was analysed using paired t-test; * denotes p<0.05. (B) Western blot analysis of ZNF217 expression in CD34+ HSPC transduced with either a control or a RUNX1-ETO vector, on day 3 of cell culture (n=1). The K562 cell line was used as a positive control (PC). GAPDH was used as a loading control for cytosolic fractions; Histone H1 was used as a loading control for nuclear fractions. Both proteins were used as a verification of subcellular fraction. (C) Western blot quantification by densitometry analysis using ImageJ. Relative protein expression was determined by normalising each cytosolic and nuclear sample to GAPDH or Histone H1, respectively, and to each corresponding ‘control’ (n=1).
Figure 3.8 – RUNX1-ETO leads to the upregulation of protein targets, in CD34+ HSPC

SWATH-MS analysis was performed on CD34+ HSPC transduced with either a RUNX1-ETO overexpression retroviral vector, or a control plasmid. Cytosolic and nuclear proteins were extracted and quantified on day 3 of cell culture. Data indicates Mean ± 1SD (where applicable) (n=3). Null expression values refer to protein samples not being detected in the SWATH-MS run for each corresponding protein.
Table 3.2 – Summary showing the effects of RUNX1-ETO expression in CD34+ HSPC on gene and protein expression in prioritised targets

§ mRNA expression was compared between RUNX1-ETO and control CD34+ HSPC by microarray analysis, performed on day 3 of cell culture (adapted from (Tonks et al., 2007))

Φ Protein expression was compared between RUNX1-ETO and control CD34+ HSPC by performing western blot/SWATH-MS analysis on cytosolic and nuclear extracts, on day 3 of cell culture. (↑) (↓) denotes upregulation or downregulation of gene/protein in RUNX1-ETO cells, respectively, as compared to control HSPC.

* Fold-change was not calculated for these samples due to the fact that protein was not detected in the control samples. (n/a denotes not applicable; CDS denotes coding sequence)

<table>
<thead>
<tr>
<th>Gene / Protein</th>
<th>CDS (kbp)</th>
<th>mRNA expression in CD34+ HSPC §</th>
<th>Protein expression in CD34+ HSPC Φ</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Regulation</td>
<td>Loc.</td>
</tr>
<tr>
<td>IRF7</td>
<td>~ 1.5</td>
<td>↑ 1.6-fold</td>
<td>n/a</td>
</tr>
<tr>
<td>IRF9</td>
<td>~ 1.4</td>
<td>↑ 2.7-fold</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ARID5B</td>
<td>~ 2.9</td>
<td>↑ 4-fold</td>
<td>n/a</td>
</tr>
<tr>
<td>ZNF217</td>
<td>~ 3.2</td>
<td>↑ 1.7-fold</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Loc. – localisation within the cell; C – cytoplasm, N - nucleus
3.3.2 Quantitative SWATH-MS analysis reveals proteomic changes arising from RUNX1-ETO expression, in CD34+ HSPC

3.3.2.1 Generation of human CD34+ HSPC expressing RUNX1-ETO for SWATH-MS analysis

This study used a primary human CD34+ HSPC model to investigate the effects of RUNX1-ETO on the dysregulation of protein expression. Changes in global protein expression were quantitated using SWATH-MS, a data-independent acquisition (DIA) technique that allows comprehensive quantification of thousands of detectable analytes in a sample (2.8.5). To generate the samples for MS analysis, three replicate pairs of CD34+ HSPC from biologically distinct sources were infected with either a control or a RUNX1-ETO-overexpression vector (2.5.2), following which proteins were separated and enriched according to their subcellular localisation: cytoplasmic or nuclear. Following MS analysis, statistical analysis was used to identify significantly changed proteins as a consequence of RUNX1-ETO expression, compared to control. Metacore™ was used to perform pathway analysis and to identify targets of interest.

3.3.2.2 Assessment of CD34+ HSPC nuclear and cytosolic extracts shows sample purity and integrity

Proteomic analysis requires high quantities of extremely pure protein. As CD34+ HSPC account for such a small proportion of total cord blood and with the requirement to extract proteins by day 3 of culture, thereby limiting the growth potential, the method had to be adapted to maximise recovery where possible. To isolate sufficient quantities of nuclear proteins from transduced cells, it was decided that if sufficient infection rate was achieved (>60%) (Figure 3.9), cells would not be sorted, thereby avoiding the concern surrounding stress response to the sorting process and further cell loss. This does, however, raise two important limitations. Firstly, the existence of a GFP expression range suggests varied incorporation rates within CD34+ HSPC, which may suggest differing overexpression levels. In addition, the remaining 40% of cells are not transduced with the plasmid and do not experience overexpression. Despite this, the strict protein requirements for SWATH-MS necessitated such compromises. Following infection, cytosolic and nuclear proteins from RUNX1-ETO and control CD34+ HSPC were extracted. Prior to SWATH-MS analysis, the overall protein quality and fractionation efficiency of each sample was assessed by performing QC assays to ensure no cross-contamination between cellular compartments. Firstly, protein extract quality was
assessed using in gel SYPRO™ Ruby Protein Blot Stain (2.6.6). This analysis showed little or undetectable cytoplasmic contamination in the nuclear fraction, and undetectable nuclear contamination in the cytoplasmic samples (Figure 3.10A - Figure 3.11). In the nuclear samples, it was possible to identify the histone proteins (~30 kDa), exclusive to this cell compartment, whilst their presence in the cytoplasm was absent.

In addition, in gel protein staining showed no degradation of any sample, with a clear separation of proteins according to size. Moreover, the RUNX1-ETO fusion protein was solely detected in the nucleus of RUNX1-ETO-transduced CD34+ HSPC whilst no protein was detected in the control samples (Figure 3.10B). Based on these assays, the samples were appropriate for SWATH-MS and downstream analysis.

3.3.2.3 SWATH-MS analysis detected over 4000 proteins in CD34+ HSPC

This study initially examined ‘housekeeping’ proteins exclusively expressed within the cytoplasmic or nuclear compartment to further confirm fractionation efficiency. As expected, GAPDH protein was found almost exclusively in the cytoplasm of both control and RUNX1-ETO expressing cells, whilst Histone H1 was solely detected in the nuclear compartment (Figure 3.11).

The protein library provides coverage for 50.9% of all human proteins annotated by UniProtKB/Swiss-Prot, which accounts for approximately 13,000 proteins (Rosenberger et al., 2014). Unfortunately, as RUNX1-ETO is a fusion protein formed through abnormal translocations, it is not detectable in this setting; however, RUNX1 levels may be used as an approximate indicator for RUNX1-ETO levels. RUNX1 was expressed in the nucleus with levels significantly higher in RUNX1-ETO infected cells compared to control (Figure 3.12). It should be noted, this is not a definitive conclusion, as both endogenous RUNX1 and the RUNX1 portion of the fusion protein RUNX1-ETO can be detected.
Figure 3.9 – Viral infection of CD34+ cells. CD34+ HSPC cells were analysed for GFP expression on day 3 of cell culture.

Representative flow cytometry histograms showing percentage of HSPC expressing GFP in (A) Control (green); (B) RUNX1-ETO (red) transduced cells. Histograms were gated to exclude cell debris and dead cells based on FCS/SSC. Background auto-fluorescence was established using HSPC subjected to the equivalent retroviral infection procedure but in the absence of retrovirus (mock) (grey). Percentage within the black marker line represents proportion of cells showing fluorescence greater than background autofluorescence at < 0.01% (n=3).
Figure 3.10 – Quality control and characterisation of CD34⁺ HSPC overexpressing RUNX1-ETO

(A) Purity of the cytoplasmic (Cyt) and nuclear (Nuc) fractionated samples was assessed by immunoblotting 1 µg of sample onto a PVDF membrane stained with SYPRO™ Ruby Protein Blot Stain. (B) Overexpression of RUNX1-ETO protein in CD34⁺ cells transduced with RUNX1-ETO, when compared to control. Each lane was loaded with 10 µg of protein. GAPDH and histone H1 were used as a loading control and as a verification of subcellular fraction for cytosolic and nuclear samples, respectively. This blot was a reprobe of Figure 3.5B, hence the same GAPDH and Histone H1 control images were used. (C) Western blot quantification by densitometry analysis using ImageJ. Relative protein expression was determined by normalising each cytosolic and nuclear sample to GAPDH or Histone H1, respectively, and to each corresponding ‘control’ (n=1).
Figure 3.11 – GAPDH and Histone H1 protein quantitation by SWATH-MS supports acceptable subcellular fraction efficiencies for cytosol and nuclear protein isolation

SWATH-MS analysis performed on cytosolic and nucleus fractions of CD34+ HSPC allowed the detection of the GAPDH protein almost exclusively in the cell’s cytoplasm, whilst Histone H1 was only detected in the cell’s nucleus, on both control and RUNX1-ETO expressing cells. The Y-axis represents the protein expression value; the X-axis represents the different cell compartments analysed (n=3).
Figure 3.12 – RUNX1 is significantly overexpressed in RUNX1-ETO transfected CD34+ HSPC

SWATH-MS analysis showed RUNX1 upregulation in the nucleus of CD34+ HSPC transfected with a RUNX1-ETO retroviral vector, on day 3 of cell culture, as compared to the nuclear compartment of control cells. The Y-axis represents RUNX1 protein expression value (arbitrary units) (n=3). Significant differences between RUNX1-ETO and control cells was analysed using paired t-test; * denotes p<0.05.
Data acquisition, filtering, and annotation was performed at the University of Manchester by Professor Anthony Whetton’s group. Three approaches were undertaken in order to study protein dysregulation as a result of RUNX1-ETO expression, in CD34+ HSPC.

Firstly, proteins were separated according to sub-cellular localisation, cytoplasmic or nuclear. For each compartment, proteins were filtered to those that were detected in all biological replicates in both control and RUNX1-ETO expressing cells (3/3 replicates), resulting in 4,635 detectable proteins, of which 2,787 were identified in the cytoplasm and 1,848 in the nucleus (explored in 3.3.2.4 -3.3.2.7).

The second approach consisted of analysing proteins exclusively expressed in RUNX1-ETO cells, without detectable protein in control cells. This analysis is possible as SWATH-MS is unable to detect certain proteins, either because there is no expression, expression level is undetectable or due errors in the detection process (discussed in 3.4). To prevent the latter from happening and influencing the results, this analysis will take into consideration proteins that failed to be detected in all three biological replicates of either control or RUNX1-ETO. This allowed the identification of 52 aberrantly expressed proteins in RUNX1-ETO cells, from which 29 were localised in the cytoplasm and 23 in the nucleus (3.3.2.8).

Lastly, the third approach will analyse all proteins detected in both the cytoplasm and nucleus of control and RUNX1-ETO cells to identify proteins that could have suffered a shift in normal protein localisation as a result of RUNX1-ETO expression. This analysis will considerer the expression of 926 proteins, explored in 3.3.2.9.

Taken together, the above data demonstrates highly efficient sub-cellular fractionations in control or RUNX1-ETO expressing cells. In addition, SWATH-MS was able to quantify over 4,500 proteins suitable for downstream targets analysis.

3.3.2.4 Exploratory data analysis

To assess the relationship between control and RUNX1-ETO samples, two methods were employed. Firstly, principal component analysis (PCA) is a useful tool which allows the visualisation of patterns form large and complex datasets. Additionally, hierarchical clustering, consisting of a statistical method used to assign objects into groups, or clusters, allows the highlighting of patterns in the protein expression dataset, based on sample type (RUNX1-ETO
vs. control CD34+ HSPC). This analysis was based on protein expression according to their sub-cellular localisation (4,635 proteins – 2,768 cytosolic and 1,848 nuclear proteins).

Regarding cytosolic proteins, Figure 3.13A and Figure 3.14A show the clustering of RUNX1-ETO samples, indicating that these cells have a protein expression profile similar to each other, whilst the control samples cluster separately. Nuclear protein expression data, on the other hand, show a similarity between 2/3 RUNX1-ETO samples and, once more, no relationship between control samples.

Altogether, this analysis indicates a distinction between control and RUNX1-ETO-expressing CD34+ HSPC; however, control samples did not cluster together, indicating variable patterns of expression, along with sample variability, which might be sourced from the fact that samples were prepared from different cord blood donations. Moreover, the fact that RUNX1-ETO samples were shown to cluster together suggests that this is a result ‘oncogenic’ induction, considering that all samples were subjected to the same viral transduction. For this reason, a paired parametric statistical test with ANOVA was performed to determine statistically significant changing proteins, arising from the expression of the fusion protein RUNX1-ETO.

### 3.3.2.5 Statistical model used for data analysis

Following normalization (log2) of the protein expression dataset, a one-way ANOVA statistical test was employed to identify significantly changed proteins in the cytoplasm and nucleus of CD34+ HSPC, arising from the expression of the fusion protein RUNX1-ETO. Uncorrected p-values were considered as this study represents an exploratory investigation due to low sample size. An uncorrected cut-off p-value of < 0.05 was applied to generate two lists of significantly changing cytosolic and nuclear proteins between CD34+ HSPC expressing RUNX1-ETO and control cells (Supplementary Table 2). This resulted in a combined list of 257 proteins significantly dysregulated as a consequence of RUNX1-ETO expression, from which 183 were present in the cytoplasm and 74 were found in the nucleus (Figure 3.15A-B).
Figure 3.13 – Principal Component Analysis (PCA) of the protein expression data

Protein expression data was analysed based on its cellular localisation: (A) cytoplasm or (B) nucleus. Protein expression from each sample is represented by coloured dots, according to each biological replicate. Each dot represents individual cord blood samples transduced with retroviral vectors. For each replicate of the experiment, different cord blood was used, and three replicates were performed per condition (n=3).
Figure 3.14 – Hierarchical clustering of proteins expressed in the cytoplasmic and nuclear compartment of CD34+ HSPC

Heatmap of Control and RUNX1-ETO-expressing CD34+ HSPC biological replicate samples (n=3). Protein expression values were uploaded into Partek® and a Euclidean correlation was applied as the distance metric to cluster the proteins. Columns represent proteins; rows represent samples, defined by experimental condition and replicate number. Dendrogram trees show the hierarchy of clusters for both samples and genes. The colour bar (key) represents the z-score from mean expression, calculated for each protein. The heatmaps were divided into cytosolic and nuclear proteins. Red indicates upregulation and blue indicates downregulation of proteins in RUNX1-ETO CD34+ HSPC when compared to control cells.
**Figure 3.15 – Differential protein expression analysis between CD34+ HSPC expressing RUNX1-ETO and control**

**Left panel:** Volcano plots of proteins differentially expressed in the cytosol and nucleus of RUNX1-ETO cells, when compared to control. The Y-axis represents statistical significance (- Log10 p-value), whilst the X-axis indicates fold-change (Log2 fold-change) of protein expression in cord blood derived CD34+ cells transfected with either a RUNX1-ETO or a control vector (n=3). (A) 183 cytosolic proteins were found to be statistically significantly changed in RUNX1-ETO cells compared to control, whereas (B) 74 proteins in the nucleus were identified using the same approach. The significance cut-off (p<0.05) is highlighted with a dotted line. Colour coding is based on fold change, in which blue indicates repressed proteins in RUNX1-ETO CD34+ HSPC when compared to control, whilst proteins in red were found to be overexpressed in the same cells.

**Right panel:** A hierarchical clustering algorithm was used to cluster differentially expressed proteins in control and RUNX1-ETO CD34+ HSPC. Columns represent samples, defined by experimental condition; rows represent proteins. Dendrogram trees show the hierarchy of clusters for both samples and genes. The colour bar (key) represents the z-score from mean expression, as it’s been calculated for each protein. Red indicates up-regulation, blue indicates down-regulation of a specific protein. Heatmap showing differential expression of (C) 183 cytosolic and (D) 74 nuclear proteins.
3.3.2.6 Expression of RUNX1-ETO disrupts normal protein expression in CD34+ cord blood derived HSPC

Comparably to the analysis performed to identify differentially expressed genes changes in RUNX1-ETO CD34+ HSPC, the 257-protein list was uploaded into Metacore™ for pathway and downstream target analysis. Firstly, to understand the underlying biological processes and pathways which might be regulated by RUNX1-ETO, functional Enrichment Analysis was performed (3.3.1.1). GO analysis identified several altered processes in cells overexpressing RUNX1-ETO (Figure 3.16A). Interestingly, these included ‘Metabolic processes’, ‘Biogenesis’ and ‘Immune response’. Even though disrupted metabolic processes have not been widely described in AML t(8;21), leukaemic cell proliferation has been shown to require the up-regulation and rewiring of metabolic pathways to promote anabolic cell growth (Rashkovan and Ferrando, 2019; Hanahan and Weinberg, 2011; Robinson et al., 2021).

To support GO analysis, an alternative approach determining the most significant dysregulated pathways was performed (Figure 3.16B). The most significantly influenced processes included ‘Cell cycle regulation’, ‘Myeloid differentiation’, and changes in the NF-kB pathway. Aberrant activation of the NF-kB pathway has been described in AML, as its been shown to regulate cell survival and apoptosis (Zhou et al., 2015). These processes are similar to the changes observed in the transcriptome of RUNX1-ETO cells, described above, and agree with phenotypical changes observed in vitro and in AML t(8;21) patients (3.3.1.1). Furthermore, this analysis demonstrates the wide range of processes implicated in the development of AML, arising as a consequence of the expression of the RUNX1-ETO fusion protein.

To clarify the classes of proteins dysregulated by RUNX1-ETO, proteins were analysed for their attributed function in Metacore™: TF, binding proteins, enzymes, receptors, general proteins, receptor ligands, regulators, channel proteins, RAS superfamily and transporters (Figure 3.17). Binding proteins and enzymes were found to represent more than half of the total number of proteins significantly altered by the expression of RUNX1-ETO. This supports my analysis of RUNX1-ETO dysregulated mRNA (3.3.1.1).
Figure 3.16 – Functional Enrichment Analysis of protein changes observed in normal haematopoietic development vs. RUNX1-ETO expressing CD34+ HSPC

Enrichment Analysis using Metacore™ of 250 proteins identified to be significantly dysregulated in CD34+ HSPC expressing RUNX1-ETO when compared to control. Top 10 Significantly dysregulated (A) GO processes and (B) Pathway Maps, ranked by - Log (p-value). Processes can further be clustered into (1) Developmental, (2) Immune response-related or (3) Asthma-related processes. A False Discovery Rate (FDR) of 0.05 was applied to both analyses.
Figure 3.17 – Classification of differentially expressed proteins due to the expression of RUNX1-ETO in CD34+ HSPC

Pie chart indicating the classification of the differentially expressed genes using Metacore™. 257 proteins were found to be significantly dysregulated in RUNX1-ETO expressing cells when compared to control.
3.3.2.7 Identification of aberrantly expressed TF in CD34+ HSPC expressing RUNX1-ETO

Pathway analysis described above can be supported by changes arising from TF dysregulated expression. This study was able to identify 16 statistically significantly dysregulated TF proteins arising from the expression of the RUNX1-ETO fusion protein (Table 3.3). From these, 11 TFs were found to be upregulated in CD34+ HSPC expressing RUNX1-ETO, as compared to control, whilst 5 were shown to be downregulated under the same conditions. Most of the TFs were localised in the nucleus, as expected; however, 4 were found to have their expression significantly dysregulated in the cytoplasm of CD34+ cells (Table 3.3). To identify the level of TF connectivity with other proteins from the dataset, this study used the Metacore™ algorithm ‘Interactome Transcription Factors’ on the significantly perturbed proteins (257 proteins). This analysis identified 4 overconnected TF, all present in the nucleus of CD34+ HSPC, detailed in Table 3.3. CBFβ and RUNX1 were found to be two of the TF with the highest number of interactions with other proteins from the dataset. Both proteins belong to the CBF complex and are responsible not only for promoting the expression of other TF involved in cell differentiation, but also growth factors and proliferation and survival regulators. RUNX1 was shown to be equally overexpressed in RUNX1-ETO-expressing CD34+ HSPC, on day 3 of cell culture, as compared to control HSPC. Even though there was an increase in CBFβ protein expression arising from the expression of RUNX1-ETO, there was no difference in CBFβ mRNA expression levels under the same conditions, as compared to control CD34+ HSPC.

This analysis was also able to identify PU.1, a master regulator of myeloid cells, as a TF responsible for the higher number of changes observed at the protein level in CD34+ HSPC upon the expression on RUNX1-ETO compared to control cells. PU.1, a known gene/protein to be dysregulated in t(8;21), was also identified in the mRNA analysis (3.3.1.2). In this context, PU.1 mRNA expression was concordant with protein expression levels, showing similar levels of downregulation in RUNX1-ETO-expressing CD34+ HSPC.

This analysis also identified C/EBPβ as a TF with the highest influence on other proteins from the dataset. This protein is a member of the CCAT enhancer-binding protein (C/EBP) family, which have been known to play important roles in proliferation and differentiation, including the suppression of myeloid leukemogenesis (Ramji and Foka, 2002). C/EBPβ specifically has been shown to reprogram B cells to the myeloid lineage, in mouse splenic and bone marrow cells (Heavey et al., 2003).
Table 3.3 – Identification of RUNX1-ETO induced TF protein dysregulation in CD34+ HSPC

TF protein expression was determined by performing SWATH-MS analysis on cytosolic and nuclear extracts of CD34+ HSPC transfected either with a RUNX1-ETO retroviral vector, or a control plasmid, on day 3 of cell culture. To identify significantly overconnected genes within the dataset, the Metacore™ ‘Interactome Transcription Factors’ algorithm was used. ✓ and X denotes if the gene was identified or not (respectively) by the algorithm. p-values were as described in 2.10.

§ Normalised microarray expression (log₂) was compared between RUNX1-ETO and control CD34+ HSPC by microarray analysis, performed on day 3 of cell culture (adapted from (Tonks et al., 2007)). (↑) (↓) denotes upregulation or downregulation of corresponding gene in RUNX1-ETO CD34+ HSPC, respectively, as compared to control.

RNASeq data from the TCGA 2013 dataset (Ley et al., 2013) (GSE13159) was used to determine mRNA expression in AML M2 patients, divided into t(8;21), expressing the fusion protein RUNX1-ETO, and non-t(8;21) patients. mRNA expression in normal human HSC were obtained from (Rapin et al., 2014) (GSE42519). Each gene has a corresponding Affymetrix ID unique to the Hu133A GeneChip® (Affymetrix®).

Fold-change (FC) values represent the regulation of each gene when compared to control cells (green – upregulated; red – repressed, when compared to control cells). (↑) denotes upregulated or (↓) downregulated gene expression in patients with t(8;21) as compared to normal undifferentiated HSC. (n/a denotes not applicable; CDS denotes coding sequence; Loc. denotes localisation within the cell; C denotes cytoplasm, N denotes nucleus).

<table>
<thead>
<tr>
<th>Protein</th>
<th>UniProtID</th>
<th>p-value</th>
<th>FC</th>
<th>Loc.</th>
<th>Metacore™</th>
<th>CD34+ HSPC™</th>
<th>Literature Review</th>
<th>AML M2 t(8;21) vs non t(8;21)</th>
<th>AML t(8;21) vs HSC</th>
<th>CDS (kbp)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NFKB2</td>
<td>Q00653</td>
<td>3.30 x 10⁴</td>
<td>3.2 C</td>
<td>X</td>
<td>-</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>TRIM28</td>
<td>Q13263</td>
<td>7.16 x 10⁴</td>
<td>1.5 C</td>
<td>X</td>
<td>-</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>PU.1</td>
<td>P17947</td>
<td>1.11 x 10⁻³</td>
<td>-1.9 N</td>
<td>✓</td>
<td>1.453 x 10⁻¹⁰</td>
<td>-1.76</td>
<td>(Vangala et al., 2003; Eder et al., 2016)</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>KMT2A</td>
<td>Q03164</td>
<td>5.97 x 10⁻³</td>
<td>1.6 N</td>
<td>X</td>
<td>-</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>ZBTB48</td>
<td>P10074</td>
<td>6.90 x 10⁻³</td>
<td>1.4 N</td>
<td>X</td>
<td>-</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>LRRFIP1</td>
<td>Q32MZ4</td>
<td>7.01 x 10⁻³</td>
<td>1.31 C</td>
<td>X</td>
<td>-</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>EDF1</td>
<td>O60869</td>
<td>1.10 x 10⁻³</td>
<td>1.4 N</td>
<td>X</td>
<td>-</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Protein</td>
<td>UniProtID</td>
<td>p-value</td>
<td>FC</td>
<td>Loc.</td>
<td>Metacore\textsuperscript{32}</td>
<td>CD34\textsuperscript{+} HSPC\textsuperscript{3}</td>
<td>Literature Review</td>
<td>AML M2 t(8;21) vs. non t(8;21)</td>
<td>AML t(8;21) vs. HSC</td>
<td>CDS (kbp)</td>
</tr>
<tr>
<td>---------</td>
<td>-----------</td>
<td>---------</td>
<td>----</td>
<td>------</td>
<td>----------------------------</td>
<td>-----------------</td>
<td>-----------------</td>
<td>--------------------------</td>
<td>-----------------</td>
<td>----------</td>
</tr>
<tr>
<td>CEBPβ</td>
<td>P17676</td>
<td>1.17 x 10\textsuperscript{-2}</td>
<td>-2.1</td>
<td>N</td>
<td>✓</td>
<td>1.39 x 10\textsuperscript{4}</td>
<td>✓</td>
<td>-1.6</td>
<td>-</td>
<td>↓</td>
</tr>
<tr>
<td>RB1</td>
<td>P06400</td>
<td>1.54 x 10\textsuperscript{-2}</td>
<td>2.4</td>
<td>N</td>
<td>✗</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>RUNX1</td>
<td>Q01196</td>
<td>2.08 x 10\textsuperscript{-2}</td>
<td>1.7</td>
<td>N</td>
<td>✓</td>
<td>1.53 x 10\textsuperscript{-32}</td>
<td>✓</td>
<td>1.91</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>CBFβ</td>
<td>Q13951</td>
<td>2.99 x 10\textsuperscript{-2}</td>
<td>1.4</td>
<td>N</td>
<td>✓</td>
<td>1.571 x 10\textsuperscript{5}</td>
<td>No change</td>
<td>(Roudaia et al., 2009)</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>BAZ2A</td>
<td>Q9UIF9</td>
<td>3.66 x 10\textsuperscript{-2}</td>
<td>1.4</td>
<td>N</td>
<td>✗</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>TTF2</td>
<td>Q9UNY4</td>
<td>3.71 x 10\textsuperscript{-2}</td>
<td>-1.8</td>
<td>N</td>
<td>✗</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>HIRA</td>
<td>P54198</td>
<td>3.80 x 10\textsuperscript{-2}</td>
<td>-1.4</td>
<td>N</td>
<td>✗</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>ZBTB49</td>
<td>Q6ZSB9</td>
<td>3.89 x 10\textsuperscript{-2}</td>
<td>-1.2</td>
<td>N</td>
<td>✗</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>TSHZ1</td>
<td>Q6ZSZ6</td>
<td>4.49 x 10\textsuperscript{-2}</td>
<td>1.6</td>
<td>C</td>
<td>✗</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>
However, no studies have been performed to determine the role of this protein in the development of AML t(8;21). In CD34+ HSPC, this gene was found to be repressed both at the mRNA and the protein level, by 1.6- and 2.1-fold, respectively, upon the expression of RUNX1-ETO (Figure 3.18). Moreover, the CEBPB gene was seen to be significantly downregulated in AML patients harbouring the t(8;21), as compared to other FAB-M2 diagnosed patients (Table 3.3). For this reason, C/EBPβ has been identified as an interesting target and its role in the development of AML t(8;21) will be determined in Chapter 5.

3.3.2.8 RUNX1-ETO leads to the aberrant expression of 53 proteins

As described above, this study identified proteins whose expression was found to be significantly (statistically) dysregulated in RUNX1-ETO expressing CD34+ HSPC, as compared to control. However, statistical analysis is not possible in all samples in which there is no detectable protein expression due to missing expression values. Hence, an additional method of analysis was applied based on proteins solely detected in RUNX1-ETO cells, with no protein being detected in control cells. In total, 53 proteins, of which 29 were expressed in the cytoplasm and 24 in the nucleus, were identified in RUNX1-ETO expressing CD34+ HSPC but could not be detected in control cells under the same conditions.

Of these proteins, 5 were found to be TFs or regulators of transcription, 4 of which detected in the cytoplasm and 1 in the nucleus of RUNX1-ETO CD34+ HSPC (Table 3.4). The fact that these proteins were found to be in cytoplasm, as opposed to the nucleus, could suggest that their normal function is somehow impaired in these cells; however, they were not detected in control cells at all, indicating that are either being expressed earlier in RUNX1-ETO cells, given that analysis was performed on day 3 of blood cell development, or RUNX1-ETO is somehow promoting the expression of these proteins (Figure 3.19A). Whilst it’s appreciated that TF would only be expected to have transcriptomic consequences within the nucleus, aberrant expression of a TF, which is not normally expressed, could lead to mis-appropriate shuttling in certain situations. However, as protein is generated within the cytoplasmic compartment of the cell, a measure of background detection of nuclear proteins can occur. Nevertheless, as these TF were not detected in control cells, they are aberrant in expression in RUNX1-ETO AML. ATXN7 and MIER1 were identified as the most highly expressed in RUNX1-ETO CD34+ HSPC.
Figure 3.18 – C/EBPβ expression profile in CD34+ HSPC expressing RUNX1-ETO

(A) Normalised microarray data showing log₂ CEBBP mRNA expression in normal human CD34+ HSPC expressing RUNX1-ETO compared to control on day 3 of cell culture (n=4) (adapted from Tonks et al., 2007). Significant differences between RUNX1-ETO and control cells was analysed using paired t-test; * denotes p<0.05. (B) C/EBPβ protein expression profile was obtained by performing SWATH-MS analysis on CD34+ HSPC transfected with either a control or a RUNX1-ETO; this protein was found to be expressed solely in the nucleus (Paired T-test; * denotes p < 0.05)
Table 3.4 – TF detected exclusively in RUNX1-ETO expressing CD34+ HSPC

§ mRNA expression (log₂) was compared between RUNX1-ETO and control CD34+ cells by microarray analysis, performed on day 3 of blood cell development (adapted from Tonks et al., 2007).

Φ Protein expression profiles were obtained by performing SWATH-MS analysis on CD34+ HSPC transfected with either a control or a RUNX1-ETO expressing vector. Cytosolic and nuclear proteins were extracted on day 3 of blood cell development. ‘Detected’ indicates protein detection in RUNX1-ETO CD34+ HSPC. (TReg denotes transcriptional regulator).

<table>
<thead>
<tr>
<th>Protein</th>
<th>UniProtID</th>
<th>Localisation</th>
<th>Classification</th>
<th>mRNA expression in CD34+ HSPC §</th>
<th>Protein expression in CD34+ HSPC Φ</th>
<th>CDS kbp</th>
</tr>
</thead>
<tbody>
<tr>
<td>ATXN7</td>
<td>O15265</td>
<td>Cytoplasm</td>
<td>TReg</td>
<td>No change</td>
<td>Detected</td>
<td>~ 2.7</td>
</tr>
<tr>
<td>ZNF395</td>
<td>Q9H8N7</td>
<td>Cytoplasm</td>
<td>TReg</td>
<td>Not detected</td>
<td>Detected</td>
<td>~ 1.4</td>
</tr>
<tr>
<td>REL</td>
<td>Q04864</td>
<td>Cytoplasm</td>
<td>TF</td>
<td>No change</td>
<td>Detected</td>
<td>~ 1.8</td>
</tr>
<tr>
<td>IRF9</td>
<td>Q00978</td>
<td>Cytoplasm</td>
<td>TF</td>
<td>↑</td>
<td>Detected</td>
<td>~ 1.4</td>
</tr>
<tr>
<td>MIER1</td>
<td>Q8N108</td>
<td>Nucleus</td>
<td>TReg</td>
<td>Absent from chip</td>
<td>Detected</td>
<td>~ 1.7</td>
</tr>
</tbody>
</table>
Figure 3.19 – Target protein and mRNA expression profiles of TF found exclusively in RUNX1-ETO CD34+ HSPC

(A) Protein expression profiles were obtained by performing SWATH-MS analysis on CD34+ HSPC transfected with either a control or a RUNX1-ETO expressing vector. Cytosolic and nuclear proteins were extracted on day 3 of blood cell development. Data indicates Mean ± 1SD (n=3). (B) Normalised microarray data showing log2 mRNA expression of the identified target genes, where data was available. mRNA expression in normal human CD34+ HSPC and cells transfected with a RUNX1-ETO expressing vector was measured on day 3 of blood cell development (n=4) (adapted from (Tonks et al., 2007)). Significant differences between RUNX1-ETO and control cells was analysed using paired t-test; * denotes p<0.05.
Given the differences observed in protein expression between RUNX1-ETO and control CD34+ HSPC, this study analysed the corresponding mRNA expression levels to determine if these differences could be observed at the transcriptomic level, or if these are transcriptionally driven. This analysis was unable to detect two genes, ZNF385 and MIER1: even though the microarray chip used contained the probe-set for the ZNF385 transcript, this was not detected. Moreover, no probe-set for the MIER1 transcript can be found on the chip used (Hu133A). Previously, IRF9 was shown to be significantly upregulated in RUNX1-ETO cells, as compared to control CD34+ HSPC (3.3.1.2); however, this analysis showed that whilst IRF9 mRNA was present in control CD34+ HSPC, no protein was detected in these cells. Furthermore, no differences were observed in the mRNA levels of the ATXN7 and REL genes in RUNX1-ETO expressing CD34+ HSPC, when compared to control cells (Figure 3.19B).

A subsequent feasibility assay was performed in order to ensure that overexpression/knockdown studies were possible, in the model systems used. All genes analysed presented a CDS below 3kbp, making them suitable for subsequent functional studies (Table 3.4).

This analysis showed significant differences between mRNA and resulting protein expression, thus strengthening previous observations that mRNA analysis can be insufficient to infer protein expression. The fact that these proteins were found to be solely detected in RUNX1-ETO cells, even though its mRNA can also be detected in control cells, could suggest that changes are post-transcriptionally driven; however, is in unlikely that these are direct transcriptional targets of RUNX1-ETO. This study found that no gene expression correlated with its corresponding protein, except IRF9. Interestingly, IRF9 had already been identified as a target of interest when analysing the transcriptomic profile of CD34+ HSPC expressing the fusion protein RUNX1-ETO (3.3.1.2); the fact that is was identified as being expressed solely in RUNX1-ETO cells suggests that this protein may play a role in the development of AML t(8;21). Nevertheless, all targets are of interest, as their expression seems to be promoted either directly or indirectly by RUNX1-ETO.

3.3.2.9 RUNX1-ETO leads to a shift in protein sub-cellular localisation in CD34+ HSPC.

I next examined whether the expression of RUNX1-ETO was associated with a change in protein sub-cellular localisation in CD34+ HSPC. A total of 926 proteins were identified as
being expressed in both the cytoplasm and the nucleus of control and RUNX1-ETO-expressing CD34+ HSPC. Considering the total amount of protein and its relative sub-cellular localisation, 33 proteins were determined to be localised differently to proteins in normal CD34+ HSPC as a consequence of RUNX1-ETO expression (Figure 3.20A).

The relative expression of 17 proteins was found to be higher in the cytoplasm of RUNX1-ETO-expressing cells, as compared to the cytoplasm of control cells. The EEA1, POLR2A, ALB and RIC8B proteins exhibited the greatest abundance in the cytosolic compartment of RUNX1-ETO cells, whilst in control cells, these proteins were found predominantly in the nucleus (Figure 3.20B). Conversely, 16 proteins were identified as being more abundant in the nucleus of RUNX1-ETO cells, compared to control, possible indicating a translocation from the cytoplasm to the nucleus, as a consequence of the expression of this fusion protein.

Of these, KRT77, SNRPD3 and PLAA showed the lowest detectable levels in the cytoplasm of RUNX1-ETO cells, when comparing to normal protein expression (Figure 3.20B). Due to the increased variability arising from these studies, of the 33 proteins, this study examined the ones that presented with the lowest variation between replicates. A cut off SD of ±10% was applied, resulting in 9 proteins that presented the lowest variation between all three replicates of RUNX1-ETO-expressing and control CD34+ HSPC. These proteins showed no significant difference in total protein expression between RUNX1-ETO and control CD34+ HSPC (Figure 3.21A), indicating that, even though the same protein is being expressed at similar levels, the sub-cellular localisation was different. Of these, 5 proteins were possibly translocated to the cytoplasm as a result of RUNX1-ETO expression, whilst 4 presented a higher prevalence in the nucleus of the same cells, when compared to control (Figure 3.21B-C). Mechanisms underlying these changes will be discussed in 3.4.

