Investigating the Mechanism of Bone Marrow Failure Observed in Patients with Acute Myeloid Leukaemia

Katharine Ailsa Hodby

Submitted in partial fulfilment of the requirements of the Degree of Doctor of Philosophy

August 2017

Statement of Originality

I, Katharine Ailsa Hodby, confirm that the research included within this thesis is my own, or that where it has been carried out in collaboration with, or supported by others, that this is duly acknowledged below and my contribution indicated. Previously published material is also acknowledged below.

I attest that I have exercised reasonable care to ensure the work is original, and does not to the best of my knowledge break any UK law, infringe any third party's copyright or other Intellectual Property Right, or contain any confidential material.

I accept the College has the right to use plagiarism detection software to check the electronic version of the thesis.

I confirm that this thesis has not been previously submitted for the award of a degree by this or any other university.

Signature:

Date:

Details of Collaboration

I performed all of the experimental work described in this thesis, with the following exceptions which are listed below.

General

Collection of AML patient samples and controls from Bart's Hospital was done by Sameena Iqbal and the Tissue Bank Team. Any samples which were stored prior to use were processed as described in Section 2.2.2 by members of the Tissue Bank Team.

FISH based staining and analysis was performed by Marianne Grantham, Cytogenetics Department, Royal London Hospital.

Whilst all sample preparation was performed by me, sequencing of DNA for the *NPM1* exon 12 region was performed by the Sequencing Laboratory, London Research Institute.

Chapter 3: Study of patients with AML who present without trilineage haematopoietic failure

A database of clinical characteristics and blood counts for all patients presenting to Bart's Hospital between 1997 and 2012 with AML was constructed by Janet Matthews, Data Collection Officer, Bart's Cancer Institute. The data were retrieved from the Bart's Cancer Clinical Database.

Trephine preparation and H&E staining was done by Andrew Clear, Postdoctoral Research Assistant, Bart's Cancer Institute.

Chapter 5: Enumerating Stem Cells in Primary Diagnostic Bone Marrow samples from Patients Presenting with Acute Myeloid Leukaemia

Immediate analysis of fresh AML and control bone marrow samples before May 2012 was performed by Dr F. Miraki-Moud.

Chapter 7: Using RNA-Seq to compare the transcriptomes of normal HSCs from AML and control bone marrow samples

Quality control of RNA samples, cDNA and prepared DNA libraries using the relevant Tape Station kit were performed by the staff of the Genome Centre, BCI.

Library preparation, quantification and pooling were performed by me. Sequencing of the prepared libraries was performed at the High Throughput Genome Centre, Oxford.

Bioinformatics Analysis of RNA-Seq Data was performed by Ai Nagano, Bioinformatics Group, BCI. This included generation of the Principal Component Analysis plot shown in Figure 7-10.

Publications

Miraki-Moud, F; Anjos-Afonso, F; Hodby, KA; Griessinger, E; Rosignoli, G; Lillington, D; Jia, L; Davies, JK; Cavenagh, J; Smith, M; Oakervee, H; Agrawal, S; Gribben, JG; Bonnet, D and Taussig, DC (2013). Acute myeloid leukemia does not deplete normal hematopoietic stem cells but induces cytopenias by impeding their differentiation. *Proc Natl Acad Sci U S A* **110**, 13576-13581.

Miraki-Moud F, Ghazaly E, Ariza-McNaughton L, Hodby KA, Clear A, Anjos-Afonso F, Liapis K, Grantham M, Sohrabi F, Cavenagh J, Bomalaski JS, Gribben JG, Szlosarek PW, Bonnet D, Taussig DC (2015). Arginine deprivation using pegylated arginine deiminase has activity against primary acute myeloid leukemia cells in vivo. *Blood* **125**, 4060-4068.

Meeting Abstracts

Farideh Miraki-Moud, PhD, Linda Ariza-McNaughton, MSc, Essam A Ghazaly, MD, Katharine A Hodby, Phuong Luong, Andrew Clear, PhD, Fareeda Sohrabi, John Bomalaski, John Gribben, MD, DSc, Dominique Bonnet, PhD, Peter Szlosarek and David C Taussig, PhD, MRCP, FRCPath (2013). Arginine Deprivation With Pegylated Arginine Deiminase Induces Death Of Acute Myeloid Leukaemia Cells In Vivo. Abstract 1458, ASH New Orleans, 2013.

Abstract

Patients with Acute Myeloid Leukaemia (AML) present with the signs and symptoms of bone marrow failure. This finding spans the genetic and phenotypic diversity of the disease. The mechanism which underlies it is poorly understood.

This thesis explores the effect of AML on the normal haematopoietic stem cell (HSC) population, using primary human diagnostic bone marrow samples. Previous work from our group suggested that AML induces a state of quiescence in HSCs, producing a differentiation block responsible for the observed cytopenias[1]. Reversal of this process might offer an alternative to the current treatment of patients with palliative transfusions.

I have developed a flow cytometry-based technique to differentiate normal HSCs from leukaemia cells, selecting cells with the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ expression signature. Validation of this technique by assessment of sorted cells by FISH and PCR, suggests it is successful in 73% of AML samples. In a further 25% of samples, it selects for a population significantly enriched for normal HSCs.

We used this panel to investigate the concentration of HSCs at AML diagnosis, compared to controls. We show that there is no significant difference between HSC concentration at AML diagnosis (n=38, median [HSC] 2.5 cells/µl) and controls (n=24, median [HSC] 2.4 cells/µl). HSC concentration was not significantly affected by AML karyotype, patient age or gender.

However, those patients presenting with a low HSC concentration at diagnosis (<0.1 HSC/µl) were found to have a significantly worse outcome both in terms of overall and relapse-free survival, an effect apparently independent of age, gender and underlying karyotype. HSC concentration at diagnosis with AML may therefore represent a new independent prognostic marker.

We then studied CD33 expression patterns on HSCs within Core Binding Factor mutated AML (n=37) at diagnosis, and found its expression to be significantly lower than on HSCs within controls (n=9) (17% versus 58%, p=0.005). CD33 expression on HSCs from AML samples rose significantly from diagnosis to remission (n=16) (17% to 58%, p=0.0001). This mirrors previous findings from our group using $CD34^{low}$ AML samples, and is, we believe, the first time that the antigenic signature of normal HSCs has been shown to be modified

by the presence of AML. However, an in vitro assay to test the significance of these changes in terms of the cytotoxicity of GO towards normal HSCs did not demonstrate a significant difference between HSC subgroups.

Finally, we attempted to investigate the mechanism by which AML might induce HSC quiescence by studying the comparative transcriptomes of HSCs from $CD34^{low}$ AML (n=6) and controls (n=6) by RNA-Seq, using direct cell to cDNA synthesis, followed by amplification. A first attempt resulted in poor quality data, with a significant proportion of reads mapping to non-coding DNA regions. A repeat approach, using utilising immediate RNA extraction post sorting resulted in significantly better quality data Bioinformatics analysis revealed differential expression of 6 genes between the 2 datasets (*GNPDA1*, *ADGRG3*, *MIAT*, *WDR31*, *RP11-244H3.1* and *RXFP1*). GO enrichment studies using David highlighted a number of pathways including the TNF signalling pathway (p=0.003; after Benjamini-Hochberg correction p=0.51). Validation of these findings by independent qPCR, and functional exploration of enriched signalling pathways remains outstanding.

Acknowledgements

The work described in this thesis was undertaken whilst I was a Clinical Research Fellow at the Centre for Haemato-Oncology at Bart's Cancer Institute. Funding for the three year PhD program was undertaken by Cancer Research UK.

I would like to thank David Taussig for providing the framework for this project, submitting the initial idea to the CRUK funding body and for his support and guidance over the first two years of my PhD. I would also like to thank Gabriella Ficz for agreeing to rescue me for the last year of the project, and for her help and guidance in steering the project to exciting new areas. I would also like to thank John Gribben for his wise oversight, and organisational support.

A massive debt of thanks is owed to Farideh Miraki-Moud, for teaching me all of the laboratory skills I know, and for her endless patience and generosity of advice when things went wrong in the lab. In particular, she developed many of the protocols adapted for handling small numbers of cells: without her assistance, I doubt there would be any results at all!

To Rebecca Pike and the Flow Laboratory team for their assistance and good humour with endless, endless sorting sessions; to Sameena Iqbal and the Tissue Bank team for their collection of samples, and my endless requests for vials; and to Janet Matthews in the Clinical Database team for her assistance in supplying patient information. Additional thanks to Emily Sanderson, Robert Petty, James Heward and Jeff Davies for providing me crash courses in Statistics and Genetics in the last year.

To CRUK and QMUL for their unquestioning support during my maternity leave, and for easing my transition back to the world of work after 12 months of changing nappies.

Finally my thanks to all of the patients and healthy donors who donated the tissue required for this project, and the clinical staff involved in collection and consent procedures.

Table of Contents

Statement of Originality **1**

Details of Collaboration **2**

General *2*

Chapter 3: Study of patients with AML who present without trilineage haematopoietic failure *2*

Chapter 5: Enumerating Stem Cells in Primary Diagnostic Bone Marrow samples from Patients Presenting with Acute Myeloid Leukaemia *2*

Chapter 7: Using RNA-Seq to compare the transcriptomes of normal HSCs from AML and control bone marrow samples *3*

Publications **4**

Meeting Abstracts **4**

Abstract **5**

Acknowledgements **7**

Table of Contents **8**

Table of Figures **16**

Table of Tables **20**

Chapter 1 Introduction **24**

1.1 Background *24*

1.2 Normal Haematopoiesis *24*

1.2.1 Introduction 24

1.2.2 Site of Haematopoiesis 25

1.2.3 The hierarchy of haematopoietic development 25

1.3 The Haematopoietic Stem Cell *27*

1.3.1 The HSC concept 27

1.3.2 Physical location of HSCs 27

1.3.3 Regulation of Stem Cell Behaviour 29

1.3.4 Stem Cell Mobilisation and Homing 31

1.3.5 Methodology for studying stem cells 33

1.4 Acute Myeloid Leukaemia *35*

1.4.1 Introduction 35

1.4.2 Incidence 36

1.4.3 Predisposing Factors to Disease Development 36

1.4.4 Clinical Presentation 36

1.4.5 Diagnosis 37

1.4.6 Classification Systems for AML 41

1.4.7 Current Therapy and Clinical Trials 41

1.4.8 Prognosis 42

1.4.9 The Leukaemia Stem Cell Model 43

1.5 Why does the development of AML lead to a failure of normal haematopoiesis? *44*

1.5.1 Introduction 44

1.5.2 The interaction between HSCs and AML: summary of previous work by the group which led to the hypothesis that acute myeloid leukemia does not deplete normal haematopoietic stem cells but induces cytopenias by impeding their differentiation 45

1.5.3 The interaction between AML and stromal cells may affect the niche and therefore HSC development indirectly 49

1.5.4 Caveat: How normal are the "normal" HSCs in acute myeloid leukaemia samples? The pre-leukaemic stem cell concept 51

Chapter 2 Materials and Methods **53**

2.1 Samples and Ethics *53*

2.2 Sample Handling Techniques *53*

2.2.1 Collection of Fresh Bone Marrow Aliquots for Immediate Analysis 53

2.2.2 Tissue Bank Processing of Samples for Delayed Analysis 54

2.2.3 Processing of G-CSF mobilised peripheral blood and cord blood 54

2.2.4 Tissue Bank storage of bone marrow serum samples 55

2.2.5 Cell Thawing 55

2.3 Cell Counting *56*

2.3.1 Automated Assessment of Cell Numbers 56

2.3.2 Manual Counting of Cell Numbers 56

2.4 Flow Cytometry *56*

2.4.1 Background 56

2.4.2 Fluorescence Activated Cell Sorting (FACS) 57

2.4.3 Specific Uses of Flow Cytometry in this Thesis 58

2.4.4 Fluorochrome Panel Selection 58

2.4.5 Specific Techniques 58

2.5 Tissue Culture Methods *61*
2.5.1 MS5 Culture 61
2.6 In vitro assays of stem cell function *62*
2.6.1 Methylcellulose Assay (for the presence of CFU) 62
2.6.2 Long Term Culture Initiating Cell Assay 63
2.7 Histological techniques *64*
2.7.1 Preparation of Slides 64
2.7.2 Immunocytochemistry and Microscopy 64
2.7.3 FISH Staining 65
2.8 Genetic Techniques *67*
2.8.1 DNA Extraction modified for small cell numbers 67
2.8.2 RNA extraction modified for small cell numbers 67
2.8.3 DNA sequencing (for the *NPM1* gene) 68
2.8.4 Quantitative Real Time Polymerase Chain Reaction (qPCR) 70

Chapter 3 Study of patients with AML who present without trilineage haematopoietic failure 75

3.1 Introduction *75*
3.1.1 Mechanisms of presentation with AML 75
3.1.2 Summary of previous work suggesting bone marrow failure in AML is due to induced HSC quiescence 75
3.1.3 Normal erythroid development 79
3.1.4 Normal platelet development pathway 80
3.2 Aims and Objectives *82*
3.3 Specific Methods *83*
3.3.1 Database analysis 83
3.3.2 Patients and samples 84
3.3.3 Flow Cytometry 84
3.3.4 Morphological Techniques 87
3.3.5 Genetic Analysis of samples for clonality identification 87
3.3.6 Megakaryocyte Enumeration by Trephine Analysis 87
3.4 Results *89*
3.4.1 Database analysis results 89
3.4.2 Erythroid precursor lineage identification and quantification 98
3.4.3 Attempts to study megakaryocyte lineage and concentration at diagnosis with AML 103
3.5 Discussion *108*
3.5.1 Summary of database analysis 108

3.5.2 Clonality studies of erythroid precursors suggests red cells are derived from normal precursors and not from the leukaemic clone (in *NPM1* and CBF mutated AML samples), but were not possible with megakaryocytes 109
3.5.3 Studies of erythroid precursor levels 110
3.5.4 Studies of platelet precursor levels 111
3.5.5 Further experimental work 112
3.5.6 Summary 112

Chapter 4 Developing a methodology for the identification of normal stem cells within AML Samples 114

4.1 Introduction 114
4.1.1 The identification of normal stem cells and progenitors within $CD34^{low}$ expressing AML samples 114
4.1.2 Difficulties in identifying normal stem cells within AML samples 114
4.1.3 Antigenic markers reported to show differential expression between normal stem and leukaemic cells 115

4.2 Aims and Objectives 123

4.3 Chapter specific methods 124
4.3.1 Flow Based Analysis of Stem Cell Marker Expression 124
4.3.2 FACS Sorting 124
4.3.3 In vitro assays of stem cell function 126
4.3.4 Assays to test for the presence of leukaemia associated mutations 126
4.3.5 Clinical data 127
4.3.6 Statistical analysis 127

4.4 Results 128
4.4.1 Antigenic panel design for stem cell identification 128
4.4.2 Validation of the Stem Cell Panel: Early optimisation 137
4.4.3 Validation of the Stem Cell Panel: testing on AML bone marrow samples 142
4.4.3.1.1 Cytogenetic 143
4.4.4 Does the composition and behaviour of the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells within AML samples at diagnosis have clinical significance? 158

4.5 Discussion 166
4.5.1 General points 166
4.5.2 Further experimental applications of the panel 167
4.5.3 Why is there a difference in the panel's ability to separate from different cytogenetic risk groups? 167
4.5.4 What is the definition of a normal HSC? The issue of pre-leukaemic stem cells 168

4.5.5 Is there a clinical association between dirty and pure sorts? 169
4.5.6 What is the significance of the failure of cells to grow in culture? 169
4.5.7 Concerns over the validity of the experimental design: LTC appears to preferentially select the growth of normal haematopoietic stem cells over AML cells, and the strength of this selection may differ between samples 171
4.5.8 Further experiments 174

Chapter 5 Enumerating Stem Cells in Primary Diagnostic Bone Marrow samples from Patients Presenting with Acute Myeloid Leukaemia 176

5.1 Introduction 176
5.2 Aims and Objectives 177
5.3 Specific methods 178
5.3.1 Processing of Fresh Samples 178
5.3.2 Subsequent Stem Cell Identification of Stored, Frozen BM Samples 179
5.3.3 Clinical Outcome Data 180
5.3.4 Statistical analysis of data 180
5.4 Results 182
5.4.1 Samples processed for fresh BM analysis 182
5.4.2 Summary of treatment outcomes for AML patients 196
5.4.3 Comparison of normal stem cell concentrations between AML and control samples 197
5.4.4 Subanalysis of AML samples 198
5.4.5 Control Sample Subanalysis 206
5.5 Discussion 207
5.5.1 Potential Clinical Applications 207
5.5.2 Variation of stem cell concentrations 209
5.5.3 Issues specific to AML samples 215
5.5.4 Further experiments 216

Chapter 6 Exploring the effect of AML on CD33 expression on normal stem cells 219

6.1 Introduction 219
6.1.1 CD33 expression patterns 219
6.1.2 Anti CD33 therapy in AML 220
6.1.3 Data from our group showing CD33 expression patterns within the normal HSC population of $CD34^{low}$ AML samples 223
6.2 Aims and Objectives 225
6.3 Chapter Specific Methods 226

6.3.1 Patient Samples 226
6.3.2 CD33 Expression Analysis 226
6.3.3 GO Exposure Experiment 226

6.4 Results *229*
6.4.1 CD33 expression patterns within $CD34^{+}38^{-}ALDH^{high}$ stem cells of CBF AML samples 229
6.4.2 The effect of CD33 expression of HSCs on susceptibility to GO therapy 233

6.5 Discussion *237*
6.5.1 General points 237
6.5.2 Why might CD33 expression show variable patterns on normal HSCs in the context of AML? Is this a cause or an effect? 237
6.5.3 Does the differential expression of CD33 on HSCs have a clinical consequence? 238
6.5.4 Potential further experimental work 241

Chapter 7 Using RNA-Seq to compare the transcriptomes of normal HSCs from AML and control bone marrow samples 243

7.1 Introduction *243*
7.1.1 Difficulties with HSC work 243
7.1.2 Previous work assessing the HSC transcriptome in the context of AML development 244
7.1.3 RNA-Seq 251

7.2 Aims and Objectives *256*

7.3 Attempt 1: Library Preparation involving cDNA synthesis directly from cell lysate *257*
7.3.1 Specific Methods 257
7.3.2 Results 269

7.4 Attempt 2: Using RNA as input material *290*
7.4.1 Methods 291
7.4.1.3 cDNA synthesis using SMART-Seq 294
7.4.2 Results 296

7.5 Discussion *316*
7.5.1 Was failure of selective amplification of polyadenylated mRNA within the SmartSeq kit the cause of genomic contamination of the first experiment? 316
7.5.2 Reasons for failure of validation of selected genes of interest by flow based protein assessment and qPCR based analysis 318
7.5.3 Using RNA as a starting material leads to better quality data 320
7.5.4 Further experimental work 322

Chapter 8 Discussion 326

8.1 Summary of main findings *326*

8.2 Key areas for further experimental effort expanding on the work described within this thesis *327*

8.2.1 Does stem cell concentration at diagnosis have an independent prognostic impact? 328

8.2.2 Do poor risk karyotype AMLs exert a different effect on normal HSCs to good or intermediate disease samples, as suggested by their different behaviour in vitro? 328

8.2.3 What is the mechanism of BM suppression in AML? 329

8.3 New avenues for further research in the broader field of malignancy-associated bone marrow failure *333*

8.3.1 How does the mechanism of AML–associated bone marrow failure compare with that seen in other malignancies? 333

8.3.2 Does AML just affect HSCs, or other levels of normal haematopoiesis? 334

8.3.3 Is the effect of AML on normal HSCs the same irrespective of AML subtype? 335

8.3.4 Are the effects of AML limited to normal haematopoiesis precursors, or does it have the ability to modulate the micro environmental stromal environment? 335

8.3.5 What does "normal" mean in the context of HSCs anymore anyway? 335

8.4 Final comments *336*

Chapter 9 References **337**

Appendix 1: Stock Solutions **357**

Appendix 2: Methodology for the AML-Stem Cell Co-Culture Assays **359**

App 2.1.4 Processing of Cells prior to Analysis 360

App 2.2. The 7 week in vitro conditioned media assay *361*

App 2.2.1 Introduction 361

App 2.2.2 Plate set up 361

App 2.2.3 MS5 preparation 362

App 2.2.4 AML and GMPB cells plated for preparation of conditioned media 362

App 2.2.5 HSC sorting and plating 362

App 2.2.6 Plate Management weeks 2 to 5 363

Appendix 3: Genes with significantly altered expression between AML-exposed and Control HSCs, as detected by RNA-Seq Attempt 1 **364**

Appendix 4: Genes with significantly different expression between AML-exposed and Control HSCs, as detected by RNA-Seq Attempt 2 **377**

Appendix 5: Abbreviations **385**

Table of Figures

Chapter 1 Introduction

Figure 1-1 Key cytokines in haematopoietic development 26

Figure 1-2A: The effect of AML on mouse haematopoietic populations in a xenograft model; B: Classification of 8 NPM1 mutated AML samples based on CD34 and CD38 expression; C: Assessment of normal residual haematopoiesis in humans with $CD34^{low}$ AML; D: The percentage of HSPCs in cell cycle was lower in subtype A AMLs than controls 48

Chapter 2 Materials and Methods

Figure 2-1 Flow gating strategy used to select single, live cells 61

Chapter 3 Study of patients with AML who present without trilineage haematopoietic failure

Figure 3-1 Hypothesis of induced differentiation block in HSCs in AML 76

Figure 3-2 Pathway showing a proposed theory of haematopoietic suppression at multiple steps of the haematopoietic pathway by AML blasts 78

Figure 3-3 Summary of erythroid development 80

Figure 3-4 Flow gating strategy used to identify erythroid precursors from bone marrow samples .. 85

Figure 3-5 A: Histogram illustrating presentation Hb levels for 529 patients with AML; B: Histogram illustrating the relationship between presentation Hb level and age at diagnosis; C: Bar chart showing effect of 19 different AML cytogenetic abnormalities on Hb level at diagnosis; D: Kaplan-Meier curve showing the effect of gender-adjusted Hb at diagnosis with AML on overall survival 92

Figure 3-6 A: Histogram illustrating platelet counts for 529 patients at diagnosis with AML; B: Histogram showing the relationship between presentation platelet count and age at diagnosis; C: Bar chart showing effect of 19 different AML cytogenetic abnormalities on platelet level at diagnosis; D: Kaplan-Meier survival curve showing the effect of platelet count at diagnosis with AML on overall survival 96

Figure 3-7 showing morphological appearances of sorted $CD235a^{+}71^{+}3^{-}$ cells. A: cytospin stained with MGG; B: immunofluorescence staining for the presence of haemoglobin within $CD235a^{+}71^{+}3^{-}$ cells; C: $CD235a^{-}71^{-}3^{+}$ cells (negative control) 98

Figure 3-8 A: Bar-chart showing the relationship between HSC concentration & Hb at diagnosis; B: Bar chart showing the relationship between HSC and erythroblast concentrations at diagnosis in AML and control samples; C: Bar chart showing trephine megakaryocyte concentration at diagnosis with AML and controls; D: Bar chart showing the relationship between HSC concentration and platelet count at diagnosis 105

Chapter 4 Developing a methodology for the identification of normal stem cells within AML samples

Figure 4-1 Representative flow plots showing ALDH activity against side scatter for a normal donor and 3 AML samples. A: normal donor; B: a patient with t(8;21) AML; C: a patient with inv(16) AML and D: a patient with AML with normal cytogenetics. 116
Figure 4-2 Flow plots illustrating the three patterns of ALDH expression within AML samples 131
Figure 4-3 Flow plots illustrating typical patterns of CLL1 expression within a $CD34^{+}$ AML and a $CD34^{low}$ AML sample 132
Figure 4-4 Flow plots showing characteristic patterns of TIM3 expression. 133
Figure 4-5 Flow plots showing CD33 expression within an AML sample 133
Figure 4-6 Flow plots showing typical CD99 expression patterns within two AML samples 134
Figure 4-7 Flow plots showing typical CD300 expression 135
Figure 4-8 Flow plots showing typical CD47 expression pattern 135
Figure 4-9 Flow plots illustrating how combining staining with ALDH and CLL1 can further define a putative stem cell population in a $CD34^{+}$ AML 136
Figure 4-10 Typical Sanger Sequencing traces showing A: normal Exon 12; B: Type A mutation; C: Type D mutation- 144
Figure 4-11 Sorting FACS plots for 4 primary AML samples, illustrating key steps of consecutive gating strategy 148
Figure 4-12 FACS sorting plots illustrating the diverse appearances of 4 poor risk samples 149
Figure 4-13 FISH staining of 4 sorted samples. A: Pat with t(8;21) i: Colonies from $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells after LTC ii. Unsorted blasts; B: Pat with t(15;17) i: Colonies from $CD34^{+}CD38^{-}ALDH^{high}CLL1^{-}$ cells after LTC ii. Unsorted blasts; C: Pat with complex karyotype i: Colonies from $CD34^{+}CD38^{-}ALDH^{high}CLL1^{-}$ cells after LTC ii. Unsorted blasts; D: Pat with complex karyotype including del(7q) i. Colonies from $CD34^{+}CD38^{-}ALDH^{high}CLL1^{-}$ cells after LTC ii. Unsorted blasts 152
Figure 4-14 Results of PCR gel after digestion for TKD mutation 157
Figure 4-15 Kaplan-Meier curves showing the effects of A: growth versus no growth on overall survival; B: Mixed versus pure composition of the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ population on overall survival; C: Mixed versus pure composition of the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ population on relapse free survival; D: Mixed versus pure composition of $CD34^{+}38\text{-}ALDH^{high}CLL1^{-}$ population within NPM1 mutated samples only 165

Chapter 5 Enumerating Stem Cells in Primary Diagnostic Bone Marrow samples from Patients with Acute Myeloid Leukaemia

Figure 5-1 illustrating sequential gating strategy used to analyse fresh and frozen AML samples. .. 181
Figure 5-2 Kaplan-Meier plots showing the effects of A: cytogenetics on overall survival; B: age on overall survival; C: cytogenetics on relapse free survival; D: age on relapse free survival 199
Figure 5-3 A: showing calculated $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ concentrations in control samples; all AML samples; AML samples excluding 7 samples where the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ population was known to contain leukaemia cells, and AML samples where analysis has shown the $CD34^{+}38^{-}$

ALDH^{high}CLL1^{-} population to be free of leukaemia-associated mutations; B: Bar chart comparing stem cell concentrations of 38 AML samples between male and female patients; C: Bar chart showing the relationship between stem cell concentration of 38 AML samples and age; D: Bar chart showing the relationship between stem cell concentration of 38 AML samples and cytogenetic risk group 200

Figure 5-4 A: Kaplan-Meier Plot showing effect of HSC concentration at diagnosis on overall survival; B: Kaplan-Meier plot showing effect of HSC concentration on relapse free survival; C: Bar chart showing association between HSC number at diagnosis and days from induction start to neutrophil recovery; D: Bar chart showing association between HSC number at diagnosis and days from induction start to platelet recovery 205

Figure 5-5 Bar charts showing the effect on HSC concentration in control patients of A: gender and B: age 206

Figure 5-6 Kaplan-Meier plots showing the effect of stem cell concentration for those patients with intermediate risk cytogenetics on A: overall survival and B: relapse free survival 208

Chapter 6 Exploring the effect of AML on CD33 expression on normal stem cells

Figure 6-1 showing CD33 expression within the CD34^{+}38^{-} cells of CD34^{low} AML and controls 224

Figure 6-2 A: CD33 expression within the HSCs of CBF AML samples versus controls; B: CD33 expression within the HSCs of inv(16) AMLs, t(8;21) AMLs and controls; C: changing pattern of CD33 expression on normal HSCs between diagnosis and remission for 16 CBF AML patients; D: CD33 expression within the HSC compartment of 4 t(8;21) AML patients at diagnosis, first remission and relapse 232

Figure 6-3 Bar chart illustrating the Median Florescence Intensity of CD33 expression of unselected AML cells subfractions of the CD34^{+}38^{-}ALDH^{high} HSCs from AML samples, and the same subfractions of CD34^{+}38^{-}ALDH^{high} HSCs from normal BM 239

Chapter 7 Using RNA-Seq to compare the transcriptomes of normal HSCs from AML and control bone marrow samples

Figure 7-1 illustrating RhoA expression within the CD34^{+}38^{-} HSCs of AMLs and controls 250

Figure 7-2 The method of mRNA selection and cDNA synthesis involved with use of the Clontech SmartSeq kit 253

Figure 7-3 The process of DNA library production utilising the Nextera XT kit 254

Figure 7-4 HS 1000 DNA ScreenTape traces showing A: yield of cDNA from SmartSeq kit from 1000 cells when (i) immediate RNA extraction and subsequent processing through SmartSeq protocol (ii)cells lysed direct cDNA synthesis performed without prior RNA extraction; B: comparing the cDNA yields when i: 8 cycles ii: 9 cycles and iii: 10 cycles were used to amplify the cDNA resultant from the processing of 1000 cells 271

Figure 7-5 showing the effect of clean-up at different stages of library preparation. Four HS DNA ScreenTape 1000 traces are shown where A. cDNA trace after SmartSeq preparation showing a small peak of adapter present at 60bp; B: library appearance after Nextera XT prep with no

additional clean up C: library after 1:1 clean up D: library appearance if an additional 1:1 clean-up is performed on cDNA pre-library formation 272
Figure 7-6 Scatter plots illustrating results of validation experiments. Graphs A-C show results from flow-based protein expression assessment; Graphs D-F show results from PCR based gene expression assessment. A: CD33 expression on $CD34^{+}38^{-}$ HSCs from AML and control sample; B: CD85g (LILRA2) expression on $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ HSCs from AML and controls; C: CD266/TWEAKR expression on $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ HSCs from AML and controls; D: Relative expression of CD33 gene in $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ HSCs from AML and controls; E: Relative expression of U2AF1 gene in $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells from AML and controls; F: Relative expression of HIST1H2BC gene in $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells from AML and controls..... 289
Figure 7-7 Scatter plot comparing the RNA quantities extracted per cell for all 12 samples, expressed relative to lowest value in the group..... 300
Figure 7-8 Optimisation of the number of PCR cycles used for cDNA amplification after first strand synthesis..... 302
Figure 7-9 SmartSeq appearances of 2 test samples and positive and negative controls, all cycled at 18 cycles. A: successful trace; B: Possibly overcycled sample; C: Positive control; D: Negative control 303
Figure 7-10 Principal Component Analysis illustrating weak clustering of the 12 samples into two groups consistent with their origin 307
Figure 7-11 Gene enrichment analysis of differentially expressed genes from HSCs from AML samples and control, generated by the DAVID functional annotation chart option. Enriched categories of molecular function, cellular compartments, and biological processes are shown 309
Figure 7-12 KEGG generated figure showing key genes involved in TNF signalling 312
Figure 7-13 KEGG generated figure showing key genes involved in cell cycle control 312
Figure 7-14 KEGG generated figure showing the key genes involved in Notch signalling 323

Table of Tables

Chapter 1 Introduction

Table 1-1 Normal peripheral blood parameters for adult males and females 25

Table 1-2 summarising antigen expression patterns seen on human and murine HSCs 33

Table 1-3 The MRC 2010 Modified Cytogenetic classification of AML, based on the clinical outcome data from the MRC AML trials 38

Chapter 2 Materials and Methods

Table 2-1 FISH probes with expected signal patterns in normal and mutated cells 66

Table 2-2 showing the production of standards for NPM1 assessment assay 73

Chapter 3 Study of patients with AML who present without trilineage haematopoietic failure

Table 3-1 showing the relationship between normal platelet levels and haemoglobin at diagnosis .. 97

Table 3-2 Clinical and genetic characteristics of 10 patients with preserved haemoglobin at diagnosis which underwent sorting and clonality assessment of their erythroblast populations 99

Table 3-3 Percentage of NPM1 mutated cells within sorted erythroid, lymphoid and blast populations from NPM1 mutated AML samples 100

Table 3-4 FISH analysis of $CD235a^{+}CD71^{+}$ cells from AML samples with known karyotypic abnormalities 101

Table 3-5 illustrating clinical characteristics, diagnosis blood counts and calculated stem cell and erythroblast concentrations in 20 AML diagnosis samples 106

Table 3-6 showing clinical characteristics, diagnosis blood counts and calculated stem cell and erythroblast concentrations at staging in 7 control samples 107

Chapter 4 Developing a methodology for the identification of normal stem cells within AML samples

Table 4-1 AML sample details and summary of results of antigen-based separation of $CD34^{+}38^{-}$ populations 129

Table 4-2 Sample details for AMLs sorted for $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells and grown in 2 week methylcellulose culture 139

Table 4-3 Sample details for the samples for $CD34^{+}38^{-}ALDH^{high}CLL1^{low-}$ cells prior to 7 week LTC-IC assay 140

Table 4-4 LTC-IC attempt 3 results 141

Table 4-5 LTC-IC attempt 4 using new MS5 stocks 142

Table 4-6 AML samples sorted for $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations- sample information 146

Table 4-7 Summary of FACS sorting for $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations, LTC results and assessment of sorted populations for AML-associated mutations 150

Table 4-8 Analysis of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells and their progeny for AML-associated mutations by FISH 153

Table 4-9 illustrating the sorted cell numbers and results for samples undergoing purity analysis by qPCR for NPM1 mutation status. 155
Table 4-10 Cross tabulation for comparison of sample distribution within Growth versus No Growth of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells, and age categories 160
Table 4-11 Cross tabulation for comparison of sample distribution within Growth versus No Growth of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells, and cytogenetic risk categories 160
Table 4-12 Cross-tabulation showing comparison of sample distribution within Mixed versus Pure $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations and age categories 161
Table 4-13 Cross-tabulation showing comparison of sample distribution within Mixed versus Pure $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations and cytogenetic risk categories 162
Table 4-14 Cross-tabulation showing comparison of sample distribution within the Mixed versus Pure $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations and age categories 162
Table 4-15 Cross-tabulation showing comparison of sample distribution within the Mixed versus Pure $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations and cytogenetic risk categories 162
Table 4-16 Cross-tabulation showing samples distribution within Mixed versus Pure $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations and age categories for NPM1 mutated samples only 163
Chapter 5 Enumerating Stem cells in Primary Diagnostic Bone Marrow samples from Patients with Acute Myeloid Leukaemia
Table 5-1 Control bone marrows analysed fresh for [$CD34^{+}$ cell] between 2012 and 2015 183
Table 5-2 Newly diagnosed AML cases processed fresh for [$CD34^{+}$ cell] at diagnosis between 2012 and 2014 184
Table 5-3 Sample details of AML samples quantified for [$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$] cells at diagnosis by co-analysis of fresh and frozen samples 186
Table 5-4 Sample details of control bone marrows studied for [$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$] at diagnosis by co analysis of fresh and frozen samples 189
Table 5-5 Treatment details and clinical outcome data for 45 AML patients 190
Table 5-6 showing calculated $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ stem cell concentrations for the 45 AML samples enumerated 193
Table 5-7 showing calculated concentration of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ stem cells for 24 control samples 195
Table 5-8 Comparison of characteristics of the patients within the AML and control cohorts 196
Table 5-9 Cross tabulation for comparison of sample distribution within diagnostic HSC concentration and age categories 201
Table 5-10 Cross tabulation for comparison of sample distribution within diagnostic HSC concentration and Cytogenetic Risk categories 202
Table 5-11 Cross tabulation for comparison of sample distribution between HSC concentration and age categories 203

Table 5-12 Cross tabulation for comparison of sample distribution between HSC concentration and cytogenetic risk categories 203
Table 5-13 Comparison of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ concentration at diagnosis, and apparently functional stem cell concentration by 7 week culture within the same $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ subpopulation, of 4 control samples 215
Table 5-14 illustrates the variability in technical replicates of previously frozen samples used for $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cell quantification and the effect this has on quantification of the $[CD34^{+}38^{-}ALDH^{high}CLL1^{-}]$ at diagnosis 217
Chapter 6 Exploring the effect of AML on CD33 expression on normal stem cells
Table 6-1 Clinical characteristics and sample availability of CBF AMLs 231
Table 6-2 Details of AML ($CD34^{low}$ and CBF) and control samples sorted for stem cell populations and subsequently exposed to GO 235
Table 6-3 illustrating results for all samples from ELDA Hall analysis of stem cell frequency and pairwise analysis of results 236
Chapter 7 Using RNA-Seq to compare transcriptomes of normal HSCs from AML and control bone marrow samples
Table 7-1 Primer Pair sequences for SYBR Green qPCR 267
Table 7-2 Clinical information for 6 $CD34^{low}$ AML and 6 controls which underwent sorting for HSCs prior to RNA-Seq based transcriptome profiling 273
Table 7-3 Qubit concentrations of cDNA library pre and post 1:1 cleaning 274
Table 7-4 showing Library concentrations as estimated by Qubit 275
Table 7-5 Library Concentrations as estimated by qPCR 275
Table 7-6 Volumes used in final library pooling to ensure equal representation of all 12 libraries in sequencing 276
Table 7-7 Number of reads mapped to coding and non-coding regions for all 12 samples 277
Table 7-8 Genes selected for Validation by qPCR and flow-based assessment 280
Table 7-9 showing sample details for the 6 AML samples and 6 controls used in this experiment for sorting of HSC populations and RNA extraction 287
Table 7-10 Sample information for the 6 $CD34^{low}$ AML and 6 control samples used for RNA extraction in the repeat RNA-Seq experiment 298
Table 7-11 showing sorted number of cells, relative quantification of RNA by qPCR and volumes added for each sample to SmartSeq experiment 299
Table 7-12 Concentrations of cDNA libraries post SmartSeq preparation, estimated by Qubit 304
Table 7-13 Concentrations of final DNA libraries post Nextera XT processing, estimated by Qubit . 304
Table 7-14 Concentrations of final DNA libraries post Nextera XT processing, estimated by Kappa qPCR 305
Table 7-15 Volumes used in final DNA library pooling to ensure equal representation of all 12 libraries in sequencing. Relative concentrations were determined by Kappa qPCR results 305

Table 7-16 Percentage of RNA-Seq reads mapping to coding and non-coding regions.....306
Table 7-17 Enrichment of genes within specific biochemical pathways differentially activated between HSCs from AML and control samples, as generated by KEGG analysis 310
Table 7-18 illustrating 13 genes which showed fold changes >2 between HSCs from AML and controls samples in BOTH RNA-Seq experiments.....315
Appendix 3 Genes with significantly altered expression between AML-exposed and Control HSCs, as detected by RNA-Seq Attempt 1
Table App 3-1 Genes upregulated in HSCs from Controls versus AML 364
Table App 3-2 Genes upregulated in HSCs from AML samples versus controls.....371
Appendix 3 Genes with significantly altered expression between AML-exposed and Control HSCs, as detected by RNA-Seq Attempt 1
Table App 4 -1 Genes upregulated in HSCs from Controls versus AML 377
Table App 4-2 Genes upregulated in HSCs from AML samples versus controls.....381

Chapter 1 Introduction

1.1 Background

Normal haematopoiesis is the process by which the blood cells needed for survival are made in the bone marrow. In many haematological disorders, but AML in particular, the process of normal haematopoiesis is disrupted. The mechanism by which this occurs is poorly understood. Patients presenting with AML normally present with the problems resultant from a failure of normal blood production: overwhelming sepsis, bleeding or anaemia-driven fatigue and shortness of breath.

Visual inspection at the time of diagnosis normally shows the healthy bone marrow architecture to be completely effaced with a mass of malignant blast cells. However, that the capacity for normal haematopoiesis persists is shown by the fact that the majority of patients, once remission is obtained, regain normal blood counts three to four weeks after the completion of chemotherapy.

With some notable exceptions (acute promyelocytic leukaemia or APML providing the prime example[2]), survival rates for AML have shown only modest improvements over the last few decades in younger patients[3], with little change in the figures in patients presenting over the age of 60[4]. Much of the improvement noted is credited to the improvements seen in supportive care given to patients undergoing intensive chemotherapy[5]. The backbone of treatment for the majority of patients remains similar to that used in the 1970s. There is a need for new approaches (and effective new drug therapies) to improve prognosis.

If the mechanism by which haematopoietic failure in AML is better understood, this might give a new therapeutic approach to treating patient with indolent disease, with the aim of reducing their transfusion requirements and improving their quality of life.

1.2 Normal Haematopoiesis

1.2.1 Introduction

In the human adult, normal haematopoiesis occurs in the bone marrow in a tightly controlled homeostatic process, derived to maintain normal levels of blood cells for function. Steady state requires the production of approximately 10^{11} to 10^{12} new cells daily. Normal cell numbers found in the peripheral blood in health are given in Table 1-1.

Table 1-1 Normal peripheral blood parameters for adult males and females. Modified from Bain, Blood Cells (2008)[6]

PARAMETER	MALE	FEMALE
HAEMOGLOBIN (g/dL)	13.5-18.0	11.5-16.0
PLATELETS ($x10^9$/L)	150-400	150-400
WBC ($x10^9$/L)	4.00-11.0	4.0-11.0
NEUTROPHILS ($x10^9$/L)	2.0-7.5	2.0-7.5
LYMPHOCYTES ($x10^9$/L)	1.0-4.5	1.0-4.5
MONOCYTES ($x10^9$/L)	0.2-0.8	0.2-0.8
EOSINOPHILS ($x10^9$/L)	0.04-0.4	0.04-0.4
BASOPHILS ($x10^9$/L)	<0.1	<0.1

1.2.2 Site of Haematopoiesis

In embryonic development, primitive blood formation occurs in aggregates of blood cells in the yolk sac, called blood islands[7]. As development progresses, haematopoiesis migrates to the spleen, liver and lymph nodes. As bone marrow forms, it eventually assumes the task of forming most of the blood cells for the entire organism. However, maturation, activation, and some proliferation of lymphoid cells occur in secondary lymphoid organs (spleen, thymus, and lymph nodes)[8]. In children, haematopoiesis occurs in the marrow of the long bones such as the femur and tibia, but by adulthood, the areas responsible for active haematopoiesis have shrunk to comprise mainly the pelvis, cranium, vertebrae, and sternum[9]. In some disease states, the liver, thymus, and spleen may resume their haematopoietic function (extramedullary haematopoiesis), causing these organs to increase in size substantially[10]. Occasionally, for example in patients with thalassemia major and intermedia, haematopoietic development may occur even in areas where the resultant tissue disrupts the surrounding organs[11,12].

1.2.3 The hierarchy of haematopoietic development

The current paradigm for haematopoietic development describes a pluripotent haematopoietic stem cell (HSC), maintained in small numbers, capable of differentiating into all of the different mature blood cell lineages required for normal life. Estimates of murine HSC frequency have been generated by testing the ability of limiting dilutions of murine bone marrow cells to rescue haematopoiesis in lethally irradiated murine recipients. This work suggests the frequency of long term repopulating stem cells in mice is around 1 in 10,000[13,14]. Anecdotally, human HSC frequency is thought to be much lower, but is probably underestimated by xenograft transplant models (D Taussig, personal communication). Haematopoietic stem cells are thought to reside within a hypoxic marrow niche, maintained in an undifferentiated and largely quiescent state, partly due to interactions with the surrounding

stromal cells[15]. However, HSCs retain the ability to leave their bone marrow niche and mobilise via peripheral blood, as part of normal homeostasis. These cells maintain their levels in bone marrow by largely asymmetric divisions[16]. The pool of stem cells is heterogeneous, and is comprised of long term cells which multiply infrequently, and short term renewing cells, which replicate much more frequently[17]. Their daughter progeny begin a pathway of differentiation to myeloid-erythroid and lymphoid progenitors (the so called Common Myeloid Erythroid Progenitor (CMP)[18] and Common Lymphoid Progenitor (CLP)[19] subtypes), but begin to lose the ability to self-renew[20-22]. The influence of further cytokines and growth factors in the bone marrow are thought to lead to further cell differentiation, as shown in Figure 1-1

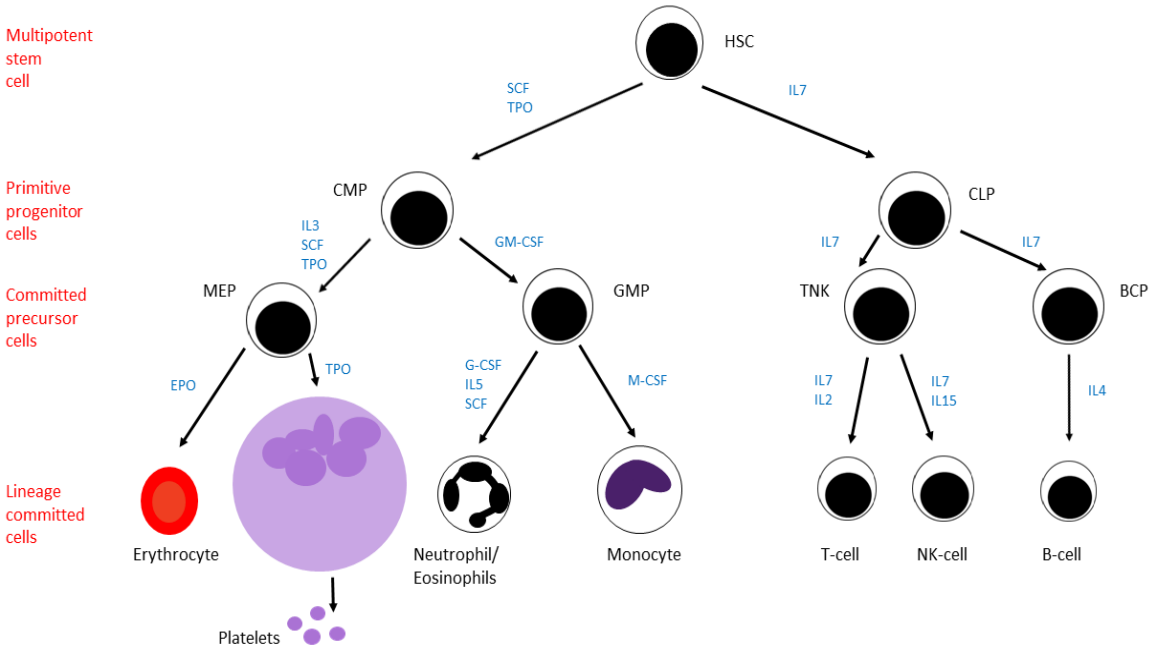

Figure 1-1 Key cytokines in haematopoietic development. From Robb, Oncogene, 2008[23]

These different commitment stages can be identified genetically by sequential expression of different transcription factors, and phenotypically by changes in surface antigen expression. However, the idea of step-wise progression towards fate determination has recently been challenges by advances in the world of single cell transcriptomics. Recent papers, involving massively parallel sequencing of bone marrow progenitor cells, suggests commitment to fate-determining transcriptional programs occurs at the very earliest stages of differentiation[24,25].

During haematopoietic development, precursor cells migrate from the subendosteal region of bone marrow, towards a more central location. The resultant mature blood cells exit the bone marrow environment through a dense network of vascular sinuses. The average cellular

lifespan varies significantly between cell types from five to ten days (platelets[26] and neutrophils[27]), to three months (erythrocytes[28]) to many years (memory lymphocytes[29]). Senescent blood cells are removed from the circulating blood volume as they pass through the spleen.

1.3 The Haematopoietic Stem Cell

1.3.1 The HSC concept

Over 40 years ago, Till and McCulloch demonstrated that a single precursor cell exists in bone marrow that is capable of both extensive self-renewal and multi-lineage differentiation[30]. The HSC pool has been characterised phenotypically and is separable into distinct subpopulations based on both phenotype and function[22]. Long term reconstituting HSCs (LT-HSC) have the greatest self-renewal capacity, and give rise to all the haematopoietic lineages throughout life. The immediate progeny of LT-HSC are short term HSCs, which also are capable of generating all lineages of blood cells, but do so for a period of only 8 to 12 weeks. The proportions of HSCs that are LT-HSC vary with the site and age of the individual under investigation[31,32].

1.3.2 Physical location of HSCs

1.3.2.1 The Stem Cell niche concept

HSCs appear to circulate freely, but have little function outside specific anatomically defined locations. The concept of a niche as a specialised microenvironment housing stem cells was first proposed by Schofield over 30 years ago[33]. As the site of haematopoiesis changes during vertebrate development, the nature of this stem cell niche must also change[34]. The adult bone marrow niche has received most attention. As summarised in an excellent review of the area, there exists experimental evidence for key roles for both osteoblasts and vascular endothelial cells in the formation of the HSC niche[34]. The key experiments which led to these conclusions are summarised below.

1.3.2.2 The role of the osteoblast in the stem cell niche

The physical location of HSCs close to the bone surface was first demonstrated in 1975[35]. Morphological evidence for the presence of HSC niches in close association with the endosteum was provided more recently when cells with HSC and progenitor activity and/or phenotype were shown to localize close to the endosteal lining of bone-marrow cavities in trabecular regions of long bones, whereas more differentiated haematopoietic progenitors were found mainly in the central bone-marrow region[36,37].

The ability of osteoblastic cells to support HSC populations was first demonstrated by osteoblastic cell lines, which were shown to secrete cytokines capable of supporting HSCs[38]. Osteoblast activity is in part regulated by parathyroid hormone (PTH) or locally produced PTH-related protein (PTHrP), acting via the PTH/PTHrP receptor. Calvi et al developed a transgenic mouse model with a constitutively active PTH/PTHrP receptor, which resulted in upregulation of both osteoblast and HSC numbers[39].

In a second study, using a transgenic murine model with conditional inactivation of the BMP receptor type IA, which is normally expressed on osteoblasts lining the endosteum, Zhang et al demonstrated a simultaneous increase in the number of both osteoblasts and repopulating HSCs. The same study used histological examination of bone structure and staining for N-cadherin expression, to reveal a small subset of spindle shaped osteoblasts, which appeared architecturally in close contact with HSCs. A small subset of HSCs was shown to express N-cadherin asymmetrically on the surface which abutted the osteoblasts. As the cadherin family had already been shown to have a role in ovarian stem cell niche formation by knock-out experiments in Drosophila, this led to the hypothesis that N-cadherin mediated interactions between a subset of osteoblasts and HSCs might play a similar function in mammals[37].

1.3.2.3 The role of the vascular endothelial cell in the stem cell niche

A second school of thought proposes a key role for vascular endothelial cells in the formation of the HSC niche. Histological studies of murine bone marrow showed a large proportion of $CD150^+$ HSCs to be attached to the fenestrated endothelium of bone-marrow sinusoids[36]. Although vascular endothelial cells that are isolated from various adult non-haematopoietic organs have little or no ability to maintain HSCs in vitro, bone marrow sinusoidal endothelial cells (BMECs) are functionally and phenotypically distinct[40]. Indeed BMECs constitutively express cytokines such as CXC-chemokine ligand 12 (CXCL12) (also known as Stromal Derived factor 1 (SDF1))[41,42], the ligand for the CXCR4 receptor, and adhesion molecules such as endothelial-cell (E)-selectin[43] and vascular cell-adhesion molecule 1 (VCAM1) that are important for HSC mobilisation, homing and engraftment.

These two distinct niches may prove not to be mutually exclusive, but indeed may co-exist. By the use of label retaining studies (involving pulsed bromodeoxyuridine (BrdU)), those HSCs located in the endosteal niche have been shown to divide infrequently, and therefore may represent the most dormant subtype of HSCs[44]. Those cells that are found in contact with vascular endothelial cells comprise a more mixed distribution of cell types, including some that divide relatively frequently. Theoretically, the location of these $CD150^+$ murine HSCs, in close

proximity to sinusoids, would enable them to be constantly sensitive to blood-borne hormone levels, a position that would enable a rapid response to haematological stress.

1.3.3 Regulation of Stem Cell Behaviour

1.3.3.1 Maintenance of a quiescent population of stem cells is vital for long term haematopoietic survival

Multiple studies of cell cycle kinetics have identified the co-existence of pools of HSCs with distinct behavioural patterns: highly quiescent stem cells that remain metabolically inactive with little transcriptional activity; and those which divide more frequently, and are responsible for replenishing the pool of haematopoietic progenitor cells by division and differentiation into mature blood cells[20].

Relative quiescence is thought to have a protective role in ensuring the longevity of the HSC pool, perhaps by minimising stresses of cellular metabolism and genome replication[45]. Amongst other adaptations, HSCs appear to have developed a distinct response mechanism to DNA damage, possibly rendering them more tolerant to insults, but also more susceptible to mutational change. Double stranded DNA breaks occurring at the G_o stage of the cell cycle are subject to immediate repair by non-homologous end joining, which is more error prone than the response of other cell types to the same insult[46,47].

Experimentally induced defects in HSC dormancy by murine conditional knock-out experiments (see Section 1.3.3.1.1), frequently result in defects in stem cell self-renewal and stem cell pool exhaustion. Serial transplantation experiments in mice are eventually limited by haematopoietic failure, suggesting a requirement for repeated repopulation is unsustainable[48].

The factors which modulate the balance in these two distinct behaviour patterns have been the subject of a number of excellent reviews[49]. Experimental data suggest there are numerous factors which control HSC divisional status: as well as those intrinsic to stem cells themselves, signalling through cell-cell interactions within the niche, as well as secreted cytokines, may play key roles.

Murine studies suggest that the proportion of HSCs in cycle changes dramatically over the lifetime of an organism, presumably in response to its immediate haematopoietic needs. During foetal development, 95 to 100% of the HSCs located in the liver are in cycle at any one time point; by four weeks after birth, the proportion of those cycling in the bone marrow has fallen dramatically to less than 5%[32]. These cells cycle extremely infrequently (every 145 days)[50].

1.3.3.1.1 Genes identified as key regulators in HSC quiescence

Studies from murine knock-out models have suggested a key role for a number of genes in the maintenance of HSC quiescence. Knock-out of the *p21*[51], *Gfi1*[52], *Pten*[53] and *Foxo* genes[54] all resulted in an early expansion of the cycling pool of HSCs, followed by subsequent exhaustion of the population, as illustrated by serial transplantation assays. The same techniques have identified negative regulators of quiescence, such as the Myeloid EL1-like gene (*ELF4*), whose knock-out results in an increase in stem cell quiescence, expansion of the HSC pool and increased resistance to fluorouracil (5-FU) [55]. The majority of these genes are thought to exert their effects by controlling the rate of HSC entry into the cell cycle.

Wilson et al demonstrated a potentially key role for *c-myc* in the regulation of HSC number. Using a murine conditional *c-myc* knock-out model, they noted that reduced expression of the gene was associated with quiescence of the progenitor population, but a ten-fold increase in HSC numbers. Forced overexpression had the opposite effect. As c-myc has a role in the regulation of a variety of cell adhesion molecules, a potential pathway for its effect has been described[56].

1.3.3.1.2 Signalling interactions between stem cells and their microenvironment

Signalling interactions between HSCs and their immediate microenvironment have also been shown to have a vital role in the regulation of HSC behaviour. Arai et al demonstrated the importance of the Tie2-Angiopoietin-1 signalling axis in the maintenance of stem cell quiescence. The ligand Angiopoietin-1 is thought to be produced by niche osteoblasts, whilst HSCs positive for its receptor, Tie2, were found to be relatively quiescent. In vitro work involving the exogenous application of Angiopoietin-1 to HSC populations resulted in enhanced HSC adhesion, maintenance of an immature phenotype and maintained BM repopulating ability[57].

Other interactions thought to be important for HSC maintenance include that between c-mpl (expressed on HSCs) and thrombopoietin (TPO). Murine knock-out experiments resulting in reduced c-mpl expression resulted in a reduction in HSC function[58,59].

1.3.3.1.3 Metabolic adaptations and quiescence

HSCs are known to have a distinct metabolism, in part adapted to the hypoxia observed in the BM niche. Quiescent cells are more dependent on anaerobic respiration to meet their metabolic needs; as cells divide and mature, mitochondrial aerobic respiration becomes more important. While such relative hypoxia is thought to have a protective effect on the longevity

of HSCs, it may also have a role in the regulation of stem cell behaviour through signalling via the Hif pathway[60,61].

Ito et al described a key role for metabolism in the control of stem cell quiescence and differentiation, in studies of the promyelocytic leukaemia-peroxisome proliferator activator-receptor δ (PML-PPAR-δ) pathway, which controls fatty acid oxidation (FAO). Knockout of key components of the PML-PPAR-δ pathway, or drug-based inhibition of FAO, led to loss of HSC maintenance. FAO also appears to have a role in the determination of cell fate after HSC division: its inhibition resulted in the symmetrical commitment of daughter cells, in comparison to PPAR-δ activation which led to asymmetrical division[62].

1.3.4 Stem Cell Mobilisation and Homing

1.3.4.1 Normal HSC mobilisation in health and disease

Although the vast majority of HSCs are located in the bone marrow in a state of shielded quiescence, as part of normal homeostasis, small numbers mobilise and migrate throughout the body in peripheral blood. This process has been shown to follow a circadian rhythm[63], possibly modulated by the action of the sympathetic nervous system nerve endings, located in close proximity to the putative HSC niche within bone marrow[64].

No definitive reason why stem cells mobilise in this manner has been proven, though one can postulate it may represent an evolutionary adaption to ensure HSC survival in the case of inadequate bone marrow stromal support or localised damage. HSC circulation has also been noted to increase in times of physiological stress, such as illness or pregnancy. In work with a particular relevance for this thesis, Ailles et al have shown that normal HSCs are easily detectable in the peripheral blood of patients newly diagnosed with AML. In experiments designed to identify the presence of leukaemia stem cells (LSCs) in peripheral blood samples of patients with AML of abnormal karyotype, the authors found the number of LTC-IC cells generating colonies of normal karyotype (presumed to be normal HSCs) was ten-fold higher than those seen in controls[65].

1.3.4.2 Regulation of HSC homing to bone marrow

One of the key characteristics of HSCs is an ability to recognise and return to their bone marrow niche. This is the case in normal physiology, but is also clearly demonstrated both in xenograft transplant experiments, and clinically during autologous or allogeneic bone marrow transplantation.

Key pathways thought to regulate the homing process include the CXCR4/CXCL12 interaction. CXCR4 (CD184) is an alpha-chemokine receptor expressed on the surface of HSCs, known to bind to the ligand CXCL12 or SDF1. CXCL12 is highly expressed on bone marrow niche cells, and as such as been shown to be a potent chemokine directing HSC homing. A secondary regulatory pathway is provided by interplay between cKit and Stem Cell Factor (SCF or Kit ligand). cKit (also known as CD117) is a cell surface receptor universally expressed on HSCs[66], whose ligand SCF is highly expressed in BM stromal tissue. Roles for this interaction in stem cell homing and adhesion within the niche have been put forward[67].

1.3.4.3 Regulation of HSC mobilisation

Study of the factors responsible for stem cell mobilisation has revealed roles for a variety of cytokines, including granulocyte-colony stimulating factor (G-CSF), granulocyte-macrophage colony stimulating factor (GM-CSF), thrombopoietin (TPO), Kit and Flt3 ligand. As well as being of biological interest, this has allowed the development of drugs to promote HSC egress from bone marrow, prior to stem cell collection procedures for human transplantation.

The mechanism of action of G-CSF has been the most investigated to date. It appears able to suppress the activity of mature osteoblasts, as well as suppressing CXCL12 expression in mesenchymal stromal and endothelial cells. As such, it interferes with one of the main signalling axes responsible for HSC retention in the BM[68]. Experiments involving conditional knockouts of CXCR4 in the BM suggest this is the key mechanism responsible for the action of G-CSF[69].

Increased understanding of this process has resulted in significant clinical impact over the last decade, as peripheral blood based collection of HSCs by leucophoresis has overtaken the traditional bone marrow harvest as the preferred source of HSC collection for autologous and allogeneic bone marrow transplantation. Excess G-CSF mobilised peripheral blood (GMPB) collected in this manner also provides an excellent source of enriched human HSCs for laboratory work.

In addition to G-CSF, other drugs used clinically to enhance HSC mobilisation prior to stem cell collection include Plerixafor, a partial CXCR4 agonist which antagonises the binding of SDF-1 to its cognate receptor. A number of clinical trials have demonstrated its effectiveness when given in combination with G-CSF in mobilising HSCs[70].

1.3.5 Methodology for studying stem cells

1.3.5.1 Identification using flow cytometry

Identification of HSCs in bone marrow samples is complex for a number of reasons. They are rare cells, of rather indistinct morphological appearance. HSCs share a common appearance with lymphocytes, being small cells with a high nuclear to cytoplasmic ratio and an agranular cytoplasm.

One of the most commonly utilised techniques for stem cell identification in bone marrow and blood samples is based on flow cytometric analysis, which enables the simultaneous interrogation of large numbers of cells, thus rendering it an excellent technique for studying cells of a rare subtype.

HSCs are known to have a low forward (FSC) and side scatter (SSC) profile (see Section 2.4). Further phenotyping has identified a number of cell surface markers commonly associated with cells that exhibit stem cell behaviour. For human HSCs, these include CD34, CD38, CD90, CD133, CD105, CD45, and c-kit[71,72]. HSCs should be negative for the markers that are used for detection of lineage commitment (Lin-). Negative selection using a cocktail of lineage-specific markers can be used to enrich a sample for stem cells (see Section 2.2.3.1).

The field (and literature) is complicated by the fact that there exists cross-species variation in antigen expression on HSCs. Table 1-2 summarises some of the main difference in cell surface phenotype between human and murine HSCs[73].

Table 1-2 summarising the different antigen expression patterns seen on human and murine HSCs

ANTIGEN	HUMAN HSC	MURINE HSC
CD34	POSITIVE	LOW/NEGATIVE
SCA1	NO HUMAN HOMOLOG	POSITIVE
THY 1.1	POSITIVE	POSITIVE/LOW
CD38	LOW	POSITIVE
C-KIT	POSITIVE	POSITIVE
LIN	NEGATIVE	NEGATIVE

The experimental work within my PhD has solely involved human samples. Henceforth, unless clearly stated otherwise, further discussion of stem cell phenotype will be limited solely to that of the human HSC.

Flow-based sorting experiments of different fractions of cord blood, followed by transplantation assays into mice, have shown the $CD34^+$ fraction of cells to be highly enriched

for stem cells, and the $CD34^{+}38^{-}$ fraction even more so. Some cells capable of long term repopulation of murine models exist within the $CD34^{+}38^{+}$ fractions, but the bulk of these cells are more mature progenitors[74]. Indeed, functional stem cells are not solely limited to cells that express CD34. The Bonnet group recently published a study of $CD34^{-}$ cells, showing them to contain a subset of extremely primitive stem cells[75].

We, along with most groups in the field, accept the approximation that the $CD34^{+}38^{-}$ group of cells comprises a population highly enriched for HSC behaviour. However, it must be stressed that the cells that share this phenotype are by no means homogenous, but contain cells at a variety of stages of maturity and early lineage commitment[21,76].

1.3.5.2 Assays for Stem Cell Function

Flow cytometry helps us to identify cells likely to be enriched for stem cell behaviour, but the best tests of stem cell nature, as defined by the abilities of long-term survival, differentiation and repopulation, are provided by a series of in vitro and in vivo assays. These are described in brief below, along with a summary of their strengths and weaknesses.

1.3.5.2.1 Short Term in vitro Culture Assays

Short term in vitro assays are semi-solid assays, which can identify and quantify lineage-restricted progenitors in standardised conditions. Their immobilised progeny accumulate in tight colonies with specific characteristics of composition, size and colour. Several types of progenitors (erythroid, granulocytic, macrophagic and megakaryocytic) can differentiate simultaneously in a given culture[77]. Within a given lineage, progenitors with different levels of maturity can be identified dependent on sensitivity to cytokines, time to generate differentiated cells and colony size[78]. The growth media utilised in such assays is unable to support cell growth for greater than three weeks, and therefore these assays are not appropriate for testing for the presence of immature stem cells, as these have to undergo a much higher number of divisions (>15) over a longer period (>5 weeks) before producing the differentiated progeny this assay is designed to detect[76].

1.3.5.2.2 Long Term in vitro Culture Assays

A long term in vitro assay describes a system designed to identify immature progenitors, with an assay period extending beyond four weeks. This allows immature progenitors to complete differentiation, and rules out any contribution from lineage-committed cells present at the start of the assay. A common feature in these assays is the presence of a feeder layer of stromal cells, designed to secrete a source of regulatory factors, mimicking the complexity of the bone marrow stromal microenvironment[79]. Spontaneously immortalised murine bone

marrow derived stromal cell lines (such as MS-5, S17 and AFT024) are popular, because they support both myeloid and lymphoid development[80]. The presence of LTC-IC in the wells is identified retrospectively by the output of colony forming cells in the culture at five to eight weeks[81]. An initial stem cell concentration can be estimated by a serial dilution assay. One of the main issues with this assay is its lack of reproducibility between laboratories, due in part to variations in stromal cell lines and growth media utilised[76].

1.3.5.2.3 In vivo Murine Transplant Model

The use of a mouse transplant model allows the identification of long term repopulating stem cells, identified by their ability to reconstitute haematopoiesis in mice over a period of months, rather than simply weeks. To survive, transplanted stem cells also have to exhibit the key ability of homing, from the site of injection into the mouse tail, to the host bone marrow microenvironment.

Murine experimentation has identified three types of HSC behaviour in the transplant model: short lived Colony Forming Unit-Spleen (CFU-S) cells home to the spleen, and start producing cells by day twelve[82]; the Marrow Repopulating Activity (MRA) cells[83]; and finally the long term reconstituting cells, which produce differentiated myeloid and lymphoid lineages months after transplantation.

The first immunodeficient mouse model, sufficiently tolerant to allow engraftment of human HSCs, was described by Kamel-Reid and Dick in 1988[84]. However, the haematopoietic development observed in this first model was not normal- NOD SCID mice are strikingly more permissive to the development of B cells, with a noticeable absence of T and NK cells derived from the transplanted HSCs. Genetic manipulation of these models (such as the NOD SCID β2-microglobulin-null murine strain)[85] have led to increased permissiveness towards T and NK cellular differentiation.

1.4 Acute Myeloid Leukaemia

1.4.1 Introduction

Acute myeloid leukaemia is a rare, aggressive cancer, characterised by overproliferation of myeloblasts in the bone marrow leading to a failure of normal haematopoiesis.[86] The clinical features were first described two centuries ago by the French physician Alfred-Armand-Louis-Marie Velpeault, but it was the development of the light microscope by Virchow in the middle of the last century that meant its pathology could begin to be understood. Today, although the diagnosis of AML can still be made just by the examination of a stained blood film, additional

information is routinely gained by the use of immunological, cytogenetic and molecular analysis of bone marrow samples.

1.4.2 Incidence

Acute Myeloid Leukaemia is a rare cancer. 2013 figures from The Office of National Statistics suggest that it represents 0.8% of all new cancer cases in the UK, and 34% of all of all new leukaemia diagnoses. This translates to approximately 3000 new cases annually in the UK, with a male: female incidence ratio of 12:10[87].

Paediatric AML cases are rare. The incidence increases with age in adulthood, most markedly after the age of 60[88,89].

1.4.3 Predisposing Factors to Disease Development

For the majority of patients with AML, no predisposing factors are identified. There is a clear association between myeloid disorders such as myelodysplasia (MDS) and myeloproliferative disorders (MPD), and later progression to AML. Prior drug exposure to chemotherapy, especially the alkylating agents and anthracyclines, is associated with the development of therapy-related leukaemias several years later, often with characteristic genetic rearrangements[90]. Likewise, exposure to ionising radiation, such as that observed post-Chernobyl, is associated with an increased incidence of AML[91].

Rarely, germline mutations in *RUNX1* and *CEBPA* can lead to a familial predisposition to MDS/AML[92], as can more profound karyotypic abnormalities such as Down's syndrome, which is associated with a ten to eighteen fold increased lifetime risk of AML[93].

1.4.4 Clinical Presentation

Patients with AML almost always present with the signs and symptoms of an abnormal blood count: fatigue or shortness of breath due to anaemia; overt bleeding or skin petechiae due to a reduced platelet count and/or disseminated intravascular coagulation; or overwhelming infection due to a lack of functional neutrophils.

Rarely, presentation can be precipitated by the symptoms of leucostasis driven by an elevated white cell count: namely headache, blurred vision, cerebrovascular events or shortness of breath from pulmonary infiltration.

Very occasionally, patients can develop symptoms or signs from extramedullary blast infiltration: skin rashes, gingival hypertrophy or chloromas.

1.4.5 Diagnosis

1.4.5.1 Morphology

Blood tests are generally highly abnormal, with suppressed levels of normal blood cells, and either a low or a high WCC. A blood film stained with May-Grünwald-Giemsa (MGG) may be diagnostic if it reveals the presence of circulating blasts. A bone marrow aspirate plus or minus trephine biopsy is normally performed to confirm the diagnosis. Morphological examination of the marrow contents will show a predominance of immature blast-like cells, characterised by a nucleus with open chromatin, a high nuclear-cytoplasmic ratio and (dependent on the subtype of AML) granular cytoplasm. Further classification can be made if other cellular subtypes are visible in excess, or show signs of dysplasia.

The blast percentage and assessment of degree of maturation and dysplastic abnormalities in the neoplastic cells should be determined, if possible, from a 200-cell leukocyte differential performed on a peripheral blood smear, and a 500-cell differential performed on marrow aspirate smears stained with Wright Giemsa or MGG.

In the past, further classification of an AML sample, and confident distinction from acute lymphoblastic leukaemia (ALL) was done with the use of cytochemical stains such as myeloperoxidase, Sudan black and nonspecific esterase stain; however nowadays, these have largely been superseded by the use of flow cytometry.

1.4.5.2 Flow cytometry

Although not part of the World Health Organisation (WHO) diagnostic criteria[94], flow cytometry in the clinical setting is a crucial tool in the diagnosis of AML[95]. Immediate analysis of bone marrow and blood samples can facilitate distinction between AML and ALL samples in those cases of morphological uncertainty, as well as further subtyping AML dependant on the myeloid maturation markers co-expressed on the blasts. In recent years, with the development of multicolour flow panels, it can also be used to identify aberrant blast antigen expression patterns, which can then be used for assessment of Minimal Residual Disease (MRD) in subsequent post-therapeutic samples[96].

Standard diagnostic flow laboratories identify the presence of blasts based on expression of CD45, CD34, CD117 and HLA-DR, as well as basic FSC and SSC characteristics. Myeloid blasts may typically co-express myeloid maturation markers such as CD33 and CD13, but be free of those associated with the lymphoid lineage.

1.4.5.3 Cytogenetic Abnormalities

The trail-blazing work of Janet Rowley and colleagues in the 1970s resulted in the identification of recurrent karyotypic changes in AML samples. Nowadays, bone marrow samples at diagnosis are routinely sent for G banding and fluorescence in situ hybridisation (FISH) studies for the commonest chromosomal translocations associated with AML[97]. Cytogenetic abnormalities carry key prognostic information for patients: the t(15;17) translocation, associated with the formation of the PML RARA fusion protein, carries an extremely good prognosis; but is now treated very differently to other AML subtypes[98]. Translocations of 8;21 and inversion 16 are commonly seen, and also are associated with improved five year survival and superior response to certain drugs, such as anti-CD33 therapy[3,99]. AMLs associated with multiple different cytogenetic changes are associated with older age and preceding MDS, and carry a significantly poorer prognosis[100]

The MRC 2010 classification system for AMLs, based on presentation cytogenetics into good, intermediate and poor risk subgroups is summarised below in Table 1-3[101]. This has been used throughout this thesis, whenever cytogenetics-based risk stratification has been required during analysis

Table 1-3 The MRC 2010 Modified Cytogenetic classification of AML, based on the clinical outcome data from the MRC AML trials. Adapted from Grimwade et al[101]

RISK GROUP	CHROMOSOMAL ABNORMALITY
GOOD	t(8;21)(q22;q22) t(15;17)(q22;q22) inv(16)(p13q22)/t(16;16)(p13;q22)
INTERMEDIATE	Everything else
POOR	abn(3q) [excluding t(3;5)(q21∼25;q31∼35)] inv(3)(q21q26)/t(3;3)(q21;q26) add(5q), del(5q), −5, −7, add(7q)/del(7q) t(6;11)(q27;q23), t(10;11)(p11∼13;q23) t(11q23) [excluding t(9;11)(p21∼22;q23) and t(11;19)(q23;p13)] t(9;22)(q34;q11), −17/abn(17p), Complex (≥ 4 unrelated abnormalities)

1.4.5.4 Genetic abnormalities

AML is a disorder associated by the acquisition of multiple, complex genetic mutations. The seminal Cancer Genome Atlas consortium study sequenced 200 AML genomes covering the cytogenetic spectrum of the disease. These samples each harboured on average more than ten mutations, with in excess of 200 genes being identified as recurrent targets[102]. Clinical practise often lags behind laboratory research, but screening for common genetic mutations known to have prognostic significance appears to be growing in importance as a method for risk stratifying patients with normal cytogenetics. Many clinical laboratories will routinely screen for the presence of the *NPM1* mutation[103] and the *FLT3* internal duplication[104]. Characteristics associated with these mutations are discussed in more depth below. Other genes found commonly mutated in AML samples include *CEBPA*, *KIT*, *N-RAS*, *MLL*, *WT1*, *IDH1&2*, *TET2* and *DNTMT3A*[102].

1.4.5.4.1 *NPM1* Mutations

The nucleophosmin (*NPM1*) gene encodes for a multifunctional nucleo-cytoplasmic shuttling protein, which is normally localized in the nucleolus. Mutated NPM1 protein locates aberrantly to the cytoplasm of leukaemia; hence the term NPM-cytoplasmic positive AML[105].

NPM1 mutations occur in 25 to 35% of all AML patients, but are seen much more frequently (50 to 60%) in samples of normal karyotype. Mutations are strongly associated with de novo AML, and are rarely seen in AML associated with previous chemotherapy exposure, myelodysplasia or myeloproliferative disorders.

The mutation is characteristically heterozygous, with preservation of the wild-type allele. Although multiple gene mutations have been described, the vast majority affect Exon 12. Type A mutations represent a duplication of a TCTG tetranucleotide at positions 956 to 959 of the reference sequence, and account for 75 to 80% of cases. Mutations B and D are responsible for 10% and 5% of cases respectively; other mutations are very rare[106].

Evidence from recent papers which involve sequencing for early mutations in pre-leukaemic HSCs (see section 1.5.4) suggest that *NPM1* mutations are not seen in this population of cells, which are more often characterised by alterations in genes such as *DNMT3A*[107,108]. However, the acquisition of *NPM1* mutation appears to be a key, later, transformational step in towards the leukaemia phenotype. Once acquired, its expression remains remarkably stable over the disease course[109,110]. Such persistence and stability in a mutation makes it an excellent candidate for MRD assessment[111].

NPM1 mutant AML was a provisional entity in WHO 2008 classification of myeloid disease[112] and now in the recent 2016 reclassification is a distinct entity[113].

1.4.5.4.2 *FLT3* mutations

Fms-like tyrosine kinase 3 (*FLT3*), also known as CD135, is a receptor tyrosine kinase expressed on the surface of early haematopoietic cells. Normal signalling through the FLT3 receptor is thought to be important in the development of stem and progenitor cells.

Mutations in the *FLT3* gene are seen in some 25 to 30% of newly diagnosed adult AML cases[104,114]. Two patterns of mutation are observed: internal tandem duplications (*FLT3*/ITD mutations) in or near to the juxtamembrane domain, and point mutations resulting in single amino acid substitutions with the tyrosine kinase domain (*FLT3*/TKD). Both cause constitutive activation of the receptor[115].

Patients with *FLT3*/ITD mutated AML often present with a similar phenotype, namely a high peripheral WCC and a hyperproliferative bone marrow appearance on morphology[116]. Whilst sensitivity to induction chemotherapy is comparable to other AML patients, *FLT3*-ITD mutation is associated with an increased relapse rate and short remission times[115]. The *FLT3*-TKD mutation is less common and does not share the same clinical phenotype[117].

In comparison to *NPM1* mutations, alterations in the *FLT3* gene are generally viewed as later events in the evolution of AML. When the Majeti lab studied *FLT3*-mutated AML samples for the presence of pre-leukaemic stem cells, the *FLT3* mutation was universally absent, suggesting this is not an initiating mutation[118]. Similarly, in a study of paired diagnosis and relapse samples, the mutated *FLT3* allelic burden varied significantly, suggesting a degree of instability in the mutation[119].

In addition to their prognostic significance, *FLT3* mutations are of interest in the world of precision medicine, as they offer a potential therapeutic drug target. A number of FLT3 inhibitors exist and have been used in the clinical trial setting. The first generation FLT3 inhibitors (examples include sunitinib, midostaurin, and lestaurtinib) were initially designed for use in other solid organ malignancies, and despite promising in vitro activity against AML samples, have largely been disappointing both in terms of efficacy and tolerability when given to patients[115]. However, second-generation FLT3 inhibitors have begun to emerge, such as quizartinib (AC-220), crenolanib (CP-868596), and PLX3397. These have a greater specificity for FLT3 receptor, and as such offer the promise of potentially greater efficacy and fewer off target side defects[120].

The MRC-sponsored, UK based AML trials have included the investigation of several FLT3 inhibitors in recent years (from which formal data remain unpublished at this time). AML-17 involved randomisation of those patients found to have *FLT3* mutations to receive CEP701[121]; AML-18 involved randomisation to receive the second generation FLT3 inhibitor AC220, either over three courses or as a year-long maintenance therapy[122], and finally, AML-19 includes the Phase II pilot study of Ponatinib for those with *FLT3* mutations[123].

1.4.6 Classification Systems for AML

AML is a highly heterogeneous disorder, with individual samples showing a wide range of morphological, cytogenetic and molecular aberrances. Developing a classification that recognises these differences has been an importance goal for more informative clinical research and patient prognostication. The most recent classification system was issued by the WHO in 2008, and relies heavily on cytogenetic information to group samples with similar disease evolution and prognosis[112]. Previously, classification was based on the morphological and cytochemical appearance of blasts, using the French-American-British (FAB) system[124].

1.4.7 Current Therapy and Clinical Trials

The backbone of treatment given with curative intent remains chemotherapy. For those patients fit enough to receive it, intensive chemotherapy is given in blocks. An initial induction block aims to achieve morphological remission, followed by two to three consecutive consolidation blocks, attempting to remove any residual leukaemia cells There exist global variations in favoured induction regimens, but in the UK, most patients will receive a combination of daunorubicin and cytarabine (plus or minus etoposide) as a first therapy.

Chemotherapy given with curative intent comes with significant side effects. These include further myelosuppression, alopecia, infertility, nausea and anthracyline-induced cardiotoxicity.

Typically, 75% of patients will achieve a morphological remission after a first induction course, and will regain the ability for normal haematopoiesis four to five weeks after the initiation of chemotherapy[101]. Dependant on age, transplant availability and pre and post treatment prognostic indicators, patients will either proceed with further consolidation chemotherapy, or an allogeneic bone marrow transplant.

The intensity of side effects seen with standard chemotherapy, combined with its relative lack of efficacy in certain disease subtypes, mean that for frailer patients, it is not an appropriate therapeutic option. Thus many patients presenting with AML may be offered a palliative

approach with early treatment of infections and regular blood and/or platelet infusions, in combination with the use of historically well tolerated drugs such as hydroxyurea to limit the speed of disease progression. In more recent years, attempts have been made by the MRC to improve treatment options available to this group, including wide scale testing of novel agents[4,122].

It is worth noting that APML, characterised by a distinct morphological appearance and associated with the t(15;17) cytogenetic abnormality, is now approached very differently to other AMLs. When recognised and treated promptly, and the bleeding disorders often observed at diagnosis due to disseminated intravascular coagulation (DIC) are corrected, these leukaemias are associated with an excellent prognosis[98]. This relatively pathologically-homogenous group of leukaemias respond extremely well to non-standard chemotherapy with ATRA[125] and Arsenic trioxide, which are much better tolerated than standard cytotoxics.

In the UK and parts of the Commonwealth, the majority of new patients diagnosed with AML are recruited into the current age-appropriate MRC trial. In the past, these trials have been designed to try and answer a number of key treatment decisions: the best induction regimen; the most appropriate number of consolidation courses of chemotherapy; which patients should receive an allograft in first remission. They have been used for rapid assessment of the efficacy of a number of drugs which showed laboratory promise: Mylotarg, Everolimus, and Zarnestra being a few examples. More recently, samples gathered have also been used to assess the feasibility of flow cytometry and molecular genetics as means of rapid MRD assessment of patient responses to therapy[121,122,126].

In Europe, similar national trial conglomerates include GIMENA (Italy), PETHEMA (Spain) and the HOVON group (Netherlands). In the US, the CALGB (Cancer and Leukaemia Group B) and the SWOG (South Western Oncology Groups) cooperatives run similar trials.

1.4.8 Prognosis

The following are currently viewed as important prognostic markers at the time of disease diagnosis: age, presenting WCC, the presence of particular cytogenetic abnormalities, development of secondary AML, and the mutational status of *FLT3*, *NPM1*, *CEBPA* and *IDH1&2*[97]. Predictors of overall outcome after therapy include the blast count after induction chemotherapy and the absence of minimal residual disease (MRD)[96,111].

Current estimates of five year survival for all patients at diagnosis with AML are between 40 and 50%, but with significant variation dependent upon age and disease characteristics at

presentation. Study of the long term survival data from sequential MRC AML trials between 1978 and 2009 would suggest that 5 year survival has improved significantly during this period for those patients under the age of 60[3], but with much more modest improvements for those aged 60 and over[4]. Improvements in supportive care and aggressive treatment of sepsis have improved treatment-induced mortality. Sequential multinational studies have led to the honing of cytotoxic chemotherapy regimens, but with the notable exception of APML, the backbone of chemotherapy for AML remains the same as it was three decades ago[5].

1.4.9 The Leukaemia Stem Cell Model

1.4.9.1 The LSC Concept

The idea that a cancer stem cell exists which shares some of the characteristics of a normal stem cell has been postulated for some time, but the first experimental evidence for its existence came from the study of leukaemia samples. In 1997, Bonnet and Dick transplanted $CD34^{+}38^{-}$ cells from leukaemic samples into NOD SCID mice, and showed that these fractions were capable of developing into leukaemia blasts and self-propagating[127]. Furthermore, these cells can go on to produce the disease in a secondary mouse transplant experiment, indicating the ability of self-renewal in the primary recipient. These cells were described as leukaemia stem cells (LSCs).

In a parallel to normal haematopoiesis, AML is often described as a loose hierarchy, in which a small number of self-renewing LSCs give rise to a large population of more mature blasts, which lack self-renewal capacity. LSCs are thought to also share the characteristics of quiescence and apoptotic resistance with normal HSCs.

One reason for the frequent failure of conventional chemotherapy to prevent relapse may be that the infrequently dividing LSCs are more chemoresistant than the majority of the blast population.

Although the LSC is thought to maintain the reservoir of disease, confer chemoresistance and be the cell type responsible for delayed relapse, we cannot assume that all leukaemia initiating mutations occur in the normal healthy HSC fraction of haematopoietic cells. Indeed, a mutation event in a partially committed progenitor cell may re-confer self-renewal abilities. Alternatively, although an initial mutation may occur in a stem cell, and persist indefinitely in this pool, development to blast crisis may only occur with the addition of further mutations, as appears to be the case in CML[128].

1.4.9.2 Phenotypic Identification of the LSC

Improved efficiency of the initial xenograft transplant model (use of the NSG murine strain, or intrafemoral injection of transplanted cells), allowed experimental work which showed that LSC cells exist not just within the $CD34^{+}38^{-}$ fraction of bone marrow, but can share the phenotype of a more committed cell[129]. LSC phenotype can also vary dependant on the phenotypic appearance of the leukaemia: for example, in AML samples characterised by low expression of CD34, LSC activity is restricted to the $CD34^{-}$ compartment, whereas in $CD34^{+}$ AML samples, LSCs are found in both $CD34^{+}$ and 34^{-} fractions[130].

A number of other antigens commonly expressed on LSCs have been reported. These are of interest both for the experimental identification of LSCs and also their therapeutic targeting[131,132]. The most valuable targets are those which are differentially expressed on LSCs and normal HSCs.

1.5 Why does the development of AML lead to a failure of normal haematopoiesis?

1.5.1 Introduction

Many acquired haematological disorders, malignant and non-malignant, present with the symptoms resultant from disruption in normal haematopoietic development, but the mechanism or mechanisms by which these diseases disrupt haematopoiesis are in general poorly understood. Both haematological and non-haematological malignancies such as metastatic breast carcinoma or melanoma presenting with heavy bone marrow infiltration can result in abnormal blood counts: any clinician should be fully aware that this is not an association solely seen with AML. However, the extent of bone marrow infiltration with abnormal disease alone is not predictive for the effect on a patients' blood count: for example, many patients with CLL will present with a marrow effaced with lymphocytes, and yet maintain normal haematopoiesis.

Personal observation of presentation bone marrows with AML suggests that even with fairly limited disease bulk (i.e. 20 to 30% blasts), myelopoiesis is much more profoundly affected than it would be from a comparable disorder.

The recent literature in this area is sparse and therefore open to conjecture. As AML is a disease of the developing myeloid lineage, it could be argued that healthy myeloid cells and blasts are more likely to be competing for resources (such as niche space, nutrients, appropriate cytokine signalling), than if the diseased cell was derived from another lineage.

Are the healthy myeloid cells sensitive to negative feedback signalling from the blast population which might impede their growth? Might exosome secretion from blast cells affect the normal haematopoietic population? Or is the effect a grosser one, such as a change in pH within the stromal milieu to which the normal tissue is sensitive? Or is it that, in bone marrows which have developed AML, even the "normal" cells are aberrant (see Section 1.5.4).

Secondary questions include the following. Where in the pathway of normal haematopoiesis do these effects take place? Are *all* cells along the haematopoietic pathway of development sensitive to the effect of AML, or are some more sensitive than others? The fact that patients commonly present with trilineage haematopoietic failure might suggest that those cells at the top of the hierarchy are most likely affected, but this does not mean subsidiary effects are not seen on partially committed progenitors as well.

Indeed, there may be more than one or indeed multiple mechanisms concurrently responsible for the observed failure of haematopoiesis in any one patient. AML is an incredibly heterogeneous disorder, as described above, but a strongly recurrent theme across almost all patients is the bone marrow failure associated with it. Are the mechanism or mechanisms by which this occurs universal across the diversity of cytogenetic and genetic abnormities associated with the development of the disease?

Clinical observation suggests that the effect of AML on normal haematopoiesis is generally reversible. Assuming a patient enters clinical remission after induction chemotherapy, normal bone marrow function is generally regained three to four weeks after the completion of therapy.

1.5.2 The interaction between HSCs and AML: summary of previous work by the group which led to the hypothesis that acute myeloid leukemia does not deplete normal haematopoietic stem cells but induces cytopenias by impeding their differentiation

As early as 30 years ago, it was observed that AML cells had the ability to disrupt growth of haematopoietic colonies by the secretion of an apparently diffusible substance[133,134]. Although we recognise that inhibition of normal haematopoiesis by AML may occur at many steps along the cellular development hierarchy, our laboratory to date has focussed on the interaction between AML and normal HSCs. Our work has used both a murine xenograft model of AML engraftment, and the study of primary human bone marrow samples[1].

1.5.2.1 Murine experiments

111 unirradiated NSG mice were transplanted with 10 human AML samples to determine the effect of AML engraftment on the mouse $CD45^+$ cells. The mice were sacrificed at different time-points post transplantation to study the effect of the developing AML on their haematopoiesis. As the AML engrafted, there were three phases of haematopoietic activity observed in the mouse bone marrow, termed early, mid and late. By the midphase of AML engraftment, progenitor cell numbers had fallen in the mouse bone marrow ($p<0.0001$) in comparison to controls, whilst HSC levels were maintained. In the late phase of engraftment, both progenitor and HSC numbers had fallen significantly. Figure 1-2A summarizes this data.

Assays of HSC function were also performed on the murine $CD45^+$ cells extracted from bone marrow at the midphase, to back up the immunophenotyping data. Colony-formation and repopulation assays, as well as xenograft secondary engraftment studies suggested that the murine $CD45^+$ fraction is enriched with long term repopulating cells at this time-point.

HSC cycling at the midphase was tested by BrdU incorporation assay. In the two samples tested, murine stem cells showed significantly reduced BrdU incorporation ($p=0.007$ and $p= 0.04$), suggesting reduced cycling.

These findings are consistent with a hypothesis that AML induces BM failure by impeding differentiation at the HSC-progenitor transition.

The ten samples used in this study had all previously been shown to be capable of engrafting in a mouse xenograft model, and represented a range of cytogenetic abnormalities. All had a similar effect on murine haematopoiesis, which gives weight to the idea of a universal theory of bone marrow suppression by AML. However, the use of an interspecies transplant model is open to a number of significant criticisms. Any observed changes in stem cell behaviour might be due simply to the transplantation process, rather than the interaction between AML cells and murine HSCs.

1.5.2.2 Primary Human Samples

Human correlation of these observations from mouse based studies was provided by studies of HSC and progenitor numbers in the bone marrows of 16 patients with a rare subtype of AML.

The term $CD34^{low}$ AML is here used to describe a rare subtype of AML, where the blasts are characterized by low expression of CD34. Diagnostic clinical laboratories will typically describe a $CD34^-$ AML as one where less than 20% of the blast cells express CD34. This phenotype is

highly associated with the presence of the *NPM1* mutation[105]. However, our group has previously subclassified this group into three groups defined by the phenotypic flow appearances (see Figure 1-2B). Subgroup A contains a much more restricted $CD34^+$ population (<0.5% of all cells). Previous transplantation studies have shown this region to be free of LSCs, and enriched for normal stem cell activity[130,135]. Quite why normal karyotype *NPM1*-mutated AMLs exhibit such a range of phenotypes is not clear, but the Subtype A group was notably free of associated mutations in genes such as *FLT3*, which were more commonly observed in samples with higher CD34 expression.

The subtype A group of AML samples (henceforth referred to as $CD34^{low}$) encompasses 40% of *NPM1* mutant normal karyotype AMLs, and a small proportion of *NPM1* wild type AMLs as well. Because both the $CD34^+38^-$ HSC subset and the $CD34^+38^+$ progenitor subset have previously been shown to be clear of leukaemic cells, they represent an excellent sample group to study the effect of AML on normal HSCs[135].

In 16 patients with $CD34^{low}$ leukaemias, the HSC ($CD34^+38^-$) and progenitor ($CD34^+38^+$) populations were quantified at diagnosis and compared to controls. Whilst the progenitor numbers were significantly reduced in AML samples, the number of normal HSCs was not statistically different between AML and control samples. Assessment of cycling status of the HSCs was determined by Ki67 staining. A lower percentage of HSCs from AML BM were in cycle in comparison to controls (p=0.002), a finding mirrored by the murine xenograft study. LTC assays and murine engraftment tests showed the $CD34^+$ fraction within the AML samples to be more enriched in LTC-IC and repopulating cells than controls. These findings are summarised in Figure 1-2 C and D.

The combination of this murine xenograft and human data led the group to develop the hypothesis that acute myeloid leukaemia does not deplete normal haematopoietic stem cells numbersbut induces cytopenias by impeding their differentiation. The mechanism by which this inhibition occurs remains unknown.

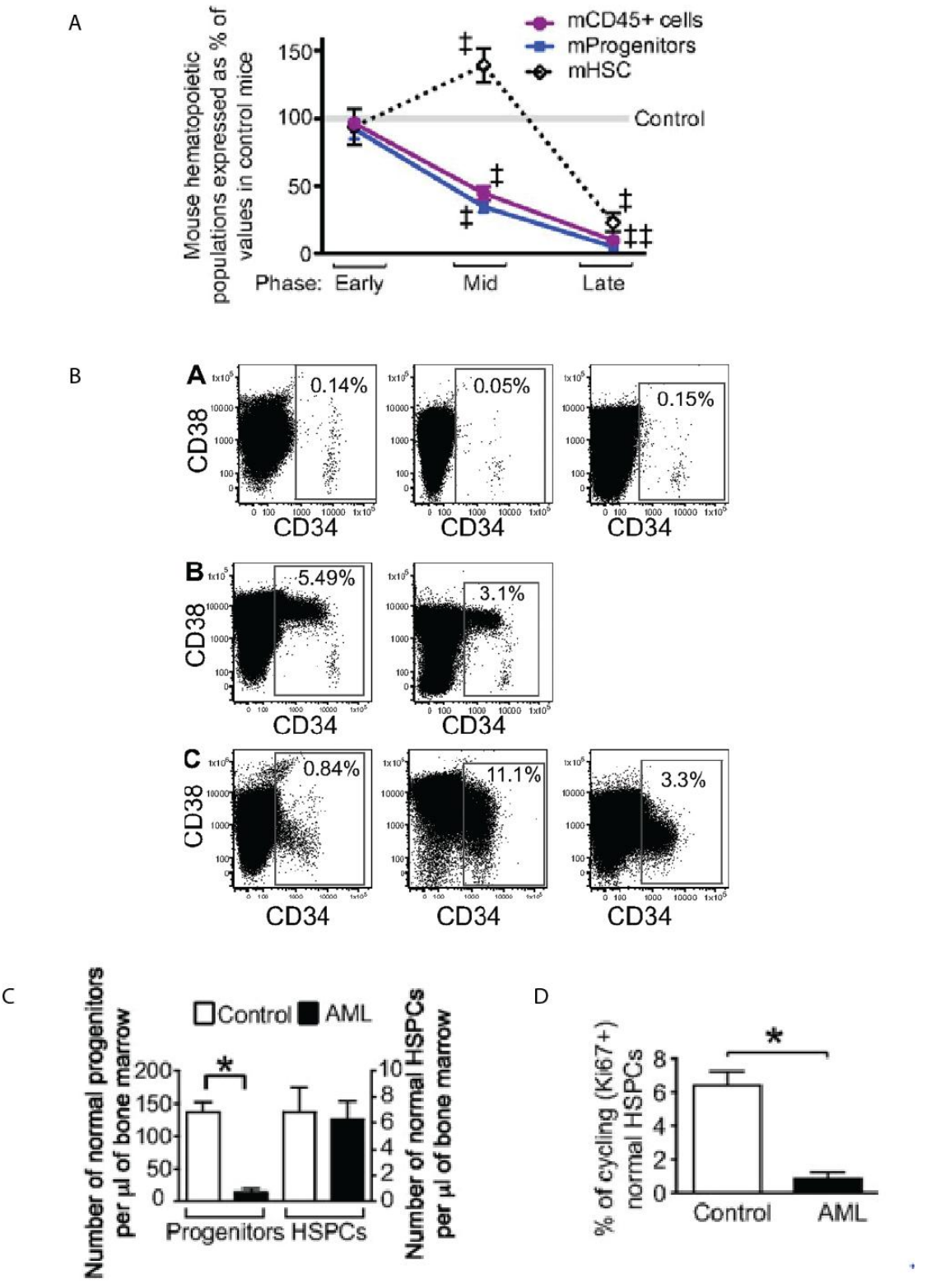

Figure 1-2A The effect of AML on mouse haematopoietic populations in a xenograft model showing number of $CD45^+$ cells, progenitors and HSCs in mice with AML as a percentage of values in controls. The grey line represents controls. Adapted from Miraki-Moud et al, PNAS, 2013[1]. B: Classification of 8 *NPM1* mutated AML samples based on CD34 and CD38 expression. Subtype A samples have <0.5% of cells which express CD34. Subtype B samples have >0.5% of cells which express CD34, but the $CD34^+38^-$ population is small (<0.1%) and distinct from the $CD34^+38^+$ and $CD34^-$ populations. Subtype C samples are more heterogeneous. The $CD34^+$ fraction is greater than 0.5%, but there is no distinct, small $CD34^+38^-$ population. Diagram from Taussig et al, 2010. C: Assessment of normal residual haematopoiesis in humans with $CD34^{low}$ AML. Numbers of phenotypically defined HSPCs are preserved (p=0.6) in the BM of 16 patients with subtype A AML, whereas progenitors are reduced in comparison with 42 controls (p<0.0001). D: The percentage of HSPCs in cell cycle was lower in subtype A (n=7) AMLs than controls (n=9) (p=0.002). Adapted from Miraki-Moud et al, PNAS, 2013[1]

1.5.2.3 Corroborative data

Since this paper was published, encouraging corroborative findings have been published by an unrelated group. Cheng et al developed a non-irradiated murine model of AML, by retroviral-mediated introduction of the *MLL-AF9* fusion gene into B6-Ly5.2 (CD45.2) mice. The resultant leukaemia cells were harvested, and introduced into B6-Ly5.1 (CD45.1) mouse recipients, which subsequently developed AML within 21 days. Flow-based analysis of the CD45.1 normal haematopoietic populations showed relative preservation of LT-HSC numbers in comparison to more committed progenitors, until the last stages of AML development. BRDU incorporation studies suggested that these HSCs are more quiescent than those extracted from untransplanted mice.

Gene expression analysis of the murine HSC populations performed at Days 7 and 14 showed differential expression of a number of quiescence-associated genes, including the transcription factor *Egr3*, which was overexpressed in murine stem cells exposed to AML, relative to controls[136]. Members of the Egr family are transcription factors, whose previously described roles include the control of HSC[137] and lymphocye proliferation[138,139].

Further evidence was presented for a potential role for Egr3 expression in mediating the effect of AML on HSCs. Retrovirally induced overexpression of *Egr3* in murine stem cells led to a quiescent phenotype both in vitro and in a murine transplant model. The reverse effects where observed when *Egr3* expression was reduced in murine stem cells by shRNA knock-down.

1.5.3 The interaction between AML and stromal cells may affect the niche and therefore HSC development indirectly

An alternative, but by no means mutually exclusive theory to explain the changes seen in haematopoiesis in AML is that, rather than solely directly inhibiting normal HSCs, AML cells might induce changes in the microenvironment and bone marrow niche such that it is no longer able to support normal haematopoietic development.

Support for this concept is provided by recent work looking at the effect of AML on mesenchymal stromal cells isolated from human AML diagnostic BM samples. The growth of these cells, as assessed in vitro by colony-forming unit fibroblast activity and population doubling time, was significantly lower than controls. The cells also exhibited reduced osteogeneic differentiation, and differential DNA methylation. Importantly, mesenchymal stromal cells also showed a reduced ability to support normal haematopoietic development as

evidence by reduced growth of normal $CD34^+$ cells in a 7 week culture assay[140,141]. Similar findings have also been described in murine models[142,143].

More radically, some sources have suggested that acquired mutations within stromal cells might themselves be a driver for leukaemia development. In a murine model, selective deletion of the miRNA processing endonuclease *Dicer1* in mesenchymal osteoprogenitors resulted in abnormal haematopoietic development, characterised by MDS-like changes and AML development[144]. Blau et al investigated the genetic makeup of MSCs on 96 AML samples and 36 controls. They found chromosomal changes within the MSCs of patients who had developed leukaemia to be significantly more common than within controls, and particularly associated with the development of poor risk AMLs. However, the actual patterns of chromosomal aberrations seen in AML and stromal cells were different. Whether there exists a causal relationship between these findings is not clear[145].

1.5.3.1 Potential mechanisms for signalling between AML and HSCs: do exosomes play a role?

Our previous work described observed changes in HSC and progenitor behaviour in the context of AML development, but did not include an explanation for how these changes might be brought about, or what signalling mechanisms might be important. As detailed above, others have subsequently proposed a role for *Egr3* in mediating the observed HSC quiescence[136].

An interesting hypothesis is that exosome secretion from AML blasts might prove a means of modulating HSC and/or stromal cell behaviour that does not require direct cell-cell contact[133].

Exosomes are tiny (30 to 100nm diameter), cell-derived vesicles, which contain protein, RNA and miRNA. They were first identified in reticulocytes, but have since been associated with a range of different cell types[146]. Exosomes have been found in the plasma, bone marrow and urine in healthy humans[147].

Exosomes are produced either by the fusion of vesicular bodies with the cell surface membrane, or by budding of the plasma membrane itself. Whilst the normal physiological functions of exosomes are still being explored, it is clear that the vesicles retain the ability to fuse with other cells, deposit their contents into the recipient cell, and as a result, modulate cell behaviour.

However, exosome secretion is not simply observed in normal cells, but also by a range of malignant cells. It has been hypothesized that exosome secretion is one way in which cancer cells can modulate their microenvironment to a more permissive state.

The presence of exosomes in fluid surrounding both AML primary samples and cell lines has been demonstrated by density-gradient centrifugation of culture medium and subsequent transmission electron microscopy studies. The transfer of membrane proteins in real time was visualized by staining AML cells with N-Rh-PE, which was imaged localizing to the cell membrane before subsequent budding-off. Fusion of these exosomes with stromal cells was illustrated by exposure of murine stromal cells to PKH26-labelled exosomes from the HL-60 AML line. Subsequent RNA analysis of the murine cells after 48 hours showed the presence of human derived CXCR4 transcripts within the cells[148]. The transfer of exosome contents to stromal cells has been shown to reduce their expression of SCF and CXCL12. The same paper also identified a number of exosome-induced changes seen directly in HSCs including CXCR4 and cKit downregulation, as well as repression of transcription factors such as c-Myb and CEBP-β. Whether it is via the loss of supporting signals from stromal cells, or direct changes induced in normal HSCs, the secretion of exosomes may be one way in which AML induces quiescence in normal HSCs.

Other authors have reported similar changes in stromal cell behaviour induced by exosomes, as well as the abrogation of these effects when extravesicular transport was inhibited. One paper identified a particular role for MiR-7977 secretion within exosomes in the induction of HSC quiescence[149].

1.5.4 Caveat: How normal are the "normal" HSCs in acute myeloid leukaemia samples? The pre-leukaemic stem cell concept

There have been a number of excellent recent papers examining the HSC fraction within AML samples for the presence of early, leukaemia associated mutations. Evidence for the existence of a pre-leukaemic stem cell population dates back to clinical observation of one patient, in remission from t(8;21) associated AML. $Lin^-CD34^+CD38^-CD90^+$ HSCs isolated from the patient up to 150 months after therapy, produced normal myeloid colonies in methylcellulose in vitro clonal progenitor cell assays, yet contained detectable AML1-ETO transcripts [150].

The Majeti group from Stanford separated $Lin^-CD34^+38^-TIM3^-$ cells from six patients with normal karyotype *FLT3*-mutated AML. Having identified genetic mutations in AML by exome sequencing, they went on to study the separated HSCs and their progeny (derived from

subsequent murine transplantation) for the presence of the same mutations by targeted sequencing. In five of these samples, 32 of a total of 51 mutations found in the leukaemia cells were also found to be present in the residual HSCs. Moreover, 7 of 13 mutations in genes recurrently mutated in AML were also present in the residual HSC population, with the notable absence of *FLT3*-ITD in all five cases[151].

A similar study was performed by John Dick's laboratory, using *NPM1*/*DNTMT3A* mutated AML samples. In this case, HSCs, common myeloid progenitors (CMPs) and mature lymphocytes were flow sorted from the leukaemia sample and the resultant cell fractions sequenced for the presence of mutations. *DNTMT3A* mutations were commonly found in all normal cellular fractions, whereas the *NPM1* mutation was found only in the blast populations. The authors highlight the fact that the *DNTMT3A* mutated HSCs appear at a competitive advantage in serial transplantation assays over non-mutated cells[107].

Clinical data supporting the idea of pre-malignant HSC mutations come from Jaiswal *et al,* who performed whole exome sequencing on blood samples from 17182 people (previously recruited into a national Type II Diabetes surveillance study and therefore unselected for haematological disease), looking for the presence of mutations within 160 genes recurrently identified in haematological malignancy[108]. Mutations were common (most commonly in *DNTMT3A*, *TET2* and *ASXL1*), with an increase in incidence with rising age. 9.5% of patients aged 70 to 79 were found to have one or more mutations in the studied gene set. Moreover, in long term follow up studies, the presence of a somatic mutation was associated with a significantly increased risk of developing a haematological malignancy (hazard ratio 11.1). For those patients who had serial samples available, sequential studies revealed persistence of the mutations over time. Similar findings have been made by others[152], and have led to the concept of clonal haematopoiesis of indeterminate potential (CHIP)[153].

.

Chapter 2 Materials and Methods

2.1 Samples and Ethics

Samples of peripheral blood and bone marrow from patients with AML and controls (obtained pre-treatment from patients with Stage I and II lymphoma and normal blood counts) were obtained by the Tissue Bank maintained by the Centre for Haemato-Oncology at Bart's Cancer Institute. This tissue bank is maintained according to the Human Tissue Act 2004 (Licence no. 12199). All patients had to sign consent forms to allow storage of specimens for research purposes after appropriate counselling, in accordance with the Declaration of Helsinki.

Both cord blood and G-CSF mobilised peripheral blood (GMPB) were used as sources of enriched stem cells for optimisation of experimental techniques, controls where stated, and for the source of stem cells within the AML Co-Culture assay. GMPB was obtained with consent from the excess donations of healthy donors and patients undergoing primed stem cell collection at Bart's Hospital. Cord cells were derived from donated umbilical cord samples from the Delivery Suite at the Royal London Hospital.

This project was approved by the East London and City HA local research ethics committee 2 with REC reference 06/Q0604/100 titled "Characterisation of leukaemia stem cells and their interactions with normal bone marrow cells".

2.2 Sample Handling Techniques

2.2.1 Collection of Fresh Bone Marrow Aliquots for Immediate Analysis

A standard operating procedure was used to ensure that the BM samples used for quantification of HSCs were obtained from the first pull of the aspirate and in a small volume (less than 0.5ml) to reduce haemodilution. Samples were decanted into eppendorfs containing 100µl Anticoagulant Citrate Dextrose Solution (Haemonetics, Cat: 426C). Samples with visible clots at the time of analysis were discarded. 200µl of fresh BM was mixed with 4ml ammonium chloride solution (Stem Cell Technologies, Cat: 07850), and incubated at 4^{o}C for 10 minutes to induce red cell lysis. 1ml FCS was added and the sample pelleted (1500rpm, 5 minutes). The cells were re-suspended in 50 µl 2% HAG, mixed well and incubated for 20 minutes at 4^{o}C. 5µl of CD34-PerCP (BD, Cat: 345803 (clone 8G12)) and CD38-Pecy7 (BD, Cat: 335825 (clone HB7) were added, and the sample incubated for 30 minutes at 4^{o}C. The sample was then washed in 4ml 2% PBS, and re-suspended in 300µl 2% PBS/DAPI/DNase. Cell counting was facilitated by the addition of 50µl CountBright™ Absolute Counting Beads (Life Technologies, Cat: C36950). Analysis was performed using a BD Biosciences LSR Fortessa.

2.2.2 Tissue Bank Processing of Samples for Delayed Analysis

Bone marrow aspirate samples were collected into EDTA collecting tubes (from multiple suppliers) at Bart's Hospital. On arrival in the Tissue Bank, extraction of mononuclear cells was performed following the Departmental SOP. In brief, samples were layered on top of 4ml Lymphoprep^{TM} (Stem Cell Technologies, Cat: 07851) in a 15ml Falcon tube, and centrifuged with brakes off (1500rpm, 25 minutes). The resultant mononuclear cell layer was removed using a pipette, and added to a Falcon tube containing 9ml of RPMI Medium (Sigma, Cat: R8758). Cell count and viability was assessed using the Beckmann Coulter Vicell XR Cell Viability Counter.

The sample was then centrifuged (1300rpm, 10 minutes), and the resultant pellet resuspended in 1ml freeze mix (7 parts RPMI; 1 part DMSO; 2 parts FCS), before gradual freezing to -198^{o}C.

2.2.3 Processing of G-CSF mobilised peripheral blood and cord blood

Excess aliquots of leucophoresis blood obtained from patients or donors undergoing G-CSF mobilised peripheral blood harvests at St Bartholomew's Hospital, London, were obtained after obtaining written informed consent. Alternatively, samples of cord blood were obtained post-delivery at the Royal London Hospital, after written informed parental consent.

These samples were diluted 1:4 with PBS, and overlaid onto Lymphoprep^{TM} (Fresnius, Cat 2015-06). The samples were centrifuged with breaks off (1550rpm, 30 minutes) at room temperature, to produce separation of red cells and a layer enriched in mononuclear cells. After removal of the mononuclear layer using a pipette, the cells were washed in PBS and re-pelleted. Pelleted cells were re-suspended in 10ml ammonium chloride (Stem Cell Technologies Cat 07850) and incubated at 4^{o}C for 10 minutes to induce red cell lysis. Lysis was terminated with the addition of 1ml FBS, the sample pelleted and re-suspended in 10ml 2% PBS. Cell number was counted and viability assessed using the Beckmann Coulter Vicell XR Cell Viability Counter.

If cells were frozen at this juncture, a freezing mix was made, consisting of 95% FBS to 5% DMSO. Cells were frozen to a maximum concentration of 100x10^{6} cells/ml.

2.2.3.1 Lineage Depletion of GMPB or cord samples to enrich for the presence of HSCs

Samples of cord or GMPB can be enriched for the presence of immature cells using negative selection for cells expressing lineage specific antigens (Lin depletion). This technique utilises the EasySep Human Progenitor Cell Enrichment Kit (Stem Cell Technologies; Cat: 19056) and the EasySep magnet (Stem Cell Technologies; Cat: 18000).

Unselected mononuclear cells were pelleted and resuspended at a concentration of $100x10^6$ cells/ml in 2% PBS, to a maximum volume of 2ml in 5ml polystyrene tubes (BD Falcon, Cat: 352052). EasySep Human Progenitor Cell Enrichment Cocktail (Stem Cell Technologies, Cat: 19056) was added at a concentration of 50µl/ml; the suspension was mixed and incubated (15 minutes, room temperature). EasySep Magnetic Nanoparticles (Stem Cell Technologies, Cat: 19056) were left at room temperature for at least 15 minutes, and mixed thoroughly to ensure even distribution. The nanoparticles were then added at a concentration of 50µl/ml to the tube, mixed and left to incubate (15 minutes, room temperature).

2% PBS was then added to the polystyrene tube to bring the total tube volume to 2.5 ml, and the tube gently mixed. The tube (without cap) was placed in the EasySep Magnet (Stem Cell Technologies, Cat: 18000), and left for 10 minutes at room temperature. Leaving the tube in the magnet, the contents of the tube were then decanted into a 50ml tube. The magnetic beads, bound to unwanted, Lin-expressing cells, remained in the polystyrene tube.

The yield of Lin-negative immature cells was increased as follows. The 5ml polystyrene tube was removed from the magnet, and the cells and beads re-suspended by adding 2.5ml 2% PBS. The tube was replaced in the magnet, incubated for 10 minutes at room temperature, and contents decanted into a 50ml Falcon tube. This was repeated twice.

Cell number and viability was then assessed using the Beckmann Coulter Vicell XR Cell Viability Counter.

2.2.4 Tissue Bank storage of bone marrow serum samples

Bone marrow samples were collected into SST tubes (multiple manufacturers) and centrifuged (3000rpm, 3 minutes). The resultant serum supernatant was aliquoted off, and frozen at -20^{o}C.

2.2.5 Cell Thawing

On removal from liquid nitrogen storage, cells were rapidly thawed at 37^{o}C in a water bath for one to two minutes. In order to minimise cell loss through cellular clumping, 500µl DNase was added immediately post-thawing, and the vial left for 2 minutes at room temperature. The vial contents were immediately diluted in 20ml 2% PBS, to limit the toxicity from DMSO. The cell suspension was then centrifuged (1500rpm, 5 minutes) and re-suspended to the required volume.

2.3 Cell Counting

2.3.1 Automated Assessment of Cell Numbers

Automated assessment of cell number and viability was performed using the Beckmann Coulter Vicell XR Cell Viability Counter. This is a video imaging system, which automates assessment of the ability of live cells to extrude trypan blue dye. The cell numbers in 30 fields were assessed, and an average calculated. The machine is calibrated to assess cell numbers most accurately at concentrations between $0.5x10^6$ and $100x10^6$ cells/ml, so samples were diluted appropriately before analysis.

2.3.2 Manual Counting of Cell Numbers

For low cellular concentrations, below the resolution of automated counting, manual assessment of cell concentration using a haemocytometer was performed. A 10µl aliquot of cells was added to 90µl trypan blue (P2100102, Cell Viability Inc), and mixed well, to produce a 1:10 dilution. 10µl of this mixture was injected into the top of a 1/400 mm^2 haemocytometer (Weber Scientific, Cat: 3048-12), under a coverslip. Using a microscope, cell numbers were counted in each of the four large corner squares of the haemocytometer and a mean taken. A mean cell number of α calculated using this technique, gives a cellular concentration (post-dilution) of $\alpha x10^4$ cells/ml; and therefore pre-dilution with trypan blue, a cellular concentration of $\alpha x10^5$ cells/ml.

2.4 Flow Cytometry

2.4.1 Background

Flow cytometry is a technique designed to measure the characteristics of single cells flowing through a detector system. Fluorescent antibodies are used to identify the expression of specific cell surface or internal glycoproteins. Single cell suspensions are introduced to the cytometer in a cell-free buffer called the sheath fluid, which flows towards a laser. Laminar flow of the sheath fluid forces the cells to line up in single file as they approach the laser. The bound antibody-fluorophore absorbs the laser energy and subsequently releases it in the form of a specific wavelength of light as cells pass through the laser. The emitted light is detected by an optical system sensitive to various wavelengths, allowing information on multiple surface markers to be read simultaneously. The specificity of detection is controlled by optical filters, which block certain wavelengths while transmitting others. Photodetectors convert the light signal into a current, whose voltage has an amplitude proportional to the total amount of light received. These data are then plotted graphically.

Basic information on cell size is represented by the forward scatter parameter (FSC), quantifying light scattered at a small angle and detected by a sensor on the opposite side of the 488-nm blue laser. Light that scatters off the cell at a 90^{o} angle is called side-scatter (SSC), and reflects cell granularity.

Fluorochromes are dyes, which accept light energy at a given wavelength (in this case from a laser) and re-emit it at a longer wavelength, processes called excitation and emission. At a basic level, a good fluorochrome will absorb and emit light at significantly different wavelengths, a phenomenon known as the Stokes shift, which will enable the machine to clearly differentiate emitted light, from the unabsorbed light emitted by the laser itself.

A variety of fluorochromes are commercially available that individually emit light of specific different wavelengths, whilst absorbing light of the same wavelength. This enables us to study a number of different parameters simultaneously, which is a key principle of the work in Chapters 4 and 5. An ideal fluorochrome has an extremely narrow emission spectrum, read by only one detector, but in practise, this does not occur in a system adapted for multi-colour analysis. Because the emission spectra of fluorochromes overlap to different degrees, a single detector may see fluorescence originating from more than one fluorochrome; this spill-over of readings from other fluorochromes in the detectors not assigned to them is removed using an algorithm called compensation. Cells stained with a single fluorochrome are run through the detector separately, and the signal detected in the channels not assigned to them is noted and removed from the final analysis.

2.4.2 Fluorescence Activated Cell Sorting (FACS)

Fluorescence Activated Cell Sorting (FACS) uses the principles of flow-cytometric based cell labelling to separate cells of different types.

Labelled single cell suspensions are introduced to the FACS machine, and laminar flow of the sheath fluid ensures the cells approach the lasers individually. After cells are interrogated by the laser in the flow chamber, the single cell stream is broken accurately into tiny droplets by a fine nozzle vibrating at ultrasonic frequency. The scatter and fluorescence signal is compared to the sort criteria set on the instrument. If the particle matches the selection criteria, the fluid stream is charged as it exits the nozzle of the fluidics system.

The speed of flow sorting depends on several factors including particle size and the rate of droplet formation.

2.4.3 Specific Uses of Flow Cytometry in this Thesis

Flow cytometry has been used heavily in this project for several purposes: to quantify HSCs in diagnostic AML samples; to characterise the expression of a variety of antigens both on the surface and cytoplasm of stem cells; to identify stem cells prior to sorting, and to identify erythroid progenitors in bone marrow samples.

The number of cells stained varied dependant on the individual experiment, and the rarity of the cell type under investigation. In general, when working with valuable bone marrow samples, all available thawed cells were stained, and the whole sample analysed.

All flow cytometry analysis was performed on a BD LSR Fortessa, equipped with 405, 488, 561 and 641 nm lasers**.**

2.4.4 Fluorochrome Panel Selection

Combinations of fluorochromes were always optimised before performing multicolour flow cytometry. Up to seven colours were used simultaneously in some experiments. The particular combinations utilised in this study were chosen based on local experience of best discrimination of the rare stem cell phenotype, in combination with knowledge of the optimal stains for the machine which we use in the lab. We attempted to pick the brightest colours for rarely expressed antigens, and to minimise spill-over between important discriminating antigens. For commonly studied antigens (i.e. CD34), a range of fluorochromes were commercially available to select from, whereas others presented a more restricted selection (i.e. ALDH activity, as assayed by the Aldefluor^{TM} reagent kit, can only be assessed in the FITC channel).

2.4.5 Specific Techniques

2.4.5.1 Surface Staining

All staining procedures on live cells were undertaken at 2 to 8^{o}C to avoid antibody capping and internalisation. For any staining procedure involving more than one antibody, the cells were pre-stained with a solution of human γ globulin (at a concentration of 50µl 2% HAG/1.5x10^{6} cells) for 20 minutes at 4^{o}C to reduce non-specific binding of antibodies. Each antibody was optimised to determine the best volume for surface staining, which ranged from 1 to 20µl/10^{6} cells. The samples were then incubated for 30 minutes in the dark. The cells were then washed in 4ml 2% PBS and re-suspended in 300µl 2% PBS/DAPI to allow for live/dead cellular discrimination. Cells were kept in the dark until analysis to reduce fluorochrome bleaching.

2.4.5.2 Intracellular Assessment of Aldehyde Dehydrogenase (ALDH) Activity

Internal ALDH activity was assessed by the Aldefluor™ reagent kit. The activated reagent is a fluorescent substrate for ALDH, which freely diffuses into viable cells. In the presence of ALDH, the reagent is converted into a substrate which is internalised within cells. The amount of fluorescence produced is proportional to the ALDH activity within cells and is measurable in the FITC channel of a cytometer. Active efflux of the reaction product is inhibited by an efflux inhibitor in the Aldefluor™ assay buffer. Cell staining was optimised at a concentration of 5µl Aldefluor™ reagent per $1x10^6$ cells, prior to incubation for 60 minutes at 37°C in a water bath. To minimise efflux, all further handling of cells was performed at 4°C, and analysis was undertaken as rapidly as possible (normally within two hours of Aldefluor™ staining). Further washing and suspension of cells was undertaken in Aldefluor™ buffer rather than PBS, for the same reason. Similarly, when using HAG to minimise non-specific antibody binding, optimisation experiments showed that the cells remained $ALDH^{bright}$ if HAG diluted in Aldefluor™ buffer was used, rather than with 2% PBS as standard.

2.4.5.3 Intracytoplasmic Staining

Any intracellular staining technique involves membrane permeabilisation to allow the antibodies to pass into the cytoplasm. In this context, DAPI cannot be used for live/dead discrimination of cells, as it normally stains dead cells by passing through non-intact membranes, and binding AT rich portions of DNA. Live/dead discrimination is therefore performed using a fixed viability dye (Fixed Viability Dye eFluor 780 (E Biosciences, Cat: 65-0865)), before surface staining.

For $1x10^6$ cells, cells were re-suspended in 1ml of diluted Fixed Viability Dye (1µl Fixed Viability Dye, premixed with 1 ml PBS), left to incubate at 4°C for 30 minutes, and then washed in 2% PBS.

Cell fixation and permeabilisation was then undertaken using the BD Fixation/Permeabilization Solution Kit (BD, Cat: 554714). Cells were resuspended and membranes permeabilized by the addition of 100µl Cytofix/Cytoperm per $1x10^6$ cells, followed by incubation for 15 minutes at room temperature. All subsequent washing of cells was performed using 1ml PermWash buffer per $1x10^6$ cells.

Internal blockade of non-specific binding was performed by the addition of a goat serum/glycine mix (50µl per $1x10^6$ cells of 10% normal goat serum, 0.3M glycine) to cells for 20 minutes at 4°C. Internal primary and secondary antibody staining was performed at room

temperature, with incubation for 30 minutes, followed by careful washing to remove unbound antibody.

Extracellular surface antigen staining was performed after the completion of internal staining as previously described, albeit with the continued use of PermWash buffer for the wash steps.

2.4.5.4 Controls

Unstained controls were used to set the voltages in the flow cytometer. Isotype controls were used in experiments where there was not expected to be a significant population of target antigen negative cells. Fluorescence minus one (FMO) controls were also used for certain experimental conditions.

Compensation was set using single stained controls, where the cells were known to express the target antigens. For experiments utilising fixed cells, the compensation cells also underwent fixation before staining.

2.4.5.5 FACS

All cell sorting for these experiments were performed on a BD FACSAria II^{TM}, supplied with 405, 488, 561 and 641 nm lasers. Sorting tiny numbers of cells from large samples involves a number of challenges to optimise sort purity.

Live cells were stained as described, and resuspended in a stock solution of 2% PBS/DAPI/DNase at a concentration of 8x10^6 cells/ml. This concentration had previously been optimised by our group as appropriate for sorting out cells of rare phenotype from AML samples. DNase was added to the pre-sort suspension to reduce cellular clumping, and for the same reason cells were filtered immediately prior to sorting to remove cellular debris (Cell Strainer Cap, BD Falcon, Cat: 352235), to reduce the chance of nozzle blockage. An 85µm nozzle was used for all sorts, and a slow flow rate of between 10 to 30µL/minute was used at all times.

To increase cell survival, cells were sorted into polypropylene tubes (BD Falcon, Cat 352063) containing 2% PBS, in order to remove the electrostatic charge from the sorted droplets prior to further processing.

Cells stained with ALDH reagent were sorted into ALDH buffer media to reduce ALDH efflux prior to purity assessment.

The purity of the sorted population was always checked immediately post sort, irrespective of cell numbers collected. Between sorting and purity checked, the machine was flushed for a minimum of five minutes to reduce cellular debris. If the purity of the sorted population was found to be below 90%, resorting was undertaken.

Sorted cells were then pelleted (1500rpm, 10 minutes) and the supernatant removed using a Gilson Safe Aspiration Centre.

2.4.5.6 Flow Cytometric Analysis

All analysis was performed using the FlowJo v 10 package. Figure 2-1 illustrates the gating strategy applied to the test samples. A first gate defined live cells from debris on the basis of FSC and SSC. A second gate excluded doublets from the analysis, on the basis of SSC-A and SSC-width. A third gate was applied on the basis of DAPI/fixed viability dye and SSC-A to exclude dead cells from the analysis. Therefore all subsequent analysis of fluorochrome intensity was performed on compensated, live-singlet cells.

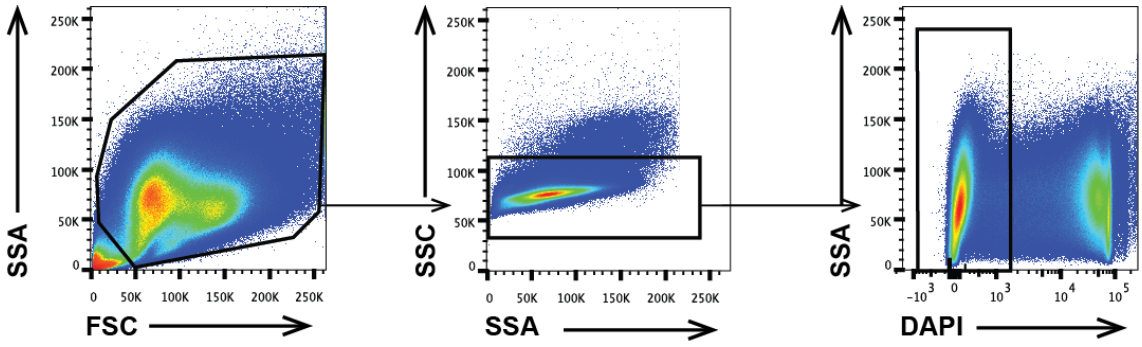

Figure 2-1 Illustrating the successive flow gating strategy used to select single, live cells

2.5 Tissue Culture Methods

2.5.1 MS5 Culture

Unless otherwise stated, all liquids were pre-warmed to 37^{o}C in a water bath prior to cell contact. To reduce the chances of cell-line infection, antibiotics were added to all media to a concentration of 1% (Streptomycin/Penicillin (Sigma, Cat: P4333)).

Cells from the MS-5 cell line (DSMZ, Cat ACC-441) were initially thawed, washed and resuspended in 20ml of 10% IMDM. Cells were pipetted into a flask (75cm^{2} Cell Culture Flask, Coning Cat: 430720U), and incubated at 37^{o}C, 5% CO2. MS-5 cells are surface-adherent, and therefore flasks were placed flat in the incubator. Flasks were inspected daily for cellular appearance, and split before a confluence of 80% was reached. To split cells, the adherent cell layer was washed with PBS twice, to remove traces of FBS. 3ml of trypsin-EDTA (Sigma, Cat:

T4299) was added to the flask, and left to incubate at 37^{o}C for 4 minutes. Flask contents were visually inspected to confirm loosening of the adherent stromal cells, and further activity of the trypsin inactivated by the addition of 10ml 10% IMDM. Cells were centrifuged (1300rpm, 5 minutes), resuspended in 10% IMDM, and split into the required number of flasks for ongoing culture.

MS-5 cells were kept in culture for a minimum of 2 weeks from thawing prior to their use as a feeder layer in the LTC-IC assay. Cell lines were not passaged indefinitely: after two months in continuous culture, cells were discarded, as local experience had shown their effectiveness as a feeder line to be reduced over repeated passaging (personal communication, Dr Anjos-Afonso).

If required, cells were frozen to a concentration of 5x10^{6} cells/ml in freeze vials in a mixture of 19 parts FCS to 1 part DMSO.

2.6 In vitro assays of stem cell function

2.6.1 Methylcellulose Assay (for the presence of CFU)

Bone marrow samples were thawed, counted, stained and sorted (under sterile conditions) as described above. Cells were centrifuged (1500rpm, 10 minutes), and the supernatant aspirated using the Gilson Safe Aspirator Station, leaving a small volume (approximately 50µl) of residual fluid in the tube. The pelleted cells were resuspended, and transferred to a tube containing 4ml Methylcellulose (MethoCult H4435 enriched, Stem Cell Technologies, Cat: 04435). Cells were carefully mixed throughout the methylcellulose using a standardised technique with a 1ml pipette and a 16G blunt hub needle (Monoject Blunt Needle, Kendall, Cat: 8881202322), and left to settle for 2 minutes. 1ml of the cells and methylcellulose mixture was transferred using the same syringe and needle to a small culture dish (Culture Dishes for Optimal Colony Growth in Methylcellulose Based Assay, Stem Cell Technologies, Cat: 27150), trying to avoid the introduction of bubbles. Plates were set up in triplicate for each experimental condition. Into one, larger 10cm diameter tissue culture dish were placed two small culture dishes containing methylcellulose, along with one open small culture dish containing 6ml sterile water to maintain an appropriate humidity. The plates were then incubated at 37^{o}C, 5% CO_2 for two weeks.

After two weeks, the myeloid and erythroid colonies visible under the microscope were counted using a graticule. Plates with visible colonies were then re-plated to assess growth over a further two weeks. For test conditions which had produced more than five colonies per

plate, one of the three available replicated plates was used. If there were five or less colonies per plate, colonies from all three plates were pooled. Using a 1ml pipette and 16G blunt needle, cells and methylcellulose were removed, and placed into a 50ml Falcon tube. The plate was washed with 1ml 2% PBS, and the washings added to the Falcon tube. This was repeated a further two times. The Falcon tube was topped up to a volume of 50ml with 2% PBS, centrifuged for 1500rpm for 5 minutes, and the supernatant discarded. The pelleted cells were washed again with 50ml 2% PBS, pelleted and re-suspended in 500µl PBS and counted using a Haemocytometer. 40,000 cells were then added to a tube containing 4ml methylcellulose, mixed well and re-plated as described above. The re-plated colonies were then incubated for a further two weeks.

At the end of two further weeks culture, cell colony numbers were assessed. Cells were removed from the culture medium and washed as described, prior to placing onto slides for FISH based analysis.

2.6.2 Long Term Culture Initiating Cell Assay

A 96 well flat bottom cell culture plate (Costar, Cat: 3598) was coated with collagen using the following method. 50µl of Collagen Solution (Stem Cell Technologies, Cat: 04902) was added to each well, and left to dry for two minutes before the excess collagen was removed. The plate was left open under a hood to dry for at least an hour. Subsequently, each well was washed with 100µl PBS.

Meanwhile, MS-5 cells were prepared at a concentration of $1x10^6$ cells/ml in IMDM/10%FCS, and placed in a 50ml Falcon Tube. The cells were irradiated at a dose of 70Gy (RadSource 2000, X-ray emitter). Cell viability 1 hour post-irradiation was assessed using a Beckmann Coulter Vicell XR Cell Viability Counter. Cells were pelleted, and re-suspended in MyeloCult™ H5100 (Stem Cell Technologies, Cat: 05100) to a concentration of 125,000 cells/ml. 100µl of cell/H5100 suspension was added to 60 wells per plate, having filled each of the outside wells with 100µl water.

The following day, HSCs were sorted as previously described. Cells were pelleted (1500rpm, 10 minutes), diluted and distributed at different concentrations between wells in triplicate.

The plates were then placed in a long term incubator at 37°C, 5% CO_2. Weekly, for five weeks of culture, the media was changed as followed. Taking care to not disturb the bottom of each well, 65µl of media was removed from the top of each well using a multichannel pipette. This was replaced with 75µl of MyeloCult™ H5100.

At the end of week 5 of culture, the media was removed from each well, using a Gilson Safe Aspiration Station, having taken care not to disturb the 10 to 20µl at the bottom of each well containing cells. Using a Gilson Distriman pipette, 100µl of MethoCult H4435 enriched Methylcellulose medium was added to each well, and the plates returned to the incubator. After a further two weeks culture at 37^{o}C, each individual well was examined for the presence of myeloid and erythroid colonies and their numbers recorded. If growth was noted, cells were removed, washed and placed on slides for FISH analysis as described in Section 2.7.1

2.7 Histological techniques

2.7.1 Preparation of Slides

If greater than 5000 cells were collected, slides were made as follows. Cells were pelleted and re-suspended in 150 to 600µl 2% PBS, dependant on the number of slides required. Cell numbers did not exceed 500,000 per slide, to prevent overcrowding. Using a clean glass slide clipped into a slide holder, 150µl of 2% PBS/cell suspension was added to the well. Slides were spun (8500rpm, 6 minutes) using a Stat Spin Cytofuge 2. Excess liquid was removed using a Gilson Safe Aspirator Station, and the slides allowed to air-dry.

If handling less than 5000 cells, the cells were resuspended after initial centrifugation in 50µl 2% PBS. This was removed using a pipette, and transferred onto a dry slide. The slides were air-dried without spinning.

2.7.2 Immunocytochemistry and Microscopy

2.7.2.1 Background

Immunocytochemistry involves the use of fluorescent antibodies to identify the presence and location of particular antigens in a cell, when studied using a laser microscope designed to detect the fluorescence. There are two forms of immunocytochemistry: direct and indirect. Only indirect methods, which involve consecutive antibody staining were used in this project, and so will be described below.

A primary antibody specific to the antigen of interest is applied to a specimen. After binding, its presence is revealed by staining with a secondary antibody against the immunoglobulin of the species used for primary antibody production. Primary antibodies are generally unlabelled, whereas secondary antibodies contain a tag for detection. The use of a preliminary blocking step prevents non-specific binding of antibodies. The choice of blocking buffer is based on the secondary antibody used. Dissolving the blocking agent in saponin increases the cells' permeability.

2.7.2.2 Methodology

After drying, slides were fixed and permeabilized. 50µl Cytofix/Cytoperm Buffer (BD, Cat: 554722) was applied to the slide, and left to incubate in a damp box for 30 minutes. Slides were rinsed in PBST (0.1% Tween-20 (Sigma P1379) in PBS).

Slides were then blocked with donkey serum (Sigma, Cat: D9663) (0.1% saponin, 5% donkey serum in PBS). 50µl of 5% donkey serum was added to slides, which were left to incubate for 30 minutes in a damp box. Slides were then rinsed in PBST.

Primary antibody concentration was optimised, based on manufacturers' instructions. 50µl of antibody was applied to slides, which were then incubated for 1 hour. Slides were washed twice, in between 5 minute agitation periods on a Stuart Scientific Platform shaker STR6.

Secondary antibody staining was performed using a labelled monoclonal antibody against the species used in the primary antibody. Secondary antibody concentration was also optimised for each experimental set up. 50µl of secondary antibody was added to slides, which were left to incubate for 1 hour in a damp box. Slides were then washed three times with PBST as described above. The slides were covered during the washing process to reduce light exposure.

Slides were then mounted with 1 drop of Vectashield Hardset™ Mounting Medium in Dapi (Vector, Cat: 1500), and a coverslip added. Examination was performed using the Confocal Zeiss Laser Scanning Microscope SM 5100.

2.7.3 FISH Staining

I am extremely grateful for the assistance of Marianne Grantham of the Cytogenetics department, Royal London, who performed all the FISH based staining and analysis detailed in this thesis. Her technical expertise was invaluable, as was her assistance in probe selection to detect less common karyotypic abnormalities. Her patience in dealing with the considerable challenges of FISH staining of small numbers of cells was also much appreciated.

FISH (Fluorescence In Situ Hybridisation) describes a technique where fluorescent probes bind to complementary sequences of DNA or RNA within cells. Subsequent examination with fluorescent microscopy can then be used to detect the presence of the probes. FISH based analysis of bone marrow, using probes specific for commonly found translocations or mutations, is part of the routine AML diagnostic process.

In brief, probes are added to cells fixed onto slides, and are left to hybridise for 12 hours, before the removal of unbound probe by washing. For DNA binding probes to anneal to complimentary sequences, the chromosomal material inside the cells has to be accessible (cells either in interphase or metaphase).

Table 2-1 FISH probes used in this thesis with expected signal patterns in normal and mutated cells

PROBE NAME	MUTATION DETECTED	PROBE DESIGN	NORMAL CELL SIGNAL	MUTATED CELL SIGNAL
Cytocell CBFB-MYH11 (LPH 022)	Inv(16)	DUAL COLOUR, DUAL FUSION	2 RED, 2 GREEN	2 FUSION, 1 RED, 1 GREEN
Cytocell PML-RARA (LPH 023)	t(15;17)	DUAL COLOUR, DUAL FUSION	2 RED, 2 GREEN	2 FUSION, 1 RED, 1 GREEN
Cytocell AML1/ETO (RUNX1/RUNX1T1) (LPH 026)	t(8;21)	DUAL COLOUR, DUAL FUSION	2 RED, 2 GREEN	2 FUSION, 1 RED, 1 GREEN
Kreatech DEK/NUP 214 t(6;9)	t(6;9)	DUAL COLOUR, DUAL FUSION	2 RED, 2 GREEN	2 FUSION, 1 RED, 1 GREEN
Cytocell MLL (KMT2A) (LPH 013)	+11	DUAL COLOUR, BREAK APART	2 FUSION	3 FUSION
Cytocell Del(7q) (LPH 025)	Del(7)	DUAL COLOUR	2 RED, 2 GREEN	1 RED, 2 GREEN
Cytocell BCR/ABL/ASS1 (LPH 038)	Trisomy 9q34	TRIPLE COLOUR,DUAL FUSION	2 RED, 2 AQUA	3 RED, 3 AQUA

A variety of different probe designs exist. Fusion probes detect translocations which bring two genes of interest together. They normally involve the use of two different coloured probes, which when brought together in close physical proximity, produce a third colour visible by eye. Break-apart probes are designed to bind sequences which under normal circumstances are found close together. If a translocation separates the two regions, the probes produce two distinct colours.

Table 2-1 details the probes used in this study, as well as the specific changes they were used to detect and the expected signal patterns in normal and mutated cells. Some probes were used for their standard commercial purpose i.e. the Cytocell CBFB-MYH11 probe for the detection of inv(16). Imaginative use of other probes allowed identification of other, rarer abnormalities. For example, the MLL break-apart probe, which normally binds to chromosome 11, was in this case used not to detect mutations within the MLL gene specifically, but trisomy 11.

2.8 Genetic Techniques

2.8.1 DNA Extraction modified for small cell numbers

DNA extraction from small populations of cells was performed using the QIAamp DNA Blood Mini Kit (Qiagen, Cat: 51104), designed for purification of up to 12µg DNA. All components listed below are included within this kit.

After washing, and pelleting of cells (1500rpm, 10 minutes), the supernatant was removed using the Gilson Safe aspiration centre. Cells were gently resuspended in 100µl PBS and 100µl AL lysis buffer (Cat: 1014594), and transferred into a 1.5ml Eppendorf tube. After vortexing for 30 seconds, 40µl Proteinase K (0.1mg/ml) (Cat: 1045166) was added, the tube briefly vortexed and then put to incubate for 10 minutes at 70^{o}C. 210µl of 100% Ethanol (Sigma, Cat: E7023) was then added, the tube vortexed and 500µl of the solution added to the spin column. The column was spun (8000rpm, 1 minute), and the supernatant discarded. The spin column was placed in a clean collection tube, and 500µl of wash buffer AW1 (Cat: 1014790) added, before centrifugation (8000rpm, 1 minute). The collection tube was again replaced, and a further 500µl of wash buffer AW2 (Cat: 1014592) added, prior to centrifugation (14000rpm, 3 minutes).

The spin column was then placed in a clean 1.5ml eppendorf tube, and between 20 and 50µl of AE (Cat:1014574), pre warmed to 45^{o}C, was added (volume added is dependent on initial cell numbers collected: for samples of between 100 to 5000 cells, 20µl of AE was added; for samples up to 50,000 cells, 50µl of AE was added). The tube was then left to incubate at room temperature for 5 minutes, and then centrifuged (8000rpm, 1 minute). The resultant DNA concentration was measured using the Nanodrop ND-1000 Spectrophotometer, prior to PCR.

2.8.2 RNA extraction modified for small cell numbers

Cells for RNA extraction were pelleted (3000rpm, 10 minutes, 4^{o}C), and the excess supernatant removed using the Gilson Safe Aspirator Station. Cells were kept on ice during RNA extraction, which was performed under a hood to reduce contamination of samples. RNA extraction was performed using the RNeasy MicroKit, designed for extraction of up to 45µg RNA (Qiagen, Cat: 74004).

To each sample, 75µl of buffer RLT (Cat: 1015750) (premixed in a ratio of 100:1 with 6-Mercaptoethanol (Sigma, Cat: M6250)), was added and the mixture vortexed. The liquid was then added to a QIAshredder spin column (Qiagen Cat: 79654), and centrifuged (13,000rpm, 2

minutes). An additional 75µl of RLT/6-Mercaptoethanol was added to the spin column and centrifugation repeated.

150µl of 70% Ethanol was then added to the filtrate, before transfer to an RNeasy MinElute spin column (Cat: 1026497), and repeat centrifugation (8000rpm, 15 seconds). The flow-through was then discarded. 350µl of buffer RW1 (Cat: 1014567) was then added to the column, centrifuged (13,000rpm, 15 seconds) and the flow-through discarded. This step was repeated to reduce 280nm contamination. A mixture of 10µl DNase I stock Solution and 70µl Buffer RDD (Cat: 1023460) was added to the top of the column, and left to incubate (room temp, 15 minutes). 350µl of buffer RW1 was then added to the column, before centrifugation (13000rpm, 15 seconds). The spin column was then placed in a fresh 2ml collection tube, and 500µl of buffer RPE (Cat: 1017974) added to the column, before centrifugation (13,000rpm, 15 seconds). The flow-through was discarded, before the addition of 80% ethanol to the spin column, which was then centrifuged (13,000rpm, 2 minutes). The spin column was then placed in a clean, dry collection tube, and centrifuged (13,000 rpm, 5 minutes) with the lid open. The spin column was then placed in a 1.5ml collection tube, 14µl of RNase free water (Cat: 1017979) warmed to 45^{o}C was added to the spin membrane, and incubated at room temperature for two to three minutes. Elution of the RNA was then performed by centrifugation (13,000rpm, 1 minute). The resultant RNA was then stored at -80^{o}C before subsequent use.

2.8.3 DNA sequencing (for the *NPM1* gene)

2.8.3.1 PCR for NPM1 Exon 12

Cells for sequencing were pelleted and DNA extracted as described in Section 2.8.1. The resultant DNA concentration was measured using the Nanodrop ND-1000 Spectrophotometer, prior to PCR.

Standard PCR for the *NPM1* gene Exon 12 was performed as follows. DNA was diluted to a concentration of 100ng/µl with AE solution. Forward and reverse primers with the following sequences were ordered from Sigma- *NPM1* Forward (F) Primer: 5' cttaaccacatttcttttttttttttttccag 3'; *NPM1* Reverse (R) Primer: 5' ggacaacatttatcaaacacggtag 3'.

Master mix was made up in the following proportions for each sample: 20µl Reddy mix PCR MasterMix (ThermoScientific, Cat: AB-0575/LD/A); 1µl F Primer (10µM); 1µl R primer (10µM); 2µl Water. 1µl of DNA (100ng/µl) was added to 24µl of master mix, and the sample mixed by pipetting.

A negative control was also made up containing Forward and Reverse primers, ReddyMix and 3µl water. The DNA and master mix solution was placed into small PCR tube, and run on the following PCR cycle (using MJ Research PTC 225 Peltier Thermal Cycler): 95^{o}C for 5 minutes; 94^{o}C for 30 seconds, 62^{o}C for 30 seconds, 72^{o}C for 1 minute for a total of 40 cycles; 72^{o}C for 10 minutes; hold at 12^{o}C.

2.8.3.2 Confirming the Production of a PCR Product

The PCR product was run on an agarose gel to check for the presence of a band prior to sequencing. A 2% agarose gel was first made up as follows. 1g agarose (Ultrapure^{TM} Agarose, Invitrogen, Cat: 15510-027) was dissolved in 50ml TBE buffer by heating in a microwave for 6 minutes until completely dissolved. 5µl of Gel Red Nucleic Acid Stain (Biotium, Cat: 41003-1) was added to the gel, which was then poured into the plate mould. Bubbles were removed, and a comb inserted into the liquid gel, which is then allowed to set at room temperature.

When dry, the plate was placed in a horizontal electrophoresis gel tray, and TBE was added until the gel was completely covered with buffer. The comb was removed, and 5µl of samples were added to the wells. 5µl of a DNA ladder (ExatGene 100bp PCR DNA ladder (Fisher Scientific, Cat: BP2571100)) was added to either end of the gel. The gel plate voltage was set to 95V, and the plate left for 30 minutes at room temperature. The presence of a band around the 198bp weight suggested a positive PCR reaction, and allowed us to proceed to DNA purification and sequencing.

2.8.3.3 DNA Purification

DNA purification was performed using the Montage PCR Centrifuge Filter Device (Merck Millipore, Cat: P36461). 350µl of Millidrop purified water was added to the device column, followed by the PCR product to be purified. The column was centrifuged (1000g, 15 minutes) at room temperature. 20µl of water was then added to the column membrane; the column was inverted and placed into a clean eppendorf tube, which was then centrifuged (1000g, 2 minutes). The resultant DNA concentration in the eppendorf was measured using the Nanodrop ND-1000 Spectrophotometer, and the DNA stored at -20^{o}C.

2.8.3.4 Preparation for sequencing (addition of sequencing primer)

Sequencing of the purified product was then performed as follows, having diluted the DNA product to a concentration of 10ng/µl.

A master mix was produced by mixing the following in the given proportions (sufficient for 12 samples): 6µl BigDye Terminator v3.1 (Applied Biosystems, Cat: 4337454); 45µl Big Dye

terminator Buffer (x5) (Applied Biosystems, Cat: 4337454); 2µl *NPM1* Reverse primer (sequence as in Section 2.8.3.1); 175µl Water.

For each sample, 1µl of DNA was mixed with 19µl master mix, before running the following PCR cycle: 96^{o}C for 1 minute; 96^{o}C for 30 seconds, 50^{o}C for 15 seconds, 60^{o}C for 4 minutes for a total of 25 cycles; hold at 12^{o}C.

2.8.3.5 Sequencing

The amplified PCR product was then taken to Sequencing Laboratory at the London Research Institute for sequencing by capillary electrophoresis. The protocol for sample processing followed there is described in brief below.

The PCR product was loaded onto a 96 well plate, and cleaned in an automated process, performed on a Beckman Coulter Biomek FX machine. PCR products were positively selected for by adherence onto charged magnetic beads (Chemagic SEQ Pure Kit). The beads were then washed twice in buffer, and the cleaned DNA eluted off the beads by washing in Millipore water.

The plates were then transferred to a 96-capillary Applied Biosystem 3730XL DNA Analyser, where the DNA strands were sequenced by capillary electrophoresis overnight.

2.8.3.6 Analysis

Software analysis was performed using the BioEdit Sequence Alignment Editor. Commonly found mutations in the *NPM1* gene are detailed in Section 1.4.5.4.1.

2.8.4 Quantitative Real Time Polymerase Chain Reaction (qPCR)

2.8.4.1 Background

The real-time polymerase chain reaction is a laboratory technique based on the polymerase chain reaction (PCR), which is used to amplify and simultaneously detect or quantify a targeted DNA molecule. For assessment of gene expression as measured by mRNA transcript numbers, an RNA sample is first reverse-transcribed to complementary DNA (cDNA) with reverse transcriptase, before qPCR is undertaken.

There exist two common methods for the detection of products in quantitative PCR: non-specific fluorescent dyes that intercalate with any double-stranded DNA; and sequence-specific DNA probes, consisting of oligonucleotides that are labelled with a fluorescent reporter, which permits detection only after hybridization of the probe with its

complementary sequence to quantify messenger RNA (mRNA) and non-coding RNA in cells or tissues. Both have been utilised in different experiments in this thesis.

DNA-binding dyes (such as SYBR Green) bind to all double-stranded (ds) DNA in a PCR reaction, causing fluorescence of the dye. An increase in DNA product during PCR therefore leads to an increase in fluorescence intensity, which is measured at each PCR cycle. By nature of their design, these probes will also bind to non-specific PCR products (such as primer-dimers), which can potentially interfere with accurate quantification of the intended target sequence.

The fluorescent reporter probe method uses a DNA-based probe with a fluorescent reporter at one end and a quencher of fluorescence at the opposite end of the probe. The close proximity of the reporter to the quencher prevents detection of its fluorescence; breakdown of the probe by the 5' to 3' exonuclease activity of the Taq polymerase breaks the reporter-quencher proximity and thus allows unquenched emission of fluorescence, which can be detected after excitation with a laser. An increase in the product targeted by the reporter probe at each PCR cycle therefore causes a proportional increase in fluorescence due to the breakdown of the probe and release of the reporter**.**

Fluorescent reporter probes detect only the DNA containing the probe sequence; therefore significantly increasing detection specificity, and enabling target quantification even in the presence of non-specific DNA amplification. Fluorescent probes can be used in multiplex assays, based on specific probes with different-coloured labels, provided that all targeted genes are amplified with similar efficiency. The specificity of fluorescent reporter probes also prevents interference of measurements caused by primer-dimers. However, fluorescent reporter probes do not prevent the inhibitory effect of the primer-dimers, which may depress accumulation of the desired products in the reaction.

2.8.4.2 SYBR Green analysis

qPCR reactions using SYBR green regents were set up as follows. A master mix was made to a total volume of 9µl per well as follows: 6µl Sso Advanced Universal SYBR Green Supermix (BioRad, Cat: 7725274); 0.125µl F primer (10µM); 0.125µl R primer (10µM); 2.75µl H_2O. 9µl of master mix was added to each well, followed by 1µl of cDNA, to a total well volume of 10µl. Reactions were set up in triplicate, along with appropriate negative controls. 384 well, clear white wells (BioRad HSP 3805) were used and sealed with compatible seals (Microseal B Seals, BioRad MSB1001). After sealing, the plate was centrifuged (1500rpm, 1 minute).

The following analysis cycle was run on the BioRad C1000 Touch Thermal Cycler: 95°C for 2 minutes; 95°C for 5 seconds, 60°C for 30 seconds for a total of 40 cycles; melt curve 65°C to 95°C, with 0.5°C increments every 5 seconds.

2.8.4.3 Taqman assays

qPCR reactions using commercially available Taqman probe and primer sets were set up as follows.

A master mix was made up in the following proportions for each sample: 1µl 20xProbe/Primer set; 10µl Taqman 2xPCR Master mix (Thermofisher, Cat: 4304437) and 7µl H_2O. 18µl master mix and 2µl cDNA were added to a 96 well PCR plate (MicroAmp Fast Optical 96 Well Plate, Applied Biosystems, Cat: 4346906), to a total well volume of 20µl. Each sample was set up in triplicate, along with appropriate negative controls. Plates were sealed and centrifuged at 1500rpm for one minute.

Samples were run on a AB StepOne Plus analyser on the following cycle: 50°C for 2 minutes; 95°C for 10 minutes; 95°C for 15 seconds, 60°C for 1 minute, for a total of 40 cycles.

2.8.4.4 Analysis

Data from the qRT-PCR reaction were converted into relative quantities (RQ). This allowed comparison of mRNA quantities between samples without the requirement to set up a standard curve.

The cycle threshold (CT) for each target gene and an endogenous control gene (in most cases, GAPDH) was recorded for each sample under investigation. This is the cycle at which the exponential phase of the PCR reaction crosses a user defined threshold, which was standardised across all plates run. Triplicate CT values for each gene were averaged. Any replicate reactions with a CT standard deviation of >0.5 were discarded from further analysis.

The difference between the average CT of a target gene and the endogenous control gene was calculated by the formula:

$$\Delta CT = CT\ target - CT\ endogenous\ control$$

This step normalises the gene expression values. Expression was then compared to that of a user-defined calibrator sample, by the following calculation:

$$\Delta\Delta CT = \Delta CT\ test\ sample - \Delta CT\ calibrator\ sample$$

In most of the work found within this thesis, the calibrator sample was derived from the average of all of the control samples' ΔCT values.

The logarithm base 2 scaled data are then converted into a linear scale by the formula:

$$Relative\ quantification\ (RQ) = 2^{-\Delta\Delta CT}$$

The significance of differences in gene expression between two groups of samples were then be assessed by a t-test (for parametric data), or Mann-Whitney test (for non-parametric data).

2.8.4.5 Purity Assessment of Sorted HSCs from AML samples with sequenced NPM1 Mutations

This method uses qPCR to assess the proportion of DNA, and therefore cells, within a given sample which carry a known mutation of the *NPM1* gene. This allows us to assess the percentage purity of a sorted, "normal" population within an appropriately selected AML sample. The relative sensitivity of this method means it is ideally suited to assessing the purity of small sorted population of cells, as results can reliably be obtained for cell numbers as low as 750.

2.8.4.5.1 Preparation of a Standard curve

Standards were produced using stock DNA samples known to have the same point mutation as the sample for which purity was being assessed. A stock solution with a concentration of 60ng/µl was made up. Using the assumption that one cell contains $6x10^{-3}$ng DNA, this allowed the production of the standards as detailed in Table 2-2.

Table 2-2 showing the production of standards for NPM1 assessment assay

STANDARD	DNA CONCENTRATION (ng/µl)	DNA VOLUME ADDED (µl)	H_2O VOLUME ADDED (µl)	EQUIVALENT CELL NUMBER IN 5µl DNA SOLUTION
1	60			50000
2	6	20µl STANDARD 1	180	5000
3	1.2	30µl STANDARD 2	120	1000
4	0.12	20µl STANDARD 3	180	100
5	0.012	20µl STANDARD 4	180	10
6	0.006	50µl STANDARD 5	50	5

2.8.4.5.2 Sample preparation

The cells whose purity was assessed were pelleted (3800rpm, 5 minutes), washed with 4ml 2% PBS, and underwent DNA extraction using the QIAamp DNA Blood Mini Kit (Qiagen, Cat:

51104), as described in Section 2.8.1. Samples were diluted to a DNA concentration that fell below that of the highest standard concentration (i.e. less than 60ng/µl).

2.8.4.5.3 PCR Reaction

The following probes and primers were ordered from Applied Biosystems, via the Assay-by-Design Service.

Forward primer (common for all *NPM1* mutations) 5' gtgttgtggttccttaaccacat 3'; Reverse primer (mutation specific) for the *NPM1* mutation A TCTG 5' tcctccactgccagacagag 3'; Probe (Taqman MGB probe, FAM dye-labelled) 5' ttttccaggctattcaagat 3'.

Master mix was made up in the following proportions for each sample: 10µl 2xTaqMan Universal Master Mix (Applied Biosystems, Cat: 4304437); 1µl Forward Primer (10µM); 1µl Reverse Primer (10µM); 2.5µl Probe (2µM); 0.5µl H_2O. Master mix for the control Albumin was made up in the following proportions for each sample: 10µl 2xTaqMan Universal MasterMix; 1µl Albumin Probe and Primer Mix (x20) (Applied Biosystems, Cat: 4351372); 4µl H_2O.

All samples were run in triplicate. Both master mix and DNA samples were vortexed before addition to the PCR plate (MicroAmp Fast Optical 96 Well Plate, Applied Biosystems, Cat: 4346906).

15µl of appropriate master mix was added to each well, followed by 5µl of DNA solution. The plate was centrifuged to 1500rpm briefly to remove air bubbles, and analysed on an Applied Biosciences 7900HT Fast Real Time PCR System, using the following program: 50^{o}C for 2 minutes; 95^{o}C for 10 minutes; 95^{o}C for 15 seconds, 60^{o}C for 1 minute, for a total of 40 cycles.

2.8.4.5.4 Analysis

Results were calculated as the percentage of DNA and therefore by extrapolation, percentage of cells, within the sample which were *NPM1* mutant positive. This was calculated by using the average CT of triplicate readings, and then using this result to read from the standard curve the number of cells this CT represents. The same process was undertaken for both *NPM1* mutated DNA, and Albumin.

The results are then expressed as

$$percentage\ NPM1\ mutant\ cells = \frac{mean\ calculated\ number\ of\ NPM1\ mutated\ cells\ from\ standard\ curve}{mean\ calculated\ number\ of\ cells\ from\ Albumin\ standard\ curve}\ x\ 100$$

Chapter 3 Study of patients with AML who present without trilineage haematopoietic failure

3.1 Introduction

3.1.1 Mechanisms of presentation with AML

The vast majority of patients with AML present with the signs or symptoms of bone marrow failure. These include shortness of breath and fatigue resultant from anaemia; bleeding or bruising resultant from a low platelet count, and sepsis resultant from reduced or dysfunctional neutrophils.

Whilst other haematological malignancies can present in a fashion attributable to bone marrow failure (this is by no means a mechanism specific to AML), they also commonly present via very different mechanisms. Some leukaemias, including ALL, are detected during incidental blood tests. ALL can often cause systemic malaise or neurological manifestations, rather than simply overt cytopenias.

3.1.2 Summary of previous work suggesting bone marrow failure in AML is due to induced HSC quiescence

In Section 1.5.2 of this thesis, I summarise previous work from our group investigating the mechanism by which AML appears to impede normal haematopoiesis. Our group has chosen to focus on the interaction between AML blasts and normal HSCs. Studies of human samples and xenograft models have suggested that, whilst absolute numbers of normal HSCs appear preserved in AML diagnosis samples in comparison to controls, their behaviour is modulated to a state of induced quiescence. In contrast, the number of downstream progenitor cells is significantly lower than in controls. These observations led to the theory that haematopoietic failure in AML is due to the induction of quiescence in normal HSC populations, leading to a block in further differentiation[1]. This is summarised pictorially in Figure 3-1.

AML is an extremely heterogeneous disease, characterised by a diversity of cytogenetic and genetic abnormalities between patients, and observed blast phenotypes. The hypothesis put forward above is potentially a universal explanation that could encompass the bone marrow failure observed across this wide diversity of samples.

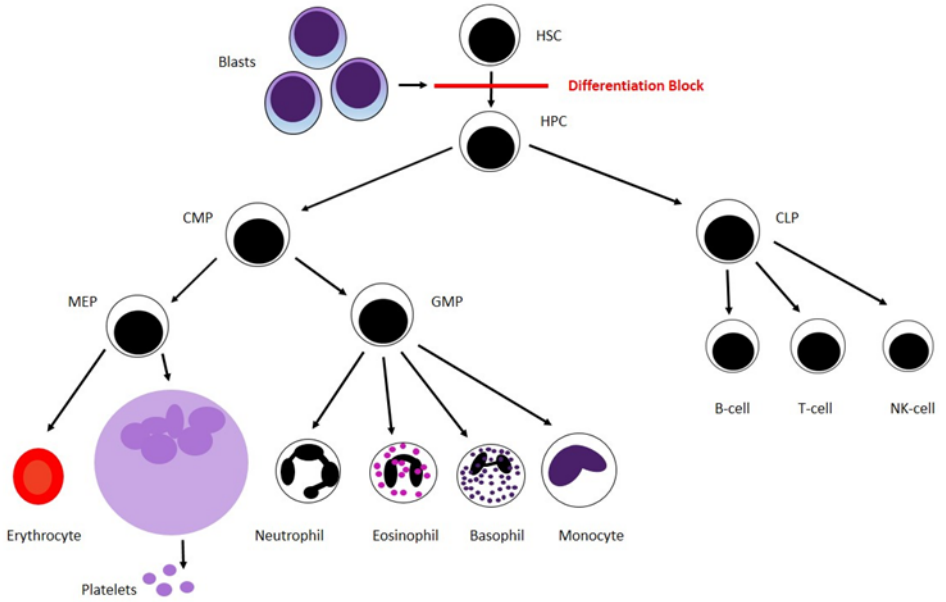

Figure 3-1 A hypothesis of induced differentiation block in HSCs in AML. HSC- haematopoietic stem cell; HPC- haematopoietic progenitor cell; CMP- common myeloid progenitor; MEP- megakaryocyte-erythroid progenitor cell; GMP- granulocyte-macrophage progenitor cell; CLP- common lymphoid progenitor

3.1.2.1 Do AMLs which present with haematopoietic lineage preservation invalidate this hypothesis?

One potential flaw in a hypothesis that seeks to explain AML-associated bone marrow failure solely by an observed effect on the HSC population is that the pattern of cytopenia development is not universal across all patients. Although most present with suppression of normal red cells, platelets and neutrophils (henceforth referred to as trilineage failure), we have also diagnosed patients with preservation of one or more of these lineages. If AML causes only HSC quiescence in all samples, theoretically, one might expect to observe the same pattern and time course of haematopoietic failure across all subjects. A parallel situation is provided by patients who receive myeloablative chemotherapy or radiotherapy during bone marrow transplantation conditioning, who universally drop their normal counts in a highly predictable fashion.

In looking for parallels in other diseases associated with HSC failure, it is clear that trilineage haematopoietic failure at diagnosis is not always observed, although the reasons for this are murky. Myelodysplasia (MDS) and acquired aplastic anaemia (AAA) are bone marrow failure syndromes thought to be due to HSC failure. However, the diagnosis of aplastic anaemia only requires the suppression of two of the three haematopoietic lineages[154]. Similarly, trilineage cytopenias are often not seen in MDS (an excellent example being that of MDS associated with 5q- deletion, associated with refractory anaemia, a preserved platelet count and the presence of dysplastic megakaryocytes in the bone marrow)[155,156].

Below, I have summarised a number of potential explanations for the differential lineage failure sometimes observed in AML. This is a novel area, with little published literature to guide conjecture.

3.1.2.1.1 Interaction with different diseases

Not all patients who develop AML might do so on the background of an entirely normal haematopoietic system, a fact which might affect their time to presentation, as well as their full blood count at the time they seek medical assistance. Examples include patients with pre-existent anaemia, due to iron deficiency or haemoglobinopathies, who might present with markedly low haemoglobin with relative preservation of platelets if HSC quiescence was induced by AML.

In a similar fashion, the association between myeloproliferative disorders and subsequent development of AML is well known[157,158]. Patients with *JAK-2* positive thrombocytosis who develop AML might maintain a normal or even elevated platelet count, whilst their red cell and neutrophils plummet. Those with Polycythaemia Rubra Vera may demonstrate a similar pattern with respect to red cell levels.

3.1.2.1.2 Do some AMLs retain the ability to differentiate into apparently normal functional cells?

A very different possibility is that some AML LSCs might retain the ability to differentiate into cells of similar morphological appearances to at least some of the terminally differentiated cells required for normal bone marrow function.

Certain leukaemia subtypes are associated with the preservation or proliferation of certain mature cell subtypes. Obvious examples include AML M4Eo, highly associated with inv(16) and eosinophilia[159]; or AMLs with inv(3), t(3;3) and balanced t(3q21) rearrangements, which have been shown to present with higher platelet counts than other AMLs[160].

Information on lineage involvement of mature cells in AML is scarce, and published work normally involves small series of patients[161,162].

FISH staining on fresh bone marrow samples of patients with t(8;21)[163] and inv(16)[164], concurrent with morphological identification of cells, has occasionally shown the development of cells from a leukaemic clone into eosinophils of mature appearances. Using the same methodology, AMLs with inv(3) have been shown to differentiate into a variety of different mature cell forms[165]. Falini's team used laser-dissection techniques on bone marrow samples

of AML patients with known *NPM1* mutations to dissect out visible megakaryocytes and perform single cell mutational analysis. This revealed the presence of the mutated gene in megakaryocytes of three patients[166].

3.1.2.1.3 Is bone marrow failure in AML not solely due to induced quiescence on the HSC population?

Our group has focussed on the interaction between AML and the activity of the normal HSC within the bone marrow. However, it is perfectly possible that this is not the only step in the pathway of normal haematopoiesis that is affected by the presence of malignant blasts in the marrow. This concept is illustrated pictorially in Figure 3-2 .

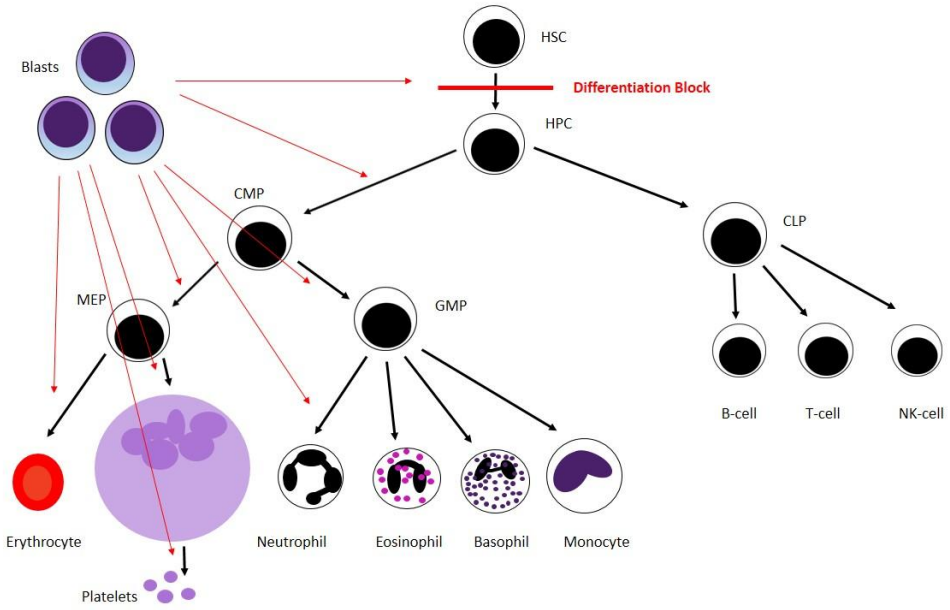

Figure 3-2 Pathway showing a proposed theory of haematopoietic suppression at multiple steps of the haematopoietic pathway by AML blasts

This hypothesis suggests that AML might inhibit either differentiation or division at multiple steps in normal development, by competition for niche space, resources, or appropriate cytokine signalling. These effects might not be consistent across all AML subtypes: dependant on the AML phenotype, different lineages might be differentially altered. For example, might an AML M7 have more impact on normal platelet production than other leukaemias? Might the interference on normal erythropoiesis be more marked from an AML of M6 subtype?

Again, this hypothesis is not mutually exclusive with the observations made by our group with regard to the effect of AML on normal HSCs, but might explain the non-uniform patterns in count suppression we observe clinically.

3.1.3 Normal erythroid development

3.1.3.1 Morphologically described maturation

Morphological examination of the bone marrow has identified a number of discrete, sequential steps in erythroid development. The first stage to be clearly identifiable is the pronormoblast; followed by the appearance of early, intermediate and late normoblasts (or erythroblasts); reticulocytes; and finally mature red cells, which are extruded into the blood stream. During this process of maturation, the cells shrink in size, and assume a biconcave disc shape, which facilitates efficient oxygen transfer. The nucleus is also extruded: an evolutionary adaptation thought to contribute to gas transfer ability and to improve cell mobility[6].

The use of flow cytometry has identified a series of surface antigens expressed on maturing cells belonging to the erythroid lineage. Commonly utilised markers in clinical practice include CD71 (transferrin receptor), which is reported to be expressed throughout development, but is downregulated as cells mature to erythrocytes[167], and CD235a (glycophorin A) whose expression increases during maturation[168,169].

3.1.3.2 In vitro assessment of red cell maturation

In the in vitro setting, the earliest committed erythroid progenitors are the Burst-Forming Unit-Erythroid (BFU-E) cells, which divide relatively slowly, and mature into the more rapidly proliferative Colony Forming Unit-Erythroid (CFU-E) cells. Over a period of two to three days, these cells divide and differentiate to assume the form of mature erythrocytes.

3.1.3.3 Homeostatic control of erythropoiesis

The process of normal erythropoiesis is under tight homeostatic control and is exquisitely sensitive to the effect of the hormone erythropoietin, produced in the kidney. Erythropoietin appears to particularly affect CFU-E cells, promoting their proliferation and terminal differentiation. However, in an acute setting, the response to erythropoietin is limited by the fact that each CFU-E can only undergo a further four to five divisions[170]. Therefore, under conditions of extreme stress, upstream activation also has to be initiated by inducing division of the BFU-E, which is an erythropoietin-independent process[171]. Work on rodent models suggests that in this context of “stress-haematopoiesis”, BFU-E mobilise to the spleen, where their proliferation is enhanced[172]. Hormonal signals which appears to have a role in promoting this process include glucocorticoids such as cortisol[173], SCF (Kit ligand) [174] and BMP4[175]. The key steps in erythropoiesis, including cytokines and transcription factors with defined roles, are summarised in Figure 3-3.

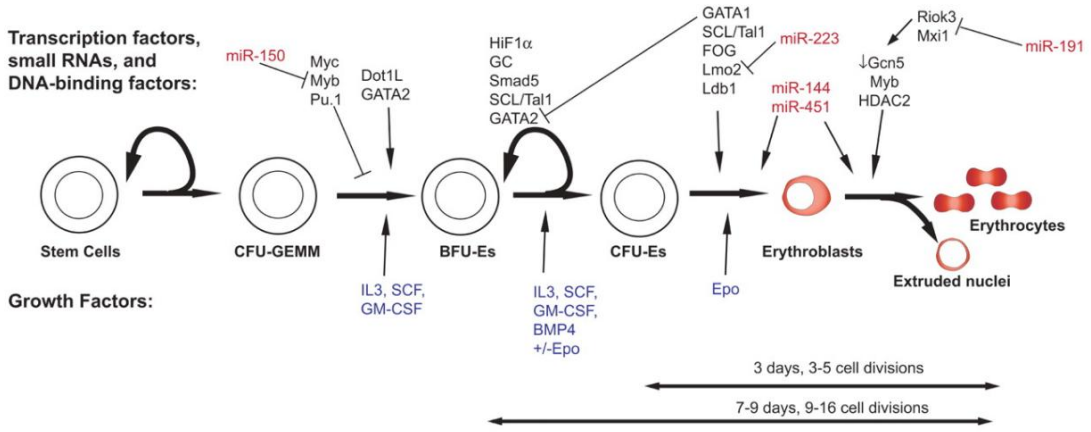

Figure 3-3 Summary of erythroid development from Hattangadi, Blood 2011[170]

3.1.4 Normal platelet development pathway

Like erythropoiesis, normal platelet production occurs in the bone marrow under tight homeostatic control. CMP or MEP cells can differentiate to form megakaryocytes: rare, large (50 to 100µm diameter) cells of distinct morphological appearances, responsible for the production of platelets. The bone marrow osteoblastic niche is thought to attract immature precursor cells and promote megakaryocyte formation. This area is rich in collagen, which has been shown, through α2β1 integrin binding, to promote differentiation towards the megakaryocytic fate whilst preventing the formation of proplatelet projections required for platelet release[176,177]. In this location, cells become polyploid via the process of endomitosis[178]. This is thought to be promoted by the action of thrombopoietin (TPO), and may represent an adaptation to the need to produce significant amounts of mRNA and protein for platelet production, without the need to undergo mitosis[179].

These partially mature megakaryocyte precursors are subsequently drawn to a separate perivascular niche location via expression of SDF1 and upregulation of the receptor CXCR4. This area seems to provide an environment conducive to late stage megakaryocytic development[41]. In this site, megakaryocytes extend long branching processes called proplatelets into the sinusoidal blood vessels of the bone marrow. The formation of these extensions requires the coordinated assembly of β1-tubulin[180], myosin[181] and F-actin[182], which not only act as structural supports, but also shuttle organelles and proteins to the developing platelets at the end of the projections.

Mature platelets bud off from the end of these branched processes. These are enucleate, small cells (1 to 3µm diameter). It is estimated that the process of megakaryocyte development from polyploidisation to the release of mature platelets takes two or three days in rats[183], and double that time in humans.

3.1.4.1 Phenotypic identification of cells of megakaryocytic lineage

Surface markers commonly used to identify cells of the megakaryocytic pathway include CD41 (αIIβ integrin, also known as platelet GPIIb) and CD61 (β3 integrin, also known as platelet GPIIIa)[184,185]. Both of these markers remain highly expressed on mature platelets as well as immature bone marrow forms. It is also worth highlighting that these markers are not absolutely restricted to this lineage[186].

3.1.4.1.1 Homeostatic control of platelet production

The influence of TPO is key to the formation of megakaryocytes. Its activity has been linked to a number of roles in megakaryocyte development and maturation, and as a result, it represents a key component of in vitro culture systems for megakaryocyte studies[187]. TPO receptor agonists such as romiplostin and eltrombopag have demonstrated clinical efficacy in restoring platelet counts in the setting of refractory ITP[188,189]. However, there is clearly redundancy within this control system, as murine knock-outs of the TPO receptor, c-mpl, have demonstrated the ability to maintain normal platelet counts[190].

3.2 Aims and Objectives

The first aim of this chapter was to determine the extent of clinical bone marrow failure in patients with AML. We hoped to determine if all patients present with the same pattern of bone marrow failure, and if there are differences, identify any obvious associations.

This aim was met by the construction and subsequent interrogation of a database comprising clinical information on all de novo AML patients presenting to St Bartholomew's Hospital in the 15 years. As the cancer referral centre for the largest NHS trust in the country, and the current highest recruiter of patients into the national MRC AML trials, this represents an extremely valuable local resource.

The second aim of this chapter was to determine how some patients present with AML with preserved levels of either red cells or platelets. For this cohort of patients, I wished to determine if these cells were derived from the leukaemia itself, or normal unmutated haematopoietic cells. Levels of neutrophils at presentation were not included, due to documented difficulties in distinguishing neutrophils from leukaemic cells on an automated blood count.

This aim was met by the development of a flow-based method for separation of nucleated red cell and platelet precursors from bone marrow samples. Subsequent examination of FACS sorted precursor cells was performed either using PCR or FISH based techniques to look for the presence of the leukaemia associated mutation.

3.3 Specific Methods

3.3.1 Database analysis

I am very grateful for the work of Janet Matthews, Haematology Data Manager, in compiling a database of all patients newly diagnosed with AML at Bart's Hospital between January 1997 and September 2012. The data for this table was retrieved from the Bart's Clinical Cancer Database.

The following were recorded for all patients where data were available: age at diagnosis; sex; karyotype and FAB classification of AML; presenting haemoglobin (g/dL), WCC ($x10^9$/L) and platelet count ($x10^9$/L). Clinical survival data and relapse/remission status were recorded where available. Final censoring of patients was performed in June 2016.

Presenting FBC parameters were measured from peripheral blood using a Sysmex analyser, as is standard clinical practise in Bart's. Haemoglobin (g/dL) and platelet count ($x10^9$/L) were logged for each patient. Other parameters generated as standard by a Sysmex included a total white cell count, as well as a full leucocyte differential. However, in the presence of blasts, these parameters often do not accurately represent the blood film appearances, and therefore a decision was made not to include measurement of neutrophil levels within this dataset. Some of these patients had records of manually differentiated blood films, but again, my own clinical experience suggests distinguishing malignant and non-malignant cells in this setting is a non-exact science.

FAB classification was made after morphological examination of the bone marrow by the Haematopathology team. Karyotypic information was provided by cytogenetic and FISH based analysis of samples run as part of routine clinical practise at the Royal London Hospital.

All data were anonymised in keeping with University-wide data protection regulations.

3.3.1.1 Statistical analysis

The majority of statistical analysis was performed using the Prism software package.

All data sets were tested for normality of distribution by the D'Agostino & Pearson Omnibus Normality Test. For parametric datasets, means and standard deviations are quoted. In the determination of statistical significance between two data sets, a Student's t-test was applied. If data were not normally distributed, medians and interquartile ranges are quoted. In the determination of statistical significance between two non-parametric data sets, a Mann-

Whitney t-test was applied. If the differences in the averages of more than two groups were being compared, either an ANOVA test (for parametric datasets with equal variance) or a Kruskal-Wallis test (for non-parametric data) was applied.

3.3.2 Patients and samples

Diagnostic bone marrow samples from selected AML patients were obtained from the tissue bank maintained by the Department of Haemato-Oncology at St Bartholomew's Hospital. Written informed consent was obtained in accordance with the Declaration of Helsinki. These were used for both FACS sorting and subsequent genetic analysis of erythroid precursors, as well as flow based-identification and enumeration of erythroid precursor and stem cell populations.

Flow and immunohistochemical techniques were optimised on control bone marrows. Control marrows were defined as those obtained during the work-up of newly diagnosed Stage I-III lymphoma patients with not only a full blood count within normal parameters, but also an unaffected bone marrow as confirmed by aspirate, flow cytometry and trephine histopathology.

Bone marrow trephine slides previously stained in H&E were examined for megakaryocyte numbers for both AML patients and controls were selected from an Institute bank comprising of bone marrow trephine samples collected between 2005 and 2011.

3.3.3 Flow Cytometry

3.3.3.1 Staining for Erythroblasts

As detailed in Section 3.1.3, mature red cells extrude their nuclei before entering the circulation. Therefore, when attempting to determine the clonality of red cells in AML samples which had presented with a relatively preserved haemoglobin level, I attempted to distinguish erythroblast precursors at as late a stage of maturity as possible, before the loss of their nuclei. A flow-based strategy was adopted which identified erythroblasts ($CD235a^{+}71^{+}$ cells), but aimed to exclude mature erythrocytes (which have lower CD71 expression levels). Co-expression of these markers by leukaemia blasts themselves is rare (with the exception of AMLs of the M6 subtype)[191].

Bone marrow vials were thawed, counted and incubated with 2% HAG as described in Section 2.4.5.1**.** Cells were incubated with 2.5µl/1.5x10^{6} cells CD235a PE (Biolegend, Cat: 349105 (Clone HI264)), 5µl/1.5x10^{6} cells CD71-APC (Biolegend, Cat: 334107 (Clone CY1G4)) and CD3-

FITC (BD, Cat: 555332 (Clone UCHT1)). After incubation, cells were washed and resuspended in DAPI/2% PBS as previously described, and sorted using a BD FACSAria II^{TM}.