Most of the proteins possibly mis-localised because of RUNX1-ETO expression were identified as enzymes and binding proteins; however, no abnormal TF protein shifts were observed in this analysis (Table 3.5). Nevertheless, these might be involved in protein synthesis and post-translational modification processes. The fact that these proteins were shown to be translocated into a different subcellular compartment, as a result of RUNX1-ETO expression, could indicate that their expression is altered, and promoting the expression of genes involved in the leukaemogenic process.
For each detected protein, the relative percentage of cytosolic and nuclear protein was quantified based on the total amount of protein, both in control and RUNX1-ETO expressing CD34+ HSPC. To determine the difference in protein expression between RUNX1-ETO and control cells in the cytoplasmic compartment, the percentage of cytosolic proteins present in RUNX1-ETO cells was subtracted from the equivalent in control cells; the same approach was used to analyse nuclear protein localisation. (A) Protein expression was analysed for 926 molecules, and differences in RUNX1-ETO-expressing compared to control cells calculated for each one; lines denote a 15% cut-off, above which there is a possible translocation of at least 15% of the total amount of protein to the cytoplasm of RUNX1-ETO cells (Barretina et al.), whereas proteins identified below the line indicate an increase of protein expression in the nucleus of RUNX1-ETO cells, as compared to control. Each dot indicates mean (n=3). (B) Proteins with a shift of at least 15% were plotted in an independent graph; as shown in (A), red proteins indicate a protein shift to the cytoplasm in RUNX1-ETO-expressing cells, while blue proteins show a higher protein expression in the nucleus of these cells, compared to normal CD34+ HSPC. Data indicates mean ± 1SD (n=3).
Figure 3.21 – RUNX1-ETO leads to a shift in localisation of 9 proteins

Protein expression profiles were obtained by performing SWATH-MS analysis on CD34+ HSPC transfected with either a control or a RUNX1-ETO expressing vector. Cytosolic and nuclear proteins were extracted on day 3 of blood cell development. (A) To calculate total protein expression, cytosolic and nuclear protein expression values were combined, and expression analysed between RUNX1-ETO and control cells, showing no significant differences between the two conditions. However, when analysing each compartment, it was possible to see that (B) 5 proteins presented a higher prevalence in the cytoplasm of RUNX1-ETO cells, as compared to control, indicating a possible shift in normal protein localisation. Conversely, (C) 4 proteins were shown to be more expressed in the nucleus of RUNX1-ETO cells, as opposed to control. Dots represent each one of the three biological replicates. Line indicates Mean ± 1SD (n=3).
Table 3.5 – Identification of ‘shifted’ proteins as a result of RUNX1-ETO expression

§ mRNA expression (log₂) was compared between RUNX1-ETO and control CD34⁺ cells by microarray analysis, performed on day 3 of blood cell development (adapted from (Tonks et al., 2007)).

Φ Protein expression profiles were obtained by performing SWATH-MS analysis on CD34⁺ HSPC transfected with either a control or a RUNX1-ETO expressing vector. Cytosolic and nuclear proteins were extracted on day 3 of blood cell development.

(↑) or (↓) denotes upregulation or downregulation of gene/protein in RUNX1-ETO cells, respectively, as compared to control HSPC.

<table>
<thead>
<tr>
<th>Protein</th>
<th>UniProtID</th>
<th>Classification</th>
<th>Function</th>
<th>mRNA expression in CD34⁺ HSPC §</th>
<th>Protein expression in CD34⁺ HSPC Φ</th>
</tr>
</thead>
<tbody>
<tr>
<td>RIC8B</td>
<td>Q9NVN3</td>
<td>GEF</td>
<td>Intracellular signalling pathways</td>
<td>Not detected</td>
<td>No change</td>
</tr>
<tr>
<td>MRPL15</td>
<td>Q9P015</td>
<td>Mitochondrial protein</td>
<td>Protein synthesis in the mitochondria</td>
<td>Not detected</td>
<td>No change</td>
</tr>
<tr>
<td>XPNPEP1</td>
<td>Q9NQW7</td>
<td>Enzyme</td>
<td>Maturation/degradation of peptides</td>
<td>No change</td>
<td>↓</td>
</tr>
<tr>
<td>RRM1</td>
<td>P23921</td>
<td>Enzyme</td>
<td>Cell proliferation/migration, tumour/metastasis development</td>
<td>No change</td>
<td>↑</td>
</tr>
<tr>
<td>CHMP1A</td>
<td>Q9HD42</td>
<td>Binding protein</td>
<td>Multivesicular bodies formation and sorting</td>
<td>No change</td>
<td>↓</td>
</tr>
<tr>
<td>MANF</td>
<td>P55145</td>
<td>Binding protein</td>
<td>Protects cells against ER-stress induced damage</td>
<td>No change</td>
<td>↓</td>
</tr>
<tr>
<td>NMT1</td>
<td>P30419</td>
<td>Enzyme</td>
<td>Post-translational modifications</td>
<td>No change</td>
<td>No change</td>
</tr>
<tr>
<td>UBAP2L</td>
<td>Q14157</td>
<td>Binding protein</td>
<td>Ubiquitination</td>
<td>No change</td>
<td>↓</td>
</tr>
<tr>
<td>PLAA</td>
<td>Q9Y263</td>
<td>Binding protein</td>
<td>Ubiquitination</td>
<td>No change</td>
<td>↓</td>
</tr>
</tbody>
</table>

GEP – guanine nucleotide exchange factor; ER – endoplasmic reticulum
In summary, this analysis demonstrated that some proteins might be mis-localised and a
shifted to a different subcellular compartment, arising from the expression of RUNX1-ETO. However, the number of proteins identified was small, and the changes in localisation were not
significant. Even though it is unlikely that these are a major mechanism of action by RUNX1-ETO, it is possible that these genes/proteins contribute to the t(8;21) phenotype.

In conclusion, this chapter was identified two targets of interest for further study and possible mediators of the RUNX1-ETO phenotype. The roles of ZNF217 and C/EBPβ will be explored in subsequent Chapter 4 and Chapter 5.

3.4 Discussion

Expression of RUNX1-ETO leads to widespread disruption in gene and protein function; however, the exact mechanisms through which this fusion protein is able to promote leukaemogenesis remains unclear. Even though this AML subtype presents a favourable outcome, additional target discovery and a deeper understanding of the underlying processes that lead to the development of the disease are necessary. This chapter’s main aim was to identify changes in gene transcription and/or protein expression as a consequence of RUNX1-ETO expression in human CD34+ HSPC. The expression of this fusion protein results in a block in normal cell differentiation; since this is a transcriptionally regulated process, I specifically focused on TF dysregulation. Transcriptome analysis identified ZNF217, also confirmed by SWATH-MS, as a potential target of interest and possible mediator of the block in differentiation observed in RUNX1-ETO expressing cells. Given that mRNA alone does not reliably predict changes in protein expression, quantitative proteomics was also performed in parallel. SWATH-MS allowed the quantitation of changes in protein expression in the cytosol and nucleus of RUNX1-ETO expressing cells. This approach identified C/EBPβ as a further target of interest in the context of t(8;21).

3.4.1 Transcriptional dysregulation imposed by RUNX1-ETO expression

Previously, microarray studies performed by Tonks et al. identified 380 dysregulated genes as a consequence of RUNX1-ETO expression in human CD34+ HSPC (Tonks et al., 2007). However, this analysis was performed using an unsupervised approach, focusing exclusively on the highest fold changing genes, and not on global changes to pathways. This current study was able to improve upon the previous analysis and identify further transcriptomic changes, by
analysing changes within pathway contexts and focusing specifically on TFs. The analysis used two online platforms: Metacore™ and IPA®. One of Metacore’s™ most useful tools is the Enrichment Analysis workflow, which calculates enriched p-values in different types of gene sets within the uploaded dataset. These gene sets originate from curated pathways, networks or related genes derived primarily from literature evaluation and from the GO lexicon (Cirillo et al., 2017). An ‘Enrichment Analysis’ performed on the 380 differentially expressed genes identified them as being implicated in multiple and complex cellular pathways, including those regulating cell proliferation. In addition, this study identified several differentially expressed pathways involved in the progression of AML. These included changes in the β-catenin pathway, which have previously been described in AML t(8;21) - several studies have shown that β-catenin is essential for the clonogenic growth of RUNX1-ETO-expressing HSPC and leukaemic cells (Zhang et al., 2013; Ysebaert et al., 2006). Furthermore, β-catenin has been described as playing a key role in the development of AML (Wang et al., 2010b), and cooperates with several chromosomal aberrations driving this disease (Müller-Tidow et al., 2004). Dysregulated pathways also included the NF-κB pathway, previously described in AML to be associated with disrupted apoptosis in several cell lines (Testa and Riccioni, 2007). Moreover, constitutive activation of NF-κB has been observed in AML CD34+ HSPC, contrasting to normal bone marrow cells (Birkenkamp et al., 2004; Braun et al., 2006; Guzman et al., 2001), leading to changes in tumour formation and maintenance (Bassères and Baldwin, 2006). Lastly, pathway analysis identified significant changes in the normal transcriptional regulation of granulocytic development as reported by Tonks et al. in CD34+ HSPC (Tonks et al., 2004).

In order to determine the most significantly represented gene/protein classes disrupted due to RUNX1-ETO expression, this study used the Enrichment Analysis in Metacore™. Using the z-score (a bidirectional measure of a difference from the mean of involved proteins within a pathway) (Shi et al., 2010), this analysis demonstrated that TFs were found to be the most represented gene class to be dysregulated as a result of RUNX1-ETO expression, following ‘Others’. The latter was not considered to be of interest as it is too generic and does not represent a specific gene class. The reason the ‘Others’ category was found to be the most significant, relies on the fact that Metacore™ was unable classify 221 genes, which represent 55% of the dataset. As z-scores can be negative, since it’s the absolute value that indicates the significant of the analysis, the greatest deviant z-scores from the mean correspond to the most meaningful results. For this reason, this analysis alone can lead to misleading results, as the
‘Others’ category will contain genes from different classes, hence a more detailed classification of genes is often necessary.

Having determined that TFs were one of the most represented gene class dysregulated as a consequence of RUNX1-ETO expression, this study aimed to identify specific TF whose activity could be enhanced due to the presence of RUNX1-ETO. Similarly to Metacore™, IPA® allows the identification of key transcriptional regulators according to overall gene expression, based on the Ingenuity Knowledge Base. To generate a more robust analysis, considering that Metacore™ and IPA® are established on different databases, this study combined both approaches, thus generating a 35-gene list. Metacore™ was able to identify 25 overconnected TF, whilst IPA® identified 31; 22 TF were shown to be identified by both approaches. The fact that some TF were only identified by one of the programs relies on the use of different databases within the programmes.

Among the TF identified were PU.1, a known master regulator required for the development of mature myeloid and lymphoid cells, and CEBPA, a gene that has been demonstrated to play a role in normal granulocytic development (Pang et al., 2018; Tenen, 2001). Both of these TFs have been extensively studied in AML, and validate the approach undertaken here to identify overconnected TF. Importantly, this approach identified other TFs not yet described as having a leukaemogenic role in AML. These included ZBTB20, which normally acts as a transcriptional repressor and is generally associated with B-cell lymphoma, not AML (Peterson et al., 2019). Another example is SMARCA1, expressed in a broad range of normal tissues, and reported to modulate the Wnt/β-catenin pathway. Recently, it has been shown that SMARCA1 contributes to the pathogenesis of leiomyosarcomas, possible due to epigenetic modifications (Patil et al., 2018), as well as to that of gastric carcinomas (Takeshima et al., 2015). However, it’s role in the development of AML t(8;21) has not been described to date. Taken together, this analysis identified the differential expression of genes whose activities could possibly be modulated and consequently the expression of genes involved in cell differentiation and proliferation.

Following identification of TF of interest, the TCGA (Ley et al., 2013) and MILE (Kohlmann et al., 2008) studies were analysed to determine the mRNA expression of each target in patients diagnosed with AML t(8;21), as compared to normal HSCs, in an attempt to refine the 35-gene list. Genes that presented contradictory patterns of expression, shown to be
upregulated in our model but downregulated in t(8;21) patients, or vice-versa, were excluded from further analysis. Furthermore, technical aspects were also considered. These included sequence size, as overexpression studies are more difficult to perform using genes with a bigger CDS, as well as its clinical significance. Using this approach, I identified four potential targets that might be implicated in the development of the phenotype observed in RUNX1-ETO expressing cells – IRF7, IRF9, ARID5B and ZNF217. Since mRNA levels are not powerful predictors of corresponding protein expression, this was assessed through western blotting. To further validate protein expression analysis, SWATH-MS results were analysed for each of targets (see below).

This approach identified two members of the Interferon Regulatory Factor (IRF) family, IRF7 and IRF9, implicated in mediating type I interferon (IFN) response. IRF7 has been shown to be involved in antibacterial and antiviral innate immunity. In the CD34+ HSPC overexpressing RUNX1-ETO, IRF7 was shown to be upregulated at the mRNA level, as well as in t(8;21) AML patient samples. The IRF7 protein was equally found to be upregulated in RUNX1-ETO cells, as compared to control, verified by western blotting. However, this protein was not detected in the SWATH-MS analysis, hence it was eliminated from further analysis.

IRF9 is another member of IRF family, and has been shown to be upregulated in both CD34+ HSPC and t(8;21) AML patient samples. Similarly to IRF7, IRF9 is a TF that mediates the type I IFN response by regulating downstream expression of interferon-stimulated genes, namely STAT1 and 2, with which it forms a complex that shuttles to the nucleus to allow IRF9 to act as a TF (Tsuno et al., 2009). IRF9 has also been shown to be involved in regulating cell proliferation (Weihua et al., 1997) and tumour formation (Luker et al., 2001). High IRF9 expression has been shown to display anti-proliferative effect in ovarian adenocarcinoma (Provance and Lewis-Wambi, 2019); the contrary effect has been proven in cells where the expression of IRF9 is absent or low, in prostate cancer (Erb et al., 2013). In breast cancer, IRF9 upregulation has shown to confer resistance to the chemotherapeutic agent paclitaxel (Conze et al., 2001). Even though its role in breast cancer has been well established, whether IRF9 plays an important part in the pathogenesis of AML is still unclear. I found it to be upregulated both at the mRNA and protein level, validated through western blotting and SWATH-MS.

ARID5B, or AT-rich domain 5B, is a chromatin-modifier transcriptional regulator present in chromosome 10 found to be significantly upregulated in CD34+ HSPC expressing RUNX1-ETO, by 4-fold. Expression of the ARID5B protein, however, showed no difference in
RUNX1-ETO cells, as compared to control HSPC, contrasting with SWATH-MS analysis where ARID5B was shown to be detected in cells expressing the fusion protein RUNX1-ETO, whilst no protein was detected in control samples. Under normal circumstances, the ARID5B protein recognizes the core DNA sequence motif AAT(C/T), influencing cell growth and differentiation of B-lymphocyte progenitors. (Whitson et al., 1999; Lahoud et al., 2001). ARID5B knockout mice exhibit abnormalities in B-lymphocyte development and ARID5B mRNA expression has been shown to be upregulated in other haematological malignancies such as acute promyelocytic leukaemia and acute megakaryoblast leukaemia (Lahoud et al., 2001; Paulsson et al., 2010; Bourquin et al., 2006). In addition, increasing evidence has shown that mutations and single nucleotide polymorphisms of the ARID5B gene are associated with the development of acute lymphoblastic leukaemia, influencing treatment outcome, and may also be associated with paediatric AML. (Lin et al., 2014; Emerenciano et al., 2014)

Lastly, ZNF217 is an oncogenic protein that has numerous functions in various human cancers (discussed below), especially considering that it is located in a chromosomal region frequently amplified in human cancers, in 20q13 (Collins et al., 1998; Tabach et al., 2011). This study confirmed that ZNF217 was upregulated in CD34+ HSPC expressing RUNX1-ETO, compared to control. Upregulation was confirmed at the protein level by western blotting and SWATH-MS. ZNF1217 possesses multiple roles, such as binding to specific DNA sequences to regulate target gene expression, is a component of the human histone deacetylase complex CoREST-HDAC and can be found in complexes with other transcriptional co-repressors including the C-terminal binding protein and the methyltransferase G9a (Cowger et al., 2007; Krig et al., 2007; Thillainadesan et al., 2012; Thillainadesan et al., 2008). Even though ZNF217 mainly acts in transcriptional repressor complexes, it positively regulates the expression of specific target genes, making it a double-faced transcriptional regulator (Cohen et al., 2015). Several studies indicate that ZNF217 interferes with multiple intracellular signalling networks to reprogram cancer cells, by altering cell cycle, cell growth, and disrupting anti-proliferative signalling (Cohen et al., 2015). Moreover, ZNF217 has been shown to interfere with normal apoptotic pathways at early stages of tumour progression and conferring resistance to chemotherapy in breast cancer cells (Thollet et al., 2010; Huang et al., 2005). Overexpression of ZNF217 in breast and ovarian cancer promoted cell proliferation, whereas the opposite effect was shown when silencing it in prostate, colorectal, ovarian and breast cancer (Li et al., 2014; Thollet et al., 2010; Szczyrba et al., 2013; Li et al., 2015). In primary cells, ZNF217 has been shown to interfere with the expression of genes involved in differentiation and organ
development, while at the same time dysregulating genes involved in the repression of cell differentiation and maintenance of cancer stem cells (Krig et al., 2007; Vendrell et al., 2012; Littlepage et al., 2012). In primary mammary cells, overexpression of ZNF217 increased the formation of cells displaying self-renewal ability, an important feature observed in leukemic stem cells (Vendrell et al., 2012; Littlepage et al., 2012).

This study used both western blotting and SWATH-MS to measure protein expression in CD34⁺ HSPC; however, results were not concordant, with results being different between the two approaches. Even though both approaches were able to identify an increased or detectable expression of three targets in CD34⁺ HSPC (IRF9, ARID5B and ZNF217), levels of regulation varied significantly; of these, only ZNF217 had a similar pattern of protein expression. Even though its level of upregulation wasn’t the same (10-fold upregulation in western blot vs. 3-fold upregulation detected through SWATH-MS, in RUNX1-ETO cells compared to control), the trend was the same. The inconsistencies observed between the two techniques may rely on the fact that western blot is typically considered a semi-quantitative technique, as it provides a relative comparison of protein levels, usually to a loading control, and not an absolute measure of quantity. Moreover, it could be related to the quality of the antibodies used. This poses the main advantage in using SWATH-MS instead of western blotting, as it provides a level of specificity unavailable when performing western blot. Furthermore, SWATH-MS allows the detection of thousands of proteins in a single run, maximizing the information obtained per sample, while significantly decreasing the time used for each experiment (Jayasena et al., 2016). The main disadvantage of SWATH-MS is cost, which can constrain the use of MS-based techniques.

In summary, ZNF217 was the only target that consistently showed to be overexpressed, both at the mRNA and protein level. Furthermore, based on literature research, overexpression of ZNF217 has been linked to several hallmarks of cancer, but its dysregulation has not been explored in the context of haematopoiesis and AML. This study will, therefore, address this in Chapter 4.

3.4.2 Dysregulation of the proteome arising from RUNX1-ETO expression

To date, studies performed to identify putative target genes in RUNX1-ETO expressing cells have mainly focused on analysing the cells transcriptomic profiles (Nafria et al., 2020; Ptasinska et al., 2019; Martinez-Soria et al., 2018; Ptasinska et al., 2012). However, proteomic
profile analysis presents a more realistic picture of cell physiology, as compared to genomic studies (Singh and Sharma, 2020).

Over the past few years, several studies have been performed using MS techniques in an attempt to elucidate the mechanisms underlying RUNX1-ETO-mediated AML. Schoenherr et al. applied MS analysis to the Kasumi-1 cell line, which expresses the fusion protein RUNX1-ETO, to analyse the incidence of proteolytic cleavage of cathepsin G and neutrophil elastase and to identify contributing proteases on a proteome-wide scale (Schoenherr et al., 2019). Previously, gene expression studies performed by Lo et al. allowed the identification of RASSF2 as a potential target gene shown to be downregulated in RUNX1-ETO leukaemic blasts, compared to control (Lo et al., 2012). Following up on this study, Stoner et al. used MS analysis to identify the RASSF2- proximal proteome in the Kasumi-1 cell line, revealing its association with Rac GTPase-related proteins (Stoner et al., 2020). Lastly, a study performed by Singh et al. aimed at using an MS-based approach to identify changes in protein expression using the U937 cell line, with inducible RUNX1-ETO expression. The authors showed that RUNX1-ETO leads to drastic change in the cell’s proteomic profile, and focused on the increased expression of NM23 (Singh et al., 2010). The authors showed that NM23 expression is normally suppressed by C/EBP proteins; however, upregulation of RUNX1-ETO blocks this effect, leading to increased levels of NM23, which in turn promoted a block in differentiation and increased proliferation (Singh et al., 2010). Even though these studies contributed to a further understating of the mechanism through which RUNX1-ETO leads to the development of AML, their main focus relied on AML cell lines, and not on primary cell material.

For this study, CD34+ HSPC were transduced with a control or a RUNX1-ETO-overexpressing vector, following which cytosolic and nuclear proteins were extracted and quantified. Before proceeding, extracts were validated for fraction purity, to ensure little or no cross contamination between cytosolic and nuclear fractions, as well as for RUNX1 expression, an indicative of RUNX1-ETO overexpression.

In order to detect all the proteins within our samples, the Pan Human Library was used, which allows the detection and quantification of 50.9% of all human proteins annotated by UniProtKB/Swiss-Prot, corresponding to approximately 13,000 proteins (Rosenberger et al., 2014). Within this study, SWATH-MS analysis detected over 6,000 proteins in the model system used. From these, 4,635 were found to be detected in all three replicates in cytoplasm/nucleus of control and RUNX1-ETO cells, making them suitable for statistical
analysis. The remaining detected proteins were missing expression values, making statistical analysis impossible. This can be attributed to two factors: failure of detection or true protein absence. It is inappropriate to assume that because a protein was not detected by SWATH-MS, is it not being expressed in a cell at the time of analysis. The reasons for samples ‘missing’ some, but not all, protein expression values, can be technical or biological. Technical, and stochastic, reasons can include instrument variation, protein expression levels too close to the noise level, or pre-analytical variation in sample preparation (Collins et al., 2017).

Nevertheless, this study statistically analysed 4,635 proteins, of which 2,787 were detected in the cytoplasm of CD34+ HSPC and 1,848 were present in the nucleus. An exploratory analysis was performed using PCA plots and heat maps, comparing general protein expression between control and RUNX1-ETO-expressing cells. These showed a certain degree of clustering, more noticeable in the cytosolic proteins, as compared to the nuclear samples, in RUNX1-ETO-trasduced cells. Variations in sample clustering can be attributed to different factors. Even though cord blood is an invaluable source of material for the study of blood-related diseases, it is a source of variability. Furthermore, cells originated from different donors have been shown to grow at different rates, depending on a number of donor-related factors (Siegel et al., 2013; Kang et al., 2018; Yap et al., 2000). Moreover, these differences can also be attributed to background noise, due to low expression of certain proteins, thus resulting in very low signal. Nonetheless, of the 4,635 proteins, 257 were found to be statistically dysregulated (p<0.05) because of RUNX1-ETO expression; 183 were localised to the cytoplasm and 74 to the nucleus. This analysis was performed in an exploratory manner and without performing statistical corrections, as the number of samples (n=3) is insufficient to confer corrected statistical significance. Downstream analysis was performed on the 257 differentially expressed proteins with a specific focus on TF changes, as previously performed for the transcriptomic dataset. Once more, the first step of analysis consisted of performing an Enrichment Analysis using Metacore™ (Figure 3.16). This analysis identified the significantly dysregulated proteins as being implicated in several cellular processes, including ‘Myeloid differentiation’ (Nafria et al., 2020; Ichikawa et al., 2013), ‘NF-kB pathway’ (Zhou et al., 2015), ‘Anti-apoptotic’ (Goyama et al., 2013) and ‘Neutrophil chemotaxis’ pathways (Tonks et al., 2004; Tijchon et al., 2019), known to be implicated in AML t(8;21) pathophysiology. Analysis of changes in the transcriptome and proteome of CD34+ HSPC, as a consequence of RUNX1-ETO expression, yielded similar results, as cell differentiation and apoptosis were highly represented in both datasets.
Following pathway analysis, this study analysed and clustered the 257 significantly dysregulated proteins previously described. Comparably to the analysis performed above (3.4.1), TFs were one of the classes with the highest representation within the dataset. Subsequently, this study focused on TF changes arising from RUNX1-ETO expression, by analysing their overconnectivity with other proteins from the dataset, leading to the identification of 4 overconnected TF: CBFβ, C/EBPβ, PU.1 and RUNX1. As expected, RUNX1 was identified in this analysis, potentially as a result of RUNX1-ETO expression, known to influence normal gene expression. Subsequently, the algorithm also identified CBFβ, a TF regulating the transcription of several haematopoietic-specific targets, including myeloperoxidase, CSF1 receptor, IL-1 and GM-CSF (Downing, 2001). Upon binding to RUNX1, this complex is responsible for not only the transcriptional activation in early HSC, but also expression of growth factors and proliferation and survival regulators. Similarly to the analysis performed above (3.4.1), PU.1 was once more identified as one of the most overconnected TF within the dataset.

Lastly, C/EBPβ was identified as one of the TFs with the highest influence on other dysregulated proteins, as identified in 3.3.2.7. Another member of the CEBP family, this TF has been shown to play an important role in macrophage development (Lekstrom-Himes and Xanthopoulos, 1998; Poli, 1998; Ramji and Foka, 2002). The process through which C/EBPβ is activated is complex, since it is regulated upon different levels, including transcriptionally, translationally and post-translationally. Firstly, the CEBPB gene does not possess any introns, and its mRNA can originate three distinct protein isoforms based on three alternative AUG translation initiation sites (Descomb and Schibler, 1991; Timchenko et al., 1999). C/EBPβ LAP* (also named full length, or C/EBPβ-1), C/EBPβ LAP (C/EBPβ-2) and C/EBPβ LIP (C/EBPβ-3). The latter is thought to function as a transcriptional repressor, since it lacks the transcription activating domain but still processes the ability to dimerize and bind to DNA (Descomb and Schibler, 1991). Alternatively, CEBP proteins can be regulated through protein-protein interactions with proteins important for cell proliferation, including Cdk2, Cdk4 or the Rb protein (Ji and Studzinski, 2004; Charles et al., 2001; Wang et al., 2001). Lastly, C/EBPβ has several phosphorylation sites, which are important for intracellular localisation and transcriptional activity (Metz and Ziff, 1991; Chinery et al., 1997; Piwien Pilipuk et al., 2003; Berg et al., 2005).
To study the role of C/EBPβ in differentiation, several studies have used HL-60 cells, an APL cell line, and 1,25D, a drug capable of promoting the differentiation of these cells. These groups noted an increase in C/EBPβ expression once the drug was added to the cells, contributing to the cell’s differentiation into the monocytic lineage, but no difference was observed in the expression levels of other members of the CEBP family, including C/EBPα and C/EBPε (Ji and Studzinski, 2004; Studzinski et al., 2005; Marcinkowska et al., 2006). Furthermore, the LAP and LIP isoforms of C/EBPβ seem to increase more rapidly in forced-differentiating HL-60, as compared to control HL-60 cells, suggesting that these isoforms have dominant roles over LAP*. Moreover, even though LIP functions as a transcriptional repressor, the authors suggest that C/EBPβ LIP plays a role in the repression of genes unnecessary in differentiating cells, such as those involved in cell proliferation (Marcinkowska et al., 2006). The same trend was observed by Pham et al. in peripheral blood monocytes (Pham et al., 2007).

The authors showed that human macrophages contain high constitutive levels of functional C/EBPβ, unlike peripheral monocytes. In addition, PMA-treated THP-1 cells were shown to induce CEBPB expression; however, even though the authors observed that these transitions were not associated with major changes in CEBPB mRNA levels, which remained stable throughout the differentiation process, there was an increase in C/EBPβ protein levels, particularly regarding the LAP and LAP* isoforms. In this study, both C/EBPβ mRNA and protein were downregulated in cells expressing RUNX1-ETO suggesting that this fusion protein is inhibiting CEBPB at the transcriptional level. Currently there is no information whether this change in C/EBPβ expression could impact growth and development of normal haematopoietic cells, this will therefore be addressed in Chapter 5.

### 3.4.2.1 Alternative analysis of RUNX1-ETO-induced proteomic changes in CD34+ HSPC

An alternative approach analysed proteins exclusively expressed in RUNX1-ETO cells, but undetectable in control samples. These were not identified above as it was not possible to perform a statistical analysis on undetected samples. This study found that 53 proteins were found to be exclusively expressed in RUNX1-ETO cells, from which 29 were found exclusively in the cytoplasm and 24 in the nucleus.

Of these, this study identified 4 TFs found to be expressed in the cytoplasm, including ATXN7, ZNF395, REL and, interestingly, IRF9. The IRF9 transcript had previously been identified as a target of interest (3.3.1.2) as it had been identified as one of the most overconnected genes within the mRNA dataset. However, western blotting analysis detected
nuclear levels of IRF9 protein, even if no nuclear protein was detected by SWATH-MS. ATXN7, is a member of the STP3/TAF9/GCN5 histone acetyltransferase (STAGA) complex. Even though some additional studies have been performed in colorectal and clear cell renal carcinomas, the functional consequences arising from mutations in this gene have not been determined yet (Gotoh et al., 2014; Kalvala et al., 2016). Gene expression studies have found that the expression of ZNF395 is significantly increased in several cancer types, including renal cell carcinoma, osteosarcoma, and Ewing sarcoma (Skubitz et al., 2006; Tsukahara et al., 2004; Dalgan et al., 2007). REL is a proto-oncogene known to play a role in cell differentiation by being part of the NF-κB complex, and its role has already been extensively studies in the context of AML (Guzman et al., 2001; Nakagawa et al., 2011; Zhou et al., 2015). However, since these TFs were identified in the cytoplasm, it is unlikely that they possess any transcriptional roles. Furthermore, additional western blotting analysis would be necessary to assess nuclear protein expression.

Moreover, one TF was identified in the nucleus. MIER1 is a transcriptional regulator differentially expressed in breast carcinoma cell lines and tumours (Paterno et al., 1998; McCarthy et al., 2008). However, these targets were not pursued in this current study, as this posed as an alternative way of analysing the proteomic changes induced by RUNX1-ETO expression.

The last approach analysed proteins that were detected in both the cytoplasm and nucleus of control and RUNX1-ETO CD34+ HSPC, by analysing relative protein expression in each compartment, regarding to total protein expression, to identify mis-localised proteins. These shifts in proportion of expression at each subcellular localisation may help identify proteins differently localised between RUNX1-ETO and control cells. Previously, several proteins have been shown to promote cancer development due to their abnormal localisation within a cell such as EGFR and β-catenin, normally localised in the plasma membrane, that upon translocation to the nucleus leads to the development of pancreatic cancer (Cohen, 1983). Therefore, this study analysed the expression levels of 926 proteins, and looked at their relative sub-cellular localisation. Of these, 33 proteins were shown to be localised differently in RUNX1-ETO cells, as compared to control, by at least 15% of relative protein expression, indicating that, in RUNX1-ETO cells, there was an accumulation of certain proteins in the cytoplasm or nucleus, as compared to the same compartment of control cells. This cut-off was based on potential validation experiments through western blotting techniques, since a smaller
difference might be harder to detect. Additionally, due to the high variability arising from the source of cord blood material and viral preps, an additional cut off SD of ±10% was applied (n=3). This led to the identification of 9 proteins potentially translocated/retained as a consequence of RUNX1-ETO expression. Of these, 5 proteins showed a higher expression in the cytoplasm of RUNX1-ETO cells, possible indicating protein retention in this compartment, as compared to control cells; and 4 were identified as being more abundant in the nucleus of RUNX1-ETO cells. However, most of these proteins were identified as enzymes and binding proteins, as no TFs were identified using this approach. For this reason, proteins were not examined further. Even so, mis-localisation of these proteins can lead to TF significant dysregulation and, therefore, impaired normal cellular process and leukaemogenic development, as a result of RUNX1-ETO expression.

It is also important to acknowledge that the current proteomics study represents an exploratory pilot programme and, as such, the small sample size and inability to study the consequences of RUNX1-ETO in the cells’ proteomic profile is recognised due to cost concerns. The sample size is sufficient to examine changes to inform further research and day 3 of culture was selected as the consequence of RUNX1-ETO is a differentiation block in early developmental stages. Secondly, due to the scarcity of umbilical cord blood donations as a result of the COVID-19 pandemic, it was not possible to validate these findings within the same CD34+ context. However, owed to the proteomic approach, there is less concern regarding validation as mRNA-based approaches would suffer.

In conclusion, this chapter’s main aim was to identify changes in TF expression arising from the expression of the fusion protein RUNX1-ETO, in CD34+ HSPC. This aim was divided into two main objectives. The first one consisted of analysing previously generated microarray data, to identify changes at the mRNA level that could explain the dysregulation observed in these cells. This led to the identification of ZNF217 as a target of interest. Its role in the haematopoietic process and in the development of AML will be explored in Chapter 4. However, since mRNA is not always an accurate predictor of protein expression, I also performed SWATH-MS analysis on the same cells, to study the cells’ proteomic profile. This analysis identified C/EBPβ, a member of C/EBP family of TF, as a potential mediator of the leukaemogenic process of AML t(8;21). The effects of its dysregulation in haematopoiesis will be addressed in Chapter 5.
Chapter 4

ZNF217 Promotes Normal Human Myeloid Development
4.1 Introduction

Previously, I identified ZNF217 as being overexpressed in RUNX1-ETO expressing cells compared to control (Chapter 3). ZNF217 is a TF whose normal function is to regulate gene expression through binding to multiple regulatory complexes, including the human histone deacetylase complex (CoREST-HDAC) (Cowger et al., 2007; Thillainadesan et al., 2012; Thillainadesan et al., 2008). However, this protein has a known oncogenic role in solid tumours, often attributed to the fact that its gene is located in a region frequently amplified in human cancers (20q13 chromosome) (Collins et al., 1998; Tabach et al., 2011). Increased copies of ZNF217 have been reported in multiple tumours, including breast, colorectal and ovarian cancer (Plevova et al., 2010; Fang et al., 2010; Rahman et al., 2012), and have been linked to a poorer outcome in colon and ovarian carcinomas (Peiró et al., 2002; Rooney et al., 2004).

In solid tumours, ZNF217 has been shown to interfere with several cell processes including disruption of cell proliferation (Li et al., 2014; Thollet et al., 2010) through changes in TGF-β-dependent anti-proliferative signalling (Massagué, 2008; Ikushima and Miyazono, 2010), cell cycle (Li et al., 2014) and apoptosis (Huang et al., 2005). For example, ZNF217 overexpression was able to confer drug resistance, as well as promote cell proliferation in vitro and increase tumour growth in vivo, in a breast cancer model (Thollet et al., 2010). Conversely, downregulation of ZNF217 resulted in a decrease in cell proliferation, increased drug sensitivity, decreased colony formation efficiency and cell migration and invasion in breast, ovarian and colorectal cell lines (Sun et al., 2008a; Thollet et al., 2010; Li et al., 2015). ZNF217 is also able to bind to the promoter of genes involved in cell differentiation and organ development, resulting in an arrest in the cell differentiation process and maintenance of breast cancer stem cells (Krig et al., 2007; Vendrell et al., 2012; Littlepage et al., 2012). In normal primary mammary epithelial cells, ZNF217 overexpression led to an increase in the formation of mammospheres with an increased self-renewal potential and was associated with the repression of an adult stem cell expression signature (Littlepage et al., 2012; Vendrell et al., 2012; Nguyen, 2018). Moreover, Mao et al. described that ZNF217 was upregulated in glioma stem cells (GSC) when compared to non-GSC. This study showed that forced differentiation of GSC led to a decrease in the levels of ZNF217 and its knockdown inhibited their proliferation, while reducing stem-like populations (Mao et al., 2011).
Altogether, these studies have established a link between the tumorigenic process and the overexpression of ZNF217 in solid tumours. However, whether ZNF217 has a role in perturbing haematopoietic growth and development has not been determined.

4.2 Hypothesis and Aims

This study hypothesises that ZNF217 expression plays an important role in haematopoiesis and that modulating its expression will disrupt normal human myeloid development, thus leading to AML initiation. This chapter’s main objective is to understand the role of ZNF217 in haematopoietic development and its possible contribution to the development of AML, using a normal human primary HSPC model and AML cell lines. This will be achieved by performing the following aims:

**Determine the expression of ZNF217 during normal human myeloid development and in AML patient blasts**

Analysis of publicly available transcriptomic datasets will be performed to determine ZNF217 mRNA expression in human HSPC subsets and across different AML subtypes.