The CD235a^{+}CD71^{+}CD3^{-}DAPI^{-} population of cells was selected using the gating strategy illustrated in Figure 3-4, and sorted into 2% PBS. Sort purity was checked, and if this fell below 90%, resorting was undertaken.

Subsequent analysis was performed using the FlowJo Version 10 software.

Sorted cells were made into slides as described in Section 2.7.1, for MGG staining, immunocytochemical staining for haemoglobin and FISH analysis.

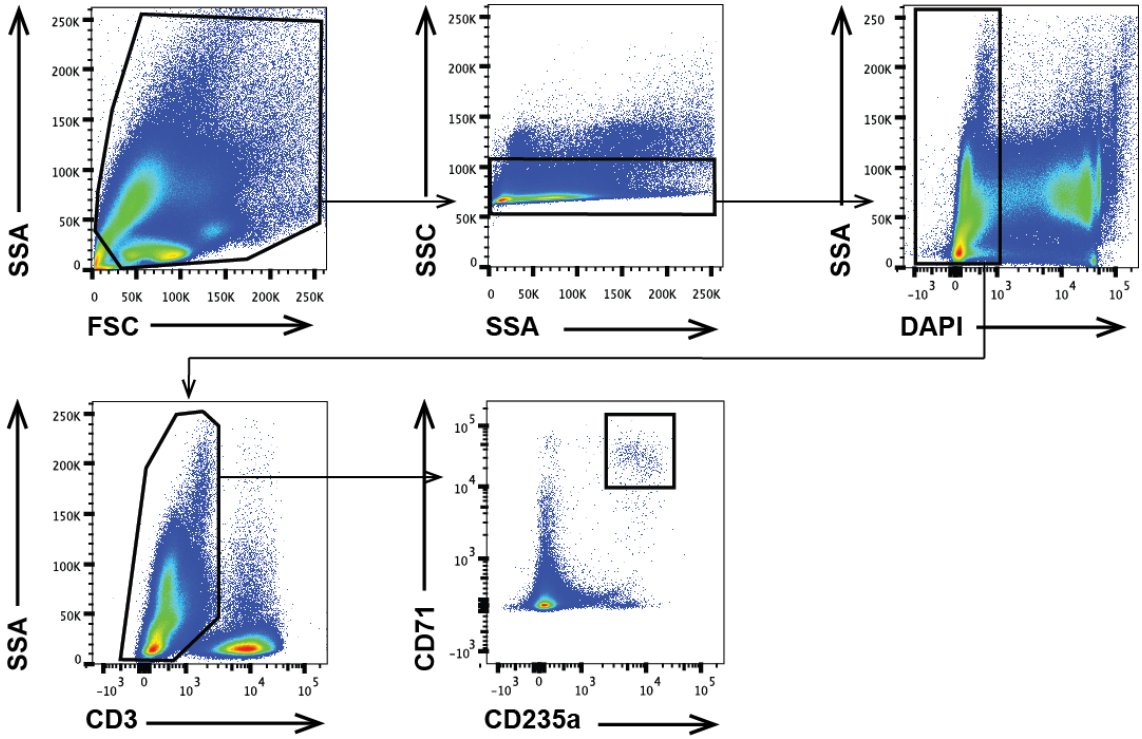

Figure 3-4 showing the flow gating strategy used to identify erythroid precursors from bone marrow samples

3.3.3.2 Analysis of previously analysed samples for erythroid precursor populations and stem cell concentrations

Chapter 5 details work on fresh BM samples designed to calculate the concentrations of normal HSCs at diagnosis with AML. In brief, AML and control bone marrow samples were analysed fresh using CountBeads, stained with CD34 and CD38, and the concentration of CD34^{+} cells in each fresh sample (cells/µl) was calculated. At a later date, a frozen sample of the same bone marrow was thawed, and using staining with CD34, CD38, ALDH and CLL1, the concentration of CD34^{+}38^{-}ALDH^{high}CLL1^{-} cells at diagnosis was calculated. Full methodological details are given in Sections 5.3.1 and 5.3.2.

For 20 of these AML samples and 7 control samples, at the time of thawing and staining with CD34, CD38, ALDH and CLL1 for HSC numbers, a separate aliquot was used to identify and quantify erythroid precursors.

Thawing, enumeration and staining of BM samples for the HSC population was performed as described in Section 5.3.2. $3x10^6$ cells were pelleted and resuspended in 2% HAG at a concentration of 50µl 2% HAG/$1.5x10^6$ cells. Cells were incubated at 4^oC for 20 minutes. Antibodies to CD34-PerCP, CD235a-PE and CD71-APC were then added at a concentration of 5µl/1.5 $x10^6$ cells, and left to incubate for 30 minutes at 4^oC. The cells were then washed with 4ml 2% PBS, and pelleted.

Cells were re-suspended in 300µl of a stock solution of 2% PBS/DAPI/DNase and filtered before analysis on a BD Fortessa.

3.3.3.2.1 Analysis of the frozen sample

Erythroid precursor numbers in the frozen sample were determined by sequential gating for live, $CD235a^+CD71^+$ cells. $CD34^+$ cell numbers in the same sample were also recorded.

The ratio of $CD235a^+CD71^+$ cells to $CD34^+$ cells within the same, thawed sample was calculated.

The original concentration of $CD235a^+CD71^+$ erythroid precursor cells (cells/µl) at diagnosis was calculated as shown below.

$$[CD235a^+CD71^+] =$$
$$[CD34^+\ cell]\ in\ fresh\ sample\ \times \left\{\frac{CD235a^+CD71^+\ cells}{CD34^+cells}\right\}\ in\ frozen\ sample$$

The concentrations of $CD34^+38^-ALDH^{high}CLL1^-$ stem cells in the same sample were calculated as shown in Section 5.3.2.2.

3.3.3.3 Staining for megakaryocytes

Bone marrow vials were thawed, counted and incubated with 2% HAG as previously described. Cells were then incubated with CD45-APCy7 (BD, Cat: 557833 (Clone 2D1)) and CD41a-APC (BD, Cat: 561852 (Clone HIP8)) at concentrations of 5µl/$1.5x10^6$ cells for CD45-ApcCy7, and 10µl/$1.5x10^6$ cells for CD41a-APC. After incubation, cells were washed and resuspended in DAPI/2% PBS as previously described, and sorted on a BD FACSAria II™.

3.3.4 Morphological Techniques

3.3.4.1 MGG staining of slides prior to morphological assessment

Slides were transferred to the Haematology Laboratory at St Bartholomew's Hospital, and underwent automated staining with May-Grunwald-Giemsa stain. Slides were air-dried, and mounted in oil, prior to examination under the microscope.

3.3.4.2 Immunostaining for the presence of haemoglobin

Slides were fixed, dried, and permeabilised as previously described.

Slides were then blocked with donkey serum (Sigma, Cat: D9663)(0.1% saponin, 5% donkey serum in PBS).

Primary antibody staining was performed using a mouse monoclonal IgG1 anti-Haemoglobin antibody (AbCam, Cat: ab55081) at a concentration of 2µg/ml.

Secondary antibody staining was performed using Alexafluor 546 Donkey anti-mouse IgG (Life Technologies, Cat: A10036) at a 1:100 dilution (in donkey serum).

3.3.5 Genetic Analysis of samples for clonality identification

3.3.5.1 FISH based assessment of sorted cells

FISH staining and analysis was performed by Marianne Grantham at the Cytogenetics Department, Royal London Hospital, on slides prepared as described in Section 2.7.1. The following probes were used: Vysis RUNX1/RUNX1T1 DF Fish Probe Kit (Abbott Molecular, Cat: 08L70-020), CBFB t(16;16) inv 16 Break Probe (Kreatech, Cat: KBI-10304) and DEK/NUP 214 t(6;9) Fusion Probe (Kreatech, Cat: KBI-10306).

3.3.5.2 Confirmation of clonality of sorted cells by assessment for the presence of the NPM1 mutation

For those samples with a previously identified and sequenced *NPM1* mutation, sorted $CD235a^{+}71^{+}$ populations were assessed for the proportion of cells containing the *NPM1* mutation by the method described in Section 2.8.4.5.

3.3.6 Megakaryocyte Enumeration by Trephine Analysis

Within the Bart's Tissue Bank, there exists a small collection of bone marrow trephine samples taken at diagnosis from patients with AML.

Samples were obtained by the use of a trephine needle placed into the posterior iliac crest to remove a solid core of cortex and marrow of at least 2cm length. Samples underwent fixation with formalin, decalcification, and wax embedding prior to sectioning and staining with haematoxylin and eosin (H&E).

Selected, anonymised slides were then scanned onto the Pannoramic Viewing software for analysis. Sample quality was ensured by the analysis of trephines of a length of greater than 1cm, with an absence of artefact or significant clot.

Slides were studied at a uniform magnification of x20. Megakaryocytes were identified by their morphological appearance by one observer (KH) and tagged. The stromal area examined (excluding bony cortex) was mapped out and its size calculated by the Pannoramic system in mm^2.

Megakaryocyte concentrations were expressed as the number of megakaryocytes present per mm^2 of marrow stroma.

3.4 Results

3.4.1 Database analysis results

Data were collated on new patients presenting with AML to Bart's Hospital between January 1997 and August 2012. In this period of 15 years, 529 new cases of AML were identified where the full blood count (FBC) at diagnosis was documented.

3.4.1.1 Patient characteristics

The basic clinical profile of these 529 patients are summarised below.

The incidence of AML with age does not follow a normal distribution, a finding which mirrors previously reported series. Median age at diagnosis is 56, with a range of 16 to 86 years.

There is a slight predominance of males in this cohort, with a calculated male:female ratio of 1.2:1 (288/529 (54%) are male and 241/529 (46%) are female). This is also consistent with previously published epidemiological studies of AML incidence[192].

Karyotypic information was available for 489 of the 529 patients (92%) diagnosed between 1997 and 2012. In the cases where data were not available, this was either because cytogenetic analysis on diagnostic samples had not been requested, or because standard analysis assays had failed. Of these 489 patients, 100 (20.4%) had cytogenetic profiles consistent with good risk disease, 270 (55.2%) were classified intermediate risk and 119 (24.3%) were of poor risk. The frequency of these cytogenetic abnormalities is broadly in keeping with published datasets[101,193].

Morphological classification of samples according to the FAB classification system was recorded in 448/529 cases (85%). Cases were split across the different morphological categories as follows: 28 (6.2%) patients were of FAB M0 appearance; 82 (18.3%) patients were FAB M1; 99 (22.1%) patients were FAB M2; 68 (15.2%) patients were FAB M3; 87 (19.4%) patients were FAB M4; 63 (14.1%) patients were FAB M5; 12 (2.7%) patients were FAB M6, and 9 (2.0%) FAB M7. Again, these figures are in keeping with previously published data series[124].

3.4.1.2 Haemoglobin levels at diagnosis with AML

3.4.1.2.1 Haemoglobin levels are lower than expected normal range at diagnosis in the vast proportion of patients

Haemoglobin levels at diagnosis, are, unsurprisingly, substantially lower than the expected values for normal healthy individuals. Figure 3-5 A illustrates the range of haemoglobin levels

seen at diagnosis with AML across the whole cohort. Interestingly, this follows a normal distribution pattern, as do haemoglobin levels within a healthy population.

The mean haemoglobin level at diagnosis across the cohort is 9.2g/dL, with a standard deviation of 2.1. The range of observed haemoglobin levels is wide (between 3.4 to 15g/dL).

3.4.1.2.2 Age

The effect of age on presenting haemoglobin was explored. This was on the basis that older patients might have a lower tolerance to anaemia, due to co-morbidity: a factor which might precipitate their earlier presentation to medical attention.

In fact, analysis of the relationship between age and haemoglobin levels at diagnosis using Spearman's rank order correlation gave a non-significant Spearman's correlation coefficient of 0.035 (p=0.42). The relationship between age and presentation haemoglobin is shown graphically in Figure 3-5 B.

3.4.1.2.3 Gender

Gender is known to influence haemoglobin levels in health (see Table 1-1), with the normal range for men (13.0 to 18.0g/dL) being slightly higher than that seen in women (11.5 to 16.0g/dL). Given these differences, presenting haemoglobin patterns for men and women were independently assessed. The median haemoglobin at presentation for men is 9.4g/dL (interquartile range 7.9 to 10.6g/dL), and for women 9.3g/dL (interquartile range 7.9 to 10.4g/dL). There is no statistical difference between these two results (Mann-Whitney t-test, p=0.35).

However, it is worth noting that the level of haemoglobin at diagnosis was not abnormal for all patients within this dataset. Allowing for gender-based differences in the lower limit of the normal range, 32/529 (6.0%) of all patients had a normal haemoglobin at diagnosis. 10/288 (3.4%) of male patients had a haemoglobin of greater than 13.0g/dL, whilst a higher proportion of women (22/219 (10.0%)) had a haemoglobin greater than 11.5g/dL.

3.4.1.2.4 Cytogenetic abnormalities do not have an effect on haemoglobin levels at presentation

Samples were grouped into good, intermediate or poor risk cohorts based on their cytogenetics. Within these groups, there were very slight differences in mean presenting haemoglobin levels only: 9.1g/dL for good risk patients (standard deviation 2.2, n=100); 9.3g/dL for intermediate risk patients (standard deviation 2.1, n=270); 9.3g/dL for poor risk

patients (standard deviation 2.0, n=119). There is no statistical significant difference between the means of these three groups (One way ANOVA, p=0.81).

To ensure I had not missed the haemoglobin-preserving (or depleting) effect of a particular translocation by pooling samples together into three cytogenetic risk groups, samples were subdivided into 19 categories with known prognostic impact, and the median haemoglobin levels for each were determined. These results are shown in Figure 3-5 C.

There is no significant difference observed between the median haemoglobins at presentation of these different cytogenetic risk categories (Kruskal-Wallis, p=0.89).

3.4.1.2.5 There are insufficient data available for this cohort of patients on the status of genes commonly mutated in AML to assess their impact on presentation haemoglobin

Ideally, one would analyse this dataset for the impact of common gene mutations found in AML on presentation haemoglobin. These would include *NPM1, FLT3, CEBPA, KIT, N-RAS, MLL, WT1, IDH1/2, TET2 and DNTMT3A*, but also in this context it would also be interesting to know the study the effects of mutation within the *JAK2* gene.

Unfortunately, this dataset comprises of patients diagnosed between 1997 and 2012, the majority of which predate even the introduction of routine *NPM1* and *FLT3* screening at diagnosis (see Section 3.5.5 for further discussion of this area). We did not have sufficient funds to enable the targeted sequencing of these samples for the listed mutations.

However, clinical notes were obtainable for 22 of 32 patients who presented with a haemoglobin within normal limits for gender. None of these patients had a prior diagnosis of PRV or MPD before presentation with AML.

3.4.1.2.6 Does presentation haemoglobin have a prognostic impact?

Survival data were obtained for 528 patients within this cohort. Analysing patients as two cohorts (low haemoglobin at diagnosis corrected for gender (<11.5g/L for females, <13.5g/dL for men, n=497) and normal haemoglobin at diagnosis corrected for gender (≥11.5g/dL for females, ≥13.5g/dL for men, n=31)), there was no significant difference in overall survival observed by Kaplan-Meier analysis (log rank test, Chi Square 1.4, p=0.23). This is illustrated in Figure 3-5 D.

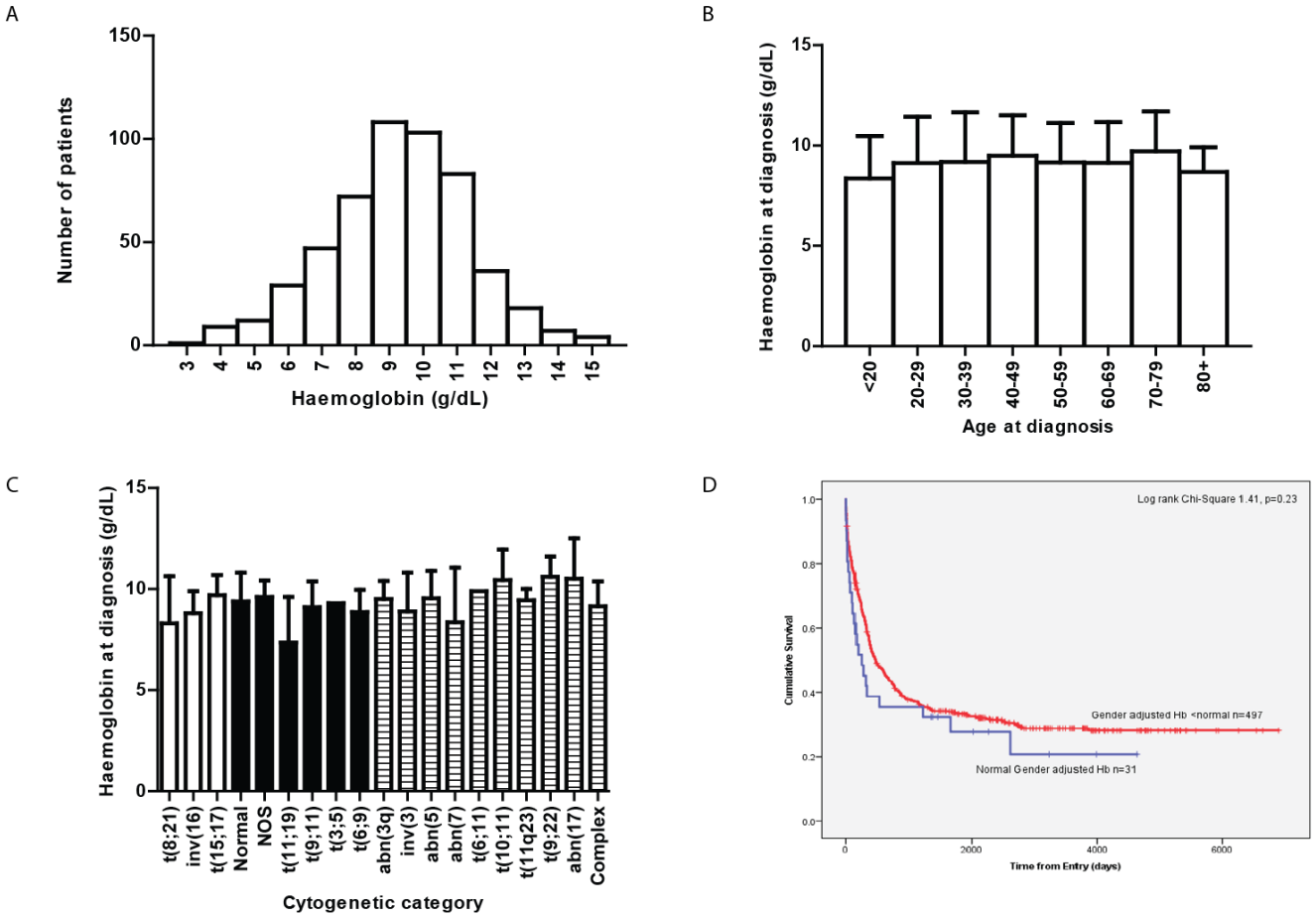

Figure 3-5 A: Histogram illustrating presentation Hb levels for 529 patients with AML; B: Histogram illustrating the relationship between presentation Hb level and age at diagnosis. Mean Hb and SD are plotted for each age range; C: Bar chart showing effect of 19 different AML cytogenetic abnormalities on Hb level at diagnosis, following the classification by Grimwade et al (2010). Good risk subcategories shown are t(8;21) n=23, inv(16) n=14, t(15;17) n=63; intermediate risk subcategories shown are normal cytogenetics n=19, abnormalities not otherwise classified or NOS n=58, t(11;19) n=2, t(9;11) n=8, t(3;5) n=1, t(6;9) n=4; poor risk subcategories shown are abn(3q) n=3, inv(3) n=2, abn(5) n=8, abn(7) n=24, t(6;11) n=1, t(10;11) n=4, t(11q23) n=14, t(9;22) n=5, abn(17) n=2, complex n=48. If a sample contained two or three cytogenetic abnormalities which meant its grouping was not clear (for example, coexistent deletion 5 and 7), it was excluded from analysis. Those with 4 or more abnormalities were placed in the complex cytogenetics cohort. Median Hb and IQR are plotted for all samples; D: Kaplan-Meier curve showing the effect of gender-adjusted Hb at diagnosis with AML on overall survival (normal Hb for males >13.0g/dL, and >11.5 for females)

3.4.1.3 Platelet levels at diagnosis with AML

3.4.1.3.1 Platelet levels are lower than the expected normal range in the majority of patients with AML, but they do not follow a normal distribution

Analysis of the 529 patients within the Bart's cohort showed, as expected, platelet levels in patients at diagnosis are substantially lower for the majority of individuals than the normal range in health. The median platelet count at diagnosis is $52x10^9$/L (interquartile range of 27-99). The range of observed counts is wide (3 to $567x10^9$/L).

However, in contrast to the pattern seen with haemoglobin levels, platelet levels in this cohort of AML patients do not follow a normal distribution pattern, as illustrated in Figure 3-6 A. The proportion of patients presenting with a normal platelet count is higher than the corresponding figure for haemoglobin levels. 75/529 (14.2%) of all patients have a platelet count greater than $150x10^9$/L.

Although $150x10^9$/L is widely considered the lower limit of the normal range for platelets, symptomatic bleeding is extremely rare at this level. A platelet count of $100x10^9$/L is taken as indicative of full count recovery in clinical trials, and within this cohort, 131/529 (24.6%) have levels greater than this.

3.4.1.3.2 Age

The effect of age on presenting platelet count was explored. This was based on the hypothesis that older patients might have a lower tolerance to thrombocytopenia, and possibly present earlier with bleeding manifestations, due to diverse factors such as increased tissue fragility and falls risk.

However, in a reflection of the pattern seen with haemoglobin levels and age, there appears to be no correlation between age and platelet count at diagnosis. Spearman's rank order correlation gave a non-significant coefficient of 0.049 (p=0.26). These data are shown graphically in Figure 3-6 B.

3.4.1.3.3 Gender

Unlike haemoglobin, the normal ranges in health for platelet levels are identical between men and women. The platelet levels at diagnosis with AML were also examined for any effect of gender. The median platelet count for men was $48.5x10^9$/L (interquartile range 26 to 83; range 6 to $567x10^9$/L). The median platelet count for women was $56x10^9$/L (interquartile range 29 to 111; range 3-$385x10^9$/L). The difference between these two results just fails to reach significance by Mann-Whitney t-test (p=0.07).

3.4.1.3.4 Patients with good risk cytogenetics have a lower platelet count at diagnosis than other groups

As in Section 3.4.1.2.4, samples were placed into good, intermediate and poor cytogenetic risk groups; and subsequently analysed for the effect of this classification on platelet levels at diagnosis. The median platelet level of good risk samples was $37.5x10^9$/L (interquartile range 16.5-61; range 6-226); of intermediate risk samples was $59x10^9$/L (interquartile range 31-106; range 7-567); and of poor risk samples was $51x10^9$/L (interquartile range 29-103; range 4-426). Kruskal-Wallis testing suggests the differences in platelet level between these groups are highly significant (p=0.0001). Sub-analysis using Dunn's Multiple Comparison Test suggests that the observed differences in platelet level between the good and intermediate risk groups, and between good and poor risk groups are highly significant.

3.4.1.3.4.1 Subgroup analysis suggests this may be due to the effect of t(15;17)

The effect of individual cytogenetic abnormalities on platelet count was explored, to ensure the effect of a particular translocation was not lost in pooling samples into three clinical risk groups. Samples were subdivided as described in Section 3.4.1.2.4, and median platelet count calculated for each category. Results are illustrated in Figure 3-6 C.

Kruskal-Wallis testing of these datasets suggest cytogenetic subclassification has a significant effect on platelet count at diagnosis (p=0.0002). However, further analysis using Dunn's Multiple Comparison Test reveals that it is only the difference in platelet counts between patients with t(15;17) (with a median platelet count of $22x10^9$/L) and those with normal cytogenetics (with a median platelet count of $66x10^9$/L) that reaches independent significance.

However, it is worth noting that the small numbers of samples which contain abnormalities of the long arm of chromosome 3 (t(3;5), inv(3) and abn (3q)) (n=6)) do appear to have relatively preserved platelet counts at diagnosis, in keeping with published data, even if the sample size here is too small to achieve statistical significance[194].

3.4.1.3.5 Preservation of platelet counts at diagnosis with AML does not appear to be due to pre-existing myeloproliferative disease.

As explained in Section 3.4.1.2.5, unfortunately, there were minimal data available for the patients within this cohort regarding the status of commonly mutated genes in AML.

However, given the potential effect of a prior diagnosis of essential thrombocytopenia (ET) on platelet counts, I attempted to review the clinical records of the 75 patients with platelet counts preserved above 150 at diagnosis. Notes were accessible for 44 of these. One patient

had a previous diagnosis of ET, and one a recent diagnosis of CMML. Therefore only 5% of those with a preserved platelet count at diagnosis with AML had a preceding history of MPD.

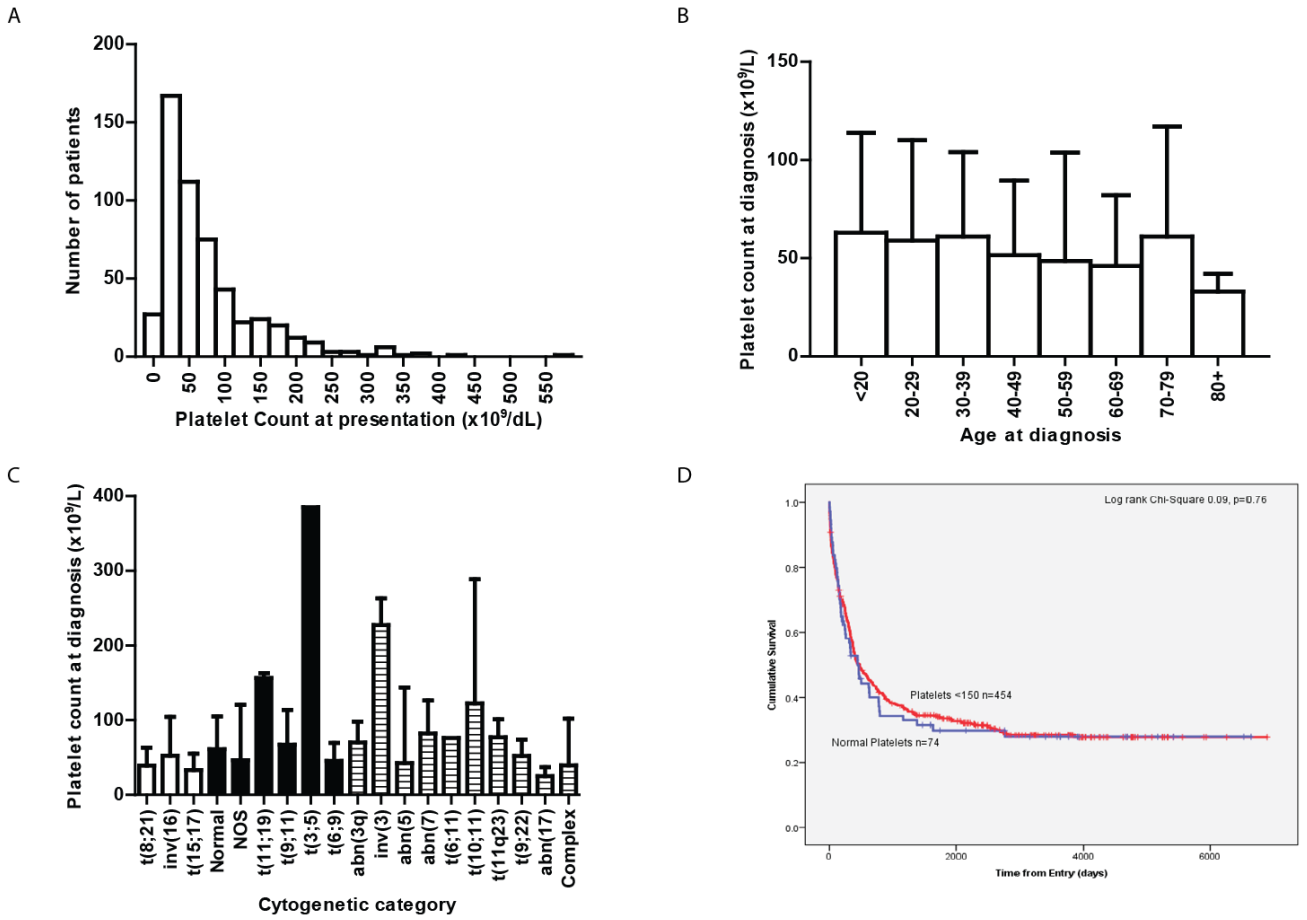

Figure 3-6 A: Histogram illustrating platelet counts for 529 patients at diagnosis with AML; B: Histogram showing the relationship between presentation platelet count and age at diagnosis. Median platelet counts and IQR are plotted for each age range; C: Bar chart showing effect of 19 different AML cytogenetic abnormalities on platelet level at diagnosis, following the classification by Grimwade et al (2010). Good risk subcategories shown are t(8;21) n=23, inv(16) n=14, t(15;17) n=63; intermediate risk subcategories shown are normal cytogenetics n=19, abnormalities not otherwise classified or NOS n=58, t(11;19) n=2, t(9;11) n=8, t(3;5) n=1, t(6;9) n=4; poor risk subcategories shown are abn(3q) n=3, inv(3) n=2, abn(5) n=8, abn(7) n=24, t(6;11) n=1, t(10;11) n=4, t(11q23) n=14, t(9;22) n=5, abn(17) n=2, complex n=48. If a sample contained two or three cytogenetic abnormalities which meant its grouping was not clear (for example, coexistent deletion 5 and 7), it was excluded from analysis. Those with 4 or more abnormalities were placed in the complex cytogenetics cohort. Median platelet count and IQR are plotted for all samples; D: Kaplan-Meier survival curve showing the effect of platelet count at diagnosis with AML on overall survival

3.4.1.3.6 Platelet count at diagnosis with AML does not have a prognostic impact on overall survival

Survival data were obtainable for 528 patients within this cohort. Analysing patients as two cohorts (low platelets at diagnosis (<150x10^9/L, n=454) and normal platelets at diagnosis (≥150x10^9/L, n=74)), there was no significant difference in overall survival observed by Kaplan-Meier analysis (log rank test, Chi Square 0.09, p=0.76). These data are illustrated graphically in Figure 3-6 D.

3.4.1.4 Is there a relationship between preservation of haemoglobin and platelet levels at diagnosis with AML?

We have identified patient cohorts with preserved haemoglobins and platelets independently. Table 3-1 highlights the interrelation between these parameters, and identifies an extremely small cohort of patients (1%) who present with both haemoglobin and platelet counts entirely within the normal range.

Table 3-1 illustrating the relationship between normal platelet levels and haemoglobin at diagnosis

PARAMETER	PLATELETS NORMAL (≥150x10^9/L)	PLATELETS LOW (≤150x10^9/L)	TOTAL
HAEMOGLOBIN NORMAL (FOR GENDER)	5 (1%)	26 (5%)	31 (6%)
HAEMOGLOBIN LOW (FOR GENDER)	70 (13%)	428 (81%)	498 (94%)
TOTAL	75 (14%)	454 (86%)	529

In the absence of obvious bone marrow failure (at least with regard to red cell and platelet lineages), how did these patients actually present? Clinical notes were accessible for four out of five of these patients. In three cases, patients presented with persistent and recurrent infections, presumably due to dysfunctional neutrophil populations. In the final case, presentation was prompted by abdominal pain from a chloroma.

3.4.1.5 Summary of results from clinical data

Study of presentation counts from this large series of patients does confirm that patients present with AML with evidence of bone marrow failure, given that 99% of patients identified within these cohorts presented with either or both an abnormally low haemoglobin or platelet count. However, as Table 3-1 shows, the pattern in which these parameters fall is not universal. This is illustrated by the fact that 5% of patients presented with normal haemoglobin but low platelets; and 13% of patients presented with normal platelets and a low haemoglobin.

Whilst haemoglobin levels in these patients fall at diagnosis with AML, but maintain a normal distribution within the cohort, platelet levels follow a different pattern, with a small and possibly distinct number of patients who appear to be able to maintain platelet production.

Examination of factors such as age, gender and cytogenetics has not been able to further identify potential drivers which might explain these patterns. The one exception to this is the confirmation in this dataset that the t(15;17) translocation is associated with significantly lower platelet levels at diagnosis. However, this observation may be due to increased peripheral consumption of platelets due to the DIC associated with this AML subtype, rather than relatively reduced marrow production.

3.4.2 Erythroid precursor lineage identification and quantification

3.4.2.1 Confirmation of red cell lineage in patients with preserved haemoglobin levels

3.4.2.1.1 Dual staining for CD235a and CD71 positivity reveals a population of erythroid precursor cells within normal bone marrow samples

Staining of normal control bone marrow samples with CD235 and CD71 produced a clear population of dual positive cells, which represented an average of 3% of all viable cells sorted (see Figure 3-4). Published flow based data would suggest this population contains nucleated erythroblasts, before extrusion of their nuclei and the final transformation to mature red cells (processes associated with CD71 expression downregulation).

A B C

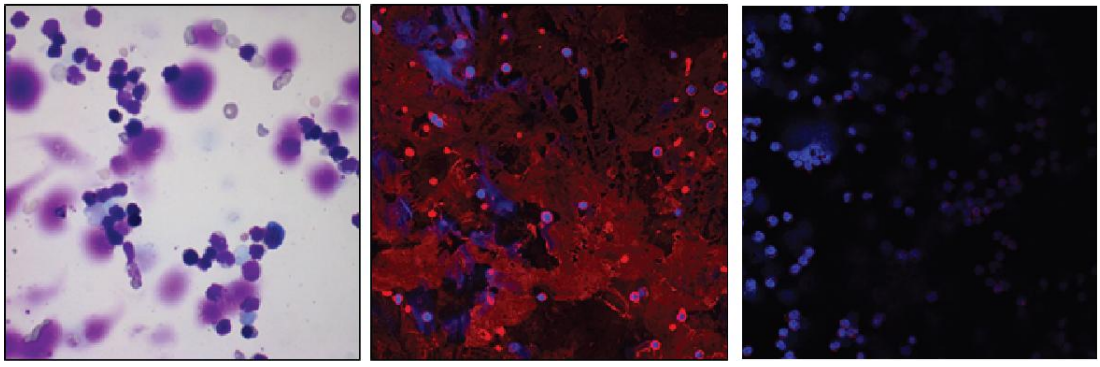

Figure 3-7 showing morphological appearances of sorted $CD235a^{+}71^{+}3^{-}$ cells. A: cytospin stained with MGG showing classic erythroid cells on the background of significant background staining presumably from cell lysis in cytospin process; B: immunofluorescence staining for the presence of haemoglobin within $CD235a^{+}71^{+}3^{-}$ cells; C: $CD235a^{-}71^{-}3^{+}$ cells (negative control)

The technique was validated by morphological examination of slides of these cells stained with MGG. Whilst slides consistently contained high artefact of cells lysed by the cytospin process, the nucleated cells in the slides had the morphological appearance of red cell precursors (as shown in Figure 3-7 A).

Secondary evidence of the erythroid nature of the sorted cells was obtained by staining the cells for the presence of intracellular haemoglobin, which is present even in nucleated red blood cell precursors. As with the MGG staining, there was consistently high background staining on the slides (presumably due to cell lysis during the cytospin process depositing free haemoglobin on the slides). However, intact nucleated cells had high uptake of the haemoglobin stain, suggesting an erythroid origin (Figure 3-7 B).

3.4.2.1.2 CD235a^{+}71^{+} cells exist within AML samples, and appear cytogenetically normal

Having optimised and validated a flow based technique to sort CD235a^{+}71^{+} erythroblasts from normal bone marrow samples, we then used the same technique to identify residual erythroblast populations within diagnostic AML samples.

Samples with a haemoglobin level of greater than 10g/dL, a known detectable leukaemia associated genetic lesion and a diagnostic bone marrow sample available in the Bart's Tissue Bank were identified and underwent sorting as described in Section 3.3.3.1. Ten samples were studied in total. Their clinical characteristics are detailed in Table 3-2. Seven had karyotypic abnormalities, and three had a normal karyotype but a mutation in the *NPM1* gene. Those samples with karyotypic abnormalities underwent sorting, with the resultant sorted cells undergoing fixation and FISH staining with relevant probes. The three *NPM1* mutated samples underwent immediate DNA extraction of sorted populations of cells and quantitative PCR for the presence of the *NPM1* mutation as detailed in Section 2.8.4.5.

Table 3-2 Clinical and genetic characteristics of 10 patients with preserved haemoglobin at diagnosis which underwent sorting and clonality assessment of their erythroblast populations

PATIENT DETAILS			**GENETICS**	
PATIENT ID	**AGE**	**SEX**	**KARYOTYPE**	***NPM1* STATUS**
6318	33	F	NORMAL	TYPE A
6580	59	M	NORMAL	TYPE A
7033	55	F	NORMAL	TYPE A
6317	22	M	t(8;21)	
8750	47	M	t(8;21)	
9220	55	F	t(8;21)	
6988	55	M	INV(16)	
7036	70	F	INV(16)	
7248	27	F	INV(16)	
7696	32	F	INV(16)	

3.4.2.1.2.1 The $CD235a^{+}71^{+}$ cells within *NPM1* mutated samples are genetically normal and not part of the leukaemia clone

Assessment of the clonality of the sorted $CD235a^{+}71^{+}$ cells derived from the three AML samples with preserved haemoglobin levels at diagnosis and known *NPM1* mutation was performed. Results are shown in Table 3-3. Percentages of cells within the $CD235a^{+}71^{+}$ population containing *NPM1* mutated DNA were extremely low (1 to 15%), comparable with the sorted negative lymphocyte controls, and compatible with accepting a sort purity of greater than 90%. $CD235a^{-}71^{-}3^{-}$ cells (likely blasts) were separated as a positive control for each sample; $CD235a^{-}71^{-}3^{+}$ lymphocytes were separated concurrently as a negative control.

We conclude from these figures that the $CD235a^{+}71^{+}$ population of erythroblasts from these *NPM1* mutated AMLs does not contain the mutation at significant levels, and therefore that the resultant mature erythrocyte population has not been derived from the malignant blast clone.

Table 3-3 Percentage of *NPM1* mutated cells within sorted erythroid, lymphoid and blast populations from *NPM1* mutated AML samples

PATIENT ID	% *NPM1* MUTATED $CD235a^{+}71^{+}3^{-}$ CELLS (ERYTHROBLASTS)	% *NPM1* MUTATED $CD235a^{-}71^{-}3^{-}$ CELLS (BLASTS)	% *NPM1* MUTATED $CD235a^{-}71^{-}3^{+}$ CELLS (LYMPHOCYTES)
6580	15	87	5
7033	10	114	1
6318	1	79	40

3.4.2.1.2.2 The $CD235a^{+}71^{+}$ cells within CBF mutated samples are genetically normal and not part of the leukaemia clone

Seven AML samples with identified karyotypic abnormalities underwent sorting for their $CD235a^{+}71^{+}$ erythroblast populations. Of these, four had inv(16) and three t(8;21). The results of FISH analysis of the sorted populations (and also concurrent analysis of the $CD235a^{-}71^{-}3^{-}$ population of presumed blasts which acted as a positive control) are shown in Table 3-4.

In six samples, slide quality was sufficient and enough cells present to make FISH analysis possible. In all cases, sorted $CD235a^{+}71^{+}$ cells were free of the clonal translocation which characterised the leukaemic cell population, and therefore it is assumed that their mature red cell populations were also derived from cells of normal karyotype rather than the leukaemia clone.

Table 3-4 FISH analysis of $CD235a^+CD71^+$ cells from AML samples with known karyotypic abnormalities

PATIENT ID	KARYOTYPE	FISH PROBE	PROPORTION OF POSITIVE $CD235a^+71^+$ CELLS	PROPORTION OF POSITIVE $CD235a^-71^-$ CELLS
6988	INV(16)	Cytocell CBFB-MYH11	0/100	74/100
7036	INV(16/0	Cytocell CBFB-MYH11	0/100	50/100
7248	INV(16)	Cytocell CBFB-MYH11	0/100	80/100
7696	INV(16)	Cytocell CBFB-MYH11	0/100	75/100
6317	t(8;21)	Cytocell AML1/ETO (RUNX1/RUNX1T1)	INSUFFICIENT CELLS ON SLIDE (671)	8/10
8750	t(8;21)	Cytocell AML1/ETO (RUNX1/RUNX1T1)	0/100	77/100
9220	t(8;21)	Cytocell AML1/ETO (RUNX1/RUNX1T1)	0/50	48/50

3.4.2.2 There is no relationship between haematopoietic stem cell concentration and haemoglobin levels at presentation

Chapter 5 details a large body of work identifying HSC concentrations in bone marrow samples at diagnosis with AML. To summarise the main findings, there was no statistically significant difference found between the concentration of HSCs in AML diagnostic BM samples (with a range of presenting blood counts) and controls (all of whom had normal blood counts). Nevertheless, the same patient series did identify that very low stem cell concentrations at diagnosis (<0.1 HSCs/µl) identified a subgroup of AML patients with a significantly worse overall survival.

Within this cohort of 38 patients, and using the same categorisation of samples into low stem cells at diagnosis ([HSC]<0.1/µl)(n=8) and high stem cells at diagnosis ([HSC]>0.1µl)(n=30), there is no significant difference found between these groups in observed haemoglobin levels at diagnosis (t test, p=0.8). This result is shown in Figure 3-8 A.

3.4.2.3 Does the concentration of $CD235a^+71^+$ erythroblasts at diagnosis with AML have clinical significance?

20 diagnostic AML samples and 7 control samples, which had been analysed whilst fresh and their $CD34^+$ concentration quantified, were subsequently analysed concurrently for their concentrations of $CD34^+38^-ALDH^{high}CLL1^-$ stem cells and $CD235a^+71^+$ erythroid precursors. Patient clinical details, including presenting full blood counts, calculated [$CD34^+$] in fresh sample, [$CD34^+38^-ALDH^{high}CLL1^-$] and [$CD235a^+CD71^+$] are shown in Table 3-5 (AML samples) and Table 3-6 (Control samples).

3.4.2.3.1 There is a significant correlation between the concentrations of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ stem cells and $CD235a^{+}71^{+}$ erythroblasts at diagnosis with AML

AML samples were divided, as described in Chapter 5, between those with low concentrations of HSCs at diagnosis (<0.1 cells/µl) and relatively high concentrations of HSCs at diagnosis (>0.1 cell/µl). Those with a low HSC concentration had a significantly lower number of erythroid precursors present at diagnosis (median 0.7 cells/µl, n=4), compared to those with relatively high HSC levels (median 174 cells/µl, n=20)(p=0.04, Mann Whitney t-test).

3.4.2.3.2 There is no relationship between the concentrations of $CD235a^{+}71^{+}$ erythroblasts in the bone marrow and haemoglobin levels at diagnosis with AML

Spearman's correlation test was used to analyse these datasets for a possible relationship between erythroblast concentration at diagnosis with AML and presentation haemoglobin level. No significant correlation was found (Spearman's r correlation coefficient 0.14, p=0.56).

3.4.2.3.3 There is no difference in erythroblast concentrations between AML samples and controls

In the group's previously published work into the effect of AML on the haematopoietic hierarchy in $CD34^{low}$ AMLs, whilst HSC concentrations were preserved between controls and AML samples, levels of precursor cells ($CD34^{+}38^{+}$ cells) were significantly lower in AML samples than controls (see Section 1.5.2)[1].

Although CD34 is expressed on many erythroblasts at the start of their maturation, its expression wanes as that of CD235a increases, and therefore we would not expect all erythroblasts to fall within the $CD34^{+}38^{+}$ subset. Nevertheless, as a precursor population, one might expect the erythroblast population to fall in a manner consistent with the hypothesised differentiation block.

In fact, in this dataset of 20 patients, whilst there is no significant difference in HSC concentration at diagnosis between AML and control samples ([HSC] AML samples median 2.4 cells/µl (n=20), [HSC] control samples median 2.3 cells/µl (n=7) (p=0.72, Mann Whitney t test)), there is also no significant difference observed between erythroblast concentrations ([erythroblast] AML samples median 3.7 cells/µl (n=20); [erythroblast] control samples, median 89.6 cells/µl)(p=0.79, Mann Whitney t test). These data are shown graphically in Figure 3-8 B.

3.4.3 Attempts to study megakaryocyte lineage and concentration at diagnosis with AML

3.4.3.1 FACS based separation of megakaryocytes from bone marrow samples previously stored in Bart's Tissue Bank by Standard Methodology is not possible

Previous published work describing the separation of megakaryocytes from bone marrow samples suggests these cells are a rare, $FSC^{high}CD41^{+}61^{+}$ population in the bone marrow.

Repeated attempts to develop a methodology capable of separation of intact megakaryocytes from control bone marrows by FACS based sorting proved unsuccessful. FACS staining of normal control bone marrows revealed a population of cells of $FSC^{high}CD41a^{+}$ cells in small numbers (300 to 400 sorted cells per approximately every $10x10^{6}$ nucleated bone marrow cells). However, after sorting, slide preparation and MGG staining, the morphology of the cells examined did not correlate with the distinctive appearance of megakaryocytes. This was not a surprise given the previously published literature on this area.

Review of the available literature on megakaryocyte separation frequently describes technical difficulties attributed to cell rarity, fragility and tendency to aggregate[185]. Groups which have published on FACS analysis of megakaryocytes collect and process bone marrow in specific buffers designed to reduce megakaryocyte aggregation; followed by density gradient centrifugation designed to enrich for the megakaryocyte population[185]. As all the BM samples identified for this study have been pre-frozen by our standard Tissue Bank technique (Section 2.2.2), and new cases are of sufficient rarity to preclude analysis of bone marrows before freezing, we have come to the conclusion that we are unable to separate megakaryocytes successfully from the BM samples which we have stored in this patient subgroup by flow based sorting techniques. An additional factor which added to concerns over FACS sorting accuracy was the documented adherence of platelets (also $CD41a^{+}$) to blast cells[195]. As a result, further attempts to utilise FACS to identify the lineage of platelet production in these patients were postponed.

3.4.3.2 Examination of diagnostic bone marrow trephine samples suggests that megakaryocyte concentration in AML are significantly reduced compared to controls

17 trephines taken from AML patients at diagnosis and 9 control bone marrows were studied for megakaryocyte numbers as described in Section 3.3.6 using the Pannoramic system. The median platelet count of the AML samples was $100x10^{9}$/L, with a range of between 13 to $169x10^{9}$/L. 9 of the AML samples were derived from patients with intermediate risk cytogenetics and 8 represented patients with poor risk cytogenetics.

Calculated megakaryocyte concentrations were significantly lower in AML bone marrow trephines than in control samples (AML samples, median [megakaryocyte] 2.9/mm^2 versus control samples, median [megakaryocyte] 13.8/mm^2; p=0.0014 by Mann-Whitney t-test). There was no statistically significant difference observed between the megakaryocyte concentrations of different cytogenetic risk group samples (intermediate risk, median [megakaryocyte] 3.6/mm^2 versus poor risk [megakaryocyte] 2.8/mm^2; p=0.44 by Mann-Whitney t-test)

3.4.3.2.1 There is no relationship between the concentrations of megakaryocytes in the bone marrow and platelet levels at diagnosis with AML

Spearman's correlation test was used to analyse this dataset for a possible relationship between megakaryocyte concentration at diagnosis with AML and presentation platelet levels. No significant correlation was found (Spearman's r correlation coefficient 0.07, p=0.78).

3.4.3.2.2 There is no relationship between haematopoietic stem cell concentration and platelet count levels at diagnosis

In a parallel with Section 3.4.2.2, the potential relationship between HSC concentration and platelet levels at diagnosis with AML was explored. The same cohort of 38 patients, whose HSC concentration at diagnosis was identified by the work performed in Chapter 5 were examined, and the same separation of the cohort into samples with HSC concentrations above and below 0.1 HSC/µl was used. As was the case with haemoglobin levels, HSC concentration at diagnosis does not appear to have a significant effect on platelet level at diagnosis (Mann-Whitney t test p=0.11) (see Figure 3-8 D).

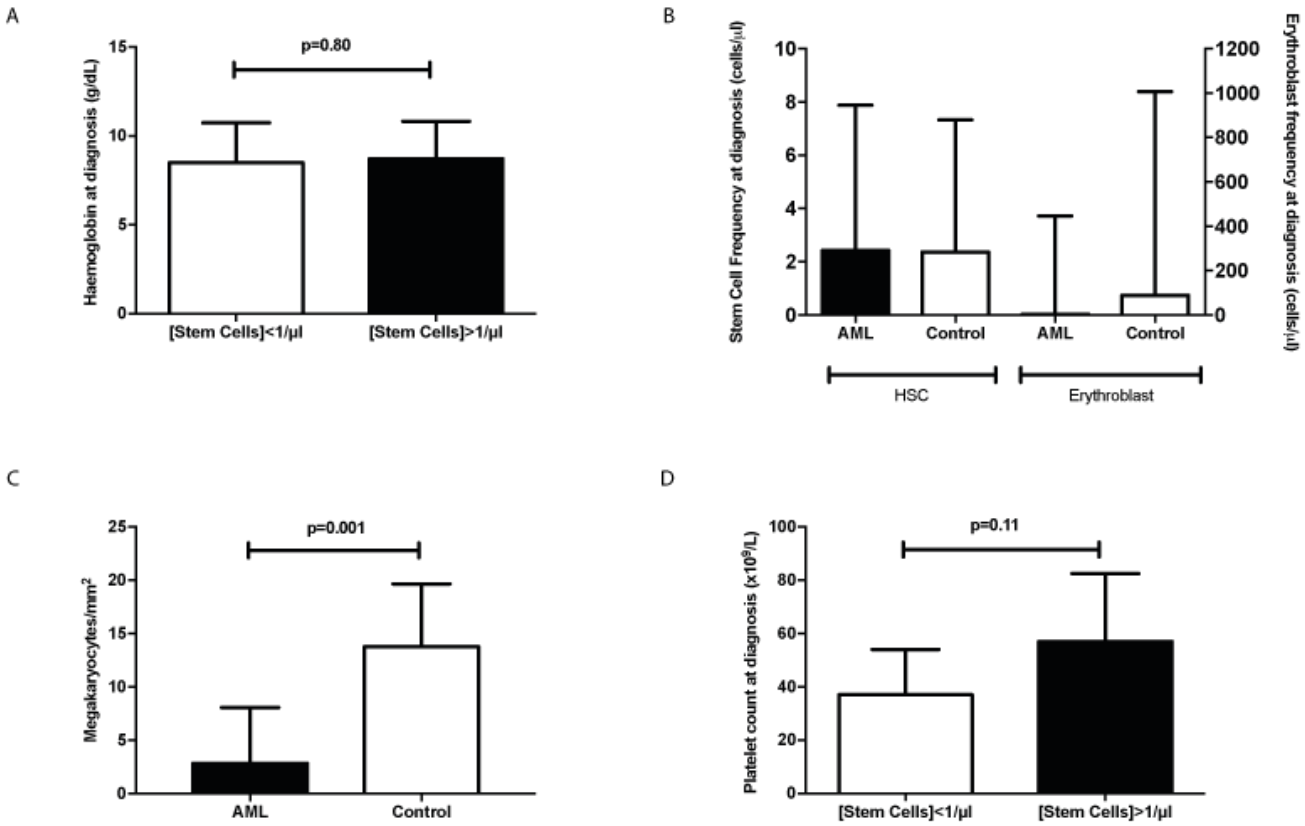

Figure 3-8 A: Bar-chart showing the relationship between HSC concentration & Hb at diagnosis. [HSC]<0.1/µl, n=8, mean Hb 8.5g/dL; [HSC]>0.1/µl, n=30, mean Hb 8.7g/dL. Mean and SD are plotted; B: Bar chart showing the relationship between HSC and erythroblast concentrations at diagnosis in AML and control samples C: Bar chart showing trephine megakaryocyte concentration at diagnosis with AML and controls. Median and IQR are plotted; D: Bar chart showing the relationship between HSC concentration and platelet count at diagnosis. [HSC]<0.1/µl, n=8, median platelets $37x10^9$/L; [HSC]>0.1/µl, n=30, median platelets $57x10^9$/L. Median platelet count and IQR are shown

Table 3-5 illustrating clinical characteristics, diagnosis blood counts and calculated stem cell and erythroblast concentrations in 20 AML diagnosis samples

PAT ID	AGE	SEX	GENETICS			DIAGNOSIS BLOOD COUNT			FRESH SAMPLE	FROZEN SAMPLE	
			KARYOTYPE	NPM1 STATUS	FLT3 STATUS	HB (g/dL)	WCC ($x10^9$/l)	PLTS ($x10^9$/L)	[$CD34^+$ CELLS] (CELLS/µl)	ESTIMATED ORIGINAL [$CD34^+38^-$ $ALDH^{high}CLL1^-$] (CELLS/µl)	ESTIMATED ORIGINAL [$CD235a^+CD71^+$] (CELLS/µl)
7774	62	M	NORMAL	TCTG		8.1	0	37	9.5	0.08	1.42
7793	56	F	13	TCTG		11.1	4	10	15.7	0.01	0.06
7916	62	F	NORMAL	WT		6.3	38	76	18.0	4.26	242.95
7955	61	M	NORMAL	WT		9.3	14	19	2943.6	8.31	3.48
7967	67	F	NORMAL	CCTG		10	2	18	328.5	2.06	3.93
8169	27	F	t(11;19)	WT		9.7	0	110	420.4	1.09	1382.02
8199	45	F	NORMAL	TCTG		6.6	22	41	31.8	1.29	157.99
8429	48	M	NORMAL	TCTG	FLT3-TKD	6.1	3	39	20.1	0.03	2.98
8452	27	M	t(11;17)	WT	WT	8.6	6	75	28.6	8.27	190.10
8791	59	M	COMPLEX	WT		7.4	31	43	797.6	5.36	2.93
8932	43	M	t(8;21)	WT	WT	7.2	3	54	67997.4	1.57	1.91
9067	67	M	NORMAL	CTTG	WT	7.4	143	172	18.6	0.04	0.02
9154	63	M	COMPLEX	WT	WT	10.3	4	129	46787.9	10.38	0.00
9669	41	M	INV(16)	WT	WT	11.1	12	70	21448.1	2.45	291.04
9672	61	M	DEL(7q)	WT		6.9	45	20	35640.0	2.41	0.00
9815	42	M	NORMAL	WT	WT	7.7	5	105	37365.7	71.81	1432.54
9900	61	F	NORMAL	WT	WT	9.1	1	239	3004.1	6.65	899.22
9963	63	F	COMPLEX	WT	WT	10.8	5	57	17705.4	17.17	1096.20
9964	24	1	NORMAL	TCTG	FLT3-ITD	13.9	77	80	14867.1	6.72	3.01
10203	67	1	NORMAL	WT	FLT3 ITD & TKD	7.8	99	27	147.0	1.24	498.69

Table 3-6 showing clinical characteristics, diagnosis blood counts and calculated stem cell and erythroblast concentrations at staging in 7 control samples

PAT ID	Age	Sex	DIAGNOSIS BLOOD COUNT			FRESH SAMPLE ANALYSIS	FROZEN SAMPLE ANALYSIS	
			HB (g/dL)	PLTS ($x10^9$/L)	WCC ($x10^9$/L)	[$CD34^+$ CELLS] (CELLS/µl)	ESTIMATED ORIGINAL [$CD34^+38^-ALDH^{high}CLL1^-$] (CELLS/µl)	ESTIMATED ORIGINAL [$CD235a^+71^+$](CELLS/µl)
7926	42	M	11.2	527	14.3	59.54	2.37	89.62
7957	63	M	12.5	295	5.2	1.58	0.03	2.60
8066	27	F	13.5	316	7.5	141.09	1.36	0.36
8115	42	M	16.5	318	8.1	353.99	13.80	1006.11
8510	21	F	13.1	587	7.4	232.21	3.07	192.17
TEMP1735	30	M	14.6	252	12.7	4.50	0.01	1.75
10115	50	M	15	294	12.6	1368.94	7.33	3174.38

3.5 Discussion

In the first chapter of this PhD, I have attempted to investigate a subset of AML samples which do not fit the pattern of trilineage haematopoietic failure classically associated with this disease. The existence of these patients presents a challenge to the previous hypothesis put forward by our group, namely that haematopoietic failure in AML is due to induced quiescence and differentiation block of the HSC population[1]. If such a universal pattern were the case, one would naively expect the same pattern of count deterioration to be observed in all patients at diagnosis.

3.5.1 Summary of database analysis

Analysis of the blood counts presenting counts at the time of diagnosis with AML for over 500 patients presenting to Bart's Hospital would suggest that trilineage failure is very strongly the normal pattern, with 81% of patients presenting with both haemoglobin and platelet levels below normal for age. Only 1% of patients were diagnosed with AML whilst maintaining a normal haemoglobin and platelet level.

Perhaps surprisingly, age and gender had no bearing on the levels of these parameters at presentation.

Detailed cytogenetic subanalysis showed no effect of different cytogenetic risk groups or specific translocations on haemoglobin levels at presentation.

The patterns of platelet count at diagnosis were more interesting, with a definite skewing of platelet counts (rather than preservation of a normal distribution), and a suggestion there is a substantial, and possibly separate, population of patients (15%) who present with a normal platelet count at diagnosis. However, I was unable to identify a clear cause for these patterns. Cytogenetic subanalysis identified only that patients with good risk disease were more likely to have a lower platelet count at diagnosis than those with intermediate and poor risk disease. Unsurprisingly, subgroup analysis showed this effect is largely due to the inclusion of patients with t(15;17). As these patients often present with DIC, it is impossible to determine if the lower platelet counts observed in this cohort are due to inefficient production or peripheral consumption.

No cytogenetic abnormalities were significantly associated with preservation of platelet counts in this cohort, although note is made of the patients with t(3;5), with a platelet count of 385×10^9/L, and the 2 patients with inv(3) with platelets of 192 and 263×10^9/L at diagnosis. Such

preservation of counts is keeping with small previous dataseries associating AML with abnormalities of 3q and relative thromobocytosis, although the small numbers of patients with these abnormalities in this cohort prevent these results from acheiving statistical significance[194]. .

Analysis of survival data for this group shows that neither gender-adjusted haemoglobin nor platelet level at diagnosis appears to have a significant effect on overall survival.

3.5.2 Clonality studies of erythroid precursors suggests red cells are derived from normal precursors and not from the leukaemic clone (in *NPM1* and CBF mutated AML samples), but were not possible with megakaryocytes

FACS based sorting of nucleated erythroid precursors was performed in AML samples with relative preservation of their red cell levels and known genetic mutations, to identify the clonality of the maturing red cell precursors. In all cases, these appeared to be of normal karyotype, and were therefore assumed not to be part of the leukaemic clone. In these patients with preserved haemoglobin levels at diagnosis, we have no evidence to suggest that normal counts are maintained by terminal differentiation of blasts.

However, although erythroblasts do not appear to contain *NPM1* or CBF mutations in the ten samples studied, these findings do not absolutely exclude the possibility of other AMLs carrying alternative cytogenetic or genetic mutations maturing into red cells of normal appearances. In order to confirm this, it would be worth performing FACS separation and analysis of erythroblasts from AMLs representing a wider range of genetic anomalies.

We attempted the same technique to study the genotype of megakaryocytes in bone marrow, in these patients with preserved platelet counts at diagnosis. Frustratingly, FACS based attempts to separate megakaryocytes proved unsuccessful, due we believe to their large size, and the need for specialist handling of bone marrow samples from the time of receipt (see Section 3.4.3.1). Other factors which might make FACS based separation of nucleated platelet-lineage precursors from AML BM difficult include the observed adherence of $CD41^+$ platelets to blasts, which would make confident distinguishing of megakaryocytes from blasts practically impossible[195]. Reassuringly (?!), similar difficulties of using flow based sorting techniques to separate intact megakaryocytes have been well documented by other groups[185].

Given the failure to adequately separate megakaryocytes from bone marrow samples by FACS sorting, a number of alternative approaches were considered.

By virtue of their distinct morphology, megakaryocytes are clearly identifiable in trephine sections. A significant number of AML patients diagnosed at Bart's Hospital have had bone marrow trephines taken at diagnosis, some of which are held for research purposes.

Genetic investigation of megakaryocytes located within trephine sections has been undertaken apparently successfully in the past: as referenced in Section 3.1.2.1.2, Falini's group reported laser dissection of megakaryocytes from diagnostic AML bone marrow trephines, which subsequently tested positive for the *NPM1* mutation. However, it is worth noting that no other group has been able to replicate this approach. The same group also have utilised staining for the presence of cytoplasmic NPM1 protein, as a surrogate marker for the presence of the *NPM1* mutated gene in studies of bone marrow sections[166]. However, using immunohistochemical techniques to identify the clonality of individual cells is not a widely available technique, and the antibodies are not commercially available.

In conjunction with Marianne Grantham in the Cytogenetics Department of the Royal London Hospital, we did attempt to see if we could utilise chromogenic *in situ* hybridisation (CISH) staining of bone marrow trephine sections of patients with known karyotypic abnormalities to identify whether megakaryocytes stain positive for the leukaemia associated mutations. On a first attempt staining AML bone marrow samples with the t(6;9) translocation, using the Kreatech DEK/NUP 214 t(6;9) Fusion Probe, probe binding was rendered inefficient by the fact that trephine sections processed at Bart's Hospital are universally fixed in formalin, with resultant DNA denaturation. This makes any attempt to identify clonality in these fixed samples by genetic means extremely difficult.

One approach we have not attempted would involve the sorting of platelets, followed by RNA extraction, reverse transcription, and interrogation of the resultant cDNA for mutations in genes such as *NPM1*. Issues with this could include documented difficulties relating to platelet adherence to other cell types, as well as generalised low transcription levels in platelets (indeed as the NPM1 protein is thought to be involved in nuclear-cytoplasmic shuttling, one wonders whether it has a function in enucleate platelets).

3.5.3 Studies of erythroid precursor levels

Having established that erythroblasts in AML appear to be derived from normal haematopoietic cells (at least in the cases of CBF and *NPM1* mutated samples), we then proceeded to calculate their concentration at diagnosis.

There was a wide variation in observed concentrations of erythroblasts between samples, but this did not correlate with presentation haemoglobin. Similarly, due possibly to the wide range of results, there was no significant difference observed between erythroblast concentration in AML and control samples. There was, however, a significant relationship between HSC concentration at presentation with AML and erythroblast levels (those with very low levels of HSCs had significantly lower erythroblast concentrations).

Perhaps it is unsurprising that we fail to observe a very clearly defined pattern between these parameters. Is it possible that under conditions of stress haematopoiesis, the more mature erythroblast forms (BFU-E) might retain the ability to move to extramedullary sites for the final stages in replication as we observe in rodent models?[172,173] The cells marked out as $CD235a^{+}71^{+}$ represent a range of stages in erythroblast maturation: it is possible that within these cohorts, certain populations are preserved or able to expand, and others are more affected. Could the final stages in red cell development such as nuclei extrusion and release into the bone marrow also be blocked by the presence of AML, therefore resulting in a failure to maintain peripheral counts?

One experimental concern on review of the data is whether the use of CD235a and CD71 by themselves is sufficient to exclude mature red cells from analysis. As mature red cells are present in very high relative concentrations in bone marrow samples, this would certainly significantly reduce the quality of the generated data. In the initial design of these experiments, I had hoped that the initial processing of the bone marrow samples (by density gradient centrifugation using Ficoll^{TM}, followed by freezing- see Section 2.2.2) would be sufficient to remove the vast majority of mature red cells. Secondarily, CD71 expression is reported to be downregulated on mature erythroid cells, and therefore I had hoped that selection of the $CD235a^{+}71^{+}$ dual positive population would be sufficient. In retrospect, the flow-based differentiation of erythroblasts would be improved by the addition of a step requiring the positive selection of cells with an intact nucleus.

3.5.4 Studies of platelet precursor levels

As megakaryocytes are easily identifiable in bone marrow trephine sections, I attempted to enumerate megakaryocytes in AML samples and compared these to controls. This approach suggested that megakaryocyte levels were significantly lower in AML patients than trephines taken from normal controls (2.9 versus 13.8 megakaryocytes/mm^2, p=0.001). This is in contrast to numbers of erythroid precursors, which did not appear to show significant reduction in their levels in bone marrow aspirates from AML patients when compared to controls.

In a parallel to the work with erythroid precursors, I attempted to analyse the different patterns observed in different cytogenetic risk groups. Here, I had no *NPM1* mutated or CBF samples: all the samples were either of intermediate or poor prognoses. There was no significant difference found between the megakaryocyte concentrations within these two risk groups, and insufficient samples were available to further subcategorise.

Again, in a situation analogous to the erythroid data, there is not a significant correlation between platelet count at diagnosis and megakaryocyte number in the bone marrow at diagnosis. Could this imply that platelet levels lag behind changes in megakaryocytes? Or could megakaryocytes similarly enter a state of induced quiescence similar to that observed in stem cells? Or are the initial changes in peripheral platelet count observed in AML more due to peripheral destruction initially than a failure in bone marrow production?

3.5.5 Further experimental work

Given unlimited time and resources, there are a number of further experiments I would wish to perform.

Whilst we failed to find an association between cytogenetic risk groups of AML and lineage preservation, in the current era of genomic sequencing, it would certainly be interesting to screen this cohort for commonly associated mutations in AML. This would allow us to identify if any of these mutations have a bearing on presentation blood counts, either because of leukaemia cells retaining the ability to differentiate into mature cells, or conferring variable inhibitory ability of the blasts on the development of cells of different lineages. However, it is noteworthy that the recently published study of 1540 patients from three German AML trials did not mention any observed association between particular genetic mutations and presentation counts[196].

Other alternative approaches could include the study of the blasts' ability to produce cytokines with known roles in lineage differentiation. This could be performed either at a transcriptional level, or by ELISA based assays of the bone marrow stromal milieu. Do, for example, those AMLs which present with a preserved platelet count secrete TPO themselves, or influence surrounding stromal cells to increase local production?

3.5.6 Summary

In conclusion, this chapter has confirmed the dominant pattern of bone marrow failure in the vast majority of AML patients. However, it is clear the patterns of lineage deterioration do vary

between patients, and the reasons for this are not clear. Evidence presented in this chapter would suggest this is not due to leukaemia differentiation, or intrinsic properties of the AML linked to its cytogenetic abnormalities.

However, I do not believe that the work presented in this chapter invalidates the group's hypothesis that the dominant mechanism of bone marrow failure seen in AML is that of induced quiescence of the stem cell population. However, I would argue that it is possible that different AML samples may well induce blocks on differentiation and maturation at a number of different stages in development, and the variable extent to which these occur may affect the fashion in which blood counts fall at presentation.

Further experimental work would be required to try and address these questions as detailed in Section 3.5.5 above. However, due to time and financial limitations in this work, we made the decision to focus on the interaction between HSCs and AML: work which is detailed in the following chapters.

Chapter 4 Developing a methodology for the identification of normal stem cells within AML Samples

4.1 Introduction

Our group historically has been interested in the fate of normal HSCs during the development of AML, and how any disruption in their behaviour may affect downstream haematopoiesis. Key to this work is the accurate identification of normal stem cells from within AML samples, and their differentiation from cells belonging to the leukaemic clone.

4.1.1 The identification of normal stem cells and progenitors within $CD34^{low}$ expressing AML samples

Previous work from our group summarized in Chapter 1.5.2.2 involved the identification of normal $CD34^{+}38^{-}$ stem cells from human primary bone marrow samples of the rare $CD34^{low}$ phenotype AML[1]. Separation of these HSCs from the leukaemia clone was facilitated by the fact that in this rare subtype of AML (highly associated with a normal karyotype, *NPM1* gene mutations and an absence of concurrent *FLT3* mutation), the $CD34^{+}$ cells found in the marrow are free of the leukaemia-associated mutation[130,135]. When the normal HSC concentration at diagnosis was calculated in these samples, it was found to be unchanged when compared to controls. In contrast, the number of $CD34^{+}38^{+}$ progenitor cells (also free of the *NPM1* mutation) was significantly reduced. Moreover, assessment of the cycling status of cells within the $CD34^{+}38^{-}$ bone marrow fraction suggested these cells are quiescent in comparison to controls[1]. These observations led to the development of our group's theory that AML induces a differentiation block in normal haematopoietic development between the HSC and progenitor stage.

4.1.2 Difficulties in identifying normal stem cells within AML samples

Identification of HSCs in bone marrow in health is made difficult because of their relative rarity. There is a lack of a universally agreed panel of markers to identify human HSCs, although it is widely accepted that the $CD34^{+}38^{-}$ fraction of bone marrow is the most enriched for cells with long term repopulating potential. However, there is considerable heterogeneity within this fraction, which contains not just cells with long term repopulating potential, but also short-term repopulating cells, and even cells beginning the pathway to differentiation.

This situation is complicated considerably when attempting to identify normal stem cells from AML samples, due to the overlap in phenotypes between the HSC and the leukaemia stem cell (LSC). In AML samples, many downstream blasts lacking LSC properties may also fall into the

$CD34^{+}38^{-}$ fraction, and due to the hyperproliferative nature of the disease, will constitute the vast bulk of cells seen at diagnosis.

A final consideration in the identification of normal HSCs is the possibility that exposure to AML may directly modulate their normal surface antigen expression. Such a shifting phenotype is not a widely-documented phenomenon, although a changing surface phenotype of long term repopulating cells after 5-FU chemotherapy has been documented in mice[197]. However, in Chapter 6 of this thesis, we demonstrate that CD33 expression within the normal HSC fraction of AMLs with t(8;21) and inv(16) mutations is temporarily down-regulated at diagnosis, returning to normal levels in remission. It is therefore possible that the expression of other key antigenic identifiers of cells with long-term repopulating ability may be similarly altered.

4.1.3 Antigenic markers reported to show differential expression between normal stem and leukaemic cells

There is considerable interest in isolating phenotypic markers which are differentially expressed between HSC and leukaemic blasts or LSCs. Not only do these markers represent a useful research tool, but they also present a target for antibody therapy, if relatively overexpressed on LSCs[131,198-200]. Given the diversity of AML samples, reflected in their varying phenotypic appearances, it is not surprising that there is no single, universal marker reported to distinguish normal HSCs from blasts in all cases.

A number of methods have been used to identify differentially expressed markers in HSCs and blasts: these include bulk gene expression profiling[198], which helps to identify differential expression of known extracellular antigens, phage display technology[201] and the signal sequence trap strategy methodology which identifies mRNAs coding for cell surface proteins by virtue of their ability to redirect a constitutively active version of c-mpl to the cell surface[202].

Irrespective of the manner of their identification, the biological function of many of these markers is often poorly understood. Therefore the reasons for the differential expression patterns remain unclear.

A summary of the best characterised extra and intracellular markers for separating HSCs and blasts is found below.

4.1.3.1 Aldehyde Dehydrogenase

Aldehyde dehydrogenase (ALDH) is a cytosolic enzyme responsible for the oxidization of a variety of aldehydes, including Vitamin A. In healthy bone marrow, it is expressed at high levels in $CD34^{+}38^{-}$ HSCs and is thought to be responsible for their observed resistance to alkylating agents, such as cyclophosphamide[203]. An assay system designed to measure ALDH activity by the generation of a fluorescent substrate was developed and described first in 1999[204]. Utilising this system, approximately 80% of the $CD34^{+}38^{-}$ fraction stain $ALDH^{high}$ [205]. As cells mature (and thus CD38 expression increases), ALDH activity reduces[205,206].

In a key paper for this thesis, Gerber et al described the study of ALDH activity levels in 20 patients with newly diagnosed AML (the majority with inv(16) or t(8;21) karyotypes)[207]. The $CD34^{+}38^{-}$ compartments within these bone marrow samples were shown to have three sub-populations of cells based on ALDH activity: $ALDH^{high}$, $ALDH^{int}$ and $ALDH^{low}$. These staining patterns are illustrated in Figure 4-1. In 16 of the 20 patients studied, the following pattern was seen. The $CD34^{+}38^{-}ALDH^{high}$ populations were small, devoid of the leukaemia-associated mutation by FISH, and capable of reconstituting normal haematopoiesis when transplanted into mice. FISH analysis of the $CD34^{+}38^{-}ALDH^{int}$ and $ALDH^{low}$ populations showed these to contain cells of leukaemic origin. Transplantation of the $CD34^{+}38^{-}ALDH^{int}$ cells showed these to be capable of leukaemic engraftment of mice, suggesting this population might include LSCs.

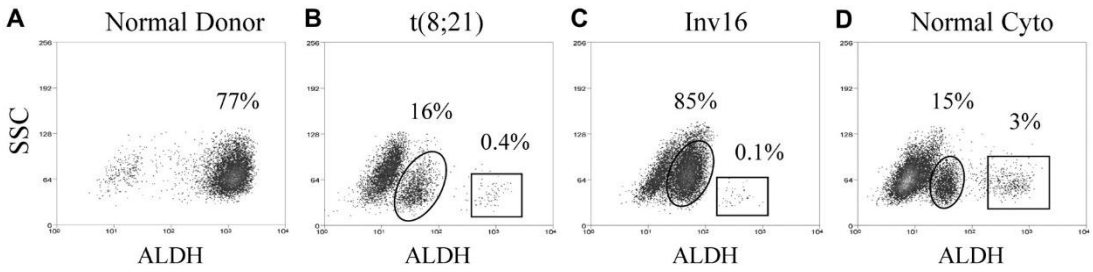

Figure 4-1 Representative flow plots showing ALDH activity against side scatter for a normal donor and 3 AML samples. A: normal donor; B: a patient with t(8;21) AML; C: a patient with inv(16) AML and D: a patient with AML with normal cytogenetics. Figure from Gerber et al, 2012[207].

Of the other four patients, two (one with normal karyotype and *FLT3*-ITD mutation; one with complex cytogenetics) did not contain a separate, small $CD34^{+}38^{-}ALDH^{high}$ population: most of the $CD34^{+}38^{-}$ cells had high ALDH activity (see below). The other two patients showed the phenotypic appearances we would describe as $CD34^{low}$, and therefore the diagnostic leukaemia marker was only found in the $CD34^{-}$ cells[130].

This paper went on to look at the clinical follow-up data available on 13 of the initial cohort, to see if ALDH expression could be used as an MRD marker. In this group, all of whom achieved a

morphological remission, the presence of a persistent $CD34^{+}38^{-}ALDH^{int}$ population in follow up samples was highly predictive for early relapse.

Further validation for this approach was provided by Schuurhuis et al, who tested ALDH expression patterns using a range of AML samples (although with a high proportion of $CD34^{low}$ AMLs)[205]. In this paper, they also make the observation that high expression of ALDH is often associated with the absence of CLL1 expression (see Section 4.4.1.3).

Schuurhuis et al went on to look at mRNA expression levels of the different isoforms of ALDH within $CD34^{+}38^{-}$ cells from normal bone marrow and AML samples. Normal stem cells showed significantly higher levels of the ALDH1A1 isoform than putative leukaemia stem cells (as defined in this paper by FSC/SSC^{high}, $CLL1^{+}$)[205].

As already mentioned, 2 of the 20 patients in the Gerber series did not have a distinct population of $ALDH^{high}$ cells within the $CD34^{+}38^{-}$ subgroup, due to overlap with a blast population with a similarly high ALDH expression. It has been noted in several studies that high ALDH expression within AML samples is associated with a poorer prognosis. Cheung et al studied 43 AML samples, and found the majority (29 samples) to have undetectable or rare cells (<5%, which should encompass all normal stem cells) within the $SSC^{low}ALDH^{high}$ gate. However, in 14 cases, a much larger $SSC^{low}ALDH^{high}$ population of cells was observed (median 15%, range 6 to 48%). Amongst other clinical variables, high ALDH expression of AML samples was significantly associated with poor risk cytogenetics[208].

Similar findings were reported by Ran et al, who studied ALDH expression within 68 AML samples. They reported that the frequency of $ALDH^{high}$ expressing cells within the samples was significantly associated with reduced survival ($p=0.025$) and also poor risk cytogenetics ($p<0.05$)[209].

4.1.3.2 CD33

CD33 is a transmembrane receptor frequently, but not exclusively, expressed on myeloid cells. It is known to be a member of the sialic acid binding family of receptors, although its physiological function in myeloid cells remains undetermined. The extracellular component comprises of two immunoglobulin domains, and the intracellular compartment contains immunoreceptor tyrosine-based inhibitory motifs that are implicated in inhibition of cellular activity[210].

CD33 is highly expressed in 90 to 95% of AML samples[211,212]. Over ten years ago, cross linking of the receptor with antibodies was shown to cause significant apoptosis in AML samples[213]. Increasing interest in the development of targeted molecular therapies led to the production of Gemtuzumab Ozogamicin (GO) (marketed in the UK as Mylotarg by Wyeth): an anti-CD33 antibody, conjugated to calicheamicin.

Although CD33 expression is classically associated with maturing myeloid cells, it has also been shown to be expressed at low intensity on normal, undifferentiated stem cells[214]. However, as we demonstrate in Chapter 6 of this thesis, CD33 expression appears to be downregulated on normal stem cells at diagnosis with AML (illustrated by studies of both $CD34^{low}$ and core binding factor-mutated AML), and therefore is a candidate marker for use in separation of HSCs and blasts.

4.1.3.3 CD47

CD47 is an immunoglobulin-like cell surface marker that interacts with the SIRPα receptor on macrophages to inhibit phagocytosis. CD47 has been implicated in a variety of different functions across cell types, including neutrophil migration and T cell co-stimulation. In a paper exploring the role of CD47 on mobilized stem cells, Jaiswal et al showed CD47 expression on mobilized cells in human cord and G-CSF mobilized cells to be significantly higher than normal controls[215]. Administration of G-CSF to mice led to a 4 fold increase in CD47 expression on murine HSC $cKit^+$ cells.

Observations from a CD47 knockout murine mouse model had previously described a protective role for CD47 in preventing macrophage-mediated removal of red cells and platelets. Using mice transplanted with murine HSCs heterozygous for the CD47 knockout- IAP +/-, Jaiswal et al showed that the transplanted mice tended to lose their donor granulocyte chimerism over time, suggesting the same protective role may also be the case for HSCs.

The same group simultaneously reported flow data on CD47 expression from 13 cases of AML (11 with normal cytogenetics; 1 inv(16) and 1 complex), and found the CD47 expression to be higher within the $CD34^+38^-90^+Lin^-$ fraction of AML samples than controls[131]. Gene expression data from 285 AML samples suggest that levels of CD47 expression are similar between different subsets of AML, apart from patients with the t(8;21) translocation, which are associated with lower expression levels. Higher CD47 expression is strongly associated with mutations within the *FLT3* gene. This paper described the FACS sorting of one AML sample (*FLT3* mutated) to $Lin^-CD34^+38^-CD47^{low}$ and $CD47^{high}$ subsets. The $CD47^{low}$ subset was capable

of reconstituting normal haematopoiesis when transplanted into mice, with the resultant cells being free of the *FLT3* mutation, which was present in the $CD47^{high}$ cells.

4.1.3.4 CD99

CD99 (also known as MIC2) is an O-glycosylated transmembrane protein. It is expressed across all leucocyte subsets, but is also found in endothelial cells and fibroblasts. In oncology, its expression is best characterized in Ewing's sarcoma and ALL.

Zhang et al studied the CD99 expression patterns on a variety of myeloid derived tumours, including AML, MDS and chloromas. Using immunostaining of bone marrow trephines from 30 patients with AML, the authors found 13 of these to be positive for CD99 expression[216]. More recently, a group from the Memorial-Sloan Hospital studied CD99 expression in 78 AML bone marrow samples by flow cytometry, and found 81% of these to be positive[200]. Limiting dilution xenotransplant experiments of the $CD34^{+}38^{-}90^{-}45RA^{+}99^{high}$ and $CD34^{+}38^{-}90^{-}45RA^{+}99^{low}$ cells of AML samples suggest that cells with LSC function reside in the $CD99^{high}$ expressing subset. It is interesting that study of a larger patient cohort (309 patients recruited to the ECOG 1900 study) suggested that CD99 transcript expression positively correlated with overall survival ($p=0.001$). Neither of these papers described the CD99 expression patterns within the $CD34^{+}38^{-}$ fraction of normal bone marrow.

In their key 2012 paper, describing the evolution of the pre-leukaemic stem cell, Jan et al used a flow panel which included CD99 to separate out normal stem cells from 6 patient samples with *FLT3* mutated AML. The $CD34^{+}38^{-}TIM3^{-}99^{-}$ cell population was sorted out from these (presumably highly selected) samples, and the resultant cells found to be clear of the leukaemia defining *FLT3* mutation[151].

4.1.3.5 CD300

CD300 (also known as IREM-1) is a type I transmembrane protein first identified in dendritic cells[217]. Work by Korver et al resulted in the production of a monoclonal antibody against the extracellular portion of CD300, which enabled a full exploration of its expression on blood and bone marrow cells. CD300 appears to be expressed on blood cells of the myeloid lineage, but not on lymphocytes or erythroid precursors. Of 54 diverse AML samples, 39/54 (72%) were found to be positive for IREM-1 expression by flow cytometry analysis. In 24 AML samples, where the IREM expression of the $CD34^{+}38^{-}$ subfraction was studied, 13/24 (54%) were found to be positive. In contrast, in 11 normal bone marrow samples, all $CD34^{+}38^{-}$ subfractions were negative for IREM1 expression. The paper then goes on to investigate the utility of IREM-targeted mAbs for the treatment of AML[218].

4.1.3.6 CLL1

C-type lectin-like molecule-1 or CLL1 (also known as Clec12A) was first identified in 2004 by members of the Schuurhuis laboratory, searching for novel AML surface targets for antibody therapy[201].

They used Phage Display technology to identify single chain Fv fragments from a large naive library capable of reacting with a panel of AML cells. Subsequent expression cloning identified the antigen recognized by the antibody as CLL1, a previously uncharacterized transmembrane glycoprotein. The function of CLL1 is still not known, but structural analysis suggested a possible role as a signalling receptor, based on an intracellular domain containing both immunotyrosine-based inhibition and YXXM motifs.

In the same paper, the group describe the subsequent production of a human anti-CLL1 IgG1 monoclonal antibody, generated from the single chain Fv fragment. A conjugated form of this antibody was then used to investigate the expression pattern of CLL1 across haematopoietic tissues. CLL1 expression appears to be restricted to the cells of the myeloid lineage. It is absent on $CD34^{+}38^{-}$ stem cell fractions of bone marrow, but increasingly expressed on more mature subsets such as $CD34^{+}38^{+}$ progenitors. 68 of 74 AML samples assayed for CLL1 were positive for its expression.

The same group went onto publish details of CLL1 expression within the $CD34^{+}38^{-}$ fraction of both AML and normal bone marrow samples[219]. Using a large cohort of AML samples, they found the $CD34^{+}38^{-}$ fraction to be positive for CLL1 expression in the vast majority (77 of 89). The $CD34^{+}38^{-}$ fraction of normal bone marrow samples was universally negative for CLL1 expression. They proposed a role for CLL1 as an MRD marker, having found that the persistence of a $CD34^{+}38^{-}CLL^{+}$ population after the completion of chemotherapy was associated with an increased relapse rate. Other groups have proposed a similar role after corroborative studies[220].

4.1.3.7 TIM3

T-cell Immunoglobulin Mucin 3 was originally found as a surface marker on $CD4^{+}$Th1 lymphocytes in mice, and is known to be an importance regulator in Th1 cell immunity and tolerance induction[221].

Kikushige et al[198] separated out the $CD34^{+}38^{-}$ cell fractions of 12 patients with AML and 5 controls, and performed cDNA microarray analysis looking for differential expression of genes, with the aim of identifying surface markers expressed more strongly on AML than normal

HSCs. Genes coding for surface markers with a greater than eight fold difference in their expression between the two test groups included *CLL1*, *CD96* and *TIM3*.

Expression of TIM3 in 33 AML samples, crucially representing all FAB subtypes, was then assayed by flow cytometry. Within the $CD34^{+}38^{-}$ cells in these AML samples, TIM3 expression was high in M0, M1, M2 and M4 subtypes. TIM3 expression was low in M3 AMLs, and more variable in the M5 and M7 cases studied. Transplant experiments of sorted cells into mice from 4 AML samples suggest that the LSC activity (as measured by the ability to produce AML in mice), resides in the $CD34^{+}38^{-}TIM3^{+}$ rather than $CD34^{+}38^{-}TIM3^{-}$ fraction.

The Weissman laboratory has also investigated TIM3 expression in AML. Jan et al tested TIM3 expression on 22 patient AML samples, representing a range of diagnostic karyotypes[222]. They found the TIM3 expression within the $Lin^{-}CD34^{+}38^{-}90^{+}$ cells to be significantly higher in AML-derived samples than controls. Transplantation of human HSCs, separated by their TIM3 expression status, into an immunodeficient murine model, revealed preferential engraftment from the $TIM3^{low}$ fraction, suggesting that this is this fraction that is relatively enriched for the presence of long term repopulating cells.

Analysis of the Weissman AML gene expression data suggested that although TIM3 expression was raised across all AML subtypes, it was significantly higher in AML samples with mutations in the core binding factors, and *CEBPA* gene.

Crucially, they also went on to test the ability of TIM3 expression to differentiate between normal HSCs and LSCS. In this paper, they describe the selection of 8 primary AML samples. 6 of these samples were sorted into $Lin^{-}34^{+}$ fractions and then subdivided into $TIM3^{+}$ and $TIM3^{-}$ subgroups. The purity of these populations was then tested as follows. 3 of 6 samples were cultured in methylcellulose, and the resultant colonies tested for the presence of the leukaemia-associated mutation (*FLT3* or inv(16)). In all 3 cases, at the end of the culture period, the molecular mutation was not detected in the $TIM3^{-}$ fraction progeny. The other 3 samples (2 with *FLT3*-ITD; one with *MLL* rearrangement) were transplanted into mice, and the resultant human haematopoietic populations studied. Studies of the progeny cells from the TIM3- fractions showed one to be free of the *FLT3* mutation, one contain a mixture of *FLT3*-ITD and wild type cells, and one to contain a significant proportion of cells still positive for the MLL translocation.

4.1.3.8 Summary of reported attempts to use differential surface antigen expression in the separation of normal HSCs and AML cells

None of the authors of these papers, recognizing the heterogeneity of AML samples, argue that their antigenic marker of preference will distinguish HSCs from AML cells in all cases. Each group within the field tends to use their own preferred antigen or combination of antigens, based on local experience. For example, our laboratory has to date favoured the use of CD34 and CD38 expression to identify normal HSCs from the $CD34^{low}$ subset of AMLs. The Majeti/Weissman group in Stanford, who have published extensively in this area and have produced some of the seminal papers in this field, have variously used a combination of CD34, 38, in combination with CD90, CD47 and TIM3 to differentiate normal HSCs from AML blasts in a variety of leukaemia subtypes.

4.2 Aims and Objectives

The primary aim of the work described in this chapter was to identify a methodology to separate normal HSCs from AML cells across all disease subtypes.

The first objective was to test a range of the antigenic markers previously reported to be able to distinguish HSCs and LSCs, and assess if we could reproduce these findings in our laboratory.

The secondary objective was to combine the most promising of these markers together to construct a panel of antigenic markers, capable of differentiating HSCs and LSCs in as many unselected samples as possible.

The final objective was to validate this approach using a cohort of AML samples, chosen to reflect the genetic and phenotypic diversity of the disease.

4.3 Chapter specific methods

4.3.1 Flow Based Analysis of Stem Cell Marker Expression

Diagnostic bone marrow and peripheral blood samples from patients diagnosed with $CD34^{high}$ and $CD34^{low}$ AML were obtained from the Bart's Tissue Bank with prior patient consent.

Samples were thawed at 37°C in a water bath for 2 minutes. 500µl DNase was then added to each tube to reduce cell clumping, and cells incubated at room temperature for 2 minutes. Samples were washed with 20ml 2% PBS. Cells were centrifuged (1500 rpm, 5 minutes) and re-suspended in 2ml 2% PBS. Cell counting was performed using the Beckmann Coulter Vicell XR Cell Viability Counter.

$2x10^6$ cells were pelleted and re-suspended in Aldefluor™ Assay Buffer (Stem Cell Technologies, Cat: 01702) to a concentration of $1x10^6$ cells/ml. Activated Aldefluor™ reagent (Stem Cell Technologies, Cat: 01703) was added at a concentration of 5µl reagent/ml. The suspension was mixed well, and incubated for 60 minutes at 37°C in a pre-warmed water bath.

Samples were pelleted and re-suspended in 2% HAG pre-diluted in Aldefluor™ Assay Buffer at a concentration of 50µl 2% HAG/$1.5x10^6$ cells. Cells were incubated at 4°C for 20 minutes. CD34-PerCP and CD38-Pecy7 were added to all samples, plus varying combinations of the following antibodies under investigation: CD33-FITC (BD, Cat: 555625 (Clone HIM 3-4)), CD33-APC (BD, Cat: 551378 (Clone WM53)), CLL1-PE (BD, Cat: 562566 (Clone 50C1)), CD300-PE (Biolegend, Cat: 340604 (Clone UP-DZ)), CD47-FITC (BD, Cat: 556045 (Clone B6H12)), TIM3-APC (R&D Systems, Cat: FAB2365A (Clone 344823)) and CD99-PE (BD, Cat: 555689 (Clone TÜ12)). Antibodies were added at a concentration of 5µl/$1x10^6$ cells, with the exception of CD47, CD99 and CD300, which were added at 20µl/$1x10^6$ cells. All were left to incubate for 30 minutes at 4°C. The cells were then washed with 4ml Aldefluor™ Assay Buffer, and pelleted.

Cells were re-suspended pre-analysis into 300µl stock solution of Aldefluor™ Assay Buffer/DAPI, and the sample analysed using a BD Biosciences LSR.

4.3.2 FACS Sorting

4.3.2.1 Standard Method

Samples were thawed at 37°C in a water bath for 2 minutes, and 500µl DNase added to each tube to reduce cell clumping, and incubated at room temperature for 2 minutes. Samples were washed with 20ml 2% PBS. Cells were centrifuged (1500rpm, 5 minutes) and re-suspended in

2ml 2% PBS. Cell counting and viability assessment was performed using the Beckmann Coulter Vicell XR Cell Viability Counter.

Cells were then pelleted and re-suspended in 2% HAG at a concentration of 50µl 2% HAG/1.5x10^6 cells. Cells were incubated at 4^{o}C for 20 minutes.

CD34-PerCP and CD38-Pecy7 were added at a concentration of 5µl/1.5 x10^6 cells, and left to incubate for 30 minutes at 4^{o}C. The cells were washed with 20ml 2% PBS, and pelleted.

Cells were re-suspended in a stock solution of 2% PBS/DAPI/DNase (to distinguish live and dead cells) at a concentration of 8x10^6 cells/ml, and filtered immediately pre-sort to reduced machine blockages.

Cell sorting was performed on a BD Biosciences Aria 2. Cells were sorted into 2ml 2% PBS in polypropylene tubes (BD Falcon, Cat: 352063). Purity was assessed by centrifuging and re-suspending cells in 2% PBS/DAPI/DNase prior to re-analysis. If sort purity was assessed as less than 90%, resorting was performed.

Sorted cells were then pelleted (3000rpm, 10 minutes) and the supernatant removed using a Gilson Safe Aspiration Centre.

4.3.2.2 Sorted Cells (ALDH Method)

Samples were thawed, washed, and viable cells counted as described above.

Cells were pelleted and re-suspended in Aldefluor™ Assay Buffer to a concentration of 1x10^6 cells/ml. Activated Aldefluor™ reagent was added at a concentration of 5µl reagent/ml. The suspension was mixed well, and incubated for 60 minutes at 37^{o}C in a pre-warmed water bath.

Samples were pelleted and re-suspended in 2% HAG pre-diluted in Aldefluor™ Assay Buffer at a concentration of 50µl 2% HAG/1.5x10^6 cells. Cells were incubated at 4^{o}C for 20 minutes. Antibodies to CD34-PerCP, CD38-Pecy7 and CLL1 were then added at a concentration of 5µl/1.5x10^6 cells, and left to incubate for 30 minutes at 4^{o}C. The cells were then washed with 10ml Aldefluor™ Assay Buffer, and pelleted.

Cells were resuspended pre-sort into a stock solution of Aldefluor™ Assay Buffer/DAPI/DNase to a concentration of 8x10^6 cells/ml, and filtered immediately pre-sort to reduced machine blockages.

Sorting, purity assessment and pelleting were performed as described above: the only exception being that cells were sorted into Aldefluor™ Assay Buffer rather than 2% PBS.

4.3.3 In vitro assays of stem cell function

4.3.3.1 Methylcellulose assay for 2 and 4 week culture

Please see Section 2.6.1.

4.3.3.2 Long term Culture Initiating Cell Assay

Please see Section 2.6.2.

4.3.4 Assays to test for the presence of leukaemia associated mutations

4.3.4.1 FISH

Slides were made as previously described and transported to the RLH for staining and analysis by Marianne Grantham, head of the Cytogenetics Department. The technique used for FISH staining is described in Methods 2.7.3.

The following probes were used for analysis of sorted populations of cells in this chapter: Cytocell CBFB-MYH11 Probe (Cat: LPH 022) for detection of inv(16); Cytocell PML-RARA Probe (Cat: LPH 023) for detection of t(15;17); Cytocell AML1/ETO (RUNX1/RUNX1T1) Probe (Cat: LPH 026) for the detection of t(8;21); Kreatech DEK/NUP 214 t(6;9) Probe for the detection of t(6;9); Cytocell Del(7q) Probe (Cat: LPH 025) for the detection of del(7); Cytocell MLL (KMT2A) Probe (Cat: LPH 013) for the detection of trisomy 11 and Cytocell BCR/ABL/ASS1 (Cat: LPH 038) for the detection of Trisomy 9q34.

Table 2-1 contains a full description of the expected signal patterns seen with the FISH probes used within this thesis.

4.3.4.2 PCR based assays

4.3.4.2.1 qPCR Based Assay for the presence of *NPM1* mutation

Please see Section 2.8.4.5. for full methodology.

4.3.4.2.2 Assay for *FLT3*-TKD mutation

This assay is performed as part of the Molecular Genetics screen performed on all patients with newly diagnosed normal karyotype AML at Bart's Hospital. The methodology follows that described by Huang et al, and facilitates the simultaneous detection of mutations within the *NPM1* and *FLT3* genes [223].

As this test was only required for assessment of one patient's sorted cells, DNA was extracted as described in Section 2.8.1, and the sample processed as per the Hospital's SOP by Suzanne McElwaine, Head of the Molecular Genetics Department at the RLH.

In brief, a duplex assay was set up using two sets of fluorochrome-tagged primers in the same tube: one primer being specific for the juxtamembrane coding sequence involved in the ITD mutations, and one for the kinase domain sequence. After amplification, the fragments were separated by capillary gel electrophoresis and the fluorochrome signals analysed using a software program such as GeneScan. Subsequent EcoRV digestion of the PCR products would cleave the kinase domain fragment unless a TKD mutation at D835 (or I836) is present.

4.3.5 Clinical data

Data pertaining to treatment details, time to count recovery and clinical outcome were obtained by reference to hospital records. Final censoring of patients for survival analysis took place in April 2016.

Full data tables showing treatment and outcome for all relevant samples are to be found in Chapter 5 (Table 5-3 and Table 5-5).

4.3.6 Statistical analysis

The majority of statistical analysis was performed using the Prism software package. All data sets were tested for normality of distribution by the D'Agostino & Pearson omnibus normality test. If data were not normally distributed, a Mann-Whitney t-test was applied in the determination of statistical significance.

Clinical outcome data analysis was performed using the SPSS statistical package. Overall survival and relapse-free survival data were displayed graphically using Kaplan-Meier curves, and differences in outcome analysed by the log-rank assay.

Due to small sample size, multivariate analysis of clinical outcome data was not undertaken. The distribution of samples within categories of variables of known significance was analysed by the Fisher's Exact Test (for 2x2 contingency table), or the Freeman-Halton extension of the Fisher's Exact Test (for 3x2 contingency tables).

4.4 Results

4.4.1 Antigenic panel design for stem cell identification

4.4.1.1 Sample Information

AML diagnostic samples from 52 different patients were used to test a potential panel of 7 antigenic markers. These samples were a mixture of leucophoresis aliquots, bone marrow and peripheral blood. They represented samples from good, intermediate and poor risk karyotype AMLs, as shown in Table 4-1. Of these, 37/52 (71%) were from patients with $CD34^{+}$ AML, and 15/52 (29%) from patients with $CD34^{low}$ AML. The relative overrepresentation of $CD34^{low}$ AML samples in this group reflects the ease of accessing samples from a bank of leucophoresis samples from $CD34^{low}$ AML patients maintained by the group.

All samples were obtained from the Bart's Tissue Bank, with consent for their use having been previously obtained in accordance with the Declaration of Helsinki.

4.4.1.2 *The utility of individual antigens in separating out putative normal stem cell fractions from the $CD34^{+}38^{-}$ fraction of a range of AML samples*

All results are summarised in Table 4-1. The results pertaining to individual antigens are considered below.

4.4.1.2.1 ALDH expression

49 AML samples were tested for ALDH expression. Of these, 31/49 (63%) contained a small, discrete $ALDH^{high}$ population (of putative stem cells) within the $CD34^{+}38^{-}$ subfraction. In 12/49 samples (25%), the leukaemia sample itself showed high ALDH expression, therefore negating the potential usefulness of ALDH expression as a mechanism of detecting stem cells. In the remaining 6/49 samples (12%), although the $CD34^{+}38^{-}$ cells within the leukaemia were of either low or intermediate ALDH expression, there were no cells visible in the $ALDH^{high}$ gate, despite successful positive controls within the same ALDH staining batch. This might suggest an absence of stem cells within these samples. Examples of all three staining patterns are given in Figure 4-2.

Subanalysis focusing on the $CD34^{+}$ AML samples tested was then performed. These are the samples of the most interest to us, being the ones where CD34 and CD38 expression alone are insufficient to identify normal HSCs. 19/35 (54%) of $CD34^{+}$ AML samples contained a small, discrete $ALDH^{high}$ population within the $CD34^{+}38^{-}$ population. 11/35 (31%) of the leukaemia samples showed high ALDH expression, and 5/35 showed no cells with $ALDH^{high}$ expression.

Table 4-1 AML sample details and summary of results of antigen-based separation of $CD34^{+}38^{-}$ populations. Key: + antibody possibly separates out 2 populations within the $CD34^{+}38^{-}$ population; -: antibody does not separate out populations within the $CD34^{+}38^{-}$ population; Blank: not tested

SAMPLE INFORMATION			ANTIGEN						
ID	**CD34 STATUS**	**CYTOGENETICS**	**ALDH**	**TIM3**	**CLL1**	**CD33**	**CD99**	**CD300**	**CD47**
2069	POS	INV(16)	+	-	+	-	-	-	
2338	NEG		+	+	+	+	+	-	
2367	NEG		+	-	+	+			
3978	POS		LEUK HIGH	-	+	-	-	-	
4496	NEG		NO CELLS	-	+	+	+	-	
6217	POS	t(6;11)	LEUK HIGH		-	+			
6321	POS		LEUK HIGH		+	+			
6753	POS		NO CELLS	+	+	-	+	-	-
6796	POS	t(15;16)	LEUK HIGH	-	+	+	-		
7033	NEG	NORMAL	+	+	+	+	+	-	-
7192	POS	DEL 7q	+	+	+	-	-		
7522	POS	t(5;15)	LEUK HIGH	-	-	-	-		
7620	NEG	NORMAL	+	+					
7667	POS	NORMAL	+	+	+	+	-		
7781	POS		+	-	+	-			
7820	POS	t(9;22)	+	+	+	-	+	-	
7828	POS	COMPLEX	LEUK HIGH	-	-	-			-
7829	POS	+13	+	+	+	+	+	-	-
7840	POS	NORMAL	NO CELLS	-	+	-	-	-	
7843	POS		+	-	+	+			
7847	POS		LEUK HIGH		-	-			
7882	POS	INV(16)	+	-					
7913	POS	DEL 7q	NO CELLS		+				-
7967	POS	NORMAL	LEUK HIGH	-	+				-
7968	NEG	NORMAL	LEUK HIGH	-	-	-			
8145	POS	t(15;16)	+	-					-

SAMPLE INFO			ANTIGEN						
ID	**CD34 STATUS**	**KARYOTYPE**	**ALDH**	**TIM3**	**CLL1**	**CD33**	**CD99**	**CD300**	**CD47**
8148	POS	+13	+	-	+	-	+	-	
8169	NEG	t(11;19)	+	-	+	-		-	
8361	POS	t(6;9)	+	-	+	+			-
8458	NEG	NORMAL	+	-	-	+			
8492	POS	t(8;21)				-		-	-
8627	POS		NO CELLS	-					
9066	POS	t(9;22)	LEUK HIGH	-					-
9068	POS	DEL 7	+	-	+	+			
9332	POS	NORMAL	NO CELLS	-	+	-			
9346	NEG	+21	+	+	+	+			
9552	POS		LEUK HIGH	-	-	-	-		
9555	NEG		+	+	+	+	-	-	
9584	POS	DEL 5q	+	-	+	-			
9653	POS	DEL 1q	+	+	+	+			
9668	POS	NORMAL	+	-	+	-	-		-
9669	POS	INV(16)	+	-	+	-			
9672	POS	DEL 7q		-					
9713	NEG	NORMAL	+	+	+	+	+		-
9716	NEG	NORMAL	+	+	+	+			
9717	POS	DEL 7 and +21	+	+	+	-			
9718	POS		+	-	+	+			
9719	NEG		+	+	+	-	+	-	
9725	NEG	NORMAL	+	+	+	+			
9815	POS	NORMAL	+	-	+	-			
TEMP 0580	POS		LEUK HIGH	-	+	-			
TEMP 1497	NEG		+	+	+	+		-	

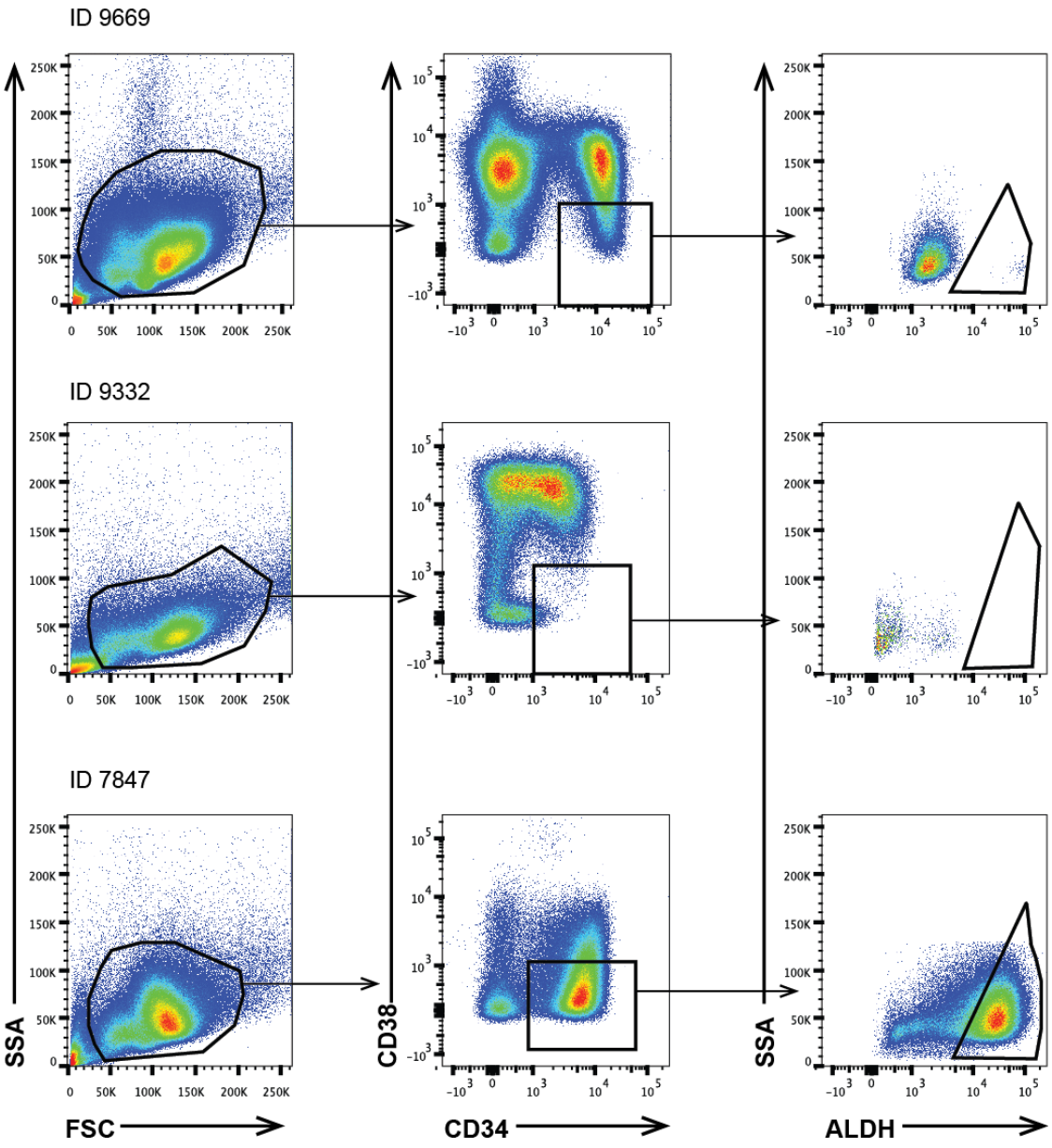

Figure 4-2 Flow plots illustrating the three patterns of ALDH expression within AML samples. Sample 9669 is an inv(16) AML sample where there is a clear, separate ALDH^{high} population of putative stem cells. Sample 9332 is an AML sample of normal karyotype where there are no apparent cells in the ALDH^{high} gate. Sample 7847 is a poor risk karyotype AML with hyperdiploidy, where the leukaemia cells are ALDH^{high}

It is interesting to observe the patterns of ALDH expression within the CD34^{+}38^{-} fraction of CD34^{low} AMLs. As we might expect, in these samples, the majority of the CD34^{+}38^{-} cells stain ALDH^{high}. This was the case in 12/14 samples studied. In 1/14 samples, all of the CD34^{+}38^{-} cells were ALDH^{high}. In 1/14 samples, there were no cells seen in the ALDH^{high} gate of the CD34^{+}38 subfraction.

4.4.1.2.2 CLL1

44 samples were tested for CLL1 expression. 38/45 (84%) of the samples showed differential CLL1 expression within the CD34^{+}38^{-} populations, which generally separated into two discrete

populations separated by one or two logs of staining intensity. Example plots are shown in Figure 4-3.

Subanalysis of the $CD34^+$ AML samples showed that in 26/31 (84%) of the samples tested, the $CD34^+38^-$ cells separated into two populations. In the $CD34^{low}$ AML samples, 12/14 (86%) showed a range of expression.

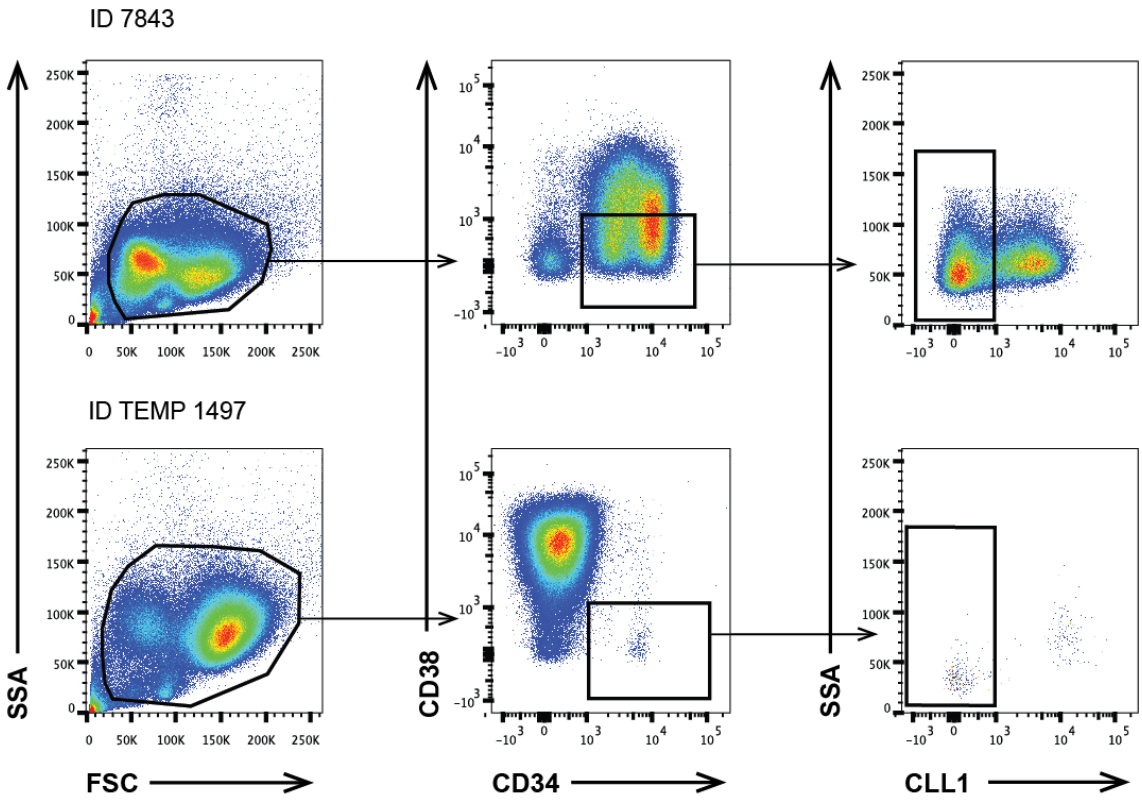

Figure 4-3 Flow plots illustrating typical patterns of CLL1 expression within a $CD34^+$ AML (ID 7843) and a $CD34^{low}$ AML sample (Temp 1497)

4.4.1.2.3 TIM3

46 samples were tested for TIM3 expression. Of these, 17/47 (36%) of AMLs demonstrated a range of TIM3 expression within their $CD34^+38^-$ subpopulations, which might be utilised to separate the cells into separate groups dependent on isotype gating. However, in only two cases did TIM3 staining mark separate populations of cells separated by greater than one log of intensity of staining. Both these staining patterns are illustrated in Figure 4-4.

Subanalysis of the $CD34^+$ AML samples showed that in 7/32 (22%) of the AML samples tested, the $CD34^+38^-$ cells showed a spread of TIM3 expression. In the $CD34^{low}$ AML samples, 10/15 (67%) showed a range of expression.

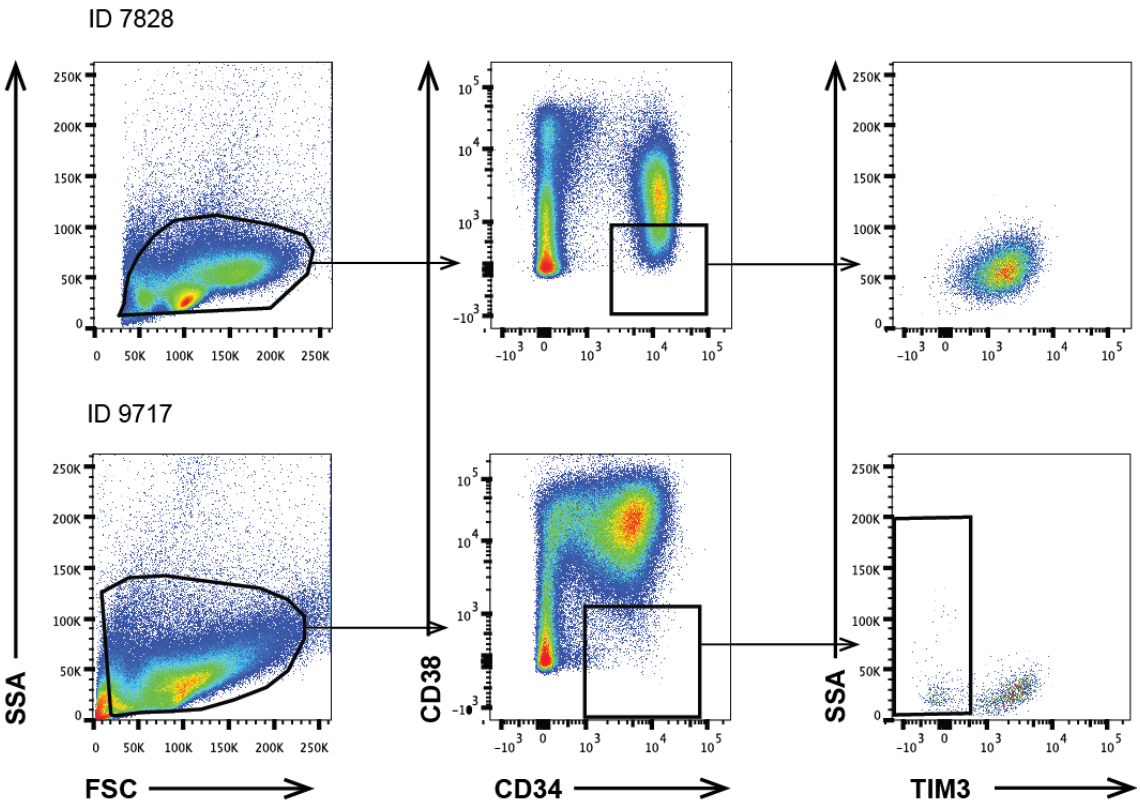

Figure 4-4 Flow plots showing characteristic patterns of TIM3 expression. Sample 7828 represents the most common pattern of TIM3 expression, with no separation within the $CD34^{+}38^{-}$ subfraction of cells. Patient 9717 is a rare example where TIM3 expression appears to distinguish separate populations within the $CD34^{+}38^{-}$ cells of a $CD34^{+}$ AML.

4.4.1.2.4 CD33

43 AML samples were tested for CD33 expression. Of these, 20/43 (47%) of the AML samples of the showed a range of CD33 expression within the $CD34^{+}38^{-}$ cells, which might be utilised to separate out the population into separate groups. Differences between positive and negative staining cells were not as clearly defined as with CLL1 or ALDH expression (see Figure 4-5).

Subanalysis of the $CD34^{+}$ AML samples showed 10/30 (33%) to have differential expression of

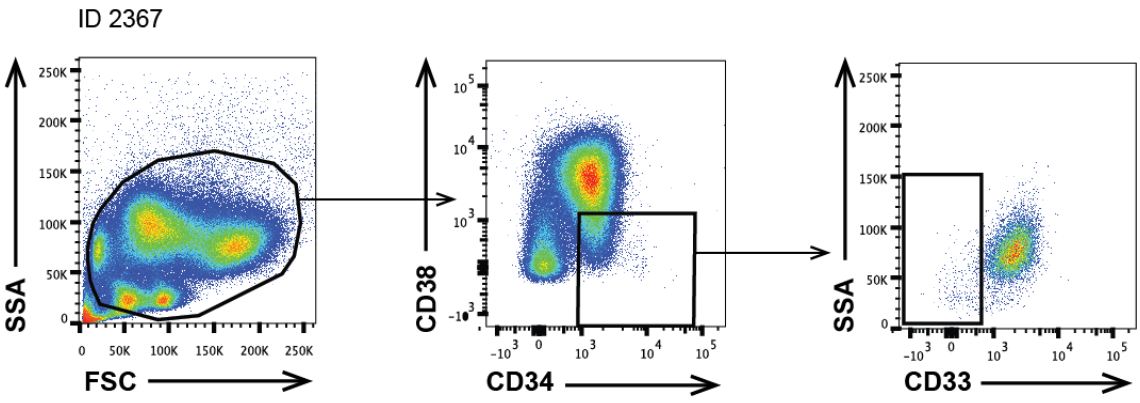

Figure 4-5 Flow plots showing CD33 expression within an AML sample. Here, CD33 defines a possible separation (gate defined by FMO) within the $CD34^{+}38^{-}$ population

CD33 within the $CD34^{+}38^{-}$ subpopulation. Within the $CD34^{low}$ AML samples tested, 10/13 samples (77%) showed differential expression.

4.4.1.2.5 CD99

19 AML samples were tested for CD99 expression. Of these, 9/19 (47%) showed distinct populations of CD99 positive and negative cells within the $CD34^{+}38^{-}$ fraction.

13 samples of $CD34^{+}$ AML were tested, and of these, 4/13 (31%) showed differential CD99 expression within the $CD34^{+}38^{-}$ cells. Within the $CD34^{low}$ AML samples, 5/6 (83%) showed the same pattern. Typical flow appearances are shown in Figure 4-6.

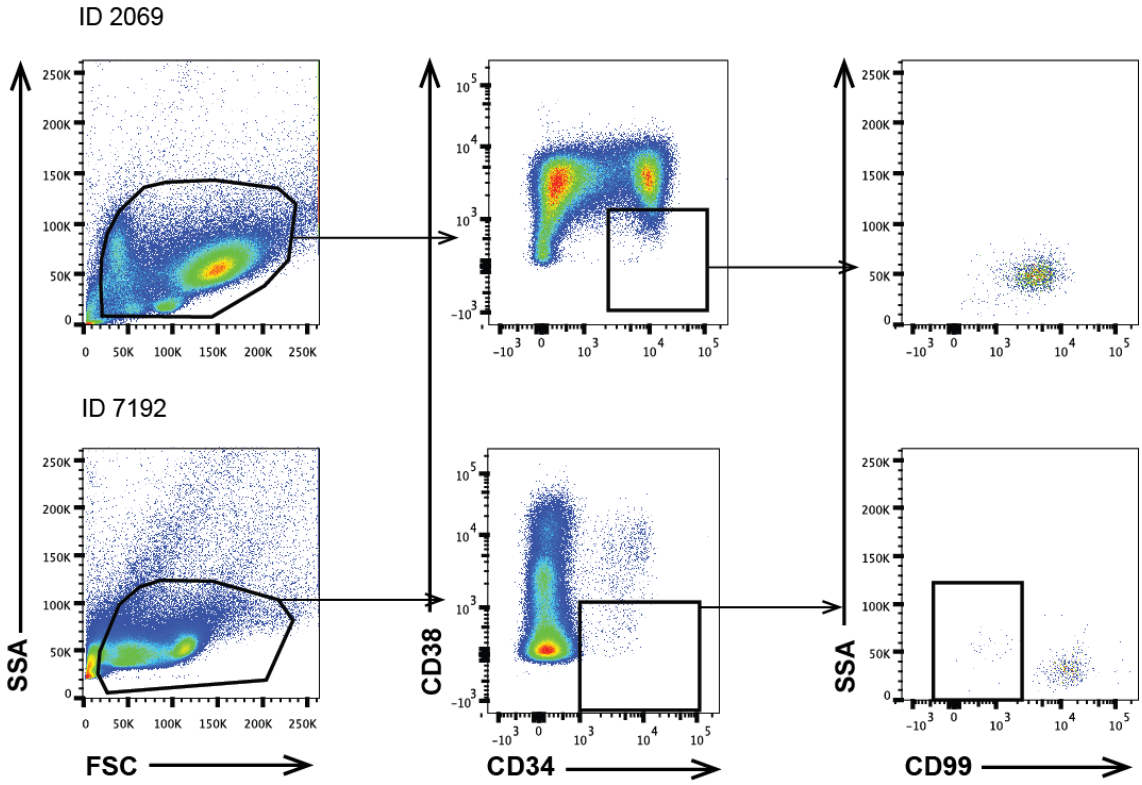

Figure 4-6 Flow plots showing typical CD99 expression patterns within two AML samples. ID 2069 is a $CD34^{+}$ AML where CD99 does not clearly demarcate a separate population within the $CD34^{+}38^{-}$ cells. ID 7192 is a $CD34^{low}$ sample, where a more stringent separation is observed

4.4.1.2.6 CD300

CD300 expression was tested in 15 samples (8 of $CD34^{+}$ AML, and 7 of $CD34^{low}$ AML). CD300 did not appear to show a range of expression within the $CD34^{+}38^{-}$ fraction of any of these samples (see Figure 4-7).

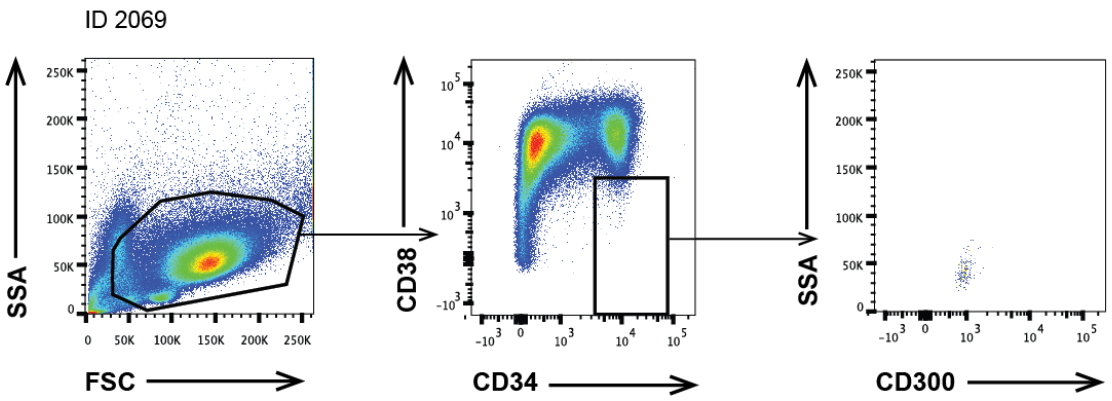

Figure 4-7 Flow plots showing typical CD300 expression, with no further resolution of the $CD34^{+}38^{-}$ population within a $CD34^{+}$ AML sample

4.4.1.2.7 CD47

CD47 expression patterns were tested in 12 samples (ten with $CD34^{+}$ AML and two with $CD34^{low}$ AML). As with CD300, CD47 failed to show a range of expression in the $CD34^{+}38^{-}$ subfractions in any of these samples (see Figure 4-8).

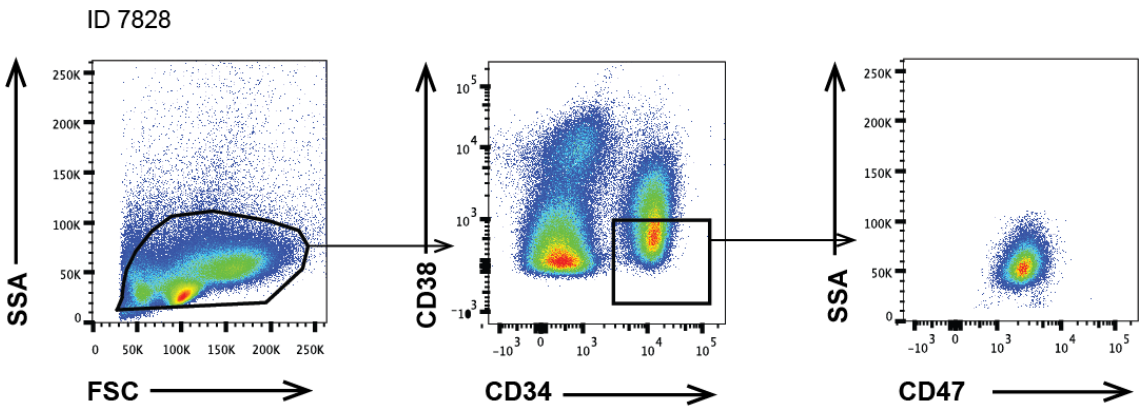

Figure 4-8 Flow plots showing typical CD47 expression pattern, with no further resolution of the $CD34^{+}38^{-}$ population within a $CD34^{+}$ AML

4.4.1.2.8 What makes a useful marker?

The potential usefulness of an individual antigen in helping to separate out normal stem cells from the $CD34^{+}38^{-}$ fraction of AML samples was assessed by two criteria. One is the number and diversity of samples in which it appears to separate out cells of different staining intensity. CLL1 and ALDH appeared to be the most effective with regarding to this effect, marking out putative populations in 84% and 54% of $CD34^{+}$ AML samples respectively. CD33 and CD99 also appeared to have some utility, marking out populations in 33% and 31% of samples respectively.

In addition, the best antigenic markers stain populations of cells with clearly different intensities of staining (ideally at least one log difference between positive and negative populations). This allows for unequivocal gating of populations, which is especially important for this panel, as it was designed to help us to identify and enumerate normal HSCs in diagnostic samples (see Chapter 5). By this criteria, ALDH and CLL1 are again the most valuable markers, consistently marking out populations of cells within the $CD34^{+}38^{-}$ subfraction of $CD34^{+}$ AMLs, with clear differential intensity in staining between positive and negative fractions. CD33 and CD99 also identified differential expression of staining in approximately a third of all $CD34^{+}$ samples, but in both cases, this staining represented more of continuum, rather than distinct populations. The same was also true for TIM3.

Staining with CD47 and CD300 did not help to further define the cells within the $CD34^{+}38^{-}$ fraction of AML samples, regardless of subtype.

4.4.1.3 Combining ALDH and CLL1 staining leads to a more clearly defined population of putative stem cells in $CD34^{high}$ samples

Could simultaneous staining with more than one antigen help to resolve a putative stem cell population within the $CD34^{+}38^{-}$ fraction of a greater proportion of $CD34^{+}$ AML samples? As ALDH expression and CLL1 appeared to be the most promising antigenic markers based on their single staining characteristics, analysis of co-stained samples was then undertaken. When CLL1 is used in combination with ALDH staining, a clear $CD34^{+}CD38^{-}ALDH^{high}CLL1^{-}$ population was identifiable in 100% of the 19 samples where the leukaemia was $ALDH^{low}$, but also in 5 of the 11 samples where the AML showed high ALDH expression. Thus the addition of CLL1 to ALDH enables us to identify a putative stem cell fraction in 24/35 (69%) of all $CD34^{+}$ AML sample (see Figure 4-9).

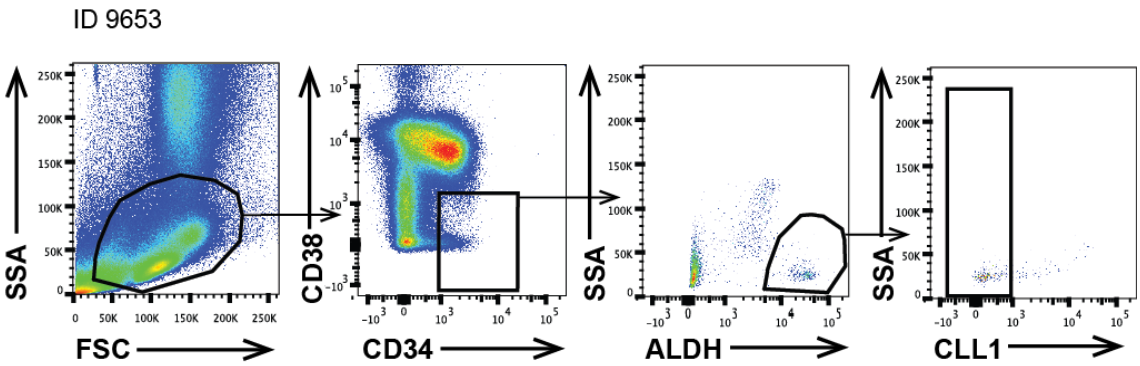

Figure 4-9 Flow plots illustrating how combining staining with ALDH and CLL1 can further define a putative stem cell population in a $CD34^{+}$ AML

These flow plots help us to select markers which may identify a discrete stem cell population, but does not prove the cells we predict are normal genuinely are. However, it is encouraging to note that when we study the expression patterns seen within the $CD34^{+}38^{-}$ fraction of $CD34^{low}$ AML samples, which we known to be enriched in stem cell activity and free of leukaemia-associated genetic mutations, that the vast majority of these cells are $ALDH^{high}$ and CLL1 negative.

4.4.1.4 Final Panel Selection

As a result of the data shown above, we constructed a panel to identify normal stem cells from AML samples comprising fluorochromes staining for expression of CD34, CD38, ALDH and CLL1. In single antigen studies, CD33 also appeared to show differential expression across the $CD34^{+}38^{-}$ subfraction (in 33% of $CD34^{+}$ AMLs). However, as we demonstrate in Chapter 6, the intensity of CD33 expression appears to fluctuate in the $CD34^{+}38^{-}$ cell subfraction in AML samples between diagnosis and remission. As we were constructing a panel to stain and then count stem cells (comparing between AML samples and controls), gating would have to be consistent across all samples. Thus the inclusion of a marker that may vary in intensity between the two groups may make unbiased interpretation of results difficult.

CD99 was another candidate marker for inclusion in the panel. As one might expect, the $CD34^{+}38^{-}$ stem cells within $CD34^{low}$ AMLs were CD99 negative in 5/6 samples tested. The $CD34^{+}38^{-}$ cells within $CD34^{+}$ AML samples showed differential CD99 expression in 4/13 (30%) of cases tested. However, at the time of panel construction, CD99 was only commercially available conjugated to the PE fluorochrome. This was also the case for CLL1, which appeared to be the superior marker for distinguishing putative stem cells in a wider range of samples, and hence CD99 was excluded.

4.4.2 Validation of the Stem Cell Panel: Early optimisation

4.4.2.1 Early attempts to assess the purity of putative stem cell fractions by FISH show encouraging results

The purity of the putative stem cell fractions identified by the above technique was assayed, having selected two different AMLs with known karyotypic abnormalities to which FISH probes were commercially available.

Patient 7013 had AML with associated inv(16). Bone marrow staining and subsequent sorting generated 7292 $CD34^{+}38^{-}ALDH^{high}$ cells, which were subsequently fixed onto slides. Using the Cytocell CBFB-MYH11 Probe (Cat: LPH 022) for detection of inv(16), 0/50 cells examined

contained the fusion product; a positive control slide made from unsorted bone marrow cells from the same patient showed the fusion signal in 10/10 cells. In this patient, as predicted by Gerber et al[207], the use of ALDH expression alone is sufficient to separate out normal cells from the leukaemic clone within the $CD34^{+}38^{-}$ subfraction.

Patient 8361 was known to have developed AML with a t(6;9) translocation. Staining and subsequent sorting generated 2109 $CD34^{+}CD38^{-}ALDH^{high}CLL1^{-}$ cells. FISH staining using the Kreatech DEK/NUP 214 Probe for the detection of t(6;9), showed that 10/10 of the sorted cells gave a normal signal pattern. Analysis of the blast population ($CD34^{+}CD38^{-}ALDH^{int}$) cells revealed 10/10 cells who all gave the same positive signal indicating the presence of t(6;9). This would suggest in patient 8361, this flow based method of sorting led to successful separation of normal cells from the leukaemic population.

Technically, using FISH to analyse the purity of sorted populations of stem cells in this manner is a significant challenge, even for the most skilled practitioners. Even using the technique described in Section 2.7.1 to generate fixed slides from small numbers of sorted cells, FISH analysis becomes difficult when cell numbers fall below 2000 per slide. Numbers of 100,000 per slide are preferred for optimal confidence in reporting. Thus samples which yield very small numbers (less than 1000) of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells cannot be confidently analysed using this method.

4.4.2.2 Testing the stem cell nature of sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells by prolonged in vitro culture

The next step in validating the sorting approach described above was to test the stem cell nature of the sorted cell fraction, as determined by the cells' ability to survive in vitro culture conditions and generate colonies of normal karyotype. This approach has a secondary benefit: assuming cells grow in culture, the resultant amplified colonies can be tested for the presence of the leukaemia-specific mutation at the end of the culture period.

This was first attempted by short term culture of sorted cells in methylcellulose for two weeks, and subsequently over a longer 7 week culture period, as detailed below.

4.4.2.2.1 Methylcellulose Assay

4.4.2.2.1.1 Optimisation of the Methylcellulose 2 week Cell Culture Assay

Sensitivity of the methylcellulose assay in my hands was tested by sorting $CD34^{+}CD38^{-}ALDH^{high}CLL1^{-}$ cells from GMPB, and growing these in culture for 2 weeks. After 2 weeks, having added 200 cells per plate in triplicate, a mean of 43 myeloid and 21 erythroid colonies were

noted. After replating of these cells, and culture for a further 2 weeks, means of 11 myeloid and 0.2 erythroid colonies were calculated.

4.4.2.2.1.2 Methylcellulose Culture on HSCs sorted from AML samples- a 2 week culture period may not be sufficient to demonstrate colony production from quiescent stem cells

Three different AMLs were then stained and sorted as described in Section 4.3.2.2. The number of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells collected from each sample are shown below in Table 4-2.

Table 4-2 Sample details for AMLs sorted for $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells and grown in 2 week methylcellulose culture

ID	KARYOTYPE	NPM STATUS	$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ CELLS SORTED	CELLS PER PLATE
8361	t(6;9)	WT	2556	100 and 200
9068	DEL 7	WT	168	42
8458	NORMAL	MUT	559	126

After 2 weeks of culture, no growth was visible on any of the plates from all 3 different AML samples. The plates were left unaltered at 37°C, and examined each week for the presence of colonies. At 5 weeks, colonies were visible on the plates containing cells from patient 8361. At an initial concentration of 100 $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells per plate; a mean of 1 myeloid and 1 erythroid colony was seen per plate. At an initial concentration of 200 cells per plate, a mean of 4 myeloid and 1 erythroid colonies were counted. Subsequent FISH analysis for the t(6;9) translocation showed a normal signal, with no evidence of the translocation within these populations. No growth was observed on the plates from patients 8458 or 9068.

This delayed growth of colonies is fascinating, as it is in marked contrast to the behaviour of the same $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ population from GMPB, which generated multiple colonies within two weeks. This observation gives further credence to the idea that the stem cells from within AML samples are in a state of induced quiescence.

4.4.2.2.2 7 week Long Term Culture on HSCs sorted from AML samples- assay optimisation

The results from the above experiment led us to hypothesise that a two to four week culture period may be inadequate to test the true stem cell nature of HSCs that have been induced into a state of quiescence by previous exposure to AML. The methylcellulose assay is not really designed to support the growth of any stem cells for a period greater than two weeks, as they generally require the cytokines released by supporting stromal cells to support a more prolonged culture period. Therefore a 7-week Long Term Culture Assay was set up to assess

the growth and purity of cells sorted from three different AML samples, one control bone marrow and one GMPB sample (Table 4-3). As well as plating the putative sorted stem cell fraction of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells, for the three AML samples, 1000 cells from each of the blast populations were also collected and plated to assess their behaviour in long term culture.

Table 4-3 Sample details for the samples for $CD34^{+}38^{-}ALDH^{high}CLL1^{low-}$ cells prior to 7 week LTC-IC assay

SAMPLE INFO				$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ CELLS SORTED	CELLS PER WELL
PATIENT ID	SAMPLE TYPE	KARYOTYPE	GENETICS		
7820	AML	t(9;22)		1055	50, 100, 200
9584	AML	MONOSOMY 5		528	10, 50, 100
7829	AML	NORMAL	*NPM1* MUT	2013	50, 100, 200, 300
7865	CONTROL			162	50
GMPB	CONTROL			393	33, 100

After 7 weeks, no colonies were seen in any wells, including those containing the positive control cells. The absence of growth from any section of the culture, including the positive controls of GMPB or control bone marrow, pointed to a technical issue with the assay as the cause of this result.

Therefore, a repeat LTC assay was set up using HSCs derived from normal bone marrow to test a number of variables which might have led to this assay failure. Any potential deleterious effect on cell viability by staining for ALDH expression was tested for by the simultaneous staining and sorting of the samples for CD34 and CD38 only (as detailed in Section 4.3.2.1), as well as concurrently for CD34, CD38, ALDH and CLL1 (as detailed in Section 4.3.2.2).

My own inexperience in LTC-IC assays was tested as another variable. Two of 96 well plates were set up by an experienced operator (F Miraki-Moud), who also provided all the required weekly media changes. A further two plates were set up by myself in duplicate.

Finally, the effect of different irradiation sources for MS5 preparation was also investigated. Two 96 well-plates were plated with MS5 cells irradiated at our newly installed irradiator at the BCI. A further two 96 well plates were plated using cells treated at the London Research Institute irradiator. The same dose of 70Gy was used for both conditions.

Once again, at the end of 7 weeks' culture period in a dedicated incubator, there was no growth observed in any of the wells, in any of the test conditions.

Given this result, a third Long-Term Culture-Initiating Cell Assay was set up having ordered replacement stocks of fresh Collagen and H5100 Myelocult™ media. This time, GMPB was used as an enriched source of stem cells. $CD34^+$ cells were sorted as per Section 4.3.2.1 (with no CD38 antibody added to reduce any potential inhibitory effect of residual CD38 antibody on stem cell viability)[224]. Cells were distributed in triplicate between wells in a range of concentrations. At the end of the culture period, a few, scattered colonies were visible (see Table 4-4).

Table 4-4 LTC-IC attempt 3 results

NUMBER OF $CD34^+$ CELLS PER WELL	PRESENCE OF COLONIES IN WELLS		
	WELL 1	WELL 2	WELL 3
18233	+	0	0
9116	+	0	0
4558	0	+	+
1367	0	0	0
911	0	0	0
456	0	+	0
230	0	0	0
138	0	0	0
46	0	0	0

The growth of even a few scattered haematopoietic colonies was promising. However, these results show a very poor degree of concordance between technical replicates, and also a much lower sensitivity than was expected. As described in a recent publication, we had previously been able to generate to generate colony growth reliably from plating a mere five cells per well[1].

Our final hypothesis to explain the failure of the LTC assay was that the MS5 cells (used as a source of stromal-induced cytokines to support the growth of stem cells over the 7 week period) had become ineffective with age and repeated passaging (D Taussig: personal communication). Mycoplasma testing of our stocks of the cell line had been repeatedly negative. Fresh replacement stocks were ordered from the same supplier (DSMZ, Cat: ACC441), and the assay repeated. Lineage-depleted GMPB was sorted for $CD34^+$ cells, which were plated in a range of concentrations as demonstrated in Table 4-5 .

The universally positive growth as expected at the end of a 7 week culture period, from a source of enriched stem cells, suggested strongly that the previous failure of our LTC assay had been due to an issue with the MS5 stromal feeder line.

Table 4-5 LTC-IC attempt 4 using new MS5 stocks

NUMBER OF $CD34^+$ CELLS PER WELL	PRESENCE OF COLONIES IN WELLS		
	WELL 1	WELL 2	WELL 3
12705	+	+	+
6352	+	+	+
3176	+	+	+
1588	+	+	+
635	+	+	+
423	+	+	+
211	+	+	+

Management of the feeder line throughout these experiments had been as described in Section 2.5.1. Once the cells reached 70 to 80% confluence in the culture flasks, cells were split. However, once thawed, cells on occasion been continually grown for periods of greater than two months, and therefore had undergone serial passaging. Other groups limit the number of serial passages to five or six to avoid an "exhaustion phenotype" (personal communication: F Anjos-Afonso). From this point on, we adopted a similar policy. All subsequent LTC assays were performed using our new replacement stocks. Given the technical difficulties this had led to, all subsequent LTC-IC assays were performed with the presence of a positive control of $CD34^+$ sorted cells from either Lin depleted GMPB or cord cells.

4.4.3 Validation of the Stem Cell Panel: testing on AML bone marrow samples

Having optimised the 7 week LTC-IC assay in our laboratory, we were then finally able to proceed to test a range of diagnostic bone marrow AML samples by sorting their $CD34^+38^-$ $ALDH^{high}CLL1^-$ cells, attempting to amplify them in LTC and then testing the resultant populations of cells for the presence of leukaemia-specific mutations. This approach enabled us both to prove the stem cell nature of the fractions, as well as the testing the validity of the proposition that these antigens separate out "normal" stem cells from the AML bulk population.

4.4.3.1 Identification of appropriate AML samples for sorting

A cohort of AML patients with diagnosis BM samples stored in the tissue bank was identified. For all samples, hospital records were analysed for the recorded cytogenetic karyotype at diagnosis, as well as any molecular genetic analysis performed. Samples were selected if one could potentially differentiate the normal population from the AML cells based on a detectable leukaemia-specific mutation.

4.4.3.1.1 Cytogenetic abnormalities

Samples associated with common cytogenetic abnormalities, detectable by FISH probes routinely used by the RLH Cytogenetic laboratory, were preferentially chosen. In a few instances, probes (such as the Kreatech DEK/NUP 214 t(6;9) Probe) were specifically purchased and optimised for use in this project by Marianne Grantham.

For patients with multiple cytogenetic abnormalities, FISH probes were chosen which would recognise the dominant clone, as this seemed the most sensitive method of screening for the presence of blasts.

4.4.3.1.2 Molecular Genetic Mutations

As discussed at length in Section 1.4.5.4.1, assessment of the constitution of the $CD34^{+}38^{-}$ $ALDH^{high}CLL1^{-}$ population by PCR based analysis for a leukaemia-associated mutation was assumed to be the optimal technique available. This is in part because the sensitivity of this test to low input cell numbers meant pre-test amplification of the sorted $CD34^{+}38^{-}$ $ALDH^{high}CLL1^{-}$ cell populations was not required. In addition, in the case of the *NPM1* mutation, the fact that this mutation occurs early in leukaemogenesis, tends to be stable throughout disease evolution[111], and therefore present in the vast majority of leukaemia cells, makes it an excellent candidate for assessing sort purity (in comparison to a secondary mutation which might only be detected within one clone)[106]. Finally, an assay for assessing the purity of an *NPM1* sorted sample had already been developed (see Section 2.8.4.5) and was therefore possible to perform in-house. As a result of these factors, strenuous efforts were made to detect all samples carrying the *NPM1* mutation.

Hospital records were searched for previous clinical assessment of *NPM1* and *FLT3* status for a large cohort of AML samples known to be stored in the tissue bank.

The methodology adopted by the RLH Molecular Genetics department for clinical screening of samples for *NPM1* mutation (as detailed in Section 4.3.4.2.2)[223] detects the presence of an 4 base pair insertion within Exon 12 of the *NPM1* gene, but does not involve sequencing the actual mutation (knowledge which is required for the primer selection during the qPCR assay described in Section 2.8.4.5).

Therefore, we undertook independent screening of 37 samples for *NPM1* mutation within our laboratory by the Methodology detailed in Section 2.8.3, both sequencing those samples which had been detected as having an *NPM1* mutation by the hospital lab (n=24), as well as assaying samples which no result was available (n=13).

Typical traces showing the normal sequence within NPM1 Exon 12, Type A and Type D mutations are shown in Figure 4-10.

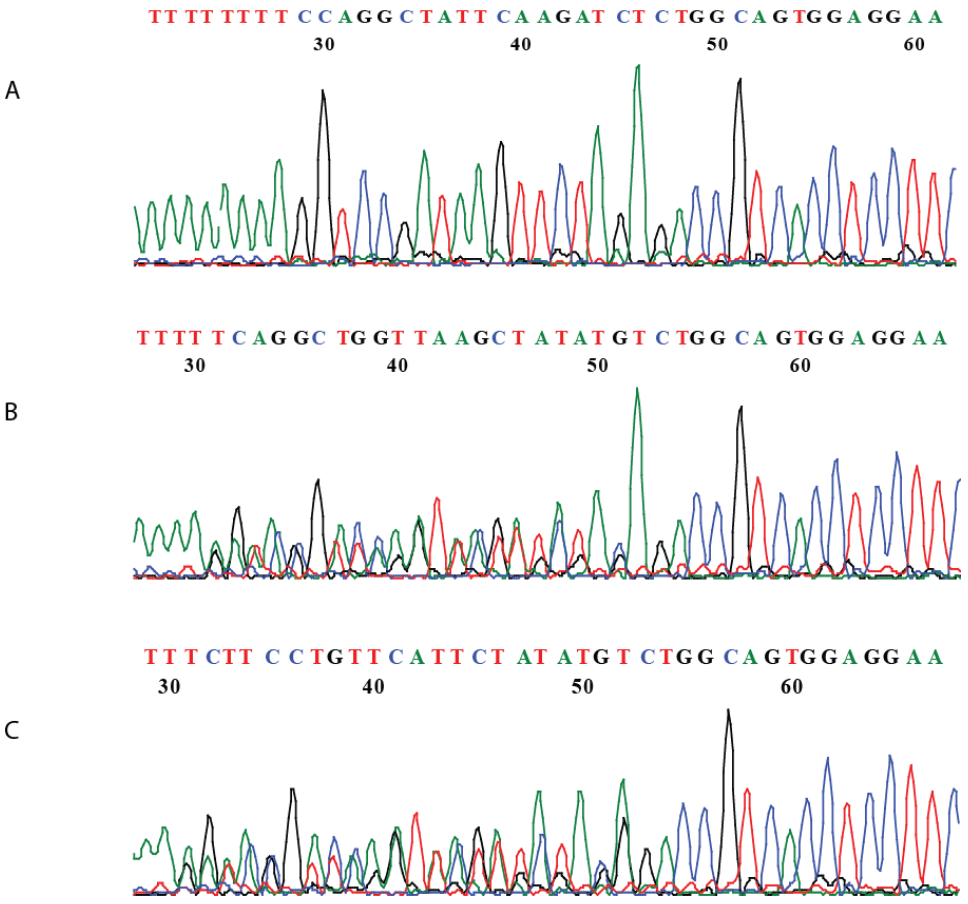

Figure 4-10 Typical Sanger Sequencing traces showing A: normal Exon 12; B: Type A mutation- an insertion of a TCTG tetranucleotide; C: Type D mutation- an insertion of a CCTG tetranucleotide

Of 24 samples assayed by both methodologies, concordant results were found in 22 of 24 patients. In two (Patient ID 9713 and 10203), results were found to be positive by the RLH assay, but subsequent local sequencing failed to reveal a mutation. The reason for this discrepancy is not obvious, given the general view that *NPM1* mutations are early events that are found within the vast majority of leukaemic cells, and therefore both tests should be sufficiently sensitive to pick up a mutation if present.

Results of screening, and comparison with hospital records for the 35 samples used for validation are shown in Table 4-6.

4.4.3.2 Selected AML sample details

In total, 35 AML samples were sorted in an attempt to validate if the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ compartment consisted solely of normal HSCs. The number tested was restricted by sample availability and processing costs. All of these AML samples were associated with mutations

(either karyotypic abnormalities, or mutations of *NPM1* and/or *FLT3*), which meant we could attempt to assess the "normality" of the sorted stem cells. Of these samples, 4 were of good risk cytogenetics, 21 intermediate and 10 poor. Biological parameters pertaining to these samples are shown in Table 4-6.

Subsequent analysis is organised with reference to the karyotypic risk group of the underlying AML samples. A summary of the results of sorting, growth in LTC and purity assessment of the resultant populations is shown in Table 4-7. Examples of the variety of phenotypic appearances shown by the AML samples sorted are illustrated in the FACS plots in Figure 4-11 and Figure 4-12.

4.4.3.3 Good risk AML samples

4.4.3.3.1 $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells are normal and free of the leukaemia-associated mutation in all cases studied

In all 4 good risk patients, colonies were observed after 7 weeks LTC of sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cell fraction. This supports the hypothesis that these cells retain a stem cell phenotype.

FISH based analysis of the resultant colonies found these to be clear of the leukaemia associated mutation in all cases. These results are shown in Figure 4-13 and Table 4-8.

4.4.3.4 Intermediate risk AML samples

4.4.3.4.1 Sample processing

21 samples with intermediate risk AML were analysed. This is a complex collection of results to explain for two reasons. Firstly, some of the samples with known *NPM1* mutations were analysed immediately after undergoing sorting. This was to increase the number of samples it was possible to process (due to the costs and the labour-intensive nature of the 7 week LTC), and was possible due to the sensitivity of the *NPM1* PCR assay. Secondarily, the intermediate cytogenetic risk group encompasses a range of cytogenetic and molecular mutations, and thus samples were analysed by either PCR or FISH-based techniques.

Table 4-6 AML samples sorted for $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations- sample information

PATIENT DETAILS			DIAGNOSIS BLOOD COUNT			AML Genetic Information					AML PHENOTYPE
						KARYOTYPE	HOSPITAL GENERATED			LAB GENERATED	CD34 STATUS (% of LIVE CELLS $CD34^{+}$)
ID	AGE	SEX	HB (g/dL)	WCC ($x10^9$/L)	PLATELETS ($x10^9$/L)		*NPM1*	*FLT3* ITD	*FLT3* TKD	*NPM1*	
7774	62	M	8.1	0	37	NORMAL				Y-TCTG	1.6
7777	39	M	8.9	15	83	NORMAL				Y-TCTG	1.4
7793	56	F	11.1	4	10	ADD 13				Y-TCTG	56.2
7820	66	M	12.6	9	52	t(9;22) BIPHENOTYPIC					85.4
7829	22	M	11.1	5	82	+13; t(3;14)				Y-TCTG	82.9
7912	72	F	7.3	3	19	NORMAL				Y-TCTG	0.2
7968	51	F	12.0	13	180	+13				Y-TCTG	2.6
8199	45	F	6.6	22	41	NORMAL				Y-TCTG	0.1
8361	38	F	9.5	41	53	t(6;9)	WT	MUT	WT		42.5
8429	48	M	6.1	3	39	NORMAL	MUT	WT	MUT	Y-TCTG	0.0
8458	49	F	11.2	?	35	NORMAL	MUT	WT	MUT	Y-TCTG	0.0
8613	28	M	10.7	52	53	+12	WT	WT	WT		
8618	51	M	10.5	19		NORMAL				Y-TCTG	0.1
8791	59	M	7.4	31	43	COMPLEX					0.5
8932	43	M	7.2	3	54	t(8;21)	WT	WT	WT		67.1
8993	83	M	6.5	45	25	TRISOMY 11					8.5
9066	59	F	6.8	12	84	DEL 5	WT	WT	WT	WT	63.2
9067	67	M	7.4	142.6	172	NORMAL	MUT	WT	WT	Y-CTTG	9.3
9154	63	M	10.3	4	129	COMPLEX	WT	WT	WT	WT	2.6

PATIENT DETAILS			DIAGNOSIS BLOOD COUNT			AML Genetic Information					AML PHENOTYPE
						KARYOTYPE	HOSPITAL GENERATED			LAB GENERATED	CD34 STATUS (% of LIVE CELLS $CD34^{+}$)
ID	AGE	SEX	HB (g/dL)	WCC ($x10^9$/L)	PLATELETS ($x10^9$/L)		*NPM1*	*FLT3* ITD	*FLT3* TKD	*NPM1*	
9583	44	M	8.3	7.6	9	HYPERDIPLOID	WT	WT	WT	WT	0.9
9584	84	F	5.5	1.9	71	DEL5Q;HYPERDIP	WT	WT	WT	WT	42.0
9669	41	M	11.1	12	70	INV(16)	WT	WT	WT	WT	15.6
9672	61	M	6.9	45	20	DEL 7q				WT	49.8
9713	36	M	5.5	19	100	NORMAL	MUT	WT	MUT	WT	0.4
9717	24	F	7.4	1.5	23	-7 AND +21					12.8
9719	61	M	8.9	6	123	NORMAL	MUT	MUT	WT	Y-TCTG	0.1
9850	59	M	8.3	56	101	NORMAL	MUT	MUT	WT	Y-TCTG	0.8
9878	48	M	11.8	73	47	NORMAL	MUT	MUT	WT	Y-TCTG	0.0
9962	66	F	7.0	230	40	NORMAL				Y-TCTG	0.6
9963	69	F	10.8	5	57	DIF CLONES INC. DEL 7Q AND 11	WT	WT	WT	WT	0.8
9964	24	M	13.9	77	80	NORMAL	MUT	MUT	WT	Y-TCTG	0.2
10139	22	M	8.1	10.5	23	t(15;17)	WT	MUT	WT	WT	82.9
10147	36	F	5.4	2.3	54	COMPLEX WITH 5Q					21.3
10192	57	M	7.0	56	121	11P DEL	MUT	MUT	WT	Y-TCTG	0.3
10212	30	F	10.7	9.8	52	t(15;17)	WT	MUT	WT	WT	7.2

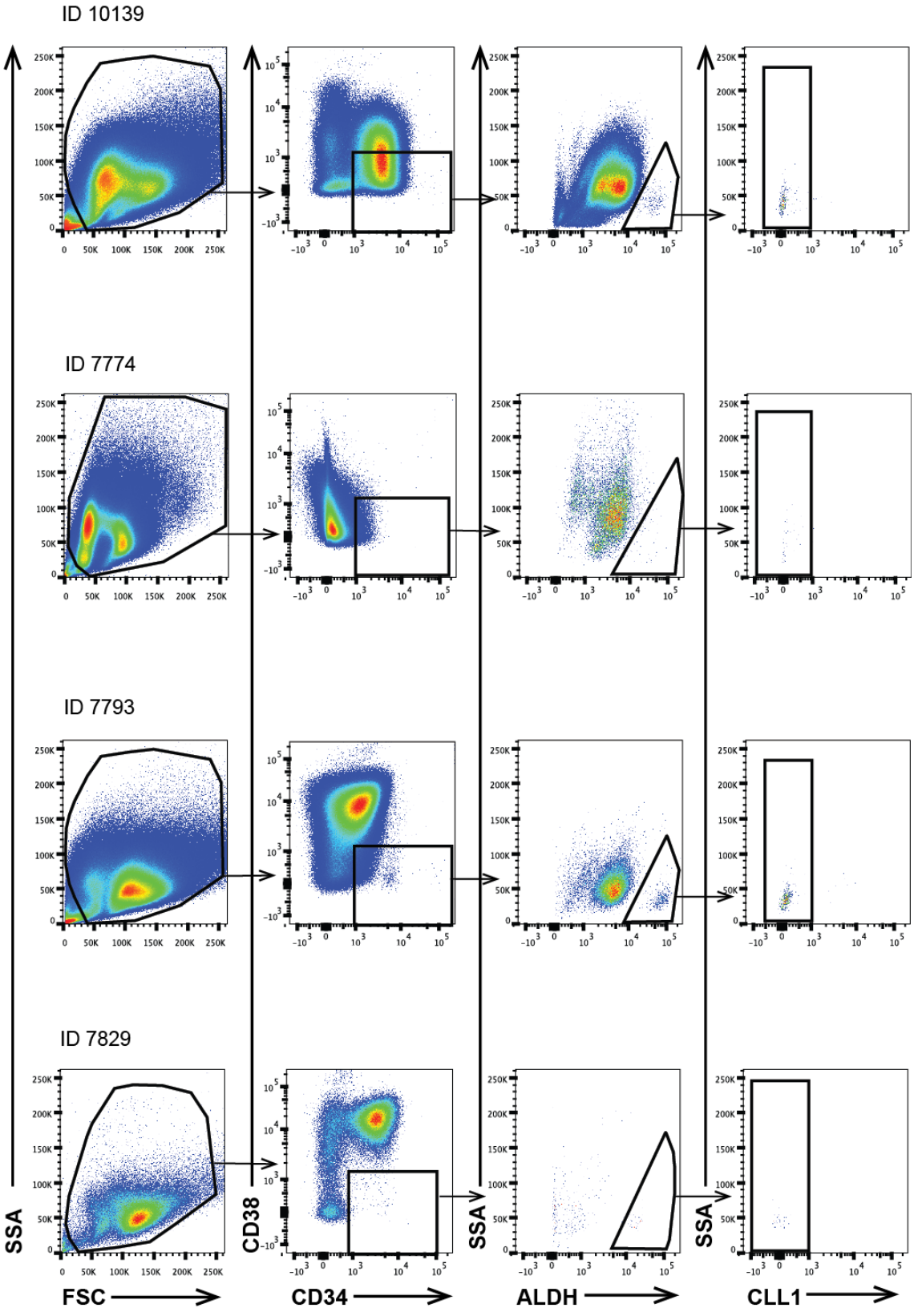

Figure 4-11 Sorting FACS plots for 4 primary AML samples, illustrating key steps of consecutive gating strategy ID 10139 is a good risk sample with t(15;17); Samples 7774 and 7793 both are intermediate risk samples illustrating the different phenotypic appearances of *NPM1* mutated AML; Sample 7829 has t(3;14) and add13 but is *NPM1* WT

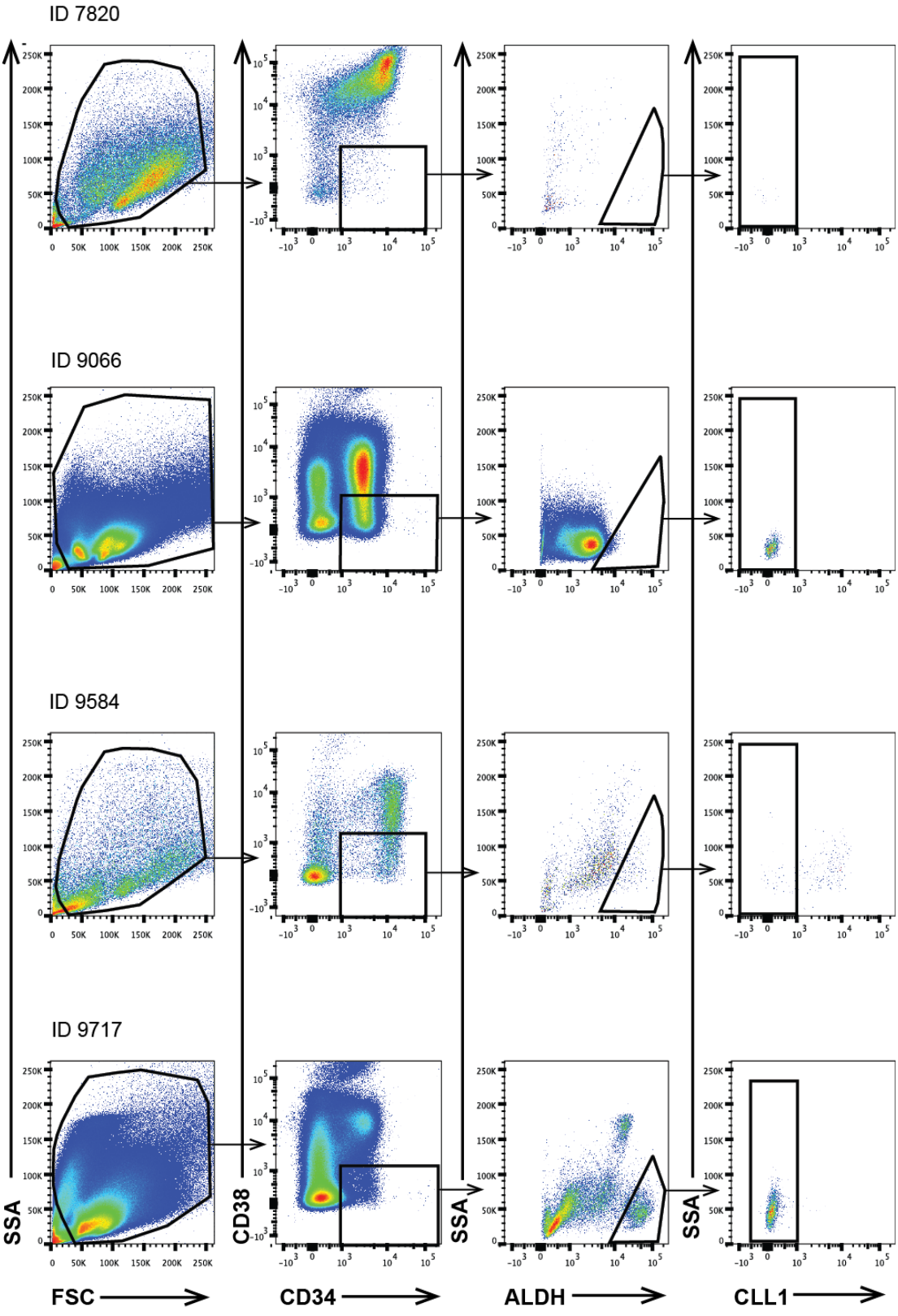

Figure 4-12 FACS sorting plots illustrating the diverse appearances of 4 poor risk samples. ID 7820 is a biphenotypic AML with t(9;22); ID 9066 has del(5); ID 9584 shows hyperdiploidy and del(5); ID 9717 has del(7) and add(21)

Table 4-7 Summary of FACS sorting for $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations, LTC results and assessment of sorted populations for AML-associated mutations

SAMPLE INFO			ASSESSMENT OF STEM CELL FUNCTION OF $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ CELLS						ASSESSMENT OF $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ CELLS FOR AML ASSOCIATED MUTATION				
										TIMING OF ASSESSMENT			
ID	KARYOTYPE	RISK	LTCIC ASSAY ATTEMPTED?	DID LTC-IC POS CONTROL WORK?	TOTAL $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells	NO OF WELLS	POSITIVE WELLS AFTER 7 WEEKS LTC	DID AML CELLS GROW?	ABLE TO ATTEMPT ASSESSMENT?	IMMEDIATE ANALYSIS	LTC WITH SUBSEQUENT COLONY ANALYSIS	METHOD	RESULTS
10212	t(15;17)	GOOD	Y	Y	546	2	2	Y	Y		+	FISH	PURE
10139	t(15;17)	GOOD	Y	Y	371	2	2	Y	Y		+	FISH	PURE
8932	t(8;21)	GOOD	Y	Y	462	3	3	Y	Y		+	FISH	PURE
9669	INV(16)	GOOD	Y	Y	947	3	3	Y	Y		+	FISH	PURE
7793	+13	INT	Y	Y	2417	3	3	N	Y		+	NPM PCR	PURE
10192	DEL 11P	INT	Y	Y	1000	3	0	Y	Y	+	UNABLE	NPM PCR	PURE
9962	NORMAL	INT	Y	Y	1153	2	2	N	Y		+	NPM PCR	PURE
8458	NORMAL	INT	Y	Y	141	2	2	Y	Y		+	NPM PCR	PURE
7968	NORMAL	INT	Y	Y	9000	4	4	Y	Y	+	+	NPM PCR	DIRTY
9850	NORMAL	INT	Y	Y	1000	3	3	Y	Y	+	+	NPM PCR	ENRICHED
9713	NORMAL	INT	Y	Y	1749	2	2	Y	Y		+	*FLT3*-TKD PCR	PURE
8993	+11	INT	Y	Y	1080	3	3	Y	Y		+	FISH	PURE
8613	+12	INT	Y	Y	1625	4	1	Y	Y		+	G BANDING	UNABLE TO ASSESS
7829	+13; t(3;14)	INT	Y	N									
7774	NORMAL	INT	N						Y	+		NPM PCR	PURE
7912	NORMAL	INT	N						Y	+		NPM PCR	ENRICHED
8429	NORMAL	INT	N						Y	+		NPM PCR	PURE

SAMPLE INFO			ASSESSMENT OF STEM CELL FUNCTION OF $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ CELLS						ASSESSMENT OF $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ CELLS FOR AML ASSOCIATED MUTATION				
										TIMING OF ASSESSMENT			
ID	KARYOTYPE	RISK	LTCIC ASSAY ATTEMPTED?	DID LTC-IC POS CONTROL WORK?	TOTAL $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells	NO OF WELLS	POSITIVE WELLS AFTER 7 WEEKS LTC	DID AML CELLS GROW?	ABLE TO ATTEMPT ASSESSMENT?	IMMEDIATE ANALYSIS	LTC WITH SUBSEQUENT COLONY ANALYSIS	METHOD	RESULTS
8618	NORMAL	INT	N						Y	+		NPM PCR	PURE
9719	NORMAL	INT	N						Y	+		NPM PCR	ENRICHED
7777	NORMAL	INT	N						Y	+		NPM PCR	ENRICHED
9964	NORMAL	INT	N						Y	+		NPM PCR	PURE
9878	NORMAL	INT	N						Y	+		NPM PCR	ENRICHED
8199	NORMAL	INT	N						Y	+		RNA SEQ	PURE
9067	NORMAL	INT	N						Y	+		RNA SEQ	PURE
8361	t(6;9)	INT	N						Y	+		FISH	PURE
9717	-7 and +21	POOR	Y	Y	1637	2	2	N	Y		+	FISH	ENRICHED
9963	3 CLONES INC DEL 7q & 11q+	POOR	Y	Y	1003	3	3	Y	Y		+	FISH	PURE
8791	COMPLEX	POOR	Y	Y	1211	3	3	Y	Y		+	FISH	PURE
9066	DEL 5	POOR	Y	Y	589	2	0	Y	N				
9583	HYPERDIPLOID	POOR	Y	Y	1074	3	0	N	N				
9672	DEL 7q	POOR	Y	Y	15	3	0	N	N				
9154	COMPLEX	POOR	Y	Y	16	2	0	Y	N				
9584	DEL 5q; HYPERDIPLOID	POOR	Y	N					N				
7820	t(9;22)	POOR	Y	N					N				
10147	COMPLEX	POOR	N	Y	8	NP			N				

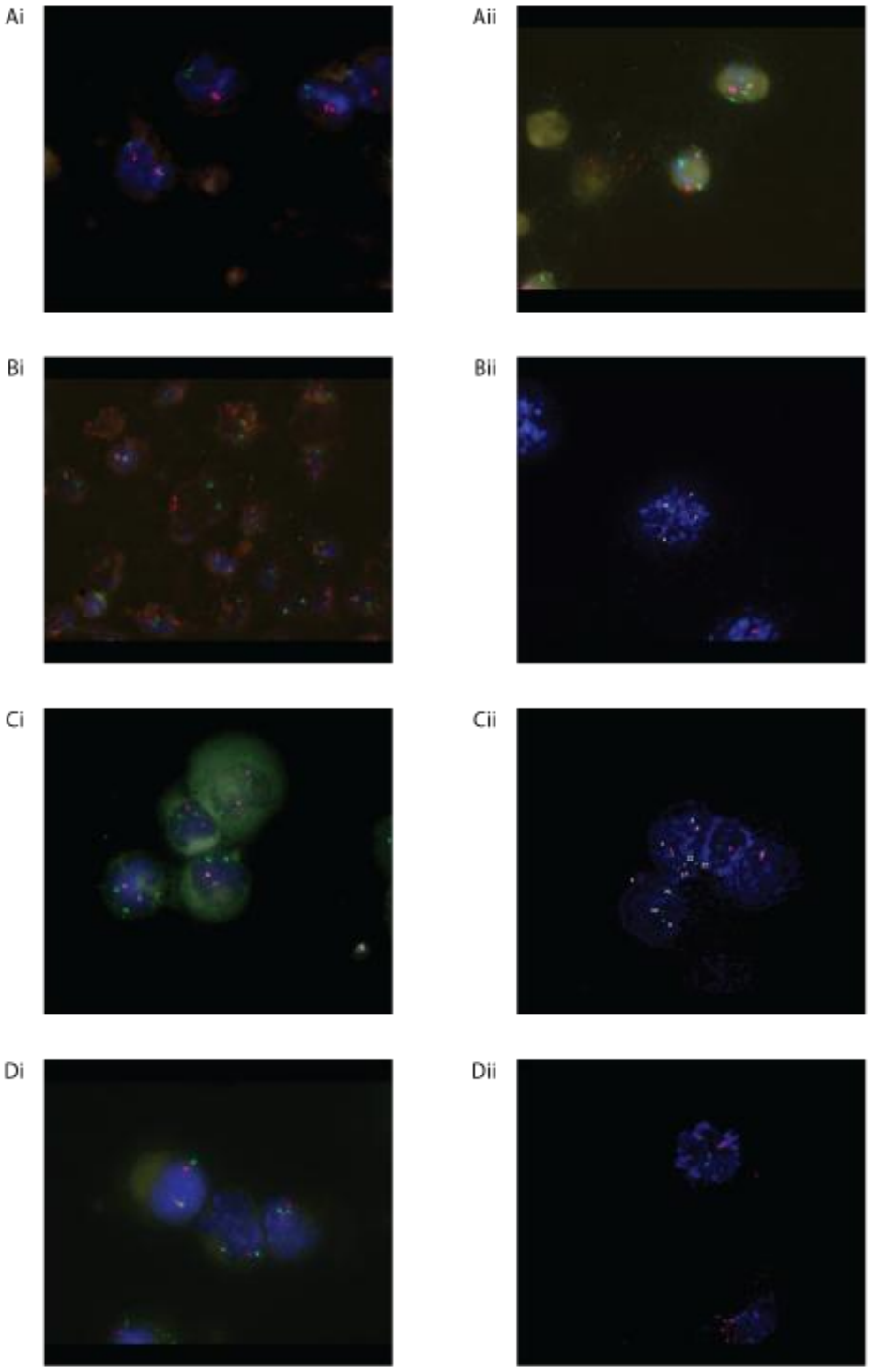

Figure 4-13 FISH staining of 4 sorted samples. A: Pat ID 8932 with t(8;21) i: Colonies from $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells after LTC ii. Unsorted blasts; B: Pat ID 10212 with t(15;17) i: Colonies from $CD34^{+}CD38^{-}ALDH^{high}CLL1^{-}$cells after LTC ii. Unsorted blasts; C: Pat ID 8791 with complex karyotype including trisomy 9q34 i: Colonies from $CD34^{+}CD38^{-}ALDH^{high}CLL1^{-}$cells after LTC ii. Unsorted blasts; D: Pat ID 9963 with complex karyotype including del(7q) i. Colonies from $CD34^{+}CD38^{-}ALDH^{high}CLL1^{-}$ cells after LTC ii. Unsorted blasts

Table 4-8 Analysis of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells and their progeny for AML-associated mutations by FISH

ID	KARYOTYPE	RISK GROUP	PROBE	$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ CELLS MADE INTO SLIDES IMMEDIATELY		UNSORTED CELLS MADE INTO SLIDES IMMEDIATELY		COLONIES FROM $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ CELLS AFTER LTC		COLONIES FROM UNSORTED CELLS AFTER LTC	
				CELLS PER SLIDE	CELLS WITH AML-ASSOCIATED FISH SIGNAL	CELLS PER SLIDE	CELLS WITH AML-ASSOCIATED FISH SIGNAL	CELLS PER SLIDE	CELLS WITH AML-ASSOCIATED FISH SIGNAL	CELLS PER SLIDE	CELLS WITH AML-ASSOCIATED FISH SIGNAL
9669	INV(16)	GOOD	Cytocell CBFB-MYH11			79484	90 OF 100	175000	0 OF 100	16000	NA
10212	t(15;17)	GOOD	Cytocell PML-RARA			150000	100 OF 100	250000	0 OF 100	90000	3 OF 100
10139	t(15;17)	GOOD	Cytocell PML-RARA			NOT MADE	FROM DIAGNOSTIC REPORT 95 OF 100	NOT RECORDED	0 OF 100	NOT RECORDED	5 OF 100
8932	t(8;21)	GOOD	Cytocell AML1/ETO			250000	94 OF 100	196000	0 OF 100	21000	NA
8613	TRISOMY 12	INT	ATTEMPTED G BANDING					NOT RECORDED	ASSAY FAIL	NOT RECORDED	ASSAY FAIL
8993	TRISOMY 11	INT	Cytocell MLL (KMT2A)			100000	100 OF 100	50000	0 OF 10	25000	NA
8361	t(6;9)	INT	Kreatech DEK/NUP214	2070	0 OF 10	9635	10 OF 10				
8791	COMPLEX INC TRISOMY 9q34	POOR	Cytocell BCR/ABL1/ASS1			250000	18 OF 100	144000	0 OF 100	24500	NA
9963	COMPLEX INC DEL 7q	POOR	Cytocell Del(7q)			250000	33 OF 100	34000	0 OF 100	110295	NA
9717	COMPLEX INC DEL 7 & +21	POOR	Cytocell Del(7q)			NOT MADE	FROM DIAGNOSTIC REPORT 10 OF 10	NOT RECORDED	21 OF 100	NO GROWTH	

In summary, 10/21 samples underwent LTC of the sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ fraction. The other eleven underwent immediate DNA extraction and analysis following FACS separation.

Of the ten samples which underwent LTC after sorting, eight of the samples grew successfully in 7 week LTC. One failed to grow (10192), and for one (7829) the entire assay, including positive controls, failed to grow (see Section 4.4.2.2.2), therefore precluding further analysis. Of the eight samples which grew successfully, seven were tested successfully for the presence of the leukaemia-associated mutation (five for an *NPM1* mutation, one for *FLT3*-TKD mutation and one by FISH for AML-associated trisomy 11). The final sample underwent G banding to look for the presence of AML-associated Trisomy 12 (in the absence of a relevant FISH probe), which was technically unsuccessful (in that the cells failed to grow in pre-banding culture in the RLH Cytogenetics lab).

Eleven further samples underwent immediate analysis of their sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations, rather than further culture. Of these, eight underwent PCR-based assessment for the presence of a leukaemia-associated *NPM1* mutation, two underwent RNA-Seq analysis, and their transcriptomes were subsequently analysed for reads mapping to the *NPM1* gene (see Chapter 7). One underwent FISH for the AML-associated t(6;9) translocation. In addition, three of the samples which underwent LTC (10192, 7968 and 9850) had sorted sufficient $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells to allow not just LTC, but also parallel immediate analysis.

4.4.3.4.2 Analysis of *NPM1* mutated samples

In total, 16 samples were analysed. 8/16 samples were sorted, with the resultant $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations undergoing immediate DNA extraction and analysis. In 2/16 samples, the sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations underwent RNA extraction immediately, followed by RNA-Seq analysis. A further 6/15 samples were sorted, with the sorted populations then undergoing 7 week LTC-IC assay. 5 of these samples grew colonies and were analysed. For 3/6 of these samples which underwent LTC-IC, sufficient cells were also sorted at the time of plating to facilitate parallel immediate qPCR analysis.

Of these 16 samples, 15/16 had Type A mutations. Sample 9067 had a Type D mutation.

4.4.3.4.2.1 qPCR based assessment of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations from AMLs with known *NPM1* mutation

Results of purity analysis and the number of cells processed for each sample which underwent qPCR based analysis are summarised in Table 4-9.

Table 4-9 illustrating the sorted cell numbers and results for samples undergoing purity analysis by qPCR for NPM1 mutation status Key: * no growth of unsorted cells in LTC, so DNA for positive control obtained from tissue bank stores. ** number of cells not recorded.

ID	SORTED CELLS FOR DNA EXTRACTION				% CELLS CONTANING *NPM1* MUTATION			
	PROCESSED IMMEDIATELY POST SORT		COLONIES DERIVED AFTER LTC		PROCESSED IMMEDIATELY POST SORT		COLONIES DERIVED AFTER LTC	
	$CD34^+38^-ALDH^{high}CLL^-$	UNSELECTED	$CD34^+38^-ALDH^{high}CLL^-$	UNSELECTED	$CD34^+38^-ALDH^{high}CLL^-$	UNSELECTED	$CD34^+38^-ALDH^{high}CLL^-$	UNSELECTED
7774	2358	455635			0	61		
7777	582	371610			21	99		
7793		*	**	NO GROWTH		105*	0**	
7912	398	342642			34	71		
7968	9000	15000	410000	160000	81	73	35	109
8429	141	89076			0	106		
8458		50000	155000	175000		81	0	44
8618	34	15087			0	65		
9719	44	30715			21	95		
9850	1000	300000	5000	70000	18	39	6	97
9878	1056	198006			20	92		
9962		250000	635000	No growth		101	0	
9964	273	503342			0	68		
10192	1000	260000			0	5		

Of the 14 samples tested by qPCR by this methodology, 8/14 of the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations were entirely free of the leukaemia-associated *NPM1* mutation. Of the remaining six samples, in five cases (7777, 7912, 9719, 9850 and 9878), the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells contained a significantly lower proportion of the *NPM1*-associated mutation than the unsorted AML cells. In only one case (7968), the sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations contained the same proportion of mutated cells as the unsorted population.