**Determine the effects of ZNF217 overexpression and knockdown on myeloid colony forming ability and self-renewal**

Limiting-dilution colony forming assays will be performed on transduced human CD34\(^+\) HSPC and compared to control. Replating assays will be performed to assess the cells’ self-renewal potential.

**Determine the effect of ZNF217 overexpression or knockdown on normal human haematopoietic growth, differentiation, and development.**

CD34\(^+\) HSPC will be transduced with a ZNF217-overexpression or -knockdown vector and grown in bulk liquid culture. Changes in monocytic and granulocytic growth will be assessed by analysing the cells’ lineage and cell surface markers using multicolour flow cytometry.

**To determine the effect of ZNF217 overexpression or knockdown on AML cell growth, proliferation and apoptosis.**
AML cell lines will be transduced with a ZNF217-overexpression or -knockdown vectors and grown in bulk liquid culture. Changes in cell growth will be evaluated by following the cells’ proliferative ability over time. Furthermore, transduced cells will be assayed to determine if ZNF217 dysregulated expression results in changes in cell cycle and apoptotic frequency.

4.3 Results

4.3.1 Expression of ZNF217 mRNA increases throughout myeloid cell development

In order to determine the expression of ZNF217 throughout haematopoietic development, mRNA expression data for normal human haemopoietic cell populations was analysed. ZNF217 mRNA expression significantly increases throughout granulocytic development, particularly in metamyelocytes, band cells and polymorphonuclear cells, as compared to immature HSC, by 2.2-, 3.3- and 6.6-fold, respectively (Figure 4.1A). Regarding monocytic differentiation, ZNF217 mRNA levels significantly increase by 2.3-fold in mature monocytes, as compared to HSC, though levels were lower compared to granulocytes (Figure 4.1A). Analysis of the mouse homolog Zfp217 also showed upregulation in mature granulocytes and monocytes, with a significant increase of 2.6-fold observed in both populations (Figure 4.1B). These observations suggest that ZNF217 is transcriptionally regulated during blood cell differentiation, and differentially expressed according to specific cell types.

4.3.2 ZNF217 is variably expressed across AML subtypes

As previously introduced, AML is a highly heterogenous disease characterised by multiple molecular and cytogenetic abnormalities (1.2.2). To determine ZNF217 mRNA expression in different subtypes of AML, publicly available transcriptomic datasets were analysed. Firstly, this study compared ZNF217 mRNA expression in different subsets of AML with that of cell subsets at different stages of haematopoietic development. As Figure 4.2A shows, ZNF217 expression is highly variable across AML. There were no significant differences in ZNF217 mRNA expression between undifferentiated HSC and the multiple AML subtypes analysed. However, this expression was significantly lower in all AML subtypes as compared to polymorphonuclear cells, suggesting that expression of ZNF217 might influence normal granulocytic development.
Figure 4.1 – ZNF217 mRNA expression in normal haematopoiesis

(A) Normalised microarray data (log2) showing ZNF217 mRNA expression in distinct human haematopoietic cells subsets based on cell surface marker expression. Normal human haematopoiesis data derived from GSE24759 (Rapin et al., 2014; Svendsen et al., 2016) Data indicates Mean ± 1SD. Statistical analysis was performed using one-way ANOVA with Bonferroni’s multiple comparisons test; ** denotes p<0.01; *** denotes p<0.001; **** denotes p<0.0001, vs HSC (Probeset 203739_at).

(B) Normalised microarray data (log2) showing ZFP217 mRNA expression, ZNF217 mouse homolog, across different murine haematopoietic cell types. Mouse normal haematopoiesis data derived from GSE60101 (Lara-Astiaso et al., 2014). Data indicates Mean ± 1SD. Statistical analysis was performed using one-way ANOVA with Bonferroni’s multiple comparisons test; ** denotes p<0.01; *** denotes p<0.001; **** denotes p<0.0001, vs HSC (Probeset 1437414_at)

HSC – Haematopoietic Stem Cell; MPP – Multipotential Progenitors; CMP – Common Myeloid Progenitor cell; MEP – Megakaryocyte / Erythroid Progenitor cell; GMP – Granulocyte / Monocyte Progenitors; Early PM - Early Promyelocyte; Late PM - Late Promyelocyte; MY - Myelocyte; MM – Metamyelocytes; BC – Band cell; PMN – Polymorphonuclear cell; Mono – Monocytes; Gran – Granulocytes; Macro – Macrophages.
Figure 4.2 – ZNF217 mRNA expression levels are variable across AML subtypes

(A) Normalised microarray data (log2) showing ZNF217 mRNA expression in normal human haematopoietic developmental subsets vs. AML subtypes. Normal human haematopoiesis data derived from GSE422519 (Rapin et al., 2014; Svendsen et al., 2016); human AML cells derived from GSE13159 (Haferlach et al., 2010; Kohlmann et al., 2008), GSE15434 (Klein et al., 2009), GSE61804 (Metzelder et al., 2015), GSE14468 (Wouters et al., 2009; Taskesen et al., 2015; Taskesen et al., 2011) and TCGA (Ley et al., 2013) (data analysed using Bloodspot’s algorithm Bloodpool (Bagger et al., 2016)). Normalised microarray data (log2) showing ZNF217 mRNA expression in (B) AML according to FAB subtype and (C) AML FAB2, categorised into t(8;21) and non-t(8;21), within the TCGA dataset (Ley et al., 2013) All studies were analysed using the 203739_at probeset. Data indicates mean ± 1SD. Significant differences were analysed by one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01; *** denotes p<0.001; **** denotes p<0.0001.

**AML normal** – AML with normal karyotype; **AML complex** – AML with complex karyotype.
Interestingly, analysis of transcriptomic data from the TCGA dataset according to AML FAB subtype, shows that \textit{ZNF217} expression, whilst highly variable, remains similarly expressed across different developmental AML subtypes (Figure 4.2B). Furthermore, no difference was observed in AML M2 patients expressing the fusion protein RUNX1-ETO (Figure 4.2C).

Subsequently, I analysed the expression of \textit{ZNF217} in patients diagnosed with \textit{de novo} AML using the TCGA 2013 RNAseq dataset (Ley et al., 2013), and stratified them according to high and low quartiles, to represent high and low \textit{ZNF217} expression, respectively (Figure 4.3A). For this analysis, patients with APL and those who did not receive standard chemotherapy treatment were excluded from the dataset. Having established these parameters, this study analysed the OS and disease-free survival of patients with high and low \textit{ZNF217} expression. Additionally, the association between differential \textit{ZNF217} expression and disease subtypes, molecular and cytogenetic abnormalities, and patient characteristics were assessed.

Regarding patient outcome, this analysis showed that patients with higher \textit{ZNF217} expression were associated with a lower OS, as compared to patients with lower mRNA expression, with a median survival of 17 compared to 30 months., although this was deemed not significant (Figure 4.3B). Furthermore, patients with high \textit{ZNF217} expression were found to possess a significantly reduced disease-free survival, accompanied by a hazard ratio of 2.2 (95% CI[1-4.8]), suggesting increased risk of relapse (Figure 4.3C).

Having determined that higher \textit{ZNF217} expression is associated with reduced disease-free survival, this study aimed at determining the relationship between disease characteristics, such as FAB subtype or molecular abnormalities, as well as patient attributes, and \textit{ZNF217} expression. Firstly, \textit{ZNF217} expression was found to not correlate with PB blasts or WBC in AML patients (Figure 4.4A, C). A significantly higher proportion of BM blasts was detected in patients with lower \textit{ZNF217} expression (Figure 4.4B), as well as a higher mutation count (Figure 4.4D). Both these factors are indicative of a poorer prognosis, despite lower \textit{ZNF217} expression being associated with better disease-free survival. Moreover, dysregulated \textit{ZNF217} expression was significantly correlated with AML FAB subtype (Figure 4.5A). Lastly, high \textit{ZNF217} was not associated with patient characteristics, such as age or gender (Figure 4.5B-C). Altogether, these results suggest a potential role for \textit{ZNF217} mis-regulation in the development of AML, regardless of subtype.
Figure 4.3 – High levels of ZNF217 is associated with poorer disease-free survival of AML patients

(A) Normalised (log2) ZNF217 mRNA expression according to upper (n=27) and lower quartiles (n=27). Significance denoted by One-way ANOVA; **** p<0.0001. (B) Overall survival and (C) disease-free survival analysis of AML patients stratified according to upper and lower ZNF217 mRNA expression quartiles. For overall survival curves, ZNF217 upper quartile n=27; ZNF217 lower quartile n=27. For disease-free survival curves, ZNF217 upper quartile n=26; ZNF217 lower quartile n=27. Untreated patients and t(15;17) AML patients (that present a different treatment regime compared to other AML subtypes) were excluded from this analysis. Statistical analysis was performed using the Long-Rank test between high and low ZNF217 expression groups. Data obtained from TCGA. (Ley et al., 2013) using cBioPortal (www.cbioportal.org).
Figure 4.4 – High ZNF217 expression is associated with decreased BM blasts and lower mutation count

Relationship between high and low ZNF217 expression and several clinical attributes of AML patients. Data obtained from TCGA (Ley et al., 2013) using cBioPortal (www.cbioportal.org). ZNF217 upper quartile (n=27); ZNF217 lower quartile (n=27). Graph showing the percentage of (A) peripheral blood (PB) blasts; (B) BM blasts; (C) white blood count (WBC); and (D) mutation count of AML patients according to ZNF217 high (n=27) and low (n=27) expression. Statistical differences between high and low ZNF217 expression in AML patients was analysed using Mann-Whitney test; Data represents mean ± 1SD; * denotes p<0.05; ** denotes p<0.01.
Figure 4.5 – Dysregulated ZNF217 expression is not associated with patient characteristics

Relationship between high and low ZNF217 expression and AML patient characteristics. Data obtained from TCGA (Ley et al., 2013) using cBioPortal (www.cbioportal.org). ZNF217 upper quartile (n=27); ZNF217 lower quartile (n=27). Graph showing the (A) number of patients across the different FAB subtypes of AML (M0-M7), (B) age at diagnosis (years) and (C) proportion of male and female patients according to ZNF217 high (n=27) and low (n=27) expression. Statistical differences between high and low ZNF217 expression in AML patients was analysed using Mann-Whitney test; Data represents mean ± 1SD; * denotes p<0.05.
4.3.3 Overexpression of ZNF217 promotes human myeloid development

The above data shows that ZNF217 expression increases with myeloid cell development and the previous chapter demonstrated that this gene and its corresponding protein are upregulated in RUNX1-ETO-expressing HSPCs compared to control vector. To determine the effect of ZNF217 overexpression on myeloid growth, differentiation, and self-renewal, I ectopically expressed this protein as a single abnormality in the CD34+ HSPC in vitro model. Knockdown studies were subsequently performed in 4.3.4.

4.3.3.1 Generation of ZNF217 overexpressing CD34+ HSPC

A lentiviral vector harbouring the ZNF217 CDS and co-expressing GFP as a selectable marker was used to infect CD34+ HSPC. To confirm ZNF217 protein expression, western blot was performed on transfected HEK 293T cells. As shown in Figure 4.6A, cells transfected with ZNF217 had higher ZNF217 protein expression compared to cells transduced with control vector. A significantly lower infection rate was observed in CD34+ HSPC transduced with a ZNF217-overexpression plasmid, primarily due to plasmid size (Figure 4.6B). For this reason, it was not technically feasible to isolate sufficient nuclear protein from HSPC for western blot.

Assays which required a pure population, such as colony forming efficiency (2.3.4), as well as cell cycle (2.7.5), apoptosis (2.7.6) and morphological analysis (2.3.5), necessitated FACS sorting to ensure non-transduced cells wouldn’t influence the interpretation of the data. However, for assays which facilitate sub-selection of the population of interest (i.e. flow cytometry based), it is possible to gate through the GFP population prior to analysis (2.7.4). It is appreciated that this mixed culture could enable non-transduced cells to influence transduced cells; however, sorting this population would not allow the recovery of enough cells to perform such assays. Additionally, puromycin selection was not feasible in these cells, as this process takes approximately 48-72h, in which analysis of the cells would not be possible.
Figure 4.6 – Transduction of ZNF217 in HEK 293T cells and human HSPC

(A) Western blot showing ZNF217 protein expression in HEK 293T cells transduced with ZNF217 plasmid or control plasmid (n=1). The K562 cell line was used as a positive control (PC). GAPDH was used as a loading control (n=1). (B) Summary data showing percentages GFP positivity in control and ZNF217-overexpression CD34⁺ HSPC cultures, on day 3 of culture. Data indicates Mean ± 1SD (n=5).
4.3.3.2 ZNF217 overexpression suppresses myeloid colony formation and self-renewal

Previous studies have shown that ZNF217 overexpression in primary mammary epithelial or breast cancer cells is able to promote the formation of mammospheres with increased self-renewal potential (Vendrell et al., 2012; Littlepage et al., 2012). Initially, to determine the effect of ZNF217 overexpression on myeloid colony forming ability and self-renewal potential, a colony assay, followed by subsequent replating was performed on GFP+ cells isolated by FACS (Figure 4.7), thus ensuring a pure transduced population. Following 7 days of culture, ZNF217 overexpression led to a significant 35% reduction colony-forming ability compared to control cells (Figure 4.8A). Furthermore, a subsequent serial replating assay showed a significant 60% reduction (Figure 4.8B) in the cells’ self-renewal potential as a result of ZNF217 expression, as compared to control cells. These results indicate that ZNF217 impairs myeloid colony formation, with a reduction in the cells’ self-renewal ability. However, it is important to note that the current assay does not allow the discrimination between different types of colonies, including monocytic- and granulocytic-derived cells. As the media used does not support erythroid cell development, it is possible to affirm that these are not present in culture. Additional assays using semi-solid culture medium, such as methylcellulose, would be necessary to definitively make further conclusions.

4.3.3.3 ZNF217 overexpression suppresses the growth of myeloid cells

Previous studies have shown that upregulated ZNF217 expression leads to the suppression of breast, colon and teratocarcinoma cancer cell differentiation, leading to tumorigenesis (Krig et al., 2007). I hypothesise that ZNF217 overexpression may have an anti-differentiation phenotype on myelopoiesis. To analyse the effect of ZNF217 overexpression in myeloid development, specific time-points were selected at an appropriate frequency to fully capture these changes, previously optimised within the research group. Moreover, cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Overexpression of ZNF217 led to a 2.2-fold decrease in the growth of myeloid cells following 13 days of culture (Figure 4.9A). To analyse the impact on different myeloid lineages, the CD13 and CD36 cell surface markers were used to discriminate between monocytic, granulocytic and erythroid populations. ZNF217 caused a significant decrease in monocytic cell growth, compared to control cells (3.5-fold on day 13) (Figure 4.9B). A slower growth rate was equally observed in the granulocytic lineage, in which ZNF217-overexpressing cells showed a 2-fold decrease in growth compared to control. (Figure 4.9C).
Figure 4.7 – Expression and enrichment of GFP+ populations in human CD34+ HSPC

(A) Representative flow cytometric histograms showing GFP expression in control and ZNF217 transduced cultures (pre-sorted cultures on day 3 of culture). Plots generated based on the “Non-debris” gate described in Figure 2.4. Background auto fluorescence was established using CD34+ cells subjected to the equivalent viral infection procedure but in the absence of virus (mock). Mock HSPC – grey; Transduced HSPC - green (control) / red (ZNF217 OE) (B) Representative histogram for control culture post sorting for GFP+ and GFP- cell populations. Plots generated based on the “Non-debris” gate described in Figure 2.4.
Figure 4.8 – ZNF217 overexpression inhibits myeloid colony formation and decreases self-renewal capacity

(A) Colony forming efficiency of control and ZNF217-overexpression cultures following 7 days of growth in liquid culture containing IL-3, SCF, G- and GM-CSF (2.3.4). Data indicates Mean ± 1SD (n=5). Significant difference was analysed by paired T-test; ** denotes p < 0.01. (B) Self-renewal potential assessed by a single replating round of control and ZNF217 cultures, in the same conditions as previously. Data indicates Mean ± 1SD (n=3). Significant difference between ZNF217-expressing cells and control was analysed by paired t-test; * denotes p<0.05; ** denotes p<0.01.
Figure 4.9 – ZNF217 overexpression inhibits monocytic and granulocytic growth during myeloid development

(A) Cumulative fold-expansion of control and ZNF217-overexpressing GFP+ cells grown over 13 days in culture medium containing IL-3, SCF, G- and GM-CSF (Supplementary Figure 2). (B) Cumulative fold-expansion of monocytic cells (CD13+ CD36+) transduced with a control of ZNF217-overexpressing vector. (C) Cumulative fold-expansion of granulocytic cells (CD13- CD36-) transduced with a control of ZNF217-overexpressing vector. (D) Cumulative fold-expansion of erythroid cells (CD13+ CD36+) transduced with a control of ZNF217-overexpressing vector. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Figure 2.4 and Figure 2.5. Data indicates Mean ± 1SD (n≥3) Significant differences between ZNF217-expressing cells and control at each time-point were analysed by paired t-test; * denotes p<0.05; (this data has also been represented as dot plots, within the supplementary section [Supplementary Figure 30]).
Whilst statistical significance is achieved late (10/13 days of cell culture) the trend is consistent throughout the experiment, suggesting a biological effect is present from onset.

However, the growth in erythroid cells was not significantly altered by ZNF217 overexpression in these cells (Figure 4.9D). This was not unexpected as the media in which these cells are grown promotes monocytic and granulocytic development, but not erythroid, due to the absence of the EPO, essential for terminal erythroid differentiation. These observations suggest that ZNF217 overexpression influences the growth of myeloid cells, particularly the cells committed to the monocytic and granulocytic lineages.

4.3.3.4 ZNF217 overexpression promotes myeloid differentiation

To determine the consequences of ZNF217 overexpression on myeloid development, I analysed the expression of differentiation cell surface markers in combination with lineage discriminators over time using flow cytometry (2.7.4). As shown in Figure 4.10A, ZNF217 overexpression led to a significant 1.6-fold (day 13) increase in the percentage of monocytic cells in culture compared to control. Concomitantly, the proportion of granulocytic cells was significantly lower throughout development (Figure 4.10B). Conversely, there was an increase in erythroid-committed cells in the ZNF217 culture compared to control, within the same timeframe (Figure 4.10C). Altogether, these observations suggest that the balance between lineages in culture is influenced by ZNF217 overexpression, by reducing the granulocytic population at the expense of monocytes, during normal myeloid development of human HSPC.

In order to analyse the consequences of ZNF217 overexpression on monocytic and granulocytic differentiation, this study assessed the expression of the differentiation cell surface markers CD34, CD11b, CD14 and CD15. CD34 is expressed in HSPC and progressively declines as HSPC differentiate (Caux et al., 1989). Moreover, as myeloid cell development progresses, there is an increase in the expression of CD11b (Tonks et al., 2004). Monocytic cells upregulate CD14 expression during development, whilst granulocytes show an increase in CD15 (Tonks et al., 2004).
Figure 4.10 – ZNF217 overexpression disrupts the balance between granulocytic and erythroid populations during myeloid cell development.

Summary data showing the proportion of (A) monocytic (CD13+ CD36+), (B) granulocytic (CD13+/CD36-) and (C) erythroid cells (CD13- CD36+) in control and ZNF217-overexpressing cultures. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Figure 2.4 and Figure 2.5. Data indicates Mean ± 1SD (n≥4). Significant differences between ZNF217-expressing cells and control at each time-point were analysed by paired t-test; * denotes p<0.05; (this data has also been represented as dot plots, within the supplementary section [Supplementary Figure 31]).
In monocytes, ZNF217 overexpression significantly altered monocytic development (Figure 4.11), resulting in a 40% reduction in the percentage of cells expressing the CD34 marker on day 3 (Figure 4.11A). Upregulation of CD11b was increased in ZNF217 overexpressing monocyctic cells throughout development and, by day 10, it was 1.5-fold higher compared to control (Figure 4.11B). CD14 was also upregulated throughout development (Figure 4.11C). Additionally, a significant increase in cell granularity (SSC) on days 10 and 13 of development was further observed (Figure 4.11D).
Figure 4.11 – ZNF217 overexpression promotes monocytic differentiation

(A) Summary data of CD34+ expression in terms of percentage in monocytic cells (CD13+ CD36+) over time for control and ZNF217-overexpressing cells. Summary data showing (B) CD11b and (C) CD14 expression in terms of MFI in monocytic cells over time for control and ZNF217-overexpressing cells. (D) Complexity of monocytic cells was measured by changes in SSC for control and ZNF217-overexpressing cells on days 10 and 13 of myeloid differentiation. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Figure 2.4 and Figure 2.5. Data indicates Mean ± 1SD (n≥3). Significant differences between ZNF217-expressing cells and control at each time-point were analysed by paired t-test; * denotes p<0.05, ** denotes p<0.01, *** denotes p<0.001; (for representative flow cytometry plots please refer to Supplementary Figure 3, Supplementary Figure 4, Supplementary Figure 5; this data has also been represented as dot plots, within the supplementary section [Supplementary Figure 32]
Analysis of morphological features showed an increase in the percentage of cells displaying late monocytic features, consistent with immunophenotypic analysis (Figure 4.13). These results suggest that overexpression of ZNF217 promotes monocytic differentiation which may give rise to the decreased proliferation of this lineage observed in Figure 4.10A.

ZNF217 overexpression was shown to also perturb granulocytic lineage development (Figure 4.12). CD34+ was only modestly impacted by ZNF217 (Figure 4.12A). CD11b was shown to be upregulated in ZNF217 cells (1.3-fold on day 13; Figure 4.12B) suggesting that differentiation was again promoted in this lineage an observation supported by an increase in cell granularity (SSC) (Figure 4.12D), however, ZNF217-overexpressing cells had a lower expression level of CD15 (1.5-fold decrease by day 13), compared to control cells (Figure 4.12C) confounding this interpretation. Additionally, morphological analysis of granulocytic cells showed no difference in ZNF217-expressing cultures, as compared to control, with an equal proportion of cells displaying late morphological features (Figure 4.13).

Taken together, these results suggest that the overexpression of ZNF217 in CD34+ HSPC leads to a pro-differentiation phenotype, as evidenced by upregulation of monocytic and, to some extent, granulocytic developmental markers, and morphology.

4.3.3.5 ZNF217 overexpression in HSPC induces changes in cell cycle

Having determined that ZNF217 leads to a decrease in colony forming ability and self-renewal potential, two possible explanations for this were tested: decreased proliferative capacity and/or increased apoptosis. ZNF217-transduced cells showed a significant decrease in the proportion of cells in cycle (S + G2/M), indicating an accumulation of cells in G1 phase supporting the hypothesis that ZNF217 impacts proliferative capacity (Figure 4.14A).

In order to assess whether ZNF217 overexpression led to an increase in apoptosis, an Annexin V-based assay was performed. No significant difference in apoptosis was observed for ZNF217-transduced cells compared to control in the days examined (Figure 4.14B). These results indicate that the reduced colony forming ability of ZNF217-expressing cells and reduced growth ability of HSPC is not a result of cell death, but a consequence of the decrease in the proportion of cells entering cell cycle, suggesting that these cells are dividing at a slower rate, as compared to control cells.
Figure 4.12 – ZNF217 overexpression promotes granulocytic differentiation by upregulating monocytic markers

(A) Summary data of CD34+ expression in granulocytic cells (CD13+/- CD36-). Summary data showing (B) CD11b and (C) CD15 expression in granulocytic cells. (D) side scatter analysis (SSC) of transduced cells on days 10 and 13 of myeloid differentiation. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Figure 2.4 and Figure 2.5. Data indicates Mean ± 1SD (n≥3). Significant differences between ZNF217-expressing cells and control at each time-point were analysed by paired t-test; * denotes p<0.05, ** denotes p<0.01. (for representative flow cytometry plots please refer to Supplementary Figure 6. Supplementary Figure 7. Supplementary Figure 8; this data has also been represented as dot plots, within the supplementary section [Supplementary Figure 33]).
Figure 4.13 – ZNF217 overexpression alters cell morphology during myeloid development

(A) Representative images showing the morphology of HSPC transduced with control and ZNF217 OE cultures were analysed on day 13 of differentiation by cytopsinning 3x10⁴ cells and staining them with May-Grünwald-Giemsa differential stain. (B) Differential counts of all cultures with morphology categorised into granulocytic and monocytic-committed cells. Approximately 500 cells were counted in total and each population count normalised to total number of cells, according to Figure 2.2. Granulocytic cells were further divided into myelocytes, metamyelocytes and band/segmented cells. Data indicates mean ± 1SD (n=3). Significant difference between ZNF217 OE cultures and control were analysed by paired T-test; ** denotes p<0.01.
Figure 4.14 – ZNF217 overexpression in CD34+ HSPC significantly disrupts normal cell cycle.

(A) Representative histograms showing cell cycle distribution of normal HSPC (Control) and ZNF217-expressing cells after 6 days of culture. Initial gating was applied as described in Figure 2.6 The percentage of cells in S + G₂/M is indicated by mean ± 1SD (n=3) Significant difference between ZNF217-expressing cells and control was analysed by paired t-test; * denotes p<0.05. (B) Summary data showing the effect of ZNF217 overexpression in apoptosis, on CD34+ HSPC following 6, 10 and 13 days of culture. Initial gating was applied as described in Figure 2.7 (i) Early apoptotic cells are characterised by Annexin V⁺ / PI⁻; (ii) Apoptotic cells are characterised by Annexin V⁺ / PI⁺. Data indicates Mean ± 1SD (n=3).
Whilst feasible, earlier time-points would not be practical for three main reasons. Firstly, an apoptosis assay requires a minimum of 50,000 cells and hence would use a substantial proportion, if not all, of sorted cells. Secondly, the CD34<sup>+</sup> model exhibits the greatest growth potential between days 3 and 6 of culture, in which 15-20-fold growth is frequently observed; taking cells on day 3 would substantially reduce the available cells for subsequent analysis. Lastly, as cells have only just been collected from cell sorting, it is likely that the stress incurred by this process would mask any biological effects. For these reasons, day 6 is the earliest suitable time-point for the apoptotic assay.

Altogether, these results indicate that ZNF217 has a pro-differentiation role, in concordance with the mRNA analysis described above (4.3.1), in which the expression of ZNF217 increases with myeloid differentiation. Overexpressing ZNF217 in HSPC suggests that this favours monocytic development, although granulocytic development might also be influenced by its expression. More importantly, given these observations, it is unlikely that overexpression of ZNF217 contributes to leukaemogenic development through change in normal myeloid cell differentiation.
4.3.4 Knockdown of ZNF217 does not impact myeloid differentiation

The data above shows that ZNF217 overexpression was shown to promote normal myeloid development by significantly increasing monocytic and granulocytic cell differentiation. Previously, several authors have also addressed the effect of ZNF217 downregulation in solid tumours. Sun et al. used shRNA-mediated KD vectors in ovarian cancer cell lines, and showed that ZNF217-KD cells presented a decrease in cell growth and in colony formation efficiency (Sun et al., 2008a). Similar observations were made by Li et al. in colorectal cancer, in which the ZNF217-siRNA mediated KD lead to a decrease in cell proliferation and impaired cell migration and invasion in the colorectal cancer cell line SW480 (Li et al., 2015). Altogether these results indicate that the ZNF217-mediated cell proliferation and invasion can be attenuated by employing KD. However, the consequences of ZNF217 KD in haematopoiesis have not been determined yet. Therefore, this study aimed at determining the effect of ZNF217 KD on myeloid growth and differentiation of CD34+ HSPC, using shRNA constructs.

4.3.4.1 Selection and validation of ZNF217 shRNA constructs

To reduce the levels of ZNF217 expression in normal human CD34+ HSPC, lentiviral vectors based on shRNA constructs targeting ZNF217, encoding both a GFP and puromycin were used (Table 2.1). To initially validate KD, ZNF217 was KD in K562 cells. The level of KD achieved was similar across all constructs used, with a reduction of at least 50% in the levels on endogenous ZNF217 protein expressed in these cells (Figure 4.15A-B).

In CD34+ HSPC, it was possible to achieve over 60% infection efficiency for the different shRNA constructs (Figure 4.16). For subsequent assays, shZNF217 #4 was discarded, as its expression resulted in cell death (data not shown), possibly due to off-target effects.

4.3.4.2 Knockdown of ZNF217 leads to a decrease in CD34+ HSPC colony forming ability

To determine the effect of ZNF217 KD on colony forming ability of myeloid cells, transduced CD34+ cells were sorted based on GFP+ expression and plated by limiting dilution, as previously described (4.3.3). Following 7 days of culture, all three ZNF217-KD cultures had a significant decrease in colony forming ability of approximately 50%, compared to control cells (Figure 4.17A).
Figure 4.15 – ZNF217 shRNA constructs successfully reduced the levels of endogenous ZNF217 in K562 cells

(A) Western blot of ZNF217 total protein levels in K562 cells transduced with a scrambled shRNA vector, as well as three different ZNF217 shRNA constructs (n=1). K562 parental cells were used as a positive control (PC). GAPDH was used as a loading control. TRCN numbers have been described in Table 2.1. (B) Western blot semi-quantification by densitometry using ImageJ. Relative protein expression was determined by normalising each sample to GAPDH and to shRNA control.
Figure 4.16 – Infection efficiency of transduced CD34+ HSPC using shRNA lentiviral constructs

Summary data showing the percentage of GFP⁺ cells in control and ZNF217 KD CD34⁺ HSPC cultures, analysed on day 3 of culture. Gating strategies were applied as described in Figure 2.4. Data indicates Mean ± 1SD (n=4).
Figure 4.17 – ZNF217 knockdown disrupts myeloid colony formation of HSPC

(A) Colony forming efficiency of control and three ZNF217-knockdown cultures following 7 days of growth in liquid culture containing IL-3, SCF, G- and GM-CSF (2.3.4). Data indicates Mean ± 1SD (n=4). Significant difference between shZNF217 cultures and control was analysed by one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01, *** denotes p<0.001. (B) Self-renewal potential was assessed by a single replating round of control and ZNF217-KD colonies, in the same conditions as previously. Data indicates Mean ± 1SD (n=3) Significant differences between shZNF217 cultures and control were analysed by one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01; *** denotes p<0.001.
Furthermore, a replating assay showed no difference in self-renewal ability between control cells and two ZNF217 KD cultures; however, shZNF217 #2 showed a 75% reduction in colony forming-ability (Figure 4.17B). Altogether, reduction of ZNF127 levels led to a decrease in the cells’ colony forming ability, with no loss of self-renewal potential for two of the constructs used.

4.3.4.3 Knockdown of ZNF217 disrupts myeloid cell growth and development

To determine the consequences of reducing endogenous levels of ZNF217 in CD34+ HSPC, cells were cultured as described above (4.3.3.3) and subsequent growth, differentiation, and survival of control and ZNF217 KD cultures were analysed throughout 13 days of culture. Having established that ZNF217 overexpression supressed the growth of HSPC (Figure 4.9), this study hypothesised that reducing ZNF217 expression would increase the proliferative ability of these cells. Figure 4.18 shows the effect of ZNF217 KD on myeloid cell growth, representing two biological replicates (see 4.4.3 for more details). KD of ZNF217 in these cells failed to induce significant differences in the growth of cells in bulk liquid culture (Figure 4.18A). As previously described, the CD13 and CD36 cell surface markers were used to discriminate between monocytic, granulocytic and erythroid populations, and to analyse the growth and differentiation of each lineage. KD of ZNF217 led to an increase in monocytic cell growth, in cultures transduced with shZNF217 #1 and #3, by 1.5- and 2.5-fold on day 13, respectively (Figure 4.18B). No meaningful differences were noted in both granulocytic and erythroid growth (Figure 4.18C-D). Taken together, these results suggest that KD of ZNF217 disrupts the normal growth of monocytic cells, potentially associated with an immature phenotype.

Similarly to the approach described above (4.3.3.4), cell surface markers were used to analyse individual lineages. No significant changes were observed in terms of lineage balance regarding the granulocytic or erythroid lineages, following ZNF217 KD (Figure 4.19B-C). However, there was a significant increase in the proportion of monocytic cells in ZNF217 KD cultures as compared to control throughout culture (1.6-fold for shZNF217 #2, on day 13) (Figure 4.19A).
Figure 4.18 – Knockdown of ZNF217 promotes monocytic cell growth in myeloid development

(A) Cumulative fold-expansion of control and three ZNF217-KD constructs in terms of GFP positivity in culture medium containing IL-3, SCF, G- and GM-CSF, grown over 13 days (Supplementary Figure 9). (B) Cumulative fold-expansion of monocytic cells (CD13+ CD36+) transduced with a control or ZNF217-KD vectors. (C) Cumulative fold-expansion of granulocytic cells (CD13+ CD36+) transduced with a control or ZNF217-KD vector. (D) Cumulative fold-expansion of erythroid cells (CD13- CD36+) transduced with a control of ZNF217-KD vector. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Figure 2.4 and Figure 2.5. Data indicates Mean ± 1SD (n=2) (no statistical test was employed); (this data has also been represented as dot plots, within the supplementary section [Supplementary Figure 34]).
Figure 4.19 – ZNF217 knockdown disrupts the balance between the monocytic and granulocytic populations in CD34+ HSPC during myeloid cell development

Summary data showing percentage of (A) monocytic (CD13+ CD36+), (B) granulocytic (CD13+/− CD36−) and (C) erythroid cells (CD13− CD36+) in control and three ZNF217-KD cultures. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Figure 2.4 and Figure 2.5. Data indicates Mean ± 1SD (n≥4). Significant differences between shZNF217 cultures and control at each time-point were analysed by one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01; (this data has also been represented as dot plots, within the supplementary section [Supplementary Figure 35]).
Following the assessment of the proportion of cells committed to each lineage in each condition, the effect of ZNF217 KD on the normal differentiation of monocytes and granulocytes was further assessed, using the cell surface markers previously described (4.3.3.4). Figure 4.20 shows the effect of ZNF217 KD in the monocyctic-committed cells. No significant changes were observed for any of the parameters analysed, including expression of CD34, CD11b and CD14. In contrast, a significant decrease in cell granularity was observed in ZNF217 KD cells, on days 10 and 13 of cell culture. However, even thought this was statistically significant, biologically this was very modest and unlikely to show a real effect. Similarly, no significant differences were observed regarding granulocytic development upon KD of ZNF217 (Figure 4.21).

Altogether, these results indicate that KD of ZNF217 impairs the colony forming ability of myeloid cells; however, it does not play an endogenous role in myeloid development, and the reduced monocytic growth is not a result of altered differentiation. Moreover, even though ZNF217 expression was shown to promote monocytic differentiation, expression of this TF is not essential for monocytic development.