4.4.3.4.2.2 RNA-Seq based assessment of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations from AMLs with known *NPM1* mutation

In Chapter 7, I describe the sorting of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations from six AML samples, RNA extraction and subsequent transcriptome profiling. Two of these samples (ID 8199 and 9067) had mutations within the *NPM1* gene. Detailed analysis of the sequence of the RNA-Seq reads mapping to Exon 12 of the *NPM1* gene suggested no evidence of the mutation, and therefore we concluded that these sorted populations were free of the leukaemia-associated mutation.

4.4.3.4.3 Analysis of *FLT3*-TKD mutated samples

Although several of the intermediate risk samples contained either a *FLT3*-ITD or TKD mutation (or both) in conjunction with a mutation in the *NPM1* gene, only one sample contained an isolated mutation in the *FLT3* gene as the sole AML-associated mutation. Patient 9713 developed an AML associated with *FLT3*-TKD mutation, detected on clinical screening at diagnosis.

DNA was extracted from the resultant colonies from both sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ and unsorted cells after 7 weeks LTC, as well as unsorted AML cells at the time of FACS separation. After initial PCR for the *FLT3* gene, the resultant DNA was then incubated with the EcoRV restriction enzyme, which normally will cleave the *FLT3* gene within the TKD domain (see Section 4.3.4.2.2).

Figure 4-14 shows the resultant PCR gel after sample processing. The presence of a slow band suggests the co-presence of the TKD mutation, which prevents protein cleavage. However the presence of one, fast band in the DNA extracted from the colonies derived from the sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells suggests that these cells are free from the leukaemia-associated mutation.

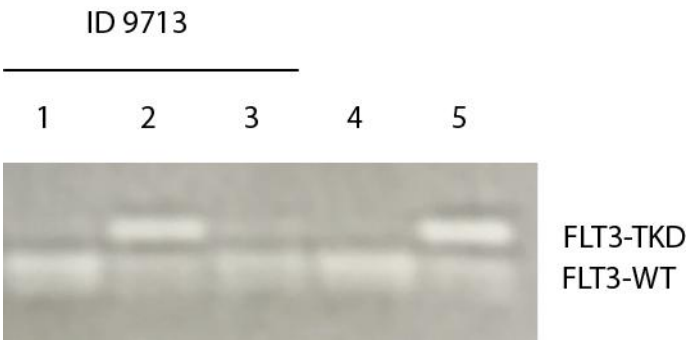

Figure 4-14 Results of PCR gel after digestion for TKD mutation. The presence of 2 bands represents a point mutation within the TKD region. 1: Colonies derived from sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ stem cells after 7 weeks in culture; 2: DNA extracted from unsorted AML sample, uncultured; 3: DNA from unsorted AML, grown in culture for 7 weeks; 4: negative control; 5: positive control

4.4.3.4.4 Analysis of samples with karyotypic abnormalities by FISH

Three samples containing karyotypic abnormalities associated with intermediate risk AML (but *NPM1* and *FLT3* WT) were sorted and their $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cell populations analysed. One (ID 8361) was analysed immediately; in the other two cases, amplification of the sorted cells via 7 week LTC was attempted. The results are summarised in Table 4-8.

In cases 8361 and 8993, the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cell population was free of the AML-associated mutation. Sample 8613 contains an AML with trisomy 12 as its only identifiable genetic lesion. As no appropriate FISH probe was available to test for this, the resultant cells were extracted after 7 week culture, placed in Cytogenetic media and sent for standard Cytogenetic analysis by G banding. Unfortunately, these cells failed to grow in the media used by the RLH Cytogenetics Lab pre-staining, and therefore G banding analysis was not possible for this sample.

4.4.3.5 Poor risk AML Samples

Of ten patients with poor risk disease, nine sorted sufficient $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells to make amplification of cells by LTC viable. Sorting of sample 10147 yielded only 8 cells after two hours of processing. Therefore sort efficiency could not be verified, and LTC was not attempted.

Of the nine remaining samples, three yielded colonies after 7 weeks LTC, and the results of their analysis are shown in *Table 4-8*. In two cases (9584 and 7820), all growth in the assay, including positive controls failed (see Section 4.4.2.2.2). In an additional four samples, the sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells did not yield colonies in LTC. This finding is discussed in more detail in Sections 4.4.4 and 8.2.2.

Samples 8791 and 9963 were analysed by appropriate FISH probes, and the colonies derived from LTC were found to be clear of the AML-associated karyotypic abnormality.

Sample 9717 represents an unusual clinical case, where a 21 year old with no previous chemotherapy exposure presented with an AML associated with both deletion 7 and trisomy 21. Tests for Fanconi's anaemia were negative. Analysis of the cells derived from 7 week culture of sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells showed 21/100 of these cells contained the deletion 7 associated with the AML. FISH on unsorted cells at the time of diagnosis suggested that 10/10 cells examined carried the chromosomal abnormalities, suggesting at least that the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ subset was enriched for normal cells.

4.4.3.6 Summary

In summary, we have definitive data on the purity of the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ fraction sorted from AML cells in 26 samples (4 good risk; 19 intermediate risk; 3 poor risk).

Of the 4 good risk samples, 4/4 of the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ fractions were pure. Of the intermediate risk samples, in 13 of the samples, the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ fraction was clear of the leukaemia associated mutation. In 5 of the samples, the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ subset was "enriched" for normal cells, but the leukaemia associated mutation was detected at low level. In one sample, the sorting strategy demonstrably failed, with the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ fraction containing high levels of the leukaemia associated *NPM1* Mutation. In the poor risk samples, 2 samples when sorted for the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ fraction were clear of the AML associated karyotypic abnormality. The $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ fraction of a third sample was "enriched" for the presence of normal cells.

Therefore, in some 73% (19/26) AML samples, the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ fraction was pure of the AML associated mutation. In a further 25% (6/24) of samples, the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ fraction was "enriched' for the presence of normal cells. These samples may contain frank blasts, or may contain stem cell fractions where some cells contain a pre-leukaemic mutation.

4.4.4 Does the composition and behaviour of the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells within AML samples at diagnosis have clinical significance?

4.4.4.1 Does the absence of growth of sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations have prognostic significance?

20 AML samples were sorted, and the viability of their $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ fractions tested in technically successful (defined by assays where the positive controls yielded colonies) LTC experiments. Of these, 4 contained good risk cytogenetic mutations, 9 intermediate, and 7

poor. Growth of colonies from the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations was observed in 15 samples; of the 5 where LTC did not yield cells, 4 were sorted from poor risk samples and only 1 from an intermediate risk AML. Does an apparent in vitro absence of functional stem cells in these 5 samples have a bearing on the observed clinical outcome of the patient?

4.4.4.1.1 There is no association between absence of growth of sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells within AML samples and overall survival

Figure 4-15 A is a Kaplan-Meier plot comparing overall survival data between samples which showed growth from their $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells, and those that did not. There is no significant difference between these two groups (log rank p=0.49).

4.4.4.1.2 No growth in LTC is significantly associated with disease refractoriness to initial induction chemotherapy

Of the 20 AML samples successfully tested by LTC assay, 17 were treated at diagnosis with intensive induction chemotherapy. 2 patients died within 30 days of diagnosis, leaving 15 survivors who had their response to initial chemotherapy examined by morphological review of bone marrow samples.

3/5 (60%) of the samples where the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ fraction failed to grow in LTC showed evidence of refractory leukaemia at the end of Induction. This compares to 0/10 (0%) of those samples in which colonies were identified at the end of 7 weeks. The association of lack of growth in LTC and refractory disease is statistically significant (Fisher's Exact Test, p=0.022).

Potential confounding factors which might affect the chance of disease remission are age[225] and cytogenetics[225,226]. In an attempt to examine this small dataset for potential confounding factors, the observed distribution of samples between the Growth versus No Growth of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells and age categories was analysed by Fisher's Exact Test (see Table 4-10). Distribution of samples within these categories was not statistically different to that which was observed across the whole cohort (2 sided Fisher's Exact Test, p=0.242).

When examining the potential confounding effect of cytogenetics within these groups, a 3x2 contingency table was constructed (see Table 4-11) and examined by the Freeman-Halton extension of the Fisher's Exact Test. Although this did not reach formal statistical significance (2 sided test, p=0.056), a very strong skew in the distribution of these samples towards the poor cytogenetic risk group is very obvious.

Table 4-10 Cross tabulation for comparison of sample distribution within Growth versus No Growth of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells, and age categories

		AGE CATEGORY		
		<60	**>60**	**TOTAL**
BEHAVIOUR IN LTC	**NO GROWTH**	3	2	5
	GROWTH	9	1	10
	TOTAL	12	3	15

A number of large clinical studies have examined the effect of cytogenetics on the chance of remission induction. A retrospective review of patients treated in CALGB trials between 1985 and 2006, using the European Leukaemia Net Cytogenetics guidelines to risk-stratify patients[97], showed within those with poor risk cytogenetics, patients aged under 60 had an overall CR rate of 50% (n=818), and aged over 60 a CR rate of 39% (n=732)[225]. These figures are comparable to those observed in the No Growth in LTC group, which suggests the difference may simply reflect these patients' underlying cytogenetic abnormality. Further investigation is warranted.

Table 4-11 Cross tabulation for comparison of sample distribution within Growth versus No Growth of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells, and cytogenetic risk categories

		CYTOGENETIC RISK CATEGORY			
		GOOD	**INTERMEDIATE**	**POOR**	**TOTAL**
BEHAVIOUR IN LTC	**NO GROWTH**	0	1	4	5
	GROWTH	3	6	1	10
	TOTAL	3	7	5	15

4.4.4.1.3 An absence of growth in LTC does not preclude normal haematopoietic recovery if remission is achieved

An absence of colony growth from the sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations of cells in 7 week culture would imply an absence of normally functioning haematopoietic stem cells within the bone marrow, if this in vitro assay is an model predictive for in vivo behaviour. Therefore, one could hypothesise these patients might exhibit delayed recovery of normal haematopoiesis after induction chemotherapy, or indeed post-chemotherapy aplasia.

Of the five samples where the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ fraction failed to produce colonies in LTC, three did not regain normal blood counts after induction, but in all of these cases, this was also in the context of refractory leukaemia. However, in the two cases (patients 10147 and 9583) where there was no evidence of residual disease, count recovery occurred within the normal

time frame (neutrophils >1.0x10^9/L at days 26 and 17, and platelets >100x10^9/L at days 25 and 32 for Patients 10192 and 9583 respectively).

4.4.4.2 Does the detection of a leukaemia-associated mutation within the sorted CD34$^+$38$^-$ ALDH^{high}CLL1$^-$ population ('Mixed' sort) have prognostic significance?

As summarised in Section 4.4.3.6 stated above, we have definitive data on the purity of the sorted CD34$^+$38$^-$ALDH^{high}CLL1$^-$ population in 26 AML samples. Nineteen of these contained no evidence of the leukaemia-associated mutation; six contained a significantly lower proportion of leukaemia cells than unsorted fractions; and one contained the same frequency of mutated cells as the matched unsorted sample.

Previous data series discovered a significant association between AML samples with high ALDH expression and poor overall survival[208,209]. Although my approach (using a combination of markers to separate out stem cells from AML rather than using ALDH expression alone) might reduce this impact, I had hypothesised that an AML which shared the same phenotypic appearance as a normal stem cell might have a worse overall outcome.

However, in this dataset, there is no association between the presence of a leukaemia associated mutation within the CD34$^+$38$^-$ALDH^{high}CLL1$^-$ population and a poor clinical outcome. If anything, the trend appears to be in the other direction.

4.4.4.2.1 There is no significant association between overall survival and the presence of a leukaemia-associated mutation within the CD34$^+$38$^-$ALDH^{high}CLL1$^-$ population

Figure 4-15 B is a Kaplan-Meier plot comparing the overall survival of patients when split into two groups dependant on the purity of the sorted CD34$^+$38$^-$ALDH^{high}CLL1$^-$ populations. There is no significant difference between the two cohorts in terms of overall survival (p=0.17).

Table 4-12 Cross-tabulation showing comparison of sample distribution within Mixed versus Pure CD34$^+$38$^-$ ALDH^{high}CLL1$^-$ populations and age categories. Fisher's Exact Test, p=1.00

		AGE CATEGORY		
		<60	**>60**	**TOTAL**
ASSESSMENT OF CD34$^+$38$^-$ALDH^{high}CLL1$^-$ CELLS	**MIXED**	5	2	7
	PURE	14	5	19
	TOTAL	19	7	26

Table 4-13 Cross-tabulation showing comparison of sample distribution within Mixed versus Pure $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations and cytogenetic risk categories. Freeman-Halton extension of the Fisher's Exact Test, p=0.62

		CYTOGENETIC RISK CATEGORY			
		GOOD	**INTER**	**POOR**	**TOTAL**
ASSESSMENT OF $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ CELLS	**MIXED**	0	6	1	7
	PURE	4	13	2	19
	TOTAL	4	19	3	26

There is no obvious skewing of sample distribution between the Mixed versus Pure sort groups and the potential confounding factors of age and cytogenetics (see Table 4-12 and Table 4-13).

4.4.4.2.2 There is no significant association between relapse-free survival and the presence of a leukaemia-associated mutation within the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ population

Of this set of 26 patients, 23 received chemotherapy with curative intent. 4 died in induction. All of the remaining 19 patients achieved remission after cycle 1 of chemotherapy, and of these, data are available for the assessment of relapse free survival in 18.

Comparison of relapse-free survival for the "Mixed" (n=5) and "Pure" (n=13) $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ sample groups again showed no significant difference between the two cohorts (see Figure 4-15 C).

Table 4-14 Cross-tabulation showing comparison of sample distribution within the Mixed versus Pure $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations and age categories. Fisher's Exact Test, p=1.00

		AGE CATEGORY		
		<60	**>60**	**TOTAL**
ASSESSMENT OF $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ CELLS	**MIXED**	4	1	5
	PURE	11	2	13
	TOTAL	15	3	18

Table 4-15 Cross-tabulation showing comparison of sample distribution within the Mixed versus Pure $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations and cytogenetic risk categories. Freeman-Halton extension of the Fisher's Exact Test, p=0.16

		CYTOGENETIC RISK CATEGORY			
		GOOD	**INTER**	**POOR**	**TOTAL**
ASSESSMENT OF $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ CELLS	**MIXED**	0	4	1	5
	PURE	3	10	0	13
	TOTAL	3	14	1	18

As before, there is no obvious skewing of sample distribution between the "Mixed" versus "Pure" sort groups and the potential confounding factors of age and cytogenetics (see Table 4-14 and Table 4-15).

4.4.4.2.3 The presence of leukaemia-associated mutation within the $CD34^+38^-ALDH^{high}CLL1^-$ population might be associated with improved overall survival within the *NPM1* mutated cohort (but does not reach clinical significance)

The results illustrated in Sections 4.4.4.2.1 and 0 seem particularly surprising given that the "Mixed" samples are either of Intermediate (n=6) and Poor (n=1) risk cytogenetics, and contain no good risk samples (although I note that this distribution of samples (see Table 4-13) is not sufficiently skewed to reach significance when analysed by the Freeman-Halton extension of the Fisher's Exact test). Nevertheless, as all of the samples with mixed $CD34^+38^-ALDH^{high}CLL1^-$ populations within the Intermediate group carry the *NPM1* mutation, an overall survival analysis was performed comparing the Mixed (n=6) versus Pure (n=10) samples which carry the *NPM1* mutation. This is displayed graphically in Figure 4-15 D. Although the difference in survival between the two cohorts does not quite achieve statistical significance by log rank test (p=0.07), an interesting trend toward increased survival in those samples with Mixed $CD34^+38^-ALDH^{high}CLL1^-$ populations is displayed, which is more marked than when considering all the cohort together. As before, analysis of the distribution of samples within these groups between age categories is not significantly skewed (see Table 4-16, Fisher's Exact Test, p=1.00).

Table 4-16 Cross-tabulation showing samples distribution within Mixed versus Pure $CD34^+38^-ALDH^{high}CLL1^-$ populations and age categories for *NPM1* mutated samples only

		AGE CATEGORY		
		<60	**>60**	**TOTAL**
ASSESSMENT OF $CD34^+38^-ALDH^{high}CLL1^-$ CELLS	**MIXED**	4	2	6
	PURE	7	3	10
	TOTAL	11	5	16

This (non-significant) trend towards an improved overall survival with a Mixed $CD34^+38^-ALDH^{high}CLL1^-$ population becomes more intriguing when considering the *FLT3* mutational status of the samples in question. Unfortunately, we only have an incomplete dataset available (see Table 4-6). FLT3 mutational status is documented for 8 of the *NPM1* mutated samples. Of those with data available, 3/3 of those samples with Mixed $CD34^+38^-ALDH^{high}CLL1^-$ populations were *FLT3*-ITD mutated, as opposed to 2/5 of those with Pure $CD34^+38^-ALDH^{high}CLL1^-$ sorts. It is well documented that the overall survival of those patients with concurrent *NPM1* and *FLT3*-

ITD mutations is inferior to those with *NPM1* mutations alone[103,227]. Further discussion of this result, and other potential explanations for the observed patterns are found in Section 4.5.5.

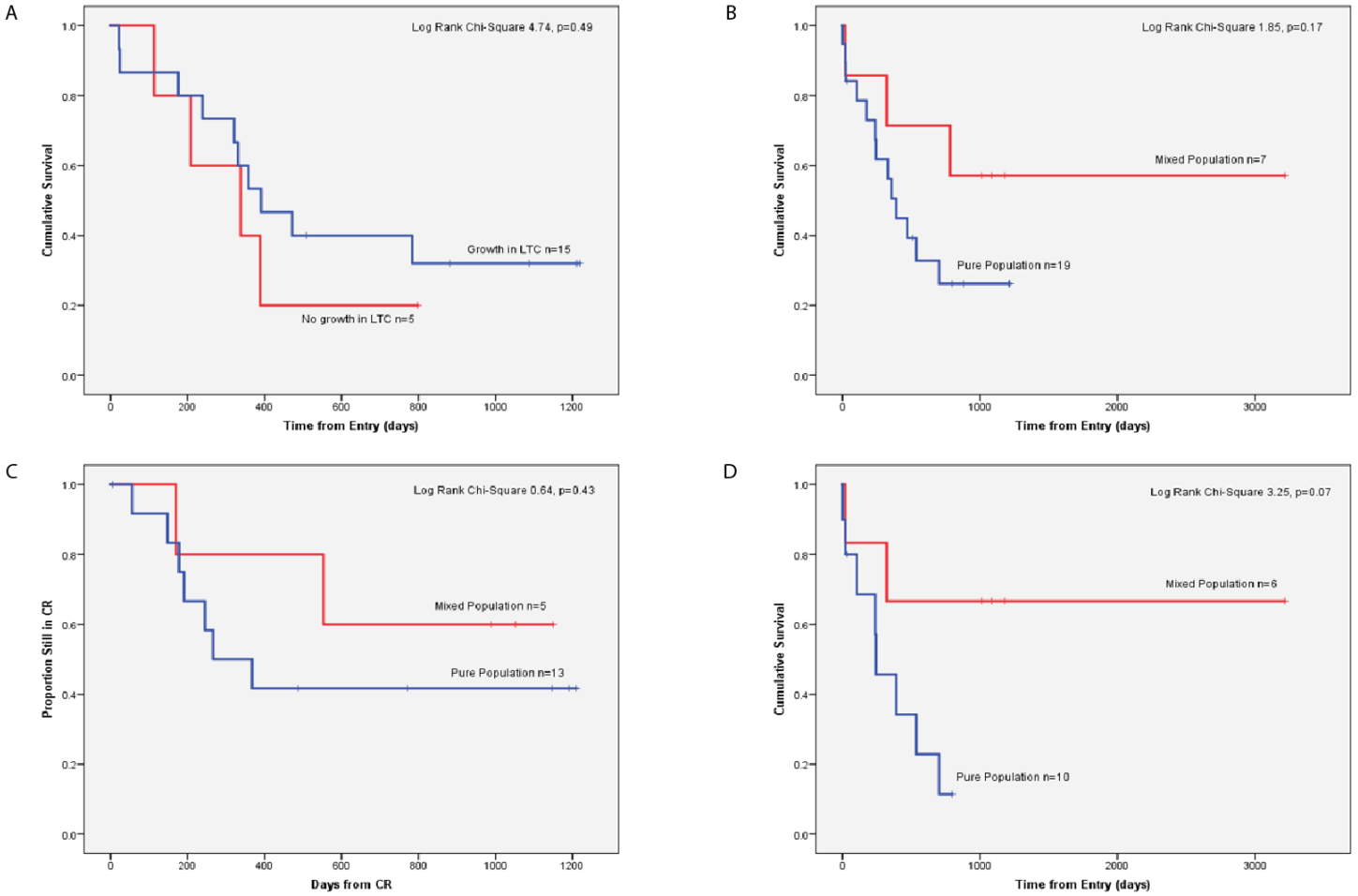

Figure 4-15 Kaplan-Meier curves showing the effects of A: growth versus no growth on overall survival; B: Mixed versus pure composition of the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ population on overall survival; C: Mixed versus pure composition of the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ population on relapse free survival; D: Mixed versus pure composition of $CD34^{+}38$-$ALDH^{high}CLL1^{-}$ population within *NPM1* mutated samples only

4.5 Discussion

4.5.1 General points

In this chapter, I describe attempts to design and validate a panel of antigenic markers, capable of distinguishing normal stem cells from leukaemia cells in unselected AML samples. In doing this, we had to overcome a number of technical hurdles. Optimising FACS sorting conditions to facilitate the accurate separation of tiny numbers of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells (which comprised on average 0.01% of cells within AML samples); identifying methods of reliably assessing the genetic makeup of very small numbers of sorted cells, and assessing stem cell functionality within the 7 week LTC assay posed the three biggest experimental challenges.

As discussed in Section 2.4.5.5, sorted samples were accepted if greater than 90% of the sorted cells fulfilled the initial sort criteria on reanalysis. This approach is widely accepted as a pragmatic approach to limit repeated sorting and where possible preserve cell viability. Nevertheless, this also means there is an inherent acceptance that up to 10% of sorted cells may therefore be leukaemic in nature, simply due to failed physical separation by the FACS machinery.

The panel of antigens selected to identify normal HSCs appeared to successfully mark out cells with a normal genotype in 73% (19/26) of the AML samples. In a further 25% (6/24) of samples, the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ fraction appears "enriched" for normal cells, containing a significantly lower proportion of cells carrying the leukaemia-associated mutation than the unsorted population.

Interestingly, predicting whether the panel of antibodies is able to select successfully for normal stem cells based solely on FACS appearances is extremely difficult. A good illustration of this is provided the FACS plots of ID 9717 (see Figure 4-12), where there was excellent demarcation of the $CD34^{+}38^{-}$ gate into high, intermediate and low ALDH-expressing populations (with little evidence of overlap), and yet 21% of the final sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ population carried a leukaemia-associated mutation.

However, the evidence would suggest that this panel can reliably separate normal stem cells from patients with good risk cytogenetic abnormalities (4/4 samples tested). In intermediate risk samples tested, 69% (13/19) of samples successfully separated; 26% (5/19) were enriched for the leukaemia associated mutation, and in 5% the technique failed (1/19). Analysis of the panel's efficacy in separating normal stem cells from AML cells in poor risk samples was much

more limited, as an inability to amplify several samples in LTC meant only three samples grew sufficiently to be analysed by FISH. Of these, two samples when sorted for the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ fraction were clear of the AML associated karyotypic abnormality. The $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ fraction of a third sample was "enriched" for the presence of normal cells.

4.5.2 Further experimental applications of the panel

This panel has subsequently been used for the majority of the experimental work on normal stem cells in the context of AML documented in the rest of this thesis. It is key to the work detailed in Chapter 5, where stem cell numbers at diagnosis were calculated in a range of sample karyotypes. Without its use, any study would be limited to the cohort of $CD34^{low}$ AMLs we have described already.

However, as documented above, it is almost impossible to predict from the FACS based appearances of samples whether they would sort pure or not using this approach. Therefore, wherever possible, the results of each sort are validated by testing the extracted cells for the presence of a leukaemia-specific mutation. For key experiments, where I was generally unable to sort sufficient cells to both perform the experiment and validate the population purity, I relied on using samples where the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells are almost certainly clear of the leukaemia-associated mutation: namely those with good risk cytogenetic abnormalities and $CD34^{low}$ samples. These AMLs therefore represent the study groups used in Chapters 6 and 7.

4.5.3 Why is there a difference in the panel's ability to separate from different cytogenetic risk groups?

It is interesting to hypothesise why there are differences in this panel's ability to separate out normal HSCs from different AML samples.

The most obvious explanation is that good risk cytogenetic AML samples do not share the phenotypic appearance of normal stem cells as defined by this panel. Indeed, Gerber et al showed that ALDH expression patterns alone were sufficient to separate out normal and leukaemia cells from CBF-mutated samples[207]. The cell of origin of the leukaemia might play a part: those leukaemias derived from mutations occurring in the HSC compartment might be more likely to have an overlapping phenotype than those evolving from a partially committed progenitor[228].

Alternatively, does the varying sensitivity of the analysis techniques contribute to the differences in separation ability seen? 5/6 of the samples which failed to sort pure carry *NPM1*

mutations. Does the high sensitivity of this assay make it more likely to detect the presence of contaminating leukaemia cells than FISH?

Or finally, is purity (for those samples analysed after LTC), more a reflection of the ability of the AML itself to survive in LTC culture, than the composition of the initial $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ population? More discussion of this possibility is to be found in Section 4.5.7.

4.5.4 What is the definition of a normal HSC anyway? The issue of pre-leukaemic stem cells

Evidence for the existence of pre-leukaemic stem cells in AML samples was first published in 2012, after the start of work towards this thesis. The key initial papers described whole genome sequencing of sorted single cells of HSC phenotype from selected AML samples[107,118,151]. Analysis of these cells led to a hypothesis of the accumulation of sequential mutations within the HSC fraction, which eventually results in transformation to cells of the leukaemia blast phenotype. If chemotherapy proves effective against the leukaemic clone and remission is achieved, any residual pre-leukaemic HSCs may either persist in remission, or alternatively begin the process of accumulation of further mutations which may result in eventual relapse.

This field brings a new dimension to the study of "normal stem cells" in AML. These pre-leukaemic stem cells may retain the ability to survive long-term in the bone marrow and differentiate into mature cells, but the mutations they carry might confer a competitive growth advantage over other, non-mutated HSCs[107]. Key unanswered questions include the following. Are pre-leukaemic stem cells to be found universally in all AML samples? What proportion of the "normal" HSC fraction is made up of pre-leukaemic stem cells? Do these cells ever differ phenotypically from the normal HSC fraction?

Acquisition of the *NPM1* mutation is thought to represent a key step in the transformation to AML, as it normally detected in all leukaemic cells within samples which contain the mutation; is an accurate MRD marker, and is normally retained at clinical relapse[111]. However, as discussed in Sections 1.4.5.4.1 and 1.5.4, the key early papers in the field suggest that the *NPM1* mutation is largely missing from the pre-leukaemic HSC population, where mutations in genes such as *DNMT3A* are more commonly found[107,108]. It therefore seems unlikely that the detection of low levels of *NPM1* mutated cells within the HSC fraction (as seen in samples 9850, 7912, 7777, 9878 and 9717) represents the presence of pre-leukaemic stem cells rather than true leukaemia cells, but this remains a question worth asking. Experimentally, one could

approach this question by transplanting these "Mixed" $CD34^+38^-ALDH^{high}CLL1^-$ populations into a permissive xenograft model, and examining the resultant $CD45^+$ cells both morphologically and by flow for the presence of leukaemic blasts, although the inefficiencies of the xenograft transplant model might limit the usefulness of such an approach[229].

4.5.5 Is there a clinical association between dirty and pure sorts?

As the results presented in Section 4.4.4.2 shows, there were no statistically significant correlations found between clinical outcome (as measured by overall survival and relapse-free survival) and the composition of the $CD34^+38^-ALDH^{high}CLL1^-$ cell population. However, there did appear to be a trend when considering *NPM1* mutated samples (albeit one which did not reach overall significance on log rank testing) towards improved outcome, when the $CD34^+38^-ALDH^{high}CLL1^-$ cells contained the leukaemia associated mutation

This finding is surprising, given that high ALDH expression in AML samples (and so presumably an increased chance of a "Mixed" population of $CD34^+38^-ALDH^{high}CLL1^-$ cells) has been associated with poor clinical outcome in at least two independent datasets[208,209].

It is important to stress that any conclusions are limited by the very small size of this dataset. Although treatment pathways have not changed dramatically during the period of sample acquisition (2008 to 2014), the structure of the AML-17 trial (into which many of the later patients were recruited) favours an increased frequency of early allogeneic transplant in first remission[230]. Five of the seven samples where the $CD34^+38^-ALDH^{high}CLL1^-$ cell populations contained a leukaemia-associated mutation received an allogeneic transplant, of which four were in first remission. This trend for early transplantation might have an effect on the overall survival, which might be a significant confounder in this small dataset.

4.5.6 What is the significance of the failure of cells to grow in culture?

One striking result from analysis of the poor risk cytogenetic samples is that a significant number of the samples sorted and then grown in LTC, showed no colony development from their $CD34^+38^-ALDH^{high}CLL1^-$ populations. In contrast, in all of the four good risk AMLs studied, the wells plated with $CD34^+38^-ALDH^{high}CLL1^-$ cells grew colonies. Similarly, in eight of the nine intermediate risk AML samples, colonies were observed within the wells plated with $CD34^+38^-ALDH^{high}CLL1^-$ cells. Could it be that in this group of patients, the residual ability for normal haematopoiesis is substantially reduced, even at presentation? This clearly has profound implications, as it suggests, even in the presence of chemosensitive disease, that these patients might struggle to recover normal haematopoiesis without transplantation.

Why would these cells fail to develop colonies? Once the technical difficulties with the MS5 feeder lines described in Section 4.4.2.2.2 had been resolved, the assay appeared to be extremely sensitive, generating reproducible results with $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cell inputs as low as two per well (see Section 6.4.2).

There are a number of possible explanations for this observation. Poor risk patients, who often present with a disease with a slower tempo of evolution[226], might simply contain no "normal" stem cells at the point of presentation. This could either be because these stem cells have themselves accumulated inactivating genetic mutations, or that their interaction with AML cells has rendered them functionally useless. We know that recovery with incomplete counts is more common in patients with poor risk cytogenetics[231]. However, this clearly does not entirely model the in vivo behaviour of these AMLS (see section 4.4.4.1.3), as the two cases that achieved leukaemic remission after induction, recovered normal blood counts briskly.

Alternatively, poor risk AML may be a more powerful inhibitor of normal haematopoietic stem cells than good or intermediate risk samples, inducing such a degree of quiescence within the HSC compartment, that the cells within it are incapable of regeneration over the 7 week LTC period.

A final explanation could be that within these poor risk samples, the sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ population contains a significant proportion of leukaemic blasts alongside the normal HSC fraction. During the 7 week LTC, the presence of blasts might impede recovery of normal haematopoietic function. However, several lines of evidence make this theory less likely. In two of the four samples where the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells failed to produce colonies, wells of unsorted AML cells plated at the same time produced colonies. If the AML itself is capable of surviving the culture period, why do contaminating AML cells not similarly produce colonies in the wells plated with $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells? Secondarily, I have already described 7 samples where the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells generated mixed populations of normal and leukaemia-associated cells in culture, suggesting that in these samples, AML cells and normal HSCs can co-exist in vitro.

In the absence of the ability to amplify the sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells of these samples by LTC, it is difficult to address the question as to whether these cells contain leukaemia-associated mutations. Poor risk samples often provide the most challenging samples to analyse due to the complexity of the cytogenetic rearrangements seen. If amplification of sorted cells to facilitate standard cytogenetic analysis or FISH is not possible, alternative approaches need

to be explored. One possibility is single cell sorting of the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ population, and subsequent whole genome sequencing. Alternatively, PCR-based analysis (analogous to the method used for the *NPM1* mutated samples) could be used if an appropriate mutation was identified for this cohort. One candidate is the *p53* gene, which is commonly mutated in patients with poor risk cytogenetics. One study of de novo AML patients reported the incidence of *p53* mutations to be 7% across all subtypes, but rising to 59% in those with complex karyotypes. However, unlike the *NPM1* mutation, mutations in the *p53* gene can occur along the length of the gene, and so each sample would have to be individually sequenced before appropriate primers could be identified. In addition, serial sample analysis suggests *p53* mutations are frequently lost on relapse, suggesting unlike the *NPM1* mutation, this may represent a clonal marker, rather than a primary mutational event. This has implications for the sensitivity of such an assay[232].

If more sophisticated analysis of the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations in these cases did show evidence that a significant number of cells contained the leukaemia-associated mutation (suggesting this panel was not able to separate normal stem cells from AML in this cohort), one could reapproach the design of the panel very specifically for poor risk cytogenetic samples. Most of the reported markers in Section 4.1.3 were identified using panels of AMLs representing normal or good risk disease. One approach would be to mimic the work of Kikushige et al[198] by exploring differential gene expression between sorted normal $CD34^{+}38^{-}$ cells and the $CD34^{+}38^{-}$ fraction in a range of poor risk AMLs. Such an experiment could then screen for differentially expressed surface markers that might be more appropriate for this cohort of samples.

4.5.7 Concerns over the validity of the experimental design: LTC appears to preferentially select the growth of normal haematopoietic stem cells over AML cells, and the strength of this selection may differ between samples

In this chapter, the use of the 7 week LTC assay was key to the experimental set up: enabling not just assessment of the functionality of the sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells, but also facilitating the amplification of sorted cells prior to genetic analysis. A key assumption was that if a population of sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$cells contained a mixture of normal and leukaemia cells, both populations would amplify in LTC and be detectable at the end of 7 weeks. However, close analysis of the results in Table 4-7, Table 4-8 and Table 4-9 shows that the LTC system appears to preferentially support the growth of normal haematopoietic stem cells over their leukaemic counterparts. This was an unexpected finding.

In the assay design for the LTC, at the time of plating sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$cells, two to four wells of unsorted cells (containing approximately 50,000 cells per well) from the same sample were also added. This enabled us to observe if each AML sample was capable of survival in the LTC assay. The absence of cell growth in these wells at the end of the 7 week period would therefore raise concerns that assessment of the colonies from the sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells might be unrepresentative of the purity of the initial sorted population.

In addition, for the majority of samples characterised by a karyotypic abnormality detectable by FISH analysis, slides were made of the unsorted sample at the start of the experiment for use as a positive control during FISH staining. These slides were stored at $4^{o}C$ for 7 weeks until the LTC was completed, and then analysed alongside the slides from colonies derived from the relevant population of sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells. This was done to ensure there was a slide available for a positive control in each case, even if the unsorted cells failed to yield colonies in LTC.

Therefore, the majority of samples had two potential slides for use as a positive control: a slide made at the start of the experiment from unsorted cells, and one made at the end from the colonies derived from unsorted cells. However, because of cost limitations around the staining of slides with FISH probes (and because I was unaware of the potential issue of preferential growth of normal cells over AML colonies), for the majority of these samples, I only used the slides made at the start of the experiment as positive controls.

Samples 10139 and 10212 (both of which represent APML) had slides both made from unsorted cells at the start of the experiment, and also from the colonies grown from unsorted cells. In these cases, both slides were analysed alongside the colonies derived from the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells. In both cases, the proportion of cells within the colonies grown from sorted AML cells which contained the leukaemia-specific mutation was markedly lower than on the slides made at the start of the experiment (Sample 10139: 5% from 95%; Sample 10212: 3% from 100%).

It is also noteworthy in the case of sample 9717 (a sample with complex cytogenetics which included del(7) and +21), that there was no growth observed in any wells plated with unsorted cells (as discussed in Section 4.4.3.5). This raises the concern that cells carrying the leukaemic mutation lack the ability to survive in vivo culture, seriously impacting the validity of any purity assessment of this sample. Interestingly however, assessment of the colonies derived from

$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells showed 21% of these cells to carry the leukaemia-associated del(7q). Quite why these cells proved able to survive in these conditions and not in others is unclear. In this patient, could this mutation mark a population of cells of pre-leukaemia status?

Similar patterns of selection for normality are also observed in the cohort of intermediate risk samples tested for the presence of *NPM1* or *FLT3* mutations after LTC. In all of these cases, DNA was extracted from unsorted cells at the start of the LTC, as well as from colonies derived after 7 weeks.

In the case of sample 9713, an AML with an associated *FLT3*-TKD mutation, visual examination of the PCR gel shown in Figure 4-14 shows a much stronger slow band (indicative of *FLT3*-TKD mutated PCR product) in Well 2 (DNA from unsorted AML cells processed immediately) than in Well 3 (DNA from unsorted AML cells after 7 week culture).

NPM1-mutated AMLs behave similarly. In the case of samples 7793 and 9962, assessment of the sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ colonies at the end of culture suggested these were entirely free of the leukaemia-associated mutation, which was present in 100% of the cells analysed before sorting. However, the unsorted AML cells failed to grow in culture. It is unclear if this absence of detectable mutated *NPM1* in the colonies derived from sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells is because initial sorting was successful, or simply due to the inability of the AML to grow in culture. Samples 7968 and 9850 sorted sufficient $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells at the start of the experiment to allow immediate analysis, as well as LTC followed by analysis of the derived colonies. In both cases, the estimated percentages of leukaemia cells within the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations were significantly higher in the cells analysed immediately post-sorting (for samples 7968: 81 versus 35%; for sample 9850: 18 versus 6%).

4.5.7.1 Preferential ability of different subtypes of AML to survive in vivo culture in mice has also been documented

Is there an alternative experimental design which could remove these sources of potential error? I had utilised an in vitro method of HSC amplification largely due to cost and sensitivity issues, but review of the literature would suggest that permissive murine xenograft models may well exhibit a similar bias.

Pearce et al explored the causes of the observed variability of AML sample engraftment into NOD/SCID mice. 59 primary samples, representing a range of karyotypic abnormalities, were transplanted into xenograft models at the same cell dose. Engraftment was assessed at 6 weeks by both morphological and flow based analysis of the murine bone marrow. The

cytogenetic risk group of the samples had a marked effect on the success of AML transplantation: 0/11 good risk karyotype samples engrafted, as compared to 18/36 intermediate samples and 5/5 poor risk samples. Interestingly, in 7 of the samples where the AML failed to engraft, there was evidence of normal human haematopoietic development instead (the co-existence of $CD45^{+}33^{+}$ myeloid and $CD45^{+}19^{+}$ lymphoid colonies). The authors do not comment on whether there was evidence of low level normal haematopoietic activity seen in those 23 samples where AML engraftment was successful[229]. By contrast, normal haematopoietic tissue consistently engrafts NOD/SCID mice[233].

4.5.8 Further experiments

I have envisaged a number of experiments which might address unanswered questions from this work.

Primarily, it would be vital to increase the size of the AML sample cohort studied. The sample size in this chapter was largely limited by the expense of lengthy sorting and LTC; however, to determine if either the final composition of the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ population, or its ability to survive in LTC, have clinical relevance, a much larger study cohort is required.

Increasing funds would also enable study of a large group of AML samples which have to date remained elusive: those with a normal karyotype, and no associated *NPM1* or *FLT3* mutation. The advent of high-throughput screening means targeted sequencing of the most commonly mutated genes in AML can now be performed in a relatively cost-effective manner. Such a panel might screen for changes within the mutational hotspots of genes such as *ASXL1, CBL, CEBPA, DNMT3A, ETV6/TEL, EZH2, GATA2, IDH1&2, KDM6A, KIT, KRAS, NRAS, RUNX1, SF3B1, SRSF2, STAG2, TET2, TP53, U2AF1 and ZRSR*. Detection of a mutation would then lead to design of primers capable of detecting the leukaemia-specific mutation within the sorted stem cell population.

One of the most interesting findings within this chapter is the apparent absence of normal stem cell function shown by a significant proportion of poor risk samples. If this finding were replicated in a wider cohort of samples, it raises key questions about how important residual stem cell functionality is in prognosis and whether such knowledge should modulate our decision making around treatment.

However, I feel it is important to stress some of my anxieties with this data. It is very difficult to explain the normal recovery of blood counts in two patients post-successful remission induction, if the in vitro data truly reflect the in vivo situation. It would be interesting to repeat

the culture of sorted $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells from these AML samples, but using a xenograft model rather than an in vitro assay, to see if the same patterns are observed.

As already discussed in Section 4.5.6, it has proved difficult to determine the composition of the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations as these cells failed to amplify. Either the development of a PCR based assay for assessment of a pooled population of cells (for any of the genes detailed above found to be mutated within the AML blasts), or the adoption of single cell sorting and whole genome sequencing might help to answer this question.

Finally, to try and answer if the in vivo model is a good model for assessing stem cell functionality in vivo, I would be interested in specifically investigating the small cohort of patients which have been rendered hypoplastic or aplastic (but with blast clearance) after chemotherapy. It would be fascinating to identify their pre-chemotherapy stem cell populations, and document their behaviour in vivo, to see if the resultant aplasia could have been predicted.

Chapter 5 Enumerating Stem Cells in Primary Diagnostic Bone Marrow samples from Patients Presenting with Acute Myeloid Leukaemia

5.1 Introduction

Our group is interested in the effect of acute myeloid leukaemia (AML) on the behaviour of normal HSCs in bone marrow, and how this interaction modulates downstream haematopoiesis. One hypothesis for the trilineage haematopoietic failure often observed in AML is that competition between AML leukaemia stem cells (LSCs) and normal HSCs might lead to death of the healthy cells, or disruption of their normal function. However, prior experimental work by our group using both mouse xenograft models and human primary diagnostic AML samples, has led us to the conclusion that HSC numbers are (at least initially) preserved, but that the proportion of cells in cycle is reduced.

The group's work on human diagnostic AML samples was limited to 16 $CD34^{low}$ AML biopsies, which, as described in Section 1.5.2.2, are ideal for this purpose, as their blast populations are solely found in the $CD34^{low}$ cell fraction. This phenotypic appearance is commonly associated with normal karyotype, *NPM1* mutated samples. However, AMLs with this mutation pattern are associated with distinct clinical[103], cytogenetic[234], molecular[235], and immunologic features[236]. As a result, they are now included as a separate entity within the WHO classification of haematological malignancies[112]. These 16 patients represent a highly selected, homogenous group of patients, which may not represent the diversity of behaviour across AML subtypes.

5.2 Aims and Objectives

The primary aim of this chapter was to utilize the techniques described and validated in Chapter 4 to enumerate HSC numbers across a wide variety of AML subtypes at diagnosis. This was to enable us to see if the observations we had previously made concerning stem cell preservation and quiescence in leukaemia samples with low CD34 expression were applicable to a greater range of AML samples[1].

A secondary aim was to investigate if there was any correlation between stem cell numbers at diagnosis and clinical outcome, as measured by time to haematopoietic recovery and overall survival.

5.3 Specific methods

5.3.1 Processing of Fresh Samples

5.3.1.1 Sample Extraction

In an attempt to limit the haemodilution which occurs with extensive BM removal, a standard operating procedure was used to ensure that the BM samples used for stem cell enumeration were obtained from the first pull of the aspirate, and in a small volume (less than 0.5ml) to reduce haemodilution for quantification of normal progenitors and HSCs. All patients presenting to St Bartholomew's Hospital with a possible diagnosis of acute leukaemia, and all patients undergoing staging bone marrows for lymphomas were requested to have a sample taken. Samples were decanted into Eppendorfs containing 100µl Anticoagulant Citrate Dextrose Solution (Haemonetics, Cat: 426C).

5.3.1.2 Sample Processing

Samples were in general analysed the same day as they were taken, and always within 24 hours. Between sample extraction and analysis, samples were kept at 4^{o}C. Any sample with a visible clot at the time of analysis was discarded.

200µl of fresh BM was mixed with 4ml ammonium chloride solution (Stem Cell Technologies, Cat: 07850), and incubated at 4^{o}C for 10 minutes to induce red cell lysis. 1ml FBS was added and the sample pelleted (1500rpm, 5 minutes). The cells were resuspended in 50µl 2% HAG, mixed well and incubated for 20 minutes at 4^{o}C. 5µl of CD34-PerCP (BD, Cat: 345803 (clone 8G12)) and CD38-Pecy7 (BD, Cat: 335825 (clone HB7)) were added, and the sample incubated for 30 minutes at 4^{o}C. The sample was then washed in 4ml 2% PBS, and resuspended in 300µl 2% PBS/DAPI/DNase. Cell counting was facilitated by the addition of 50µl CountBright™ Absolute Counting Beads (Life Technologies, Cat: C36950). Analysis was performed using a BD Biosciences LSR.

5.3.1.3 Confirmation of Clinical Details and Appropriateness for Further Investigation

As samples were often taken before a firm diagnosis of AML were made, and control samples taken as part of lymphoma staging, no further analysis was undertaken until cross-referencing with the St Bartholomew's Hospital Diagnostic Database was undertaken. Only those samples from patients with a confirmed diagnosis of AML, or those from newly diagnosed Stage I-III lymphoma patients with a full blood count within normal parameters with an unaffected bone marrow as confirmed by aspirate, flow cytometry and trephine histopathology were analyzed further. This represented approximately 50% of the samples initially received.

5.3.1.4 Analysis of the Initial Sample

Cell numbers within a sample of known volume were calculated using the CountBright System (Molecular Probes). A known volume of a calibrated suspension of fluorescent microspheres was added to a defined volume of test sample, therefore producing a known sample volume to microsphere ratio. The volume of the sample analysed can therefore be calculated by using the recorded numbers of microsphere events.

Sample analysis on the flow cytometer was followed by bead counting on a plot depicting Side scatter on a log scale versus R670/14 on a linear scale. Gating was then performed to enumerate the number of live, single $CD34^+$ cells present in the sample, by the gating strategy illustrated in Figure 5-1 A.

The concentration of $CD34^+$ cells present in the original sample was then calculated by the following formula:

$$cell\ concentration\ (cells/\mu l) = \frac{a}{b} \times \frac{c}{d}$$

Where *a:* number of cell events, *b:* number of bead events, *c:* assigned bead count of the individual CountBright batch (beads/50µl) and *d:* volume of the sample used (µl)

5.3.2 Subsequent Stem Cell Identification of Stored, Frozen BM Samples

5.3.2.1 Sample processing

At a later date, previously stored bone marrow samples, taken at the same time as the samples analysed fresh in Section 5.3.1.4, were accessed. Samples were processed, stained for CD34, CD38, ALDH and CLL1 expression and analysed as described in Section 4.3.2.2.

5.3.2.2 Analysis of the frozen sample

HSC numbers (the $CD34^+CD38^-ALDH^{high}CLL1^-$ population) in the frozen sample were determined by the sequential gating strategy shown in Figure 5-1 B. $CD34^+$ cell numbers in the same sample were calculated by the sequential gating shown in the same figure.

The original concentration of $CD34^+CD38^-ALDH^{high}CLL1^-$ cells (cells/µl) at diagnosis was calculated as shown below.

$$\left[CD34^+38^-ALDH^{high}CLL1^-\right]cells =$$

$$[CD34^+\ cell]\ in\ fresh\ sample\ \times \left\{\frac{CD34^+38^-ALDH^{high}CLL1^-\ cells}{CD34^+cells}\right\}\ in\ frozen\ sample$$

5.3.2.3 Validating the purity of the identified $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ population

As described in Chapter 4, a clear population of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells (predicted to be stem cells, free of the leukemic clone) can be identified in the majority of patients with AML. However, as discussed in Section 4.5.1, even in those samples in which a clear, separate, population of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells can be identified, those cells are not always karyotypically normal. To be able to confidently count residual stem cell numbers in diagnostic bone marrow samples, it is vital to be able to confirm the sorted cells are genuinely pure of the leukaemia-associated mutation. Therefore, in all cases possible, an attempt was made to establish the purity of the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ sorted cells from the leukaemic clone, by the most appropriate method. Techniques utilized were conventional cytogenetic analysis, FISH and PCR specific for mutations within the *NPM1* or *FLT3* genes.

5.3.3 Clinical Outcome Data

Treatment details, time to count recovery and clinical outcome data were obtained by reference to hospital records. Final censoring of patients took place in April 2016.

5.3.4 Statistical analysis of data

The majority of statistical analysis was performed using the Prism software package. All data sets were tested for normality of distribution by the D'Agostino & Pearson omnibus normality test. If data were not normally distributed, a Mann-Whitney t-test was applied in the determination of statistical significance.

Clinical outcome data analysis was performed using the SPSS statistical package. Overall survival and relapse free survival data were displayed graphically using Kaplan-Meier curves, and differences in outcome analysed by the log-rank assay. Due to small sample size, multivariate analysis of clinical outcome data was not undertaken. The distribution of samples within categories of variables of known significance was analysed by the Fisher's Exact Test (for 2x2 contingency table), or the Freeman-Halton extension of the Fisher's Exact Test (for 3x2 contingency tables).

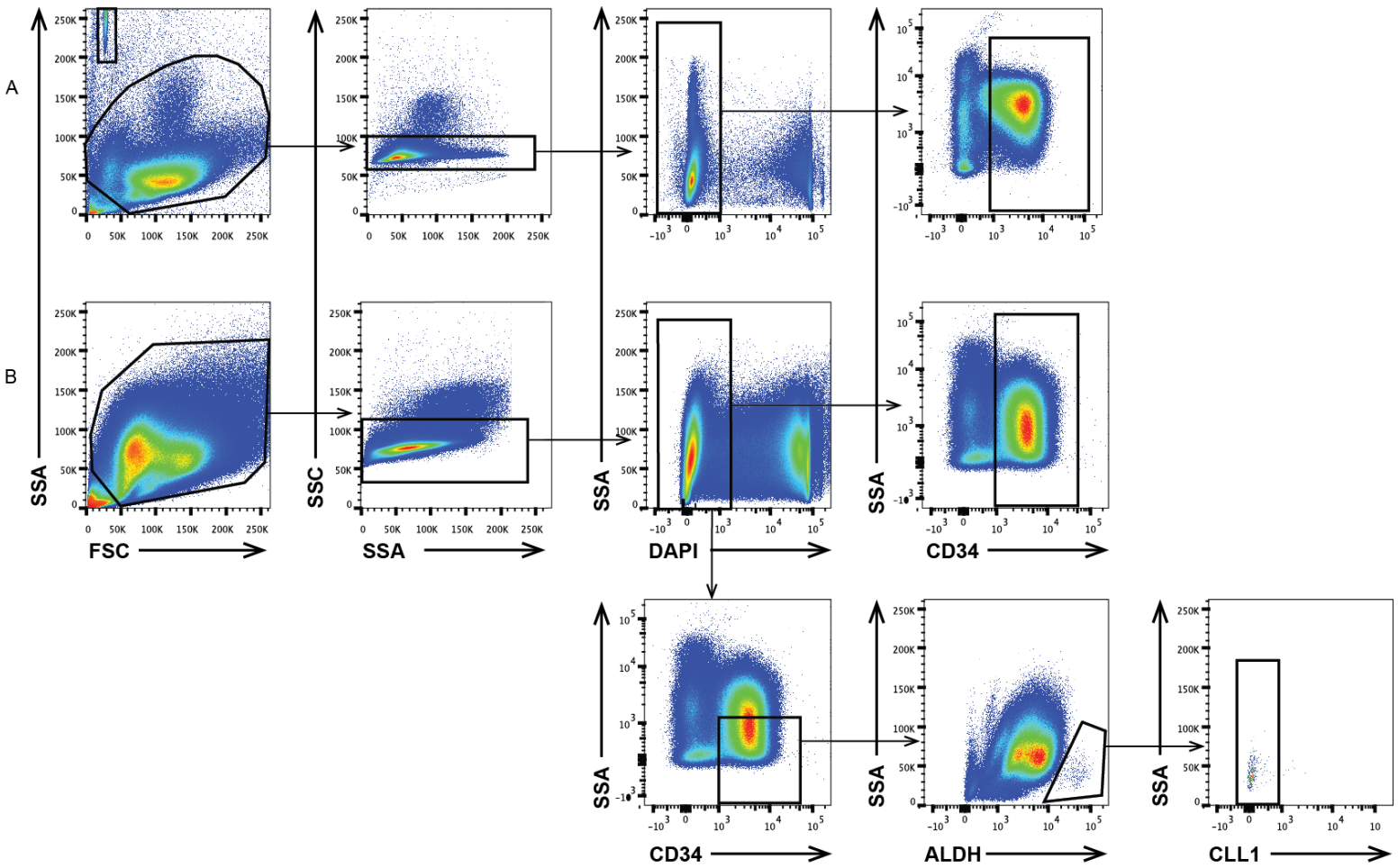

Figure 5-1 illustrating the sequential gating strategy used to analyse fresh and frozen AML samples. Sample shown is ID 10139, an APML BM. Plot A illustrates the gating strategy used on fresh samples to identify the initial [$CD34^{+}$]. CountBeads are visible on the first plot shown; Plot B illustrates the gates used on analysis of frozen samples to identify both the numbers of $CD34^{+}$ cells and $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells in the sample

5.4 Results

5.4.1 Samples processed for fresh BM analysis

Between June 2012 and August 2015, 71 fresh Eppendorf BM samples from patients were received and processed to determine the concentration of $CD34^+$ cells at diagnosis. Subsequent examination of the St Bartholomew's Integrated Diagnostic Pathology Reports revealed that 34 of these patients had been newly diagnosed with AML, and 16 represented true controls (see Table 5-1 and Table 5-2).

A further 60 AML samples and 20 control samples had previously been processed and analysed by Dr F Miraki-Moud between 2007 and 2012 (clinical details not shown).

5.4.1.1 Criteria for selection of samples for further analysis

Interrogation of the tissue bank data base suggested that 73 of these 94 AML patients had a further sample of diagnostic BM available for further evaluation. Of these, 45 were thawed, stained for the panel of markers detailed in Section 5.3.2.1 and analysed in conjunction with the data stored from their fresh samples to calculate the concentration of $CD34^+34^-ALDH^{high}CLL1^-$ cells present at diagnosis. In parallel, 24 control samples were analysed.

Numbers were limited due to the expense and the time consuming nature of these experiments. The AML samples chosen represented the full range of observed cytogenetic risk groups and phenotypic appearances. Therefore sample selection was not random: those AML samples which had an identifiable karyotype or genetic mutation, which would enable assessment of the $CD34^+38^-ALDH^{high}CLL1^-$ cell population for a leukaemia-associated mutation, were preferentially processed.

Table 5-3 and Table 5-4 summarise patient demographics and presenting FBC for the 45 AML and 24 control samples analysed. Table 5-3 also highlights the results for those AML samples where the $CD34^+34^-ALDH^{high}CLL1^-$ population was not just quantified, but also sorted and interrogated for the presence of a leukaemia-associated mutation. Table 5-5 details the treatment and clinical outcome for all 45 AML patients.

5.4.1.2 $CD34^+38^-ALDH^{high}CLL1^-$ cell concentrations for all AML and control samples processed

Table 5-6 summarises the concentration of $CD34^+38^-ALDH^{high}CLL1^-$ cells at diagnosis for the 45 AML samples assayed. Table 5-7 illustrates the same information for the 24 control samples assayed.

Table 5-1 Control bone marrows analysed fresh for [$CD34^+$ cell] between 2012 and 2015

SAMPLE ID		PATIENT DETAILS			PRESENTATION BLOOD COUNT		
TEMP ID	**ID**	**AGE**	**SEX**	**DIAGNOSIS**	**HB (g/dL)**	**WCC ($x10^9$/L)**	**PLTS ($x10^9$/L)**
TEMP2027	10146	62	F	DLBCL	8.0	14.1	661
TEMP1950	10115	50	M	DLBCL	15.0	12.6	294
TEMP2011	10145	28	M	MEDIASTINAL T CELL LYMPHOMA	12.9	16.0	517
TEMP1954	10089	80	M	MALT	11.2	8.1	312
TEMP1935	10114	44	M	DLBCL	10.3	3.5	177
TEMP1862	9974	39	M	DLBCL	11.6	14.9	417
TEMP1831	9979	70	M	FL	15.0	4.9	167
TEMP1823		30	F	HD	11.9	8.5	395
TEMP1821	9968	74	F	DLBCL	12.5	6.0	229
TEMP1735		30	M	REACTIVE	14.6	12.7	252
TEMP1728		75	F	DLBCL	12.4	11.8	341
TEMP1693	9760	64	F	FL	13.9	7.0	360
TEMP1692	9806	54	M	BURKITTS	11.2	7.6	301
TEMP1605		88	M	DLBCL	11.4	8.1	191
TEMP1449		26	F	HD	12.3	12.3	280
TEMP1503	9516	74	M	DLBCL	12.0	10.5	313

Table 5-2 Newly diagnosed AML cases processed fresh for [$CD34^+$ cell] at diagnosis between 2012 and 2014

SAMPLE INFO		PATIENT DETAILS		AML DETAILS				PRESENTATION BLOOD COUNT		
TEMP ID	**ID**	**AGE**	**SEX**	**FAB CLASSIFICATION**	**KARYOTYPE**	***NPM1***	***FLT***	**HB (g/dL)**	**WCC ($x10^9$/L)**	**PLTS ($x10^9$/L)**
TEMP2467	10782	48	M	AML-M0	NORMAL	WT	WT	8.2	2.2	32
TEMP2028	10212	30	F	AML-M3	t(15;17)	WT	MUT	10.7	9.8	52
TEMP2010	10203	67	M	AML-M5b	NORMAL	MUT	MUT	7.8	99.0	27
TEMP2009	10192	57	M	AML-M5b	DEL(11p)	MUT	MUT	7.0	56.0	121
TEMP1987	10147	36	F	AML-M6	COMPLEX INC DEL 5q			5.4	2.3	54
TEMP1982	10032	74	M	MIXED PHENOTYPE AML/ALL	HYPERDIPLOID	WT	WT	8.6	1.3	121
TEMP1949	10139	22	M	AML-M3	t(15;17)	WT	MUT	8.1	10.5	23
TEMP1867	9900	61	F	AML-M6		WT	WT	9.7	1.0	239
TEMP1853	9892	56	M	AML-M5b	NORMAL	WT	MUT	8.3	134.0	30
TEMP1848	9891	61	M	AML-M2	NORMAL	WT	WT	10.5	1.0	60
TEMP1842	9964	24	M	AML-M2	NORMAL	MUT	MUT	13.9	77.0	70
TEMP1810	9963	70	F	AML WITH MDS	COMPLEX INC DEL7q	WT	WT	10.8	5.0	57
TEMP1811	9962	67	F	AML-M4	NORMAL			7.2	178.0	29
TEMP1794	9879	58	M	AML-PLASMABLASTIC SUBTYPE	COMPLEX			9.8	5.2	204
TEMP1754	9850	59	M	AML-M4	NORMAL	MUT	MUT	7.5	49.0	70
TEMP1727	NA	24	F	AML-M4	NORMAL	MUT	WT	8.2	0.9	26
TEMP1724	9878	49	M	AML-M5b	NORMAL	MUT	MUT	11.8	73.0	47

SAMPLE INFO		PATIENT DETAILS		AML DETAILS				PRESENTATION BLOOD COUNT		
TEMP ID	**ID**	**AGE**	**SEX**	**FAB CLASSIFICATION**	**KARYOTYPE**	***NPM1***	***FLT***	**HB (g/dL)**	**WCC (x10^9/L)**	**PLTS (x10^9/L)**
TEMP1689	9786	69	M	AML-Mo	t(2;19), t(8;11)			10.9	0.9	88
TEMP1687	9653	53	M	AML-M4		WT	WT	9.7	28.0	13
TEMP1677	9815	43	M	AML-M4	NORMAL	WT	WT	7.7	3.9	80
TEMP1667	9716	70	M	AML-M4	NORMAL	WT	WT	7.9	13.2	135
TEMP1553	9668	62	F	AML-M1	NORMAL			10.3	17.0	54
TEMP1565	9713	38	M	AML-M0	NORMAL			5.5	19.4	100
TEMP1649	9717	24	F	AML WITH MDS	DEL(7) & ADD(21)			9.5	2.5	38
TEMP1627	9725	73	F	AML WITH MDS	NORMAL			10.4	4.7	CLUMPS
TEMP1637	9718	56	F	AML-M1				5.4	7.5	173
TEMP1619	9672	62	M	SECONDARY AML	DEL(7q)			8.0	37.0	21
TEMP1623	9719	62	M	AML-M2	NORMAL	MUT	MUT	8.9	6.1	123
TEMP1611	9669	42	M	AML-M4 Eo	INV(16)	WT	WT	11.1	12.2	70
TEMP1552	9655	64	M	AML-M0	NORMAL	WT	WT	7.7	0.8	23
TEMP1542	9583	44	M			WT	WT			
TEMP1537	9584	84	F	AML-M5a WITH MDS	DEL(5q), HYPERDIPLOID	WT	WT	5.5	1.9	71
TEMP1457	9467	87	M	AML-M0				9.0	1.0	38
TEMP1510	9554	39	M	AML-M4	NORMAL	WT	MUT	8.6	85.2	60

Table 5-3 Sample details of AML samples quantified for [$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$] cells at diagnosis by co-analysis of fresh and frozen samples

CLINICAL DETAILS			BLOOD COUNT AT DIAGNOSIS			AML GENETIC INFORMATION			$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ CELLS PURITY
ID	AGE	SEX	HB (g/dL)	WCC ($x10^9$/L)	PLTS ($x10^9$/L)	KARYOTYPE	NPM1 MUTATION	OTHER MUTATIONS	
7774	62	M	8.1	0	37	NORMAL	Y-TCTG		PURE
7777	39	M	8.9	15	83	NORMAL	Y-TCTG		MIXED
7793	56	F	11.1	4	10	ADD 13	Y-TCTG		PURE
7820	66	M	12.6	9	52	t(9;22)			
7829	22	M	11.1	5	82	ADD 13; t(3/14)	Y-TCTG		
7882	46	M	9.9	2	69	INV(16)			
7912	72	F	7.3	3	19	NORMAL	Y-TCTG		MIXED
7916	62	F	6.3	38	76	NORMAL	WT		
7955	61	M	9.3	14	19	NORMAL	WT		
7967	67	F	10.0	2	18	NORMAL	Y-CCTG		
7968	51	F	12.0	13	180	ADD 13	Y-TCTG		MIXED
8169	27	F	9.7	0	110	t(11;19)	WT		
8199	45	F	6.6	22	41	NORMAL	Y-TCTG		PURE
8361	38	F	9.5	41	53	t(6;9)	WT	*FLT3*-ITD	PURE
8429	48	M	6.1	3	39	NORMAL	Y-TCTG	*FLT3*-TKD	PURE

CLINICAL DETAILS			BLOOD COUNT AT DIAGNOSIS			AML GENETIC INFORMATION			$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ CELLS PURITY
ID	AGE	SEX	HB (g/dL)	WCC ($x10^9$/L)	PLTS ($x10^9$/L)	KARYOTYPE	NPM1 MUTATION	OTHER MUTATIONS	
8452	27	M	8.6	6	75	t(11;17)	WT	WT	
8458	49	F	11.2	?	35	NORMAL	Y-TCTG	*FLT3*-TKD	PURE
8618	51	M	10.5	19		NORMAL	Y-TCTG		PURE
8751	62	M	10.6	4	57	t(8;21)	WT	WT	
8791	59	M	7.4	31	43	COMPLEX			
8932	43	M	7.2	3	54	t(8;21)	WT	WT	PURE
8993	83	M	6.5	45	25	TRISOMY 11			PURE
9066	59	F	6.8	12	84	DEL 5	WT	WT	
9067	67	M	7.4	142.6	172	NORMAL	Y-CTTG	WT	PURE
9154	63	M	10.3	4	129	COMPLEX	WT	WT	
9583	44	M	8.3	7.6	9	HYPERDIPLOID	WT	WT	
9584	84	F	5.5	1.9	71	DEL5Q;HYPERDIP	WT	WT	
9669	41	M	11.1	12	70	INV(16)	WT	WT	PURE
9672	61	M	6.9	45	20	DEL 7q	WT		
9713	36	M	5.5	19	100	NORMAL	WT	*FLT3*-TKD	PURE
9717	24	F	7.4	1.5	23	-7 & +21			MIXED

CLINICAL DETAILS			BLOOD COUNT AT DIAGNOSIS			AML GENETIC INFORMATION			$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ CELLS PURITY
ID	AGE	SEX	HB (g/dL)	WCC ($x10^9$/L)	PLTS ($x10^9$/L)	KARYOTYPE	NPM1 MUTATION	OTHER MUTATIONS	
9719	61	M	8.9	6	123	NORMAL	Y-TCTG	*FLT3*-ITD	MIXED
9815	42	M	7.7	5	105	NORMAL	WT	WT	
9850	59	M	8.3	56	101	NORMAL	Y-TCTG	*FLT3*-ITD	MIXED
9878	48	M	11.8	73	47	NORMAL	Y-TCTG	*FLT3*-ITD	MIXED
9900	60	F	9.1	1	239	NORMAL	WT	WT	
9962	66	F	7.0	230	40	NORMAL	Y-TCTG		PURE
9963	69	F	10.8	5	57	DIF CLONES INC. DEL 7Q AND 11	WT	WT	PURE
9964	24	M	13.9	77	80	NORMAL	Y-TCTG	*FLT3*-ITD	PURE
10139	22	M	8.1	10.5	23	t(15;17)	WT	*FLT3*-ITD	PURE
10147	36	F	5.4	2.3	54	COMPLEX WITH 5Q			
10192	57	M	7.0	56	121	DEL 11P	Y-TCTG	*FLT3*-ITD	PURE
10203	67	M	7.8	99	27	NORMAL	WT	*FLT3*-ITD & TKD	
10212	30	F	10.7	9.8	52	t(15;17)	WT	*FLT3*-ITD	PURE
10782	48	M	8.2	2.2	32	NORMAL	WT	WT	

Table 5-4 Sample details of control bone marrows studied for [$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$] at diagnosis by co analysis of fresh and frozen samples

PATIENT INFO		CLINICAL DETAILS			PRESENTATION BLOOD COUNT		
TEMP ID	**ID**	**AGE**	**SEX**	**DIAGNOSIS**	**HB (g/dL)**	**WCC ($x10^9$/L)**	**PLTS ($x10^9$/L)**
	7865	41	M	HD	15.5	5.3	265
	7926	42	M	HD	11.2	14.3	527
	7957	63	M	DLBCL	12.5	5.2	295
	7958	22	F	HD	10.5	6.2	479
	8066	27	F	HD	13.5	7.5	316
	8113	53	M	DLBCL	13.2	8.1	432
	8115	42	M	HD	16.5	8.1	318
	8126	61	M	MALT	13.7	6.2	250
	8215	72	F	MALT	12.7	7.2	110
	8245	45	M	HD	12.9	7.8	316
TEMP 0484	8445	62	M	FL/DLBCL	15.1	7	252
TEMP 0487	8510	21	F	T CELL LYMPHOMA	13.1	7.4	587
TEMP 1503	9516	74	M	DLBCL	12.0	10.5	313
TEMP 1693	9760	64	F	FL	13.9	7.0	360
TEMP 1692	9806	54	M	BURKITTS	11.2	7.6	301
TEMP 1605	9903	88	M	GREY ZONE LYMPHOMA	11.4	8.1	191
TEMP 1728	NA	75	F	DLBCL	12.4	11.8	341
TEMP 1735	NA	30	M	REACTIVE	14.6	12.7	252
TEMP 1821	9968	74	F	DLBCL	12.5	6.0	229
TEMP 1823	9975	30	F	HD	11.9	8.5	395
TEMP 1831	9979	70	M	FL	15.0	4.9	71
TEMP 1954	10089	80	M	MALT	11.2	8.1	312
TEMP 1950	10115	50	M	DLBCL	15.0	12.6	294
TEMP 2011	10145	28	M	MEDIASTINAL T CELL LYMPHOMA	12.9	16.0	517

Table 5-5 Treatment details and clinical outcome data for 45 AML patients

PAT ID	1ST LINE THERAPY	RESPONSE TO 1ST LINE THERAPY	DAYS TO NEUTS >1.0	DAYS TO PLTS >100	FURTHER RX	ACHIEVED REMISSION?	DAYS FROM TREATMENT TO REMISSION	RELAPSE?	FURTHER RX	DAYS FROM REMISSION TO RELAPSE	STATUS AT LAST REVIEW	DAYS DIAGNOSIS TO DEATH	DAYS DIAGNOSIS TO LAST SEEN
7774	DA	REFRACTORY			HD ARAC	YES	78	YES		148	DEAD	244	
7777	AML 15-DA	CR	N/A	N/A		YES	N/A	YES	ALLO	N/A	ALIVE		3219
7793	AML 15-DA	DIED IN INDUCTION	-	-	-	-	-	-	-	-	DEAD	22	
7820	GLIVEC	N/A	-	-		NO	-	-	-	-	DEAD	139	
7829	GMALL	CR	-	-	IMMEDIATE ALLO	YES	24	NO	-	-	ALIVE		3000
7882	DA	CR	N/A	N/A		YES	34	YES	ALLO	277	ALIVE		3015
7912	AML 16- DClo	DIED IN INDUCTION	-	-	-	-	-	-	-	-	DEAD	21	
7916	SUPPORTIVE	N/A	-	-	-	-	-	-	-	-	DEAD	2	
7955	DA	CR	N/A	N/A		YES	N/A	NO	-	-	ALIVE		35
7967	AML 16-DClo	CR	N/A	N/A		YES	32	YES	AZA	449	DEAD	584	
7968	AML 15-DA	CR	N/A	N/A		YES	28	YES		170	DEAD	321	
8169	DA	CR	18	26	-	YES	26	YES	ALLO	1466	ALIVE		2756
8199	AML 15-?	CR	31	49		YES	49	NO	-	-	DEAD	105	
8361	DA	CR	23	22	ALLO	YES	23	YES		245	DEAD	358	
8429	DA & GO	CR	23	32		YES	32	YES		368	DEAD	538	
8452	AML 17-ADE & GO	CR	28	38		YES	38	YES	ALLO	385	ALIVE		2485

PAT ID	1ST LINE THERAPY	RESPONSE TO 1ST LINE THERAPY	DAYS TO NEUTS >1.0	DAYS TO PLTS >100	FURTHER RX	ACHIEVED REMISSION?	DAYS FROM TREATMENT TO REMISSION	RELAPSE?	FURTHER RX	DAYS FROM REMISSION TO RELAPSE	STATUS AT LAST REVIEW	DAYS DIAGNOSIS TO DEATH	DAYS DIAGNOSIS TO LAST SEEN
8458	AML 17-ADE and GO	CR	24	26		YES	26	YES		266	DEAD	391	
8618	DA	CR	21	30		YES	30	NO	-	-	ALIVE		35
8751	AML 16-DA & GO	CR	21	32		YES	32	YES		233	DEAD	327	
8791	AZA	REFRACTORY				NO	-	-		-	DEAD	176	
8932	AML 17-ADE & GO	CR	23	22		YES	23	NO	-	-	ALIVE		509
8993	SUPPORTIVE	N/A				NO	-	-		-	DEAD	472	
9066	AML 17-ADE	REFRACTORY			IMMEDIATE ALLO	YES	159	NO	-	-	DEAD	208	
9067	DA	DIED IN INDUCTION	-	-	-	-	-	-	-	-	DEAD	2	
9154	AML 17-DA & GO	REFRACTORY			HD CYTARABINE	YES	66	YES		140	DEAD	389	
9583	DA	CR	17	32		YES	32	YES		1	DEAD	113	
9584	SUPPORTIVE	N/A				NO	-	-		-	DEAD	28	
9669	DA	CR	22	28		YES	28	NO	-	-	ALIVE		1219
9672	AML 17-DA	REFRACTORY			LD ARAC	NO	-	-		-	DEAD	338	
9713	AML 17-DA	CR	28	64	IMMEDIATE ALLO	YES	64	NO	-	-	ALIVE		1211
9717	DA	CR	26	31	IMMEDIATE ALLO	YES	31	YES		553	DEAD	784	
9719	AML 18-ADE & GANETOSPIB	CR	31	27	ALLO	YES	31	NO	-	-	ALIVE		1181

PAT ID	1ST LINE THERAPY	RESPONSE TO 1ST LINE THERAPY	DAYS TO NEUTS >1.0	DAYS TO PLTS >100	FURTHER RX	ACHIEVED REMISSION?	DAYS FROM TREATMENT TO REMISSION	RELAPSE?	FURTHER RX	DAYS FROM REMISSION TO RELAPSE	STATUS AT LAST REVIEW	DAYS DIAGNOSIS TO DEATH	DAYS DIAGNOSIS TO LAST SEEN
9815	AML 17-DA	CR	24	23	ALLO	YES	24	NO	-	-	ALIVE		1148
9850	AML 17-DA	CR	35	35	ALLO	YES	35	NO	-	-	ALIVE		1087
9878	AML 17-DA	CR	22	26	ALLO	YES	26	NO	-	-	ALIVE		1015
9900	DA	CR	29	30	ALLO	YES	30	NO	-	-	ALIVE		929
9962	DA	CR	29	30		YES	30	YES		191	DEAD	239	
9963	RAAVA	REFRACTORY				NO	-	-		-	DEAD	331	
9964	DA	CR	21	28	ALLO	YES	28	YES		178	DEAD	703	
10139	IDARUBICIN & ATRA	CR	36	29		YES	36	NO	-	-	ALIVE		881
10147	DA	CR	30	30	ALLO	YES	30	NO	-	-	DEAD	158	
10192	AML 17-DA	CR	26	25	ALO	YES	26	NO	-	-	ALIVE		798
10203	DA	CR	21	28		YES	28	YES		258	DEAD	333	
10212	IDARUBICIN & ATRA	DIED IN INDUCTION	-	-	-	-		-	-	-	DEAD	24	
10782	DA	CR	24	30	ALLO	YES	30	NO	-	-	ALIVE		224

Table 5-6 showing calculated $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ stem cell concentrations for the 45 AML samples enumerated

PAT ID	FRESH SAMPLE ANALYSIS					FROZEN SAMPLE ANALYSIS		
	BEAD EVENTS	VOLUME (µl)	BEAD CONC (BEADS/50µL)	$CD34^{+}$ CELLS	[$CD34^{+}$ CELLS] (CELLS/µl)	$CD34^{+}$ CELLS	$CD34^{+}CD38^{-}ALDH^{high}CLL1^{-}$ CELLS	ESTIMATED ORIGINAL [$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$] (CELLS/µl)
7774	15573	100	51800	286	**9.5**	118021	993	**0.08**
7777	3511	100	51800	36	**5.3**	205865	512	**0.01**
7793	22204	100	51800	672	**15.7**	2025286	1165	**0.01**
7820	33870	100	51800	379954	**5810.9**	44468	15	**1.96**
7829	20904	200	51800	720810	**8930.8**	96217	31	**2.88**
7882	707	150	51800	89441	**43687.4**	13234278	25089	**82.82**
7912	2931	50	51800	48	**17.0**	3216	90	**0.47**
7916	35242	50	51800	614	**18.0**	23279	5500	**4.26**
7955	26653	200	51800	302919	**2943.6**	1429399	4037	**8.31**
7967	12237	200	51800	15520	**328.5**	192887	1210	**2.06**
7968	16372	200	51800	249912	**3953.5**	679	3	**17.47**
8169	2311	150	54000	2699	**420.4**	1227309	3175	**1.09**
8199	3033	150	54000	268	**31.8**	9510	387	**1.29**
8361	5772	200	54000	169064	**7908.4**	4212	9	**16.90**
8429	1998	200	52500	153	**20.1**	2286	3	**0.03**
8452	7969	200	52500	869	**28.6**	8081	2334	**8.27**
8458	1991	100	52500	1691	**445.9**	286	0	**0.00**
8618	31762	100	52000	1449	**23.7**	1495	5	**0.08**
8751	2130	50	52500	109346	**53903.0**	9371014	2428	**13.97**
8791	2258	150	52000	5195	**797.6**	37955	255	**5.36**
8932	3468	40	51000	184953	**67997.4**	2428178	56	**1.57**
8993	1851	200	52500	35128	**4981.7**	482919	747	**7.71**
9066	15365	100	52500	203200	**6943.1**	1512881	704	**3.23**

PAT ID	FRESH SAMPLE ANALYSIS					FROZEN SAMPLE ANALYSIS		
	BEAD EVENTS	VOLUME (µl)	BEAD CONC (BEADS/50µL)	$CD34^+$ CELLS	[$CD34^+$ CELLS] (CELLS/µl)	$CD34^+$ CELLS	$CD34^+CD38^-ALDH^{high}CLL1^-$ CELLS	ESTIMATED ORIGINAL [$CD34^+38^-ALDH^{high}CLL1^-$] (CELLS/µl)
9067	1821	200	51000	133	**18.6**	1200802	2339	**0.04**
9154	280	200	51000	51375	**46787.9**	4509	1	**10.38**
9583	26699	200	51500	489	**4.7**	15	0	**0.00**
9584	17018	167	51000	65792	**1180.6**	13381	41	**3.62**
9669	9161	200	51000	770534	**21448.1**	4360255	498	**2.45**
9672	10791	200	51000	1508200	**35640.0**	2544238	172	**2.41**
9713	20381	4.8	51000	182079	**94921.2**	44215	301	**646.19**
9717	11231	200	51000	159289	**3616.7**	141420	2956	**75.60**
9719	44341	200	51000	2444	**14.1**	2367	3	**0.02**
9815	3432	200	49500	518138	**37365.7**	889768	1710	**71.81**
9850	3321	200	49000	18604	**1372.5**	88684	537	**8.31**
9878	6293	200	49000	7893	**307.3**	31503	130	**1.27**
9900	18898	200	55000	206445	**3004.1**	98006	217	**6.65**
9962	7231	200	49000	333743	**11307.8**	44193	93	**23.80**
9963	15255	200	49000	1102432	**17705.4**	55689	54	**17.17**
9964	402	200	55000	21733	**14867.1**	254587	115	**6.72**
10139	22003	200	55000	375584	**4694.2**	886806	446	**2.36**
10147	9666	200	55000	20511	**583.5**	293771	1	**0.00**
10192	7920	200	55000	20714	**719.2**	10231	25	**1.76**
10203	12346	200	55000	6601	**147.0**	15001	127	**1.24**
10212	27333	200	55000	459817	**4626.3**	408141	266	**3.02**
10782	20111	200	55000	410501	**5613.2**	365403	162	**2.49**

Table 5-7 showing calculated concentration of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ stem cells for 24 control samples

PAT ID	FRESH SAMPLE ANALYSIS					FROZEN SAMPLE ANALYSIS		
	BEAD EVENTS	VOLUME (µl)	BEAD CONC (BEADS/50µl)	$CD34^{+}$ CELLS	[$CD34^{+}$ CELLS] (CELLS/µl)	$CD34^{+}$ CELLS	$CD34^{+}CD38^{-}ALDH^{high}$ $CLL1^{-}$ CELLS	ESTIMATED ORIGINAL [$CD34^{+}38^{-}$ $ALDH^{high}CLL1^{-}$] (CELLS/µl)
7865	12475	200	51800	19042	**395.3**	1250	12	**3.80**
7926	12919	200	54000	2849	**59.5**	20633	820	**2.37**
7957	32200	200	51800	196	**1.6**	154432	3305	**0.03**
7958	25101	200	51800	10466	**108.0**	10797	32	**0.32**
8066	19251	100	54000	5030	**141.1**	13630	131	**1.36**
8113	18401	100	54000	18532	**543.8**	283737	3372	**6.46**
8115	9850	100	54000	6457	**354.0**	42114	1642	**13.80**
8126	33867	60	54000	7324	**194.6**	755488	33545	**8.64**
8215	22217	50	54000	10980	**533.8**	226807	1278	**3.01**
8245	14811	100	54000	9284	**338.5**	69628	4715	**22.92**
8445	1073	200	52500	1129	**276.2**	27850	851	**8.44**
8510	13423	100	52500	5937	**232.2**	106581	1408	**3.07**
9516	31144	136	51000	535	**6.4**	91244	965	**0.07**
9760	31759	200	49000	310	**2.4**	1483	24	**0.04**
9806	20268	200	49000	6954	**84.1**	7952	55	**0.58**
9903	25143	100	51000	1138	**23.1**	363274	1038	**0.07**
TEMP1728	14119	200	49000	9251	**160.5**	34392	1065	**4.97**
TEMP1735	20068	200	49000	369	**4.5**	13800	33	**0.01**
9968	17591	200	49000	13194	**183.8**	688	9	**2.40**
9975	4148	200	49000	3738	**220.8**	7248	74	**2.25**
9979	24996	200	55000	9837	**108.2**	2766	8	**0.31**
10089	20082	200	55000	31318	**428.9**	5320	82	**6.61**
10115	6213	200	55000	30928	**1368.9**	19995	107	**7.33**
10145	22964	200	55000	2707	**32.4**	923	11	**0.39**

5.4.1.3 *Summary of the clinical characteristics of AML and control patients*

Table 5-8 directly compares the demographics of the patients whose samples were analysed for stem cell concentration at diagnosis.

In 7 of the 45 AML samples where an attempt was made to calculate the stem cell concentration at diagnosis, cytogenetic or PCR based analysis of the sorted population of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells showed them to be a mixture of normal and leukemic cells (see Table 5-3: Pat IDs 7777, 7912, 7968, 9717, 9719, 9850 and 9878). Therefore, as I am unable to estimate the actual normal stem cell concentrations in these samples, they are excluded from further analysis.

Table 5-8 Comparison of characteristics of the patients within the AML and control cohorts

SAMPLE TYPE		**CONTROL**	**AML- ALL**	**AML- EXCLUDING THOSE KNOWN TO SORT MIXED POPULATIONS**
SAMPLE NUMBER		24	45	38
AGE	**MEDIAN**	54	51	54
	RANGE	22-88	22-84	22-84
SEX	**MALE**	67% (16)	62% (28)	63% (24)
	FEMALE	33% (8)	38% (17)	37% (14)
CYTOGENETIC RISK	**GOOD**	-	13% (6)	16% (6)
	INTERMEDIATE	-	62% (28)	58% (22)
	POOR	-	24% (11)	26% (10)

5.4.2 Summary of treatment outcomes for AML patients

5.4.2.1 Induction therapy

Of the 38 AML patients analysed further, 32 (84%) underwent initial intensive induction therapy (the vast majority enrolled into the relevant MRC AML trial). 3/38 (8%) received low dose chemotherapy with palliative intent, and 3/38 (8%) received supportive therapy with blood products.

5.4.2.2 Overall survival

Kaplan-Meier curves showing the effect of cytogenetic risk group and age on overall survival in the cohort as a whole are shown in Figure 5-2 A and B respectively. The significant effect of both factors fits expected patterns, and suggests this is a representative dataset.

5.4.2.3 Remission Induction

Of the 32 patients who received high dose induction therapy with curative intent, 25/32 (78%) achieved CR after Cycle 1. 4/32 (13%) had refractory disease, and 3/32 (9%) died within 30 days of diagnosis. 3/4 (75%) of those with initially refractory disease achieved remission with salvage chemotherapy and/or allogeneic transplantation.

5.4.2.4 Relapse free survival

Of the 28 patients who did achieve remission, clinical data were obtainable for 26/28 (93%) to permit the calculation of relapse free survival. Figure 5-2 C and D show the effects of cytogenetics and age on relapse-free survival across the whole cohort. As with the data pertaining to overall survival, these figures suggest this cohort is a representative sample.

5.4.2.5 Time to Count Recovery after Induction chemotherapy

Of the 25 patients who achieved remission after initial induction therapy, data were obtainable for 21/25 (84%), to allow the calculation of time to count recovery after initiation of therapy. The median time for neutrophil recovery (using the MRC trial definition of neutrophils greater than $1.0x10^9$/L) was 24 days (range 17 to 36 days). The median time for platelet recovery (using the MRC criteria of platelets greater than $100x10^9$/L) was 30 days (range 22 to 64 days).

5.4.3 Comparison of normal stem cell concentrations between AML and control samples

Calculated HSC (as defined by $CD34^+38^-ALDH^{high}CLL1^-$ cells) concentrations at diagnosis are shown in Table 5-6 for all individual AML samples and Table 5-7 for controls.

5.4.3.1 Stem cell numbers appear to be preserved between AML samples and controls

Figure 5-3 A is a bar chart comparing stem cell concentrations at diagnosis between AML samples and controls. There is no statistically significant difference between the concentration of stem cells within control samples (n=24; median 2.4 cells/µl), and the diagnostic AML samples (n=45; median 2.5 cells/µl). When the AML samples are subclassified to either exclude those samples where the $CD34^+38^-ALDH^{high}CLL1^-$ population contains the leukaemia-associated mutation (“dirty” sort)(n=38; median 2.7 cells/µl) or only include those samples where the $CD34^+38^-ALDH^{high}CLL1^-$ population is known to be free of the AML associated mutation (“proven pure” sort) (n=17; median 2.4 cells/µl), there are still no significant differences to be found between any AML subpopulation and the control group.

5.4.4 Subanalysis of AML samples

5.4.4.1 There is no association between stem cell number at diagnosis and the age of the patient, or cytogenetic karyotype of the AML

The 38 AML samples were further analysed for factors which we hypothesised might have an effect on stem cell concentration at diagnosis. There was no significant difference observed between the stem cell concentration and patient sex (male, n=24, median $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ concentration 2.5 cells/µl; female, n=14, median $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ concentration 3.1 cells/µl), as illustrated in Figure 5-3 B.

Age does not appear to have a significant effect on stem cell concentration this cohort. This is the case when age is defined as a categorical variable (see Figure 5-3 C) (<60, n= 23, median 2.5 cells/µl; >60, n=15, median 4.3 cells/µl, p=0.23 by Mann Whitney t-test), as well as when expressed as a quantitative variable (non-significant Spearman rank coefficient of 0.17 (p=0.33)).

Finally, there is also no significant association between $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ concentration at diagnosis and karyotype (good risk, n=6, median 2.7 cells/µl; intermediate risk, n=22, median 2.3 cells/µl; poor risk, n=10, median 3.4 cells/µl) (Figure 5-3 D).