4.3.4.4 Knockdown of ZNF217 does not affect cell cycle and apoptosis

Lastly, to determine the effects of ZNF217 KD on cell cycle, this study analysed FACSorted transduced HSPC, on day 6 of culture. No difference between control and ZNF217 KD cells, was observed (Figure 4.22A). I next determined if the decrease in the cells’ colony forming ability was a result of increased apoptosis. ZNF217-KD cultures showed no difference in apoptosis on days 10 and 13 of development. These results indicate that the impact of ZNF217 KD on the cells colony forming ability and growth is not supported by changes in cells’ cell cycle or apoptotic events (Figure 4.22B).
Figure 4.20 – Knockdown of ZNF217 does not significantly alter monocytic development

(A) Summary data of CD34+ expression in terms of percentage in monocyctic cells (CD13+ CD36+) over time for control and ZNF217-KD cells. Summary data showing (B) CD11b and (C) CD14 expression in terms of MFI in monocyctic cells over time for control and ZNF217-KD cells. (D) Complexity of monocyctic cells was measured by changes in SSC for control and ZNF217-KD cultures on days 10 and 13 of myeloid differentiation. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Figure 2.4 and Figure 2.5. Data indicates Mean ± 1SD (n=4). Significant differences between shZNF217 cultures and control at each time-point were analysed by one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01; (for representative flow cytometry plots please refer to Supplementary Figure 10, Supplementary Figure 11, Supplementary Figure 12; this data has also been represented as dot plots, within the supplementary section [Supplementary Figure 36]).
Figure 4.21 – Knockdown of ZNF217 does not significantly alter granulocytic development

(A) Summary data of CD34+ expression in terms of percentage in granulocytic cells (CD13+/ CD36−) over time for control and ZNF217-KD cells. Summary data showing (B) CD11b and (C) CD15 expression in terms of MFI in granulocytic cells over time for control and ZNF217-KD cells. (D) Complexity of granulocytic cells was measured by changes in SSC for control and ZNF217-KD cultures on days 10 and 13 of myeloid differentiation. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Figure 2.5. Data indicates Mean ± 1SD (n≥4). Significant differences between shZNF217 cultures and control at each time-point were analysed by one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01; (for representative flow cytometry plots please refer to Supplementary Figure 13, Supplementary Figure 14, Supplementary Figure 15; this data has also been represented as dot plots, within the supplementary section [Supplementary Figure 37]).
Figure 4.22 – Knockdown of ZNF217 has no effect on normal cell cycle or apoptosis

(A) Representative histograms showing cell cycle distribution of normal HSPC (Control) and three ZNF217-knockdown cultures after 6 days of culture. Initial gating was applied as described in Figure 2.6. The percentage of cells in S + G₂/M is indicated by mean ± 1SD (n=3). (B) Summary data showing the effect of ZNF217 knockdown on apoptosis, on HSPC following 10 and 13 days of culture. Initial gating was applied as described in Figure 2.7: (i) Early apoptotic cells are characterised by Annexin V⁺ / PI⁻ and (ii) Apoptotic cells are characterised by Annexin V⁺ / PI⁺. Data indicates Mean ± 1SD (n=3).
4.3.5 The role of ZNF217 in AML cell lines

4.3.5.1 ZNF217 expression is variable across AML cell lines

Using the Cancer Cell Line Encyclopaedia (CLLE) (Barretina et al., 2012), the expression of ZNF217 mRNA in a cohort of AML cell lines was determined. Figure 4.3A shows ZNF217 transcript is highly expressed across all lines, including in the t(8;21) cell lines Kasumi-1 and SKNO-1. However, previous studies have reported that ZNF217 protein expression does not correlate with ZNF217 gene amplification levels (Li et al., 2015; Szczyrba et al., 2013; Renner et al., 2013; Etcheverry et al., 2010). For this reason, this study determined ZNF217 protein expression by western blot. ZNF217 protein expression level was shown to be variable across the AML cell lines tested. The expression level of ZNF217 was undetectable in Nomo-1, PLB985, TF1, NB4, HL-60 and OCI-AML5 cells but was strongly expressed in SKNO-1, HEL, OCI-AML2 and K562 cells (Figure 4.3B) making these cell lines suitable for overexpression and knockdown studies respectively. Additionally, ZNF217 was heterogeneously expressed in t(8;21) cell lines, showing a substantially lower expression in the Kasumi-1 cell line, as compared to SKNO-1; furthermore, there was no distinct pattern of expression between t(8;21) cell lines and non-t(8;21) cell lines.

4.3.5.2 Generation of ZNF217-knockdown AML cell lines

Using the same shRNA constructs (4.3.4), ZNF217 was knocked-down in the leukaemic cell lines HEL and SKNO-1, which presented the highest levels of endogenous ZNF217 protein. Transfected cell lines were selected using puromycin, confirmed through GFP positivity (Figure 4.24-Figure 4.25A) and validated by western blot (Figure 4.24-Figure 4.25B-C). Different levels of knockdown were achieved based on the cell line used. In the HEL cell line, all three constructs used were shown to possess a similar level of KD, with endogenous levels of ZNF217 reduced by c75%. In the SKNO-1 line, however, shZNF217 #1 had the weakest level of KD, with a modest c40% reduction in ZNF217 expression; the other two constructs showed similar reduction levels to HEL cells.
Figure 4.23 – ZNF217 expression in leukaemic cell lines

(A) Western blot analysis of ZNF217 endogenous expression in a cohort of AML cell lines (n=1). The K562 cell line was used as a positive control (PC); GAPDH was used as a loading control. (B) Normalised (log2) ZNF217 mRNA expression in a cohort of leukaemic cell lines. mRNA expression data derived from https://portals.broadinstitute.org/ccle/page?gene=ZNF217 (Barretina et al., 2012). Null protein values refer to undetectable levels of endogenous ZNF217.
Figure 4.24 – ZNF217 is knocked-down in the erythroleukaemia cell line HEL.

(A) Representative flow cytometric plots showing transduction efficiency in the HEL cell line transfected with (i) an shRNA control and (ii–iv) three ZNF217 knockdown constructs (shZNF217 #1, shZNF217 #2, shZNF217 #3), following puromycin selection. (B) Western blot analysis showed successful ZNF217 knockdown in all three shRNA constructs (n=1). GAPDH was used as a loading control.; K562 cells were used as a positive control (PC) (C) Western blot quantification by densitometry analysis using ImageJ. Relative protein expression was determined by normalising each sample firstly to the loading control, then to shRNA control (n=1).
Figure 4.25 – ZNF217 is knocked-down in the t(8;21) cell line SKNO-1

(A) Representative flow cytometric plots showing transduction efficiency in the SKNO-1 cell line transfected with (i) an shRNA control and (ii-iv) three ZNF217 knockdown constructs (shZNF217 #1, shZNF217 #2, shZNF217 #3), following puromycin selection. (B) Western blot analysis showed successful ZNF217 knockdown in all three shRNA constructs (n=1). GAPDH was used as a loading control. (C) Western blot quantification by densitometry analysis using ImageJ. Relative protein expression was determined by normalising each sample firstly to the loading control, then to shRNA control (n=1).
4.3.5.3 ZNF217 knockdown in AML cell lines reduces cell growth by promoting changes in cell cycle and apoptosis

I next determined the effect of ZNF217 KD on leukaemic cell growth. Growth was significantly reduced in the HEL cell line following day 4 of culture (Figure 4.26A), by 1.5-fold (with shZNF217 #3). This reduction in growth remained consistently lower until day 7, in which a 2.1- and 2.5-fold reduction was observed upon transduction with shZNF217 #1 and #3, respectively. Similarly, even though not statistically significant, a reduction in cell growth was also observed in the SKNO-1 cell line (Figure 4.27A).

In order to determine the reason behind this decreased proliferative ability, these lines were assayed for apoptosis using Annexin V binding in combination with 7-AAD exclusion. In HEL cells, the decrease in growth observed for shZNF217 #3 can be attributed to an increase in pre-apoptotic cells, on day 6 of culture, by 5-fold, as compared to control (Figure 4.26B). However, an analysis of the apoptotic rate on day 4 of culture would be necessary to further strengthen these observations. No significant differences were observed in SKNO-1 cells (Figure 4.27B). These data suggest that ZNF217 expression performs an anti-apoptotic role in HEL cells.

Additionally, the cell cycle distribution of these cells was analysed using the DNA dye PI. In the HEL cell line, ZNF217 KD did not lead to significant changes in the normal cells’ cell cycle (Figure 4.26C). However, loss of this protein resulted in substantial changes in SKNO-1 cells’ cell cycle, with a significant increase in the proportion of cells in S+G2 phase, throughout culture (Figure 4.27C). Taken together, the decrease in proliferative potential observed upon ZNF217 KD can be a result of not only an increase in apoptosis, but also due to an arrest in the (S + G2) phase, potentially enhancing cell death, suggesting ZNF217 is necessary for AML cells’ proliferation and survival. However, it is suggested that this effect occurs independently of RUNX1-ETO expression.
Figure 4.26 – ZNF217 knockdown significantly decreases HEL cell proliferation

Summary data showing (A) the growth of HEL cells with three ZNF217 knockdown constructs compared to control over 7 days of growth. Data indicates Mean ± 1SD (n=5) (2-way ANOVA statistical test with Bonferroni’s multiple comparison test - ** denotes p < 0.01; *** denotes p < 0.001; **** denotes p < 0.0001); (B) the effect of ZNF217 knockdown on Annexin V staining in HEL cells following 1, 3 and 6 days of cell growth. Bars indicate percentage of pre-apoptotic cells (Annexin V⁺ / 7-AAD⁻). Data indicates Mean ± 1SD (n=3) (One-way ANOVA statistical test with Bonferroni’s multiple comparison test - * denotes p < 0.05); (C) the effect of ZNF217 knockdown in the HEL cells’ cell cycle following 1, 3 and 6 days of cell growth. Data indicates Mean ± 1SD (n=3) (One-way ANOVA statistical test with Bonferroni’s multiple comparison test - * denotes p < 0.05; ** denotes p < 0.01; *** denotes p < 0.001; **** denotes p < 0.0001).
Summary data showing (A) the growth of SKNO-1 cells with three ZNF217 knockdown constructs compared to control over 7 days of growth. Data indicates Mean ± 1SD (n=5) (2-way ANOVA statistical test with Bonferroni’s multiple comparison test - * denotes p < 0.05; ** denotes p < 0.01; **** denotes p < 0.0001); (B) the effect of ZNF217 knockdown on Annexin V staining in SKNO-1 cells following 1, 3 and 6 days of cell growth. Bars indicate percentage of pre-apoptotic cells (Annexin V+ / 7-AAD-). Data indicates Mean ± 1SD (n=3) (One-way ANOVA statistical test with Bonferroni’s multiple comparison test - * denotes p < 0.05; ** denotes p < 0.01; *** denotes p < 0.001; **** denotes p < 0.0001); (C) the effect of ZNF217 knockdown in the SKNO-1 cells’ cell cycle following 1, 3 and 6 days of cell growth. Data indicates Mean ± 1SD (n=3) (One-way ANOVA statistical test with Bonferroni’s multiple comparison test - * denotes p < 0.05; ** denotes p < 0.01; *** denotes p < 0.001; **** denotes p < 0.0001).

Figure 4.27 – ZNF217 knockdown significantly decreases SKNO-1 cell proliferation
4.4 Discussion

ZNF217 is a TF belonging to the Kruppel-like family of TF. The gene is located on chromosome 21q13, in a region frequently amplified in various human cancers, including breast and ovarian, as compared to normal tissue and epithelial cells (Collins et al., 1998; Littlepage et al., 2012). Whilst several studies have been performed in solid tumours, the role of ZNF217 in the haematopoietic process and in the development of AML has not yet been described. The main aim of this chapter was to describe the normal expression of ZNF217 during haematopoiesis and its role in normal human myeloid development. Overexpression of ZNF217 in HSPC was found to significantly inhibit the growth of myeloid progenitor cells, associated with an increase in their differentiation potential, therefore suggesting that ZNF217 promotes myeloid development. Moreover, KD of ZNF217 had no impact in both monocytic and granulocytic maturation, suggesting that, even though expression of ZNF217 promotes differentiation, it is not essential for this process to occur.

To understand whether overexpression of ZNF217 might contribute to leukaemic development, this study aimed at analysing its expression level during normal haematopoiesis. ZNF217 gradually increases as cells differentiate into mature granulocytes and, to a lesser extent, monocytes. Expression of the mouse homologue Zfp217 revealed a similar trend, with mature monocytes and granulocytes expressing the higher levels of Zfp217 as compared to HSC. However, a study based on embryonic stem cells identified Zfp217 (homologue) as essential for maintaining the undifferentiated state cells present, associated with epigenetic regulation and m6A RNA modifications (Aguilo et al., 2015). These high expression levels tightly correlate with the regulation of several pluripotency genes, including NANOG, POU5F1 and SOX2 and upregulation of these has previously been described in other cancers (Aguilo et al., 2015). For instance, SOX2 upregulation has not only been identified in several cancers, but it also correlated with poor disease prognosis (Tam and Ng, 2014; Weina and Utikal, 2014; Wuebben and Rizzino, 2017). Ultimately, there are few studies regarding the role of ZNF217 in the haematopoietic process and often have conflicting data. These might be related to different models/techniques used.
4.4.1 High ZNF217 in AML is associated with increased risk of relapse

Analysis aimed at exploring the association between ZNF217 expression and leukaemogenesis showed that the expression of ZNF217 mRNA is highly variable across AML subtypes, but is generally lower in AML as compared to fully mature granulocytes, suggesting that ZNF217 is promoting a block in granulocytic differentiation.

To determine the relationship between ZNF217 expression and clinical attributes, the TCGA dataset was used to stratify patients according to high and low expression quartiles (Ley et al., 2013). Disease-free survival was found to be significantly lower in patients with high levels of ZNF217, indicating that these patients possess a higher risk of relapse. Similar observations have previously been made in solid tumours. Overexpression of ZNF217 levels was observed in both primary prostate carcinoma (Szczyrba et al., 2013) and colorectal carcinoma (Zhang et al., 2015), the latter further associating with poor clinical features. Similar results were observed in glioma samples, in which high ZNF217 was found to relate to poor outcome (Mao et al., 2011). However, the majority of studies regarding overexpression of ZNF217 in tumour development are associated with breast cancer, in which high ZNF217 has been documented in breast tumour samples, as compared to normal epithelial cells (Krig et al., 2010; Littlepage et al., 2012; Nguyen et al., 2014; Nonet et al., 2001; Plevova et al., 2010; Thollet et al., 2010; Vendrell et al., 2012; Collins et al., 1998). Similarly to the present study, analysis of breast cancer patients showed that high levels of ZNF217 were associated with decreased relapse-free survival (Vendrell et al., 2012; Littlepage et al., 2012; Nguyen et al., 2014), further correlating with poor OS (Littlepage et al., 2012; Nguyen et al., 2014) and prognosis (Nguyen et al., 2014; Vendrell et al., 2012), and with the development of metastases (Vendrell et al., 2012). Equally, high levels of ZNF217 protein have been identified in cervical (Zhu et al., 2009) and colorectal (Zhang et al., 2015; Li et al., 2015) cancers, as well as gastric carcinoma (Shida et al., 2014). These were found to be associated with more aggressive disease subtypes (Zhang et al., 2015; Li et al., 2015), as well as poorer OS (Li et al., 2015) and relapse-free survival (Shida et al., 2014). Given that this study identified ZNF217 as overexpressed in RUNX1-ETO-expressing cells, it is hypothesised that the block in normal cell differentiation might be mediated by ZNF217.
4.4.2 ZNF217 overexpression of promotes myeloid development in HSPC

Previous studies have demonstrated that increased levels of ZNF217 contribute to tumour development in several cancers, including breast and ovarian (Littlepage et al., 2012; Nonet et al., 2001; Plevova et al., 2010; Li et al., 2014; Li et al., 2007). However, its role in the development of leukaemia remains largely unknown. A recent study attempting to unravel the role of this TF in paediatric B-ALL has identified ZNF217 as a crucial mediator of the oncogenic role observed in these patients, and essential for the survival of B-ALL cells (Qin et al., 2019).

To determine whether ZNF217 equally contributes to the pathogenesis of AML by disrupting normal myeloid development, functional studies were performed. To determine if overexpression ZNF217 in HSPC affected haematopoiesis and cell survival, this study performed a colony forming assay. ZNF217 was shown to significantly reduce myeloid colony formation under clonal conditions, suggesting that this TF influences the proliferative ability of these cells. A subsequent replating assay yielded similar results, in which ZNF217-overexpressing cells generated significantly less colonies as compared to control cultures. These observations indicate that expression of ZNF217 in HSPC inhibits colony formation and decreases the self-renewal potential of myeloid progenitor cells, potentially associated with a more mature phenotype. Previous studies have shown that overexpression of ZNF217 in primary mammary epithelial cells resulted in enhanced formation of mammospheres associated with an increased self-renewal ability (Vendrell et al., 2012; Littlepage et al., 2012; Nguyen et al., 2014). However, this study’s results contrast with these observations and suggest, instead, that ZNF217 promotes HSPC differentiation in this system. However, once more, the models used to perform these experiments were not equivalent.

Subsequently, this study analysed the consequences of ZNF217 overexpression on myeloid development. In this context, overexpression of ZNF217 significantly reduced HSPC growth, especially within the monocytic and granulocytic populations, suggesting that ZNF217 does not act as an oncogene in a haematopoietic setting. ZNF217 has been previously associated with increased proliferation of ovarian and breast cancer cells in vitro (Li et al., 2014; Thollet et al., 2010), with similar observations demonstrated in vivo, where constitutive expression of this TF promoted the growth and rate of tumour formation (Vendrell et al., 2012; Li et al., 2014; Thollet et al., 2010; Littlepage et al., 2012), which this study did not observe. A reduction in cell growth can be a result of several factors, including changes in cell cycle, increased
apoptotic rate or differentiation processes. In other models, the increased proliferation of ZNF217-overexpressing cells was accompanied by an increase in the S phase of the cell cycle (Li et al., 2014). In this model, the decreased growth of the cells was actually associated with a significant reduction in the proportion of cells in the S / G2 phase, suggesting an arrest in normal cell cycle arising from the expression of ZNF217.

The consequences on cell differentiation upon ZNF217 expression were assessed through the expression of both monocytic and granulocytic maturation markers. Overexpression of ZNF217 promoted significant increases in the expression of monocytic markers CD11b and CD14, accompanied by a significant reduction in CD34 in early time points, indicating that the reduced proliferation of monocytic progenitors is a result of the differentiation process induced by the expression of ZNF217. The impact on granulocytic development, on the other hand, was less clear. Even though expression of ZNF217 resulted in the upregulation of CD11b in these cells, increase of CD15 was not observed, hence drawing a meaningful conclusion is not possible and additional analysis of other cell surface markers would need to be conducted. Altogether, these results suggest that ZNF217 is mainly involved in monocytic differentiation and does not contribute to leukaemic development. Additionally, these results do not agree with the analysis described above (4.4.1), which suggested an important role for ZNF217 in granulocytic development. Moreover, a role for ZNF217 in differentiation directly contrasts with preceding studies performed in other tumours. For instance, overexpression of ZNF217 in breast cancer cells was found to promote the repression of an adult stem cell expression signature (Littlepage et al., 2012). Additionally, expression of this TF in CSC was shown to be associated with the dysregulation of genes involved in the differentiation and maintenance of these cells (Krig et al., 2007; Vendrell et al., 2012; Littlepage et al., 2012). In summary, overexpression of ZNF217 in HSPC resulted in an increased differentiation potential, associated with a reduced proliferative ability and colony forming efficiency, leading to the hypothesis that overexpression of this TF is actually promoting haematopoietic development and not contributing the leukaemogenesis. These observations have been summarised in Table 4.1.
Table 4.1 - Summary of the main findings regarding the role of ZNF217 in normal human haematopoiesis

Table summary detailing the consequences of ZNF217 modulation vs control. * denotes significant differences in ZNF217 overexpression / knockdown, as compared to normal control cells.

<table>
<thead>
<tr>
<th></th>
<th>ZNF217 Overexpression</th>
<th>ZNF217 Knockdown</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Monocytes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growth</td>
<td>Decrease *</td>
<td>No change</td>
</tr>
<tr>
<td>Markers</td>
<td>Increase in CD11b and CD14 *</td>
<td>No change</td>
</tr>
<tr>
<td><strong>Granulocytes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Growth</td>
<td>Decrease *</td>
<td>No change</td>
</tr>
<tr>
<td>Markers</td>
<td>Increase in CD11b and CD15 *</td>
<td>No change</td>
</tr>
<tr>
<td>Apoptosis</td>
<td>No change</td>
<td>No change</td>
</tr>
<tr>
<td>Cell Cycle</td>
<td>Increase in G1 *</td>
<td>No change</td>
</tr>
<tr>
<td>CFU</td>
<td>Decrease *</td>
<td>Decrease *</td>
</tr>
</tbody>
</table>
4.4.3 ZNF217 knockdown in HSPC does not impact myeloid development

To determine if ZNF217 was necessary for myeloid development, an shRNA-based approach was used by lentivirally transducing HSPC with a scrambled shRNA control or three ZNF217 shRNA constructs. Since overexpression of ZNF217 promoted monocytic and granulocytic differentiation, it was hypothesised that knocking it down in the same cells would lead to a block in normal myeloid development associated with an immature phenotype. KD of ZNF217 significantly inhibited myeloid colony formation, associated with a decreased self-renewal potential. These suggest that loss of ZNF217 does not promote the retention of an immature phenotype.

The effects of ZNF217 KD on myeloid development were also assessed in liquid bulk culture. Significant differences in the overall growth were observed, with HSPC derived from frozen units presenting a 10-fold increase in growth, as compared to cells originated from fresh blood. For this reason, growth analysis solely reflects the results obtained from two sets of experiments, which did not allow statistical analysis. KD of ZNF217 was shown to exclusively impact monocytic growth, consistent with a role for this TF in monocytic development. Expression of cell surface markers showed minimal impact throughout myeloid development, suggesting that, even though ZNF271 might possess a role in myeloid differentiation, this is not critical for maturation to occur. These observations have been summarised in Table 4.1.

In conclusion, this study hypothesised the overexpression of ZNF217 observed in RUNX1-ETO expressing cells resulted in the disruption of normal myeloid development. This hypothesis was strengthened by the fact that overexpression of ZNF217 had previously been observed in several cancer types, and the assumption that ZNF217 played a key role in tumorigenesis by orchestrating tumour progression. This included the involvement of ZNF217 in the major hallmarks of cancer, including cell proliferation, replicative immortality, resistance to cell death or changes in cell cycle. However, this study found that, instead of possessing a pro-leukaemic phenotype, overexpression of ZNF217 in HSPC resulted in monocytic differentiation, inconsistent with previous studies. Moreover, even though KD of ZNF217 impacted monocytic growth, no impact was observed regarding their differentiation ability, suggesting that this TF is not essential for this process to occur.
Chapter 5

Identification of C/EBPβ Suppression by RUNX1-ETO as a Potential Mediator of a Block in Differentiation
5.1 Introduction

The t(8;21), responsible for the expression of the fusion protein RUNX1-ETO, occurs in approximately 12% of all AML cases. Ectopic expression of RUNX1-ETO in a primary human haematopoietic model has been shown to result in a block in myeloid development, associated with an increase in the cells’ self-renewal potential (Tonks et al., 2003; Tonks et al., 2004) (1.3.3.4). A subsequent study by our group demonstrated that expression of this fusion gene leads to dysregulation in the transcriptome of HSPC (Tonks et al., 2007). However, whilst mRNA studies are routinely used for target identification, it is not a powerful predictor of protein expression (Singh and Sharma, 2020). Therefore, this current study performed SWATH-MS to determine quantitative changes in the proteome of human CD34+ HSPC expressing RUNX1-ETO compared to normal control cells (Chapter 3). Expression of RUNX1-ETO in these cells resulted in the dysregulated expression of several TF, including PU.1 and CBFβ, confirming previous studies (Vangala et al., 2003; Kwok et al., 2009; Roudaia et al., 2009). However, RUNX1-ETO expression was also shown to result in the repression of C/EBPβ, a member of the C/EBP family of TF (Figure 3.18), an observation not previously reported.

Generally, members of the C/EBP family are viewed as tumour suppressor genes, due to their ability to repress cell growth and respond upon DNA damage events (1.5.1). However, in certain cell types, or depending on the protein isoform present, members of this family can possess opposing effects on the cells’ proliferative ability. Dysregulation of C/EBPα has been demonstrated in several solid tumours, including breast and lung cancer (Lourenço and Coffer, 2017). Its role in hepatocellular carcinoma and hepatoblastoma, however, has been the subject of multiple studies, as discordant results have been observed. Whilst some studies show that expression of C/EBPα is downregulated in these cancer cells and is associated with a worst OS (Tomizawa et al., 2003; Tseng et al., 2009), other studies claim that a poorer OS is linked to its overexpression (Lu et al., 2015; Gray et al., 2000). This suggests that the role of C/EBPα is context dependent and may have a tumour suppressor or promoter role (at least in liver cancers). In other solid tumours, C/EBPβ has been shown to be frequently upregulated in aggressive types of breast cancer, correlating with poor prognosis, metastasis, and high tumour grade (van de Vijver et al., 2002; Zahnow, 2009). This was further associated with an increased ratio of C/EBPβ LIP:LAP isoform ratio (Zahnow, 2009; Arnal-Estanpé et al., 2010) (1.5.3). C/EBPδ, on the other hand, is a tumour promoter in brain cancer, overexpressed in
glioblastoma cells and associated with poor prognosis (Cooper et al., 2012; Carro et al., 2010). C/EBPδ mRNA levels have been shown to be overexpressed in cancer cells expressing high levels of STAT3 which is frequently hyperactivated in cancer (Zhang et al., 2007; Silva, 2004). Additionally, expression of C/EBPγ has previously been shown to inhibit cellular senescence and promote cell proliferation in several solid tumours, further associating with poor clinical prognosis (Huggins et al., 2013).

Despite their roles as tumour promoter genes, some members of the C/EBP family have been shown to be important in normal haematopoiesis, including myeloid development. For instance, C/EBPα has been shown to be important for granulocytic and monocytic differentiation (Suh et al., 2006), and its dysregulation has been shown to block the transition from CMPs to GMPs, resulting in the loss of mature granulocytes (Zhang et al., 1997). This was subsequently shown to be associated with an increase in proliferation, increase in number of functional LT-HSC, and repopulating ability (Zhang et al., 2004b). C/EBPβ, on the other hand, is essential for macrophage and B-cell differentiation (Ruffell et al., 2009; Chen et al., 1997). Additionally, myeloid differentiation is preceded by the overexpression of C/EBPδ, followed by the upregulation of G-CSF and repression of C-Myc (Agrawal et al., 2007; Wang and Friedman, 2002). In fact, mature granulocytes express high levels of C/EBPδ (Scott et al., 1992; Gery et al., 2005). This is similar to macrophages, in which this TF is responsible for the regulation of genes associated with M1 macrophage polarisation (Banerjee et al., 2013). Lastly, C/EBPε is exclusively expressed in myeloid cells, and is required for the terminal differentiation of both neutrophilic and eosinophilic cells; however, expression of this protein and different isoforms has the ability to reprogram myeloid lineage commitment and differentiation (Koike et al., 1997).

Several studies have reported the dysregulation of members of the C/EBP family and the development of AML. Downregulation of C/EBPα has been previously shown to play a role in leukaemogenic development, and has been reported in AML, leading to a block in myeloid differentiation (Pabst et al., 2001; Rosenbauer and Tenen, 2007; Perrotti et al., 2002; Zheng et al., 2004a). Furthermore, mutations in the CEBPA gene have equally been associated with the development of AML (Paz-Priel and Friedman, 2011). Hypermethylation of another member of this family, C/EBPδ, has previously been associated with its low expression levels in AML (Agrawal et al., 2007). No association has been established to date regarding loss of C/EBPε and the development of AML, even though allelic loss of this TF was detected in AML patients.
Lastly, whilst different C/EBPβ protein isoforms have been shown to possess multiple functions, relying on their ratio to each other, there is little information regarding the role of C/EBPβ in AML (1.5.4). C/EBPβ regulates macrophage function, as well as emergency granulopoiesis and neoplastic transformation (Hirai et al., 2006; Tanaka et al., 1995a; Natsuka et al., 1992). In APL, C/EBPβ promotes ATRA-induced cell differentiation. These studies suggest that high levels of this protein result in transcriptional changes able to overcome the block in the differentiation process (Duprez et al., 2003). Similarly, ectopic expression of C/EBPβ in BCR-ABL-expressing CML cells promotes granulocytic cell differentiation (Guerzoni et al., 2006). However, the pathophysiological role of C/EBPβ in AML, including t(8;21), as well as its role in haematopoietic development remains largely unknown.

5.2 Hypothesis and Aims

This study hypothesises that downregulation of C/EBPβ expression will disrupt normal human myeloid development. This chapter’s main aim is to understand the role of C/EBPβ in haematopoietic development and its possible contribution to the development of AML, using a normal human primary HSPC model, as well as AML cell lines. This will be achieved by performing the following objectives:

**Determine the expression of CEBPB during normal human myeloid development and in AML patient blasts.**

Analysis of publicly available transcriptomic datasets will be performed to determine CEBPB mRNA expression in normal human HSPC subsets and across different AML subtypes.

**Determine the effects of C/EBPβ overexpression and KD on myeloid colony forming ability and self-renewal.**

Limiting-dilution colony forming assays will be performed on transduced human CD34+ HSPC and compared to control. Replating assays will be performed to assess the cells’ self-renewal potential.

**Determine the effect of C/EBPβ overexpression or KD on normal human haematopoietic growth, differentiation, and development.**
CD34+ HSPC will be transduced with a C/EBPβ-overexpression or -KD vectors and cultured in bulk liquid culture. Changes in monocytic and granulocytic growth will be assessed by analysing the cells’ lineage and cell surface markers using multicolour flow cytometry.

To determine the effect of C/EBPβ overexpression or KD on AML cell growth, proliferation, and apoptosis.

AML cell lines will be transduced with a C/EBPβ-overexpression or -KD vectors and changes in cell growth will be evaluated by following the cells’ proliferative ability over time. Furthermore, transduced cells will be assayed to determine if C/EBPβ dysregulated expression results in changes in cell cycle and apoptotic frequency.

5.3 Results

5.3.1 Expression of CEBPB mRNA increases throughout myeloid cell development

Members of the C/EBP family of TF, including C/EBPα, have been shown to play important roles in haematopoietic development (1.5.2); however, the expression and function of C/EBPβ in myeloid cell development has not yet been fully established. In order to determine the expression of CEBPB in human haematopoietic cells, publicly available microarray data was analysed using the online repository Bloodspot (Bagger et al., 2016) (2.9.2.1). Figure 5.1A shows that the expression of CEBPB is lower in less differentiated cells, including HSC, MPP and CMP than differentiated cell types (mature granulocytes and monocytes). This data is supported by RNASeq analysis of the CEBPB mouse homologue, which showed a similar increase in mRNA expression in differentiated granulocytes, monocytes and macrophages compared to immature HSC (Figure 5.1B). These observations suggest the transcriptional regulation of CEBPB during haematopoietic development, and a possible role in myeloid differentiation.

To validate mRNA expression data, C/EBPβ protein expression was determined in human cord blood derived HSPC throughout myeloid development (2.3.3). Total protein lysates harvested over 13 days of myeloid culture were generated and western blot analysis performed. As shown in Figure 5.2A-B, the expression of the longer C/EBPβ isoform LAP* (1.5.1) was shown to be higher in early myeloid progenitors (day 3 of culture), decreasing over time. Moreover, expression of C/EBPβ LAP was solely detected on day 3 lysates (Figure 5.2A), whilst LIP was not detected through myeloid development. These observations contrast with
mRNA expression. However, these lysates were derived from a combination of all haematopoietic populations, including monocytic, granulocytic and erythroid, and these are not equally distributed within a bulk population. To address this, C/EBPβ expression in each lineage was assessed in total protein lysates from monocytic-, granulocytic- and erythroid-committed progenitor cells. These cell subpopulations were FACS based on the expression on the cell surface markers CD14 and CD36 (2.7.4). This analysis showed that, on day 6 of culture, C/EBPβ is expressed in monocytes and erythrocytes and, to a lesser extent, in granulocytes (Figure 5.2C-D). However, day 6 granulocytic-committed cells most likely correspond to late promyelocytes which have lower C/EBPβ expression as compared to monocytes. However, these results do not correspond to the mRNA analysis, and the fact that C/EBPβ expression is shown to decline over 13 days cannot be accounted for (5.4.1). Nevertheless, this analysis suggests that C/EBPβ is differentially expressed in the different progenitor committed populations, suggesting a specific role for this protein in myeloid cell development.

5.3.2 *CEBPB* is variably expressed across AML subtypes

Heterogeneity observed in AML is generally associated with molecular and chromosomal abnormalities (1.2.2). To determine the expression of *CEBPB* mRNA in different subtypes of AML, this study analysed publicly available transcriptomic datasets, and compared them with normal undifferentiated HSC and mature granulocytes/monocytes. Figure 5.3A shows that *CEBPB* expression is significantly increased in all subtypes of AML studied, as compared to normal HSC. However, mRNA expression was found to be significantly downregulated compared to normal terminally differentiated myeloid cells, suggesting that reduced *CEBPB* expression might contribute to a block in normal haematopoietic differentiation. *CEBPB* expression was further observed to be heterogeneously expressed across FAB disease sub-types (Figure 5.3B). Interestingly, within the M2 AML subtype, *CEBPB* expression was shown to be significantly lower in t(8;21) patients as compared to non-t(8;21) FAB-M2 patients (Figure 5.3C), suggesting a specific role for this gene in the development of AML t(8;21).
Figure 5.1 – CEBPB mRNA expression in normal haematopoiesis

Normalised microarray data (log₂) showing CEBPB mRNA expression in distinct human haematopoietic cells subsets based on cell surface marker expression. Normal human haematopoiesis data derived from (A) GSE24759 (Rapin et al., 2014; Svendsen et al., 2016). (B) Normalised microarray data (log₂) showing CEBPB mRNA expression across different murine haematopoietic cell types. Mouse normal haematopoiesis data derived from GSE60101 (Lara-Astiaso et al., 2014). Data indicates Mean ± 1SD. Statistical analysis was performed using one-way ANOVA with Bonferroni’s multiple comparisons test; *** denotes p<0.001; **** denotes p<0.0001, vs HSC (Probeset Cebpb)

HSC – Haematopoietic Stem Cell; MPP – Multipotential Progenitors; CMP – Common Myeloid Progenitor cell; MEP – Megakaryocyte / Erythroid Progenitor cell; GMP – Granulocyte / Monocyte Progenitors; Early PM - Early Promyelocyte; Late PM - Late Promyelocyte; MY – Myelocyte; MM – Metamyelocytes; BC – Band cell; PMN – Polymorphonuclear cell; Mono – Monocytes; Gran – Granulocytes; Macro – Macrophages.
Figure 5.2 – C/EBPβ endogenous expression throughout haematopoiesis

(A) Western blot analysis showing C/EBPβ expression in HSPC over 13 days in culture in IMDM supplemented with IL-3 SCF, G- and GM-CSF (3S low G/GM) (n=1). HEK 293T cells transduced with an overexpression C/EBPβ vector were used as a positive control (PC). GAPDH was used as a loading control. (B) Bar chart showing densitometry analysis of C/EBPβ expression in HSPC over 13 days of culture derived from (A). C/EBPβ expression was quantified using ImageJ, normalised to loading control (GAPDH) and to HSPC day 3. (C) Western blot analysis showing C/EBPβ expression in human CD34+ HSPC, and in monocytic (Mono, CD14^high); erythroid (Ery; CD14^lowCD36^high) and granulocytic (Gran; CD14^lowCD36^low) progenitors (Tonks et al., 2007) (n=1). GAPDH was used as a loading control. Reprobed membrane provided by Dr. Rachael Nicholson. (D) Bar chart showing densitometry analysis of C/EBPβ expression in HSPC, and respective myeloid populations derived from (C). C/EBPβ expression was quantified using ImageJ, normalised to loading control (GAPDH); PC – Positive control.
This study next analysed the TCGA 2013 RNASeq dataset (Ley et al., 2013) to determine OS and disease-free survival of AML patients with high and low CEBPB expression. Patients were stratified according to CEBPB expression using the upper and lower quartiles (high and low CEBPB expression, respectively) (Figure 5.4A). Patients with APL and those who did not receive standard treatment were excluded from this analysis. High CEBPB expression patients had a median survival of 12.4 months compared to 27 months in patients with low CEBPB expression (Figure 5.4B); however, this difference was not statistically significant. High CEBPB expression was, however, associated with a significant decrease in disease-free survival (DFS of 9.25 months, compared with 39 months for patients with low CEBPB expression) (Figure 5.4C). This was further associated with an increased risk of relapse, suggested by a hazard ratio of 2.4 (95% CI[1.2-4.8]).