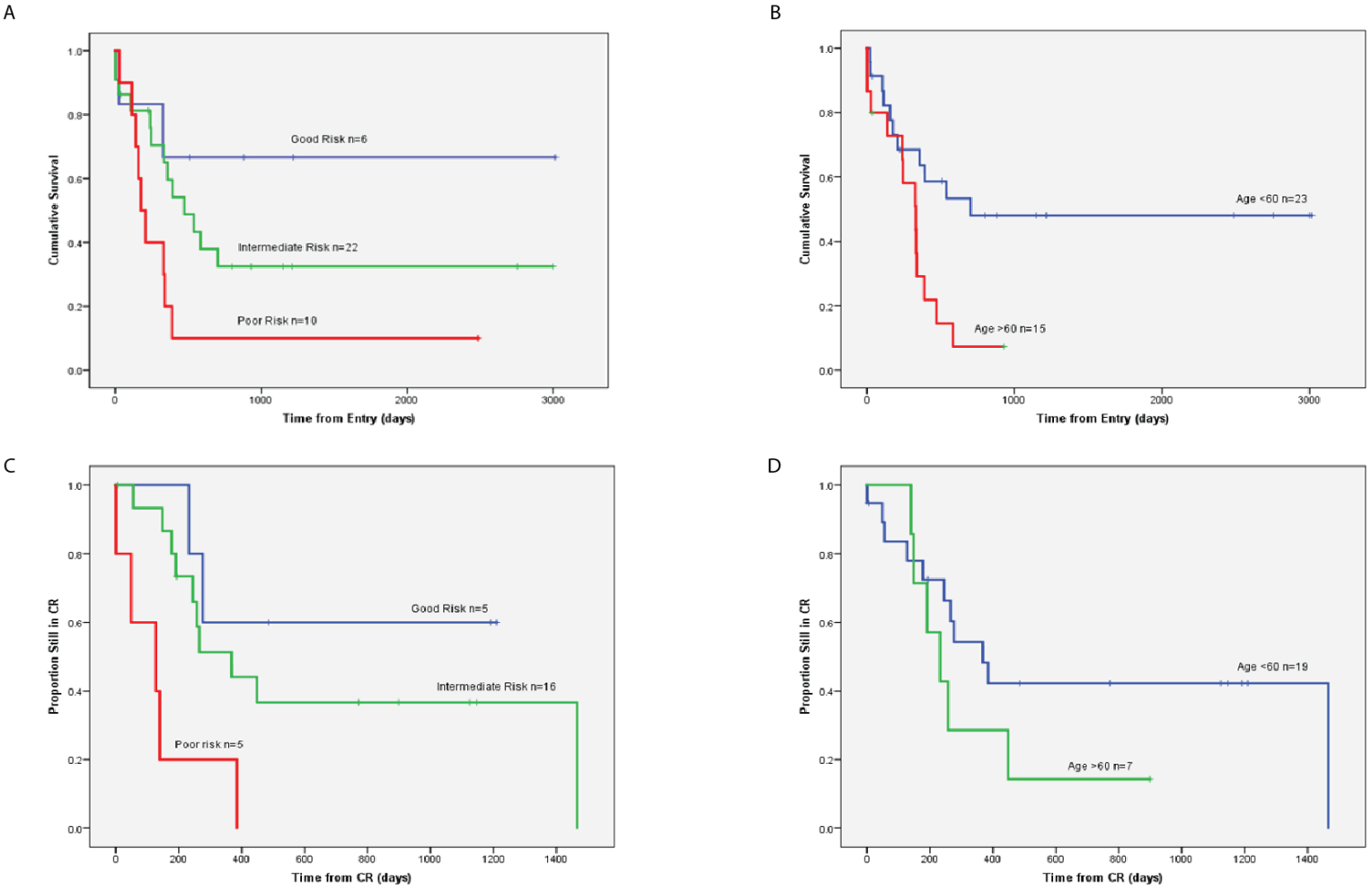

Figure 5-2 Kaplan-Meier plots showing the effects of A: cytogenetics on overall survival; B: age on overall survival; C: cytogenetics on relapse free survival; D: age on relapse free survival

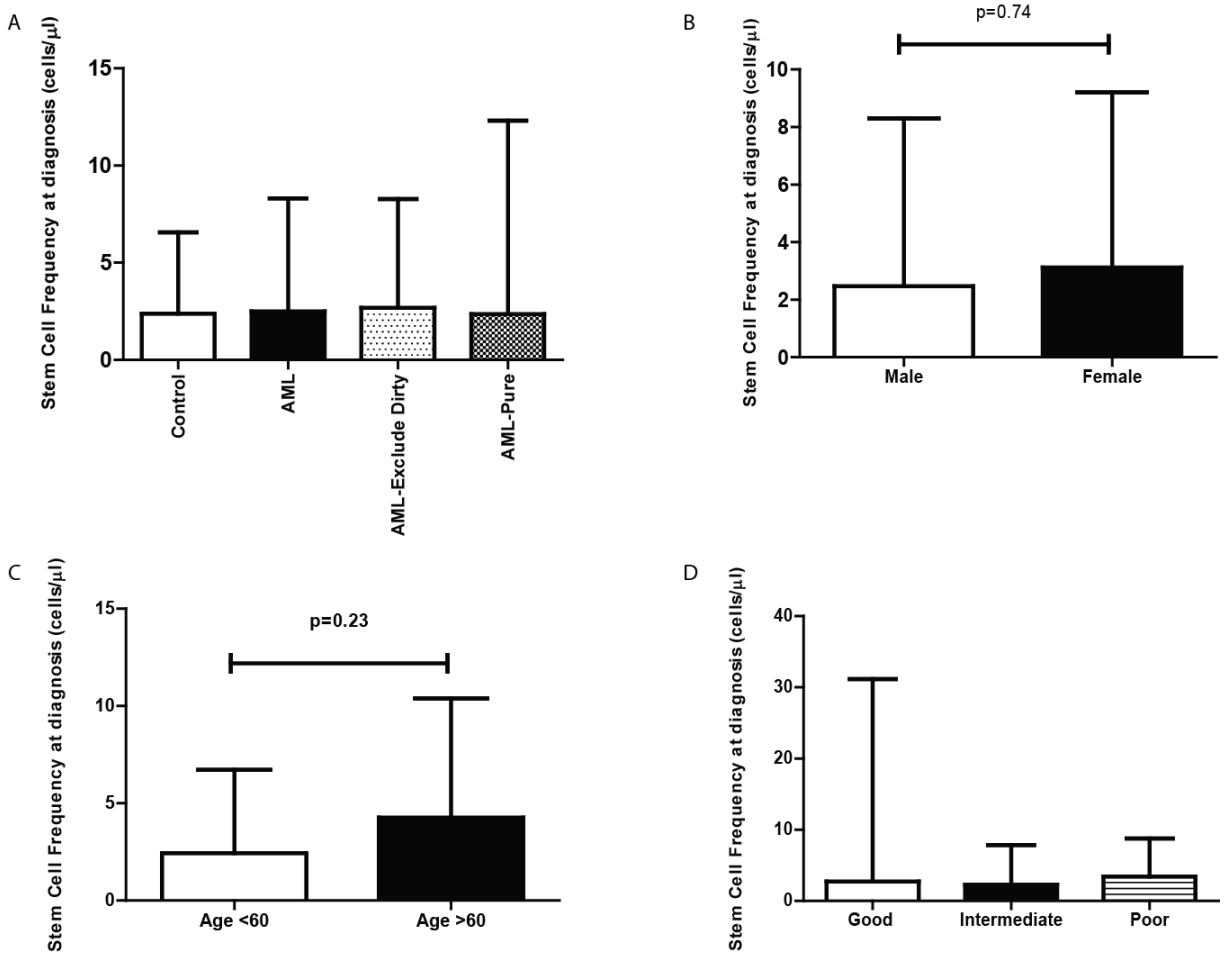

Figure 5-3 A: showing calculated $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ concentrations in control samples; all AML samples; AML samples excluding 7 samples where the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ population was known to contain leukaemia cells, and AML samples where analysis has shown the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ population to be free of leukaemia-associated mutations. Medians and IQRs are displayed; B: Bar chart comparing stem cell concentrations of 38 AML samples between male and female patients. Medians and IQRs are displayed. Mann-Whitney t-test shows no significant difference between male and female groups; C: Bar chart showing the relationship between stem cell concentration of 38 AML samples and age. Median and IQRs are displayed; D: Bar chart showing the relationship between stem cell concentration of 38 AML samples and cytogenetic risk group. Medians and IQRs are shown. There are no significant differences between groups by Kruskal-Wallis testing.

5.4.4.2 Effect of stem cell concentration on clinical outcome

5.4.4.2.1 Low stem cell concentration at diagnosis is significantly associated with reduced overall survival

There is a significant range of concentrations of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells at diagnosis within the 38 samples studied, ranging from 0 to 646 cells/µl, as shown in Table 5-6. There appears to be a cluster of samples (n=8) with a low stem cell concentration (ranging 0 to 0.1 cells/µl), with the remaining 30 samples having concentrations which range from 1 cell/µl upwards (with the majority 1 to 10 cells/µl). For the purpose of further analysis, stem cell concentration has been expressed as a categorical variable (low stem cells= $[CD34^{+}38^{-}ALDH^{high}CLL1^{-}]$ <0.1 cells/µl; high stem cells= $[CD34^{+}38^{-}ALDH^{high}CLL1^{-}]$ >0.1 cells/µl).

Figure 5-4 A is a Kaplan-Meier survival plot illustrating a significant difference in overall survival between those with low and high stem cells, with those patients with less than 0.1 cells/µl at diagnosis having a significantly lower overall survival (log rank test, p=0.01) than those with higher levels at diagnosis.

In order to validate this cut-point between the two groups, it would be appropriate to use the same parameters to analyse a second, independent sample cohort.

This cohort of samples is too small to allow for multivariate analysis to determine if stem cell number at diagnosis is a truly independent variable from the other factors shown to have a significant effect on overall survival (age and cytogenetic analysis). In an attempt to examine the data for potential confounding factors, the observed distribution of samples between age and stem cell concentration groups was analysed by Fisher's Exact Test (see Table 5-9). Distribution of samples within these categories was not statistically different to that which was observed across the whole cohort (2 sided Fisher's Exact Test, p=0.440).

Table 5-9 Cross tabulation for comparison of sample distribution within diagnostic HSC concentration and age categories

		AGE CATEGORY		
		<60	**>60**	**TOTAL**
[STEM CELL]	**<0.1 CELLS/µl**	6	2	8
	>0.1 CELLS/µl	17	13	30
	TOTAL	23	15	38

A similar method was employed to investigate the observed distribution of samples between cytogenetic risk and HSC concentration groups. A 2 by 3 contingency table was constructed

(see Table 5-10), and analysis was performed using the Freeman-Halton extension of the Fisher's Exact Test (2 sided probability, p=0.47).

Table 5-10 Cross tabulation for comparison of sample distribution within diagnostic HSC concentration and Cytogenetic Risk categories

		CYTOGENETIC RISK CATEGORY			
		GOOD	**INTER**	**POOR**	**TOTAL**
[STEM CELL]	**<0.1 CELLS/µl**	**0**	6	2	8
	>0.1 CELLS/µl	**6**	16	8	30
	TOTAL	**6**	22	10	38

This provisional analysis (limited due to the size of the sample, and the number of factors known to be potential confounders) would suggest that the distribution of samples within the low and high stem cell groups within the age and cytogenetic stratification categories reflects that seen in the cohort as a whole. This is highly suggestive that stem cell concentration is an independent variable impacting overall survival.

5.4.4.2.2 There is no associated between low stem cell number at diagnosis and increased 30 day mortality

When all 38 AML patients were considered together (irrespective of initial therapy), low stem cell concentration at diagnosis is associated with a worse outcome in terms of an increased 30-day mortality. Of those 8 with a low HSC concentration at diagnosis (<0.1 cells/µl), the 30 day mortality rate was 25% (2/8), as opposed to 10% (3/30) for those with high HSC concentration at diagnosis (>0.1 cells/µl). However, this effect does not reach statistical significance (Fisher's Exact Test, p=0.246).

This association is more pronounced when only the 32 patients deemed fit to receive intensive induction therapy, are included in analysis. The 30 day mortality for those with low stem cells remains 25% (2/8), as opposed to 4% (1/24) for the high stem cell concentration cohort. Again, however this effect does not reach statistical significance (Fisher's Exact Test, p=0.139).

5.4.4.2.3 There is no association between stem cell number at diagnosis and leukaemia resistance to induction chemotherapy

Only 4 of the 32 patients who received intensive induction therapy were refractory to initial treatment. 1/8 (13%) of those with a low stem cell concentration at diagnosis had refractory disease; 3/24 (also 13%) of the high stem cell concentration cohort responded similarly.

5.4.4.2.4 Low stem cell number at diagnosis is significantly associated with reduced relapse free survival.

Of the 32 patients who received intensive chemotherapy at diagnosis, 25 entered an immediate remission. 3 of the 4 patients with initially refractory disease obtained remission after further salvage chemotherapy. Of this group of 28 patients, data were obtainable on relapse free survival for 26. Figure 5-4 B is a Kaplan-Meier plot illustrating the differential relapse free survival for the low stem cell (n=6) and high stem cell (n=20) cohorts. Relapse free survival was significantly lower in the cohort with low stem cell numbers at diagnosis (p=0.02, log rank test).

As discussed in Section 5.4.4.2.1, this sample group is too small to facilitate multivariate analysis. As before, a Fisher's Exact Test was used to analyse the observed distribution of the 26 samples between age and stem cell concentration groups (see Table 5-11). Distribution of samples was not different to that observed across the whole cohort (p=1.00).

Table 5-11 Cross tabulation for comparison of sample distribution between HSC concentration and age categories

		AGE CATEGORY		
		<60	**>60**	**TOTAL**
[STEM CELL]	**<0.1 CELLS/µl**	5	1	6
	>0.1 CELLS/µl	14	6	20
	TOTAL	19	7	26

Table 5-12 Cross tabulation for comparison of sample distribution between HSC concentration and cytogenetic risk categories

		CYTOGENETIC RISK CATEGORY			
		GOOD	**INTER**	**POOR**	**TOTAL**
[STEM CELL]	**<0.1 CELLS/µl**	0	4	2	6
	>0.1 CELLS/µl	5	12	3	20
	TOTAL	5	16	5	26

A similar method was employed to investigate the observed distribution of samples within stem cell concentration and cytogenetic risk groups. A 2 by 3 contingency table was constructed (see Table 5-12), and analysis was performed using the Freeman-Halton extension of the Fisher's Exact Test (2 sided probability, p=0.369).

As with overall survival, this analysis (albeit even more limited due to the smaller sample numbers) suggests that stem cell concentration at diagnosis is an independent variable predicative of relapse-free survival.

5.4.4.2.5 There is no significant association between stem cell concentration at diagnosis and time to count recovery

Of the 38 AML patients studied, 32 received intensive chemotherapy with a realistic aim of inducing remission. Of these, 25 entered CR after Cycle 1, and of these, serial blood count data were available for 21/25.

Using the same categorisation of samples into low and high stem cells at diagnosis, there appears to be no association between stem cell number and time to neutrophil recovery (using the MRC criteria of neutrophil count greater than $1.0x10^9$/L) from the start of induction therapy (low stem cells, n=5, median days to neutrophil recovery 23; high stem cells, n=16, median days to neutrophil recovery 24; p=0.45 by Mann Whitney t-test) (Figure 5-4 C). Similarly, there is no association between stem cell number and time to platelet recovery (using the MRC criteria of platelets greater than $100x10^9$/L) from the start of induction (low stem cells, median days to platelet recovery 30; high stem cells, median days to platelet recovery 28.5; p=0.40 by Mann Whitney t-test) (Figure 5-4 D).

It is worth noting time to count recovery is a highly complex variable. Within this cohort of patients, a variety of induction chemotherapy regimens were used (details are listed for individual patients in Table 5-5). These have been shown to vary in the degree of myelosuppression induced. In addition, those patients who developed severe sepsis may well have prolonged time to count recovery. G-CSF use to promote neutrophil recovery is now widespread in clinical practise, but was not subject to standardised use during the whole of the period these patients were treated. In a cohort of this size, it is not possible to control for all of these variables.

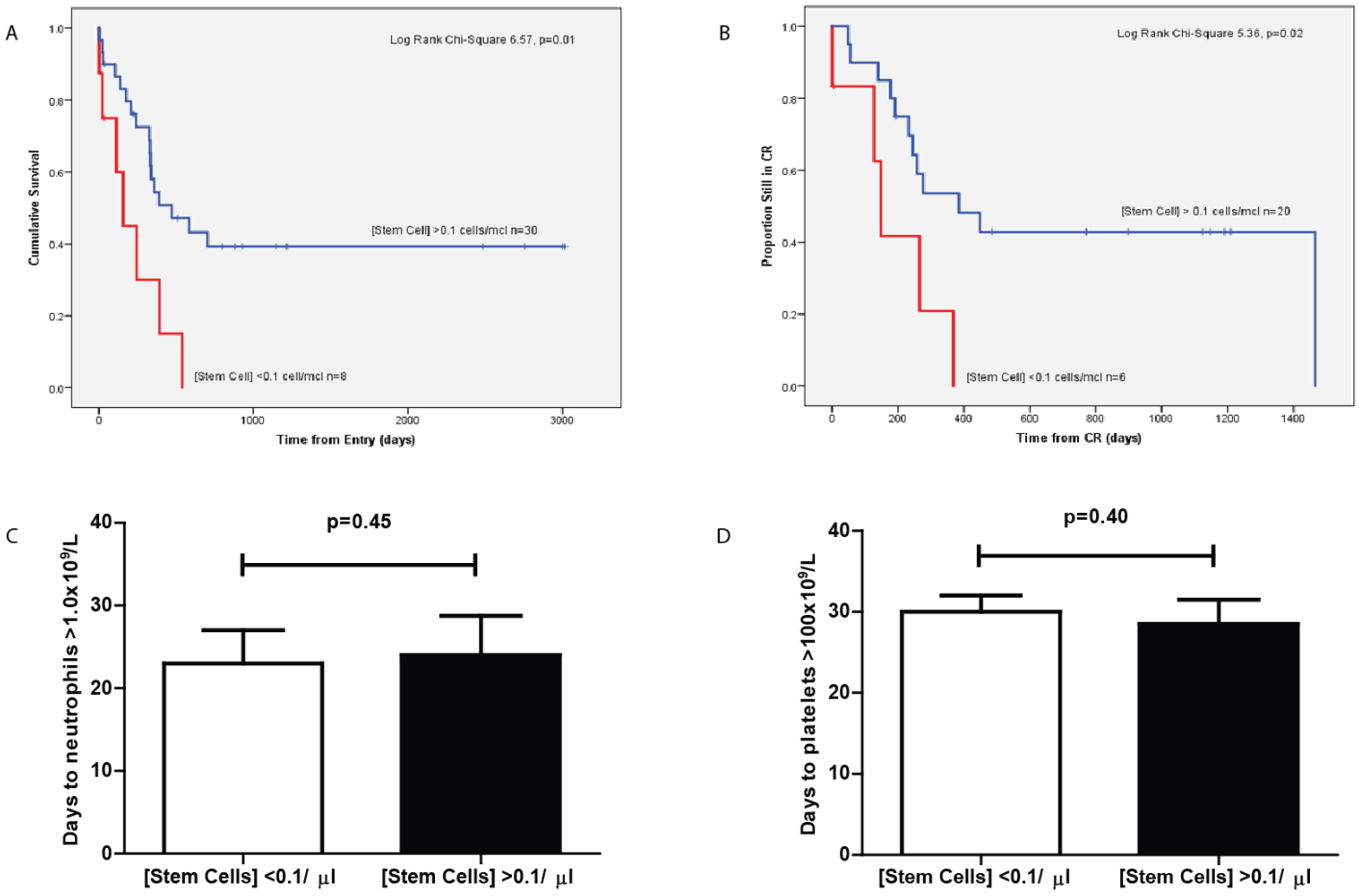

Figure 5-4 A: Kaplan-Meier Plot showing effect of HSC concentration at diagnosis on overall survival; B: Kaplan-Meier plot showing effect of HSC concentration on relapse free survival; C: Bar chart showing association between HSC number at diagnosis and days from induction start to neutrophil recovery. Samples are divided into those with [stem cell] < 0.1 cells/µl (n=5) and [stem cell]> 0.1 cells/µl (n=16). Median and interquartile range for each group are shown; D: Bar chart showing association between HSC number at diagnosis and days from induction start to platelet recovery. Samples are divided into those with [stem cell]<0.1 cells/µl (n=5) and [stem cell]>0.1 cells/µl (n=16). Median and interquartile range for each group are shown

5.4.5 Control Sample Subanalysis

One of the most striking and unexpected findings from this analysis is the considerable variation in the concentration of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells present at diagnosis within the control cohort. In these 24 supposedly normal samples, there is an observed three log fold variation in stem cell concentration, from 0.01 to 23 HSCs/µl (as shown in Table 5-7).

Further analysis was performed to investigate factors that potentially could contribute to variation in stem cell concentration at diagnosis.

5.4.5.1 There is no significant effect of age or gender on stem cell concentration within the control cohort.

There is no significant difference between the stem cell concentration for men and women (men, n=16, median 3.18 cells/µl; women, n=8, median 2.3 cells/µl) (see Figure 5-5 A). There is also no significant effect of age on stem cell concentration. This appears to be the case when age is defined as a categorical variable (see Figure 5-5 B (age<60, n= 13, median 2.4 cells/µl; age>60, n=11, median 2.4 cells/µl)), as well as when expressed as a quantitative variable (non significant Spearman rank coefficient of -0.03 (p=0.89)).

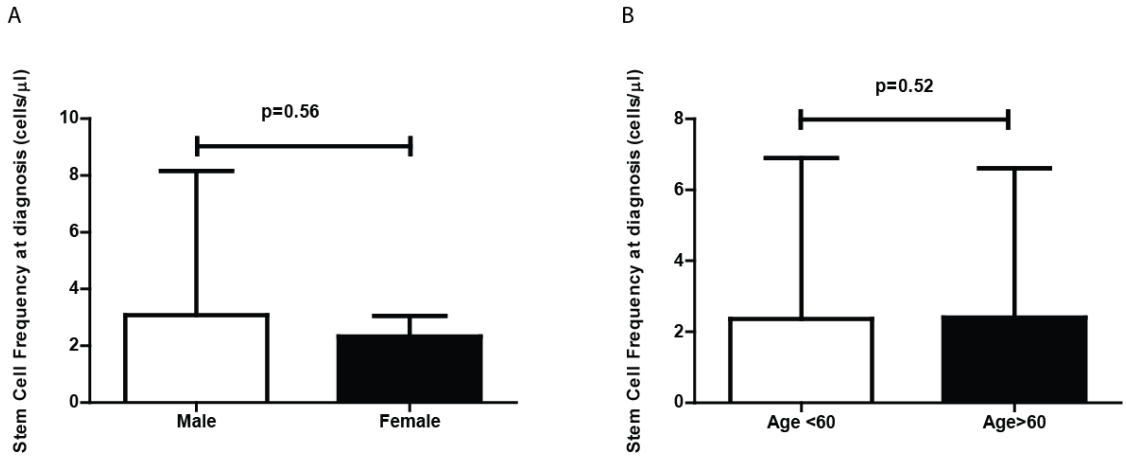

Figure 5-5 Bar charts showing the effect on HSC concentration in control patients of A: gender and B: age. Medians and interquartile range are plotted. Assessment of statistical significance between groups was performed by Mann-Whitney t-test

5.5 Discussion

5.5.1 Potential Clinical Applications

This study of a cohort of AML patients which represents a range of cytogenetic and genetic abnormalities, appears to confirm the earlier finding of our group in a more selected patient population, that normal HSC numbers are preserved at diagnosis[1].

In addition, calculation of HSC concentration at diagnosis with AML may have prognostic benefit for newly diagnosed patients. Analysis of 38 AML patients revealed this population appeared to separate into two biological groups- one cohort with a low $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cell concentration at diagnosis (<0.1 cells/µl), and one with a higher concentrations (>0.1 cell/µl). Comparative analysis of these two groups revealed that low stem cell concentration at diagnosis was significantly associated with reduced overall survival and relapse free survival. There was a non-significant trend toward increased 30 day mortality within the patients with a low stem cell concentration at diagnosis. There was no significant association found between the stem cell concentration at diagnosis, and induction failure due to refractory disease, or time to count recovery from the start of induction.

Ideally, these findings would be validated in a larger, independent cohort. However, if further study did corroborate the findings, how might this information be put to clinical use, most importantly to improve patient outcomes? It is difficult to see how one might improve the survival to 30 days with standard therapy, where patients are normally closely monitored, sepsis treated aggressively and G-CSF support is now used widely. However, this cohort also displayed a reduced relapse-free survival: therefore knowledge of stem cell number at diagnosis might potentially be a factor for influencing the decision to transplant in first remission; or indeed selecting an induction regimen such as FLAG-IDA associated with a reduced relapse risk. Such information might be particularly relevant in stratifying risk and treatment decisions in those patients with intermediate risk disease by cytogenetics. Indeed, it is this intermediate risk cohort of patients where the association between initial stem cell concentration and overall survival appears the most marked (see Figure 5-6).

However, there are certain limitations with this dataset, which restrict the strength of any conclusions drawn from its analysis. As discussed at length in Chapter 4, the panel of antibodies selected to identify and separate normal stem cells from AML cells does not work for every AML sample. The data illustrated in that chapter suggest that in 71% (17/24) of AML samples, the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ fraction retains stem cell functionality and is free of

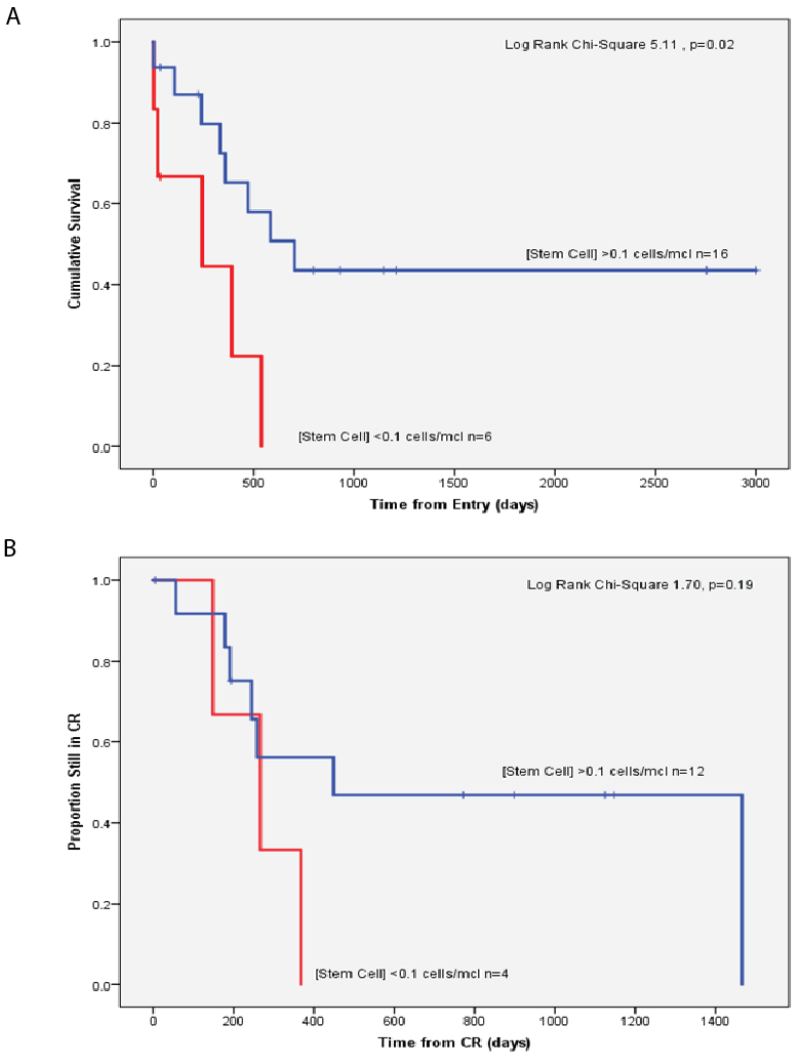

Figure 5-6 Kaplan-Meier plots showing the effect of stem cell concentration for those patients with intermediate risk cytogenetics on A: overall survival and B: relapse free survival

the AML-associated mutation. In a further 25% (6/24) of samples, the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ fraction is "enriched' for the presence of normal cells (less than 25% of all cells contain the leukaemia-associated mutation). Therefore, accurate enumeration of normal HSC populations using this methodology is not possible for all samples.

Of my initial cohort of 45 patients with enumerated cells, analysis of the genetics of the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells was attempted in 35 (78%) of the samples. 5/45 (11%) had no karyotypic or identified genetic abnormality which would facilitate separation of normal cells from AML; and the remaining 5 (11%) did have karyotypic abnormalities, but the sorted cells were used for other purposes (such as RNA extraction) which meant immediate cytogenetic analysis was not possible.

Seven of those AML samples where the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations were successfully analysed were shown to be a mixture of AML cells and normal cells, although these cells were

clearly enriched for residual normal cells, as the proportion of AML to normal cells was below 25% in all but one case. Nevertheless, all of these samples were excluded from further analysis in this chapter. The remaining analysed cohort of 38 samples is therefore comprised of samples where the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ population has been shown to be clear of the AML-associated mutation (n=17), but also includes those samples where it was impossible to determine the origin of cells, due to a lack of identifiable leukaemia-associated mutation, as well as those samples where purity analysis was technically unsuccessful.

It is interesting that when these different groups are compared (see Figure 5-3 A), there are no significant differences between their median stem cell concentrations. This is reassuring, in that the inadvertent inclusion of a "dirty" stem cell sample (for example one of the normal karyotype AMLs, with no *NPM1* or *FLT3* mutation) is unlikely to have a significant impact on the overall findings.

I chose to do the bulk of analysis using the 38 AML samples which represented the entire cohort minus the seven samples known to sort dirty (rather than only the 17 samples known to sort pure) for two reasons. Firstly, when attempting to draw correlations between stem cell numbers and clinical outcomes, the larger the cohort, the more useful any conclusions will be. Secondarily, there is a clear bias within the pure cohort to patients with good risk disease (as these have easily identifiable cytogenetic abnormalities) and therefore clinical correlations within this sample group would be limited, even if it were a larger size.

5.5.2 Variation of stem cell concentrations

One of the interesting and somewhat unexpected findings of this work was the substantial variation in stem cell concentrations observed within the control cohort of samples, which spanned three log folds between individuals. Why might this be? Is this a real finding, or a reflection of the limitations of the experimental method employed? Some potential explanations are discussed below.

5.5.2.1 Potential biological causes of variability in stem cell numbers within the control cohort

5.5.2.1.1 Variation between individuals

Experimental information on absolute stem cell number in the bone marrow in humans is surprisingly limited. No previous papers exist where attempts have been made to quantify calculate the HSC concentration in fresh BM samples as we have attempted here. We know that the process of normal haematopoiesis is an extremely tightly controlled homeostatic

pathway, and between healthy individuals, the numbers of mature blood cells (erythrocytes, platelets, neutrophils etc.) are kept within relatively narrow normal distributions. One might therefore extrapolate that the same should be true of the number of HSCs, but this does not appear to be the case from the data shown above.

Alternatively, it is possible that individuals do have widely varying numbers of stem cells maintained throughout life, and yet due to redundancy within the system, still retain the ability to maintain a normal blood count.

In in-bred mice, several quantitative trait loci have been identified which can control stem cell numbers, suggesting a genetic link between stem cell numbers and specific genes[237]. It is possible that genetic factors account for similar variation within a healthy human population.

5.5.2.1.2 Age

One might assume the number of functional HSCs in the bone marrow might decrease with age for the following reasons. Clinically, we observe that haematopoiesis becomes restricted to smaller areas of the bone marrow with increasing age. We know that with age comes reduced bone marrow cellularity[238]. We know age is associated with increased incidence of diseases associated with dysfunctional haematopoiesis, such as anaemia and MDS[239]. Younger donors for bone marrow transplant procedures will be preferentially selected over older donors[240], although the reasons for this are complex and not solely related to an increased chance of successful engraftment.

Pang et al published a study looking at the effect of age on human HSC populations. HSCs (identified as a $Lin^-CD34^+38^-90^+45RA^-$ population) were quantified as a proportion of the $CD34^+$ cells (as opposed to all cells) within previously frozen BM samples. Samples from young (aged 20 to 35) healthy donors (n=13), and older (age >65) healthy donors (n=11) were compared. Flow characterisation suggested the ratio of $Lin^-CD34^+38^-90^+45RA^-$ to $CD34^+$ cells increased with age, as the proportion of quiescent HSCs fell. In vivo xenograft culture experiments suggested that the older stem cells were myeloid skewed, in terms of the progeny they generated[241]. Whilst it is easy to question whether expressing HSCs as a proportion of total $CD34^+$ cells (which may themselves vary in number with age) is a valid approach, these findings have also been mirrored in murine studies[242].

In our similar sized cohort of patients, we found no statistical association between age and stem cell number within the control cohort.

It is worth remembering that the actual functional volume of bone marrow which is able to perform haematopoiesis is increasingly restricted with age. Therefore, irrespective of the absolute stem cell concentration within the areas sampled, the actual overall number of stem cells present in the body must therefore fall with age, even if their concentration with active haematopoietic areas remains constant.

5.5.2.1.3 Do stem cell numbers vary over time within a healthy individual?

No studies exist, in either animal models or human subjects, which track the number of HSCs in an individual's bone marrow over time. One assumes their numbers are under tight homeostatic regulation, whilst being able to respond to the challenge of illness, blood loss, pregnancy etc. which might require the increased production of blood cells.

Within healthy individuals, it is interesting to hypothesise that stem cell numbers might fluctuate even on a more frequent basis. A number of studies of both human and mice progenitor cell production suggests this process follows circadian rhythms, both in terms of production and release from the BM[243]. Clinically, we use this knowledge to try and optimise stem cell collections from BM donors by morning apheresis procedures. Although the same patterns may well not be observed for the more immature HSC fraction of cells, the timing of extraction of bone marrow samples (whilst for the vast majority being between 9am and 5pm) was not standardised.

5.5.2.1.4 Are these samples really controls?

An important consideration, and a possible cause for the variation in HSC concentration seen within the control cohort, is that these patients are not "normal". As stated in Section 5.3.1.3, control samples were only included in further analysis if they came from patients with Stage I to III lymphoma (the vast majority I and II), with no prior treatment. Examination of both aspirate and trephine by morphology, flow, cytogenetics and immunohistochemistry had to exclude any suggestion of infiltration; and blood counts at the time of biopsy had to fall within the normal range.

However, these remain patients recently diagnosed with cancer, and therefore their bone marrow might reflect changes indicative of a reactive process.

We do not have access at Bart's Hospital to bone marrow samples from entirely healthy volunteers, although to answer this question fully, it would be appropriate to consider applying for ethical and financial approval to recruit volunteers for sample collection. This would clearly have the advantage of being able to standardise the procedure (see Section

5.5.2.2). Ideally, bone marrows would be taken by an appropriately trained single operator, at a standardised time of day. This would also enable a reduction in the time to sample processing within the laboratory.

5.5.2.2 Sources of Potential Experimentally introduced variation

5.5.2.2.1 Bone marrow extraction

All samples were taken and processed according to an SOP set up within the Haematology Department at Bart's Hospital (see Sections 5.3.1.1 and 5.3.1.2). The close proximity between hospital and research laboratory, as well as the excellent service provided by the Tissue Bank at Bart's meant this process is as well standardised as possible for collection of clinical samples. However, there are clearly some potential variables within this process. These are discussed in detail below.

5.5.2.2.1.1 Operator Variation

Any clinician with experience of taking bone marrow samples knows that two different operators can take widely different quality of samples on the same patient at the same time dependent on a combination of experience and luck. Haemodilute, aparticulate draws of bone marrow can occur, even with the most experienced operators, and one would imagine that such a sample would generate a much lower stem cell concentration on analysis.

Such haemodilute draws can be attributable not just to the operator, but also to the biological state of the bone marrow (such as myelofibrosis or infiltration with a non-haematological malignancy).

5.5.2.2.1.2 Location of stem cells within the bone marrow

All samples processed were aspirates taken from the right or left iliac crest (as opposed to the sternum). However, we believe that that stem cells are not distributed uniformly throughout the bone marrow, and therefore potentially small differences in the position of the aspiration needle might significantly affect the concentration of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells within the material obtained.

5.5.2.2.1.3 Order of aspirate draw

The SOP requests that on taking a bone marrow sample for research, the very first 200µl drawn is immediately placed into an eppendorf tube for further analysis. Subsequent samples for morphological analysis, flow, cytogenetic, molecular genetics and further research samples should then be obtained from the same needle position. However, one can imagine that busy doctors, who should prioritise the diagnostic value of a bone marrow procedure to the patient,

might use the first pull for clinical purposes, and only use subsequent marrow for research processes.

Clearly, the order that samples are taken is vital, as with each repeated pull on a positioned needle, the extracted sample becomes visibly more haemodilute.

All Registrars responsible for taking bone marrows are made familiar with the research SOPs on induction, but anecdotal discussion suggests these protocols were not always rigorously undertaken, possibly in part because the full significance of changes in technique would only be apparent to those analysing the data.

5.5.2.2.1.4 Variable dilution with EDTA

The aspirated bone marrow was immediately placed into an eppendorf prefilled with 100µl EDTA for anticoagulation. Although the SOP requested the collection of 200µl only, it was clear when processing samples that the received volume of bone marrow varied between 200µl to 1ml. Therefore, the relative degree of dilution due to the additional EDTA will vary between samples, which might impact on observed stem cell concentration.

5.5.2.2.2 Time to processing

All samples included in analysis were processed within 24 hours of removal from BM extraction. However, one would imagine that cell viability could fall within this window, within increasing time of processing, even if the sample was kept at 4^{o}C.

5.5.2.2.3 Two-step analysis

As described in Sections 5.3.1.2 and 5.3.2, identification and subsequent enumeration of the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ population within these samples was a two-step process. Initial processing of the fresh sample with CountBeads enabled quantification of the $CD34^{+}$ concentration. Subsequent staining of a frozen sample allowed enumeration of the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations. Calculation of the concentration of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells in the fresh sample makes the assumption that these cells survive the processing, freezing and thawing process in exactly the proportion as the $CD34^{+}$ cells within the same sample.

This process was necessarily split for two reasons. Primarily, it was vital to accrue sufficient samples to study. Approximately 30 to 40 new cases of AML present to Bart's each year, but for a variety of reasons (patient consent, technical difficulties in bone marrow sampling, delays in sample transport, time of sampling etc. to name a few), by no means all of these provide suitable research samples. At the start of my PhD, I started analysis of all new AML and control

samples for their $CD34^+$ cell concentration. It was only after 24 months of concurrent work that I had constructed and validated the panel of antibodies which I felt had the optimal chance of distinguishing normal HSCs from AML, and hence enumerate stem cells.

Secondarily, this approach enabled me to utilise the samples previously processed between 2007 and 2012 by Dr Miraki-Moud, all of which had been quantified for $CD34^+$ cell concentration at diagnosis.

The potential variation introduced by two-step processing of samples was tested by the analysis when fresh of the last AML sample received (Pat ID 10782), with the entire panel of antibodies (CD34, CD38, ALDH and CLL1), along with Count Beads; followed by the staining of a frozen sample at a later date for the same markers. Using this technique, when stained fresh, the concentration of $CD34^+38^-ALDH^{high}CLL1^-$ cells was calculated as 1.4 cells/µl. When the concentration of the same group of cells was calculated using the method detailed in Sections 5.3.1.2 and 5.3.2, it was calculated as 2.5 cells/µl. Ideally, this comparison would be performed on more than one sample, but I was somewhat reassured that the differences in stem cell concentrations given by these two methods certainly could not explain the log fold changes in stem cell concentrations observed between different samples!

5.5.2.3 Is very accurate flow based quantification of stem cell concentration the best way of determining the functional capacity of the bone marrow for regeneration?

It is well documented that not all of the cells with long term reconstitutive potential reside in the $CD34^+38^-$ compartment within normal bone marrow, but that these markers determine a population significantly enriched for stem cell function[75]. ALDH and CLL1 expression pattern have been independently shown to mark cells with stem cell-like behaviour, but it remains an assumption that the concentration of $CD34^+38^-ALDH^{high}CLL1^-$ cells reflects the concentration of cells with true functional repopulating potential within the bone marrow. It may be, for example, that the proportion of functional stem cells within the $CD34^+38^-ALDH^{high}CLL1^-$ compartment of AMLs and control samples may significantly differ.

In Chapter 6, a number of AML and control BMs were sorted and placed into a 7 week long limiting dilution analysis to determine their functional stem cell numbers. Four of the control samples utilised in this experiment also had been quantified for their $CD34^+38^-ALDH^{high}CLL1^-$ cell concentration at diagnosis by flow based analysis. Table 5-13 summarises the comparison between the calculated stem cell concentration at diagnosis, and the concentration of cells capable of colony generation in a 7 week long culture assay within this population. It is clear from even these four samples that the concentration of functional long term culture

generating cells within this relatively tightly described phenotypic group varies significantly (between 1 in every 15 cells to 1 in every 161 cells).

Table 5-13 Comparison of $CD34^+38^-ALDH^{high}CLL1^-$ concentration at diagnosis, and apparently functional stem cell concentration by 7 week culture within the same $CD34^+38^-ALDH^{high}CLL1^-$ subpopulation, of 4 control samples

PAT ID	ESTIMATED [$CD34^+38^-ALDH^{high}CLL1^-$] BY FLOW (CELLS/µl)	LTC ESTIMATE OF 1/STEM CELL FREQ WITHIN $CD34^+38^-ALDH^{high}CLL1^-$ POPULATION		
		ESTIMATE	LOWER CI	HIGHER CI
7957	0.03	89	188	42
8510	3.07	15	29	8
8126	8.64	54	79	38
8215	3.01	161	506	51

Similarly, the results summarised in Chapter 4 raise the possibility that within the AML samples with poor risk cytogenetics, there exists a cohort of samples where $CD34^+38^-ALDH^{high}CLL1^-$ cells exist phenotypically, but appear unable to grow in the conditions of LTC, suggesting that these cells have impaired stem cell function.

5.5.3 Issues specific to AML samples

5.5.3.1 Sampling bias

There is also an inherent (and I would argue) unavoidable bias in the samples which were processed and enumerated, due to their biological nature.

Any sample with visible clot in it at the time of processing was excluded from further analysis. Coagulation of BM samples from APML patients is more likely to occur, due the susceptibility of these patients to DIC. This data series contains two APML samples within 45 samples, a relative under-representation from the expected 10%. Similarly, all samples successfully enumerated for their $CD34^+38^-ALDH^{high}CLL1^-$ population concentration at diagnosis had to have had both an eppendorf and an EDTA BM sample sent for research purposes. This requirement for relatively abundant BM effectively excludes those patients with difficult BM aspirates to obtain (i.e. those with AML M7).

5.5.3.2 Patchy infiltration with disease

It is generally assumed that the acute leukaemias, in contrast to other haematological malignancies such as myeloma or lymphoma, present with diffuse intramedullary involvement, and as a result, discrepant results from bilateral iliac biopsies are rare[94]. This assumption is corroborated by MRI appearances which show diffuse changes at diagnosis throughout the axial skeleton[244]. However, rare examples do exist where leukaemic infiltrates are patchy, and

BM attempts from a variety of sites are required to obtain a final diagnosis[245]. Thus it remains a possibility that some of these patients might yield very different results between technical replicates if bone marrow aspiration was attempted from multiple sites.

5.5.3.3 Might the development of AML change the phenotypic expression of surface markers on normal stem cells?

As discussed in Section 1.3.5.1, not all cells with long term repopulating ability reside within the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ fraction. For example, we know very immature cells may be localised within the $CD34^{-}$ population[75]. Although this distribution of cells with different phenotypic appearances exists in health, it is possible that this pattern of distribution may become more marked with the development of a disease such as AML. Changing patterns of HSC surface antigen expression with AML have not been described to date. However, in Chapter 6 of this thesis, we demonstrate that levels of CD33 expression within the $CD34^{+}38^{-}ALDH^{high}$ population of bone marrow drop significantly on development of AML. It is possible, albeit a very difficult question to address experimentally, that the expression levels of the other markers associated with long term repopulating cells might similarly fluctuate, and thus might make comparative enumeration of AML and control samples impossible.

5.5.4 Further experiments

The most pressing experiment to be undertaken would be the validation of this study on an independent cohort of patients, ideally from a different centre. I would place a particular emphasis on validating the conclusions drawn by comparison between the low and high stem cell groups within the AML, and the effect this appeared to have on overall survival, relapse-free survival, and 30 day survival.

Having now designed the appropriate test panel to identify stem cells in the majority of samples as detailed in Chapter 4, it would be desirable to process samples immediately for CD34, CD38, ALDH and CLL1 expression patterns in the presence of Count Beads, to reduce the potential error involved in the sequential processing of fresh and then frozen samples.

Increasing the sample size would clearly be advantageous to increase the confidence in any clinical correlations drawn from the data, which are somewhat limited with an AML sample size of 45, given the heterogeneous nature of the disease and treatment.

To investigate further whether the variability observed within the control cohort was a true biological phenomenon, and not due to technical and sampling issues, the study of biological

replicates (i.e. samples taken from simultaneously from the right and left iliac crest) from a limited number of patients with suitable consent would be desirable.

I have been partially able to explore the question of variability between technical replicates for a few selected patient samples. For all of the AML and control patients studied, we received only one Eppendorf for immediate processing, and therefore there is no possibility of exploring the effects of technical replicates for this step of the process. However, the additional samples which were stored in the Tissue Bank (See Section 2.2.2) were commonly divided into multiple vials at the time of processing dependent on the cell number received. For two of the samples studied, I have extracted multiple vials of the same sample for different experiments over different days, and stained them with the same panel of antibodies. Comparison of these results enables a snapshot view of the potential variability in results between technical replicates (see Table 5-14).

Table 5-14 illustrates the variability in technical replicates of previously frozen samples used for $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cell quantification and the effect this has on quantification of the $[CD34^{+}38^{-}ALDH^{high}CLL1^{-}]$ at diagnosis

PAT ID	ATTEMPT	CD34+	$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ CELLS	$[CD34^{+}]$ FROM FRESH SAMPLE (CELLS/µl)	ESTIMATED $[CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ CELLS] IN ORGINAL
7882	1	13234278	25089	43687.4	82.82
	2	5814422	10497		78.87
7926	1	132146	1844	59.5	0.83
	2	20880	882		2.51

I remain intrigued that there appeared to be no obvious relationship between HSC number at diagnosis and time to count recovery. However, it is interesting that in this cohort, there were no patients who obtained remission from leukaemia after their Induction chemotherapy, but with significantly prolonged count recovery. One patient (9713) had prolonged time to platelet recovery (64 days), but this was in the context of severe abdominal sepsis. None of the patients who survived induction chemotherapy were rendered persistently hypoplastic or aplastic. It would be interesting to specifically identify patients who showed a marked delay in count recovery after induction, despite achieving remission from leukaemic infiltration, and quantify their stem cells at diagnosis, to see if this cohort had abnormal presentation HSC concentrations. However, in an analogous situation, not all studies show a relationship between CD34 dose and time to haematopoietic recovery after BMT, and therefore, factors other than a received dose of haematopoietic progenitors are likely to play a role[246].

Finally, I remain intrigued by the observed significant variation in HSC numbers within control samples. I would be interested to repeat this experiment on healthy volunteers, to see if this

finding is reproduced in a totally normal cohort. I think it would also be fascinating (with a willing donor!) to perform serial BMs over time in one individual, to identify the degree of fluctuation in HSC concentration over time.

Chapter 6 Exploring the effect of AML on CD33 expression on normal stem cells

6.1 Introduction

6.1.1 CD33 expression patterns

6.1.1.1 In normal haematopoiesis

CD33 (also known as sialic acid binding Ig-like lectin 3 or Siglec-3), is a transmembrane receptor frequently, but not exclusively, expressed on myeloid cells. It is a member of the sialic acid binding family of receptors: other members of which include CD22 (Siglec-2), CD169 (Siglec-1) and MAG (Siglec-4). These receptors are thought to mediate cell-cell interactions and intercellular signalling through binding to sialic acid.

In addition, there exists a large subfamily of CD33-related Siglec receptors, which share significant structural homology. These receptors appear to have evolved rapidly and recently: comparisons between species have revealed little homology between mammals such as mice, monkeys and humans. This has obvious ramifications for the utility of animal models in investigating the role of these receptors in normal cell functioning[247].

The members of the CD33-related Siglec family are expressed mainly on the mature cells of the innate immune system, such as neutrophils and eosinophils[248,249]. In contrast, CD33 itself tends to be expressed earlier in development at the progenitor phase. To date, reported roles for this family of receptors include inhibition of cellular proliferation[250], apoptosis induction[251], inhibition of cell activation[210] and induction of pro-inflammatory cytokine secretion[252]. However, it is unknown at present how much overlap in function exists between structurally similar receptors.

How do CD33 and related Siglecs mediate the effects described above? Structural studies have revealed all these receptors to contain an amino-terminal V set immunoglobulin domain, capable of sialic-acid recognition. Avidity for different sialic residues appears to vary between family members. Interestingly, sialic acids are found abundantly in the membrane-region of most human cells, and therefore it is assumed that for many of these receptors, their binding sites may be occupied by *cis*, rather than *trans*, ligands[247].

The intracellular section of these proteins contains immunoreceptor tyrosine-based inhibitory motifs that are implicated in inhibition of cellular activity. On phosphorylation of these motifs

by SRC tyrosine kinases, the receptors recruit SRC homolog 2 (SH2)-domain containing proteins, including SHP1 and 2, or SOCS 3[210].

The current prevailing consensus in published literature is that CD33 is absent from the most immature HSCs, with expression occurring only at the progenitor stage onwards in development[212,253-255]. However, several groups, including our own, have reported the expression of CD33 on cells with a stem phenotype[256-258]. Taussig et al provided functional data which suggest that the $CD33^+$ fraction of sorted human $CD34^+38^-$ stem cells is capable of preferential engraftment (when compared to the $CD33^-$ fraction) in a NOD-SCID murine xenograft model[214].

6.1.1.2 In AML

CD33 expression in AML is highly ubiquitous, with 90 to 95% of clinical samples showing some level of antigen expression[211,212]. Levels of antigen expression vary significantly between patients[259,260], although the average antigen density on a typical blast cell is relatively sparse at 10^4 molecules per cell[260]. Mimicking the patterns seen in normal haematopoietic development, antigen expression is lower in the fraction traditionally though to contain LSCs ($CD34^+38^-CD123^+$ or $CD34^+38^-$). Antigen expression appears to vary predictably with disease genetics: APML is associated with high CD33 expression[261], in contrast to the other core-binding factor leukaemias, where low CD33 levels are classically observed[259]. Similarly, high levels of CD33 expression are also associated with *NPM1* and *FLT3*-ITD mutations[262].

6.1.2 Anti CD33 therapy in AML

The backbone of standard chemotherapy in AML (daunorubicin and cytarabine) has not changed in 40 years. Such intensive, untargeted cytotoxic therapy is associated with significant side effects, including marked myelosuppression, which can limit its use in older patients. As detailed in Section 1.4, even when administered at full dosage, current therapy frequently fails to cure patients, with only half surviving to five years after diagnosis. There is an absolute consensus amongst clinicians that new effective therapies are urgently required.

Interest in the development of targeted molecular therapies for haematological malignancies, combined with knowledge of the ubiquitous expression of CD33 in AML, resulted in focussed attempts to develop an anti-CD33 antibody therapy. Despite challenges with CD33 as a target for therapy (due to its relatively low abundance of expression, and slow conjugate internalisation)[260], cross linking of the receptor with antibodies was shown to cause significant apoptosis in AML samples over ten years ago[213].

Today, a number of related therapies are either in use or development, and their details are summarised below.

6.1.2.1 Gemtuzumab Ozogamicin (GO)

The first developed, and most widely studied anti-CD33 antibody in clinical practise is Gemtuzumab Ozogamicin (GO) (marketed in the UK as Mylotarg by Wyeth): a monoclonal humanised IgG4 antibody against CD33 (hP67.6), conjugated to calicheamicin[263]. The unconjugated antibody is not directly toxic to cells, but binding to the CD33 receptor facilitates internalisation of the conjugate. Subsequent lysosomal degradation results in a calicheamicin derivative capable of causing single and double stranded DNA breaks. These in turn lead to cell cycle arrest, and either DNA repair or cellular apoptosis. The central role of DNA damage in GO-induced cytotoxicity is supported by the observation that cell lines defective in DNA repair are highly sensitive to calicheamicins[264].

In vitro and in vivo resistance to GO can be mediated via increased drug efflux caused by members of the ATP binding cassette protein family, such as P-glycoprotein (ABCB1) and multidrug resistance-associated protein 1 (ABCC1)[265-268].

GO gained accelerated FDA approval for use in AML therapy following one non-randomised study, which revealed a benefit with the administration of single agent GO to patients over the age of 60 with relapsed AML[269]. Since then, the efficacy of GO has been tested in several large international trials, in a range of age groups and dosing schedules, but mainly in newly diagnosed patients in combination with standard chemotherapy.

Concerns about its potential toxicity in treatment (chiefly the development of veno-occlusive disease, but also reports of prolonged cytopenias[270,271]), had been raised since the FDA licence was approved. However, in 2010, the drug was voluntarily withdrawn in the USA, following the Phase III SWOG S0106 trial, which showed an increase in toxicity-associated mortality in those patients who received GO, without a matched benefit in terms of disease control[272]. However, in the last three years, the results of several other collaborative studies have been published, refuting these findings[3,4]. Experts remain divided over its clinical benefit[273].

6.1.2.1.1 The UK Experience

The two largest studies investigating the efficacy of GO with standard chemotherapy in treating newly diagnosed AML patients have been the MRC-sponsored AML 15 and 16 trials. The AML 15 trial, which ran between 2002 and 2006, randomised 1113 patients (predominantly aged under 60) to receive GO in induction and/or consolidation, in

combination with a variety of standard chemotherapy regimens. 557 patients were randomised to receive GO at induction, and follow-up revealed no differences in overall survival, response rate or 30-day all-cause mortality between the two treatment arms. However, subgroup analysis revealed that those patients with a favourable risk karyotype (inv(16) and t(8;21)) showed a significant improvement in overall survival if they received GO at induction (79% versus 51%)[3]. This finding was replicated in the SWOG SO106 trial[272].

Interestingly, in those patients who received GO at a dose of $3g/m^2$, the AML 15 study did not reveal significantly increased haematopoietic suppression. Those receiving GO treatment at induction did require significantly more platelet transfusions (18.9 versus 13.7 units, $p<0.001$) during course 1, and more days of antibiotics (20.4 versus 19 days, $p=0.02$). However, the time to haematopoietic recovery was similar between the 2 groups (time to neutrophil recovery 20 versus 21 days ($p=0.3$), and to platelet recovery 20 versus 19 days ($p=0.2$)). A similar effect was observed in AML 16, which used GO in an older patient cohort[4].

These promising data, showing an improvement in overall survival, without the significant toxicity reported in the SWOG trial, has led to the inclusion of GO within the protocol for both current UK based AML trials (AML 18 and 19)[122,126]. In these, all eligible patients should receive either one or two doses of GO during their induction therapy.

6.1.2.2 Vadastuximab talirine (SGN-CD33A)

Although GO remains the most widely studied anti-CD33 agent, difficulties with access and concerns over toxicity have led to the development of a number of competitors. These include Vadastuximab talirine (SGN-CD33A), marketed by Seattle Genetics. Vadastuximab is a conjugate of an anti-CD33 antibody, attached to a pyrrolobenzodiazepine dimer of the drug SGD-1882 (a DNA binding agent). Preliminary data from phase I trials demonstrating tolerability have been reported[274,275], and a Phase II clinical trial involving its use in combination with demethylating agents, to treat older patients newly diagnosed with AML, is planned to open this year.

6.1.2.3 AMG 330

AMG 330 is a bispecific T-cell engaging (BiTE) antibody, which has specificity for CD33 and CD3 antigens. In a mechanism analogous to the more well-known Blinatumomab (which binds to CD3 and CD19)[276], the design of the BiTE antibody brings $CD3^+$ cytotoxic T cells into contact with cells expressing CD33, resulting in cell destruction. One potential advantage in its design is that, as internalisation of the drug is not required for its mechanism of action, resistance to its effects cannot develop via drug efflux mechanisms[129,266].

In vitro experimental work has shown AMG 330 to be capable of lysis of AML blasts[277]. Recruitment for a Phase I clinical trial is currently underway.

6.1.2.4 Does anti-CD33 therapy affect normal haematopoiesis?

As discussed above in Section 6.1.1.1, the classical understanding of CD33 expression patterns during haematopoietic development describes antigen expression beginning only as cells enter the progenitor stage of differentiation[212,253-255]. However, several other groups, including our own, have shown cells capable of long term reconstitution to express CD33, suggesting it can be found on the most primitive of HSCs[214,256-258].

Whether HSCs express CD33 has clinical relevance for those patients receiving anti-CD33 therapy, as elimination of cells with long term reconstituting potential could lead to prolonged cytopenias (and delays in subsequent consolidation chemotherapy), or indeed permanent aplasia.

Interestingly, in the early days of GO therapy, there were several reports of its usage being associated with prolonged cytopenia in remission[270,271]. However, it is noteworthy that the data available on count recovery from the much larger cohort of patients in the AML 15 and 16 trials do not support these early concerns (see Section 6.1.2.1.1).

6.1.3 Data from our group showing CD33 expression patterns within the normal HSC population of $CD34^{low}$ AML samples

Why do we not observe more prolonged haematopoietic suppression with the use of GO clinically, given that we and others have demonstrated normal stem cells do express CD33?

In Section 1.5.2.2, we describe a cohort of AML samples defined by low CD34 expression. In these samples, it has been demonstrated that the $CD34^{+}38^{-}$ compartment retains normal stem cell activity, with no evidence of the leukaemia-associated mutation[130]. These samples are therefore ideal to use in the investigation of the effect of AML on normal HSCs[1].

Previous unpublished work from our group has shown that the CD33 expression (the percentage of CD33 positive cells, as defined by FMO) in the $CD34^{+}38^{-}$ normal stem cell fraction within these $CD34^{low}$ AML samples is significantly lower than the same compartment within control samples. Median percentage CD33 positivity within 15 $CD34^{low}$ AML samples was 18%, versus 82% within 20 control samples ($p < 0.0001$ by Mann-Whitney test) (see Figure 6-1).

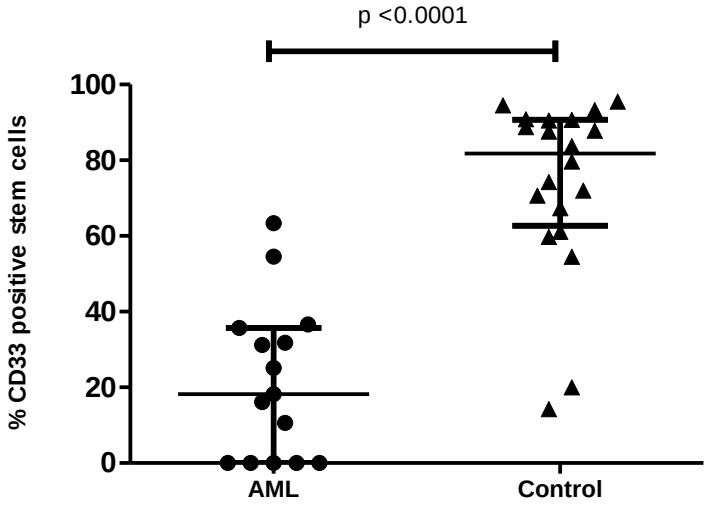

Figure 6-1 showing CD33 expression within the $CD34^{+}38^{-}$ cells of $CD34^{low}$ AML and controls. Median and interquartile range are plotted for both groups. Median CD33 expression was 18% (AML) v 82% (controls)

This is a novel finding, and I believe the first demonstration of normal HSCs illustrating an altered surface phenotype in response to exposure to AML. This led us to ask whether this altered expression pattern is limited solely to the normal HSCs found within $CD34^{low}$ AMLs, or is a more universal finding. Secondarily, given increasing interest in anti-CD33 therapy, it also raises the question as to whether these changes in expression of CD33 have clinical impact in terms of altered sensitivity to antigen-targeted therapy.

6.2 Aims and Objectives

The primary aim of the work in this chapter was to study the CD33 expression pattern within the normal HSC compartment of a range of AML subtypes, to identify if the reduced expression observed on normal HSCs from $CD34^{low}$ AMLs is a universal finding.

The first objective was therefore to study CD33 expression patterns within the $CD34^{+}38^{-}$ $ALDH^{high}$ compartment of CBF leukaemias. Previous work from Gerber[207], as well as results detailed in Section 4.4.3.3, had shown we could confidently separate normal HSCs from AML cells in this cohort of samples using this antigen panel, without the need to validate the purity of the sort. This enabled the testing of multiple samples quickly. In addition, previous trial data suggest it is this particular group of patients who gain the most clinical benefit from GO therapy[3], and therefore are the most likely to receive it if the drug becomes widely available in the future.

The secondary objective was to see if any changes in CD33 expression are reversible on remission. This was achieved by identifying remission samples from CBF AML patients, where available, and simultaneously assaying them with diagnosis samples.

The secondary aim of this chapter was to identify if a change in CD33 expression has an impact on HSC sensitivity to GO therapy. Does lower CD33 expression reduce susceptibility to the drug, and therefore have a protective effect for the normal HSC population at diagnosis with AML, when compared to controls?

Thus the final objective was to test the effect of GO on normal HSCs using an in vitro LTC assay. Sorted normal HSCs from AML samples were briefly exposed to GO, prior to assessment of residual stem cell function by a 7 week LTC assay. The effect of the drug on stem cell frequency was compared to control samples treated in the same fashion.

6.3 Chapter Specific Methods

6.3.1 Patient Samples

Diagnostic bone marrow from patients diagnosed with CBF mutated AML and $CD34^{low}$ AML along with control samples were obtained from the Bart's Tissue Bank with prior patient consent.

6.3.2 CD33 Expression Analysis

6.3.2.1 Staining

Bone marrow samples were thawed, washed and cell counted as previously described.

$4x10^6$ cells (or the total contents of the bone marrow vial if lower) were stained with activated Aldefluor™ reagent as described. HAG staining was followed by staining with CD34-PerCP, CD38-Pecy7 and CD33-APC (BD, Cat: 551378 (Clone WM53)) antibodies at a concentration of $5\mu l/1x10^6$ cells.

6.3.2.2 Flow Data Analysis

Samples were all processed on a BD Biosciences LSR Fortessa. Subsequent analysis was performed using the FlowJo Program. The number of $CD34^+CD38^-ALDH^{high}$ normal stem cells in each sample was quantified, and the proportion of $CD33^+$ cells (as defined by the Fluorescence Minus One CD33 control for each sample) recorded. For those patients with paired diagnostic and remission/relapse samples, all samples were processed on the same day to aid inter-sample comparison.

6.3.2.3 Statistical analysis

The majority of statistical analysis was performed using the Prism software package. All datasets were tested for normality of distribution by the D'Agostino & Pearson omnibus normality test. If data were normally distributed, means are quoted, and a t-test was used for analysis. If data were not normally distributed, medians are quoted, and a Mann-Whitney t-test was applied in the determination of statistical significance.

6.3.3 GO Exposure Experiment

This experiment was performed at three different times, dependent on the availability of GO. As we were unable to access GO directly from Wyeth for this work, we instead utilised the excess from vials reconstituted for patients within the AML-19 trial. In all cases, the GO was used within 72 hours of reconstitution, and before use, was kept in the dark at 4^oC.

In total, 7 AML and 7 control samples were used: the numbers were limited by the costs of the experiment, as well as the substantial time taken to sort sufficient HSCs from each sample.

6.3.3.1 Sorting of HSCs

Two slightly different sorting techniques were utilised, dependent on the subtype of AML being investigated. In the first experiment, HSCs were extracted from three $CD34^{low}$ AML samples and three matched controls, and, in an attempt to maximise the stem cell yield prior to LTC, HSCs were extracted based on a $CD34^{+}38^{-}$ phenotype only. In the subsequent two experiments, for CBF AMLs and matched controls were sorted, and as described in Chapter 4, a $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ criteria was required for separation of HSCs from AML cells in these cases.

The three $CD34^{low}$ AML and three matched control bone marrow samples were thawed, washed and cell counted as previously described. HAG staining was followed by staining with CD34-PerCP and CD38-PeCy7 at a concentration of 5µl/1.5x10^{6} cells. A further aliquot of 0.5 million cells was stained with HAG and then CD34-PerCP, CD38-PeCy7 and CD33 APC, and analysed separately to assess CD33 expression. Cells were resuspended in 2% PBS/DNase/DAPI and the $CD34^{+}38^{-}$ population sorted into 2% PBS. Purity was checked, and cells resorted if purity fell below 90% as previously described.

The four CBF AMLs (one t(8;21) and three inv(16)) and four control bone marrow samples were thawed, washed and cell counted as previously described. Samples were stained for ALDH expression as previously described. Staining with HAG dissolved in ALDH buffer was followed with staining for CD34-PerCP, CD38-PeCy7 and CLL1-PE at a concentration of 5µl/1.5x10^{6} cells. A further aliquot of 0.5 million cells was stained for ALDH expression, followed by HAG and finally stained with CD34-PerCP, CD38-PeCy7, CLL1 PE and CD33 APC, and analysed separately for CD33 expression. Cells for sorting were resuspended into ALDH/DNase/DAPI and the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells sorted into ALDH buffer as previously described. Sort purity was checked, and cells resorted if purity fell below 90%.

6.3.3.2 Drug Exposure

In both experiments, the exact number of cells sorted for each samples was recorded. Subsequent to sorting, the sample was gently mixed by pipetting and then divided into 2 tubes of equal volume. Tubes were spun (1500rpm, 10 minutes), and the cells resuspended in 1ml H5100 Myelocult™ with and without the addition of GO at a concentration of 3µg/ml for 2 hours at 37°C. Samples were then washed with 3ml H5100 Myelocult™, and centrifuged

(1500rpm, 10 minutes). The supernatant was aspirated using a Gilson pipette, and cells resuspended in 100µl H5100 Myelocult™.

6.3.3.3 LTC set up and maintenance

Cells were then plated in serial dilutions of between 2 to 800 cells per well, on plates set up for LTC conditions with collagen and MS5 as previously described. A maximum additional fluid volume of 10µl per well was added. 4 to 20 replicates for each cell dilution were added, dependant on the cell numbers available for plating.

Cells were incubated at 37°C, 5% CO_2 for five weeks, with weekly changes of H5100 media (65µl supernatant extracted and replaced with 75µl fresh H5100 Myelocult™, pre-warmed to 37°C). After 5 weeks in culture, the supernatant was aspirated from each well using a Gilson aspirator, leaving a residual volume of 10 to 20 µl. 100µl of methylcellulose was added to each well and the plates cultured at 37°C in an incubator for a further 2 weeks. Wells were then examined for the presence of colonies and recorded as either positive or negative.

6.3.3.4 Data analysis

An LTC-IC concentration was calculated for each condition using the Extreme Limiting Dilution Analysis software package[278,279]. Statistical analysis comparing the relative effect of GO treatment on HSCs from control and AML samples was performed using the Wilcoxon signed rank test.

6.4 Results

6.4.1 CD33 expression patterns within $CD34^{+}38^{-}ALDH^{high}$ stem cells of CBF AML samples

All available patient samples from patients with core binding factor mutations were tested. A total of 53 samples of diagnostic, first remission and relapse bone marrow samples from 21 patients diagnosed with CBF leukaemia were obtained from the Bart's Tissue Bank with prior patient consent. These were compared to 9 normal bone marrows.

6.4.1.1 Patient Details

Analysis of the Bart's Patient Registry identified 37 patients diagnosed with CBF leukaemias between 1997 and 2012. 13 patients had inv(16) AML and 24 t(8;21) AML. Of these, diagnosis BMs were obtainable for 22 patients (14 from t(8;21) and 8 from inv(16) patients).

The vast majority of these patients achieved a remission with standard induction therapy. Remission BMs were obtainable for 17 of these patients (10 with t(8;21) and 7 with inv(16)).

Despite the relatively favourable prognosis of the CBF leukaemia subgroup, a proportion of these patients relapsed. Of the 22 patients with obtainable diagnosis BMs, 4 patients had relapse BM samples stored in the Tissue Bank (all 4 patients had t(8;21) mutated AML). Sample availability and clinical information is summarised in Table 6-1.

6.4.1.2 CD33 expression on normal stem cells in CBF AMLs is downregulated at diagnosis in comparison to controls

22 diagnosis BM samples for patients with CBF leukaemias and 9 controls were analysed for the CD33 expression of their normal stem cell populations. Normal stem cells were identified as $CD34^{+}38^{-}ALDH^{high}$ cells. CD33-expressing stem cells were defined as those with a higher intensity of CD33 expression than the FMO control, and are expressed as a percentage of all $CD34^{+}38^{-}ALDH^{high}$ cells The mean percentage of $CD33^{+}$ normal stem cells found within the bone marrow of CBF AML samples is significantly lower than that of control bone marrow (means 17% versus 58%; p=0.005, unpaired t-test with Welch correction) (see Figure 6-2 A).

CD33 expression is lower in HSCs in the 8 samples with inv(16) than the 14 samples with t(8;21) (median 2% versus 13%, p=0.03 by Mann Whitney t test). However, allowing for multiple testing, this difference does not achieve statistical significance. When compared independently to the control samples, the stem cells from each favourable-risk cytogenetic group do display statistically significant reduced CD33 expression (inv(16) 2% versus controls

67%, p=0.001 by Mann-Whitney t-test; t(8;21) 13% versus controls 67%, p=0.01 also by Mann-Whitney -test). These results are illustrated graphically in Figure 6-2 B.

6.4.1.3 CD33 expression on normal stem cells in CBF AML increases towards baseline when patients obtain morphological remission

Of the 22 AML patients whom a diagnosis BM was obtainable, 16 also had a BM sample available which was taken at clinical remission (7 with inv(16) and 9 with t(8;21)). Of these samples, fifteen were obtained after induction chemotherapy, and one dated from 6 months post completion of chemotherapy. The percentage of $CD33^+$ normal stem cells was compared between paired diagnosis and remission samples.

The mean percentage of $CD33^+$ normal stem cells across all CBF samples at diagnosis was 13%; at remission, this figure rose to 55%. The mean average increase of $CD33^+$ stem cells was 42%. Paired two-tailed t-testing suggests this change is highly significant ($p<0.0001$). This result is illustrated graphically in Figure 6-2 C. This increase in CD33 expression remained statistically significant when the two cytogenetic groups were considered independently. The seven patients with inv(16) AML exhibited an average increase in CD33 positivity within their stem cell populations of 33% from 2 to 35% (p=0.03); the nine patients with t(8;21) AML similarly showed a slightly greater increase of 48% from 22 to 70% (p=0.002).

6.4.1.4 CD33 expression on normal stem cells in CBF AML does not fall in the same manner at relapse

Of all patients with CBF treated at Bart's, four of those who subsequently relapsed had a BM taken at that time available for analysis. All of these patients had AML associated with t(8;21).

The results of parallel analysis of diagnosis, remission (available for three of the four), and relapse samples for the CD33 expression in the HSC sunset are shown in Figure 6-2 D.

The median percentage of HSCs which are CD33 positive at diagnosis is 23%; at remission this figure rises to 91%, and falls again at relapse to 59%. These results do not reach statistical significance when analysed by Friedman's test, but in this case, the very small sample size may be relevant.

Table 6-1 Clinical characteristics and sample availability of CBF AMLs

CLINICAL CHARACTERISTICS					SAMPLE AVAILABILITY		
ID	**KARYOTYPE**	**SEX**	**AGE**	**PRESENTATION WCC**	**REMISSION SAMPLE?**	**RELAPSE SAMPLE?**	**CLINICAL OUTCOME**
3206	INV(16)	F	59	38	Y		REMISSION
3229	INV(16)	F	30	34	Y		REMISSION
3230	t(8;21)	F	44	5.3	Y	Y	RELAPSE @ 2 YEARS
4400	t(8;21)	M	49	14	Y		RELAPSE @ 5 MONTHS
5904	t(8;21)	M	34	16	N		REMISSION
6252	t(8;21)	M	53	28	Y		RELAPSE @ 2 YEARS
6434	t(8;21)	F	18	4.9	Y	Y	RELAPSE @ 1 YEAR
6461	t(8;21)	F	34	28	N		REMISSION
6527	t(8;21)	F	49	5.9	Y		REMISSION
6550	t(8;21)	M	68	6	Y		REMISSION
6555	t(8;21)	F	22	70	N		REMISSION
6720	INV(16)	F	68	127	Y		RELAPSE @ 7 MONTHS
6734	t(8;21)	M	51	6	Y	Y	RELAPSE @ 3 MONTHS
6755	INV(16)	M	65	113	Y		REMISSION
6988	INV(16)	M	55	7.2	Y		REMISSION
7013	INV(16)	F	59	61	Y		RELAPSE @ 2 YEARS
7849	t(8;21)	F	57	9.4	Y		REMISSION
7882	INV(16)	M	46	33	N		RELAPSE @ 1 YEAR
8750	t(8;21)	F	39	6.1	Y		REMISSION
8751	t(8;21)	M	62	22	N	Y	RELAPSE @ 7 MONTHS
9220	t(8;21)	F	55	41	Y		REMISSION
9445	INV(16)	F	52	95	Y		REMISSION

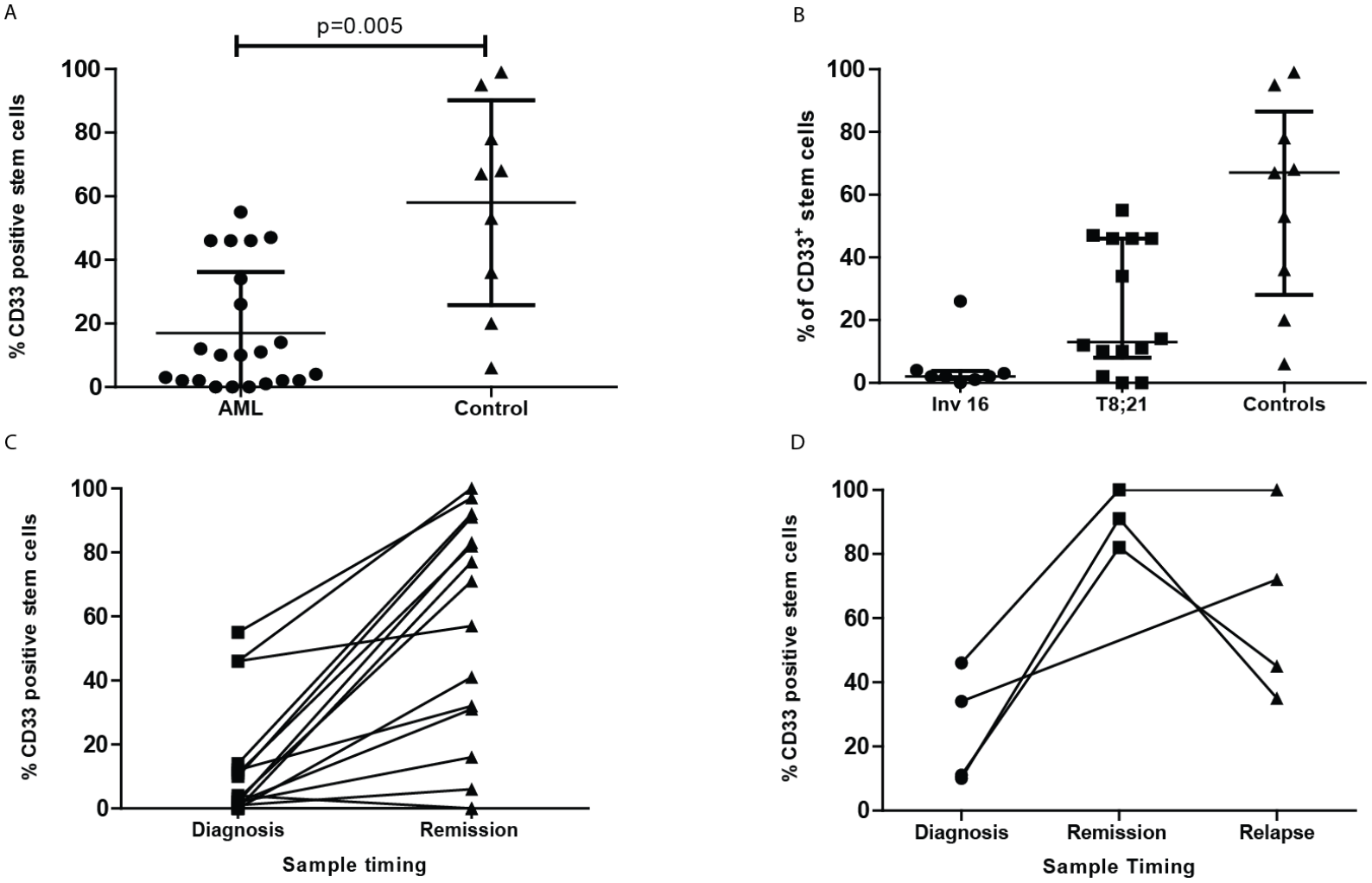

Figure 6-2 A: CD33 expression within the HSCs of CBF AML samples (n=22) versus controls (n=9). Means and SD are shown; B: CD33 expression within the HSCs of inv(16) AMLs (n=8), t(8;21) AMLs (n=14) and controls (n=9). Median and IQR are displayed for all samples; C: changing pattern of CD33 expression on normal HSCs between diagnosis and remission for 16 CBF AML patients; D: CD33 expression within the HSC compartment of 4 t(8;21) AML patients at diagnosis, first remission and relapse

6.4.2 The effect of CD33 expression of HSCs on susceptibility to GO therapy

In 3 separate experiments, a total of 7 AML and 7 control samples underwent sorting for HSCs. Extracted HSCs were exposed to GO at a concentration of 3µg/ml for 2 hours, prior to 7 week LTC. At the end of this period, stem cell concentration was calculated using the Extreme Limiting Dilution Analysis software package as described in Section 6.3.3.4.

In the first experiment, stem cells from 3 $CD34^{low}$ AMLs were extracted by selection of the $CD34^{+}38^{-}$ cells. 3 matched controls were sorted in the same way. Previous data had shown these antibodies to be capable of selecting out normal stem cells from these AMLs, and my initial concerns about sorting sufficient stem cells to perform parallel limiting dilutions analyses meant I had selected staining by this method, rather than also including ALDH and CLL1. In retrospect, I acknowledge that using the same panel of antibodies for all samples processed would have been preferable to enable comparison between different AML subtypes.

In the two subsequent experiments, HSCs from 4 CBF AMLs (3 with inv(16) and 1 with t(8;21)) were separated with an additional 4 controls by selection of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells. In all experiments, samples were split into two groups, exposed to 0 and 3µg/µl GO in H5100 respectively. Whilst all other experimental details remained unchanged, in the final experiment, knowledge of the likely frequency of stem cells within the samples meant that care was taken to plate multiple (20 wells) at the extreme low ranges of the dilution (2,5,10 cells per well) to ensure sufficient statistical power to detect any discrepancies between groups.

Clinical details of all samples and number of cells sorted are listed in Table 6-2. Final estimation of stem cell frequencies generated by the limiting dilution assays are shown in Table 6-3.

6.4.2.1 GO does not appear to have a significant effect on functional HSC numbers in either AML or control samples

In only three of these fourteen samples, did there appear to be a significant difference in the stem cell frequencies generated between the untreated and treated samples (7882, 8126 and 8215). One of these was an AML sample, and two were controls. In two of these cases, the stem cell frequency was greater in the untreated sample (as we had hypothesised). But in one (8215), the reverse was true. This is very difficult to explain. Further discussion as to why this might be the case is found below in Section 6.5.3.

6.4.2.2 Comparison of the relative effect of GO on HSCs from AML and control samples

Whole group analysis on the relative effect of GO on HSCs from AML and controls was performed using the Wilcoxon signed rank test. There was no statistically significant difference found either between estimated stem cell frequencies before and after GO treatment in the seven AML HSC samples (p=0.47), or between stem cell frequencies before and after GO treatment in the seven control HSC samples (p=0.22).

For completeness, the same analysis was performed looking solely at the $CD34^{low}$ AML samples and controls sorted for $CD34^{+}38^{-}$ cells. There was no significant difference in estimated stem cell frequencies before and after GO treatment in the three AML samples (p=1.0), or three controls (p=0.5). We also analysed the CBF AML samples and matched controls sorted for $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells. Unsurprisingly, there was no significant difference found in estimated stem cell frequencies before and after GO treatment in the four AML samples (p=0.25), or four controls (p=0.63).

Table 6-2 Details of AML ($CD34^{low}$ and CBF) and control samples sorted for stem cell populations and subsequently exposed to GO

SAMPLE INFORMATION					STEM CELL SORTING		
PAT ID	**TIME**	**TYPE**	**KARYOTYPE**	**FLOW**	**SORT CRITERIA**	**CELLS SORTED**	**CELLS PER CONDITION**
7774	DIAG	AML	N	$CD34^{low}$	$CD34^{+}38^{-}$	5809	2905
8169	DIAG	AML	t(11;19)	$CD34^{low}$	$CD34^{+}38^{-}$	4960	2480
8452	DIAG	AML	t(11;17)	$CD34^{low}$	$CD34^{+}38^{-}$	9283	4642
7926	DIAG	CONTROL	N		$CD34^{+}38^{-}$	2893	1447
8115	DIAG	CONTROL	N		$CD34^{+}38^{-}$	2898	1449
10115	DIAG	CONTROL	N		$CD34^{+}38^{-}$	18768	9384
3206	DIAG	AML	INV(16)	CBF	$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$	8999	4499
7013	DIAG	AML	INV(16)	CBF	$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$	3997	1998
7882	DIAG	AML	INV(16)	CBF	$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$	14376	3594
6434	DIAG	AML	t(8;21)	CBF	$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$	3179	1589
7957	DIAG	CONTROL	N		$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$	1968	984
8126	DIAG	CONTROL	N		$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$	28276	4713
8215	DIAG	CONTROL	N		$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$	1146	573
8510	DIAG	CONTROL	N		$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$	3598	1794

Table 6-3 illustrating results for all samples from ELDA Hall analysis of stem cell frequency and pairwise analysis of results

PATIENT DETAILS			EXPERIMENT DETAILS		RESULTS				
ID	SAMPLE	KARYOTYPE	SORT CRITERIA	GO DOSE µg/ml	1/STEM CELL FREQUENCY			PAIRWISE ANALYSIS TREATED V UNTREATED	
					ESTIMATE	LOWER CI	HIGHER CI	CHISQ	PROB (>CHISQ)
7774	AML	N	$CD34^{+}38^{-}$	0	10.1	22.5	4.5	0.225	0.635
				3	12.9	25.2	6.6		
8169	AML	t(11;19)	$CD34^{+}38^{-}$	0	12.4	24.7	6.2	3.39	0.0657
				3	4.4	9.9	1.9		
8452	AML	t(11;17)	$CD34^{+}38^{-}$	0	4.4	9.1	2.1	0.328	0.567
				3	5.9	11.6	2.9		
7926	CONTROL	N	$CD34^{+}38^{-}$	0	25.4	47.7	13.5	0.154	0.695
				3	29.9	55.7	16.1		
8115	CONTROL	N	$CD34^{+}38^{-}$	0	12.5	21.9	7.1	3.41	0.0646
				3	25.1	43.2	14.6		
10115	CONTROL	N	$CD34^{+}38^{-}$	0	13.2	27.0	6.5	0.000216	0.988
				3	13.1	26.8	6.4		
3206	AML	INV(16)	$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$	0	9.3	12.6	6.9	0.864	0.353
				3	7.6	10.3	5.6		
7013	AML	INV(16)	$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$	0	7.9	16.4	3.8	2.87	0.0902
				3	18.0	34.8	9.3		
7882	AML	INV(16)	$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$	0	7.4	10.1	5.4	7.96	0.00479
				3	14.0	18.9	10.3		
6434	AML	t(8;21)	$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$	0	5.0	11.1	2.3	1.35	0.244
				3	9.3	19.1	4.6		
7957	CONTROL	N	$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$	0	89.2	188.0	42.3	0.352	0.553
				3	124.8	287.0	54.3		
8126	CONTROL	N	$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$	0	54.4	78.5	37.6	10.8	0.00099
				3	148.4	244.1	90.2		
8215	CONTROL	N	$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$	0	160.8	506.0	51.1	5.2	0.0226
				3	39.6	76.9	20.4		
8510	CONTROL	N	$CD34^{+}38^{-}ALDH^{high}CLL1^{-}$	0	14.8	29.1	7.5	1.56	0.211
				3	26.5	49.5	14.2		

6.5 Discussion

6.5.1 General points

Flow based analysis of CD33 expression on clinical samples from 22 patients with CBF leukaemias detailed in Section 6.4.1, suggests that CD33 expression is significantly lower on normal HSCs exposed to AML, but increases significantly once remission is obtained. This mirrors an effect previously found by my group when studying the CD33 expression of the HSC compartment of $CD34^{low}$ AMLs. We chose to study CBF leukaemias partly because there exists a validated means of identifying and separating out stem cells from the malignant clone in these samples[207], and partly because the interaction between CBF AMLs and GO is of current interest, as it is these patients that derive the most clinical benefit from GO therapy[3].

6.5.2 Why might CD33 expression show variable patterns on normal HSCs in the context of AML? Is this a cause or an effect?

Why do normal HSCs show reduced CD33 expression in this context of leukaemia presentation? As the exact function of the CD33 receptor on HSCs is unknown, this is a difficult question to answer.

Increased expression of CD33 is associated with a more mature phenotype, as cells develop towards the progenitor $CD34^{+}38^{+}$ subtype. If (as we speculate in Section 1.5.2) normal stem cells in the context of AML are induced to remain quiescent and undifferentiated, it may be that a greater proportion of the cells in the $CD34^{+}38^{-}$ gate of these samples are immature, and therefore have lower CD33 expression. A more appealing, alternative hypothesis is that an induced change in CD33 expression on normal HSCs, by either the surrounding AML cells or stroma, results in a change in signalling which drives the observed stem cell quiescence and differentiation block.

One way of attempting to investigate this further is to attempt to manipulate CD33 expression experimentally (either by inducing transient gene overexpression, or knockdown) and investigate how this affects stem cell behaviour and differentiation in vitro.

This is another area where a functional in vitro model of the interplay between AML and normal HSCs would be invaluable (see Section 8.2.3.6). It would be interesting to compare the CD33 expression phenotype of HSCs exposed to AML in vitro, and see if the same patterns are replicated as we have documented in human samples. Such an experiment would not answer whether such changes are due to the influence of the AML directly on CD33, or simply reflect induced quiescence and immaturity of the normal HSC subset. However, subsequently

inducing forced over-expression of CD33 on normal HSCs, and seeing if this negated the effect of the AML on their behaviour might be more illuminating.

One of the potential difficulties in exploring the functional role of CD33 in HSC development is the observed inter-species differences, mentioned in Section 6.1.1.1. As there is no direct homolog between CD33 in humans and mice, investigating the effects of gene deletion in xenograft experiments is not possible. However, it is interesting that in the gene expression profiling experiment previously performed by our group using AML samples transplanted into a murine xenograft model (see Section 7.1.2.1 for experimental detail), one of the genes most markedly downregulated in normal murine HSCs exposed to AML was Sialic acid binding Ig-like lectin E (Siglec E), a member of the CD33 related gene family (log fold reduction in gene expression 2.32 AML versus controls, p=0.00077, Adjusted p value 0.134).

6.5.3 Does the differential expression pattern of CD33 on HSCs have a clinical consequence?

Does the reduced CD33 expression on normal HSCs in the context of AML presentation have a protective effect for these cells against the effect of GO? The results illustrated in Section 6.4.2 would suggest that this is not the case.

6.5.3.1 Clinical trial data suggest that CD33 expression level is not predictive for cytotoxicity

It is logical, in today's era of precision medicine, to hypothesise that the level of antigen expression on a cell would be predictive for the cytotoxic effect of a therapy targeted to that antigen. However, in the case of CD33 and GO, it is worth noting that the clinical data from the MRC-AML trials involving GO suggest this is not the case.

When considering the effect of GO on normal HSCs, it is worth comparing the time to count recovery after chemotherapy, or incidence of cytopenias observed, between the patients who received GO therapy in induction and later in Consolidation 3. As discussed in detail in Section 6.1.2.1.1, those receiving GO at Induction had no significant difference in time to count recovery compared to those who did not. However, the same pattern was also seen in those who were randomised to receive chemotherapy in consolidation 3. If the patterns described in Section 6.4.1.3 with respect to HSC CD33 expression increasing to baseline at remission in patients with CBF mutated AML are replicated across a cohort of wider karyotypic abnormalities, this would imply that HSCs appear unaffected by GO, despite fluctuating levels of CD33 expression.

Secondary evidence comes from the more widely studied effect of GO on AML blasts themselves. Possibly counter-intuitively, the same trial shows no correlation between the level of CD33 expression on blasts and the clinical effectiveness of GO therapy [3,280].

6.5.3.2 Why might HSCs be immune to the effect of GO?

The data illustrated in Section 6.4.2 suggest that at least in this in vitro model, normal HSCs (from AML samples and controls) are not significantly affected by the presence of GO. There exist a number of explanations for this observation, which might either represent a true biological finding, or reflect an inadequacy of the in vitro system used in this experiment.

6.5.3.2.1 Normal HSC expression of CD33 is significantly lower than on blast cells

A lack of clinical effect of GO on the behaviour of normal HSCs (compared to its effects on AML blast cells) might in part be explained by the significantly lower CD33 expression seen on normal HSCs. Figure 6-3 shows the median fluorescence intensity (MFI) of CD33 expression calculated for the 22 AML samples and 9 controls studied in Section 6.4.1. When the MFI is calculated for the subsets of $CD34^{+}38^{-}ALDH^{high}$ cells from both AML and control samples, it is clear that in both cases, this is lower in normal HSCs than in unselected AML blasts (MFI CD33

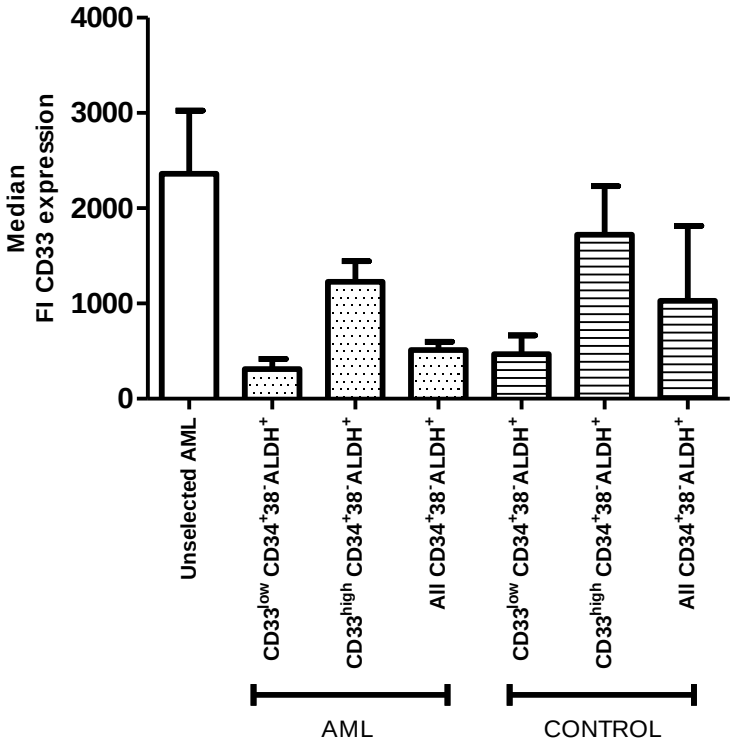

Figure 6-3 Bar chart illustrating the Median Florescence Intensity of CD33 expression of unselected AML cells subfractions of the $CD34^{+}38^{-}ALDH^{high}$ HSCs from AML samples, and the same subfractions of $CD34^{+}38^{-}ALDH^{high}$ HSCs from normal BM. Median Fluorescence Intensity and interquartile range are displayed

expression for unselected AML is 2362; HSCs from AML samples is 513, and HSCs from control samples is 1028). Kruskal-Wallis testing suggests the observed differences between these results are highly significant (p value <0.0001). Subsequent pair-wise analysis, using Dunn's multiple comparison test, suggests that the difference in MFI between unselected AML blasts and HSCs from AML is highly significant (p<0.0001), and between control and unselected AML p=0.028 (which when correcting for multiple testing is just above the level of significance (p<0.025)).

Are these differences significant enough to explain the observed lack of effect of GO on the normal HSCs?

6.5.3.2.2 Are other aspects of normal HSC biology more important than CD33 expression for protection against the effects of GO?

GO, once internalised by cells after CD33 binding, exerts its cytotoxicity by inducing DNA strand breaks in cycling cells[264]. In vitro and in vivo resistance to GO can occur by increased drug efflux by members of the ATP binding cassette protein family, including P-glycoprotein (ABCB1) and multidrug resistance-associated protein1 (ABCC1)[265-268]. Thus stem cells, which cycle relatively infrequently[281], have unique DNA repair mechanisms[282] and reduced apoptotic responses to DNA damage[283], and active drug efflux pathways[284,285], may be extremely well adapted to resist such agents when administered in isolation.

However, new potential strategies to increase the effectiveness of GO by giving it in combination with ABC transporter blockade, might potentially increase its cytotoxicity.

6.5.3.2.3 How could the in vitro experimental model be improved?

Despite the fact that the results from the in vitro experiments, showing resistance of normal HSCs to the effects of GO, do replicate what is observed clinically, it is worth pointing out that the model used could benefit from significant optimisation to be truly confident of any conclusions drawn.

Key to this is the limited availability of GO, which only became available for our use in the last 4 months of this PhD, when the AML-19 trial opened locally. As we were using excess reconstituted GO designed for patient use, experiments (which under normal circumstances would be planned well in advance to allow for growth of sufficient MS5 feeder cells, irradiation access and sorter bookings) were performed at short notice and thus sample numbers were limited. Secondarily, the method of obtaining GO meant there was often a delay between

reconstitution and in vitro use (although this was limited to a maximum of 72 hours). In clinical usage, drug stability is not guaranteed from 24 hours after reconstitution.

Because we only had access to GO three times in total, lengthy optimisation of experimental conditions was not possible. The only published paper involving GO and a LTC assay involved a two hour exposure of unselected cells from the bone marrow of a patient with CML to GO at a dose of 5µg/ml, prior to washing and plating in LTC for 5 weeks. At this dose and with this exposure method, GO significantly reduced the number of CFU-GM cells when compared to untreated controls[286]. I used an exposure dose of 3µg/ml for two hours to pre-sorted cells, as this replicated the mean C_{max} of plasma hP67.6 antibody levels observed in Phase II trials[263] (with GO being administered at a dose of 9mg/m^2, three times higher than that now used in AML-19[126]). However, it could certainly be argued that a two hour exposure time before cell plating (although reflecting the time taken to reach C_{max} clinically) is not sufficient to illustrate the potential effects on slowly cycling HSCs. Pharmacokinetic studies suggests the plasma $t_{1/2}$ is in fact 62 hours. Secondarily, whether the plasma C_{max} is necessarily good guide to the levels of GO in the bone marrow, where stem cells normally reside, is clearly debatable.

Therefore, ideally, a range of concentrations of GO would ideally be assayed for their effect on HSC viability. Secondarily, the effectiveness of the GO (given concerns about reducing stability from reconstitution) should be assayed in parallel with each experiment by testing its ability to reduce the viability of AML blast cells, either from primary samples, or an AML cell line known to be sensitive to is effects.

We had chosen to measure HSC numbers functionally by a 7 week LTC assay, as opposed to purely phenotypically by flow based appearances. An alternative approach, which would allow for a more prolonged drug exposure period, would be to place cells in culture with GO for 48 hours, and then assess cell viability by a flow based assay at the end of this period. Using this approach to assess GO sensitivity in LSCs, Jawad et al found normal $CD34^+38^-$ HSCs were insensitive to the effects of GO (albeit at a much lower concentration of 10ng/ml)[287].

6.5.4 Potential further experimental work

This work has raised a number of unanswered questions, which would be worth exploring with further experimental effort.

We have illustrated that CD33 expression is downregulated on normal HSCs in the context of AML development for both $CD34^{low}$ samples and CBF leukaemias. We have also demonstrated for the $CD34^{low}$ cohort that this change reflects a downregulation in CD33 gene expression;

ideally we would also explore gene expression patterns within the CBF AML cohort to see if the same pattern is observed.

As discussed in Section 6.5.2, it is unclear if the changes we have described in CD33 expression patterns on normal HSCs are directly induced by exposure to AML cells, or whether they simply reflect an increase in the proportion of more immature, quiescent HSCs within the $CD34^{+}38^{-}$ $ALDH^{high}$ population. One way of addressing this would be to explore the effect of transient CD33 over or underexpression on HSCs. Not only would it be interesting to observe the effect this had on HSC division kinetics and repopulation ability, but I would also be keen to study if modified cells were still influenced in the same fashion (in terms of induced quiescence) by the presence of AML.

Despite the observed changes in CD33 expression on normal HSCs in the context of AML development, the clinical evidence from recent AML trials, as well as our in vitro experimental data would suggest that HSCs from both normal and AML bone marrows are relatively insensitive to the effects of GO. However, as anti-CD33 therapy in AML undergoes a current resurgence in interest, it will be interesting to see if new, emerging therapies are associated with prolonged cytopenias and delayed count recovery. In particular, proposed trials looking at GO therapy administered in combination with drug efflux inhibitors might render normal HSCs more susceptible to its effects. As AMG 330 (by nature of its design as a BiTE antibody) is not reliant on cell cycle status, DNA repair mechanisms or drug transporters, it may prove to be more toxic to normal HSCs than GO. Our data suggest the time of greatest differential of CD33 expression between normal HSCs and AML is at diagnosis, and therefore is supportive of the use of anti-CD33 therapies at induction rather than consolidation. Clinical trial data are supportive of this[288,289].

.

Chapter 7 Using RNA-Seq to compare the transcriptomes of normal HSCs from AML and control bone marrow samples

7.1 Introduction

As discussed in Section 1.5.2, previous work from our group using both primary human samples and a xenograft transplant model of AML, suggested that whilst the concentration of HSCs in the bone marrow at diagnosis with AML is preserved, the cells themselves appear to enter a state of quiescence. We have hypothesised that this may be responsible for the corresponding fall in progenitor numbers and observed downstream failure of haematopoiesis[1].

The mechanism by which this occurs is not known. Observational work published over thirty years ago suggests cell-cell contact between AML cells and HSCs is not required[290,291]. Blasts might either affect the cycling behaviour of normal HSCs through a direct signalling pathway, or indirectly by modulating the activity of the stromal cells normally responsible for niche formation[140-143].

The almost universal observation of bone marrow failure associated with the development of AML, despite the known diversity of genetic aberrations in leukaemic populations, would suggest to me that these changes are most likely the result of the modulation of a signalling or homeostatic feedback pathway which is normally involved in the control of HSC behaviour, rather than a new acquired signalling route. Could it be possible that the proliferating blasts are resistant to inhibitory feedback mechanisms that under normal circumstances should prevent their further growth, but might concurrently inhibit normal HSC development?

Hunting for the pathway or pathways responsible for HSC quiescence in this setting has become the holy grail of my PhD over the last three years. Identification of even one of the important factors in this process might reveal a pharmacological target, through which we could reverse this process. Clinically, whilst the prevention of haematopoietic failure in this setting might not provide a long-term cure, it might result in effective disease palliation and a reduced need for transfusional support.

7.1.1 Difficulties with HSC work

A key consideration when planning experimental work in this area is the small number of HSCs available. This is particularly an issue when working with primary human samples, when ethical considerations limit the amount of bone marrow it is possible to extract from each patient.

Personal experience has shown it is possible to extract between 500 and 2000 HSCs from most bone marrow samples stored in the Bart's Tissue Bank. Using techniques modified for low cell numbers, it is possible to study both the DNA and RNA content of these cells. However, this is far less than the minimum material required for proteomic analysis via Western blotting, ELIZA or mass spectrometry. Any attempted amplification of the HSC population after sorting would, by definition, require a reversal of the quiescence we are attempting to study.

7.1.1.1 Experimental models of the AML-normal HSC interaction

In the past, as detailed in Section 1.5.2.1, we have utilised a xenograft transplant model of AML development, where NSG mice act as permissive recipients of primary human AML samples. This approach was used both in our recent publication on HSC numbers at AML diagnosis[1], as well as to date unpublished work involving the analysis of AML-exposed normal murine HSCs by gene expression array (see Section 7.1.2.1). The alternative approach of Cheng et al, who, rather than transplanting primary human AML samples into mice, developed an AML murine model by introduction of the *MLL-AF9* gene into Ly B6-Ly5.2 mice, avoids many of the concerns that the observed changes in HSC behaviour might be due to the transplantion process per se[136].

To date, we have been unable to develop an in vitro model of the interaction between normal HSCs, stromal cells and AML blasts. Over several years, our group in collaboration with members of Dr Bonnet's lab at the LRI, have made strenuous attempts to develop such an assay, but without reproducible success. Further discussion of these attempts is found in Section 8.2.3.6, and the relevant experimental protocols listed in Appendix 2.

Given these limitations, one method of screening for the activation of potentially important signalling pathways, is the study of the comparative transcriptomes of HSCs within AML samples and controls. Such an approach has been attempted by my group and others when studying murine HSCs, but not to date, with primary human samples.

7.1.2 Previous work assessing the HSC transcriptome in the context of AML development

7.1.2.1 Within a murine xenograft model

Unpublished work from my group using a murine xenograft model of AML-induced bone marrow failure has revealed differential gene expression linked to metabolic control and TGF β signalling within the normal HSC subset, when compared to controls.

Seven unirradiated NSG (Nod/Scid/il2-Gamma chain receptor null) mice were transplanted with one primary sample of AML. Once the mice had reached the midphase of AML bone marrow engraftment[1], they were sacrificed along with six untransplanted controls. Murine HSCs were extracted from bone marrow by FACS sorting using CD48, CD150, CD117 and mCD45. Subsequent RNA extraction and whole RNA genome amplification was undertaken, prior to genomic expression analysis using the Affymetrix array platform.

Significant changes in gene expression were detected in over 200 genes, with subsequent pathway enrichment analysis suggestive that many of these were involved in the control of cell metabolism (in particular the PPAR pathway) and the TGF β signalling pathway.

The data from this model clearly have some limitations when translating its relevance to the interaction between human HSCs and AML. Firstly, the cells under investigation are murine, not human HSCs, and there are well documented differences between the mechanisms that control their behaviour. Secondarily, this is an interspecies xenograft transplant model, and some of the induced changes in behaviour may well be due to the transplantation process, rather than an effect solely due to the AML per se (controls were untransplanted mice, rather than mice transplanted with normal human haematopoietic tissue). Finally, these mice were all transplanted with the same sample of AML, previously documented to be capable of causing engraftment within the NSG murine model. Therefore, it is impossible to say if any observed differences in behaviour are due to a common effect from all AML samples, or are specific to this particular sample. There is no representation of disease diversity.

The alternative approach of Cheng et al mentioned above, and in Section 1.5.2.3, involved the generation of a murine model of AML-induced BM failure, with the retroviral-driven introduction of the *MLL-AF9* gene into Ly B6-Ly5.2 mice. Subsequent transplant of murine AML cells into a Ly B6-Ly5.1 murine recipient allows the development of leukaemia without concerns of immune rejection, but provides a means of easy flow-based identification of normal HSCs from leukaemic blasts.

The authors of this paper also performed microarray analysis of the normal HSC populations of these mice, and identified relative upregulation of a number of genes associated with cell quiescence signalling, including Egr3. In a number of subsequent experiments, a role for Egr3 in mediating AML-induced HSC quiescence was mapped out. Retrovirally induced overexpression of Egr3 in murine stem cells led to a quiescent phenotype both in vitro and in a

murine transplant model. The reverse effects were observed when Egr3 expression was reduced in murine stem cells by shRNA knock-down.

7.1.2.2 My own work on primary human samples using targeted pathway analysis

Targeted pathway analysis was used to compare the expression of genes within the PPAR and TGFβ pathways between AML-exposed HSCs and controls. In the interests of brevity, I have chosen to summarise these experiments here, rather than in their own, individual chapter.

These experiments were performed using normal HSCs extracted from primary human AML and control bone marrow samples. In their completion, I gained experience in the accurate sorting of HSCs from BM samples, as well as subsequent efficient RNA extraction from limited material. Knowledge of the observed inter-sample heterogeneity amongst both the HSCs from controls, as well as different AML subtypes, was invaluable for planning the subsequent RNA-Seq work which comprises the bulk of this chapter.

7.1.2.2.1 Rationale for studying the expression of genes within the TGFβ and PPAR signalling pathways

The results from the murine xenograft work detailed in Section 7.1.2.1 highlighted the differential expression of a number of genes which contributed to two pathways: the Peroxisome Proliferator Activator Receptor (PPAR) pathway and the TGFβ signalling pathway.

Peroxisome proliferator-activated receptors are a group of nuclear receptor transcription factors that regulate the expression of genes which control cellular differentiation, development and metabolism. Three PPAR isoforms exist with differential expression across the body: α, β and γ[292]. Activated receptors heterodimerize with the retinoic acid receptor (RXR) in the nucleus, before binding to specific peroxisome proliferator hormone response elements within DNA. PPAR activity can be modulated by a number of coactivators and repressors[293], and in addition, a variety of drug modulators. In humans, hereditary loss of function disorders of all of the PPAR receptors has been reported, largely resulting in metabolic deregulation leading to lipodystrophy, or insulin resistance, rather than loss of bone marrow function[294].

Ito et al assigned a key role for the PPAR β/δ pathway in the maintenance of normal HSC quiescence and differentiation in mice. Conditional deletion of the PPAR δ gene resulted in reduced long term repopulating ability of stem cells; whereas exposure to PPAR δ agonists (GW-501516) increased cobblestoning activity of HSC and progenitor cells. Downstream inhibition of mitochondrial fatty acid oxidation (by the drug etoxomir) reduced the numbers of

HSCs and their ability to repopulate lethally irradiated mice; similarly, the addition of etoxomir to cells negated the effect of the PPAR agonist GW-501516. These data together suggest a key role for PPAR β/δ, modulated downstream by the fatty acid oxidation pathways, in regulating stem cell division and differentiation decisions[62].

TGFβ signalling is known to have a key, multifactorial role in the control of cell proliferation, differentiation and cell death, both under conditions of health, and in the context of cancer. TGFβ is a secreted protein that exists in at least three isoforms[295]. In normal cells, TGFβ signalling results in G1 stage arrest, leading to the cessation in proliferation, differentiation induction, or promotion of apoptosis[296]. Cancer cells can acquire mutations in the TGFβ signalling pathway, leading to insensitivity to its effects[297].

Through analysis of cell lysates from AML samples using western blotting techniques, our group has shown a consistent increase in TGFβ production from AML samples when compared to GMPB (D Taussig, personal communication).

TGFβ signalling has been shown to affect normal HSC behaviour within patients with multiple myeloma. Bruns et al used Affymetrix gene expression profiling on HSC fractions derived from multiple myeloma patients to show prominent changes in expression of genes within the TGFβ signalling pathway, when compared to HSCs from normal controls. Whilst Ki67 staining revealed reduced cycling within stem cell progenitors in patients with myeloma, this effect was reversed by the addition of SD-208, an inhibitor of Smad-2 phosphorylation[298].

The role of the TGFβ pathway in affecting HSC behaviour in patients with AML has not been defined, but it remains an attractive target for further investigation, as it is widely understood, and several pharmacological modifiers of both receptors and downstream signalling targets are available[299].

7.1.2.2.2 Summary of experimental method

Five $CD34^{low}$, *NPM1* mutated AML BM samples and four controls were sorted for $CD34^{+}38^{-}$ HSC populations. Four CBF AML BM samples and four controls were sorted for $CD34^{+}38^{-}ALDH^{high}$ HSC populations. Cells were sorted as previously described, and processed following the method detailed within the Qiagen RT^2 Profiler PCR assay kit.

These assays facilitate the simultaneous qPCR based analysis of expression of 84 genes within a particular pathway. The incorporation of a pre-PCR amplification step allows the analysis of very small initial RNA quantities.

RNA extraction from sorted cells was performed as per Section 2.8.2. For analysis of TGFβ pathway expression, the RT² Profiler™ PCR Array Human TGFβ Signalling Targets (Qiagen, Cat: PAHS-235Z) was used. For analysis of PPAR target expression, the RT² Profiler™ PCR Array Human PPAR Targets (Qiagen, Cat: PAHS-149Z) was used. After cDNA synthesis, samples underwent pre-amplification for the cDNA target templates, followed by loading onto 384 well plates prefilled with the appropriate primers for all 84 gene targets.

Analysis of the resultant qPCR data was performed as per Section 2.8.4.5.4.

7.1.2.2.3 Summary of results

7.1.2.2.3.1 PPAR Pathway analysis

Initially, the data from all 8 AML samples and 8 controls were analysed together. Of the 84 genes within the PPAR pathway, 10 showed upregulation (relative expression greater than two) in those HSCs sorted from AML samples when compared to those from control samples, and 12 showed downregulation (relative expression less than 0.5). However, the variability in expression between individual samples meant that only one of these results obtained statistical significance: the expression of malonyl CoA decarboxylase, which was downregulated in stem cells derived from AML samples as compared to controls.

Gene expression within the HSCs from CBF leukaemias (n=4) was then compared independently to matched controls (n=4). Four genes showed statistically significant alterations in expression: Cytochrome p450, 3-hydroxy-3methylglutaryl-CoA synthase 2 (*HMGCS2*), Peroxisome Proliferator Activated Receptor Alpha (*PPARA*) and Retinoid X Receptor Gamma.

By comparison, when the gene expression within the HSCs derived from $CD34^{low}$ leukaemias (n=4) was compared independently to matched controls (n=4), two genes were significantly downregulated in AML-exposed HSCs: Enoyl CoA hydratase/3hydroxyacyl CoA dehydrogenase and E1A binding protein p300.

7.1.2.2.3.2 TGFβ Signalling Pathway analysis

When all AML samples and controls were analysed together as a single cohort, 20 genes showed upregulation (relative expression greater than two) in AML exposed HSCs, and 7 showed down regulation (relative expression less than 0.5). However, as before, the intersample variability in results meant that only four of these changes obtained statistical significance: Epithelial membrane protein 1 (*EMP1*), FBJ murine osteosarcoma viral oncogene

homolog (*Fos*), Ras Homolog member A (*RhoA*) and Ras homolog member B (*RhoB*) all showed statistically significant upregulation when compared to control samples.

Gene expression within the HSCs from CBF leukaemias (n=4) was then compared independently to matched controls (n=4). Expression of the Aryl Hydrocarbon receptor interacting protein-like 1 (*AIPL1*), plasminogen (*PLG*) and *EMP-1* were significantly increased in AML-exposed HSCs, and expression of Inhibitor of DNA binding 1 (*ID1*) significantly downregulated in AML-exposed HSCs.

Subgroup analysis of mRNA extracted from stem cells from AMLs of the $CD34^{low}$ phenotype (n=4) also showed a small but significant rise in expression of the cyclin dependent kinase inhibitor 1B (*p27*).

7.1.2.2.4 Validation attempts

I subsequently attempted to validate the changes observed and selected one of the five genes shown to be differentially expressed when both subtypes of AML were analysed together- *RhoA*. This was partly a pragmatic decision based on the availability of commercial antibodies of proven validity, and partly because of a recent publication from the Bonnet lab had suggested that *RhoA* upregulation in a lentiviral transfection model led to cord stem cell quiescence[300].

Two methods of assessment of protein expression in HSCs were attempted. Sorted HSCs (both $CD34^{+}38^{-}$ cells from $CD34^{low}$ AMLs and $CD34^{+}38^{-}ALDH^{high}$ cells from CBF AMLs) were placed on slides, fixed and stained for RhoA expression using an anti RhoA mouse-derived IgG1 antibody (AbCam, Cat: ab54835).

Secondarily, RhoA expression was studied using the same antibody to stain fixed cells from unsorted bone marrow samples, prior to flow cytometric based analysis. Samples were analysed for the proportion of $CD34^{+}38^{-}$ cells that were positive for RhoA.

This work proved technically challenging, as it required intracellular staining with unconjugated, primary anti-RhoA antibody, followed by secondary staining with a Goat Anti-Mouse antibody, conjugated to PE (BD, Cat: 550589). The necessity for cell fixation prior to staining for intracellular RhoA expression precluded the use of CBF AML samples, where the vital stain ALDH is necessary for accurate distinguishing of normal cells from the leukaemia clone.

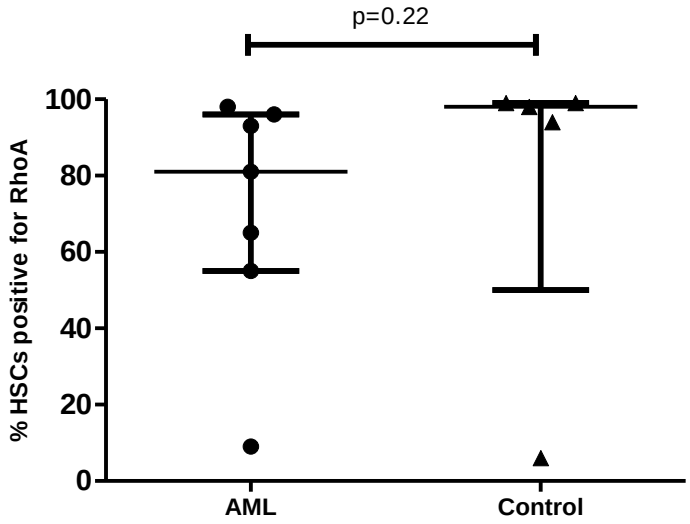

Figure 7-1 illustrating RhoA expression within the $CD34^{+}38^{-}$ HSCs of AMLs and controls. Median percentage of RhoA positive cells and interquartile ranges are plotted.

Having optimised this flow-based technique, I then assessed RhoA expression patterns on seven $CD34^{low}$ AML bone marrows and five controls. Populations were analysed for the percentage of $CD34^{+}38^{-}$ cells which were RhoA positive, based on isotype-stained controls. These data are illustrated graphically in Figure 7-1. The median percentage of RhoA positive cells within the $CD34^{+}38^{-}$ HSC population of $CD34^{low}$ AML samples was 81% (IQR 55 to 96%, n=7); for controls, this figure was 98% (IQR 50-99%, n=5). There is no significant difference between these values, as determined by Mann-Whitney t-test (p=0.22).

7.1.2.2.5 Summary

The experience of these experiments taught me several things of relevance to the RNA-Seq work that followed.

One of the strengths of this work is that it is performed on primary human samples, representing a variety of different AML subtypes. This avoids all of the inherent concerns with using a murine xenograft model summarised in Section 7.1.2.1.

However, one of the difficulties with the use of primary samples is the significant inter-sample heterogeneity in gene expression, which resulted in few genes reaching statistical significance in terms of differential expression. This was apparent even within the control samples used. Interestingly, it also appeared that there were differences between the patterns of gene expression observed between the HSCs from CBF and $CD34^{low}$ AMLs. Although the observed changes might not have any demonstrable effect of stem cell quiescence or differentiation, this was an unexpected result. The obvious response to such inter-sample variability is to

substantially increase the number of samples studied, but we were limited both by cost implications and tissue availability within the Bart's Tissue Bank.

In retrospect, the experiments would have been better designed if I had adopted a consistent approach to sorting HSCs across all the AML subtypes tested. CD34 and CD38 expression are sufficient to separate out normal HSCs from $CD34^{low}$ AML cells, but the more rigorous $CD34^{+}38^{-}$ $ALDH^{high}$ selection is required when analysing CBF mutated samples. Although the HSCs from both types of AMLs were analysed against control samples sorted by identical criteria, it is certainly possible that some of the differences in gene expression observed between the HSCs from the AML subtypes could be due to the slightly different sorting strategies employed, rather than an inherent biological distinction.

Finally, despite the limitations detailed above, the results of this experiment did highlight the lack of parallels between results generated by a murine xenograft model and primary human samples. As we believe the use of actual human samples is far more likely to represent relevant translational biology, this reinforced my desire to continue to work with these in further experiments.

7.1.3 RNA-Seq

7.1.3.1 Background

The technique of RNA-Seq (also known as WTSS or Whole Transcriptome Shotgun Sequencing) harnesses the ability of Next Generation Sequencing platforms to detect the RNA present within cells. It is most commonly used to study the mRNA transcripts within cells, but experimental techniques can be modified to focus on non-coding RNA such as miRNA[301] and tRNA[302]. Its sensitivity is such that it can be used to study a huge dynamic range of starting material- from single cells, to millions of pooled cells.

7.1.3.2 RNA-Seq versus Microarray

The dominant technology until recently for whole transcriptome analysis of cells involved the use of microarrays, which allowed the simultaneous detection of multiple gene transcripts. Microarray chips are coated with thousands of probes specific for particular genes, in tightly specified locations. A test sample of cDNA is pre-treated with chemoluminescent dyes that bind to the cDNA, and emit light on hybridisation. On probe binding to complementary cDNA, hybridisation is detected by a chemoluminescent signal, which is proportional in intensity to the degree of binding, and therefore the amount of the target within the sample. After

washing of the chip to remove non-specifically bound targets, the fluorescence of the chip is read.

In recent years, RNA-Seq based approaches have largely superseded microarrays as the technique of choice for parallel sequencing analysis. The technology behind RNA-Seq offers a number of improvements on previously available methods. Microarrays require detailed prior knowledge of the genome, so that primers can be produced for all known genes under investigation. By contrast, RNA-Seq requires no prior knowledge of the genome, making it ideal for the studies of rare species. This also enables the detection of novel genes, mutated transcripts (particularly of relevance for cancer work), rare allelic variants and splice variants. Finally, as the detection of transcripts is not dependant on chemoluminescence, RNA-Seq offers a wider dynamic range, in the absence of significant background signal[303].

7.1.3.3 RNA-Seq experimental design for small cell inputs

7.1.3.3.1 RNA extraction and cDNA synthesis

At the time of experimental design, several commercial kits optimised for the extraction of RNA and cDNA synthesis from small numbers of cells prior to library preparation were available.

The SmartSeq v4 Ultralow RNA Kit for Sequencing (Clontech, Cat 634889) is marketed for the production of cDNA from RNA or cells, for inputs of between 1 and 1000 cells. mRNA is selected by the use of primers specific for the PolyA tail of transcripts. The subsequent use of SMART® technology, which relies on the template switching activity of reverse transcriptases, enriches for full-length cDNAs and adds PCR adapters directly to both ends of the first-strand cDNA[304]. This ensures the final cDNA libraries contain the 5' end of the mRNA, and hopefully maintains a true representation of the original mRNA transcripts. This process is illustrated in Figure 7-2.

Market competitors include the Ovation V.2 RNA-Seq kit (Nugen, Cat 7109), which also offers cDNA synthesis and amplification from small RNA inputs (500pg upwards). In this case, the manufacture of cDNA involves the use of primers both specific to the PolyA tail of mRNA transcripts as well as random sequences, and therefore allows for production of transcripts from both polyadenylated and non-polyadenylated RNA.

Our final choice of the Clontech SmartSeq kit was based on discussions with other users in Oxford, UCL and BCI, who all had experience of using this kit in experiments with limited cell numbers.

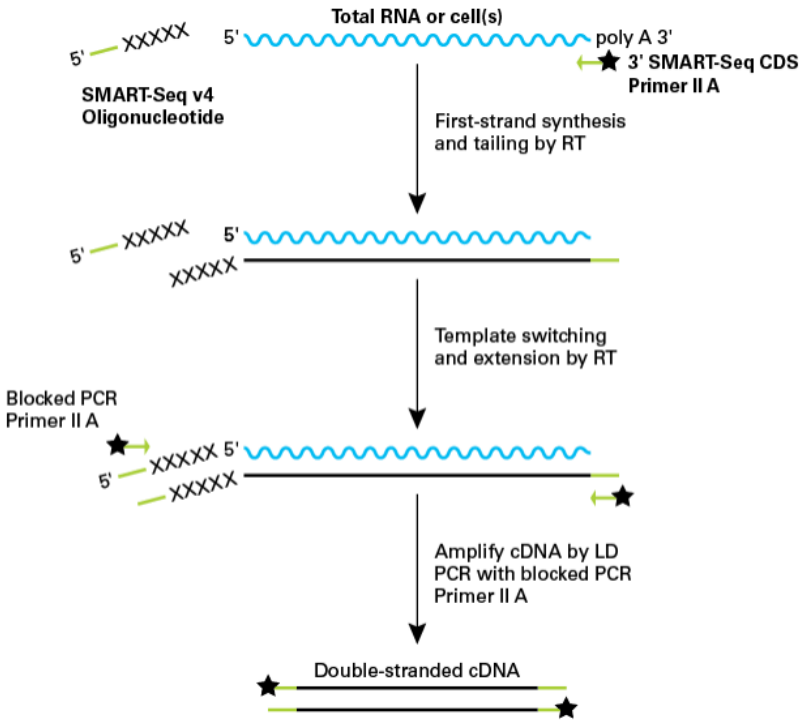

Figure 7-2 The method of mRNA selection and cDNA synthesis involved with use of the Clontech SmartSeq kit[305]

7.1.3.3.2 Library Preparation Options

Library preparation prior to NGS involves the fragmentation of DNA, followed by the fusion of the resultant fragments to adaptors capable of immobilising the strands onto a solid surface during sequencing. Initial fragmentation of DNA strands can be performed either by mechanical shearing, or enzymatic breakdown of cDNA[306].

Commonly, libraries are amplified by PCR, and undergo size selection of fragment strands prior to sequencing.

The Nextera XT DNA library preparation kit (Illumina, Cat FC 131-1024) is specifically designed for efficient library preparation from very small DNA or cDNA inputs (1ng per sample). Efficiency is increased by the use of an engineered transposome to simultaneously fragment and tag the input DNA in random places with adapter sequences. Subsequent PCR uses the adapter sequences to amplify the insert DNA. The PCR reaction also adds index sequences on both ends of the DNA. This process is illustrated in Figure 7-3.

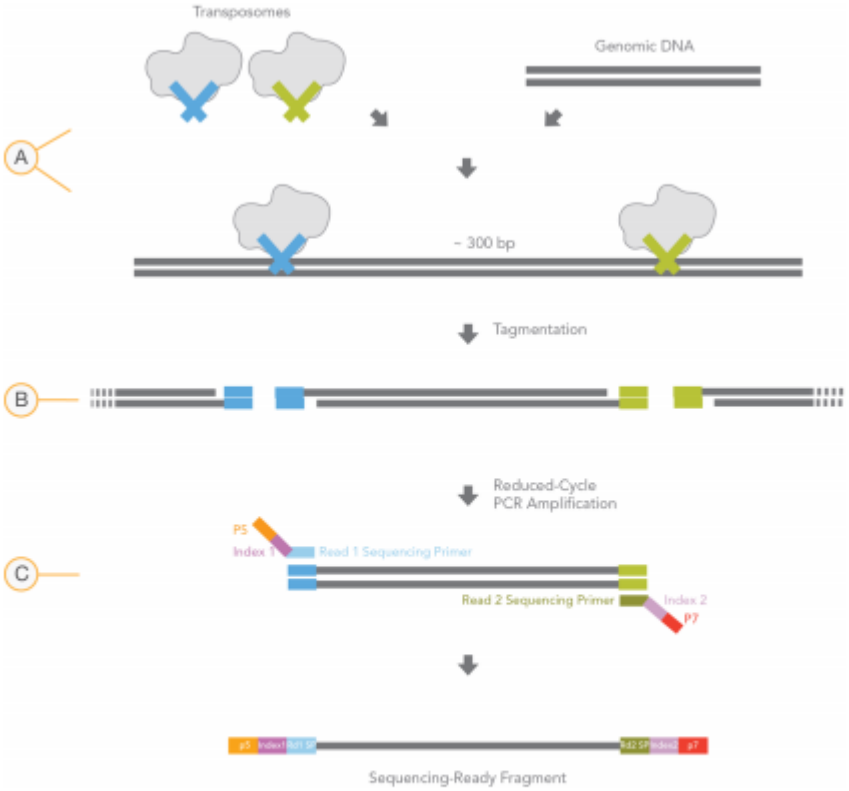

Figure 7-3 The process of DNA library production utilising the Nextera XT kit[307]

7.1.3.4 Next Generation Sequencing

The process of NGS sequencing is briefly described here. In the first step of cluster generation, the DNA library is added to the Sequencer, where the fragments bind to a cell lined with sequences complementary to their adapters. Each fragment is amplified, and distinct clonal clusters produced.

Subsequent sequencing involves the incorporation of fluorescently labelled dNTPS by DNA polymerase to a single strand of template DNA in sequential cycles. The binding of each dNTP releases a fluorescent signal which can be detected. Simultaneous sequencing of multiple strands in parallel allows collection of huge data sets for analysis[308]. Data regarding sequence are stored and then analysed via a Bioinformatics pathway.

In planning a sequencing experiment, a number of choices have to be made. Key decisions include the depth and length of reads required, as well as a choice of single or paired end reads. Allowing for a greater number of reads per sample comes at a higher cost, but allows for the detection of infrequently expressed genes, as well as mutations within transcripts. The

choice of read length is often defined by the provider of the sequencing platform, but the advantage of longer reads is an increased ease of unambiguous mapping to a specific position within the genome. The advent of paired end sequencing allowed sequencing from both ends of a DNA strand, and the automatic association of the data generated. This approach facilitates better read alignment, which is of particular value in repetitive regions of the genome.

7.2 Aims and Objectives

The work described in this chapter aimed to use transcriptome analysis to identify key regulators responsible for the observed quiescence of HSCs exposed to AML. In doing so, I hoped to identify potential therapeutic targets with the aim of reversing this process.

This first objective of this chapter was to sort HSCs from human primary $CD34^{low}$ AML BM samples and matched controls by the techniques already described in this thesis, and study to their transcriptomes via RNA-Seq analysis. This required the generation of cDNA from the RNA of sorted cells, followed by the production of DNA libraries.

A first experimental attempt generated poor quality data, with a significant number of sequencing reads mapping to non-coding DNA regions. As a result, the entire experiment was repeated with some substantial methodological changes.

The secondary objective was to validate the expression changes of genes of interest either by qPCR, or by flow-based assessment of protein expression where possible.

7.3 Attempt 1: Library Preparation involving cDNA synthesis directly from cell lysate

7.3.1 Specific Methods

7.3.1.1 Samples

Six $CD34^{low}$ and six control samples were chosen for transcriptome analysis. The decision to study only $CD34^{low}$ AMLs samples was based on the previous work on targeted pathway analysis summarised in Section 7.1.2.2, which suggested there may exist differences in gene expression patterns between the HSCs from different AML subtypes. As budget limitations meant the number of samples to be sequenced was relatively restricted, we decided to sequence the HSCs within a hopefully homogenous group of AMLs. Control BM numbers were matched, as the same previous work suggested these also exhibited relatively heterogeneous gene expression profiles.

Sample details for the six AML samples are shown in Table 7-2. As in previous work, control samples came from patients with no prior chemotherapy exposure, blood counts within normal limits and a normal BM appearance by morphology, flow and immunohistochemistry.

7.3.1.2 Sorting

Samples were thawed, washed and cell counted.

As $CD34^{low}$ AML samples had been chosen for this experiment, we could have separated normal HSCs on the basis of a $CD34^{+}38^{-}$ signature alone. Instead, we chose to use the selection criteria for normal HSCs described in Chapter 4: namely $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$. Despite the increased experimental costs this incurred, using this more rigorous phenotype for the selection of HSCs meant that if any findings were to be validated in a different subgroup of AML samples (such as the CBF leukaemias), their HSCs could be separated in the same fashion.

Samples were stained for ALDH expression as previously described. Staining with HAG dissolved in ALDH buffer was followed with staining for CD34-PerCP, CD38 PeCy7 and CLL1-PE at a concentration of 5µl/1.5x10^{6} cells. Cells were resuspended into ALDH/DNase/DAPI, and the $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells sorted into ALDH buffer. 1100 cells were sorted for each sample, and 100 cells utilised for a purity check at the end of sorting. Cells underwent resorting if purity fell below 90%.

Sorted cells underwent cell lysis and immediate cDNA synthesis, followed by amplification and clean-up, utilising the SmartSeq v4 UltraLow Input RNA Kit (Clontech, Cat 634889). The full

protocol is listed in the associated manual, but the completeness, details of the protocol are given below. Given low cell numbers, all sample processing was performed under a flow hood to reduce the possibility of contamination.

Sorted cells in a FACS tube were spun (3000rpm, 10 minutes, 4^{o}C) and the supernatant aspirated using a Gilson machine. The pellet was resuspended by flicking the tube, and the residual contents transferred to a 200µl eppendorf tube (MicroAmp Reaction Tube and Cap, AB, N801-0612). 100µl PBS was added to the FACS tube, vortexed and the liquid then transferred to the eppendorf, to transfer any residual cells and remove any residual ALDH buffer which might impede subsequent cell lysis.

7.3.1.3 RNA prep and cDNA synthesis

7.3.1.3.1 Cell Lysis and Direct RNA to cDNA synthesis

The 200µl eppendorf was spun in a microcentrifuge (3000rpm, 10 minutes, 4^{o}C), and the supernatant carefully fully removed using a Gilson aspirator loaded with a 10µl pipette tip. 9.5µl nuclease-free water was then added, and the tube vortexed to resuspend the pellet.

Cell lysis was then performed as follows. A stock solution of 10x Reaction Buffer was made by mixing 19µl 10x Lysis buffer with 1µl RNase Inhibitor to a total volume of 20µl. 1µl of Reaction buffer was added to each test sample, which was briefly vortexed, and then incubated at room temperature for 5 minutes. A negative control of 1µl Reaction Buffer and 9.5µl nuclease-free water was processed simultaneously.

Samples were placed on ice, and 2µl 3'SMART-Seq CDS Primer IIA added to each. Samples were briefly vortexed, and spun down, and then incubated at 72^{o}C for 3 minutes in a preheated thermal cycler. Samples were then placed immediately on ice for 2 minutes.

cDNA synthesis master mix was then made up in the following proportions: 4µl 5x Ultralow First Strand Buffer; 1µl SmartSeq v4 Oligonucleotide and 0.5µl RNase Inhibitor. 2µl per sample of Reverse Transcriptase was added to the master mix immediately before use.

The master mix was gently vortexed, and 7.5µl added to each reaction tube, which was then gently mixed, spun down and placed in a PTC-225 Peltier Thermal Cycler for the following PCR regime: 42^{o}C for 90 mins; 70^{o}C for 10 minutes, followed by holding at 4^{o}C.

7.3.1.3.2 cDNA Amplification by LD PCR

The resultant cDNA was amplified by the following methodology, optimised for the quantity of cDNA produced from these cells.

PCR master mix was prepared in the following proportions: 25µl 2xSeqAmp PCR Buffer; 1µl PCR Primer IIA; 1µl SeqAMP DNA Polymerase and 3µl Nuclease Free Water.

The master mix was briefly vortexed and spun down. 30µl of PCR master mix was then added to each tube containing 20µl first strand cDNA product. The samples were vortexed briefly, spun down and then placed on a PTC-225 Peltier Thermal Cycler set to the following program: 95^{o}C for 1 minute; 98^{o}C for 10 seconds, 65^{o}C 30 seconds, 68^{o}C 3 minutes for a total of 10 cycles; 72^{o}C for 10 minutes; hold at 4^{o}C.

7.3.1.3.3 Purification of Amplified cDNA using the Agencourt AMPure XP Kit

PCR products were transferred to a 1.5ml eppendorf tube. 1µl 10x Lysis Buffer was added to each sample. AMPure XP beads (Agencourt AMPure XP, A63881) (having been allowed to equilibrate to room temperature for 30 minutes before the start of the experiment) were then vortexed for 30 seconds to ensure even distribution of beads, before adding 50µl beads to each sample. Sample and beads were then mixed carefully by pipetting up and down at least 10 times. Samples were incubated for eight minutes at room temperature to allow cDNA to bind to the beads. Tubes were briefly spun down, before placing on a magnetic separation device (Dynamag^{TM}2/Magnetic Particle Concentrator, Invitrogen) for five minutes until the supernatant was completely clear. Leaving the tube on the magnet, the supernatant was removed using a pipette, and 200µl of freshly made 80% ethanol was added without disturbing the beads. After 30 seconds, the supernatant was removed, and the process repeated once. Samples were briefly spun down, and then replaced on the magnet for 30 seconds. Any residual traces of ethanol were removed by pipette.

Eppendorf tubes were left with their lids open for 2.5 minutes, to allow samples to dry. 17µl of elution buffer was then added to the pellet, which was gently pipetted to resuspend fully. The sample was then incubated at room temperature for two minutes to thoroughly rehydrate. Samples were briefly spun, and then replaced on the magnet for one minute until the supernatant was completely clean. The clean supernatant was then removed by pipette, and stored at -20^{o}C.

7.3.1.3.4 Sample QC and quantification

Sample quality control was performed at the Genome Centre, Bart's Cancer Institute.

Successful cDNA synthesis and amplification is predicted to yield no product in the negative control, and a distinct peak spanning 400bp to 10000bp, peaking at 2500bp in the positive control and yielding between 3.4-17ng of cDNA. The optimal methodology for assessing a sample's suitability for further processing would involve the use of the Agilent 2100 BioAnalyser, combined with Agilent's High Sensitivity DNA Kit (Agilent, 5067-4626). This is of the appropriate sensitivity for the predicted yield of cDNA, and covers the expected sample size range. Unfortunately, this was not available in the Institute, and therefore on the advice of staff in the Genome centre, the DNA D1000 Tape Station High Sensitivity (35 to 1000bp, sensitive to 5pg/µl), loaded onto an Agilent 4200 Tape Station, was used to assess samples (Agilent 5067-5584). "Successfully" processed samples were identified by visualisation of the start of the cDNA peak visible between 500 and 1000bp (see Figure 7-4 Aii). In retrospect, given the difficulties with the quality of the data generated with the first RNA-Seq experiment, this approximation may have been significant. It meant we were unable to make any assessment of the quality of the cDNA generated, as we could only see the smallest strands generated. In theory, a significantly degraded sample might generate some cDNA of under 1000bp length, but no peak in the 1000 to 10,000 range where full length mRNA transcripts would be expected to lie.

Since the first experiment was completed, Agilent have produced the HS D5000 DNA ScreenTape (Agilent 5067-5592), which is much more appropriate for assessment of the cDNA libraries generated by the SmartSeq kit. It covers a size range of 100 to 5000bp, and is also sensitive to a concentration of 5pg/µl. This was used in for QC of samples generated in the second RNA-Seq attempt (see Section 7.4.1.3.4).

7.3.1.3.4.1 Quantification using the Qubit dsDNA HS Assay Kit

Quantification of cDNA libraries generated by the SmartSeq kit was performed using the Qubit HS DNA reagent kit (ThermoFisher Q32851), which offers a sensitivity range of 0.2-100ng/µl.

A working solution of a 1:200 dilution of Qubit dsDNA HS Reagent with dsDNA HS Buffer was prepared fresh, and mixed well. 190µl of working solution was added to two plastic analysis tubes (provided with kit), and 10µl of kit DNA Standards 1 and 2 were added to each respectively. For each test sample, 1µl cDNA was added to 199µl of working solution in a fresh tube, prior to vortexing for three seconds. All samples and standards were left to equilibrate for 3 minutes at room temperature before analysis.

Analysis was performed using the Qubit 3.0 fluorometer. Calibration was first performed by analysis of the two standards diluted in working solution, followed by analysis of the samples.

7.3.1.3.5 Secondary Clean-Up

All test samples on initial analysis contained a small cDNA peak of 60bp size, thought to represent excess adaptor (see Figure 7-5). This was removed prior to further analysis by repeat cleaning of the cDNA sample, using a modified version of the above method with AMPure XP Beads. The residual sample (15µl) and AMPure XP beads were mixed in a 1:1 volume ratio (this ratio is optimal to remove DNA strands of less than 100bp length). Samples were then processed, washed with ethanol and eluted in 17µl of buffer as already described. Quantification using the high sensitivity Qubit test was repeated, as was visualisation by HS D1000 DNA ScreenTape.

The clean supernatant was then removed by pipette, and stored at -20^{o}C.

7.3.1.4 DNA Library Preparation

DNA Libraries were prepared using the Nextera XT DNA Library Preparation Kit (Illumina, FC-131-1024). The protocol is summarised below. All reagents are included within the kit, unless where specifically stated.

With each batch of processed samples, a negative control (compromising of the negative control processed through the SmartSeq cDNA synthesis process) also underwent the same library preparation process.

7.3.1.4.1 Index Selection

As 12 libraries were to be pooled together and sequenced simultaneously, the Illumina Low Plex Pooling Guidelines were followed when choosing compatible Index 1 (i7) and Index 2 (i5) adapters. Index sequences need to be complementary to ensure appropriate registration of samples during the sequencing process: as the sequencer uses a green laser to sequence G and T bases, and a red laser to sequence A and C bases, at each cycle at least one of two nucleotides for each colour channel needs to be read.

From within the 24 sample Nextera XT DNA Library Preparation Index kit (FC131-1001), a combination of N701 to 706 (Index 1 (i7)) and S503 to 504 (Index 2 (i5)) were identified as compatible with the pooling and simultaneous analysis of 12 samples.

7.3.1.4.2 Tagmentation of Input DNA

10µl Tagment DNA buffer was added to a 200µl eppendorf tube, followed by 5µl of the amplified cDNA (prediluted to a concentration of 0.2ng/µl). Samples were gently mixed by pipetting. 5µl of Amplicon Tagment Mix was then added to each eppendorf, and samples again mixed by pipetting. Tubes were then centrifuged (280g, 1 minute, 20^{o}C), and placed on a PTC-225 Peltier Thermal Cycler for the following program: 55^{o}C for 5 minutes, followed by 10^{o}C forever.

As soon as the samples reached 10^{o}C, the transposome was immediately neutralised by the addition of 5µl Neutralise Tagment Buffer to each well. Samples were mixed by pipette and centrifuged (280g, 1 minute, 20^{o}C), prior to incubation at room temperature for five minutes.

7.3.1.4.3 PCR Amplification

Tagmented DNA was subsequently amplified by limited PCR. Simultaneously, indexes were added to barcode each sample before pooling for sequencing.

15µl Nextera PCR master mix was added to each reaction well. 5µl of the appropriate i2 primer was added to each sample, followed by 5µl the appropriate i1 primer. Samples were mixed by pipetting. Tubes were centrifuged (280g, 1 minute, 20^{o}C), and then placed on the PTC-225 Peltier Thermal Cycler for the following program: 72^{o}C for 3 minutes; 95^{o}C for 30 seconds; 95^{o}C for 10 seconds, 55^{o}C for 30 seconds, 72^{o}C for 30 seconds for a total of 12 cycles; 72^{o}C for 5 minutes; hold at 10^{o}C.

7.3.1.4.4 PCR clean up

PCR clean-up of the library was then performed, using a bead to sample ratio designed to remove excess short fragments, which are preferentially read during the sequencing process.

Tubes were centrifuged (280g, 1 minute, 20^{o}C), and samples transferred to 1.5ml eppendorf tubes. 90µl Agencourt AMPure XP beads (having been brought to room temp for 30 minutes and resuspended by thorough vortexing) were added to each sample, and mixed by pipetting 10 times. Tubes were incubated for five minutes at room temperature to allow DNA binding to beads. The tubes were placed on a magnetic stand for two minutes until the supernatant had cleared. With the tube remaining on the stand, the supernatant was removed by pipette and the sample washed with freshly made 80% ethanol as follows. 200µl 80% ethanol was added to the tube on the magnetic stand, left for 30 seconds, and then removed by pipette. This process was repeated a second time. The sample was then left to air-dry on the stand for 15 minutes. The tubes were then removed from the magnet, and 52.5µl Resuspension Buffer

added to each tube before resuspension by gentle repeated pipetting. The resuspended beads were then incubated at room temperature for two minutes to allow for DNA elution, before placing back on the magnet for a further two minutes until the supernatant had cleared. 50µl of supernatant was then removed and stored at -20^{o}C until further notice.

QC of the resultant DNA library was performed by repeat analysis on the HS D1000 DNA Screen Tape, and the resultant DNA concentration of the library confirmed by Qubit analysis. Optimal libraries should contain a range of fragment sizes between 250 and 1000 bp.

7.3.1.4.5 Quantification

7.3.1.4.5.1 Using Cubit and HS D1000 DNA Screen Tape

Quantification of the DNA concentration of the library was performed as follows. The average size of the fragments contained within the library was determined by integration under the trace generated by the HS D1000 DNA ScreenTape. Library concentration was then calculated using the following formula:

$$Library\ concentration\ (nM) = \frac{Concentration\ as\ calculated\ by\ Qubit\ \left(\frac{ng}{\mu l}\right) x10^{6}}{average\ fragment\ length\ (bp) x650}$$

7.3.1.4.5.2 Using Kappa Primers

Individual libraries were also quantified using the KAPA Library Quantification Kit for Illumina® platforms (KAPA Biosystems, KK 4824). This contains primers designed to target the Illumina P5 and P7 flow cell oligo sequences, and therefore, when used in combination with DNA standards, provides an absolute quantification of the amount of DNA tagged for sequencing.

7.3.1.4.5.2.1 Library Dilution and Generation of a Standard Curve

Six DNA standards were provided with a range of concentrations between 0.0001pM and 20pM.

Libraries were pre-diluted to fall within the dynamic range of the assay. To reduce potential error associated with pipetting errors, two independent dilutions into 10mM Tris-HCl, pH 8.0 (to concentrations of 1/10000 and 1/20000) were made for each library. Each dilution was made in two serial steps (1/10000: 1µl undiluted library into 99µl buffer, followed by a separate dilution of 1µl into 99µl (1/10,000); 1/20000: 1µl into 99µl, followed by 1µl into 199µl). Samples were thoroughly vortexed between each dilution. Diluted libraries were kept on ice to increase their stability.

7.3.1.4.5.2.2 Reaction Set Up

6µl KAPA SYBR®FAST and Primer Premix and 2µl H_20 per sample were combined to make PCR master mix. 8mcl of PCR master mix was added to each well, followed by 2µl of DNA library or DNA standard, to a total well volume of 10µl. Each concentration was assayed in triplicate. 384 well, clear white wells (BioRad HSP 3805) were used and sealed with compatible seals (Microseal B Seals, BioRad MSB1001). After sealing, the plate was centrifuged (1500rpm, 1 min).

The following analysis cycle was run on the BioRad C1000 Touch Thermal Cycler: 95°C for 5 minutes; 95°C for 30 seconds, 60°C for 45 seconds for a total of 35 cycles; melt curve 65°C to 95°C, with 0.5°C increments every 5 seconds.

7.3.1.4.5.2.3 Data Analysis

A standard curve was calculated using the results generated from the DNA standards. This was used to convert the average Cq score for each dilution of each library to an average concentration (pM). The average size-adjusted concentration for each dilution was calculated as follows:

$$average\ size\ adjusted\ concentration\ (pM) = calculated\ average\ concentration\ (pM) x \frac{size\ of\ DNA\ standard\ (454\ bp)}{average\ fragment\ length\ of\ library\ (bp)}$$

The size adjusted concentration was then multiplied by the dilution factor to calculate the final concentration of the undiluted library. As for each library, two dilutions were assayed; the average of these was taken to calculate the working concentration used for subsequent pooling.

7.3.1.4.6 Library Pooling

The Library Normalisation step, utilising beads as described in the Nextera XT handbook, was not used. This decision was taken after discussion with the Oxford Genomics Centre and Illumina, who felt that the concentration yielded by the Bead Normalisation approach would not be sufficient for samples being sequenced on the HiSeq 4000. Therefore Library normalisation was undertaken manually using library concentrations generated by Kappa analysis.

7.3.1.5 Sequencing

Library sequencing was undertaken at the High Throughput Genomic Centre at Oxford. The pooled library was run over two lanes of an Illumina HiSeq 4000, with paired end reads of 75bp. This was calculated to give an average read depth of 40 million reads per sample.

7.3.1.6 Bioinformatic Analysis

Sequencing reads were aligned to the human genome build hg38/GRCh38 with the HISAT2 aligner. Transcript quantification was performed with HTSeqt part of the HTSeq package (version 0.6.1p1)[309], using GENCODE v23 human gene annotation (Ensembl release 81).

The read count data were filtered to keep genes that achieve at least one read count per million (cpm) in at least three samples.

Reads per kilobase per million mapped reads (RPKM) values were calculated with the conditional quantile normalisation (cqn) counting for gene length and gc content in the R statistical environment via Bioconductor packages[310].

Expression data were analysed using LIMMA (http://bioconductor.org/packages/release/bioc/html/limma.html) to fit a linear model to the expression data for each gene to detect differentially expressed genes between two groups (AML versus control samples). Differential expressed genes were gauged using LIMMA empirical Bayes statistics module. The differentially expressed genes are selected when p-value is less than 0.05 and the absolute value of log fold change is more than 1.

7.3.1.7 Validation Techniques

7.3.1.7.1 qPCR based validation

7.3.1.7.1.1 Sample selection

Excess sorted HSCs for both AML and control samples used for RNA-Seq underwent immediate RNA extraction using the Qiagen MicroColumn method detailed in Section 2.8.2, prior to storage at -70^{o}C. In several cases, insufficient HSCs were extracted to allow for both RNA-Seq analysis, and separate RNA extraction, and no further vials from the same patient were stored in the Tissue Bank. Therefore, an additional five control samples were sorted at a later date, using the same sorting and extraction methodology. Table 7-9 shows details of the samples used for qPCR based validation of the RNA-Seq data.

7.3.1.7.1.2 cDNA synthesis

First strand cDNA synthesis from RNA was performed using the Superscript®VILO™ cDNA synthesis kit (Thermofisher, Cat 11754050), supposedly optimised for maximising efficiency of RNA conversion.

cDNA synthesis was performed as follows. Into a single eppendorf tube the following reagents were mixed: 4µl 5xVILO™ Reaction mix; 2µl 10xSuperscript Enzyme mix; RNA 1 to 14µl as required; H_2O to a total volume of 20µl. The tube was gently pipetted and placed on a thermal cycler for the following PCR reaction: 25°C for 10 minutes; 42°C for 120 minutes; 85°C for 5 minutes; hold at 4°C.

7.3.1.7.2 For Taqman based probes

Master mix was made up in the following proportions for each sample: 1µl 20xProbe/Primer set; 10µl Taqman 2xPCR Master mix (Thermofisher, Cat: 4304437); 2µl cDNA; 7µl H_2O. Each sample was set up in triplicate, along with appropriate negative controls. Plates were sealed and centrifuged at 1500rpm for one minute.

Samples were run on a AB StepOne Plus analyser on the following cycle: 50°C for 2 minutes; 95°C for 10 minutes; 95°C for 15 seconds, 60°C for 1 minute, for a total of 40 cycles.

The following probe-primers were used, both tagged to FAM detectors: for CD33 Hs01076281_m1 Taqman CD33 (Thermofisher, Cat 4331182); and for U2AF1 Hs01597465_g1 Taqman U2AF1 assay (Thermofisher, Cat 4351372). A GAPDH probe in VIC-TAMRA (Thermofisher, Cat 4310884E) was used as a housekeeping gene.

7.3.1.7.2.1 For SYBR gene based probes

7.3.1.7.2.1.1 Primer design

Ten genes were selected for validation by qPCR using the SYBR green system. No Taqman probes were commercially available for these genes.

For eight of these genes, primer pairs from previous publications were selected from the Primer Bank website website[311]. Published primers were checked for suitability and specificity against the human genome using the BLAST website[312].

For the final two genes of interest, no published primer pairs were available. I am very grateful for the assistance of Dr Emily Saunderson with primer design for qPCR for the *HIST1H2BC* and *HIST1H2AH* genes.

Primer pairs were ordered from Sigma at a concentration of 100µM, using their custom qPCR probe service. Ordered sequences are detailed in Table 7-1. Primers were diluted in nuclease-free H_2O to a concentration of 10µM before use.

Table 7-1 Primer Pair sequences for SYBR Green qPCR

Gene Name	Primer Sequence (5' - 3')	
	Forward	Reverse
GNA15	CCAGGACCCCTATAAAGTGACC	GCTGAATCGAGCAGGTGGAAT
HIST1H2BC	ACACAGAGTAACTCTCCTTGCG	AAGAAGGCAGTGACCAAAGC
HIST1H2AH	CAAGCAAGGCGGTAAAGCTC	AACCCGCTCGGCATAATTAC
IDH1	TGTGGTAGAGATGCAAGGAGA	TTGGTGACTTGGTCGTTGGTG
KLF9	GCCGCCTACATGGACTTCG	GGATGGGTCGGTACTTGTTCA
LILRA2	AGCCCCAGGAAAGAACGTG	GCTGAGTGAGCTGTAGCATCT
NFIL3	AAAATGCAGACCGTCAAAAAGGA	TGACACTTCCGTTAAAGCAGAAT
NXF3	GGACACACTACGGGTCACAC	ACAGGTTCAGACCTACTGCTAAA
RGS1	TCTTCTCTGCTAACCCAAAGGA	TGCTTTACAGGGCAAAAGATCAG
S100A10	GGCTACTTAACAAAGGAGGACC	GAGGCCCGCAATTAGGGAAA
TRAF4	TATTGGGCCTGCCTATCCG	CAAAACTCGCACTTGAGGCG
ZKSCAN4	TCAAGGGTCTCAAAGTAGCCA	CTATCGTGTAAGGGCTGGGAT
GAPDH	GATTTGGTCGTATTGGGCGC	TTCCCGTTCTCAGCCTTGAC

7.3.1.7.2.1.2 SYBR green qPCR

Primer specificity and efficiency were first tested on two samples of stock cDNA from normal bone marrows. Melt-curves were inspected for the presence of multiple amplicons.

A master mix was made to a total volume of 9µl per well as follows: 6µl Sso Advanced Universal SYBR Green Supermix (Biorad, Cat: 7725274); 0.125µl F primer (10µM); 0.125µl R primer (10µM); 2.75µl H_2O. 9µl of master mix was added to each well, followed by 1µl of cDNA, to a total well volume of 10µl. Reactions were set up in triplicate, along with appropriate negative controls. 384 well, clear white wells (BioRad HSP 3805) were used and sealed with compatible seals (Microseal B Seals, BioRad MSB1001). After sealing, the plate was centrifuged (1500rpm, 1 min).

The following analysis cycle was run on the BioRad C1000 Touch Thermal Cycler: 95^{o}C for 2 minutes; 95^{o}C for 5 seconds, 60^{o}C for 30 seconds for a total of 40 cycles; melt curve 65^{o}C to 95^{o}C, with 0.5^{o}C increments every 5 seconds.

All ordered primers yielded single amplicons.

7.3.1.7.2.2 Analysis

qPCR results were analysed as described in Section 2.8.4.4.

The distribution of calculated fold changes in gene expression seen for individual samples within the AML-exposed and control HSC groups was examined for normality by the D'Agostino-Pearson normality test. Non-parametric datasets were analysed for significance by the Mann-Whitney t-test, using a cut-off value for statistical significance of <0.05.

7.3.1.7.3 Flow based analysis

BM samples of $CD34^{low}$ AML and controls were thawed, washed and counted as previously described. Samples were stained with ALDH, followed by HAG (dissolved in ALDH buffer). Samples were then stained with the following dyes at a concentration of 5µl/1.5x10^6 cells: CD34-PerCP, CD38-Pecy7, CLL1-PE and CD85g-APC (BD, Cat 562355). A second batch of the samples was stained with CD34-PerCP, CD38-PeCy7, CLL1-Alexaflour 647 (BD, Cat 562568) and Anti-TWEAK PE (BD, Cat 565731). Analysis of all samples was performed on a BD Aria II.

Analysis was performed using the FlowJo v10 package. Staining of samples was expressed as the percentage of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ HSCs which were positive (as defined by matched FMOs) for either CD85g or TWEAK expression.

7.3.2 Results

7.3.2.1 Optimisation of SmartSeq Kit

Knowledge of likely yields of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells from bone marrow samples stored in the Bart's Tissue Bank suggested that the majority should yield a minimum of 1000 HSCs on sorting. Therefore, all samples were sorted until exactly 1100 cells had been collected. Excess HSCs were sorted into a separate tube, to undergo RNA extraction for later qPCR validation.

7.3.2.1.1 Optimal input material: the use of cells rather than RNA appeared to offer more efficient cDNA production

The SmartSeq v4 kit allows for cDNA synthesis starting either from intact cells or previously extracted RNA. I first compared the efficiency of these two approaches, using the Qiagen MicroColumn method for RNA extraction.

One sample (Pat ID 10146) of normal control BM was sorted as described for $CD34^{+}38^{-}ALDH^{high}CLL^{-}$ cells. 1000 cells were collected into two separate FACS sorting test tubes. One underwent RNA extraction, producing an eluate of 14µl volume. 9.5µl of this was transferred to a 200µl eppendorf. The second FACS tube was processed as described above (Section 7.3.1.3), with cells undergoing washing, transfer to a 200µl eppendorf and resuspension in H_2O to a volume of 9.5 ml.

1µl of a 10x Reaction buffer was added (19µl 10xLysis Buffer, 1µl RNase Inhibitor) to both tubes (of RNA and intact cells), after which the tubes were vortexed, and then incubated for 5 minutes at room temperature to promote cell lysis.

All downstream processing of the samples was identical (see Section 7.3.1.3.1). Analysis of the resultant cDNA on a HS 1000 Screentape appeared to suggest a larger yield when direct cell lysis was used, rather than formal RNA extraction (see Figure 7-4 A). As a result, sorted intact cells were used in preference to pre-extracted RNA.

7.3.2.1.2 Optimal amplification cycles of cDNA: 10 PCR cycles appeared to offer consistent detection of cDNA

Optimisation of the number of PCR cycles used in cDNA amplification within the SmartSeq v4 kit was performed. Yields of cDNA can vary dependant on the input number of cells and relative size of their transcriptomes, as well as the performance of the thermal cycler used. Too few PCR cycles can lead to small yields of cDNA; too many cycles leads to a less representative library.

Sorting 1000 cells from a control BM and amplifying the cDNA with 8 PCR cycles generated no visible cDNA on HS ScreenTape. The experiment was then repeated using 9, 10 and 12 cycles on 1000 cells sorted from the same cycle. cDNA was visible at both 9 and 10 cycles (see Figure 7-4 B).

A final choice of 10 PCR cycles per sample was taken on the basis that the sorting process and the SmartSeq kit were both costly, and sample availability was limited. Less wastage would occur with the use of 10, rather than 9, cycles. We took some reassurance that the cDNA was not over amplified, as the output concentration of cDNA as assessed by HS Qubit after clean-up was at the lower end of the expected range (3.4 to 17 ng cDNA).

7.3.2.1.3 Optimal cleaning: at the end of the SmartSeq process, an additional 1:1 clean up step of cDNA was required

On analysis of all samples using the HS1000 Screentape at the end of processing, there was a consistent peak of DNA of 60bp size, thought to represent adaptor dimer. If left in situ, this was further amplified during library preparation, a process which might reduce the efficiency of sequencing.

Two different time points for repeat clean-up were compared. Repeat cleaning of the cDNA after PCR amplification using a 1:1 bead ratio effectively removed the adapter-dimer peak, but also removed some of the cDNA of larger size (and therefore overall concentration). This method was compared with using a 1:1 bead ratio to remove the same peak after the sample had undergone library preparation using the Nextera XT kit. In the illustrated case, cleaning resulted in fall in library concentration from 1.1 to 0.6ng/µl, but more importantly, incomplete removal of the unwanted peak.

Therefore, in order to ensure complete removal of adaptor dimer, the additional 1:1 bead: sample clean up step was performed after cDNA production, and before library preparation. These results are summarised in Figure 7-5. Repeat cleaning of the cDNA sample at this stage often resulted in a fall in sample concentration to below the 0.2ng/µl sensitivity of the HS Qubit assay. However, despite this, it was still possible of generate libraries of apparently sufficient concentration for sequencing.

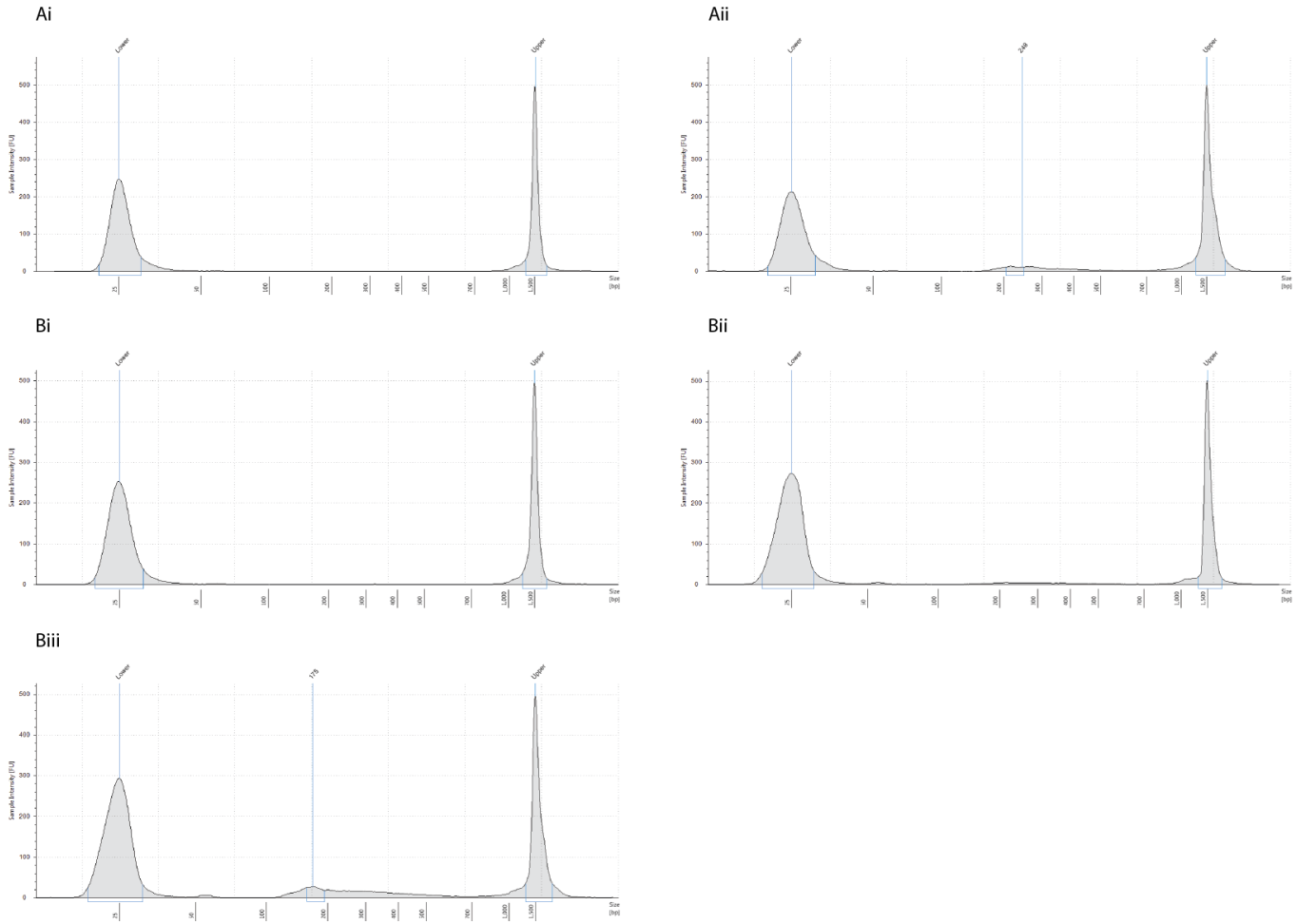

Figure 7-4 HS 1000 DNA ScreenTape traces showing A: yield of cDNA from SmartSeq kit from 1000 cells when (i) immediate RNA extraction and subsequent processing through SmartSeq protocol (ii)cells lysed direct cDNA synthesis performed without prior RNA extraction. Traces shows size of strand on x axis, Intensity/amount on y axis. Upper and lower markers are labelled; B: comparing the cDNA yields when i: 8 cycles ii: 9 cycles and iii: 10 cycles were used to amplify the cDNA resultant from the processing of 1000 cells

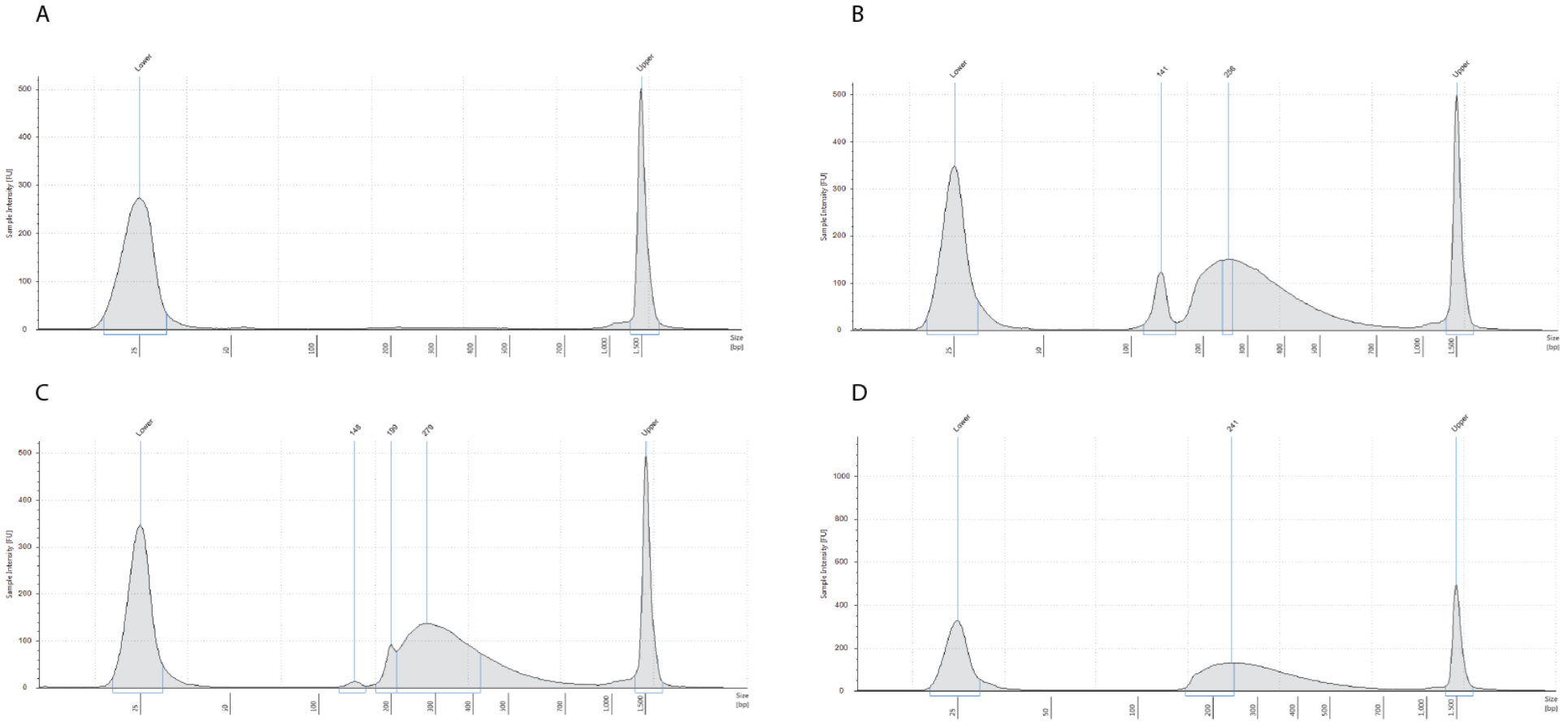

Figure 7-5 showing the effect of clean-up at different stages of library preparation. Four HS DNA ScreenTape 1000 traces are shown where A. cDNA trace after SmartSeq preparation showing a small peak of adapter present at 60bp; B: library appearance after Nextera XT prep with no additional clean up C: library after 1:1 clean up D: library appearance if an additional 1:1 clean-up is performed on cDNA pre-library formation

7.3.2.2 Sample Information

Sample information in terms of patient details, blood counts at presentation and genetic mutational status is given in Table 7-2.

Table 7-2 Clinical information for 6 $CD34^{low}$ AML and 6 controls which underwent sorting for HSCs prior to RNA-Seq based transcriptome profiling

ID	SAMPLE	SEX	AGE	PRESENTATION BLOOD COUNT			KARYOTYPE	*NPM1* STATUS
				HB (g/dL)	WCC ($x10^9$/L)	PLATELETS ($x10^9$/L)		
7793	AML	F	56	11.1	4	10	+13	NPM A
9002	AML	M	68	8.9	0	41	NORMAL	WT
9067	AML	M	67	7.4	142	172	NORMAL	NPM A
8169	AML	F	27	9.7	0	110	t(11;19)	WT
8452	AML	M	27	8.6	6	75	t(11;17)	WT
7774	AML	M	62	8.1	0	37	NORMAL	NPM A
8245	CONTROL	M	45	12.9	7.8	316	NORMAL	
7926	CONTROL	M	42	11.2	14.3	527	NORMAL	
10115	CONTROL	M	50	15	12.6	294	NORMAL	
8445	CONTROL	M	62	15.1	7	252	NORMAL	
TEMP 1728	CONTROL	F	70	12.4	11.8	341	NORMAL	
9903	CONTROL	M	88	11.4	8.1	191	NORMAL	

7.3.2.3 Library Preparation

7.3.2.3.1 Sorting and cDNA synthesis

All samples listed yielded a minimum of 1100 cells after staining and sorting for HSCs as described above.

A typical screen tape trace at the end of SmartSeq cDNA synthesis is shown in Figure 7-4 Biii. All samples at this point were assessed for DNA concentration by the Qubit HS DNA assay. After an additional 1:1 cleaning step to remove excess adaptor, all concentrations were re-assayed, and sample composition re-analysed by HS D1000 screen tape. Results are shown in Table 7-3.

In all cases, the secondary 1:1 cleaning step to remove excess adaptor resulted in a fall in the dsDNA concentration as measured by HS Qubit below the level of detection of the assay. Nevertheless, these cDNA samples remained capable of producing DNA libraries of adequate

Table 7-3 Qubit concentrations of cDNA library pre and post 1:1 cleaning

SAMPLE ID	PAT ID	PRE-CLEAN [] ng/µl	POST 1:1 CLEAN [] ng/µl	VOLUME IN LIBRARY PREP µl
1	7793	0.218	<0.2	8
2	9067	0.26	<0.2	7
3	8445	0.294	<0.2	7
4	7926	0.3	<0.2	8*
5	8452	0.318	<0.2	7
6	7774	0.308	<0.2	7
7	10115	0.218	<0.2	8
8	8169	0.188	<0.2	8
9	8245	0.274	<0.2	7
10	TEMP 1728	0.246	<0.2	7
11	9002	0.272	<0.2	7
12	9903	ND	<0.2	7

concentration for sequencing. Given that the additional cleaning step was so effective at reducing the adaptor presence, and optimisation experiments had suggested cleaning BEFORE library prep was preferable to afterwards,we proceeded on this basis. A disadvantage of this approach is that an approximation had to be made as to the amount of cDNA added to each library preparation. As detailed in Section 7.3.1.4.2, 1ng cDNA (at a concentration of 0.2ng/µl) is required for input into the Nextera XT kit. Given that the secondary cleaning step appeared to reduce the DNA concentrations by a minimum of a third, I estimated that adding between 7 and 8 µl of cDNA would be appropriate for library preparation. This appeared to me to be one of the potential sources of inaccuracy between individual libraries in the protocol, but there was no obvious alternative.

7.3.2.3.2 Library Prep

After library preparation, libraries were QCed by analysis by HS ScreenTape. An example trace is shown in Figure 7-5 D. The average BP size of the resultant library was calculated using the Integration analysis package on the HS Screen tape. This varied between 289 and 445 bp. Interestingly, (despite the same initial DNA input), the final library concentration varied quite considerably. Individual Qubit concentrations and average BP size of the library are given in Table 7-4. Indexes added are also logged. Individual libraries were also QCed with the use of the Kappa qPCR kit. Average concentrations based on these dilutions are shown in Table 7-5.

Table 7-4 showing Library concentrations as estimated by Qubit

ID	SAMPLE	INDEX 1 (i7)	INDEX 2 (i5)	AVERAGE LIBRARY SIZE (bp)	QUBIT LIBRARY [] (ng/µl)	CALCULATED LIBRARY [] (nM)
7793	AML	N701	S503	302	0.96	4.9
9067	AML	N702	S503	325	2.98	14.1
8445	CONTROL	N703	S503	445	6.8	23.5
7926	CONTROL	N704	S503	320	0.984	4.7
8452	AML	N705	S503	271	5.66	32.1
7774	AML	N706	S503	302	1.57	8.0
10115	CONTROL	N701	S504	289	4.52	24.1
8169	AML	N702	S504	345	1.2	5.4
8245	CONTROL	N703	S504	342	3.34	15.0
TEMP 1728	CONTROL	N704	S504	306	1.15	5.8
9002	AML	N705	S504	294	1.07	5.6
9903	CONTROL	N706	S504	334	3.22	14.8

Table 7-5 Library Concentrations as estimated by qPCR

ID	SAMPLE	AVERAGE LIBRARY SIZE (bp)	ESTIMATED LIBRARY [] BY qPCR (nM)			QUBIT LIBRARY [] (nM) FOR COMPARISON
			1/10000 DILUTION	1/20000 DILUTION	MEAN	
7793	AML	302	2.76	4.07	3.42	4.89
9067	AML	325	17.76	21.99	19.88	14.11
8445	CONTROL	445	26.74	29.11	27.93	23.51
7926	CONTROL	320	6.82	7.41	7.12	4.73
8452	AML	271	33.71	38.52	36.12	32.13
7774	AML	302	11.95	13.75	12.85	8.00
10115	CONTROL	289	23.73	29.33	26.53	24.06
8169	AML	345	5.24	6.50	5.87	5.35
8245	CONTROL	342	18.13	15.92	17.02	15.02
TEMP 1728	CONTROL	306	7.31	6.24	6.77	5.78
9002	AML	294	9.25	7.32	8.29	5.60
9903	CONTROL	334	21.27	23.05	22.16	14.83

7.3.2.3.3 Library Normalisation and pooling

Library normalisation was performed based on the library concentrations generated by Kappa based analysis. Pooling ratios are shown in Table 7-6.

A library concentration minimum of 2nM was required for sequencing on the HiSeq4000. The final pooled library underwent HS D1000 Screentape analysis, and was found to have an average length of 360bp. HS Qubit measured a final concentration of 1.41ng/µl, which equates

to 6nM. The final concentration was also calculated independently using Kappa primers (on library dilutions of 1/10,000, 1/20000 and 1/40000), which generated a mean concentration of 6.74nM.

Table 7-6 Volumes used in final library pooling to ensure equal representation of all 12 libraries in sequencing

PAT ID	SAMPLE	LIBRARY [] BY qPCR (nM)	RELATIVE CONCENTRATION	VOLUME ADDED (µl)
7793	AML	3.42	1.00	20.00
9067	AML	19.88	5.82	3.44
8445	CONTROL	27.93	8.18	2.45
7926	CONTROL	7.12	2.08	9.60
8452	AML	36.12	10.58	1.89
7774	AML	12.85	3.76	5.32
10115	CONTROL	26.53	7.77	2.57
8169	AML	5.87	1.72	11.64
8245	CONTROL	17.02	4.98	4.01
TEMP 1728	CONTROL	6.77	1.98	10.09
9002	AML	8.29	2.43	8.24
9903	CONTROL	22.16	6.49	3.08

7.3.2.4 Sequencing data

7.3.2.4.1 A high proportion of reads map to intronic and intergenic regions, highly suggestive of gDNA contamination

After successful sequencing at the Oxford High Throughput Genome, the resultant data were analysed by the BCI Bioinformatics Team.

Analysis of the data revealed that a large proportion of reads mapped to non-coding regions of the genome. Although this occurred in all samples, the proportion of total reads mapping to coding regions varied significantly from sample to sample. These data are shown in Table 7-7.

Potential explanations for this rather unexpected and disappointing result are detailed in length in Section 7.5.1. In brief, external contamination appeared an unlikely explanation given the consistent success of negative controls at all stages in processing.

The SmartSeq kit design is dependent upon the selection of polyadenylated RNA only for cDNA synthesis and subsequent amplification. As we had chosen to use cells as our starting material (rather than pre-extracted total RNA), the specificity of this selection process is vital to exclude the amplification of non-polyadenylated RNA or DNA, which will also be present in the cell

lysate. Thus inadvertent amplification of internal gDNA might be one explanation of the received results.

Table 7-7 Number of reads mapped to coding and non-coding regions for all 12 samples

ID	SAMPLE	READS MAPPED TO CODING REGIONS	READS MAPPED TO INTRONIC REGIONS	Intronic/coding+Intronic
7793	AML	3515643	24869043	0.876
9067	AML	13988713	17771459	0.600
8452	AML	12592268	12553038	0.499
7774	AML	13060130	12446822	0.488
8169	AML	4154727	30407212	0.880
9002	AML	10515967	11655531	0.526
8445	CONTROL	3169227	43637766	0.932
7926	CONTROL	3983312	26909355	0.871
10115	CONTROL	17746555	13054093	0.424
8245	CONTROL	14750523	17906626	0.548
TEMP 1728	CONTROL	5886733	22759573	0.795
9903	CONTROL	10954671	19330724	0.638

Clontech have published QC data examining the efficiency with which their kit produces sequencing data mapping to coding regions of DNA. They have explored the effect of different starting amounts of material, as well as the quality of data generated from cells versus RNA. When 1000 HeLa cells were used as starting material for direct cDNA synthesis and library generation, 75% of eventual reads mapped to the coding regions of the genome. However, it is interesting to note that this specificity is significantly less than observed when cell input numbers are smaller. When the kit is used for single cell processing, an impressive 97% of reads map to coding regions. The reasons for this deterioration in performance are not expanded upon in the company literature[306].

We hypothesised that as the concentration of DNA and RNA within the reaction increases, as would be expected as cell input numbers rise, the efficiency of selection for polyadenylated RNA might fall. It is also interesting to hypothesise whether the relative amount of RNA per input cell might have a bearing on the quality of data generated. It would seem reasonable that the transcriptome of a quiescent HSC is likely to be significantly smaller than that of a metabolically active HeLa cell, and therefore the ratio of DNA to RNA within this cell type is likely to be considerably higher (see Section 7.5.1.2).

7.3.2.4.2 Internal QC of sort purity

We have previously shown that $CD34^{+}38^{-}$ cells sorted from $CD34^{low}$ AMLs generate normal haematopoietic colonies, both in in vitro and murine transplant work. We have also independently sorted the $CD34^{+}38^{-}$ cell populations from *NPM1* mutated $CD34^{low}$ AML samples, and shown the leukaemia-associated mutation to be absent in these cells by immediate qPCR-based analysis.

Because cell numbers in this experiment were highly limited, and all sorted $CD34^{+}38^{-}$ $ALDH^{high}CLL1^{-}$ cells were either used for immediate cDNA synthesis or extraction of RNA for subsequent validation, DNA-based analysis to confirm the presumed normal genotype of the sorted cells was not performed at the time of sorting.

Nevertheless, an internal QC of sorting efficiency was performed by looking in detail at three of the AML samples known to have a *NPM1* Type A mutation (Patients 7793, 9067 and 7774). Examination of the reads which mapped to Exon 12 of the *NPM1* gene for each of these samples showed normal sequences, with no evidence of the leukaemia associated mutation.

7.3.2.4.3 Differential gene expression

Even allowing for the high proportion of reads which mapped to intronic regions, the read depth per sample was high enough for a reasonable number of reads for each sample to have mapped to coding regions (between 3.5 to 17 million reads per sample). Although we do not presume that all of these reads relate to cDNA made from polyadenylated RNA, if one assumes that the genomic contamination was random, and therefore evenly distributed throughout the genome, genes which appear to be differentially represented between the two sample cohorts are certainly worthy of further investigation.