To further elucidate the relationship between CEBPB expression and clinical outcome, the co-occurrence with disease characteristics and known prognostic indicators was analysed. High CEBPB expression was found to be associated with a lower percentage of PB blasts (Figure 5.5A), as well as increased WBC (Figure 5.5C), although no differences were observed in BM blast proportion (Figure 5.5B). Moreover, patients with high CEBPB expression were found to be significantly associated with a lower mutation count, as compared to low CEBPB expression (Figure 5.5D). Additionally, high CEBPB mRNA expression significantly correlated with higher age at diagnosis (Figure 5.6A). There were no significant differences in patient gender (Figure 5.6B). Interestingly, t(8;21) was solely linked to low expression levels (Figure 5.7A), which was also significantly associated with mutations in both the RUNXI and ETO genes (Figure 5.7B). These observations further strengthen the hypothesis that C/EBPβ downregulation plays a role in the development of AML t(8;21).
Figure 5.3 – CEBPB mRNA expression levels are variable across AML subtypes

(A) Normalised microarray data (log$_2$) showing CEBPB mRNA expression in normal human haematopoietic developmental subsets vs. AML subtypes. Normal human haematopoiesis data derived from GSE422519 (Rapin et al., 2014; Svendsen et al., 2016); human AML cells derived from GSE13159 (Haferlach et al., 2010; Kohlmann et al., 2008), GSE15434 (Klein et al., 2009), GSE61804 (Metzelder et al., 2015), GSE14468 (Wouters et al., 2009; Taskesen et al., 2015; Taskesen et al., 2011) and TCGA (Ley et al., 2013) (data analysed using Bloodspot’s algorithm Bloodpool (Bagger et al., 2016)). Normalised microarray data (log$_2$) showing CEBPB mRNA expression in (B) AML according to FAB subtype and (C) AML FAB2, categorised into t(8;21) and non-t(8;21), within the TCGA dataset (Ley et al., 2013) All studies were analysed using the 212501_at probeset. Data indicates mean ± 1SD. Significant differences were analysed by one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01; *** denotes p<0.001; **** denotes p<0.0001. AML normal – AML with normal karyotype; AML complex – AML with complex karyotype.
Figure 5.4 – High levels of *CEBPB* is associated with poorer disease-free survival of AML patients

(A) Normalised (log2) *CEBPB* mRNA expression according to upper (n=27) and lower quartiles (n=27). Significance determined by one-way ANOVA; **** denotes p<0.0001. (B) Overall survival and (C) disease-free survival analysis of AML patients stratified according to upper and lower *CEBPB* mRNA expression quartiles. Untreated patients and t(15;17) AML patients (that present a different treatment regime compared to other AML subtypes) were excluded from this analysis. Statistical analysis was performed using the Long-Rank test between high and low *CEBPB* expression groups. Data obtained from TCGA. (Ley *et al*., 2013) using cBioPortal ([www.cbioportal.org](http://www.cbioportal.org)).
Figure 5.5 – High CEBPB expression is associated with increased WBC and lower mutation count

Relationship between high and low CEBPB expression and several clinical attributes of AML patients. Data obtained from TCGA (Ley et al., 2013) using cBioPortal (www.cbioportal.org). CEBPB upper quartile (n=27); CEBPB lower quartile (n=27). Graph showing the percentage of (A) peripheral blood (PB) blasts; (B) BM blasts; (C) white blood count (WBC); and (D) mutation count of AML patients according to CEBPB high (n=27) and low (n=27) expression. Statistical differences between high and low CEBPB expression in AML patients was analysed using Mann-Whitney test; Data represents mean ± 1SD; * denotes p<0.05; ** denotes p<0.01.
Figure 5.6 – Dysregulated CEBPB expression is associated with patient characteristics

Relationship between high and low CEBPB expression and AML patient characteristics. Data obtained from TCGA (Ley et al., 2013) using cBioPortal (www.cbioportal.org). CEBPB upper quartile (n=27); CEBPB lower quartile (n=27). Graph showing the (A) age at diagnosis (years) and (B) proportion of male and female patients according to CEBPB high (n=27) and low (n=27) expression. Statistical differences between high and low CEBPB expression in AML patients was analysed using Mann-Whitney test; Data represents mean ± 1SD; * denotes p<0.05.
Figure 5.7 – CEBPB expression correlates with cytogenetic and molecular abnormalities

Relationship between increased and reduced CEBPB mRNA expression and the occurrence of specific (A) cytogenic or (B) molecular abnormalities in AML patients, according to CEBPB high (n=27) and low (n=27) quartiles Data obtained from TCGA (Ley et al., 2013) using cBioPortal (www.cbioportal.org). Statistical analysis was performed using one-sided Fisher Exact test. * denotes p<0.05; ** denotes p<0.01.
5.3.3 **Knockdown of C/EBPβ influences normal myeloid development**

The above observations show that *CEBPB* expression increases throughout blood cell development and lower levels of *CEBPB* are associated with t(8;21) in AML patients. *In vitro* studies have shown that both the C/EBPβ mRNA and protein are downregulated in RUNX1-ETO-expressing CD34+ HSPC, as compared to control (Figure 3.18). For these reasons, this study next determined if C/EBPβ is required for normal haematopoietic development. C/EBPβ was knocked down in normal human cord blood derived CD34+ HSPC and the effect on myeloid cell growth, differentiation and self-renewal was assessed. Moreover, ectopic expression of this protein for overexpression studies was also performed, as described in 5.3.4.

5.3.3.1 **Generation of C/EBPβ-knockdown CD34+ HSPC**

To assess the effect of C/EBPβ KD in CD34+ HSPC, lentiviral vectors based on shRNA constructs, in combination with a GFP selectable marker were used (Table 2.1). Transduction efficiency in HSPC using the four shRNA vectors was assessed using GFP expression. As shown in Figure 5.8A, transduction frequencies ranged between 60-70% across all constructs. To validate protein KD by western blot, cells were sorted by FACS for GFP expression. On day 6 of culture, C/EBPβ LAP* and LAP isoforms were detected in normal human myeloid cells, with no detectable levels of LIP (Figure 5.8B). KD with shC/EBPβ #1 had no detectable effect on endogenous levels of the C/EBPβ protein, whilst shC/EBPβ #2 and #3 showed a reduction of 40-50% of LAP* protein expression, and a 45% reduction in LAP, as compared to shRNA control (no mammalian target) (Figure 5.8C).

5.3.3.2 **Knockdown of C/EBPβ leads to a decrease in HSPC colony forming ability**

To determine the effect of C/EBPβ KD on myeloid colony forming efficiency and self-renewal potential, a colony assay, followed by subsequent replating was performed on GFP+ sorted cells (2.3.4). Following one week of culture, all three shC/EBPβ cultures had significantly less myeloid colonies as compared to control; showing a reduction of 25%-36% (Figure 5.9A). Moreover, all three cultures showed a significant decrease in self-renewal capacity (Figure 5.9B). The fact that one of the knockdown constructs, shC/EBPβ #1, was shown to inhibit colony formation, even though no level of KD was detected (Figure 5.8), suggests off-target effects, further discussed in 5.4. Nevertheless, these observations suggest that C/EBPβ KD reduces myeloid colony formation with decreased self-renewal potential.
Figure 5.8 – C/EBPβ is knocked-down in human CD34+ HSPC

(A) Summary data showing percentages of GFP positivity in control and C/EBPβ KD CD34+ HSPC populations, analysed on day 3 of culture. Gating strategies were applied as described in Figure 2.4. Data indicates mean ± 1SD (n=4). (B) Western blot analysis showing C/EBPβ protein expression in FACS HSPC transduced with a control or C/EBPβ-targeted shRNA construct (Table 2.1), on day 6 of culture (n=1). HEK 293T transfected with a C/EBPβ-overexpression construct was used as a positive control (PC); GAPDH was used as a loading control. (C) Bar chart showing densitometry analysis regarding C/EBPβ expression in HSPC transduced with three shC/EBPβ constructs. C/EBPβ expression was quantified using ImageJ, normalised to loading control (GAPDH) and to shRNA control (n=1).
(A) Myeloid colony forming efficiency of control and three shC/EBPβ cultures following 7 days of growth in liquid culture containing IL-3, SCF, G- and GM-CSF (2.3.4). Data indicates mean ± 1SD (n=3). Significant difference between shC/EBPβ cultures and control was analysed by one-way ANOVA with Bonferroni’s multiple comparisons correction; ** denotes p<0.01, *** denotes p<0.001.

(B) Self-renewal potential was assessed by a single replating round of control and shC/EBPβ cultures, in the same conditions as previously. Data indicates mean ± 1SD (n=3). Significant differences between shC/EBPβ cultures and control were analysed by one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01; *** denotes p<0.001.

Figure 5.9 – C/EBPβ KD disrupts myeloid colony formation of HSPC
5.3.3.3 Knockdown of C/EBPβ suppresses myeloid cell growth

This study next determined the consequences of reduced C/EBPβ expression on cell proliferation and differentiation. With no influence on C/EBPβ expression levels, shC/EBPβ #1 had little effect on myeloid cell growth in a bulk population setting (Figure 5.10A). In contrast, there was up to a significant 2.8-fold reduction in HSPC growth transduced with other shC/EBPβ. Monocytic and granulocytic cumulative cell growth was significantly impaired by approximately 2- and 5-fold, respectively, when analysed on day 13 of culture in KD #2 and #3 cells (Figure 5.10B-C). Even though erythroid growth showed significant reduction in cells transduced with shC/EBPβ #2 and #3, this was deemed not significant by day 13 of culture (Figure 5.10D). However, these cultures are not designed to examine erythroid growth given the lack of EPO. These observations suggest that KD of C/EBPβ influences the growth of myeloid cells, particularly in cells committed to the monocytic and granulocytic lineages.

5.3.3.4 Knockdown of C/EBPβ influences myeloid differentiation

Differentiation in each myeloid lineage was determined using myeloid differentiation cell surface markers, including CD34, CD11b, CD14 and CD15 (4.3.3.4). Over 13 days of culture, KD of C/EBPβ resulted in a significant increase in the proportion of monocytic cells during myeloid development, culminating in an increase of 10%, 15% and 14% observed with shC/EBPβ #1, #2 and #3, respectively, as compared to control cells (Figure 5.11A). This increase was likely at the expense of significant decreases in the granulocytic population, reduced on day 13 of development by 9%, 14% and 13% in cultures transduced with shC/EBPβ #1, #2 and #3, correspondingly (Figure 5.11B). No significant differences were observed in the erythroid-committed population (Figure 5.11C); even though these cells have the ability to grow within the same media as the other haematopoietic constituents, terminal erythroid differentiation is impaired due to lack of EPO.

Regarding monocytic development, even though there was a significant impairment of growth in cells transduced with shC/EBPβ #2 and #3, analysis of monocytic differentiation markers showed a modest increase of 1.3 and 1.6-fold in the expression of the cell surface markers CD11b and CD14, respectively (Figure 5.12A-C).
Figure 5.10 – KD of C/EBPβ decreases granulocytic growth in myeloid development

(A) Cumulative fold-expansion of control and three shC/EBPβ constructs for GFP positive cells in culture medium containing IL-3, SCF, G- and GM-CSF, grown over 13 days (Supplementary Figure 16). (B) Cumulative fold-expansion of monocytic cells (CD13⁺ CD36⁺) transduced with a control or shC/EBPβ vectors. (C) Cumulative fold-expansion of granulocytic cells (CD13⁻/⁺ CD36⁻) transduced with a control or shC/EBPβ vector. (D) Cumulative fold-expansion of erythroid cells (CD13⁻ CD36⁺) transduced with a control of shC/EBPβ vector. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Figure 2.4 and Figure 2.5. Data indicates mean ± 1SD (n=4). Significant differences between shC/EBPβ cultures and control at each time-point were analysed by one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01; *** denotes p<0.001; **** denotes p<0.0001; (this data has also been represented as dot plots, within the supplementary section [Supplementary Figure 38]).
Summary data showing percentage of (A) monocytes (CD13+ CD36+), (B) granulocytes (CD13+/− CD36) and (C) erythroid cells (CD13− CD36+) in control and three shC/EBPβ cultures. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Figure 2.4 and Figure 2.5. Data indicates mean ± 1SD (n=4). Significant differences between shC/EBPβ cultures and control at each time-point were analysed by one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01; *** denotes p<0.001; (this data has also been represented as dot plots, within the supplementary section [Supplementary Figure 39]).
Figure 5.12 – KD of C/EBPβ has little impact on normal monocytic differentiation

(A) Summary data of CD34+ expression in terms of percentage in monocytic cells (CD13+ CD36+) over time for control and shC/EBPβ cells. Summary data showing (B) CD11b and (C) CD14 expression in terms of MFI in monocytic cells over time for control and shC/EBPβ cells. (D) Granularity of monocytic cells was measured by changes in SSC for control and shC/EBPβ cultures on days 10 and 13 of myeloid differentiation. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Figure 2.4 and Figure 2.5. Data indicates mean ± 1SD (n=4). Significant differences between shC/EBPβ cultures and control at each time-point were analysed by one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01; (for representative flow cytomtery plots please refer to Supplementary Figure 17, Supplementary Figure 18, Supplementary Figure 19; this data has also been represented as dot plots, within the supplementary section [Supplementary Figure 40]).
Moreover, monocytic cell surface marker upregulation was accompanied by a significant increase in the SSC of these cells, implying cells transduced with constructs #2 and #3 have an increased granularity and complexity, as compared to shRNA control cells (Figure 5.12D). Altogether, these observations suggest that the reduction in monocytic proliferation maybe a result of increased differentiation; however, this increase was modest.

Regarding granulocytic development, KD of C/EBPβ did not significantly influence the expression of the CD34 marker nor CD11b (Figure 5.13A-B). However, shC/EBPβ #2 and #3-transduced cells showed a significant upregulation of the granulocytic marker CD15 throughout myeloid development, culminating in an increase of 1.3-fold observed in shC/EBPβ #3 (Figure 5.13C). No differences were observed in terms of granularity and complexity upon KD of C/EBPβ (Figure 5.13D). These results indicate that normal levels of C/EBPβ are not critical for granulocytic development.

Morphological assessment of transduced cells was performed using May-Grünwald-Giemsa staining of cells cultured under myeloid conditions for 17 days (Figure 5.14). This analysis scores post-mitotic developmental changes which are difficult to measure by flow cytometry. The proportion of granulocytes was markedly increased in all conditions, as compared to monocytes. shC/EBPβ cultures showed a decrease in the percentage of terminally differentiated granulocytic cells. These suggest that KD of C/EBPβ may impact late granulocytic differentiation.

shC/EBPβ cultures presented a significant decrease in the percentage of monocytic population, in accordance with lineage analysis (Figure 5.11A-B). These observations support the hypothesis that KD of C/EBPβ has a role in promoting monocytic differentiation, as fully differentiated cells will lose their self-renewal potential, reaching terminal maturation earlier than shRNA control cells.

Overall, these observations do not agree with the mRNA results described above (5.3.1), which suggested a role for C/EBPβ in haematopoietic development. However, this maybe a result of several factors, including poor level of KD, as well as the interplay with other members of the C/EBP family of TF, further discussed in section 5.4.
Figure 5.13 – KD of C/EBPβ promotes granulocytic development

(A) Summary data of CD34⁺ expression in terms of percentage in granulocytic cells (CD13⁺/CD36⁻) over time for control and shC/EBPβ cells. Summary data showing (B) CD11b and (C) CD15 expression in terms of MFI in granulocytic cells over time for control and shC/EBPβ cells. (D) Granularity of granulocytic cells was measured by changes in SSC for control and shC/EBPβ cultures on days 10 and 13 of myeloid differentiation. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Figure 2.4 and Figure 2.5. Data indicates mean ± 1SD (n=4). Significant differences between shC/EBPβ cultures and control at each time-point was analysed by one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01; *** denotes p<0.001; (for representative flow cytometry plots please refer to Supplementary Figure 20, Supplementary Figure 21, Supplementary Figure 22; this data has also been represented as dot plots, within the supplementary section [Supplementary Figure 41]).
Figure 5.14 – C/EBPβ KD affects cell morphology during myeloid development

(A) Representative images showing control and shC/EBPβ cultures, analysed on day 17 of differentiation, by cytopinning 3x10^4 cells and staining them with May-Grünwald-Giemsa differential stain. (B) Differential counts of all cultures with morphology categorised into granulocytic and monocytic-committed cells, according to Figure 2.2. Approximately 500 cells were counted in total and each population count normalised to total number of cells. Granulocytic cells were further divided into myelocytes, metamyelocytes and band/segmented cells. Data indicates mean ± 1SD (n=3). Significant difference between shC/EBPβ cultures and control was analysed by one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01.
5.3.3.5 Knockdown of C/EBPβ does not result in significant cell cycle or apoptotic changes

Having determined that C/EBPβ KD significantly reduces myeloid cell proliferation, this study next determined if the reduction in proliferation could be a result of changes to the cells’ normal cell cycle, or due to an increase in apoptosis. Only one shC/EBPβ culture showed a modest accumulation of cells in the G1 phase of the cycle (5% in shC/EBPβ #2) (Figure 5.15A), suggesting a slight reduction in rate of cell cycle progression.

An Annexin-V assay was further performed to assess whether C/EBPβ induced apoptosis in these cells. No changes were observed in the apoptotic frequency of HSPC, upon KD of C/EBPβ (Figure 5.15B). Altogether, the reduced proliferation observed in shC/EBPβ cultures is not a result of either significant changes to the cells’ normal cycle nor due to apoptotic events.

5.3.4 Overexpression of C/EBPβ influences normal myeloid development

Having determined the effect of C/EBPβ KD on normal human haematopoiesis, this study then aimed at analysing the consequences of C/EBPβ overexpression using the same human cord blood model.

5.3.4.1 Generation of C/EBPβ-overexpressing HSPC

To promote the ectopic expression of the C/EBPβ LAP* protein in HSPC, a C/EBPβ overexpression construct was used (Table 2.1), coupled with a GFP selectable marker, to allow for cell sorting. HSPC were transduced with the control or overexpression construct and, on day 3 of cell culture, efficiency of infection was evaluated by flow cytometry through GFP expression. Efficiency of transduction of these cultures was c25% for cells transduced with a control vector and c18% of GFP positivity for the C/EBPβ overexpression construct (Figure 5.16A). To validate C/EBPβ overexpression in these cells through western blot analysis, cells were sorted by FACS based on their GFP expression. C/EBPβ transduced cells strongly overexpressed the C/EBPβ LAP* protein, as compared to the control cells (Figure 5.16B-C).
Figure 5.15 – KD of C/EBPβ has no significant effect on normal cell cycle or apoptosis

(A) Graph showing cell cycle distribution of normal CD34+ HSPC (Control) and three shC/EBPβ cultures after 6 days of culture. Initial gating was applied as described in Figure 2.6. Data indicates mean ± 1SD (n=3). (B) Summary data showing the effect of C/EBPβ KD on apoptosis, in HSPC following 6, 10 and 13 days of culture. Initial gating was applied as described in Figure 2.7: (i) Early apoptotic cells are characterised by Annexin V+/PI- and (ii) Apoptotic cells are characterised by Annexin V+/PI+. Data indicates Mean ± 1SD (n=3). Significant difference between shC/EBPβ cultures and control was analysed by one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05.
Figure 5.16 – C/EBPβ overexpression in human HSPC

(A) Summary data showing percentages of GFP positivity in control and C/EBPβ overexpression CD34+ HSPC populations, analysed on day 3 of culture. Gating strategies were applied as described in Figure 2.4. Data indicates mean ± 1SD (n=4). (B) Western blot analysis showing C/EBPβ protein expression in sorted cells transduced with a control or C/EBPβ-overexpression construct, on day 15 of culture (n=1). HEK 293T transfected with a C/EBPβ-overexpression construct was used as a positive control (PC); GAPDH was used as a loading control. (C) Bar chart showing C/EBPβ expression in cells transduced with a C/EBPβ-overexpression construct. C/EBPβ expression was quantified using ImageJ, normalised to loading control (GAPDH) and to control (n=1).
5.3.4.2 \textit{C/EBPβ overexpression promotes myeloid colony formation and self-renewal ability}

Having generated and validated the overexpression of C/EBPβ in human HSPC, the consequences on myeloid colony forming capacity and self-renewal ability were examined by performing a limiting dilution assay. To do this, GFP\(^+\) sorted cells were cultured in the presence of IL-3, SCF, G- and GM-CSF (2.3.4). Under clonal conditions, the colony-forming efficiency was 4-fold higher in C/EBPβ cultures compared to control cells (Figure 5.17A). Serial replating of colony forming cells showed that C/EBPβ-overexpressing cells were able to generate 2-fold more colonies compared to control cells (Figure 5.17B). These results indicate that overexpression of C/EBPβ increases myeloid colony formation with and suggest an enhanced self-renewal potential.

5.3.4.3 \textit{C/EBPβ promotes myeloid cell growth}

Having determined that overexpression of C/EBPβ promoted myeloid colony formation, the growth and differentiation of HSPC was followed over 13 days. C/EBPβ overexpression promoted the expansion of myeloid cells (4-fold) compared to control cells, in bulk liquid culture (Figure 5.18A). Using the lineage discriminator markers CD13 and CD36 to discriminate haematopoietic lineages, C/EBPβ-overexpressing monocytes displayed a 3.5-fold increase in normal growth, compared to cells transduced with a control vector (Figure 5.18B). This increase was more evident in the granulocytic population, in which cell expansion was increased by 9-fold in C/EBPβ overexpression cells compared to control (Figure 5.18C). As expected, since the media used for these experiments does not support terminal erythroid differentiation, the growth of erythroid progenitor cells was not significantly altered (Figure 5.18D).
Figure 5.17 – C/EBPβ overexpression promotes myeloid colony formation of HSPC

(A) Myeloid colony forming efficiency of control and C/EBPβ-overexpression (OE) cultures following 7 days of growth in liquid culture containing IL-3, SCF, G- and GM-CSF (2.3.4). Data indicates mean ± 1SD (n=3). Significant difference between C/EBPβ-overexpression and control cultures were analysed by paired t-test; ** denotes p<0.01. (B) Self-renewal potential was assessed by a single replating round of control and C/EBPβ-OE cultures, in the same conditions as previously. Data indicates mean ± 1SD (n=3) Significant difference between C/EBPβ-overexpression and control cultures was analysed paired t-test; * denotes p<0.05.
Figure 5.18 – Overexpression of C/EBPβ promotes monocytic and granulocytic growth in myeloid development

(A) Cumulative fold-expansion of control and C/EBPβ-overexpression constructs in terms of GFP positivity in culture medium containing IL-3, SCF, G- and GM-CSF, grown over 13 days (Supplementary Figure 23). (B) Cumulative fold-expansion of monocytic cells (CD13+ CD36+) transduced with a control or C/EBPβ-OE vector. (C) Cumulative fold-expansion of granulocytic cells (CD13+CD36+) transduced with a control or C/EBPβ-OE vector. (D) Cumulative fold-expansion of erythroid cells (CD13+CD36+) transduced with a control of C/EBPβ-OE vector. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Figure 2.4 and Figure 2.5. Data indicates mean ± 1SD (n=4). Significant differences between C/EBPβ-overexpression and control cultures at each time-point were analysed by paired T-test; * denotes p<0.05; ** denotes p<0.01; (this data has also been represented as dot plots, within the supplementary section [Supplementary Figure 42]).
5.3.4.4 C/EBPβ expression promotes myeloid differentiation

KD of C/EBPβ disrupted the balance of haematopoietic populations, by increasing the proportion of monocytes, accompanied by a decrease in granulocytes (Figure 5.11). Overexpression of C/EBPβ, on the other hand, had a contrasting effect, significantly decreasing the proportion of the monocytic population, especially in the latter stages of myeloid development, coupled with an increase in the percentage of granulocytic committed cells throughout myeloid development (Figure 5.19A-B). Furthermore, C/EBPβ also contributed to a significant decrease in erythroid population, even though this was not the focus of the study (Figure 5.19C). Since cell proliferation and differentiation are often inversely correlated, it is possible that the increase in cell proliferation arising from C/EBPβ overexpression is due to an inhibition in HSPC differentiation in the monocytic and granulocytic lineages.

Immunophenotypic analysis of monocyte-progenitor cells showed no difference in the percentage of cells expressing the haematopoietic stem-cell marker CD34 (Figure 5.20A). However, a consistent increase above control in the expression of CD11b was observed in C/EBPβ expressing cells, culminating in a 1.5-fold upregulation in the expression of this marker compared to control (Figure 5.20B). Upregulation of the monocytic marker CD14 was further observed in these cells; however, this trend was solely detected within the first 8 days of myeloid growth, decreasing to similar levels to those of control cells afterwards (Figure 5.20C). No significant differences were observed in terms of granularity or complexity of C/EBPβ-overexpressing monocytes (Figure 5.20D). These observations suggest that, despite the significant increase in monocytic growth, C/EBPβ appears to promote monocytic differentiation. This phenotype further agrees with the mRNA and protein analysis, in which C/EBPβ expression increased with myeloid cell differentiation, particularly towards monocytic committed cells (Figure 5.1, Figure 5.2).

In contrast, C/EBPβ overexpression did not significantly impact the expression of CD34 or CD11b in granulocytic progenitor cells (Figure 5.21A-B). Overexpression of C/EBPβ in this population did, however, lead to the upregulation of the granulocytic marker CD15, up until day 10 of culture, in which a 1.5-fold upregulation was observed compared to control cells (Figure 5.21C). Similarly, to the monocytic population, no significant differences were observed in terms of cell granularity / complexity in these cells (Figure 5.21D).
Summary data showing the proportion of (A) monocytic (CD13+ CD36⁻), (B) granulocytic (CD13⁺/⁻ CD36⁻) and (C) erythroid cells (CD13⁻ CD36⁺) in control and C/EBPβ-overexpressing cultures. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Figure 2.4 and Figure 2.5. Data indicates mean ± 1SD (n=4). Significant differences between C/EBPβ-expressing cells and control at each time-point were analysed by paired t-test; * denotes p<0.05, ** denotes p<0.01, *** denotes p<0.001; (this data has also been represented as dot plots, within the supplementary section [Supplementary Figure 43]).
This study next sought to assess terminal stages of development, difficult to score by flow cytometry techniques only. This analysis showed that overexpression of C/EBPβ promotes granulocytic development, with a significant increase in the percentage of cells showing mature granulocytic traits, such as segmented nuclei and cytoplasmic granules (Figure 5.22). Moreover, the percentage of monocytes and macrophages in the C/EBPβ-overexpressing culture was significantly lower compared to the control population, in agreement with the lineage imbalance described above (Figure 5.19).

Altogether, these results suggest that ectopic expression of C/EBPβ LAP* in HSPC promotes monocytic and granulocytic differentiation, although the latter to a lesser extent. Moreover, the fact that C/EBPβ LAP* overexpression in HSPC promoted differentiation suggests that its upregulation is not likely to promote leukaemogenesis through the perturbation of myeloid cell development.

5.3.4.5 Overexpression of C/EBPβ induces change in the cell cycle of HSPC

Having determined that ectopic C/EBPβ expression leads to an increase in cell growth, not associated with an arrest in differentiation, this study aimed at analysing if these changes partially occurred as a result of variations to the normal cells’ cell cycle and apoptotic frequency. Cell cycle analysis showed that overexpression of C/EBPβ led to a significant decrease by 12% in the percentage of cells in the G1 phase of the cell cycle and, conversely, to an increase of cell in S/G2 (Figure 5.23A). In addition, no significant differences were observed regarding apoptosis (Figure 5.23B). In summary, this data suggest that overexpression of C/EBPβ in human HSPC promotes the expansion of monocytes and granulocytes, whilst promoting the differentiation of both lineages in a myeloid setting.
Figure 5.20 — C/EBPβ overexpression promotes the upregulation of monocytic markers in monocytic progenitors

(A) Summary data of CD34+ expression in terms of percentage in monocytic cells (CD13+ CD36+) over time for control and C/EBPβ-overexpressing cells. Summary data showing (B) CD11b and (C) CD14 expression in terms of MFI in monocytic cells over time for control and C/EBPβ-overexpressing cells. (D) Complexity of monocytic cells was measured by changes in SSC for control and C/EBPβ-overexpressing cells on days 10 and 13 of myeloid differentiation. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Figure 2.4 and Figure 2.5. Data indicates mean ± 1SD (n=4). Significant differences between C/EBPβ-expressing cells and control at each time-point were analysed by paired t-test; * denotes p<0.05, ** denotes p<0.01; (for representative flow cytometry plots please refer to Supplementary Figure 24, Supplementary Figure 25, Supplementary Figure 26; this data has also been represented as dot plots, within the supplementary section [Supplementary Figure 44]).
Figure 5.21 – C/EBPβ overexpression promotes the upregulation of granulocytic markers in granulocytic progenitors

(A) Summary data of CD34+ expression in terms of percentage in granulocytic cells (CD13+c/ CD36–) over time for control and C/EBPβ-overexpressing cells. Summary data showing (B) CD11b and (C) CD15 expression in terms of MFI in granulocytic cells over time for control and C/EBPβ-overexpressing cells. (D) Complexity of granulocytic cells was measured by changes in SSC for control and C/EBPβ-overexpressing cells on days 10 and 13 of myeloid differentiation. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Figure 2.4 and Figure 2.5. Data indicates mean ± 1SD (n=4). Significant differences between C/EBPβ-expressing cells and control at each time-point were analysed by paired t-test; ** denotes p<0.01; (for representative flow cytometry plots please refer to Supplementary Figure 27, Supplementary Figure 28, Supplementary Figure 29; this data has also been represented as dot plots, within the supplementary section [Supplementary Figure 45]).
Figure 5.22 – C/EBPβ overexpression alters cell morphology during myeloid development

(A) Representative images showing control and C/EBPβ OE cultures, analysed on day 17 of differentiation, by cytopinning 3x10^4 cells and staining them with May-Grünwald-Giemsa differential stain. (B) Differential counts of all cultures with morphology categorised into granulocytic and monocytic-committed cells, according to Figure 2.2. Approximately 500 cells were counted in total and each population count normalised to total number of cells. Granulocytic cells were further divided into myelocytes, metamyelocytes and band/segmented cells. Data indicates mean ± 1SD (n=3). Significant differences between C/EBPβ-expressing cells and control were analysed by paired t-test; * denotes p<0.05, ** denotes p<0.01.
Figure 5.23 – Overexpression of C/EBPβ influences normal cell cycle

(A) Bar chart showing cell cycle distribution of normal CD34+ HSPC (Control) and C/EBPβ-overexpression cultures after 6 days of culture. Initial gating was applied as described in Figure 2.6. The percentage of cells in each phase of the cell cycle is indicated by mean ± 1SD (n=3). Significant difference between C/EBPβ-overexpression and control cultures were analysed paired t-test; * denotes p<0.05; ** denotes p<0.01. (B) Summary data showing the effect of C/EBPβ overexpression on apoptosis in HSPC following 10 and 13 days of culture. Initial gating was applied as described in Figure 2.7; (i) Early apoptotic cells are characterised by Annexin V+ / PI- and (ii) Apoptotic cells are characterised by Annexin V+ / PI+. Data indicates mean ± 1SD (n=3).
5.3.5 The effect of C/EBPβ mis-regulation in AML cell lines

5.3.5.1 Expression of C/EBPβ is variable across AML cell lines

In order to determine the expression of C/EBPβ in AML cell lines, thus establishing whether these were suitable to study the biological role of this protein in AML, endogenous mRNA and protein levels were assessed in a cohort of AML cell lines. The CCLE database was firstly employed to determine the expression of CEBPB mRNA in AML cell lines. Since the CEBPB transcript consists of a single exon, giving rise to a single mRNA sequence, it is not possible to determine the mRNA expression of each of the three C/EBPβ isoforms. CEBPB was found to be heterogeneously expressed in all the cell lines tested, including in the t(8;21) cell lines Kasumi-1 and SKNO-1 (Figure 5.24A). Subsequently, C/EBPβ protein expression was determined using western blotting. Figure 5.24B shows three cell lines (U937, HEL and SKNO-1) as suitable for KD studies, and an additional three lines (THP-1, OCI-AML 2 and OCI-AML5) with lower levels of endogenous C/EBPβ appropriate for overexpression studies. Moreover, this analysis allowed the detection of two C/EBPβ isoforms, LAP* and LAP, in almost all the AML cell lines tested. Additional comparison between C/EBPβ mRNA and protein showed poor association between the two (Figure 5.24A). In summary, it was possible to identify suitable AML cell lines for to study the consequences of modulating C/EBPβ expression using overexpression, which could not be achieved in this study (5.4.5), or KD systems (5.3.5.2).

5.3.5.2 Generation of C/EBPβ-knockdown cell lines

The consequences of knocking down C/EBPβ in AML cell lines on proliferation and survival was determined. Using the same shRNA constructs (shRNA control and three shC/EBPβ vectors), three cell lines were selected for the generation of C/EBPβ-KD cultures based on the endogenous levels of C/EBPβ expression. These included the t(8;21) SKNO-1, the erythroleukaemia HEL and the leukaemic U937 cell lines. Successful cell line transduction was confirmed through GFP expression, once bulk cell population achieved 100% GFP positivity (Figure 5.25A; Figure 5.26A; Figure 5.27A), without the need to select them.

KD of C/EBPβ in the generated cell lines was validated through western blotting. In all three cell lines, cells transduced with shC/EBPβ #1 showed the lowest level of KD, similarly to that observed in HSPC (5.3.3.1) (Figure 5.25B; Figure 5.26B; Figure 5.27B).
Figure 5.24 – C/EBPβ expression in leukaemic cell lines

(A) Western blot analysis of C/EBPβ endogenous expression in a cohort of AML cell lines (n=1). The HeLa cell line was used as a positive control (PC); GAPDH was used as a loading control. (B) Normalised (log₂) CEBPB mRNA expression in a cohort of leukaemic cell lines. mRNA expression data derived from [https://depmap.org/portal/gene/CEBPB?tab=characterization](https://depmap.org/portal/gene/CEBPB?tab=characterization) (Barretina et al., 2012). Null protein values refer to undetectable levels of endogenous C/EBPβ, derived from (A).
Figure 5.25 – C/EBPβ KD in the t(8;21) cell line SKNO-1

(A) Representative flow cytometric plots showing transduction efficiency in the SKNO-1 cell line transduced with an shRNA control and three shC/EBPβ constructs (shC/EBPβ #1, shC/EBPβ #2, shC/EBPβ #3). (B) Western blot analysis showing successful C/EBPβ KD in all three shRNA constructs (n=1). GAPDH was used as a loading control; HeLa cells were used as a positive control (PC). (C) Western blot quantification by densitometry analysis using ImageJ. Relative protein expression was determined by normalising each sample firstly to the loading control (GAPDH), then to shRNA control (n=1).
Figure 5.26 – C/EBPβ KD in the erythroleukaemia cell line HEL

(A) Representative flow cytometric plots showing transduction efficiency in the HEL cell line transduced with an shRNA control and three shC/EBPβ constructs (shC/EBPβ #1, shC/EBPβ #2, shC/EBPβ #3). (B) Western blot analysis showing successful C/EBPβ KD in all three shRNA constructs (n=1). GAPDH was used as a loading control.; HeLa cells were used as a positive control (PC) (C) Western blot quantification by densitometry analysis using ImageJ. Relative protein expression was determined by normalising each sample firstly to the loading control (GAPDH), then to shRNA control (n=1).
Figure 5.27 – C/EBPβ KD in the leukaemic cell line U937

(A) Representative flow cytometric plots showing transduction efficiency in the U937 cell line transduced with an shRNA control and three shC/EBPβ constructs (shC/EBPβ #1, shC/EBPβ #2, shC/EBPβ #3). (B) Western blot analysis showing successful C/EBPβ KD in all three shRNA constructs (n=1). GAPDH was used as a loading control. HeLa cells were used as a positive control (PC) (C) Western blot quantification by densitometry analysis using ImageJ. Relative protein expression was determined by normalising each sample firstly to the loading control (GAPDH), then to shRNA control (n=1).
Constructs #2 and #3 showed similar levels of KD, with a decrease of c75%, c25% and c90% in the detectable levels of C/EBPβ LAP* in SKNO-1, HEL and U937 cells, respectively (Figure 5.25C; Figure 5.26C; Figure 5.27C), as compared to cells transduced with a control shRNA vector. The LAP isoform was shown to be completely KD in SKNO-1 and U937 cells; additionally, this isoform was not detected in the HEL cell line. The C/EBPβ LIP isoform was not detected in any of the cell lines tested.

5.3.5.3 The impact of C/EBPβ knockdown on the AML t(8;21) cell line SKNO-1

Having demonstrated that C/EBPβ was successfully knocked-down in all three cell lines, the phenotypic impact of this, including cell growth and differentiation, were assessed by flow cytometry. Firstly, this study analysed the t(8;21) cell line SKNO-1. In these cells, KD of C/EBPβ resulted in a 4- and 2.6-fold significant increase in cell growth, observed in cells transduced with shC/EBPβ #2 and #3, respectively (Figure 5.28A). Given the significant impact of C/EBPβ KD in these cells, this study further analysed their DNA content, to determine cell cycle progression. This showed an increase in the proportion of cells in the S/G2 phase of the cell cycle, as compared to control (Figure 5.28B). Subsequently, these lines were assayed for apoptosis using Annexin V and 7-ADD, showing a significant decrease in the percentage of cells in early and late apoptosis in cultures transduced with shC/EBPβ #2 and #3, compared to control (Figure 5.28C). Altogether, these results indicate that the increase in cell proliferation observed upon KD of C/EBPβ can be attributed to changes in normal cell cycle, together with a decreased apoptotic rate.

To determine if KD of C/EBPβ had any effect on the differentiation status of transformed lines, shC/EBPβ-transduced cells were analysed using immunophenotypic assays. A broad panel of markers associated both with immature HSPC (such as CD34), as well as markers of monocytic (CD11b, CD14), granulocytic (CD11b, CD15) and erythroid (GlyA) differentiation was used. According to this analysis, SKNO-1 control cells are typically CD13+CD33+CD38+CD45RA+, resembling a GMP phenotype (Figure 1.1) (Figure 5.29). However, upon KD of C/EBPβ, these cells lose the expression of all myeloid markers, being defined as CD13−CD33−CD38−CD45RA−, suggesting a more immature phenotype associated with HSC or MPP.
Figure 5.28 – C/EBPβ KD promotes cell growth in SKNO-1 cells, associated with changes in normal cell cycle and apoptotic rate

Summary data showing (A) the growth of SKNO-1 cells with three shC/EBPβ constructs compared to control, over 5 days of culture. Data indicates mean ± 1SD (n=4); (B) the effect of C/EBPβ KD in normal cell cycle. Data indicates Mean ± 1SD (n=3); (C) the effect of C/EBPβ KD on normal apoptotic rate using Annexin V staining. Bars indicate percentage of pre-apoptotic cells (Annexin V+/7-AAD−) and late apoptotic cells (Annexin V+/7-AAD+). Data indicates Mean ± 1SD (n=3). Significant differences between shC/EBPβ cultures and control were analysed using one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01; *** denotes p<0.001; **** denotes p<0.0001.
Following immunophenotypic analysis, it was hypothesised that the increase in cell growth, associated with the loss of myeloid surface marker expression could be correlated with the promotion of a more undifferentiated phenotype. Undifferentiated cells are usually bigger and less internally complex than differentiated cells, thus the morphology of transduced SKNO-1 cells was assessed. This showed that SKNO-1 cells transduced with shC/EBPβ #2 and #3 were larger than control cells, or cells transduced with shC/EBPβ #1, the weakest construct (Figure 5.30A). This was further confirmed by analysing the cells average area, as well as through the forward and side scatter properties of these cells, with all three parameters showing a significant increase in cells transduced with shC/EBPβ #2 and #3, as compared to the shRNA control vector (Figure 5.30B-D).

Altogether, these results suggest that KD of C/EBPβ in the AML t(8;21) cell line SKNO-1 promotes cell growth, associated with an increase of cells in the S phase of the cell cycle, as well a decrease in the proportion of cells in early and late apoptotic stages. Moreover, KD of this protein was further associated with the ablation of myeloid cell surface marker expression which, in combination with morphological analysis, suggests that these cells show a regression in their differentiation state. However, additional cell surface markers would have to be analysed to further strengthen this hypothesis. Furthermore, subsequent studies in other t(8;21) cell lines would be necessary to definitively conclude how KD of C/EBPβ influences the cell’s development and differentiation status, upon the expression of RUNX1-ETO. Even though this was attempted in the t(8;21) cell line Kasumi-1, no meaningful results were obtained, due to technical difficulties (Supplementary Figure 46).

5.3.5.4 The impact of C/EBPβ knockdown on AML non-t(8;21) cell lines.

Previous results have suggested that low levels of C/EBPβ are associated with the development of t(8;21), and that this protein might interact with the RUNX1-ETO in the context of this disease (5.3.2). Having determined that, in the RUNX1-ETO-expressing cell line SKNO-1, KD of C/EBPβ promoted cell proliferation and was associated with an immature phenotype, this study additionally analysed the consequences of knocking-down C/EBPβ in the context on other AML subtypes, by applying the same KD vectors to HEL and U937 cells. In contrast to the effect on SKNO-1 cells, KD of C/EBPβ significantly reduced cell growth in both cell lines (Figure 5.31A; Figure 5.34A).
Figure 5.29 – The effect of C/EBPβ KD on cell surface marker expression in SKNO-1 cells

Expression of cell surface markers in SKNO-1 cells transduced with a control or shC/EBPβ constructs. An isotype control (IgG) was used to define negative marker expression (n≥3) Significant difference between shC/EBPβ cultures and control were analysed using one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; *** denotes p<0.001; **** denotes p<0.0001.
Figure 5.30 – Morphological effect of C/EBPβ KD in SKNO-1 cells

(A) Images showing the morphology of SKNO-1 cells transduced with a control or shC/EBPβ constructs (Table 2.1). Cells were stained with a combination of May-Grunwald and Gisma differential stains and scanned using a Zeiss Axioscan Z1 slide scanner, at 20x magnification (2.3.5). Scale bar indicates 50 µm. Bar charts showing (B) the area of SKNO-1 cells, as calculated using ImageJ; (C) FSC and (D) SSC of SKNO-1 cells transduced with a control or shC/EBPβ constructs. Data indicates mean ± 1SD (n=3). Significant difference between shC/EBPβ cultures and control were analysed using one-way ANOVA with Bonferroni’s multiple comparisons correction; ** denotes p<0.01; **** denotes p<0.0001.
These changes were further associated with an increase in the proportion of cells in the $G_1$ phase of the cell cycle, although these changes were deemed not significant (Figure 5.31B; Figure 5.34B). Moreover, both HEL and U937 cells presented a significant increase in the percentage of cells in early and late apoptotic stages, specifically for cultures transduced with shC/EBPβ #2 and #3 (Figure 5.31C; Figure 5.34C).

Subsequently, immunophenotypic analysis was performed to determine the effect of C/EBPβ KD on cell surface marker expression. The HEL cell line, derived from an erythroleukaemia patient, was found to be characterised as CD13$^+$CD33$^+$CD36$^+$ (Figure 5.32). KD populations were found to equally express the CD13 marker, however, shC/EBPβ #2 and #3 cultures showed a significant decreases in the expression of CD33 and CD36 and increased expression of CD41a. In the histiocytic lymphoma cell line U937, this study found that control cells were CD13$^+$CD45RA$^+$, with a heterogeneous expression of CD14, CD15, CD33, CD38 and CD64 (Figure 5.35). Upon C/EBPβ KD, these cells showed a significant decrease in the expression of CD14, CD45RA and CD38, with the simultaneous increase of CD15, CD36, CD64 and HLA-DR.

Subsequently, these cells were subjected to morphological analysis. An increase in cell size and complexity is usually observed as the differentiation process occurs. Both HEL and U937 cells showed an increase in cell size, further associated with a significant increase in the forward scatter) and side scatter properties of these cells (Figure 5.33; Figure 5.36). Additionally, KD of C/EBPβ in HEL cells, associated with shC/EBPβ #2 and #3, resulted in the development of multinucleated cells, suggestive of megakaryocytic differentiation, supported by increased expression of the CD41 marker. U937 cells, on the other hand, particularly cells transduced with shC/EBPβ #3, showed features associated with macrophage differentiation, but this was not supported by the immunophenotype of these cells.
Figure 5.31 – C/EBPβ KD significantly reduces cell growth in HEL cells, associated with changes in apoptotic frequency

Summary data showing (A) the growth of HEL cells with three shC/EBPβ constructs compared to control, over 5 days of culture. Data indicates mean ± 1SD (n=4); (B) the effect of C/EBPβ KD in normal cell cycle. Data indicates Mean ± 1SD (n=3); (C) the effect of C/EBPβ KD on apoptosis using Annexin V staining. Bars indicate percentage of pre-apoptotic cells (Annexin V+ / 7-AAD-) and late apoptotic cells (Annexin V+ / 7-AAD+). Data indicates Mean ± 1SD (n=3). Significant differences between shC/EBPβ cultures and control were analysed using one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01.; *** denotes p<0.001; **** denotes p<0.0001.
Figure 5.32 – The effect of C/EBPβ KD on cell surface marker expression in HEL cells

Expression of cell surface markers in HEL cells transduced with a control or C/EBPβ-KD constructs. An isotype control (IgG) was used to define negative marker expression (n≥3) Significant difference between shC/EBPβ cultures and control were analysed using one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; *** denotes p<0.001; **** denotes p<0.0001.
Figure 5.33 – Morphological effect of C/EBPβ KD in HEL cells

(A) Images showing the morphology of HEL cells transduced with a control or shC/EBPβ constructs (Table 2.1). Cells were stained with a combination of May-Grunwald-Gisma differential stains and scanned using a Zeiss Axioscan Z1 slide scanner, at 20x magnification (2.3.5). Scale bar indicates 50 µm. Bar charts showing (B) the area of HEL cells, as calculated using ImageJ; (C) FSC and (D) SSC of HEL cells transduced with a control or shC/EBPβ constructs. Data indicates mean ± 1SD (n=3). Significant difference between shC/EBPβ cultures and control were analysed using one-way ANOVA with Bonferroni’s multiple comparisons correction; ** denotes p<0.01; **** denotes p<0.0001.
Figure 5.34 – C/EBPβ KD significantly reduces cell growth in U937 cells, associated with changes in normal cell cycle and apoptotic rate

Summary data showing (A) the growth of U937 cells with three shC/EBPβ constructs compared to control, over 5 days of culture. Data indicates mean ± 1SD (n=4); (B) the effect of C/EBPβ KD in normal cell cycle. Data indicates Mean ± 1SD (n=3); (C) the effect of C/EBPβ KD on normal apoptotic rate using Annexin V staining. Bars indicate percentage of pre-apoptotic cells (Annexin V⁻/7-AAD⁻) and late apoptotic cells (Annexin V⁺ / 7-AAD⁺). Data indicates Mean ± 1SD (n=3). Significant differences between shC/EBPβ cultures and control were analysed using one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01.; *** denotes p<0.001; **** denotes p<0.0001.
**Figure 5.35 – The effect of C/EBPβ KD on cell surface marker expression in U937 cells**

Expression of cell surface markers in U937 cells transduced with a control or C/EBPβ-KD constructs. An isotype control (IgG) was used to define negative marker expression (n≥3) Significant difference between shC/EBPβ cultures and control were analysed using one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; *** denotes p<0.001; **** denotes p<0.0001.
Figure 5.36 – Morphological effect of C/EBPβ KD in U937 cells

(A) Images showing the morphology of U937 cells transduced with a control or shC/EBPβ constructs (Table 2.1). Cells were stained with a combination of May-Grunwald and Gisma differential stains and scanned using a Zeiss Axioscan Z1 slide scanner, at 20x magnification (2,3,5). Scale bar indicates 50 µm. Bar charts showing (B) the area of U937 cells, as calculated using ImageJ; (C) FSC and (D) SSC of U937 cells transduced with a control or shC/EBPβ constructs. Data indicates mean ± 1SD (n=3). Significant difference between shC/EBPβ cultures and control were analysed using one-way ANOVA with Bonferroni’s multiple comparisons correction; ** denotes p<0.01; **** denotes p<0.0001.
5.4 Discussion

C/EBPβ is a member of the C/EBP family of TF, and is involved in regulating several processes, such as cell growth, differentiation, inflammation, metabolism, survival and tumorigenesis (Sebastian and Johnson, 2006; Cao et al., 1991; Poli, 1998; Zahnnow, 2002; Diehl, 1998; Ramji and Foka, 2002). Previous studies have shown that, depending on cell context, the regulatory function of C/EBPβ is variable (Sebastian and Johnson, 2006). For instance, KO of C/EBPβ in mice resulted in a decrease in the number of B lymphocytes with reduced cell expansion properties (Regalo et al., 2016; Chen et al., 1997). Furthermore, C/EBPβ has been found to be essential for liver regeneration, as C/EBPβ KO mice displayed reduced hepatocytic proliferation, indicating that C/EBPβ is necessary for liver regeneration (Greenbaum et al., 1998). The same trend has been observed in several cancer models. In the BCR-ABL expressing cell line 32D, forced expression of C/EBPβ led to a decrease in cell proliferation and promoted differentiation (Guerzoni et al., 2006). Similarly, expression of C/EBPβ in the APL cell line NB4, upon ATRA treatment, was shown to result in decreased proliferative ability, associated with cell differentiation (Duprez et al., 2003). In glioblastoma and gastric cancer cells, on the other hand, decreased expression of C/EBPβ led to the inhibition of cell growth (Regalo et al., 2016; Aguilar-Morante et al., 2011).

In a haematopoietic setting, C/EBPβ is highly expressed in cells committed to the myelomonocytic lineage, including monocytes and macrophages (Williams et al., 1991; Katz et al., 1993; Haas et al., 2010; Gutsch et al., 2011). Several roles have been attributed to C/EBPβ in haematopoietic development. Depending on the system used, studies have shown C/EBPβ is involved in monocyctic differentiation (Friedman, 2007), eosinophil maturation (Nerlov et al., 1998) or granulocytic differentiation (Popernack et al., 2001; Iwama et al., 2002; Yusenko et al., 2021). In this study, RUNX1-ETO expression in CD34+ HSPC resulted in the significant downregulation of C/EBPβ mRNA and protein, as compared to control cells. The exact role of C/EBPβ in the development of t(8;21), however, has not been previously described. This chapter’s main aim was to determine the role of C/EBPβ in normal haematopoietic development, using a cord blood-derived model, and in leukaemogenic development, using AML cell lines. In summary, C/EBPβ overexpression promoted monocyctic and granulocytic differentiation in HSPC. However, no significant impact of C/EBPβ KD was observed. This latter observation could be explained by the lack of complete KD/KO of C/EBPβ, or due to redundancy in the context of other members of the C/EBP family, further
discussed in 5.4.3. In AML cell lines, the effect of C/EBPβ KD was found to be highly context-dependent: in the t(8;21) line, SKNO-1, decreased C/EBPβ levels resulted in an increased proliferative ability and reduced myeloid marker expression. In non-t(8;21) cells, however, KD of C/EBPβ reduced cell growth and induced more variable effects on the expression of markers of differentiation.

5.4.1 CEBPB expression increases during haematopoietic development

To further enlighten the role of C/EBPβ on myeloid development, specifically in the monocytic and granulocytic maturation process, CEBPB expression was determined using publicly available transcriptomic datasets in human haematopoietic progenitor cells. This analysis showed that CEBPB expression increased specifically in fully mature granulocytes and monocytes, as compared to undifferentiated HSCs, suggesting that CEBPB is involved in the development of these lineages. These observations were further corroborated with the analysis of the CEBPB expression in the mouse and agree with Tamura et al., who showed that CEBPB was upregulated in mouse mature monocytes (Tamura et al., 2017). Additionally, the authors further demonstrated that C/EBPβ is involved in the survival of non-classic (Ly6C−) monocytes as well as being essential for monocytic development (Tamura et al., 2017). However, since a single mRNA transcript generates three C/EBPβ protein isoforms, LAP*, LAP and LIP (1.5.1) transcriptional analysis cannot discriminate between the roles of these isoforms, whose activity may also be influenced by post-translational modifications, including phosphorylation, acetylation, methylation and sumoylation (Ramji and Foka, 2002; Zahnow, 2009; Tsukada et al., 2011; Hattori et al., 2003). Further, Pham et al. reported that whilst CEBPB mRNA was unchanged during differentiation process, nuclear levels of C/EBPβ increased during the maturation process of monocytes into macrophages (Pham et al., 2007).

Since transcriptomic analysis alone may not be an accurate guide to C/EBPβ protein expression, this study determined the expression of C/EBPβ protein in HSPC by western blotting analysis of cells cultured in conditions that favoured monopoiesis and granulopoiesis. In this context, C/EBPβ LAP* was shown to decrease over time. Expression of LAP was only detectable on day 3, whilst LIP was undetected throughout. These observations directly contrast with the mRNA analysis described above. Analysis of purified lineage subsets showed that C/EBPβ LAP* was highest in monocytic-progenitor cells, followed by erythroid-committed cells and granulocytic cells; however, these observations are made early in the myeloid development and later time-points would be necessary to determine the pattern.
C/EBPβ protein expression throughout the development of these lineages. Unfortunately, there was no time to perform these experiments.

C/EBPβ LAP* and LAP have previously been shown to be involved in the differentiation process of monocytes (Friedman, 2007), and upregulation of these two C/EBPβ isoforms is necessary for this process to occur (Katz et al., 1993; Gutsch et al., 2011; Pham et al., 2007; Natsuka et al., 1992; Pan et al., 1999; Studzinski et al., 2005; Zhang et al., 2011a). C/EBPβ expression has also been shown to be essential for granulopoiesis under stress conditions (Hirai et al., 2006; Popernack et al., 2001) with steady-state granulocytic development being regulated by another member of the C/EBP family, C/EBPa (Suh et al., 2006; Radomska et al., 1998).

5.4.2 Increased CEBPB expression is associated with poor prognosis in AML

Expression of CEBPB was found to be elevated across AML subtypes, but lower in t(8;21). Furthermore, increased CEBPB expression was associated with poor prognosis; however, since multivariate analysis was not conducted, its status as a prognostic factor was not established in this study. Similar observations had been previously made in solid tumours. High CEBPB expression in breast cancer was associated with the most aggressive tumour types, including metastatic breast cancer (van de Vijver et al., 2002), tumours associated with a high tumour grade (van ’t Veer et al., 2002; Ma et al., 2004; Finak et al., 2008) and an overall poorer prognosis (van de Vijver et al., 2002). Similarly, in human gastric carcinoma, C/EBPβ protein expression was found to correlate with poorer prognosis, associated with lower median survival, and the development of metastasis (Du et al., 2013). However, these observations need to be considered carefully, as C/EBPβ expression is also regulated through post-translational mechanisms (5.4.1). In fact, Kurzejamska et al. reported that loss of C/EBPβ protein promoted metastatic development of mouse mammary tumour cells (Kurzejamska et al., 2014) and in a cohort of 137 breast cancer patients decreased C/EBPβ expression correlated with poorer OS and development of metastasis within the lymph nodes (Kurzejamska et al., 2014).

5.4.3 Knockdown of C/EBPβ in HSPC does not influence normal myeloid development

To establish whether the reduced C/EBPβ expression seen in t(8;21) patients could contribute to the pathogenesis of AML, functional studies were performed using KD of
C/EBPβ in human cord blood-derived HSPC. Under clonal conditions, KD of C/EBPβ resulted in a significant decrease in myeloid colony forming efficiency and self-renewal potential. In bulk liquid culture, KD of C/EBPβ repressed HSPC monocytic and granulocytic proliferation. In BM-derived progenitor cells from mice, absence of C/EBPβ also resulted in a significant reduction in the percentage of colonies formed, as well as in the number of generated cells, in accordance with this study’s results.Interestingly, however, KO of C/EBPβ in more differentiated cells within the same model led to an increase in cell proliferation (Gutsch et al., 2011; Screpanti et al., 1995). It is plausible that the effects observed would have been different if C/EBPβ had been knocked-down in more differentiated cells, as opposed to HSCs.

Developmentally, KD of C/EBPβ in HSPC did not significantly alter the monocytic differentiation markers CD11b and CD14. Moreover, this study found that C/EBPβ was not required for the survival of mature monocytes, as suggested by Tamura et al. (Tamura et al., 2017). However, Tamura et al. applied a mouse C/EBPβ−/− model of peripheral blood monocyte survival, in contrast with the in vitro HSPC model used in this study. In granulocytic development, KD of C/EBPβ promoted CD15 expression, but did not affect the expression of other developmental markers in the panel. The study of additional granulocytic markers, such as CD16 or CD32 (Attar, 2014), would be necessary to unequivocally determine the exact role of C/EBPβ in granulocytic development. Together, this data suggests that C/EBPβ is mainly involved in monocytic development; however, it is not essential for this process to occur.

Reduced growth can be a result of perturbed cell cycle progression or increased apoptotic events. KD of C/EBPβ in HSPC was found to result in the accumulation of cells within the G1 phase of the cell cycle. Opposingly to this study, Gutsch et al. showed that, in macrophage-like cells, KO of C/EBPβ resulted in the increase of cells displaying S and G2/M phase markers (Gutsch et al., 2011), associated with an increase in proliferation. In human non-small cell lung cancer tissues, KD of C/EBPβ was associated with an increase in the proportion of cells arrested at the G2/M phase of the cell cycle (Lee et al., 2019); however, in this study, the delay in cell cycle progression was further associated with the inhibition of cell proliferation.

Overall, KD of C/EBPβ had a modest impact in the HSCP model used in this study. Several factors can explain these results. Firstly, levels of C/EBPβ were knocked-down by c50%, and it is possible that these levels are enough to support haematopoietic development. Moreover, other C/EBP isoforms might be acting compensatively in the events of C/EBPβ KD. In the haematopoietic system, C/EBPα, C/EBPβ, C/EBPδ, and C/EBPε are all expressed within the
myeloid lineage (Tolomeo and Grimaudo, 2020). Moreover, the overlapping expression of the C/EBP family members suggests that myeloid development is regulated by different combinations of C/EBP homo- and heterodimers (Tolomeo and Grimaudo, 2020).

Several studies have demonstrated that C/EBPβ is only necessary for granulocytic development in stress-derived conditions (Hirai et al., 2006; Hayashi et al., 2013), but it is necessary for the normal function and differentiation of myelomonocytic cells and macrophages (Pham et al., 2007; Cain et al., 2013). KD studies, however, did not demonstrate this, as the differentiation of both granulocytes and monocytes was not impaired upon KD of this TF. These observations have been summarised in Table 5.1.

5.4.4 C/EBPβ overexpression in HSPC promotes myeloid cell growth and development

Next, this study performed functional studies by overexpressing the longer C/EBPβ isoform LAP*, within the same HSPC model. Under clonal conditions, C/EBPβ significantly promoted myeloid colony formation. A single round of colony replating suggested that this may have, at least in part, be driven by promotion of self-renewal. Overexpression of C/EBPβ also resulted in an increase in both monocytic and granulocytic growth in bulk liquid culture with an increase in the proportion of granulocytic cells predominating over that of monocytes.

Phenotypically, overexpression of C/EBPβ upregulated both the CD11b and CD14 markers on monocytic cells. This agrees with previous studies, which showed that C/EBPβ controls CD14 expression (Pan et al., 1999; Ji and Studzinski, 2004; Zhang et al., 2011a; Xu et al., 2008) and that C/EBPβ promotes monocytic development (Pham et al., 2007; Zhang et al., 2011b; Tamura et al., 2017).

Analysis of the granulocytic population showed that overexpression of C/EBPβ resulted in the upregulation of the cell surface marker CD15 during the differentiation process, suggesting promotion of granulocytic development. Assessment of cell morphology also revealed that overexpression of C/EBPβ promotes granulocytic differentiation, with an increase in the development of band cells (mature granulocytes). These observations have been summarised in Table 5.1.
Table 5.1 – Summary of the main findings regarding the role of C/EBPβ in normal human haematopoiesis

Table summary detailing the consequences of C/EBPβ modulation vs control.
* denotes significant differences in C/EBPβ overexpression / knockdown, as compared to normal control cells.

<table>
<thead>
<tr>
<th></th>
<th>C/EBPβ Overexpression</th>
<th>C/EBPβ Knockdown</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Monocytes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Growth</strong></td>
<td>Decrease</td>
<td>No change</td>
</tr>
<tr>
<td><strong>Markers</strong></td>
<td>Increase in CB11b and CD14 *</td>
<td>No change</td>
</tr>
<tr>
<td><strong>Granulocytes</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Growth</strong></td>
<td>Decrease *</td>
<td>Decrease *</td>
</tr>
<tr>
<td><strong>Markers</strong></td>
<td>Increase in CD15 *</td>
<td>Increase in CD15 *</td>
</tr>
<tr>
<td><strong>Apoptosis</strong></td>
<td>No change</td>
<td>No change</td>
</tr>
<tr>
<td><strong>Cell Cycle</strong></td>
<td>Decrease in G1 *</td>
<td>No change</td>
</tr>
<tr>
<td><strong>CFU</strong></td>
<td>Increase *</td>
<td>Decrease *</td>
</tr>
</tbody>
</table>
5.4.5 The consequences of C/EBPβ KD in AML cell lines is context-dependant

The final part of this chapter consisted of analysing the consequences of C/EBPβ expression in AML by modulating its expression in a cohort of AML cell lines, to determine whether it influenced leukaemogenic development. KD of C/EBPβ was found to significantly promote the fold expansion on the (8;21) cell line SKNO-1 accompanied by a decrease in the proportion of cells in the G1 phase of the cell cycle and decrease in normal apoptotic frequency, all consistent with a pro-proliferative AML phenotype. Analysis of cell surface markers showed ablation of myeloid marker expression upon KD of C/EBPβ indicating that the developmental status of these cells was also affected. Interestingly, contrary results were observed in cell lines that do not express the fusion protein RUNX1-ETO. In these cells (HEL and U937), KD of C/EBPβ was found to suppress cell growth, associated with an increase in the proportion of cells in the G1 phase of the cycle and increased apoptotic frequency. Immunophenotypic marker expression was again perturbed in these lines but without any clear pattern emerging.

Previous studies have showed that forced expression of C/EBPβ promotes monocytic development of AML cell lines. For instance, addition of 1,25D to HL-60 cells resulted in forced monocytic differentiation, associated with a significant increase in the expression of C/EBPβ LAP and LIP isoforms (Marcinkowska et al., 2006). The fact that the dominant-negative LIP isoform was found to be increased in parallel with LAP makes the authors suggest that this isoform in responsible for repressing genes unnecessary for the differentiation process, including those involved in sustained proliferation. In K562 cells, a CML cell line, expression of C/EBPβ was shown to induce granulocytic development, with changes to cell morphology and promotion of gene expression associated with this cell type (Duprez et al., 2003). In these cells, the process was initiated by C/EBPβ, which led to the expression of another member of the C/EBP family, C/EBPε, known to play an important role in granulocytic differentiation (Park et al., 1999). Additionally, in APL cells, increased expression of C/EBPβ was shown to be necessary in ATRA-induced cell differentiation, and its inhibition resulted in a decreased response to ATRA (Duprez et al., 2003). Similar observations were made by Yusenko et al., who reported that ectopic expression of C/EBPβ in the APL cell line NB4 led to a decrease in cell proliferation (Yusenko et al., 2021). Therefore, this study hypothesised that KD of C/EBPβ in AML cell lines would promote cell proliferation associated with the expression of cell surface markers associated with immature cells. Ours results, however, conflict with the above-described studies, where overexpression, opposed to KD, of C/EBPβ resulted in cell...
differentiation and decreased proliferative ability. However, a few studies have been published that describe the downregulation of C/EBPβ as a potential therapeutic target. With this in line, it has been shown that C/EBPβ cooperates with MYB and the co-activator p300 to promote the expression of myeloid associated genes (Burk et al., 1993; Ness et al., 1993; Mink et al., 1996; Yusenko et al., 2021), and that MYB is highly expressed in AML cells, making it an interesting therapeutic target (Hess et al., 2006; Somervaille et al., 2009; Jin et al., 2010; Zuber et al., 2011). In a recent study, Yusenko et al. showed that inhibiting the binding of MYB to C/EBPβ resulted in the upregulation of differentiation associated genes and increased apoptosis in AML cell lines (Yusenko et al., 2021), similarly to this study’s results. Equally, expression of C/EBPβ in multiple myeloma cell lines and patient samples was shown to be increased as compared to normal B cells, and silencing it resulted in the downregulation of several TF, including those involved in the proliferative process, such as IRF4, XBP1 and BLIMP1, and apoptotic signalling, including BCL2 (Pal et al., 2009).

Additionally, this study attempted to generate stable C/EBPβ LAP* overexpression cell lines. However, this was not successful due to rapid loss of transduced cells, as determined by GFP positivity; this suggests that overexpression of C/EBPβ LAP* is incompatible with AML cell lines.

In conclusion, the data presented in this chapter shows that KD of C/EBPβ in HSPC did not result in an arrest in haematopoietic development. However, a limitation of the application of shRNA is the inability to completely remove C/EBPβ expression; it is possible that, despite the reduced levels, a far greater level of KD would be necessary to demonstrate any dependency. Additionally, compensation from alternate C/EBPβ family members could be acting as a mechanism of redundancy. Whilst the former could be further explored with CRISPR-Cas9, which would result in complete KO, the latter concern would require a complex study design to fully consider all the C/EBP family members. Furthermore, this study demonstrated that KD of C/EBPβ is context-dependant with regards to the differentiation status of the AML cell lines. These findings indicate that t(8;21) might be an important context, but attempts to reproduce this in another t(8;21) line failed, making it hard to draw a meaningful conclusion (Supplementary Figure 46). In contrast, this study provides some evidence that the overexpression seen in AML might have a role in developmental inhibition.
Chapter 6

Conclusion and Future Directions
6.1 Conclusions

AML is a disease characterised by the clonal expansion of primitive HSCs which are incapable of terminal differentiation (Döhner et al., 2015). Furthermore, AML is a highly heterogenous disease, arising from the development of several molecular abnormalities, including mutations and chromosomal rearrangements. A common abnormality is the t(8;21), occurring in approximately 12% of all AML cases, leading to the expression of the fusion protein RUNX1-ETO. Despite the fact that t(8;21) AML is associated with a good prognosis, with the majority of patients achieving CR following standard chemotherapy treatment, a significant proportion of these are deemed not suitable to receive intensive chemotherapy, especially due to their age. Consequently, half of the patients will relapse within 2.5 years of achieving CR (Marcucci et al., 2005; Hospital et al., 2014). The development of new targeted therapies, such as ATRA/ATO applied in APL (reviewed in (Sanz et al., 2019)), or Venetoclax, targeting BCL-2 (DiNardo et al., 2019), has significantly improved CR rates and OS rates of AML patients. Traditional treatment regimens for AML t(8;21) patients, however, have remained unchanged for several years. Despite the fact that these patients would potentially be eligible for targeted therapies, where appropriate, there is a pressing need to develop new therapies and improve the remission rate of t(8;21) patients.

Previous studies have demonstrated that RUNX1-ETO is able to inhibit normal myeloid differentiation in mice and patient samples (Okuda et al., 1998; Pabst et al., 2001; Nafria et al., 2020). Furthermore, ectopic expression of RUNX1-ETO in human HSPCs was shown to block myeloid development and promote their self-renewal, associated with changes in gene expression patterns (Mulloy et al., 2002; Tonks et al., 2007; Tonks et al., 2003; Tonks et al., 2004). Despite the progress in understanding the biological implications of t(8;21), the mechanism through which this fusion protein leads to the development of AML is not yet fully understood. Moreover, a better understanding of the mechanism behind this process would allow the identification of novel therapeutic targets for the treatment of this disease.

The basis for this study’s original hypothesis relied on the identification of TF, both at the mRNA and at the protein level, that could be responsible for the developmental disruption observed in t(8;21) patients, upon the expression of RUNX1-ETO. The original goal to apply proteomic analysis techniques to investigate targets of interest was achieved, as it allowed the identification of several targets known to be involved in the development of AML, including
PU.1 and CBFβ, with numerous other novel targets being equally detected. Analysis of transcriptomic data yielded similar results, with the identification of well-established TF, thus validating the approach and model used in this study. Ultimately, it was possible to identify two targets, ZNF217 and C/EBPβ, derived from the transcriptomic and proteomic analysis, respectively, as targets on interest in the context of t(8;21).

6.1.1 RUNX1-ETO-induced transcriptomic and proteomic changes in HSPC

The initial aim of this study was to re-analyse a previously generated microarray dataset, following the ectopic expression of RUNX1-ETO in human HSPCs (Tonks et al., 2007). The previous approach used an unsupervised analysis and identified changes in all classes for genes, including membrane bound proteins (CD200) and signalling molecules (γ-Catenin) (Tonks et al., 2007). I performed an ‘Enrichment Analysis’ using Metacore™ and showed that RUNX1-ETO overexpression significantly impaired megakaryopoiesis and granulocytic development, consistent with previously reported studies (Niebuhr et al., 2008; Tonks et al., 2004). Since developmental alterations are primarily a result of changes in the transcriptional process, I focused on TF dysregulation, to identify regulators of transcription that had been altered upon the expression of RUNX1-ETO. This analysis identified not only TF known to be involved in the development of AML, such as PU.1 (Nerlov and Graf, 1998; Koschmieder et al., 2005) and CEBPA (Koschmieder et al., 2005), but also TF not yet extensively described in either haematopoietic development, or AML. This included the TF ZNF217, overexpressed in RUNX1-ETO-expressing cells, and studied in Chapter 4.

Whilst most studies have focused on RUNX1-ETO-mediated transcriptional changes (Nafria et al., 2020; Ptasinska et al., 2019; Martinez-Soria et al., 2018; Ptasinska et al., 2012; Tonks et al., 2007), there is a paucity regarding analysis of the cells’ proteomic profile. Additionally, whilst transcriptomic analysis can infer corresponding protein expression, there are several factors that influence its corresponding protein expression including translational or post-translational modifications, epigenetic regulators or protein half-life. For this reason, analysis of the cells proteomic profile represents a more direct insight into the changes cells undergo under certain conditions. To address this, human HSPCs were transduced with a control or RUNX1-ETO-expressing vector, and proteins fractionated according to nuclear or cytosolic localisation (Figure 6.1A). From this analysis, this study identified 256 significantly dysregulated proteins, involved in several pathways previously identified in AML, including
dysregulation of normal myeloid differentiation (Nafria et al., 2020; Ichikawa et al., 2013) and changes in the NF-kB pathway (Zhou et al., 2015). Similarly to the transcriptomic analysis described above, I determined the candidate regulators of the observed phenotype, focusing on TF. I identified, a member of the C/EBP family of TF repressed in RUNX1-ETO cells and predicted to have the greatest influence on other dysregulated proteins. C/EBPβ was studied in Chapter 5.

6.1.2 ZNF217 does not contribute to the pathogenesis of AML

Following the identification of ZNF217 overexpression (Chapter 3), this study focused on examining its role in normal haematopoietic development. Given the role of ZNF217 in the development of several human cancers (1.4.2), it was hypothesised that it would act as an oncogene in haematopoietic cells by disrupting normal myeloid growth and development.

ZNF217 mRNA expression was shown to be highly variable across AML subtypes, but this was lower as compared to mature cells, suggesting a potential developmental role. Moreover, higher levels of ZNF217 were associated with reduced disease-free survival, indicating a possible role for ZNF217 in promoting relapse; however, no multivariate analysis was performed to conclusively determine this. Indeed, increased ZNF217 expression has previously been associated with worse patient outcome and prognosis in several solid tumours, including prostate (Szczyrba et al., 2013), colorectal (Zhang et al., 2015; Li et al., 2015), breast (Krig et al., 2010; Littlpage et al., 2012; Nguyen et al., 2014; Nonet et al., 2001; Plevova et al., 2010; Thollet et al., 2010; Vendrell et al., 2012) and ovarian (Zhu et al., 2009), amongst others, highlighting the prognostic value of ZNF217 in malignant disease. However, even though the role of ZNF217 has been extensively studied in solid cancers, its role in AML development has not been studied.

In normal haematopoiesis, ZNF217 mRNA levels increased throughout myeloid differentiation, with highest expression in mature granulocytes, suggesting a role in granulocytic development. Initially, this study sought to determine the role of ZNF217 on normal myeloid development using a human primary HSPC model. These studies showed that overexpression of ZNF217 significantly impaired myeloid colony formation, coupled with a decreased self-potential capacity. In bulk liquid culture, ZNF217 expression promoted monocytic differentiation, associated with an increased expression of the differentiation markers CD11b and CD14. These results indicate that, in contrast to its role in solid tumours,
it is unlikely that ZNF217 possesses a role in leukaemogenesis, which is characterised by a block in myeloid development. Nevertheless, my data suggests that ZNF217 might have a developmental role in normal haematopoiesis. To address this, KD studies were performed. Although KD of ZNF217 significantly impaired HSPC myeloid colony formation and self-renewal potential, no effect was observed during myeloid growth and differentiation of these cells. This suggests that, although ZNF217 can promote myeloid differentiation, this is not essential for the process to occur (Figure 6.1B).

6.1.3 KD of C/EBPβ does not contribute to the pathogenesis of AML

Chapter 5 focused on analysing the importance of C/EBPβ expression and dysregulation during normal myeloid development. In Chapter 3, analysis of proteomic data revealed that C/EBPβ is repressed upon the expression of RUNX1-ETO, in human HSPCs. Even though some studies have been performed in an attempt to determine its the role in normal human haematopoiesis, it’s role in the development of AML t(8;21) has not been described to date.

This study first determined the association between CEBPB mRNA expression and the pathogenesis of AML. This analysis showed that CEBPB expression is highly variable in different AML subtypes, however, interestingly, decreased expression of CEBPB was consistently observed in t(8;21) patients within the FAB M2 AML subtype, as compared to non-t(8;21) patients, suggesting an association between RUNX1-ETO and CEBPB in the leukaemogenic process. However, mRNA studies related to CEBPB expression need to be interpreted with caution. CEBPB mRNA results in the synthesis of three C/EBPβ isoforms: LAP*, LAP and LIP due to the existence of alternative start codons (Zahnow, 2009; Sears and Sealy, 1994). For this reason, it is impossible to infer expression of each isoform in AML patients and in normal haematopoiesis from mRNA analysis alone.

Analysis of CEBPB mRNA expression in normal human haematopoiesis revealed that it is expressed across all haematopoietic lineages, with higher expression levels in mature monocytes and granulocytes. Analysis of C/EBPβ protein that LAP* is highly expressed in monocytic progenitor cells, however expression was relatively low in granulocyte progenitors. This analysis did not detect either of the remaining two isoforms LAP and LIP. However, this analysis was performed on immature cells, and additional studies on more differentiated sub-populations would be necessary to give a more complete understanding of C/EBPβ protein expression during development. Nevertheless, these observations suggested a role for C/EBPβ
in myeloid development, particularly in the monocytic population. In agreement with these observations, upregulation of C/EBPβ has been observed in myelomonocytic cells and macrophages (Williams et al., 1991; Katz et al., 1993; Haas et al., 2010; Gutsch et al., 2011), where it influences cell proliferation (Friedman, 2007). Even though this study was only able to detect low levels of C/EBPβ in granulocyte-progenitor cells, previous studies have shown that, whilst C/EBPα, another member of the C/EBP family, is important for normal granulopoiesis (Hirai et al., 2006; Zhang et al., 1997; Zhang et al., 2004b; Radomska et al., 1998), C/EBPβ is essential for stress-induced granulopoiesis (Akagi et al., 2008; Zhang et al., 2010; Hall et al., 2012). In the present study, KD of C/EBPβ in human HSPC failed to significantly impact myeloid cell growth and development. This observation is likely due to low levels of KD achieved in these cells. Additionally, functional redundancy with other members of the C/EBP family can further account for these results, as previous studies have shown that these TF can act compensatively in the absence of a certain member (Tanaka et al., 1995a; Screpanti et al., 1995; Hu et al., 1998). In line with this, studies have suggested that C/EBPα and C/EBPβ might collaborate with each other, in an attempt to ensure a proper supply of granulocytes at all times, through the regulation of common target genes linked to granulocytic differentiation (Jones et al., 2002). Moreover, whilst some authors have claimed that KO of C/EBPβ in early progenitor cells give rise to a lower number of colonies and cells (Hirai et al., 2006), in more differentiated cells, absence of C/EBPβ results in a pro-proliferative phenotype (Gutsch et al., 2011; Screpanti et al., 1995) (Figure 6.2).

Next, I determined the consequences of C/EBPβ LAP* overexpression as a single abnormality on the myeloid development on human HSPCs. Overexpression of this isoform increased progenitor cell colony forming ability, as well as self-renewal potential observations consistent with an inhibition of differentiation. However, analysis of granulocytic and monocytic cell surface markers indicated C/EBPβ LAP* overexpression promoted the differentiation of these cells. Is it plausible that the increase in self-renewal potential is, in fact, a result of arrested terminal differentiation. Of note, increased expression of C/EBPβ in both APL and CML cells following treatment with ATRA or Imatinib, respectively, was shown to induce differentiation of AML cell lines (Duprez, 2004; Duprez et al., 2003; Guerzoni et al., 2006), whilst its KD in APL cells reduced the differentiation potential of these cells. These observations suggest that overexpression of the LAP* isoform promotes haematopoietic development in HSPC.
As KD of C/EBPβ in HSPC is unlikely to contribute to the pathogenesis of AML through the disruption of normal myeloid development, this study assessed the consequences of knocking down C/EBPβ in AML cell lines. Interestingly, KD of C/EBPβ yielded distinct results depending on the context of the transformed cell, likely a result of the cells’ developmental stage. In the t(8;21) cell line SKNO-1, KD of C/EBPβ resulted in an increase in the cells’ proliferative ability, combined with the ablation of myeloid cell surface marker expression. However, similar experiments performed in HEL and U937 demonstrated a growth inhibition associated with the increased expression of markers of differentiation. Interestingly, a recent study has reported that inhibition of C/EBPβ activity actually promoted the expression of differentiation-associated cell surface markers and genes (Yusenko et al., 2021) (Figure 6.2).

In conclusion, this study has shown that C/EBPβ downregulation is associated with AML t(8;21). However, KD of C/EBPβ in HSPCs did not support an oncogenic role for this TF in normal haematopoiesis, most likely due to insufficient level of KD, suggesting that even low levels of C/EBPβ are enough for myeloid development to occur. Moreover, the absence of a differentiation block in these cells might be a result of the interplay between C/EBPβ and other members of the C/EBP family (discussed in ).

6.1.4 Limitations of the study

The novel targets identified by applying the latest proteomic analysis platform known as SWATH-MS represent the advantage of leveraging a next-generational technique; however, early adoption requires certain concessions. Whilst a great advantage of SWATH-MS is that expression results are direct measures of protein expression, rather than mRNA level, which can be disassociated from actual protein expression due to post-translational regulation, this requires substantially greater cell number for sample preparation, limiting the analysis of rare cell subsets. Further, SWATH-MS relies upon the construction of a library for protein identification. This is often generated by each respective research site, as no standard has been adopted widely, increasing the input requirements, limiting the number of proteins considered, and potential affecting reproducibility. In this study, such requirements precluded the ability to FACS a population as insufficient material would have been recovered. Lastly, integration of proteomic and transcriptomic data at specified timepoints is complicated by the time delay between mRNA expression and protein generation.
The CD34 model used within this study was further limited by access to cord-blood material, restricting the number of CD34+ HSPC as input for experiments. Due to this, different protocols had to be implemented, depending on the experiment outcome necessary. If the majority of the population had been successfully transduced, these were deemed suitable for analysis albeit with the acknowledgement that untransduced cells were perhaps influencing those with overexpression or knockdown, potentially masking a phenotype. For certain plasmids, this proved a greater challenge as the substantial plasmid size heightened the difficult of transducing CD34, as seen in overexpression studies. The difficulty was exacerbated in preliminary investigation of double infections, which were deemed infeasible as the proportion of transduced cells harbouring both constructs were too few for subsequent studies.

A subsequent limitation of the current study relies on the fact that certain proteins of interest were not investigated in all of the desired settings, or in multiple time-points. For instance, ZNF217 was not examined in the context of normal haematopoiesis which was due to the scarcity of the material. For the same reason, it was not possible to analyse each target (ZNF217 and C/EBPβ) expression throughout the 13-day experiments performed (OE and KD). This was a consequence of the COVID-19 pandemic which prevented cord-blood donations.

A limitation of the cell line approach is the inherent separation from the clinical context of primary material. However, access to such material was not possible and cell lines represented the next best model. Unfortunately, limited cell lines harbour the t(8;21) translocation limiting the sample sizes within this study. Whilst Kasumi-1 is another t(8;21) line which could have been used, this line is very slow growing and very resistant to transduction, limiting the usefulness to this study.

Cell lines were investigated in the context of knockdown by shRNA. A more conclusive endpoint could have been reached by application of CRISPR-Cas9 which is able to generate complete knockout of target genes. However, CRISPR-Cas9 requires tremendous cost both monetarily and in labour to generate a complete knockout clone. As this study wished to examine numerous targets in multiple contexts, shRNA provided the best means to achieve this.

The colony formation assay presented herein is an in-house solution to overcome the poor reproducibility, cost, and access to material associated with Matrigel assays. However, colonies were not examined for their specific subtypes – erythroid, myeloid, granulocytic, macrophage
– as these did not represent an endpoint of interest; rather, colony formation ability was focused on the capacity of self-renewal at the second plating round.

Together, these data represent the first in-depth examination of two novel targets of interest in RUNX1-ETO AML. Whilst sample sizes and access to material limited the study, these data were generated in an attempt to provide a broad understanding of the target mechanisms, a goal which was achieved.

6.2 Future directions

Even though the current study provide insights into the role of both ZNF217 and C/EBPβ in normal haematopoietic development, additional studies would be necessary to support the role of these TF in t(8:21) AML development. Expression of ZNF217 should be knocked-down/out in HSPCs expressing RUNX1-ETO, whilst C/EBPβ should be overexpressed in the same model, to determine if either of these factors can reverse RUNX1-ETO-induced phenotype. Additionally, the fact that ZNF217, as a single abnormality, was shown to promote myeloid development, suggests that RUNX1-ETO might be inhibiting its normal function. For this reason, ChIP-Seq analysis of RUNX-ETO cells, co-transduced with a ZNF217-overexpressing vector should be performed, to determine chromatin co-occupancy of both RUNX1-ETO and ZNF217.

Regarding C/EBPβ studies, this study hypothesised that the failure to induce a significant block in myeloid differentiation was a result of two separate factors, and additional experiments would be necessary to test them. On one side, it was not possible to achieve efficient KD. To address this, CRISPR-Cas9 experiments could be applied in the same HSPC model, to try to achieve complete KO. Furthermore, considering the functional redundancy within the C/EBP family, KD, or KO, of other C/EBP proteins, such as C/EBPα or C/EBPε, in combination with C/EBPβ should be employed to determine the extent of the interactions between all the family members in haematopoietic development.

Furthermore, studies in additional models would allow the complete representation of both ZNF217 and C/EBPβ in other contexts. These would include in vivo murine studies, to match the in vitro experiments in HSPCs; patient-derived primary material, particularly those with t(8:21), with the intent to replicate the AML cell lines result; and, lastly, induced pluripotent stem cells (iPSCs), to determine the role of these TF in other contexts, for instance, in fully
differentiated monocytes or granulocytes. Additionally, it would be beneficial to target ZNF217 and C/EBPβ with clinically transferable inhibitors and activators, respectively, in the context of AML t(8;21). Unfortunately, no agents currently exist in case of C/EBPβ and the only inhibitor of ZNF217 has been abandoned (Collins et al., 2007). Furthermore, other targets could be investigated, such as ARID5B or IRF9, described in Chapter 3, to determine if these would contribute to the leukaemogenic development.
(A) Human CD34⁺ HSPC were isolated from neonatal cord blood (Stage 1) and transduced with retrovirus co-expressing either RUNX1-ETO and GFP (RE), or GFP alone (CT). Transduced cells were fractionated into cytosolic and nuclear protein fractions, and analysed by performing SWATH-MS to determine protein expression. Differentially expressed proteins were uploaded into Metacore for pathway analysis and target identification (Stage 2). (B) To determine the role of ZNF217 in normal human haematopoiesis, CD34⁺ HSPC were isolated as described in (A). Following this, functional studies using overexpression and KD constructs were applied to study the cells myeloid development.
Chapter 6

Figure 6.2 – Summary of the main findings described on Chapter 5

To determine the role of C/EBPβ in normal human haematopoiesis, CD34+ HSPC were isolated as described in Figure 6.1A. Following this, functional studies using overexpression and KD constructs were applied to study the cells myeloid development. Additionally, to determine the role of C/EBPβ in leukaemogenic development, the same KD constructs were applied to AML cell lines, with different characteristics and stages of differentiation: SKNO-1 (M2), HEL (M6) and U937 (M5).
**Supplementary Tables**

**Supplementary Table 1 – Compensation matrix**

Compensation was performed computationally post-acquisition according to the following matrix:

<table>
<thead>
<tr>
<th></th>
<th>FITC</th>
<th>PE</th>
<th>PERCP</th>
<th>APC</th>
</tr>
</thead>
<tbody>
<tr>
<td>FITC</td>
<td>1</td>
<td>-0.11</td>
<td>0.011</td>
<td>0</td>
</tr>
<tr>
<td>PE</td>
<td>-0.024</td>
<td>1</td>
<td>-0.22</td>
<td>0</td>
</tr>
<tr>
<td>PERCP</td>
<td>-0.002</td>
<td>-0.0001</td>
<td>1</td>
<td>-0.15</td>
</tr>
<tr>
<td>APC</td>
<td>0</td>
<td>0</td>
<td>-0.06</td>
<td>1</td>
</tr>
</tbody>
</table>
Supplementary Table 2 – Identification of significantly dysregulated proteins, as a result of RUNX1-ETO expression

SWATH-MS identified 256 significantly dysregulated proteins in HSPC, as a result of increased RUNX1-ETO expression.

Fold-change (FC) values represent the regulation of each protein when compared to control cells (green – upregulated; red – repressed, when compared to control cells). Statistical analysis was performed using one-way ANOVA, without p-value correction (n=3).

<table>
<thead>
<tr>
<th>Protein</th>
<th>Localisation</th>
<th>p-value</th>
<th>FC</th>
</tr>
</thead>
<tbody>
<tr>
<td>CA2</td>
<td>Cytoplasm</td>
<td>0.0053</td>
<td>6.19</td>
</tr>
<tr>
<td>GBF1</td>
<td>Cytoplasm</td>
<td>0.0237</td>
<td>4.03</td>
</tr>
<tr>
<td>MAGED2</td>
<td>Cytoplasm</td>
<td>0.0016</td>
<td>3.47</td>
</tr>
<tr>
<td>COPZ1</td>
<td>Nucleus</td>
<td>0.0022</td>
<td>3.38</td>
</tr>
<tr>
<td>ALDH1L1</td>
<td>Cytoplasm</td>
<td>0.0023</td>
<td>3.30</td>
</tr>
<tr>
<td>RNH1</td>
<td>Nucleus</td>
<td>0.0462</td>
<td>3.24</td>
</tr>
<tr>
<td>NFKB2</td>
<td>Cytoplasm</td>
<td>0.0003</td>
<td>3.17</td>
</tr>
<tr>
<td>GRN</td>
<td>Cytoplasm</td>
<td>0.0253</td>
<td>3.16</td>
</tr>
<tr>
<td>ALDH1A1</td>
<td>Cytoplasm</td>
<td>0.0066</td>
<td>3.06</td>
</tr>
<tr>
<td>CMPK2</td>
<td>Cytoplasm</td>
<td>0.0308</td>
<td>2.96</td>
</tr>
<tr>
<td>MAN1A1</td>
<td>Cytoplasm</td>
<td>0.0199</td>
<td>2.80</td>
</tr>
<tr>
<td>SERPINB9</td>
<td>Cytoplasm</td>
<td>0.0052</td>
<td>2.75</td>
</tr>
<tr>
<td>CLPP</td>
<td>Cytoplasm</td>
<td>0.0301</td>
<td>2.72</td>
</tr>
<tr>
<td>KLC1</td>
<td>Cytoplasm</td>
<td>0.0153</td>
<td>2.60</td>
</tr>
<tr>
<td>ANXA2</td>
<td>Cytoplasm</td>
<td>0.0126</td>
<td>2.58</td>
</tr>
<tr>
<td>KTI12</td>
<td>Cytoplasm</td>
<td>0.0175</td>
<td>2.58</td>
</tr>
<tr>
<td>INTS12</td>
<td>Nucleus</td>
<td>0.0159</td>
<td>2.57</td>
</tr>
<tr>
<td>NMT1</td>
<td>Nucleus</td>
<td>0.0451</td>
<td>2.55</td>
</tr>
<tr>
<td>PPP4R1</td>
<td>Cytoplasm</td>
<td>0.0004</td>
<td>2.54</td>
</tr>
<tr>
<td>EVI2B</td>
<td>Cytoplasm</td>
<td>0.0428</td>
<td>2.51</td>
</tr>
<tr>
<td>COL4A3BP</td>
<td>Cytoplasm</td>
<td>0.0058</td>
<td>2.50</td>
</tr>
<tr>
<td>NT5C</td>
<td>Cytoplasm</td>
<td>0.0019</td>
<td>2.48</td>
</tr>
<tr>
<td>RB1</td>
<td>Nucleus</td>
<td>0.0154</td>
<td>2.43</td>
</tr>
<tr>
<td>NIT1</td>
<td>Cytoplasm</td>
<td>0.0013</td>
<td>2.40</td>
</tr>
<tr>
<td>HSPB1</td>
<td>Cytoplasm</td>
<td>0.0065</td>
<td>2.36</td>
</tr>
<tr>
<td>HPGDS</td>
<td>Cytoplasm</td>
<td>0.0381</td>
<td>2.36</td>
</tr>
<tr>
<td>ICAM1</td>
<td>Cytoplasm</td>
<td>0.0242</td>
<td>2.35</td>
</tr>
<tr>
<td>Gene</td>
<td>Compartment</td>
<td>Average</td>
<td>Z Score</td>
</tr>
<tr>
<td>----------</td>
<td>-------------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>JAK1</td>
<td>Cytoplasm</td>
<td>0.0468</td>
<td>2.32</td>
</tr>
<tr>
<td>RAB2A</td>
<td>Cytoplasm</td>
<td>0.0117</td>
<td>2.17</td>
</tr>
<tr>
<td>RNASEH2B</td>
<td>Cytoplasm</td>
<td>0.0079</td>
<td>2.16</td>
</tr>
<tr>
<td>NUCB1</td>
<td>Cytoplasm</td>
<td>0.0033</td>
<td>2.15</td>
</tr>
<tr>
<td>PDCD4</td>
<td>Cytoplasm</td>
<td>0.0051</td>
<td>2.11</td>
</tr>
<tr>
<td>ZNF142</td>
<td>Cytoplasm</td>
<td>0.0259</td>
<td>2.10</td>
</tr>
<tr>
<td>LAIR1</td>
<td>Cytoplasm</td>
<td>0.0419</td>
<td>2.10</td>
</tr>
<tr>
<td>TXN</td>
<td>Cytoplasm</td>
<td>0.0109</td>
<td>2.08</td>
</tr>
<tr>
<td>BRCC3</td>
<td>Cytoplasm</td>
<td>0.0080</td>
<td>2.07</td>
</tr>
<tr>
<td>NKAP</td>
<td>Nucleus</td>
<td>0.0004</td>
<td>2.06</td>
</tr>
<tr>
<td>ARRD1C1</td>
<td>Nucleus</td>
<td>0.0396</td>
<td>2.06</td>
</tr>
<tr>
<td>CENPF</td>
<td>Nucleus</td>
<td>0.0094</td>
<td>2.05</td>
</tr>
<tr>
<td>NEDD8</td>
<td>Cytoplasm</td>
<td>0.0043</td>
<td>2.00</td>
</tr>
<tr>
<td>PYM1</td>
<td>Cytoplasm</td>
<td>0.0025</td>
<td>1.99</td>
</tr>
<tr>
<td>ZNF428</td>
<td>Cytoplasm</td>
<td>0.0029</td>
<td>1.97</td>
</tr>
<tr>
<td>SPTAN1</td>
<td>Nucleus</td>
<td>0.0154</td>
<td>1.95</td>
</tr>
<tr>
<td>USP47</td>
<td>Cytoplasm</td>
<td>0.0424</td>
<td>1.95</td>
</tr>
<tr>
<td>FAM91A1</td>
<td>Cytoplasm</td>
<td>0.0223</td>
<td>1.93</td>
</tr>
<tr>
<td>DNASE2</td>
<td>Cytoplasm</td>
<td>0.0485</td>
<td>1.90</td>
</tr>
<tr>
<td>MARCKSL1</td>
<td>Cytoplasm</td>
<td>0.0115</td>
<td>1.90</td>
</tr>
<tr>
<td>RNF214</td>
<td>Cytoplasm</td>
<td>0.0036</td>
<td>1.87</td>
</tr>
<tr>
<td>FKBP8</td>
<td>Nucleus</td>
<td>0.0046</td>
<td>1.87</td>
</tr>
<tr>
<td>ISYNA1</td>
<td>Cytoplasm</td>
<td>0.0356</td>
<td>1.86</td>
</tr>
<tr>
<td>HIST1H1B</td>
<td>Cytoplasm</td>
<td>0.0365</td>
<td>1.86</td>
</tr>
<tr>
<td>MED12</td>
<td>Nucleus</td>
<td>0.0447</td>
<td>1.86</td>
</tr>
<tr>
<td>B2M</td>
<td>Cytoplasm</td>
<td>0.0474</td>
<td>1.83</td>
</tr>
<tr>
<td>ap1g2</td>
<td>Cytoplasm</td>
<td>0.0395</td>
<td>1.82</td>
</tr>
<tr>
<td>EMG1</td>
<td>Cytoplasm</td>
<td>0.0042</td>
<td>1.82</td>
</tr>
<tr>
<td>NDUFA4</td>
<td>Nucleus</td>
<td>0.0430</td>
<td>1.80</td>
</tr>
<tr>
<td>PRKAR2A</td>
<td>Nucleus</td>
<td>0.0220</td>
<td>1.78</td>
</tr>
<tr>
<td>MRPS6</td>
<td>Cytoplasm</td>
<td>0.0024</td>
<td>1.77</td>
</tr>
<tr>
<td>ARFIP2</td>
<td>Cytoplasm</td>
<td>0.0019</td>
<td>1.76</td>
</tr>
<tr>
<td>IKBKG</td>
<td>Cytoplasm</td>
<td>0.0443</td>
<td>1.75</td>
</tr>
<tr>
<td>RUNX1</td>
<td>Nucleus</td>
<td>0.0208</td>
<td>1.73</td>
</tr>
<tr>
<td>BICD2</td>
<td>Cytoplasm</td>
<td>0.0303</td>
<td>1.72</td>
</tr>
<tr>
<td>STX7</td>
<td>Cytoplasm</td>
<td>0.0396</td>
<td>1.72</td>
</tr>
<tr>
<td>Gene</td>
<td>Compartment</td>
<td>log2 Fold Change</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>---------------</td>
<td>-----------------</td>
<td></td>
</tr>
<tr>
<td>HSP90AB4P</td>
<td>Nucleus</td>
<td>1.71</td>
<td></td>
</tr>
<tr>
<td>CCT8</td>
<td>Nucleus</td>
<td>1.70</td>
<td></td>
</tr>
<tr>
<td>PTER</td>
<td>Cytoplasm</td>
<td>1.69</td>
<td></td>
</tr>
<tr>
<td>DBNL</td>
<td>Nucleus</td>
<td>1.68</td>
<td></td>
</tr>
<tr>
<td>IFI16</td>
<td>Cytoplasm</td>
<td>1.68</td>
<td></td>
</tr>
<tr>
<td>RAC1</td>
<td>Cytoplasm</td>
<td>1.66</td>
<td></td>
</tr>
<tr>
<td>ITGB1</td>
<td>Cytoplasm</td>
<td>1.66</td>
<td></td>
</tr>
<tr>
<td>SH3BGR1</td>
<td>Cytoplasm</td>
<td>1.65</td>
<td></td>
</tr>
<tr>
<td>LGALS9</td>
<td>Cytoplasm</td>
<td>1.65</td>
<td></td>
</tr>
<tr>
<td>TSHZ1</td>
<td>Cytoplasm</td>
<td>1.62</td>
<td></td>
</tr>
<tr>
<td>DDAH2</td>
<td>Cytoplasm</td>
<td>1.61</td>
<td></td>
</tr>
<tr>
<td>PLEKHA7</td>
<td>Cytoplasm</td>
<td>1.60</td>
<td></td>
</tr>
<tr>
<td>EIF6</td>
<td>Cytoplasm</td>
<td>1.57</td>
<td></td>
</tr>
<tr>
<td>SERPINH1</td>
<td>Cytoplasm</td>
<td>1.56</td>
<td></td>
</tr>
<tr>
<td>TM9SF2</td>
<td>Cytoplasm</td>
<td>1.56</td>
<td></td>
</tr>
<tr>
<td>KMT2A</td>
<td>Nucleus</td>
<td>1.55</td>
<td></td>
</tr>
<tr>
<td>RPS27L</td>
<td>Nucleus</td>
<td>1.55</td>
<td></td>
</tr>
<tr>
<td>RPS2</td>
<td>Nucleus</td>
<td>1.55</td>
<td></td>
</tr>
<tr>
<td>ARHGDIB</td>
<td>Cytoplasm</td>
<td>1.54</td>
<td></td>
</tr>
<tr>
<td>SEC23A</td>
<td>Cytoplasm</td>
<td>1.53</td>
<td></td>
</tr>
<tr>
<td>SET</td>
<td>Cytoplasm</td>
<td>1.53</td>
<td></td>
</tr>
<tr>
<td>TRIM28</td>
<td>Cytoplasm</td>
<td>1.51</td>
<td></td>
</tr>
<tr>
<td>NADSYN1</td>
<td>Cytoplasm</td>
<td>1.51</td>
<td></td>
</tr>
<tr>
<td>DHX38</td>
<td>Cytoplasm</td>
<td>1.50</td>
<td></td>
</tr>
<tr>
<td>NDRG3</td>
<td>Cytoplasm</td>
<td>1.48</td>
<td></td>
</tr>
<tr>
<td>FKBP3</td>
<td>Cytoplasm</td>
<td>1.48</td>
<td></td>
</tr>
<tr>
<td>RPL13A</td>
<td>Nucleus</td>
<td>1.48</td>
<td></td>
</tr>
<tr>
<td>SERPINB6</td>
<td>Cytoplasm</td>
<td>1.47</td>
<td></td>
</tr>
<tr>
<td>NAMPT</td>
<td>Cytoplasm</td>
<td>1.47</td>
<td></td>
</tr>
<tr>
<td>TMEM263</td>
<td>Cytoplasm</td>
<td>1.46</td>
<td></td>
</tr>
<tr>
<td>LASP1</td>
<td>Cytoplasm</td>
<td>1.46</td>
<td></td>
</tr>
<tr>
<td>ATP6V1C1</td>
<td>Cytoplasm</td>
<td>1.46</td>
<td></td>
</tr>
<tr>
<td>FKBP1A</td>
<td>Cytoplasm</td>
<td>1.46</td>
<td></td>
</tr>
<tr>
<td>MPI</td>
<td>Nucleus</td>
<td>1.46</td>
<td></td>
</tr>
<tr>
<td>CSNK2A2</td>
<td>Cytoplasm</td>
<td>1.46</td>
<td></td>
</tr>
<tr>
<td>ACBD3</td>
<td>Cytoplasm</td>
<td>1.45</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Compartment</td>
<td>Expression</td>
<td>Fold Change</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>------------</td>
<td>-------------</td>
</tr>
<tr>
<td>IRGQ</td>
<td>Cytoplasm</td>
<td>0.0024</td>
<td>1.45</td>
</tr>
<tr>
<td>VARS</td>
<td>Cytoplasm</td>
<td>0.0287</td>
<td>1.45</td>
</tr>
<tr>
<td>LAMP1</td>
<td>Cytoplasm</td>
<td>0.0066</td>
<td>1.45</td>
</tr>
<tr>
<td>RRM1</td>
<td>Cytoplasm</td>
<td>0.0018</td>
<td>1.45</td>
</tr>
<tr>
<td>PRDX1</td>
<td>Cytoplasm</td>
<td>0.0219</td>
<td>1.44</td>
</tr>
<tr>
<td>RPE</td>
<td>Cytoplasm</td>
<td>0.0425</td>
<td>1.44</td>
</tr>
<tr>
<td>ZBTB48</td>
<td>Nucleus</td>
<td>0.0069</td>
<td>1.43</td>
</tr>
<tr>
<td>RPLP2</td>
<td>Nucleus</td>
<td>0.0083</td>
<td>1.43</td>
</tr>
<tr>
<td>CDC73</td>
<td>Cytoplasm</td>
<td>0.0218</td>
<td>1.43</td>
</tr>
<tr>
<td>EDF1</td>
<td>Nucleus</td>
<td>0.0110</td>
<td>1.43</td>
</tr>
<tr>
<td>PHAX</td>
<td>Cytoplasm</td>
<td>0.0311</td>
<td>1.43</td>
</tr>
<tr>
<td>GLB1</td>
<td>Cytoplasm</td>
<td>0.0480</td>
<td>1.43</td>
</tr>
<tr>
<td>rbm12</td>
<td>Cytoplasm</td>
<td>0.0046</td>
<td>1.42</td>
</tr>
<tr>
<td>STMN1</td>
<td>Cytoplasm</td>
<td>0.0029</td>
<td>1.42</td>
</tr>
<tr>
<td>BAZ2A</td>
<td>Nucleus</td>
<td>0.0366</td>
<td>1.42</td>
</tr>
<tr>
<td>CAP1</td>
<td>Cytoplasm</td>
<td>0.0358</td>
<td>1.41</td>
</tr>
<tr>
<td>TMED10</td>
<td>Cytoplasm</td>
<td>0.0171</td>
<td>1.40</td>
</tr>
<tr>
<td>SNX6</td>
<td>Cytoplasm</td>
<td>0.0007</td>
<td>1.40</td>
</tr>
<tr>
<td>AGPAT1</td>
<td>Cytoplasm</td>
<td>0.0449</td>
<td>1.40</td>
</tr>
<tr>
<td>APBB1IP</td>
<td>Cytoplasm</td>
<td>0.0042</td>
<td>1.40</td>
</tr>
<tr>
<td>ARHGAP32</td>
<td>Cytoplasm</td>
<td>0.0260</td>
<td>1.40</td>
</tr>
<tr>
<td>PPIA</td>
<td>Cytoplasm</td>
<td>0.0084</td>
<td>1.39</td>
</tr>
<tr>
<td>SEC23IP</td>
<td>Cytoplasm</td>
<td>0.0255</td>
<td>1.39</td>
</tr>
<tr>
<td>CBFB</td>
<td>Nucleus</td>
<td>0.0299</td>
<td>1.39</td>
</tr>
<tr>
<td>ACTR3</td>
<td>Cytoplasm</td>
<td>0.0182</td>
<td>1.39</td>
</tr>
<tr>
<td>CFL1</td>
<td>Cytoplasm</td>
<td>0.0201</td>
<td>1.38</td>
</tr>
<tr>
<td>ETFB</td>
<td>Cytoplasm</td>
<td>0.0481</td>
<td>1.38</td>
</tr>
<tr>
<td>POTEF</td>
<td>Cytoplasm</td>
<td>0.0269</td>
<td>1.37</td>
</tr>
<tr>
<td>INPP5D</td>
<td>Nucleus</td>
<td>0.0111</td>
<td>1.37</td>
</tr>
<tr>
<td>GSTO1</td>
<td>Cytoplasm</td>
<td>0.0022</td>
<td>1.37</td>
</tr>
<tr>
<td>PGK1</td>
<td>Cytoplasm</td>
<td>0.0028</td>
<td>1.36</td>
</tr>
<tr>
<td>RPL7</td>
<td>Nucleus</td>
<td>0.0405</td>
<td>1.35</td>
</tr>
<tr>
<td>MTHFD2</td>
<td>Cytoplasm</td>
<td>0.0348</td>
<td>1.35</td>
</tr>
<tr>
<td>PSMF1</td>
<td>Cytoplasm</td>
<td>0.0338</td>
<td>1.35</td>
</tr>
<tr>
<td>PSMC3</td>
<td>Cytoplasm</td>
<td>0.0166</td>
<td>1.35</td>
</tr>
<tr>
<td>UROD</td>
<td>Cytoplasm</td>
<td>0.0224</td>
<td>1.34</td>
</tr>
<tr>
<td>Gene</td>
<td>Location</td>
<td>Log2 Fold Change</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>----------</td>
<td>-----------------</td>
<td></td>
</tr>
<tr>
<td>AMPD2</td>
<td>Cytoplasm</td>
<td>0.0087</td>
<td>1.34</td>
</tr>
<tr>
<td>PLIN3</td>
<td>Cytoplasm</td>
<td>0.0346</td>
<td>1.33</td>
</tr>
<tr>
<td>NDUVF2</td>
<td>Nucleus</td>
<td>0.0352</td>
<td>1.33</td>
</tr>
<tr>
<td>MMS22L</td>
<td>Cytoplasm</td>
<td>0.0238</td>
<td>1.32</td>
</tr>
<tr>
<td>LRRFIP1</td>
<td>Cytoplasm</td>
<td>0.0070</td>
<td>1.31</td>
</tr>
<tr>
<td>NUP93</td>
<td>Nucleus</td>
<td>0.0323</td>
<td>1.31</td>
</tr>
<tr>
<td>CCT6A</td>
<td>Cytoplasm</td>
<td>0.0254</td>
<td>1.31</td>
</tr>
<tr>
<td>MRPS30</td>
<td>Cytoplasm</td>
<td>0.0145</td>
<td>1.30</td>
</tr>
<tr>
<td>PHIP</td>
<td>Nucleus</td>
<td>0.0134</td>
<td>1.30</td>
</tr>
<tr>
<td>URM1</td>
<td>Cytoplasm</td>
<td>0.0168</td>
<td>1.30</td>
</tr>
<tr>
<td>THUMPD1</td>
<td>Cytoplasm</td>
<td>0.0169</td>
<td>1.29</td>
</tr>
<tr>
<td>HEXA</td>
<td>Cytoplasm</td>
<td>0.0157</td>
<td>1.29</td>
</tr>
<tr>
<td>CASP6</td>
<td>Cytoplasm</td>
<td>0.0500</td>
<td>1.27</td>
</tr>
<tr>
<td>PFDN1</td>
<td>Cytoplasm</td>
<td>0.0318</td>
<td>1.27</td>
</tr>
<tr>
<td>CDC42BPB</td>
<td>Nucleus</td>
<td>0.0350</td>
<td>1.26</td>
</tr>
<tr>
<td>ADH5</td>
<td>Cytoplasm</td>
<td>0.0467</td>
<td>1.25</td>
</tr>
<tr>
<td>KNTC1</td>
<td>Cytoplasm</td>
<td>0.0005</td>
<td>1.25</td>
</tr>
<tr>
<td>ABCF1</td>
<td>Cytoplasm</td>
<td>0.0382</td>
<td>1.25</td>
</tr>
<tr>
<td>LPXN</td>
<td>Cytoplasm</td>
<td>0.0169</td>
<td>1.25</td>
</tr>
<tr>
<td>RPS8</td>
<td>Nucleus</td>
<td>0.0022</td>
<td>1.22</td>
</tr>
<tr>
<td>ABHD10</td>
<td>Cytoplasm</td>
<td>0.0221</td>
<td>1.22</td>
</tr>
<tr>
<td>EWSR1</td>
<td>Cytoplasm</td>
<td>0.0448</td>
<td>1.19</td>
</tr>
<tr>
<td>RBM10</td>
<td>Nucleus</td>
<td>0.0382</td>
<td>1.13</td>
</tr>
<tr>
<td>PSMB2</td>
<td>Cytoplasm</td>
<td>0.0069</td>
<td>1.12</td>
</tr>
<tr>
<td>TUFM</td>
<td>Cytoplasm</td>
<td>0.0229</td>
<td>1.10</td>
</tr>
<tr>
<td>GAR1</td>
<td>Nucleus</td>
<td>0.0138</td>
<td>-1.06</td>
</tr>
<tr>
<td>FAM96A</td>
<td>Cytoplasm</td>
<td>0.0283</td>
<td>-1.13</td>
</tr>
<tr>
<td>RPA3</td>
<td>Cytoplasm</td>
<td>0.0090</td>
<td>-1.15</td>
</tr>
<tr>
<td>NUDT5</td>
<td>Cytoplasm</td>
<td>0.0489</td>
<td>-1.16</td>
</tr>
<tr>
<td>STAG2</td>
<td>Nucleus</td>
<td>0.0162</td>
<td>-1.17</td>
</tr>
<tr>
<td>RPS21</td>
<td>Cytoplasm</td>
<td>0.0208</td>
<td>-1.18</td>
</tr>
<tr>
<td>POLR2H</td>
<td>Nucleus</td>
<td>0.0157</td>
<td>-1.18</td>
</tr>
<tr>
<td>NCBP1</td>
<td>Cytoplasm</td>
<td>0.0233</td>
<td>-1.18</td>
</tr>
<tr>
<td>CUL1</td>
<td>Cytoplasm</td>
<td>0.0285</td>
<td>-1.18</td>
</tr>
<tr>
<td>SUPT5H</td>
<td>Nucleus</td>
<td>0.0042</td>
<td>-1.19</td>
</tr>
<tr>
<td>zbtb49</td>
<td>Nucleus</td>
<td>0.0389</td>
<td>-1.19</td>
</tr>
<tr>
<td>Gene</td>
<td>Location</td>
<td>Log2 Fold Change</td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>----------</td>
<td>-----------------</td>
<td></td>
</tr>
<tr>
<td>PKM</td>
<td>Cytoplasm</td>
<td>0.0474</td>
<td></td>
</tr>
<tr>
<td>RIF1</td>
<td>Nucleus</td>
<td>0.0007</td>
<td></td>
</tr>
<tr>
<td>SNRPD3</td>
<td>Cytoplasm</td>
<td>0.0278</td>
<td></td>
</tr>
<tr>
<td>PTPN23</td>
<td>Cytoplasm</td>
<td>0.0491</td>
<td></td>
</tr>
<tr>
<td>RETSAT</td>
<td>Cytoplasm</td>
<td>0.0099</td>
<td></td>
</tr>
<tr>
<td>SERPINC1</td>
<td>Cytoplasm</td>
<td>0.0334</td>
<td></td>
</tr>
<tr>
<td>ESF1</td>
<td>Nucleus</td>
<td>0.0476</td>
<td></td>
</tr>
<tr>
<td>EXOSC9</td>
<td>Nucleus</td>
<td>0.0156</td>
<td></td>
</tr>
<tr>
<td>CORO1A</td>
<td>Cytoplasm</td>
<td>0.0097</td>
<td></td>
</tr>
<tr>
<td>NUP88</td>
<td>Cytoplasm</td>
<td>0.0405</td>
<td></td>
</tr>
<tr>
<td>ITIH2</td>
<td>Nucleus</td>
<td>0.0397</td>
<td></td>
</tr>
<tr>
<td>NDUFB11</td>
<td>Cytoplasm</td>
<td>0.0468</td>
<td></td>
</tr>
<tr>
<td>TOR1AIP2</td>
<td>Cytoplasm</td>
<td>0.0221</td>
<td></td>
</tr>
<tr>
<td>POLA2</td>
<td>Cytoplasm</td>
<td>0.0404</td>
<td></td>
</tr>
<tr>
<td>TARS2</td>
<td>Cytoplasm</td>
<td>0.0397</td>
<td></td>
</tr>
<tr>
<td>SCAF11</td>
<td>Nucleus</td>
<td>0.0381</td>
<td></td>
</tr>
<tr>
<td>INCENP</td>
<td>Nucleus</td>
<td>0.0008</td>
<td></td>
</tr>
<tr>
<td>MCMBP</td>
<td>Cytoplasm</td>
<td>0.0297</td>
<td></td>
</tr>
<tr>
<td>c8orf33</td>
<td>Nucleus</td>
<td>0.0068</td>
<td></td>
</tr>
<tr>
<td>UCK2</td>
<td>Cytoplasm</td>
<td>0.0012</td>
<td></td>
</tr>
<tr>
<td>HIRA</td>
<td>Nucleus</td>
<td>0.0380</td>
<td></td>
</tr>
<tr>
<td>LIMA1</td>
<td>Nucleus</td>
<td>0.0358</td>
<td></td>
</tr>
<tr>
<td>ACSL3</td>
<td>Cytoplasm</td>
<td>0.0406</td>
<td></td>
</tr>
<tr>
<td>HIRIP3</td>
<td>Nucleus</td>
<td>0.0042</td>
<td></td>
</tr>
<tr>
<td>PYCARD</td>
<td>Cytoplasm</td>
<td>0.0476</td>
<td></td>
</tr>
<tr>
<td>CYFIP1</td>
<td>Cytoplasm</td>
<td>0.0347</td>
<td></td>
</tr>
<tr>
<td>EXOSC4</td>
<td>Nucleus</td>
<td>0.0011</td>
<td></td>
</tr>
<tr>
<td>KPNA2</td>
<td>Cytoplasm</td>
<td>0.0006</td>
<td></td>
</tr>
<tr>
<td>RPS3</td>
<td>Cytoplasm</td>
<td>0.0164</td>
<td></td>
</tr>
<tr>
<td>PTPN6</td>
<td>Cytoplasm</td>
<td>0.0405</td>
<td></td>
</tr>
<tr>
<td>KIF20B</td>
<td>Nucleus</td>
<td>0.0366</td>
<td></td>
</tr>
<tr>
<td>MCM7</td>
<td>Nucleus</td>
<td>0.0196</td>
<td></td>
</tr>
<tr>
<td>ARRB2</td>
<td>Nucleus</td>
<td>0.0200</td>
<td></td>
</tr>
<tr>
<td>CHAF1B</td>
<td>Nucleus</td>
<td>0.0059</td>
<td></td>
</tr>
<tr>
<td>TBC1D5</td>
<td>Cytoplasm</td>
<td>0.0478</td>
<td></td>
</tr>
<tr>
<td>ITGB2</td>
<td>Cytoplasm</td>
<td>0.0074</td>
<td></td>
</tr>
<tr>
<td>Protein</td>
<td>Location</td>
<td>Value 1</td>
<td>Value 2</td>
</tr>
<tr>
<td>----------</td>
<td>----------</td>
<td>---------</td>
<td>---------</td>
</tr>
<tr>
<td>LAMTOR2</td>
<td>Cytoplasm</td>
<td>0.0283</td>
<td>-1.49</td>
</tr>
<tr>
<td>CPSF3</td>
<td>Cytoplasm</td>
<td>0.0364</td>
<td>-1.50</td>
</tr>
<tr>
<td>CAPG</td>
<td>Cytoplasm</td>
<td>0.0473</td>
<td>-1.50</td>
</tr>
<tr>
<td>EFHD2</td>
<td>Cytoplasm</td>
<td>0.0274</td>
<td>-1.52</td>
</tr>
<tr>
<td>MCM4</td>
<td>Nucleus</td>
<td>0.0245</td>
<td>-1.53</td>
</tr>
<tr>
<td>KIF22</td>
<td>Nucleus</td>
<td>0.0474</td>
<td>-1.53</td>
</tr>
<tr>
<td>MARS2</td>
<td>Cytoplasm</td>
<td>0.0310</td>
<td>-1.54</td>
</tr>
<tr>
<td>ATL3</td>
<td>Cytoplasm</td>
<td>0.0176</td>
<td>-1.54</td>
</tr>
<tr>
<td>LSP1</td>
<td>Nucleus</td>
<td>0.0086</td>
<td>-1.54</td>
</tr>
<tr>
<td>SLC1A4</td>
<td>Cytoplasm</td>
<td>0.0328</td>
<td>-1.56</td>
</tr>
<tr>
<td>NCF4</td>
<td>Cytoplasm</td>
<td>0.0125</td>
<td>-1.56</td>
</tr>
<tr>
<td>SNRPE</td>
<td>Cytoplasm</td>
<td>0.0250</td>
<td>-1.58</td>
</tr>
<tr>
<td>TAGLN2</td>
<td>Cytoplasm</td>
<td>0.0487</td>
<td>-1.58</td>
</tr>
<tr>
<td>KRI1</td>
<td>Nucleus</td>
<td>0.0006</td>
<td>-1.59</td>
</tr>
<tr>
<td>SLC16A3</td>
<td>Cytoplasm</td>
<td>0.0258</td>
<td>-1.60</td>
</tr>
<tr>
<td>MKI67</td>
<td>Nucleus</td>
<td>0.0402</td>
<td>-1.63</td>
</tr>
<tr>
<td>CKAP2</td>
<td>Nucleus</td>
<td>0.0356</td>
<td>-1.66</td>
</tr>
<tr>
<td>CARNMT1</td>
<td>Cytoplasm</td>
<td>0.0174</td>
<td>-1.67</td>
</tr>
<tr>
<td>GALE</td>
<td>Cytoplasm</td>
<td>0.0235</td>
<td>-1.70</td>
</tr>
<tr>
<td>MPO</td>
<td>Cytoplasm</td>
<td>0.0189</td>
<td>-1.71</td>
</tr>
<tr>
<td>ALOX5AP</td>
<td>Cytoplasm</td>
<td>0.0089</td>
<td>-1.76</td>
</tr>
<tr>
<td>GFAP</td>
<td>Nucleus</td>
<td>0.0355</td>
<td>-1.79</td>
</tr>
<tr>
<td>TOP2A</td>
<td>Nucleus</td>
<td>0.0042</td>
<td>-1.81</td>
</tr>
<tr>
<td>TTF2</td>
<td>Nucleus</td>
<td>0.0371</td>
<td>-1.83</td>
</tr>
<tr>
<td>MORF4L2</td>
<td>Nucleus</td>
<td>0.0362</td>
<td>-1.84</td>
</tr>
<tr>
<td>LGALS1</td>
<td>Cytoplasm</td>
<td>0.0223</td>
<td>-1.85</td>
</tr>
<tr>
<td>NDUFS6</td>
<td>Cytoplasm</td>
<td>0.0213</td>
<td>-1.88</td>
</tr>
<tr>
<td>SRP19</td>
<td>Cytoplasm</td>
<td>0.0141</td>
<td>-1.88</td>
</tr>
<tr>
<td>SPI1</td>
<td>Nucleus</td>
<td>0.0011</td>
<td>-1.89</td>
</tr>
<tr>
<td>DOK2</td>
<td>Cytoplasm</td>
<td>0.0497</td>
<td>-1.92</td>
</tr>
<tr>
<td>XPNPEP1</td>
<td>Nucleus</td>
<td>0.0327</td>
<td>-1.99</td>
</tr>
<tr>
<td>XRCC4</td>
<td>Nucleus</td>
<td>0.0339</td>
<td>-2.02</td>
</tr>
<tr>
<td>OXSM</td>
<td>Cytoplasm</td>
<td>0.0296</td>
<td>-2.06</td>
</tr>
<tr>
<td>CEBPB</td>
<td>Nucleus</td>
<td>0.0117</td>
<td>-2.09</td>
</tr>
<tr>
<td>RAD51C</td>
<td>Cytoplasm</td>
<td>0.0429</td>
<td>-2.13</td>
</tr>
<tr>
<td>MPO</td>
<td>Nucleus</td>
<td>0.0357</td>
<td>-2.15</td>
</tr>
<tr>
<td>Gene</td>
<td>Location</td>
<td>p-value</td>
<td>log2 Fold Change</td>
</tr>
<tr>
<td>-------</td>
<td>----------</td>
<td>---------</td>
<td>-----------------</td>
</tr>
<tr>
<td>CTSG</td>
<td>Nucleus</td>
<td>0.0422</td>
<td>-2.15</td>
</tr>
<tr>
<td>SSR1</td>
<td>Cytoplasm</td>
<td>0.0119</td>
<td>-2.22</td>
</tr>
<tr>
<td>WBP11</td>
<td>Cytoplasm</td>
<td>0.0439</td>
<td>-2.24</td>
</tr>
<tr>
<td>MRPS2</td>
<td>Cytoplasm</td>
<td>0.0105</td>
<td>-2.25</td>
</tr>
<tr>
<td>SAP18</td>
<td>Cytoplasm</td>
<td>0.0279</td>
<td>-2.42</td>
</tr>
<tr>
<td>AURKA</td>
<td>Nucleus</td>
<td>0.0212</td>
<td>-2.44</td>
</tr>
<tr>
<td>PRG2</td>
<td>Nucleus</td>
<td>0.0153</td>
<td>-2.60</td>
</tr>
<tr>
<td>MINA</td>
<td>Nucleus</td>
<td>0.0119</td>
<td>-2.64</td>
</tr>
<tr>
<td>CHTF18</td>
<td>Cytoplasm</td>
<td>0.0326</td>
<td>-2.71</td>
</tr>
<tr>
<td>CTSG</td>
<td>Cytoplasm</td>
<td>0.0407</td>
<td>-2.94</td>
</tr>
<tr>
<td>KRT10</td>
<td>Cytoplasm</td>
<td>0.0271</td>
<td>-3.14</td>
</tr>
<tr>
<td>APOBR</td>
<td>Cytoplasm</td>
<td>0.0286</td>
<td>-3.71</td>
</tr>
<tr>
<td>CLC</td>
<td>Cytoplasm</td>
<td>0.0165</td>
<td>-5.11</td>
</tr>
<tr>
<td>RNASE3</td>
<td>Cytoplasm</td>
<td>0.0069</td>
<td>-7.84</td>
</tr>
</tbody>
</table>
Supplementary Figures

Supplementary Figure 1 - Representative gating strategy used to immunophenotype HSPC

Representative density plots and histograms outlining the gating strategy used for the immunophenotypic analysis of HSPC, as described in XX. (A) Example gate used to exclude cell debris, based on FSC and SSC; (B) The percentage of GFP⁺ cells within a bulk population was established based on the ‘non-debris’ gate shown in (A); (C) The CD13-APC and CD36-PerCP-Cy5 markers were used to discriminate between the granulocytic, monocytic and erythroid lineages within the GFP⁺ population. Background auto fluorescence was established using cells subjected to the equivalent viral infection procedure but in the absence of virus (mock culture). (D-E) Each subpopulation was subsequently examined for differentiation markers, such as the granulocytic marker CD15 or the monocytic marker CD14. Cells stained the isotype control IgG (grey) were used to set the appropriate gating strategies.
Supplementary Figure 2 – Expression of control GFP and ZNF217 in human HSPC

Representative density plots showing the expression of GFP in control and ZNF217 overexpressing cultures, throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, as described in Supplementary Figure 1A. Background autofluorescence was established using cells subjected to the equivalent viral infection procedure but in the absence of virus (mock culture).
Supplementary Figure 3 – Expression of the stem cell marker CD34 in monocytes during differentiation, following ZNF217 overexpression

Representative contour plots showing the expression of CD34 in monocytes (CD13+ CD36+) transduced with a control or a ZNF217 overexpression vector throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP+ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
Supplementary Figure 4 – Expression of the myeloid marker CD11b in monocytes during differentiation, following ZNF217 overexpression

Representative contour plots showing the expression of CD11b in monocytes (CD13⁺ CD36⁺) transduced with a control or a ZNF217 overexpression vector throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP⁺ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
Supplementary Figure 5 – Expression of the monocytic marker CD14 in monocytes during differentiation, following ZNF217 overexpression

Representative contour plots showing the expression of CD14 in monocytes (CD13⁺ CD36⁺) transduced with a control or a ZNF217 overexpression vector throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP⁺ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
Supplementary Figure 6 – Expression of the stem cell marker CD34 in granulocytes during differentiation, following ZNF217 overexpression

Representative contour plots showing the expression of CD34 in granulocytes (CD13⁺/CD36⁻) transduced with a control or a ZNF217 overexpression vector throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP⁷ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
Supplementary Figure 7 – Expression of the myeloid marker CD11b in granulocytes during differentiation, following ZNF217 overexpression

Representative contour plots showing the expression of CD11b in granulocytes (CD13^+/− CD36^−) transduced with a control or a ZNF217 overexpression vector throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP^+ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
Supplementary Figure 8 – Expression of the granulocytic marker CD15 in granulocytes during differentiation, following ZNF217 overexpression

Representative contour plots showing the expression of CD15 in granulocytes (CD13⁺/⁻ CD36⁻) transduced with a control or a ZNF217 overexpression vector throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP⁺ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
<table>
<thead>
<tr>
<th>Days in culture</th>
<th>3</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shZNF217 #1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shZNF217 #2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shZNF217 #3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GFP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- **Control**: FSC-H
- **shZNF217 #1**: FSC-H
- **shZNF217 #2**: FSC-H
- **shZNF217 #3**: FSC-H
- **GFP**: FSC-H

**Supplementary Figures**
Supplementary Figure 9 – Expression of control GFP and shZNF217 KD in human HSPC

Representative density plots showing the expression of GFP in control and shZNF217 KD cultures, throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, as described in Supplementary Figure 1A. Background autofluorescence was established using cells subjected to the equivalent viral infection procedure but in the absence of virus (mock culture).
Supplementary Figures

Days in culture

Control

shZNF217 #1

shZNF217 #2

shZNF217 #3

CD34-PE

FSC-H

<table>
<thead>
<tr>
<th>Days in culture</th>
<th>3</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>41.03%</td>
<td>0.44%</td>
<td>0.12%</td>
<td>0.21%</td>
<td>0.58%</td>
</tr>
<tr>
<td>shZNF217 #1</td>
<td>56.25%</td>
<td>1.18%</td>
<td>0.26%</td>
<td>0.53%</td>
<td>0.67%</td>
</tr>
<tr>
<td>shZNF217 #2</td>
<td>51.35%</td>
<td>1.07%</td>
<td>0.40%</td>
<td>0.64%</td>
<td>1.38%</td>
</tr>
<tr>
<td>shZNF217 #3</td>
<td>50.00%</td>
<td>1.26%</td>
<td>0.31%</td>
<td>0.57%</td>
<td>0.38%</td>
</tr>
</tbody>
</table>
Supplementary Figure 10 – Expression of the stem cell marker CD34 in monocytes during differentiation, following ZNF217 KD

Representative contour plots showing the expression of CD34 in monocytes (CD13⁺ CD36⁺) transduced with a control or a shZNF217 KD construct throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP⁺ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
Supplementary Figure 11 – Expression of the myeloid marker CD11b in monocytes during differentiation, following ZNF217 KD

Representative contour plots showing the expression of CD11b in monocytes (CD13⁺ CD36⁺) transduced with a control or a shZNF217 KD construct throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP⁺ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
Supplementary Figure 12 – Expression of the monocytic marker CD14 in monocytes during differentiation, following ZNF217 KD

Representative contour plots showing the expression of CD14 in monocytes (CD13+CD36+) transduced with a control or a shZNF217 KD construct throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP+ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
Supplementary Figure 13 – Expression of the stem cell marker CD34 in granulocytes during differentiation, following ZNF217 KD

Representative contour plots showing the expression of CD34 in granulocytes (CD13+/ CD36-) transduced with a control or a shZNF217 KD construct throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP+ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
Days in culture | 3 | 6 | 8 | 10 | 13

Control
- 0.16%
- 0.34%
- 5.23%
- 21.48%
- 34.50%

shZNF217 #1
- 0.41%
- 0.31%
- 0.65%
- 5.85%
- 27.76%

shZNF217 #2
- 0.35%
- 1.41%
- 0.65%
- 27.94%
- 45.20%

shZNF217 #3
- 0.18%
- 0.48%
- 1.04%
- 7.32%
- 27.42%

CD11b-PE
FSC-H
Supplementary Figure 14 – Expression of the myeloid marker CD11b in granulocytes during differentiation, following ZNF217 KD

Representative contour plots showing the expression of CD11b in granulocytes (CD13⁺/− CD36⁻) transduced with a control or a shZNF217 KD construct throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP⁺ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
Supplementary Figure 15 – Expression of the granulocytic marker CD15 in granulocytes during differentiation, following ZNF217 KD

Representative contour plots showing the expression of CD15 in granulocytes (CD13+ CD36−) transduced with a control or a shZNF217 KD construct throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP+ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
Supplementary Figures

Days in culture 3 6 8 10 13

Control

56.70% 56.00% 49.55% 46.55% 44.78%

shC/EBPβ #1

56.11% 52.74% 46.05% 42.23% 36.58%

shC/EBPβ #2

60.48% 51.40% 37.04% 28.72% 22.88%

shC/EBPβ #3

53.83% 41.86% 29.79% 23.42% 19.52%

FSC-H
Supplementary Figure 16 – Expression of control GFP and shC/EBPβ KD in human HSPC

Representative density plots showing the expression of GFP in control and shC/EBPβ KD cultures, throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, as described in Supplementary Figure 1A. Background autofluorescence was established using cells subjected to the equivalent viral infection procedure but in the absence of virus (mock culture).
Days in culture

Control

shC/EBPβ #1

shC/EBPβ #2

shC/EBPβ #3

CD34-PE

FSC-H
Supplementary Figure 17 – Expression of the stem cell marker CD34 in monocytes during differentiation, following C/EBPβ KD

Representative contour plots showing the expression of CD34 in monocytes (CD13⁺ CD36⁺) transduced with a control or a shC/EBPβ KD construct throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP⁺ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
Supplementary Figure 18 – Expression of the myeloid marker CD11b in monocytes during differentiation, following C/EBPβ KD

Representative contour plots showing the expression of CD11b in monocytes (CD13+CD36+) transduced with a control or a shC/EBPβ KD construct throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP+ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
Supplementary Figure 19 – Expression of the monocytic marker CD14 in monocytes during differentiation, following C/EBPβ KD

Representative contour plots showing the expression of CD14 in monocytes (CD13+ CD36+) transduced with a control or a shC/EBPβ KD construct throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP+ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
Days in culture

<table>
<thead>
<tr>
<th>Days</th>
<th>3</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>35.89%</td>
<td>1.19%</td>
<td>0.99%</td>
<td>0.27%</td>
<td>0.08%</td>
</tr>
<tr>
<td>shC/EBPβ #1</td>
<td>34.34%</td>
<td>1.42%</td>
<td>1.11%</td>
<td>0.40%</td>
<td>0.06%</td>
</tr>
<tr>
<td>shC/EBPβ #2</td>
<td>32.17%</td>
<td>1.48%</td>
<td>1.15%</td>
<td>0.37%</td>
<td>0.04%</td>
</tr>
<tr>
<td>shC/EBPβ #3</td>
<td>34.08%</td>
<td>2.37%</td>
<td>1.71%</td>
<td>0.67%</td>
<td>0.35%</td>
</tr>
</tbody>
</table>

FSC-H
Representative contour plots showing the expression of CD34 in granulocytes (CD13^+ CD36^-) transduced with a control or a shC/EBPβ KD construct throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP^+ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
Supplementary Figures

Days in culture

<table>
<thead>
<tr>
<th></th>
<th>3</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shC/EBPβ #1</td>
<td>0.28%</td>
<td>0.54%</td>
<td>2.05%</td>
<td>9.38%</td>
<td>18.69%</td>
</tr>
<tr>
<td>shC/EBPβ #2</td>
<td>0.36%</td>
<td>0.44%</td>
<td>0.93%</td>
<td>4.84%</td>
<td>13.77%</td>
</tr>
<tr>
<td>shC/EBPβ #3</td>
<td>0.26%</td>
<td>0.93%</td>
<td>2.21%</td>
<td>9.63%</td>
<td>15.88%</td>
</tr>
<tr>
<td>CD11b-PE</td>
<td>0.43%</td>
<td>0.70%</td>
<td>1.78%</td>
<td>6.54%</td>
<td>22.17%</td>
</tr>
</tbody>
</table>

FSC-H
Supplementary Figure 21 – Expression of the myeloid marker CD11b in granulocytes during differentiation, following C/EBPβ KD

Representative contour plots showing the expression of CD11b in granulocytes (CD13+/ CD36−) transduced with a control or a shC/EBPβ KD construct throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP+ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
Days in culture

<table>
<thead>
<tr>
<th>Days in culture</th>
<th>3</th>
<th>6</th>
<th>8</th>
<th>10</th>
<th>13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>shC/EBPβ #1</td>
<td>3.65%</td>
<td>5.16%</td>
<td>8.55%</td>
<td>16.60%</td>
<td>7.75%</td>
</tr>
<tr>
<td>shC/EBPβ #2</td>
<td>4.68%</td>
<td>4.45%</td>
<td>6.59%</td>
<td>15.67%</td>
<td>6.20%</td>
</tr>
<tr>
<td>shC/EBPβ #3</td>
<td>4.28%</td>
<td>6.58%</td>
<td>17.18%</td>
<td>25.84%</td>
<td>8.50%</td>
</tr>
<tr>
<td>CD15-PE</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FSC-H</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Supplementary Figure 22 – Expression of the granulocytic marker CD15 in granulocytes during differentiation, following C/EBPβ KD

Representative contour plots showing the expression of CD15 in granulocytes (CD13+/ CD36-) transduced with a control or a shC/EBPβ KD construct throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP+ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
## Supplementary Figure 23 – Expression of control GFP and C/EBPβ in human HSPC

Representative density plots showing the expression of GFP in control and ZNF217 overexpressing cultures, throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was C/EBPβ, as described in **Supplementary Figure 1A**. Background autofluorescence was established using cells subjected to the equivalent viral infection procedure but in the absence of virus (mock culture).
Supplementary Figure 24 – Expression of the stem cell marker CD34 in monocytes during differentiation, following C/EBPβ overexpression

Representative contour plots showing the expression of CD34 in monocytes (CD13⁺ CD36⁺) transduced with a control or a C/EBPβ overexpression vector throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP⁺ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
Supplementary Figure 25 – Expression of the myeloid marker CD11b in monocytes during differentiation, following C/EBPβ overexpression

Representative contour plots showing the expression of CD11b in monocytes (CD13⁺ CD36⁺) transduced with a control or a C/EBPβ overexpression vector throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP⁺ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
Supplementary Figure 26 – Expression of the myeloid marker CD14 in monocytes during differentiation, following C/EBPβ overexpression

Representative contour plots showing the expression of CD14 in monocytes (CD13+ CD36+) transduced with a control or a C/EBPβ overexpression vector throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP+ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
Supplementary Figure 27 – Expression of the stem cell marker CD34 in granulocytes during differentiation, following C/EBPβ overexpression

Representative contour plots showing the expression of CD34 in granulocytes (CD13+/- CD36-) transduced with a control or a C/EBPβ overexpression vector throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP+ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
Supplementary Figure 28 – Expression of the myeloid marker CD11b in granulocytes during differentiation, following C/EBPβ overexpression

Representative contour plots showing the expression of CD11b in granulocytes (CD13+/CD36-) transduced with a control or a C/EBPβ overexpression vector throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP+ cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
Supplementary Figure 29 – Expression of the granulocytic marker CD15 in granulocytes during differentiation, following C/EBPβ overexpression

Representative contour plots showing the expression of CD15 in granulocytes (CD13<sup>+</sup> CD36<sup>-</sup>) transduced with a control or a C/EBPβ overexpression vector throughout 13 days of culture, as measured by flow cytometry. An initial gating to exclude cell debris was applied, followed by the selection of GFP<sup>+</sup> cells, as described in Supplementary Figure 1. Transduced haematopoietic populations were subsequently delimited through the myeloid markers CD13 and CD36. Basal expression was determined using the corresponding IgG control.
Supplementary Figure 30 – ZNF217 overexpression inhibits monocytic and granulocytic growth during myeloid development

Summary dot plots showing: (A) Cumulative fold-expansion of control and ZNF217-overexpressing HSP GFP+ cells grown over 13 days in culture medium containing IL-3, SCF, G- and GM-CSF. (B) Cumulative fold-expansion of monocytic cells (CD13+ CD36+) transduced with a control of ZNF217-overexpressing vector. (C) Cumulative fold-expansion of granulocytic cells (CD13+ CD36-) transduced with a control of ZNF217-overexpressing vector. (D) Cumulative fold-expansion of erythroid cells (CD13 CD36+) transduced with a control of ZNF217-overexpressing vector. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Supplementary Figure 1. Data indicates Mean ± 1SD (n≥3) Significant differences between ZNF217-expressing cells and control at each time-point were analysed by paired t-test; * denotes p<0.05.
Supplementary Figure 31– ZNF217 overexpression disrupts the balance between granulocytic and erythroid populations during myeloid cell development

Summary dot plots showing the proportion of (A) monocyic (CD13+ CD36+), (B) granulocytic (CD13+/ CD36−) and (C) erythroid cells (CD13− CD36+) in control and ZNF217-overexpressing cultures. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Supplementary Figure 1. Data indicates Mean ± 1SD (n≥4). Significant differences between ZNF217-expressing cells and control at each time-point were analysed by paired t-test; * denotes p<0.05.
Supplementary Figure 32 – ZNF217 overexpression promotes monocytic differentiation

Summary dot plots showing (A) CD34+ expression in terms of percentage in monocytic cells (CD13+ CD36+) over time for control and ZNF217-overexpressing cells. Summary data showing (B) CD11b and (C) CD14 expression in terms of MFI in monocytic cells over time for control and ZNF217-overexpressing cells. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Supplementary Figure 1. Data indicates Mean ± 1SD (n≥3). S Significant differences between ZNF217-expressing cells and control at each time-point were analysed by paired t-test; * denotes p<0.05, ** denotes p<0.01, *** denotes p<0.001.
Supplementary Figure 33 – ZNF217 overexpression promotes granulocytic differentiation by upregulating monocytic markers

Summary dot plots showing (A) CD34+ expression in granulocytic cells (CD13+/CD36-). Summary data showing (B) CD11b and (C) CD15 expression in terms of MFI in granulocytic cells over time for control and ZNF217-overexpressing cells. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Supplementary Figure 1. Data indicates Mean ± 1SD (n≥3). Significant differences between ZNF217-expressing cells and control at each time-point were analysed by paired t-test; * denotes p<0.05, ** denotes p<0.01.
Supplementary Figure 34 – Knockdown of ZNF217 promotes monocytic cell growth in myeloid development

Summary dot plot showing the (A) cumulative fold-expansion of control and three ZNF217-KD constructs in terms of GFP positivity in culture medium containing IL-3, SCF, G- and GM-CSF, grown over 13 days. (B) Cumulative fold-expansion of monocytic cells (CD13+CD36+) transduced with a control or ZNF217-KD vectors. (C) Cumulative fold-expansion of granulocytic cells (CD13+/CD36−) transduced with a control or ZNF217-KD vector. (D) Cumulative fold-expansion of erythroid cells (CD13−CD36+) transduced with a control of ZNF217-KD vector. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Supplementary Figure 1. Data indicates Mean ± 1SD (n=2) (no statistical test was employed).
Supplementary Figure 35 – ZNF217 knockdown disrupts the balance between the monocytic and granulocytic populations in CD34+ HSPC during myeloid cell development

Summary dot plot showing percentage of (A) monocytic (CD13+ CD36−), (B) granulocytic (CD13+/− CD36) and (C) erythroid cells (CD13− CD36+) in control and three ZNF217-KD cultures. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Supplementary Figure 1. Data indicates Mean ± 1SD (n≥4). Significant differences between shZNF217 cultures and control at each time-point were analysed by one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01.
Supplementary Figure 36 – Knockdown of ZNF217 does not significantly alter monocytic development

Summary dot plots showing (A) CD34⁺ expression in terms of percentage in monocytic cells (CD13⁺ CD36⁺) over time for control and ZNF217-KD cells. Summary data showing (B) CD11b and (C) CD14 expression in terms of MFI in monocytic cells over time for control and ZNF217-KD cells. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Supplementary Figure 1. Data indicates Mean ± 1SD (n=4). Significant differences between shZNF217 cultures and control at each time-point were analysed by one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01.
Supplementary Figure 37 – Knockdown of ZNF217 does not significantly alter granulocytic development

Summary dot plots showing (A) CD34+ expression in terms of percentage in granulocytic cells (CD13+/ CD36-) over time for control and ZNF217-KD cells. Summary data showing (B) CD11b and (C) CD15 expression in terms of MFI in granulocytic cells over time for control and ZNF217-KD cells. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Supplementary Figure 1. Data indicates Mean ± 1SD (n≥4). Significant differences between shZNF217 cultures and control at each time-point were analysed by one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01.
Supplementary Figure 38 – KD of C/EBPβ decreases granulocytic growth in myeloid development

Summary dot plots showing the (A) cumulative fold-expansion of control and three shC/EBPβ constructs for GFP positive cells in culture medium containing IL-3, SCF, G- and GM-CSF, grown over 13 days. (B) Cumulative fold-expansion of monocytic cells (CD13+ CD36+) transduced with a control or shC/EBPβ vectors. (C) Cumulative fold-expansion of granulocytic cells (CD13+/- CD36-) transduced with a control or shC/EBPβ vector. (D) Cumulative fold-expansion of erythroid cells (CD13- CD36+) transduced with a control of shC/EBPβ vector. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Supplementary Figure 1. Data indicates mean ± 1SD (n=4). Significant differences between shC/EBPβ cultures and control at each time-point were analysed by one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01; *** denotes p<0.001; **** denotes p<0.0001.
Summary dot plots showing percentage of (A) monocytic (CD13^+ CD36^+), (B) granulocytic (CD13^{+-/-} CD36) and (C) erythroid cells (CD13^- CD36^+) in control and three shC/EBPβ cultures. Cells were reseeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Supplementary Figure 1. Data indicates mean ± 1SD (n=4). Significant differences between shC/EBPβ cultures and control at each time-point were analysed by one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01; *** denotes p<0.001.

Supplementary Figure 39 – C/EBPβ KD disrupts the balance between the monocytic and granulocytic populations in CD34^+ HSPC during myeloid cell development
Supplementary Figure 40 – KD of C/EBPβ has little impact on normal monocytic differentiation

Summary dot plots showing (A) CD34+ expression in terms of percentage in monocytic cells (CD13+ CD36+) over time for control and shC/EBPβ cells. Summary data showing (B) CD11b and (C) CD14 expression in terms of MFI in monocytic cells over time for control and shC/EBPβ cells. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Supplementary Figure 1. Data indicates mean ± 1SD (n=4). Significant differences between shC/EBPβ cultures and control at each time-point were analysed by one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01.
Supplementary Figure 41 – KD of C/EBPβ promotes granulocytic development

**Summary dot plots showing (A)** CD34⁺ expression in terms of percentage in granulocytic cells (CD13⁺/CD36⁻) over time for control and shC/EBPβ cells. Summary data showing (B) CD11b and (C) CD15 expression in terms of MFI in granulocytic cells over time for control and shC/EBPβ cells. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in **Supplementary Figure 1**. Data indicates mean ± 1SD (n=4). Significant differences between shC/EBPβ cultures and control at each time-point was analysed by one-way ANOVA with Bonferroni’s multiple comparisons correction; * denotes p<0.05; ** denotes p<0.01; *** denotes p<0.001.
Overexpression of C/EBPβ promotes monocytic and granulocytic growth in myeloid development

Summary dot plots showing the (A) cumulative fold-expansion of control and C/EBPβ-overexpression constructs in terms of GFP positivity in culture medium containing IL-3, SCF, G- and GM-CSF, grown over 13 days. (B) Cumulative fold-expansion of monocytic cells (CD13+ CD36+) transduced with a control or C/EBPβ-OE vector. (C) Cumulative fold-expansion of granulocytic cells (CD13+/ CD36-) transduced with a control or C/EBPβ-OE vector. (D) Cumulative fold-expansion of erythroid cells (CD13- CD36+) transduced with a control of C/EBPβ-OE vector. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Supplementary Figure 1. Data indicates mean ± 1SD (n=4). Significant differences between C/EBPβ-overexpression and control cultures at each time-point were analysed by paired T-test; * denotes p<0.05; ** denotes p<0.01.

**Supplementary Figure 42 – Overexpression of C/EBPβ promotes monocytic and granulocytic growth in myeloid development**
Supplementary Figure 43 – C/EBPβ overexpression disrupts the balance between myeloid population during haematopoietic development

Summary dot plots showing the proportion of (A) monocytic (CD13+ CD36+), (B) granulocytic (CD13+/- CD36) and (C) erythroid cells (CD13- CD36+) in control and C/EBPβ-overexpressing cultures. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Supplementary Figure 1. Data indicates mean ± 1SD (n=4). Significant differences between C/EBPβ-expressing cells and control at each time-point were analysed by paired t-test; * denotes p<0.05, ** denotes p<0.01, *** denotes p<0.001.
Supplementary Figure 44 – C/EBPβ overexpression promotes the upregulation of monocytic markers in monocytic progenitors

Summary dot plots showing (A) CD34+ expression in terms of percentage in monocytic cells (CD13+ CD36+) over time for control and C/EBPβ-overexpressing cells. Summary data showing (B) CD11b and (C) CD14 expression in terms of MFI in monocytic cells over time for control and C/EBPβ-overexpressing cells. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Supplementary Figure 1. Data indicates mean ± 1SD (n=4). Significant differences between C/EBPβ-expressing cells and control at each time-point were analysed by paired t-test; * denotes p<0.05, ** denotes p<0.01.
Supplementary Figure 45 – C/EBPβ overexpression promotes the upregulation of granulocytic markers in granulocytic progenitors

Summary dot plots showing (A) CD34+ expression in terms of percentage in granulocytic cells (CD13+/CD36−) over time for control and C/EBPβ-overexpressing cells. Summary data showing (B) CD11b and (C) CD15 expression in terms of MFI in granulocytic cells over time for control and C/EBPβ-overexpressing cells. Cells were re-seeded at the appropriate density following each time-point, as described in 2.3.3. Gating strategies were applied as described in Supplementary Figure 1. Data indicates mean ± 1SD (n=4). Significant differences between C/EBPβ-expressing cells and control at each time-point were analysed by paired t-test; ** denotes p<0.01.
**Supplementary Figure 46 – C/EBPβ KD in the Kasumi-1 cell line**

Growth of Kasumi-1 cells with three shC/EBPβ constructs compared to control, over 10 days of culture. (n=1). Cells were selected for GFP positivity on day 2 of culture, using the puromycin selectable marker.
References


Buck, M., Turler, H. & Chojkier, M. 1994. LAP (NF-IL-6), a tissue-specific transcriptional activator, is an inhibitor of hepatoma cell proliferation. *EMBO J*, 13, 851-60.


Christiansen, D. H., Andersen, M. K. & Pedersen-Bjergaard, J. 2004. Mutations of AML1 are common in therapy-related myelodysplasia following therapy with alkylating agents and are significantly associated with deletion or loss of chromosome arm 7q and with subsequent leukemic transformation. *Blood*, 104, 1474-81.


References


Hiebert, S. W., Lutterbach, B. & Amann, J. 2001. Role of co-repressors in transcriptional repression mediated by the t(8;21), t(16;21), t(12;21), and inv(16) fusion proteins. Curr Opin Hematol, 8, 197-200.


References


References


References


References


References


Yamada, O. & Kawauchi, K. 2013. The role of the JAK-STAT pathway and related signal cascades in telomerase activation during the development of hematologic malignancies. JAKSTAT, 2, e25256.