Filtering revealed differential expression of 18025 genes. Of these, 525 showed differential expression between the two subgroups with fold changes of greater than two, and a calculated p-value of less than 0.05. No gene changes achieved an adjusted p value of significance of <0.05 (calculated using the Benjamini–Hochberg procedure). Tables showing details of the top differentially expressed genes are located in Appendix 3 of this thesis.

7.3.2.5 Validation of RNA-Seq Data

7.3.2.5.1 Choice of genes for validation

The choice of genes for attempted validation was based on a number of filters. Interest was focussed on those genes which exhibited a difference in expression between HSCs from AML

and control samples of greater than two, and ideally where the p-value of the observed changes was calculated to be less than 0.001.

Any genes with previous reported roles in cell cycle regulation, HSC maintenance or AML development or normal haematopoietic development were noted with interest.

Comparison was also made between the datasets generated by this work in primary human samples, and that previously produced by the group in the murine xenograft experiment detailed in Section 7.1.2.1. Within the top 200 differentially regulated genes from both datasets, seven appeared in both groups. Of these, only three exhibited changes in gene expression in the same direction in both datasets. Expression of the *DMD* (Dystrophin) and *SEMA7A* (Semaphorin 7A) genes was increased in AML-exposed HSCs compared to controls; *ST3GAL5* (ST3 beta-galactoside alpha-2,3-sialyltransferase 5) expression was relatively lower in the same cohorts. However, given none of these genes had described roles in the regulation of cell cycle or differentiation within either normal HSC or AML biology, none were selected for further validation.

7.3.2.5.2 Selection of Validation Methods

7.3.2.5.2.1 By qPCR

Given our concerns with regard to the quality of the RNA-Seq data detailed in Section 7.3.2.4.1, we felt it was reasonable to attempt validation by qPCR for a number of genes of interest.

At the end of sorting, we had extracted RNA from any excess HSCs not required for SmartSeq processing. As a result of the small number of available cells, this RNA was assumed to be extremely dilute (certainly below the lower level of sensitivity of HS Qubit RNA assay). To try and maximise the cDNA yield from the small quantity of RNA available, we used the Superscript®VILO™ cDNA synthesis kit (Thermofisher, Cat 11754050), with a prolonged incubation time designed to promote maximum cDNA synthesis.

Despite this step, the number of genes which we could attempt to validate expression levels on was limited. Samples were sufficiently dilute that on qPCR, even highly expressed house-keeping genes were detected relatively late (with an average Ct of 30 to 32). Therefore, highly expressed genes (as defined by RPKM values from the RNA-Seq data) were preferentially selected for validation, in the hope that levels for all samples would fall within the sensitivity of 40 cycle qPCR.

7.3.2.5.2.2 By protein expression levels

Protein-level validation of changes in gene expression was also attempted where possible. Previous work had shown that simultaneous staining of samples for HSC markers and the protein of interest, followed by flow-based analysis, was the most efficient way of approaching this question. Many of the differentially expressed genes of interest coded for nuclear-located proteins such as transcription factors. However, the experience of intracellular staining for RhoA expression during the flow cytometry assessment detailed in Section 7.1.2.2.4 had highlighted some of the difficulties of this approach for protein based validation, especially in the absence of a validated, fluorochrome-conjugated antibody. However, three genes of interest coded for cell surface proteins, with commercially available flow antibodies to assess their expression levels.

7.3.2.5.3 Validation targets

Table 7-8 shows the chosen targets and methods for validation, along with observed fold changes in gene expression between the two cohorts and statistical significance of these changes. Pre-existing experimental evidence for these genes having a potential role of interest in HSCs, AML development or cell cycle control is detailed below in Sections 7.3.2.5.3.1 and 7.3.2.5.3.2.

Table 7-8 Genes selected for Validation by qPCR and flow-based assessment

	GENE	FULL NAME	LOG CHANGE IN EXPRESSION	P VALUE	VALIDATION METHOD
UP IN HSCS FROM CONTROLS VERSUS AML	CD33	CD33	2.98	0.0046	FLOW & TAQMAN qPCR
	LILRA2	leukocyte immunoglobulin-like receptor, subfamily A (with TM domain), member 2	2.87	0.0027	FLOW
	IDH1	isocitrate dehydrogenase 1 (NADP+), soluble	1.56	0.0072	SYBR qPCR
UP IN HSCs FROM AML VERSUS CONTROLS	U2AF1	U2 small nuclear RNA auxiliary factor 1	3.42	0.0018	TAQMAN qPCR
	HIST1H2BC	histone cluster 1, H2bc	3.37	0.0122	SYBR qPCR
	S100A10	S100 calcium binding protein A10	2.61	0.0015	SYBR qPCR
	TNFRSF12A	tumor necrosis factor receptor superfamily, member 12A	2.18	0.0232	FLOW
	TRAF4	TNF receptor-associated factor 4	2.13	0.0168	SYBR qPCR

7.3.2.5.3.1 Genes relatively downregulated in HSCs exposed to AML compared to controls

7.3.2.5.3.1.1 CD33

The role of CD33 in HSC behaviour, and its pattern of expression on normal and AML-exposed HSCs are discussed at length in Chapter 6 of this thesis.

As a cell surface marker, protein based expression is readily quantified by flow cytometry, but a specific Taqman probe/primer pair is also commercially available to assay gene expression levels.

7.3.2.5.3.1.2 LILRA2 (Leucocyte Immunoglobulin-Like Receptor, subfamily A (with TM domain), member 2 (CD85H)

The Leucocyte Immunoglobulin-Like Receptors are expressed on monocytes and B cells. There exist two subfamilies: subfamily A receptors are thought to have stimulatory functions, whilst subfamily B receptors facilitate inhibitory signalling. LILRA2 has recently been implicated as holding an important role within the innate immune system. Signalling through the receptor appears vital in the recognition of pathogen-cleaved immunoglobulin molecules, and the myeloid cell activation that follows[313].

Whilst members of the LILRB subfamily have been shown to have functions within the normal HSCs of both mice and humans, and to be capable to causing HSC proliferation[314], the LILRA subfamily have no documented roles in stem cell populations or in cancer development to date.

As a cell surface protein, protein expression is quantifiable by flow cytometry.

7.3.2.5.3.1.3 IDH1 (Isocitrate Dehydrogenase 1)

The *IDH 1* and *2* genes have been identified as recurrent mutation targets in a variety of cancers. Their significance was first described in gliomas[315], and shortly afterwards in AML[316]. Current estimates suggest that 15 to 20% of AMLs carry mutations in one of these genes, with an increased incidence in AMLs of normal karyotype.

Under normal circumstances, IDH1 and 2 play key roles in the regulation of metabolism, converting isocitrate into α-ketoglutarate (αKG). As well as being a component of the Krebs cycle, αKG is required for the normal functioning of a number of dioxygenase enzymes, such as TET2 and histone demethylases.

Mutations in these genes have a dual effect: not only do they reduce the activity of the normal enzyme, but when present in the heterozygous form, the mutant enzyme can also catalyse the conversion of existing αKG into an alternative "oncometabolite" R-2-hydroxyglutarate (2HG). 2HG is a competitive inhibitor of a number of dioxygeneases[317]. As a result, such mutations results in a change in cellular metabolism, epigenetic reprogramming[318], and a block in cell differentiation. However, as patients rarely present with AML characterised solely with mutations in *IDH1* or *2*, the accumulation of further mutations is thought to be required for transformation.

In murine models, knock-in of mutated *IDH1* into myeloid cells resulted in abnormal haematopoiesis, with increased numbers of early haematopoietic progenitors and the development of extramedullary haematopoiesis[319].

There is little available literature on the effect of over- or under- expression of the wild-type *IDH1* gene in humans. However, interesting recent work using zebrafish showed that knockdown of wildtype *IDH1* resulted in a blockade of normal myeloid differentiation[320].

Primers were designed for use with the SYBR green system for qPCR based gene expression analysis.

7.3.2.5.3.2 Genes relatively upregulated in HSCs exposed to AML

7.3.2.5.3.2.1 U2AF1 (U2 small nuclear RNA auxiliary factor 1)

U2AF1 represents the small subunit of the U2 small ribo-nuclear protein, an auxiliary factor involved in pre mRNA splicing. Mutations in spliceosome-coding genes have caused significant interest in the field of cancer evolution, as a number have recently been reported in both solid and haematological malignancies. Affected genes include *SF3B1*, *SRSF2*, *U2AF1* and *ZRSR2*[321].

Spliceosome mutations have been identified in at least 50% of patients with MDS[322]. *U2AF1* mutations are found in around 9% of MDS patients, and are associated with an increased risk of transformation to AML[323]. The mutations of interest affect the Ser34 amino acid, which forms part of the zinc binding domain of the protein, and is normally highly conserved.

Functionally, mutations in the U2AF1 protein result in enhanced splicing and exon skipping[323]. In a conditional murine model with inducible expression of the mutant gene, changes in haematopoiesis were observed when the mutant allele was expressed. Murine leucocyte levels dropped (whilst red cells and platelet levels were maintained). Interestingly, after a month of mutant gene expression, $CD150^{+}cKit^{+}Lin^{-}$ stem cell proliferation (as evidenced by

BRDU incorporation and Ki67 staining) increased. However, in vitro assays of stem cell repopulating ability showed these HSCs to be at a competitive disadvantage to unmutated forms[322].

Identified targets of downstream missplicing in this setting include genes involved in DNA damage repair (ataxia telangiectasia and Rad3 related (*ATR*)), epigenetic regulation (H2A histone family Y (*H2AFY*)) and apoptosis (caspase 8)[324,325].

To date, all experimental interest has focussed on the effect of mutation on the activity and function of the U2AF1 protein. There is no literature to date which comments on the potential effects of increased expression of wild type U2AF1 on HSC behaviour or myelopoiesis.

Taqman based probe-primer sets are commercially available for gene expression validation.

7.3.2.5.3.2.2 Histone factors HIST1H2AL and HISTH2AH

One of the most striking features of this dataset is the marked enrichment within the AML exposed HSCs for expression of members of the histone family. Ten of the top thirty genes overexpressed in AML-exposed HSCs compared to controls code for histones.

Interestingly, the fact these genes appear at all within the RNA-Seq dataset is a puzzle, and possibly indicative of the failure of PolyA selection of mRNAs. Most of these gene transcripts are not polyadenylated, but instead contain a palindromic termination element, and therefore theoretically should not have undergone amplification within the SmartSeq kit.

Histones are basic proteins, responsible for the formation of the nucleosomes around which chromosomes are wrapped. Nucleosomes are comprised of two of each of the four core histones (H2A, H2B, H3 and H4). H1 is a linker histone, which binds to DNA between nucleosomes and is responsible for chromosome compaction.

Histones are thought to be vital in the regulation of gene expression: genes bound to histones are less transcriptionally active[326]. However, most research has focussed on how post-translational modifications of histones such as acetylation or methylation affect their repressor activity. There are limited published data on whether differential expression of specific histone genes are associated with different cell behaviours, and the mechanisms which control their transcription are poorly understood.

Of relevance, increased expression of *H2A* histone genes has been associated with cell senescence[327]. Could an increase in histone expression be associated with a state of acquired HSC quiescence as well?

Primers were specifically designed to assay gene expression within the SYBR green qPCR system.

7.3.2.5.3.2.3 S100A10

S100A10 (also known as p11) is part of the S100 family of proteins. S100 proteins have been reported to play roles in cell cycle progression and differentiation.

The structure of S100A10 allows it to bind internally to annexin I and II, whilst facilitating interactions with a variety of cell surface-based proteins. It has a role in shuttling protein channels to the cell surface, and therefore in the maintenance of cell polarity. Its role in neuronal development has been the most studied, where it assists in the transport of neurotransmitters to the cell surface[328].

It can also be found embedded in the cell surface membrane, where it functions as a receptor for plasminogen, mediating the conversion of plasminogen to plasmin[329]. Murine knock-outs of the *S100A10* gene have led to a subsequent reduction in neutrophil-plasmin generation, and thus impeded migration due to a reduced ability to degrade the extracellular matrix. A similar role for S100A10 in mediating invasive behaviour has been proposed for cancer cells[330]. To date, S100A10 has not been reported to have a defined role either in normal HSC, or AML development.

Primers were specifically designed to assay gene expression within the SYBR green qPCR system.

7.3.2.5.3.2.4 TNFSF12A (Tumor necrosis family receptor superfamily, member 12A)

The gene *TNFRSF12a*, also known fibroblast growth factor-inducible 14 (*Fn14*), codes for a cell surface receptor belonging to the TNF receptor family, known as TWEAKR or CD266. Its ligand is TNF-like weak inducer of apoptosis (TWEAK or TNFSF12), a member of the TNF ligand family[331].

TWEAKR is fairly ubiquitously expressed during development, before becoming more restricted in adulthood to the heart, kidney, liver and placenta[332,333]. Under normal conditions, expression levels of TWEAKR are thought to be modulated by a range of cytokines[330].

Abnormal TWEAKR expression has been reported in a number of solid organ malignancies, including pancreatic tumours[334]. AML cells have been reported to downregulate the receptor to promote their survival[335].

The receptor ligand, TWEAK, has been reported to be cytotoxic both in cancer lines and healthy monocytes[336]. However, in contrast to other members of the TNF ligand family, TWEAK seems to require the co-activation of additional signalling pathways to result in cell death. In vascular endothelial cells, TWEAK has been reported to be responsible for cell proliferation[337], as well as directing cell migration[338].

It is assumed, because of structural homology with other TNF-receptor family members, that TWEAKR signalling utilises TRAFs[338]. Downstream of this, TWEAK-R activation is thought to result in NF-κB activation, as forced overexpression of the receptor results in NF-κB activation[339].

Of interest for further validation work, there already exists a specific antibody capable of blocking the TWEAK-TWEAKR interaction[340], which has been trialled in mice models and shown to reduce intestinal damage in intestinal GvHD[341].

As a cell surface antigen, protein based expression analysis is possible by flow cytometry.

7.3.2.5.3.2.5 TRAF4 (TNF receptor-associated factor 4)

TNF receptor-associated factor 4 is a member of the TRAF family of scaffold proteins, which interact with TNF and IL1R/Toll receptors, facilitating downstream signalling into the NF-κB and MAP kinase pathways.

Despite its name, and structural similarities with other TRAF family members, there are to date no documented interactions with TRAF4 and the TNF receptor[342]. However, it has been shown to block p75, the neurotropin receptor, an action which leads to NF-κB mediated cell death[343].

Overexpression of TRF4 has been reported in a number of carcinomas, but not to date in AML.

Primers were specifically designed to assay gene expression within the SYBR green qPCR system.

7.3.2.5.4 Validation results

7.3.2.5.4.1 Protein based assessment

7.3.2.5.4.1.1 CD33 is significantly downregulated in AML-exposed HSCs compared to controls

As detailed in Section 6.1.3, previous unpublished work from our group has shown that the CD33 expression (percentage of CD33 positive cells, as defined by FMO) in the $CD34^{+}38^{-}$ HSC fraction within $CD34^{low}$ AML samples is significantly lower than the same compartment within control samples. Median percentage CD33 positivity within 15 $CD34^{low}$ AML samples was 18%, versus 82% within 20 control samples ($p < 0.0001$ by Mann-Whitney test) (see Figure 7-6 A).

7.3.2.5.4.1.2 CD85g/LILRA protein expression is not significantly different between AML-exposed HSCs and controls

A total of 6 $CD34^{low}$ AML samples and 8 controls were assessed for LILRA expression by flow cytometry staining for surface CD85g. Results were expressed as the percentage of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ HSCs which were positive (as defined by matched FMOs) for CD85g. There is no significant difference between expression in AML-exposed HSCs (0%) versus control-exposed HSCs (0.02%) ($p=0.54$ by Mann-Whitney test)(see Figure 7-6 B).

7.3.2.5.4.1.3 TNFSF12A/TWEAKR/CD266 protein expression is not significantly different between AML-exposed HSCs and controls

A total of 6 $CD34^{low}$ AML samples and 9 controls were assessed for TWEAKR expression by flow cytometry staining for surface CD266. Results were expressed as the percentage of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ HSCs which were positive (as defined by matched FMOs) for TWEAKR expression There is no significant difference between expression in AML-exposed HSCs (10.7%) versus control-exposed HSCs (3.5%) ($p=0.26$ by Mann-Whitney test)(see Figure 7-6 C).

7.3.2.5.4.2 qPCR based gene expression validation

7.3.2.5.4.2.1 Sample details

Samples which underwent sorting and subsequent RNA extraction of their $CD34^{+}CD38^{-}ALDH^{high}CLL1^{-}$ populations for use in qPCR based validation are summarised in Table 7-9. All six $CD34^{low}$ AML samples are the same as those used in the original RNA-Seq experiment. Two of the controls (8245 and 7926) yielded sufficient HSCs for both RNA-Seq processing and separate RNA extraction. The other four control samples used for RNA-Seq work had no further vials available in the Tissue Bank, and hence a further four control samples were selected solely for RNA extraction for gene validation.

Extracted RNA was converted to cDNA as described in Section 7.3.1.7.1.2., prior to qPCR.

Table 7-9 showing sample details for the 6 AML samples and 6 controls used in this experiment for sorting of HSC populations and RNA extraction. Those samples marked *had RNA extracted from them at the same time as the same original sort for RNA-Seq cells. Those marked ** underwent repeat sorting from a duplicate vial at a later date for RNA extraction. Those unmarked were control samples not used in the original experiment

ID	SAMPLE	SEX	AGE	PRESENTATION BLOOD COUNT			KARYOTYPE	NPM STATUS
				HB (g/dL)	WCC ($x10^9$/L)	PLATELETS ($x10^9$/L)		
7793**	AML	F	56	11.1	4	10	add13	NPM A
9002*	AML	M	68	8.9	0	41	Normal	WT
9067**	AML	M	67	7.4	142	172	Normal	NPM A
8169**	AML	F	27	9.7	0	110	t(11;19)	WT
8452*	AML	M	27	8.6	6	75	t(11;17)	WT
7774*	AML	M	62	8.1	0	37	Normal	NPM A
8245*	CONTROL	M	45	12.9	7.8	316	Normal	
7926**	CONTROL	M	42	11.2	14.3	527	Normal	
8508	CONTROL	M	22	14.3	3.9	197	Normal	
8749	CONTROL	F	70	13.9	7.8	193	Normal	
8984	CONTROL	M	42	16.5	6.2	198	Normal	
10030	CONTROL	F	68	12.3	7.8	228	Normal	

7.3.2.5.4.2.2 Low RNA Concentration meant several genes could not be validated due to failure to detect expression after 40 cycles of qPCR

As discussed in Section 7.3.2.5.2.1, we had deliberately selected genes with relatively high expression levels in all samples in the RNA-Seq dataset (as defined by RPKM), in order to maximise the possibility of detectable gene expression levels during a 40 cycle qPCR experiment.

Despite this, of the four genes selected for validation of gene expression levels by qPCR using the SYBR Green system and specifically designed primers, expression of three (*S100A10*, *IDH1* and *TRAF4*) were not detected in 8 of the 10 samples. As a result, interpretation of relative expression of these genes within the AML-exposed and control-exposed HSCs was not valid. Further discussion of this issue is located in Section 7.5.2.3.

7.3.2.5.4.2.3 CD33 gene expression is not significantly different between AML-exposed HSCs and controls

Relative gene expression of CD33 is lower in AML-exposed HSCs than control-exposed HSCs (RE 0.56 versus 2.00), but this change does not reach statistical significance (p=0.22, Mann Whitney t test) (see Figure 7-6 D).

7.3.2.5.4.2.4 U2AF1 gene expression is not significantly different between AML-exposed HSCs and controls

Relative gene expression of *U2AF1* is lower in AML-exposed HSCs than control-exposed HSCs (RE 0.20 versus 1.27), but this change does not reach statistical significance (p=0.15, Mann Whitney t test) (see Figure 7-6 E).

7.3.2.5.4.2.5 HIST1H2BC gene expression is not significantly different between AML-exposed HSCs and controls

Relative gene expression of *HIST1H2BC* is lower in AML-exposed HSCs than control-exposed HSCs (RE 1.75 versus 1.23), but this change does not reach statistical significance (p=0.55, Mann Whitney test) (see Figure 7-6 F).

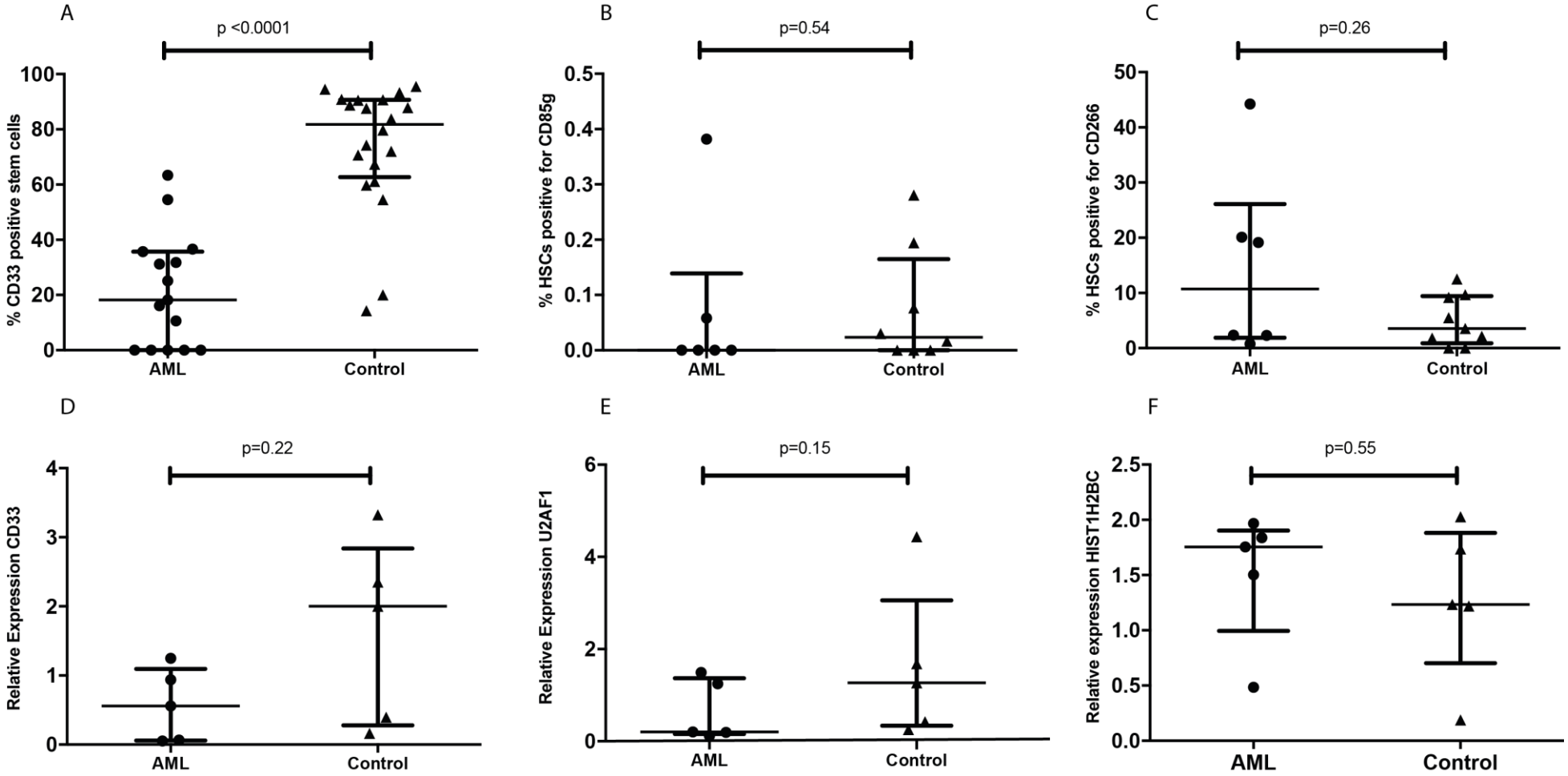

Figure 7-6 Scatter plots illustrating results of validation experiments. Graphs A-C show results from flow-based protein expression assessment; Graphs D-F show results from PCR based gene expression assessment. In all cases, median expression levels and IQR are plotted. Significance of differences in medians calculated by Mann-Whitney testing. A: CD33 expression on $CD34^{+}38^{-}$ HSCs from AML and control sample (data from F Miraki-Moud, as previously shown in Section 6.1.3); B: CD85g (LILRA2) expression on $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ HSCs from AML and controls; C: CD266/TWEAKR expression on $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ HSCs from AML and controls; D: Relative expression of *CD33* gene in $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ HSCs from AML and controls; E: Relative expression of *U2AF1* gene in $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells from AML and controls; F: Relative expression of *HIST1H2BC* gene in $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells from AML and controls

7.4 Attempt 2: Using RNA as input material

As detailed in the Discussion section of this chapter (Section 7.5), we were disappointed with the quality of the data generated by our first RNA-Seq experiment comparing the transcriptomes of HSCs from AML samples and controls. Sequencing data contained a high proportion of reads mapping to non coding regions of the genome, pointing to a potential failure in mRNA selection prior to cDNA synthesis and subsequent amplification.

We still felt the approach of using RNA-Seq to identify differentially expressed genes in primary human samples was a valid one. Moreover, given the positive feedback regarding the SmartSeq kit from other groups working in the same area, we were keen to continue working with the same kit, using a modified methodology, to see if cleaner data could be produced.

We decided to repeat the experiment using RNA extracted from sorted HSCs by traditional methods, rather than relying on cDNA synthesis directly from lysed cells.

Considerable attention was paid to the selection of the optimal method to extract high quality RNA reproducibly from small cell numbers. Below are detailed the methods for the three different techniques of RNA extraction compared, and in the Results section that follows, the QC steps taken to compare the quantity and quality of the different yields. Also detailed is the SmartSeq technique used to produce cDNA from the RNA, which was slightly modified from that described in Section 7.3.1.3, as well as sample details and eventual library concentrations.

7.4.1 Methods

7.4.1.1 Optimisation of RNA extraction methods

To reduce the possibility of sample contamination, all sampling handling was performed under a flow hood. RNA degradation was limited by the use of RNAseZap (ThermoFisher, Cat AM9780) to clean all surfaces and pipettes prior to handling. Fresh filter pipette tips were used for all stages.

7.4.1.1.1 Trizol extraction of RNA from limited cell numbers

The classical method of RNA extraction, with slight modification to optimise yields from small cell numbers, is described below.

7.4.1.1.1.1 Homogenizing Cells

Sorted cells in FACS tunes were spun down (3000rpm, 10 minutes, 4^{o}C), and the supernatant removed with a Gilson aspirator. Cells were resuspended by flicking the tube, prior to the addition of 0.9ml Trizol (ThermoFisher Cat 15596-026) and vortexing. The contents of the tube were then transferred to a clean 2ml eppendorf tube, and left to incubate at room temperature for five to ten minutes.

7.4.1.1.1.2 Phase separation

200µl chloroform (Sigma, Cat C2432) was added to the eppendorf tube, which was then vortexed. The tube was then incubated for three minutes at room temperature, before centrifugation (12000g, 15 minutes, 4^{o}C). Phase separation of phenol and chloroform layers was then visible, with DNA at the interface between the two liquids. The top layer, which contains the RNA, was removed by pipetting (avoiding disturbing the DNA) and placed into a fresh 1.5ml eppendorf tube.

7.4.1.1.1.3 RNA Isolation

RNA isolation was optimised by the use of GenElute LPA carrier (Sigma Cat 56575). In a separate tube, LPA was prediluted to 2.5mg/ml by the addition of 10µl LPA to 90µl nuclease free water. 4µl (10µg) was then added to the RNA aqueous layer, and vortexed. 0.5ml 100% isopropanolol (Sigma, Cat 278475) was then added, vortexed briefly and left to incubate for 10 minutes at room temperature. The tube was then centrifuged (12000g, 10 minutes, 4^{o}C)

7.4.1.1.1.4 RNA Washing

After centrifugation, the supernatant was gently removed by pipetting, leaving only the residual pellet behind. This was washed with the addition of 0.9ml 70% ethanol, followed by vortexing and centrifugation (7500g, 5 minutes, 4^{o}C). The residual liquid was carefully removed

using a pipette, leaving only the pellet. A repeat wash of 0.9ml 70% ethanol was added, and the pellet resuspended by vortexing. However, this time, the eppendorf was placed at -20°C overnight to improve RNA yield (personal communication, Dr J. Brown, UCL).

The following day, the eppendorf was centrifuged as before (7500g, 5 minutes, 4°C), and all traces of residual ethanol removed, leaving the pellet as dry as possible. The tube was left open to air-dry at room temperature for 10 minutes.

7.4.1.1.1.5 RNA resuspension

The pellet was resuspended with the addition of 5µl RNase-free water. The eppendorf was then incubated at 55 to 60°C for 10 minutes. The resultant RNA was stored at -70°C.

7.4.1.1.2 DirectZol method

This column-based method of RNA extraction uses the DirectZol™ RNA MicroPrep with TriReagent kit (Zymo, Cat R2061).

7.4.1.1.2.1 Cell homogenisation

Sorted cells in FACS tubes were spun down (3000 rpm, 10 minutes, 4°C), and the supernatant removed with a Gilson aspirator. Cells were resuspended by flicking the tube, prior to the addition of 150µl TriReagent and vortexing. The tube was then incubated at room temperature for 5 minutes.

7.4.1.1.2.2 RNA purification

150µl 100% ethanol was then added to the tube, followed by vortexing. The contents of the tube were pipetted into a Zymo-Spin™ IC spin column in a collection tube and spun down (16,000g, 30 seconds, room temp). The column was then transferred to a new collection tube, and flow-through discarded.

7.4.1.1.2.3 On-column DNA digestion

400µl RNA wash buffer was added to the column and centrifuged (16,000rpm, 30 seconds, 4°C). In a separate eppendorf tube, 5µl DNase I (6U/µl) and 35µl DNA digestion buffer were mixed, before adding to the top of the column. This was left to incubate at room temperature for 15 minutes. 400µl DirectZol™ RNA PreWash was added to the column and centrifuged. The flow through was discarded, and the process repeated.

7.4.1.1.2.4 RNA Elution

700µl RNA wash buffer was added to the column, and centrifuged (16000rpm, 2 minutes, 4^{o}C) to dryness. The column was transferred to an RNase free tube and 6µl nuclease free water was added to the column and left for two to three minutes before centrifugation.

7.4.1.1.3 Qiagen Microcolumn method

This column-based RNA extraction method, involving an on-column DNA digestion step, is described in full in Section 2.8.2.

7.4.1.2 RNA Quality assessment

The quantity and quality of RNA extracted from small cell numbers by these three methods was compared over a series of experiments. Using cord blood as a source of cells, aliquots of 1000 and 10,000 live (DAPI negative) cells were sorted using the Aria II FACS sorter into tubes. RNA extraction using the methods detailed above was performed in parallel.

7.4.1.2.1 RNA quality assessment using the RNA 6000 Pico Kit

Assessment of RNA yield was performed in the BCI Genome Centre, using the RNA 6000 Pico Kit (Agilent, Cat 5067-1513) and the Agilent 2100 Bioanalyser. In brief, 1µl per sample was loaded onto a chip, along with a RNA ladder. The RNA molecules separate on the basis of size as they run through a polymer matrix across an applied voltage. Dye molecules intercalate with the RNA as it runs across the chip, and are detected as they pass through a fluorescence detector.

The RNA 6000 Pico Kit offers a qualitative range of 250-5000 pg/µl. Based on the assumptions that one cell contains 10pg total RNA and extraction was 100% efficient, one could expect to detect RNA from 125 cells using a Phenol based method and 250 cells with the Qiagen kits.

7.4.1.2.2 RNA Quantification using the HS Cubit RNA Assay

RNA quantification was attempted using the HS Cubit RNA assay (Thermofisher Q32852). This assay is sensitive in the range of 250pg/µl to 100ng/µl. In brief, the kit contains a dye which fluoresces on binding with RNA. This fluorescence is proportional to the amount of RNA present within the test sample, and is detected by the analyser which calculates an RNA concentration relative to known standards.

7.4.1.2.3 RNA Quantification using qPCR

As the RNA concentration of the extracted samples was below the accuracy of even the HS Cubit analyser, relative assessment of RNA concentrations extracted from the 12 $CD34^{+}38^{-}$

ALDH^{high}CLL1^{-} samples was performed by qPCR, to ensure the same amount of starting material was used for all samples before processing using the SmartSeq kit.

cDNA was generated from 1µl aliquots of the RNA from each sample, using the methodology described in Section 7.3.1.7.1.2. qPCR for relative expression of GAPDH was performed, as described in Section 7.3.1.7.2. All samples were run in triplicate, and a mean Ct for each sample calculated.

Relative cDNA concentrations (and assuming a linear reverse transcriptase reaction, relative RNA concentrations) were calculated as follows:

$$Relative\ RNA\ concentration = 2^{-(sample\ mean\ Ct - mean\ Ct\ of\ most\ diluted\ sample)}$$

7.4.1.3 cDNA synthesis using SMART-Seq kit

The same protocol was followed as for when handling intact cells as listed in Section 7.3.1.3, with the following exceptions.

7.4.1.3.1 First strand synthesis

The relative concentrations of the 12 RNA samples were calculated as described in Section 7.4.1.2.3. 9.5µl of the most dilute sample was added to an eppendorf for further processing, and the appropriate volumes of RNA from the other samples were used to attempt the same input of RNA for each sample. Samples were made up to 10.5µl with nuclease free water. 1µl of Reaction buffer was added to each test sample, which was briefly vortexed. Samples were placed on ice, and 1µl 3'SMART-Seq CDS Primer IIA added to each. Further steps in cDNA synthesis are as already described.

7.4.1.3.2 cDNA Amplification

Amplification of the cDNA was performed as described in Section 7.3.1.3.2, with the exception that 18 PCR cycles, rather than the previous 10, were utilised.

7.4.1.3.3 cDNA clean up

cDNA clean up using AmpPure beads was performed as described in Section 7.3.1.3.3.

7.4.1.3.4 Sample QC

After sample clean up, samples were run on the HS D5000 DNA ScreenTape (Agilent 5067-5592). As discussed in Section 7.3.1.3.4, this kit is much more appropriate for assessment of the libraries generated by the SmartSeq kit. It covers a size range of 100 to 5000bp, and is also sensitive to a concentration of 5pg/µl. As illustrated by the example traces shown in Figure

App 4.3, a small peak at 60bp, thought to be related to excess primer, was not seen in these samples. Therefore the additional 1:1 clean up step was not performed.

Quantification of cDNA library concentration was performed using the Qubit HS DNA Quantification kit as described in Section 7.3.1.3.4.1

7.4.1.4 Further processing

All further library preparation, QC, quantification and pooling was performed in the fashion as the first attempt at RNA-Seq, detailed in Section 7.3.1.4 onwards.

7.4.1.5 Data analysis

Sequencing and bioinformatic analysis of data was performed as described in sections 7.3.1.5 and 7.3.1.6

Gene enrichment analysis was performed using David .Version 6.8 (https://david.ncifcrf.gov)[344,345]. Enrichment analysis of specific biological pathways was performed using the KEGG Pathway tool integral to David[346].

The dataset analysed within David contained all genes which exhibited differential expression between the two test groups with a fold change of greater than 2, and an individual p value of <0.05. Of 1013 genes that fulfilled these criteria, 810 were recognised by the software and therefore used for downstream analysis.

7.4.2 Results

7.4.2.1 Assessment of optimal RNA extraction technique for small cell numbers

Three different techniques for RNA extraction from small number of cells were rigorously compared for the efficiency and quality of the RNA yielded from 1000 cells. From experience, 1000 cells is a reasonable achievable yield of $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells from one vial of bone marrow of AML or control samples stored in the Tissue Bank.

7.4.2.1.1 The importance of RNA quality in RNA-Seq Experiments

The absolute efficiency of RNA extraction was deemed less important in this experiment than the quality of the RNA generated. This was because the SmartSeq kit is sufficiently sensitive to allow for inputs as low as one cell, and therefore low yields of RNA could be adjusted for by increasing the number of PCR cycles used for amplification of cDNA.

RNA quality measurements assess the integrity of an RNA sample, assigning it a score based on the likelihood of degradation. As polyadenylated mRNA is often the first component of total RNA to undergo degradation, maintaining RNA quality is thought to be key to ensuring the quality of data generated by RNA-Seq.

The RNA integrity number (RIN) is an algorithm applied to analyse an electrophorectic trace. It is based on a number of characteristics of the trace including the relative areas under the peaks of the 18S and 28S RNA peaks, compared to the total area under the graph. Secondarily, the height of the 28S peak is important, as it is the most prominent species and is also first degraded.

Although there are no absolute cut offs for the RNA quality required for sequencing, a consensus opinion appears to be a RIN of over 9.0 for each sample is deemed ideal to ensure good quality data. However, it is also accepted that as the quantity of available RNA falls, so does the maximum achievable RIN.

Quality assessment of RNA of low concentration is challenging, and to date, our group has not attempted it, believing the sensitivity of available assays to be insufficient to base experimental decisions upon. However other labs working with small cell numbers and performing RNA-Seq routinely assess the quality of their RNA samples using the Agilent 2100 system with an RNA 6000 Pico chip (personal communication, J Brown, UCL).

7.4.2.1.2 At 1000 cell input, Qiagen columns provide optimal reliability and reproducibility. Phenol based extraction techniques appear to offer the best quality RNA, but are highly irreproducible between technical replicates.

Cord blood cells were thawed, washed and resuspended in DAPI to a concentration of $8x10^6$ cells/ml as previously described. Live cells were selected on the basis of FSC and SSC and DAPI negativity, and sorted into polypropylene tubes containing 2ml 2% PBS. Tubes containing 1000 cells and 10000 cells were sorted.

RNA extractions using the three methods detailed above (Qiagen Microkit; DirectZol and traditional Phenol separation) were performed in parallel, and the quality of the yields generated compared by assessment of RIN on the Agilent Pico Kit 6000.

In summary, at an input of 10,000 cells, the RNA quality (as measured by RIN) yielded by all three kits appeared equivalent (mean RIN for Qiagen MicroColumn method 8.8, n=2; mean RIN for DirectZol Column method 8.4, n=1; mean RIN for Trizol 9.1, n=4). In addition, with 10,000 cell input, the reproducibility for all three methods is excellent.

In contrast, when cell inputs are reduced to 1000 cells, both the quality and importantly, the reproducibility of RNA yields, varied enormously between replicates (mean RIN for Qiagen method 6.6, n=3; mean RIN for DirectZol 0, n=2; mean RIN for Trizol 3.7, n=12). The results for Trizol-based extraction in particular proved frustratingly variable: six of the twelve extraction attempts generated detectable RNA, and in these samples, the mean RIN calculated was 7.5. However, a further six attempts generated no detectable RNA at all. Unfortunately, the reproducibility of this technique did not improve with repeated extraction attempts and increased operator experience. Small modifications to the technique (the use of increased LPA and the use of 1-Bromo-3-chloropropane in place of chloroform) made no difference to the observed reproducibility.

We concluded that the quality of RNA yielded by phase separation with Phenol might be marginally better than with the column-based techniques, but that its reduced sensitivity rendered it impossible to use on samples as difficult and costly to extract as the ones we were dealing with. An additional benefit of the use of Qiagen columns is that this technique includes an on-column DNA digestion step, which reduced concerns about genomic DNA contamination in the sequencing data.

I remain sceptical with regard to the utility of RIN based assessment of RNA quality at low levels. Whilst it is possible that the quality of RNA yielded from MicroColumns does fall as

input numbers fall (as we observe a significant fall in mean RIN between 10,000 and 1000 cell input with the Qiagen columns), it is also possible that the effect of a lower RNA concentration at the limit of detection sensitivity of the Agilent Pico 6000 chip, might result in difficulties in accurate assessment of RNA quality.

7.4.2.2 Sample Information

Clinical details pertaining to the 6 $CD34^{low}$ AML and 6 control samples which underwent RNA extraction and subsequent processing are shown in Table 7-10. All of the AML samples, and two of the controls were the same as those utilised in the first experiment. Due to lack of further BM availability in the tissue bank, a further 4 new controls were selected for use.

Table 7-10 Sample information for the 6 $CD34^{low}$ AML and 6 control samples used for RNA extraction in the repeat RNA-Seq experiment

ID	SAMPLE	SEX	AGE	PRESENTATION BLOOD COUNT			KARYOTYPE	*NPM1* STATUS
				HB (g/dL)	WCC ($x10^9$/L)	PLATELETS ($x10^9$/L)		
7793	AML	F	56	11.1	4	10	+13	NPM A
9002	AML	M	68	8.9	0	41	NORMAL	WT
9067	AML	M	67	7.4	142	172	NORMAL	NPM A
8169	AML	F	27	9.7	0	110	t(11;19)	WT
8452	AML	M	27	8.6	6	75	t(11;17)	WT
7774	AML	M	62	8.1	0	37	NORMAL	NPM A
8508	CONTROL	M	22	14.3	3.9	197		
8245	CONTROL	M	45	12.9	7.8	316		
7926	CONTROL	M	42	11.2	14.3	527		
8749	CONTROL	F	70	13.9	7.8	193		
8984	CONTROL	M	42	16.5	6.2	198		
10030	CONTROL	F	68	12.3	7.8	228		

7.4.2.3 Library Preparation

7.4.2.3.1 Sorting and cDNA synthesis

All samples underwent sorting for $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ populations as previously described. In contrast to the first experiment however, rather than sorting a consistent 1100 cells into each FACS tube, all available HSCs for each sample were sorted into a single tube. This was done to try and maximise cell yield prior to RNA extraction, on the basis that RNA quality as assessed by RIN appeared better when starting from a higher number of cells. The number of cells sorted per sample is shown in Table 7-11.

Before commencing SmartSeq cDNA processing, I needed to quantify the RNA concentration for each sample, to ensure equal inputs to the SmartSeq kit. We felt that as RNA extraction efficiency was likely to vary significantly between samples, normalising inputs solely to the basis of the number of cells which underwent RNA extraction was unlikely to be accurate. The RNA concentration of extracted cells was below the sensitivity of the HS RNA Qubit. As a result, the relative concentration of RNA of each sample was calculated by qPCR as detailed in Section 7.4.1.2.3. The results of qPCR quantification are also shown in Table 7-11.

Table 7-11 showing sorted number of cells, relative quantification of RNA by qPCR and volumes added for each sample to SmartSeq experiment

ID	SAMPLE	CELLS SORTED	MEAN Ct FOR GAPDH	VOLUME OF RNA FOR SmartSeq (µl)	RELATIVE RNA YIELD PER CELL
7793	AML	2919	33.9	4.8	1.1
9002	AML	1870	34.0	5.2	1.6
9067	AML	1100	34.9	9.5	1.5
8169	AML	3159	33.7	4.1	1.2
8452	AML	4539	33.4	3.4	1.0
7774	AML	1776	34.1	5.7	1.5
8508	CONTROL	2971	32.1	1.4	3.8
8245	CONTROL	9843	30.6	0.5	3.1
7926	CONTROL	1297	33.0	2.7	4.5
8749	CONTROL	5401	30.9	0.6	4.6
8984	CONTROL	5903	32.7	2.1	1.3
10030	CONTROL	2213	33.5	3.8	1.9

7.4.2.3.2 The amount of RNA per HSC cell is significantly lower for HSCs exposed to AML than from controls

Comparison of the yields of RNA per input cell for the 6 AML and 6 control samples was extremely interesting. As shown in Figure 7-7, the yield of RNA per AML-exposed HSC is significantly lower than that from a control HSC (relative yield 1.3 versus 3.4, p=0.02 by Mann-Whitney t-testing).

The significance of this result is briefly discussed in in Section 7.5.1.3, but the implication that the total amount of RNA per cell might be lower in AML-exposed HSCs would certainly fit with our hypothesis of induced HSC quiescence. It would be reasonable to hypothesise that a quiescent non-dividing cell would have a smaller transcriptome than one more involved in the process of differentiation.

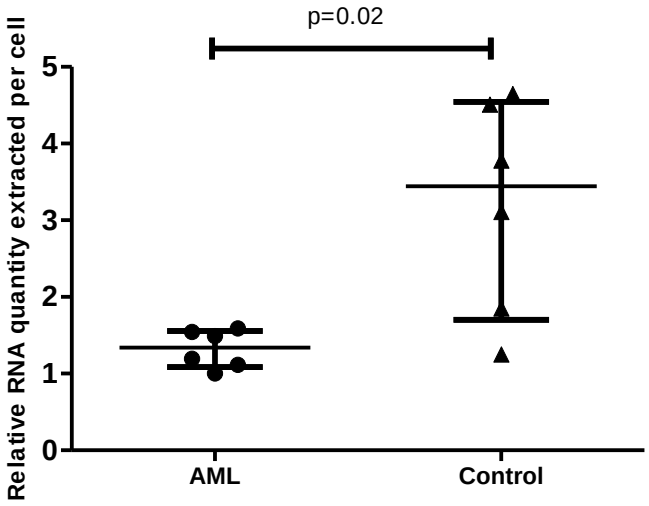

Figure 7-7 Scatter plot comparing the RNA quantities extracted per cell for all 12 samples, expressed relative to lowest value in the group. Medians are AML 1.3 versus 3.4 for control group (n=6 for both). Median and IQR plotted for each group. Statistical significance between the two groups was calculated by Mann-Whitney testing (p=0.02).

7.4.2.3.3 Optimising the number of PCR cycles

In the first experimental attempt using 1100 cells as input to the SmartSeq kit, optimisation work suggested that 10 PCR cycles would be appropriate for amplification of the product of first strand cDNA synthesis before further processing (see Section 7.3.2.1.2 and Figure 7-4). However, as discussed at length in Chapter 7, QC of the resultant amplified cDNA was comprimised by the fact that we did not at that time possess a kit capable of detecting cDNA of lengths above 1000bp.

At the time of this second repeat, QC was able to be carried out using the HS Screentape 5000, which allowed detection of cDNA strands up to 5000 bp long.

RNA extracted in the same fashion as the test samples, and quantified by qPCR assessment of GAPDH, was used for optimising experiments testing the relative appearances and sizes of cDNA libraries generated by 12 (not shown), 14, 16 and 18 cycles of PCR amplification. The results of these optimisation experiments are shown in Figure 7-8. With this sample, the use of 18 PCR cycles appeared to produce a cDNA library of the expected size and within the expected concentration yield (0.8ng/µl).

7.4.2.3.4 Normalising input RNA amounts to GAPDH expression led to overcycling in some samples

As discussed in Section 7.4.1.2.3, the relative concentration of each of the 12 RNA samples was determined by qPCR for GAPDH expression. The volume of input RNA into the SmartSeq kit was based on these normalised values, in an attempt to ensure equal RNA input.

The RNA used for experiments optimising the number of PCR cycles required (as detailed in Section 7.4.1.3.2, and results in Figure 7-8), had been quantified and normalised in the same fashion.

It is interesting therefore, that despite what was deemed to be an equivalent RNA input to the SmartSeq prep, that different samples outputted different size libraries (shown in Table 7-12). This is also illustrated in the corresponding HS TapeStation 5000 traces analysing the samples (Figure 7-9 shows an appropriately sized library, and B shows an overcycled library). As a result, several libraries appeared to be overcycled, with end yields in excess of the upper end of the expected range. As we were keen to avoid overcycling of cDNA where possible to gain a representative library, for those samples which had an end DNA yield significantly above that expected (17ng, corresponding to a concentration of greater than 1.0ng/µl), and where there was sufficient residual RNA available (n=4), SmartSeq preparations were remade, using a smaller input of RNA but the same cycle number. The concentrations of these repeat cDNA libraries are also given in Table 7-12.

7.4.2.3.5 DNA Library Preparation using the Nextera XT Kit

After processing of the cDNA libraries using the Nextera XT kit to fragment, tag with adapters and subsequently amplify the DNA, the resultant DNA libraries were analysed for both average fragment size by HS ScreenTape 1000 and concentration by HS Qubit. Results are shown in Table 7-13. As in Chapter 7, the libraries were also quantified by Kappa PCR, designed to specifically detect DNA appropriately tagged for sequencing. These results are shown in Table 7-14. There was good correspondence between the two methods, but as before, the results generated by Kappa qPCR were used to assign relative concentrations to each library, prior to final pooling (see Table 7-15).

After pooling of equivalent quantities of DNA for each of the 12 libraries, the final library concentration pre-sequencing was estimated by HS Qubit assay at 11.1ng/µl. The average fragment size of the library was also calculated by HS Screentape 1000, and was found to be 452 base pairs. This gives a calculated library concentration (using the formula in Section 7.3.1.4.5.1) pre-sequencing of 37.8nM.

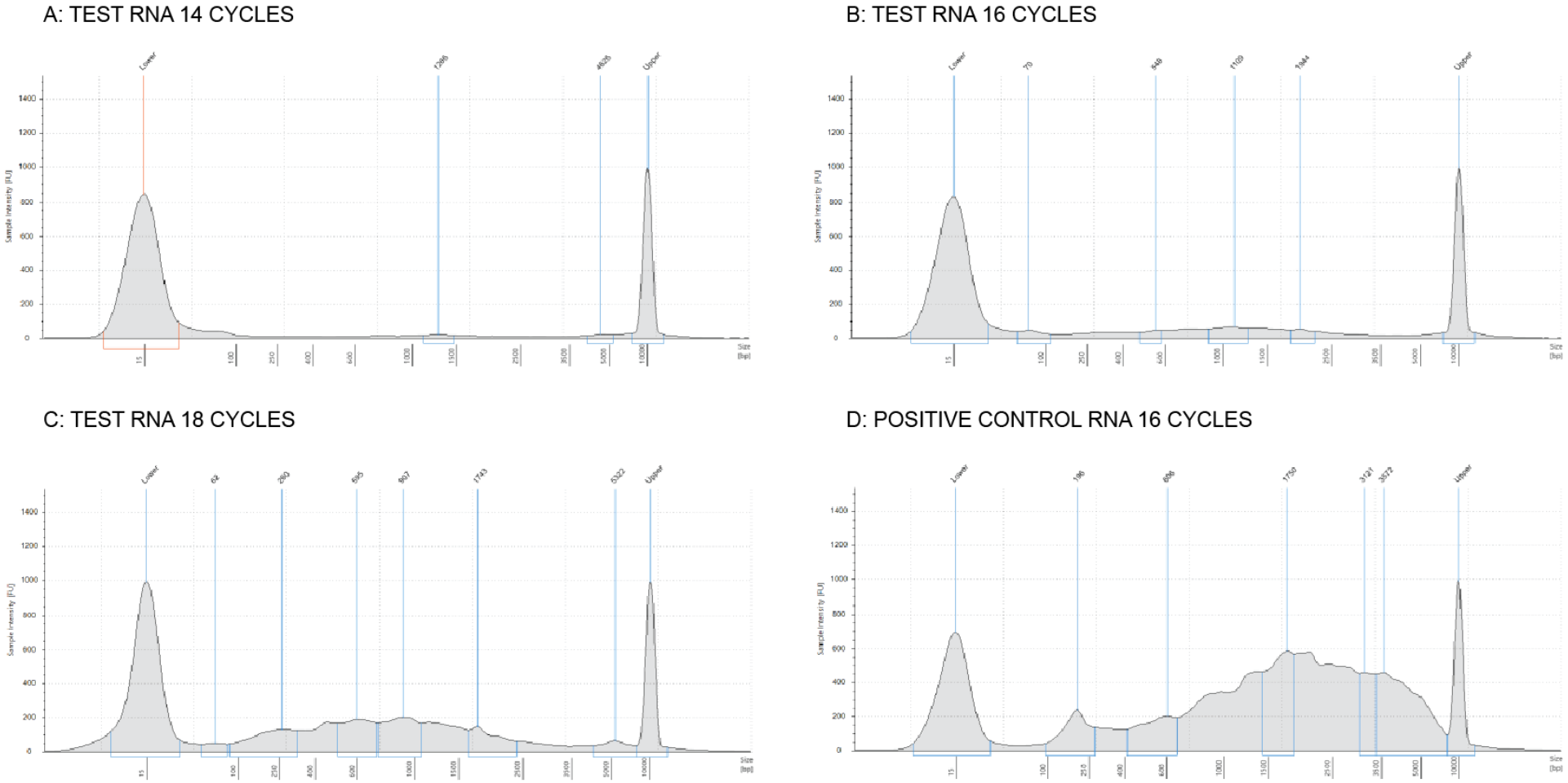

Figure 7-8 Optimisation of the number of PCR cycles used for cDNA amplification after first strand synthesis. Equal inputs of RNA (extracted from DAPI negative cells from cord blood, extracted through Qiagen microcolumns and equivalent to the planned input RNA quantities for the primary samples were amplified 14, 16 & 18 times. cDNA yields post clean-up were 0.184, 0.366 and 0.808ng/µl respectively

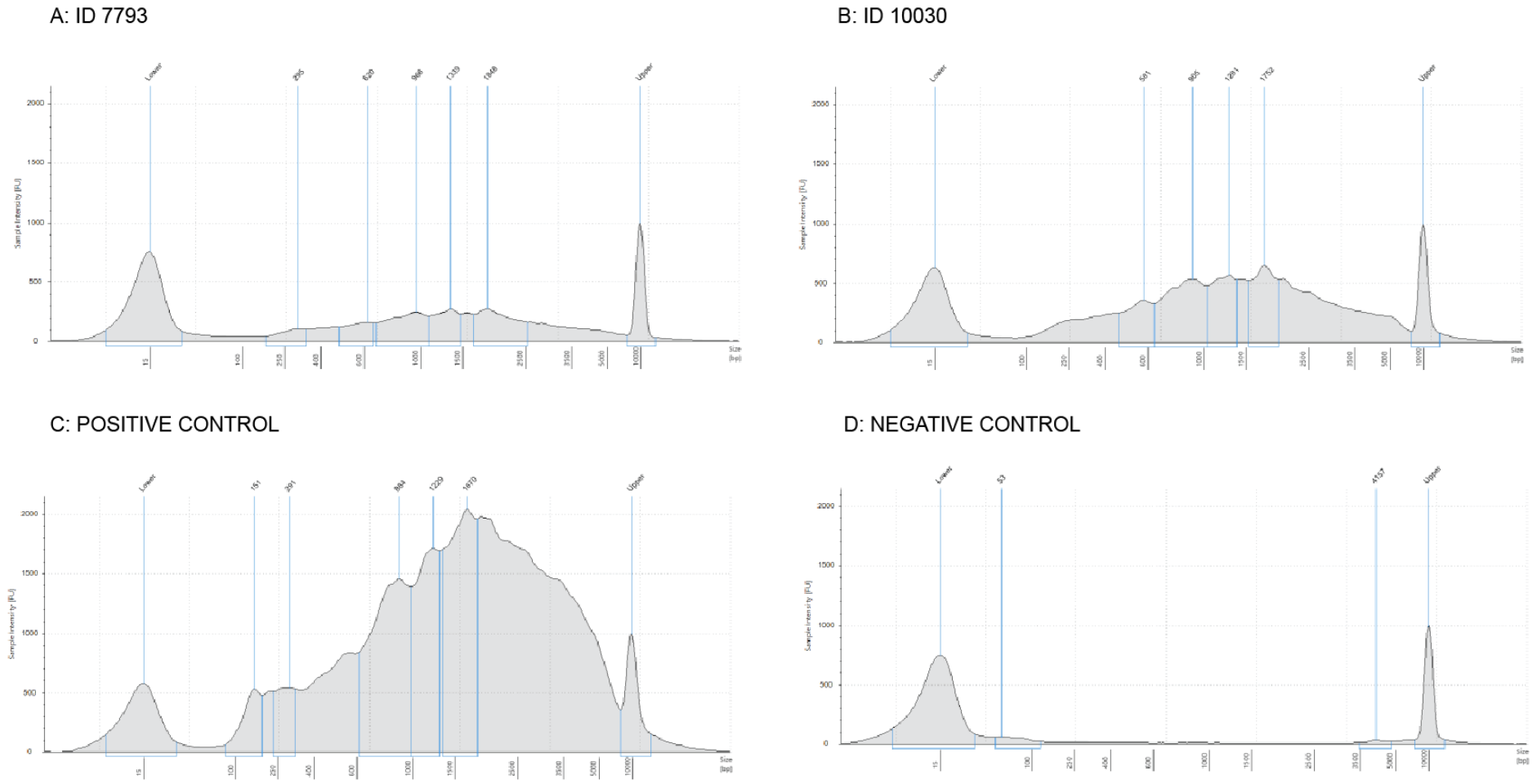

Figure 7-9 SmartSeq appearances of 2 test samples and positive and negative controls, all cycled at 18 cycles. A: successful trace ID 7793 with an end [] of 0.853ng/µl; B: Possibly overcycled sample ID 10030 with an end [] of 1.8ng/µl (above expected range); C: Positive control; D: Negative control

Table 7-12 Concentrations of cDNA libraries post SmartSeq preparation, as estimated by Qubit

SAMPLE ID	PAT ID	POST-CLEAN [] ng/µl RUN 1	POST-CLEAN [] ng/µl RUN 2	VOLUME IN LIBRARY PREP (µl)
1	8452	1.02		0.98
2	8169	1.45		0.69
3	7793	0.858		1.17
4	9002	3.02	0.668	1.5
5	7774	1.26		0.79
6	9067	2.48		0.4
7	8508	1.46	0.728	1.37
8	8245	1.12		0.89
9	7926	2.32		0.43
10	8749	1.05		0.95
11	8984	4.92	0.768	1.3
12	10030	1.8	0.504	1.98

Table 7-13 Concentrations of final DNA libraries post Nextera XT processing, as estimated by Qubit

ID	SAMPLE	INDEX 1 (i7)	INDEX 2 (i5)	AVERAGE LIBRARY SIZE (bp)	QUBIT LIBRARY [] (ng/µl)	CALCULATED LIBRARY [] (nM)
8452	AML	N701	S503	467	7.58	25.0
8169	AML	N702	S503	413	5.82	21.7
7793	AML	N703	S503	519	8	23.7
9002	AML	N704	S503	449	7.72	26.5
7774	AML	N705	S503	511	7.88	23.7
9067	AML	N706	S503	471	8.58	28.0
8508	CONTROL	N701	S504	480	3.68	11.8
8245	CONTROL	N702	S504	447	5.94	20.4
7926	CONTROL	N703	S504	472	11	35.9
8749	CONTROL	N704	S504	445	5.28	18.3
8984	CONTROL	N705	S504	472	7.1	23.1
10030	CONTROL	N706	S504	461	5.64	18.8

Table 7-14 Concentrations of final DNA libraries post Nextera XT processing, as estimated by Kappa qPCR

ID	SAMPLE	AVERAGE LIBRARY SIZE (bp)	ESTIMATED LIBRARY [] BY qPCR (nM)			QUBIT LIBRARY [] (nM) FOR COMPARISON
			1/20000 DILUTION	1/40000 DILUTION	MEAN	
8452	AML	467	26.5	28.4	27.4	25.0
8169	AML	413	26.8	22.7	24.7	21.7
7793	AML	519	23.3	26.2	24.7	23.7
9002	AML	449	30.7	34.8	32.7	26.5
7774	AML	511	22.8	27.2	25.0	23.7
9067	AML	471	45.7	47.8	46.7	28.0
8508	CONTROL	480	12.5	13.5	13.0	11.8
8245	CONTROL	447	20.5	29.6	25.1	20.4
7926	CONTROL	472	39.8	43.1	41.5	35.9
8749	CONTROL	445	27.9	28.4	28.1	18.3
8984	CONTROL	472	27.4	32.7	30.0	23.1
10030	CONTROL	461	26.0	33.3	29.6	18.8

Table 7-15 Volumes used in final DNA library pooling to ensure equal representation of all 12 libraries in sequencing. Relative concentrations were determined by Kappa qPCR results

PAT ID	SAMPLE	LIBRARY [] BY qPCR (nM)	RELATIVE CONCENTRATION	VOLUME ADDED (µl)
8452	AML	27.4	2.1	9.5
8169	AML	24.7	1.9	10.5
7793	AML	24.7	1.9	10.5
9002	AML	32.7	2.5	8.0
7774	AML	25.0	1.9	10.4
9067	AML	46.7	3.6	5.6
8508	CONTROL	13.0	1.0	20.0
8245	CONTROL	25.1	1.9	10.4
7926	CONTROL	41.5	3.2	6.3
8749	CONTROL	28.1	2.2	9.3
8984	CONTROL	30.0	2.3	8.7
10030	CONTROL	29.6	2.3	8.8

7.4.2.4 Sequencing data

7.4.2.4.1 Proportion of reads mapping to coding regions of genes is significantly improved when RNA extraction is performed prior to first strand synthesis

In attempt 1, when cells were used directly as the source of RNA for cDNA synthesis, the proportion of reads mapped to non-coding regions of the genome ranged from 42 to 88% between samples. However, when RNA extraction was performed prior to cDNA synthesis using the SmartSeq kit, the proportion of reads mapping to non-coding regions dramatically reduced to between 7 and 34% per sample. These data are shown in Table 7-16. This compares favourably with the expected data from the company literature[347].

Table 7-16 Percentage of RNA-Seq reads mapping to coding and non-coding regions for each sample

ID	SAMPLE	TOTAL NUMBER OF READS	% READS UNIQUELY ALIGNED TO GENOME	% ALIGNED READS MAPPED TO INTRONIC REGIONS
7793	AML	62261501	62	13
9002	AML	60122953	61	34
9067	AML	51128266	64	13
8169	AML	58066526	57	26
8452	AML	70935294	62	27
7774	AML	66088268	62	8
8508	CONTROL	63003682	58	21
8245	CONTROL	55627489	57	7
7926	CONTROL	58162995	61	11
8749	CONTROL	50223437	57	13
8984	CONTROL	54387966	60	20
10030	CONTROL	43635497	58	14

7.4.2.4.2 Differentially expressed genes

7.4.2.4.2.1 Principal Component Analysis

Principal component analysis of the 12 samples, performed and graphically illustrated by Ai Nagano (Bioinformatics Department, BCI) shows weak clustering into two groups as shown in Figure 7-10.

7.4.2.4.2.2 Individual genes of interest

Filtering revealed differential expression of 14624 genes between normal HSCs from AML samples and controls. Of these, 243 showed differential expression with log fold changes of

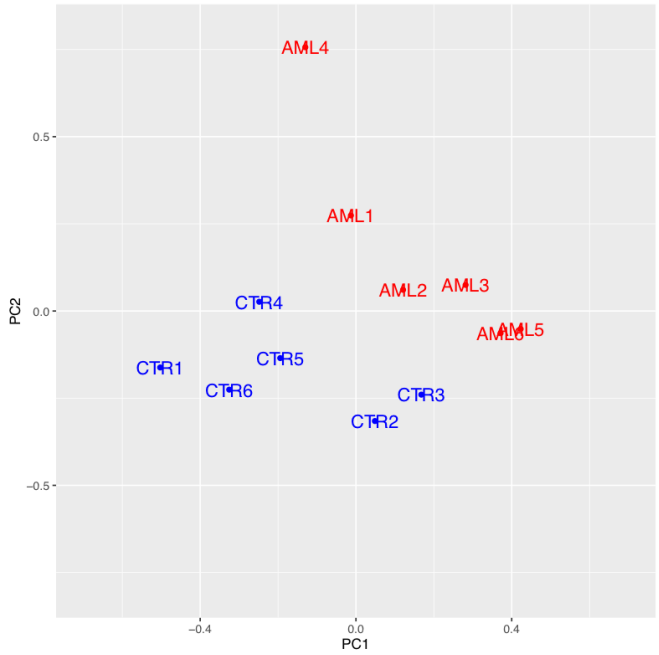

Figure 7-10 Principal Component Analysis illustrating weak clustering of the 12 samples into two groups consistent with their origin (analysis and graphic kindly supplied by Ai Nagano)

greater than two, calculated p-value of less than 0.01. These are listed in Appendix 4 of this thesis. By comparison, only 37 genes were differentially expressed by the same criteria in the original experiment.

Six genes were differentially expressed in a statistically significant fashion when allowing for multiple testing (an adjusted p value of <0.05 calculated using the Benjamini–Hochberg procedure). Of these, one gene showed significantly increased expression in control HSCs versus AML-exposed HSCs (*GNPDA1*); and five showed significantly increased expression in AML-exposed HSCs versus control HSCs (*ADGRG3*, *MIAT*, *WDR31*, *RP11-244H3.1* and *RXFP1*). Reported data on these genes are summarised below. In the first experimental attempt, no genes were significantly differentially expressed after the Benjamini-Hochberg test had been applied.

7.4.2.4.2.2.1 Genes relatively downregulated in HSCs exposed to AML compared to controls

7.4.2.4.2.2.1.1 GNPDA1 (Glucosamine-6-phosphate deaminase 1)

GNPDA1 (also known as oscillin) is an allosteric enzyme that catalyses the reversible conversion of D glucosamine-6-phosphate into D-fructose-6-phosphate and ammonium[348]. Originally identified as a sperm-derived factor responsible for calcium flux within oocytes, it is now known to be ubiquitously expressed and is therefore thought to have a housekeeping role. There are no published associations with haematopoietic development, AML or control of cell division.

7.4.2.4.2.2.2 Genes relatively upregulated in HSCs exposed to AML

7.4.2.4.2.2.1 ADGRG3 (Adhesion G protein coupled receptor G3)

ADGRG3 is an orphan receptor which acts to regulate endothelial cell migration via RhoGTPases and CDC42. Other members of the adhesion GPCR family function to control cell polarity, migration and adhesion. Up and down regulation has been noted in a variety of cancer types[349].

A related family member, GPR56, has a role in haematopoietic stem cell formation. Saito et al demonstrated high GPR56 expression in EVI1-positive AML cells. Knockdown of the gene in AML cells resulted in reduced cellular adhesion. In murine *GPR56* knock-outs, HSC numbers were significantly reduced in the BM and increased in liver and spleen and blood, with concurrent loss of their repopulating ability[350].

7.4.2.4.2.2.2 MIAT (myocardial infarction associated transcript)

MIAT (also known as RNCR2 (retinal non-coding RNA2)) is a long, non-coding RNA. These RNAs are postulated to play a role in gene expression regulation. Unlike many non-coding RNAs, MIAT is polyadenylated, and therefore would be expected to be amplified by the SmartSeq kit.

Although most studies have focussed on an association between expression and increased cardiac risk[351], MIAT expression has also been associated with a variety of cancers. In glioblastoma cases, MIAT upregulation has been linked to increased survival rates[352]. Conversely, in CLL, MIAT upregulation under the control of OCT4 is associated with more aggressive disease[353].

7.4.2.4.2.2.3 WDR31

WDR31 is a gene encoding a member of the WD repeat protein family. These are minimally conserved regions of 40-odd amino acids, which facilitate formation of multiprotein complexes. Other family members have a variety of reported roles, including cell cycle progression, signal transduction and apoptosis, but to date, none have been specifically reported for WDR31.

7.4.2.4.2.2.4 RP11-244H3.1

RP11-244H3.1 is a non-coding long RNA with no reported functions or associations to date.

7.4.2.4.2.2.5 RXFP1

RXFP1 is a human G protein coupled receptor, which is one of the relaxin receptors. There are no published associations with haematopoiesis, AML or cell cycle control to date.

7.4.2.4.3 Pathway analysis

Gene enrichment analysis performed using the DAVID functional annotation chart is summarised in Figure 7-11[354]. The input dataset contained all genes differentially expressed between HSCs from AML samples and controls, with a fold change of greater than 2, and individual p value of <0.05. Of 1013 genes, 810 were recognised by DAVID and used for further analysis.

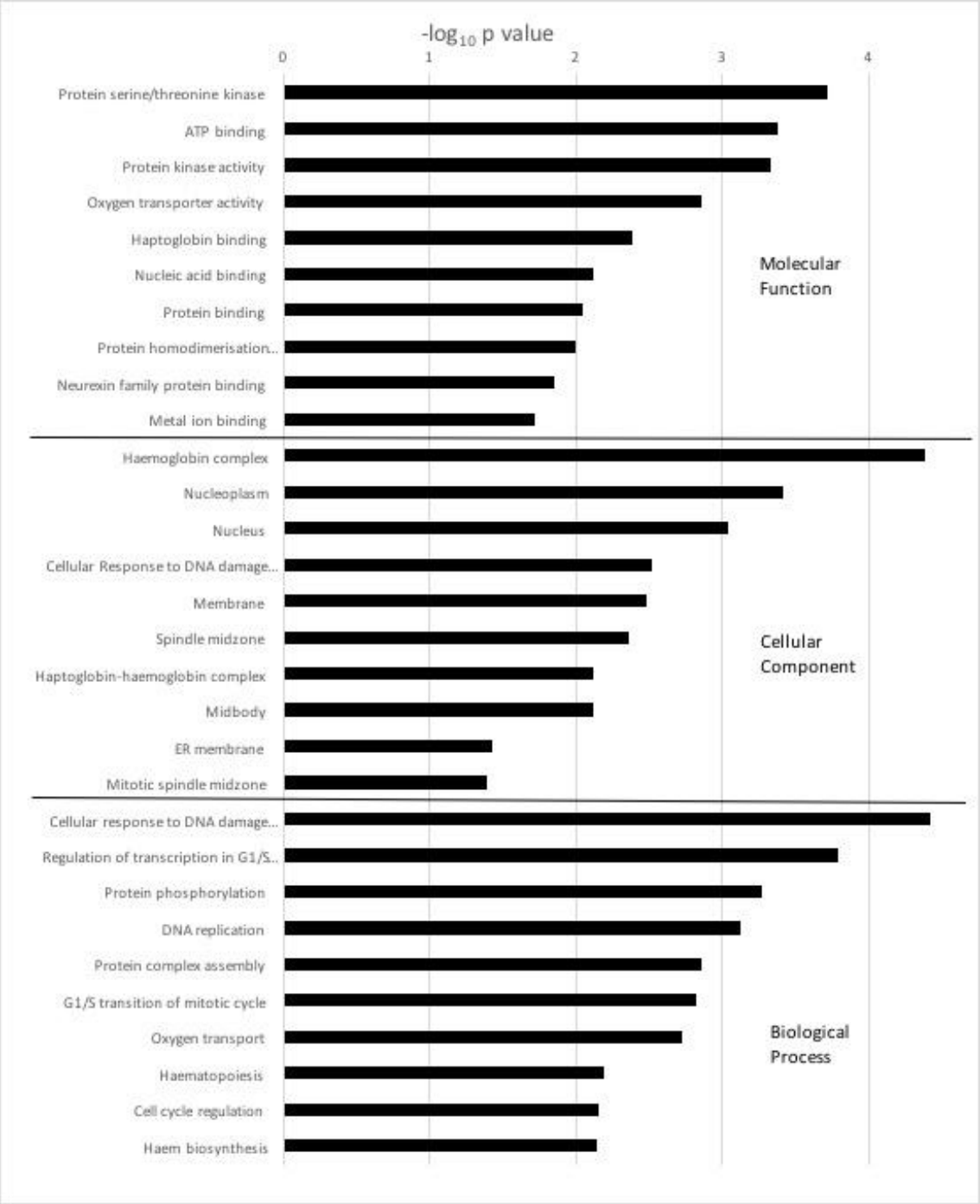

Figure 7-11 Gene enrichment analysis of differentially expressed genes from HSCs from AML samples and control, generated by the DAVID functional annotation chart option. Enriched categories of molecular function, cellular compartments, and biological processes are shown. Significance of enrichment is shown by EASE score (modified Fisher's Exact Score). The illustrated figures are NOT corrected for multiple testing

Of particular interest are the biological processes which show enrichment within this gene set. As might be expected, given the previous observed quiescence of HSCs in AML samples, there is enrichment in genes involved in control of cell cycle, especially around key points such as the G1/S transition. Similarly, given the assumed lack of downstream differentiation of HSCs, it is interesting that genes associated with haematopoiesis and haem synthesis also strongly appear.

However, it must be highlighted that none of the functions highlighted by David using this input set of genes are significant when correction is made for multiple testing using the Benjamini-Hochberg procedure. As before, one wonders if the limited number of samples along with the variation seen between biological samples may be responsible.

7.4.2.4.3.1 KEGG pathway analysis

Enrichment analysis of specific biological pathways was performed using the KEGG pathway analysis tool integral to David[346]. Table 7-17 summarises the top enhanced pathways (all with an individual EASE p value of <0.05). As before, none of these pathways retain statistical significance after the application of the Benjamini-Hochberg procedure.

I have chosen to discuss in further depth the significance of the Tumor Necrosis Factor (TNF) signalling and cell cycle control pathway.

Table 7-17 Enrichment of genes within specific biochemical pathways differentially activated between HSCs from AML and control samples, as generated by KEGG analysis. Significance of enrichment is shown by EASE score (modified Fisher's Exact Score)

TERM	NUMBER OF GENES	EASE SCORE P VALUE	P VALUE AFTER BENJAMINI CORRECTION
TNF signalling pathway	11	0.003	0.51
Non-alcoholic fatty liver disease (NAFLD)	12	0.014	0.79
Toxoplasmosis	10	0.019	0.76
Cell Cycle	10	0.025	0.76
Toll-like receptor signalling pathway	9	0.027	0.72
Fanconi anaemia pathway	6	0.033	0.72
NOD-like receptor signalling pathway	6	0.038	0.72
Sulphur relay system	3	0.043	0.72
Viral carcinogenesis	13	0.046	0.69

7.4.2.4.3.1.1 TNF signalling

11 genes within the TNF signalling pathway show significant changes in expression between the two subgroups of HSCs, making it the most enhanced pathway on KEGG analysis (p=0.003). Figure 7-12 summarises the pathway, and highlights the genes within it which show differential expression. Of the 11 genes, 8 show increased expression in HSCs exposed to AML. It is harder to plausibly explain the increased expression of three genes in the pathway (*MAPK11* (p38), *MAP2K6* and *TAB1*) in HSCs from control samples relative to AML.

A relative activation of the TNF signalling pathway in HSCs exposed to AML is certainly highly biologically plausible, and provides a potential mechanism by which HSC quiescence could be induced.

Identified over 30 years ago, TNF is produced as a type II transmembrane protein, capable of signalling with its cognate receptors both in membrane and soluble form when cleaved[355]. There are two receptors: TNF-R1 is ubiquitously expressed, whereas TNF-R2 is found on lymphoid cells[356]. TNF has been reported to have a wide range of roles dependant on cell type and activated receptor, but in general mediates pro-inflammatory behaviour, and retains ability to induce cellular apotosis[357].

Increased levels of TNF have been illustrated in the peripheral blood and bone marrow of patients with AML, and are associated with high WCC disease and poor prognosis[358]. Increased levels of TNF have been documented in patients with bone marrow failure from other causes (AA, MDS), leading to the suggestion it may be involved in the development of cytopenias in these conditions[359,360].

Murine gene knock-out studies exploring the role of TNF in HSC regulation suggested that TNF signalling (requiring both TNFrsf1b and Tnfrsf1a activation) had a suppressive effect on HSC expansion[361].

Goselink et al reported the ability of AML to inhibit the proliferation of human progenitor cells when culture was separated by transwells of a pore size of 0.45µm. Interestingly, the addition of anti-TNF α inhibitor restored normal HPC proliferation. However, the authors also noted this led to significantly increased proliferation of AML blasts, an observation they felt would limit any therapeutic benefit[362].

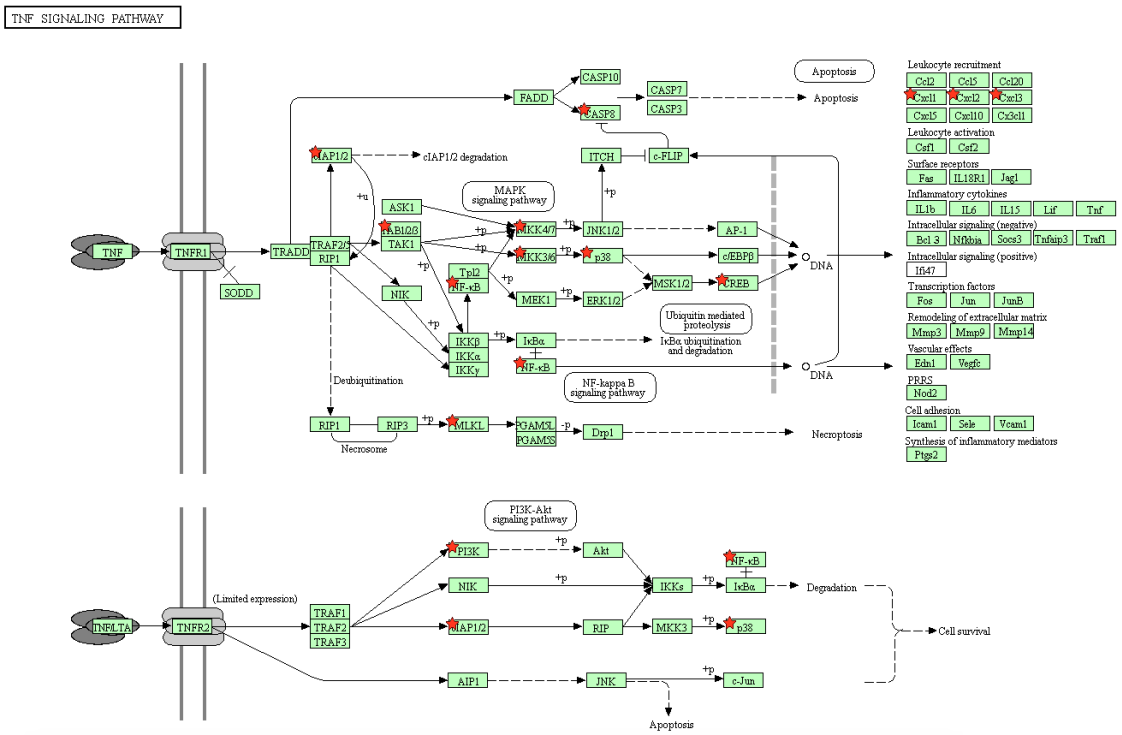

Figure 7-12 KEGG generated figure showing key genes involved in TNF signalling. Genes differentially expressed between HSCs from AML and control samples (fold change >2, EASE score p value <0.05) are highlighted with a star

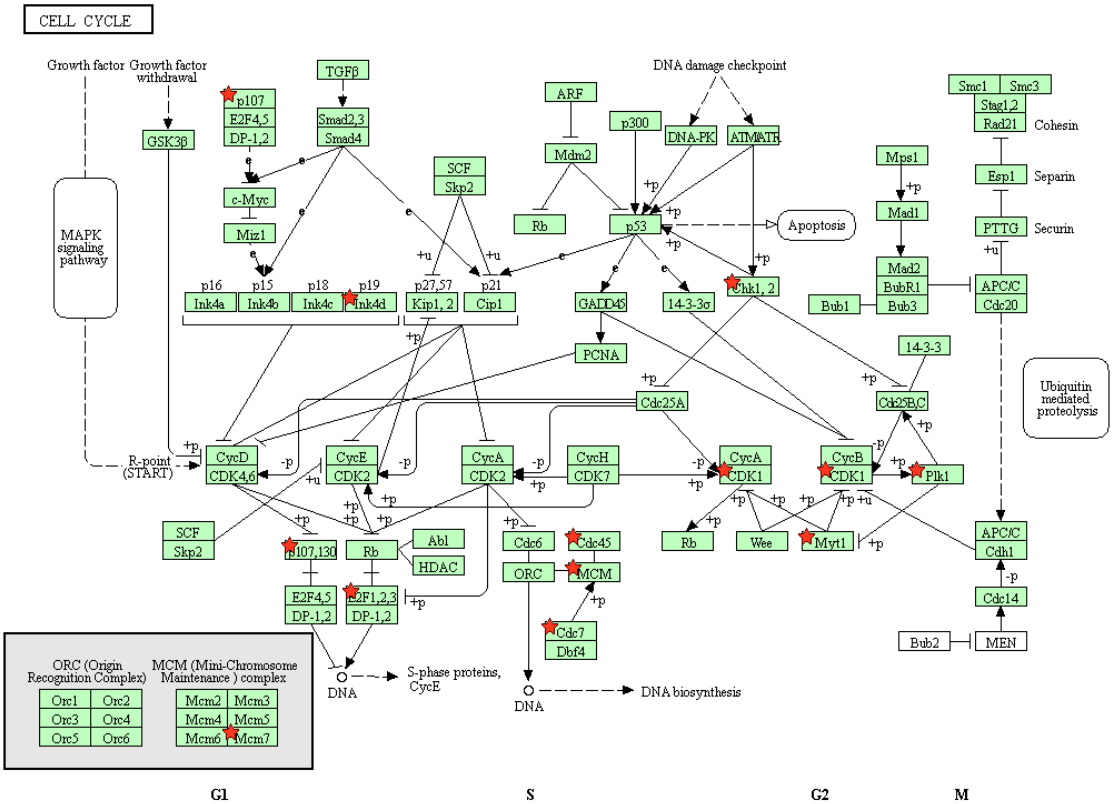

Figure 7-13 KEGG generated figure showing key genes involved in cell cycle control. Genes differentially expressed between HSCs from AML and control samples (fold change >2, EASE score p value <0.05) are highlighted with a star

7.4.2.4.3.1.2 Cell Cycle Control

As one of the key hypotheses of this thesis is that HSCs from AML samples are relatively quiescent when compared to those from controls, it is somewhat reassuring to see that the pathways of cell cycle control are highlighted as one of the key differential pathways on KEGG analysis. This is summarised in Figure 7-13. Of the 10 genes which exhibit significant differential expression, all but *CDKN2D* (part of the Ink4d complex) are more highly expressed in control cells than AML cells. This is consistent with the idea that HSCs from AML cells are more likely to be quiescent; as in general these genes are upregulated during the process of cell division. The only exception to this pattern is *RBL1*- its relative upregulation in HSCs from control samples is difficult to explain, as increased expression is normally associated with blockade of the G1/S check point[363].

7.4.2.4.4 Comparison with other Datasets

7.4.2.4.4.1 There is little overlap between the datasets produced in the first and second RNA-Seq experimental attempts

Genes which show significant differential expression between HSCs from AML samples and controls in both the first and second RNA-Seq experiments are shown in Table 7-18. Only 13 genes showed marked differences in both experiments. Of these, 10 showed changed in expression in a consistent direction in the two experiments.

Potential reasons for the lack of concordance between the two datasets are further explored in Section 7.5.2.1.

7.4.2.4.4.2 There is little overlap between this dataset and murine xenograft data

Comparison was also made between the dataset generated by this experiment, and that detailed in Section 7.1.2.1 resultant from the transplantation of mice with one human AML sample (and comparison with HSCs from untransplanted mice). Seven genes appeared in the top 200 differentially expressed genes within both datasets. However, of these, only two exhibited differential expression between HSCs from AML and controls in consistent directions in both datasets. Expression of *ANK1* was relatively increased in HSCs extracted from control BM; expression of *TMEM53* was relatively increased in HSCs from AML samples. Whilst the protein product of the *TMEM53* gene has no clear reported roles to date, the functions of ANK1 are better defined.

The *ANK1* gene codes for a protein involved in linking proteins integral to the cell membrane to the underlying actin cytoskeleton. Clinically, mutations in *ANK1* are associated with hereditary spherocytosis[364]. In health, its expression is associated with erythroid maturation

from HSCs and switching[365]. Within AML samples, microarray studies have shown its expression to be relatively reduced in *FLT3* ITD-mutated samples versus WT samples of normal cytogenetics[366].

Table 7-18 illustrating 13 genes which showed fold changes >2 between HSCs from AML and controls samples in BOTH RNA-Seq experiments (3 genes in italics showed marked fold changes in gene expression in non- consistent direction between the 2 experiments)

			Attempt 1				**Attempt 2**			
ID	**Gene**	**Description**	**Log fold change**	**Upregulated**	**P.Value**	**adj.P. Val**	**Log fold change**	**Upregulated**	**P.Value**	**adj.P. Val**
ENSG00000164120	*HPGD*	*hydroxyprostaglandin dehydrogenase 15-(NAD)*	*2.34*	*up in CTR*	*0.0377*	*0.9998*	*4.41*	*up in AML*	*0.0035*	*0.4387*
ENSG00000204644	ZFP57	ZFP57 zinc finger protein	2.00	up in CTR	0.0132	0.9998	6.06	up in CTR	0.0083	0.5495
ENSG00000149257	SERPINH1	serpin peptidase inhibitor, clade H (heat shock protein 47), member 1, (collagen binding protein 1)	1.78	up in CTR	0.0050	0.9998	5.89	up in CTR	0.0014	0.3252
ENSG00000080839	RBL1	retinoblastoma-like 1	1.37	up in CTR	0.0421	0.9998	5.24	up in CTR	0.0057	0.5015
ENSG00000065054	SLC9A3R2	solute carrier family 9, subfamily A (NHE3, cation proton antiporter 3), member 3 regulator 2	1.26	up in CTR	0.0474	0.9998	4.05	up in CTR	0.0046	0.4796
ENSG00000097046	CDC7	cell division cycle 7	1.26	up in CTR	0.0285	0.9998	5.67	up in CTR	0.0063	0.5197
ENSG00000113621	TXNDC15	thioredoxin domain containing 15	1.09	up in CTR	0.0144	0.9998	5.54	up in CTR	0.0014	0.3252
ENSG00000120833	SOCS2	suppressor of cytokine signaling 2	1.47	up in AML	0.0180	0.9998	1.72	up in AML	0.0089	0.5495
ENSG00000101400	*SNTA1*	*syntrophin, alpha 1*	*1.54*	*up in AML*	*0.0118*	*0.9998*	*5.11*	*up in CTR*	*0.0089*	*0.5495*
ENSG00000197329	PELI1	pellino E3 ubiquitin protein ligase 1	2.02	up in AML	0.0056	0.9998	3.80	up in AML	0.0030	0.4235
ENSG00000180573	HIST1H2AC	histone cluster 1, H2ac	2.27	up in AML	0.0400	0.9998	2.28	up in AML	0.0084	0.5495
ENSG00000146592	CREB5	cAMP responsive element binding protein 5	2.43	up in AML	0.0028	0.9998	4.26	up in AML	0.0033	0.4302
ENSG00000122824	*NUDT10*	*nudix (nucleoside diphosphate linked moiety X)-type motif 10*	*2.64*	*up in AML*	*0.0188*	*0.9998*	*5.16*	*up in CTR*	*0.0030*	*0.4235*

7.5 Discussion

This chapter follows our attempts to identify the mechanism or mechanisms by which AML induces quiescence in the HSC population, by studying the comparative transcriptomes of HSCs sorted from primary AML and control BM samples.

The work involved provided an excellent opportunity to learn a host of new techniques required for the handling of DNA and RNA in small amounts. The first experiment resulted in data of disappointing quality, due to heavy contamination with reads mapped to intronic DNA. Perhaps unsurprisingly, these results failed to be replicated in validation experiments. A repeat experiment, using extracted RNA as source material for subsequent library preparation, has yielded a very different dataset, of substantially improved quality in terms of proportion of reads mapped to coding regions of the genome.

Data from the second experimental attempt were received after my PhD laboratory time was completed; hence, I have been unable to perform validation experiments on the genes of interest which have been highlighted by this work.

7.5.1 Was failure of selective amplification of polyadenylated mRNA within the SmartSeq kit the cause of genomic contamination of the first experiment?

As discussed in Section 7.3.2.4.1, our first attempt at studying the comparative transcriptomes of sorted HSCs from AML and control samples was frustrated by the fact that a large proportion of sequenced reads mapped to genomic rather than coding DNA regions.

The reason for this is not entirely clear, but our eventual favoured hypothesis was that relying on selection of polyadenylated transcripts by the SmartSeq kit prior to first strand synthesis was not specific enough to prevent the concurrent amplification of DNA. As we used intact cells as the direct input material to the SmartSeq kit (as opposed to pre-extracted RNA), the quality of the data generated is entirely dependant on this specificity of this amplification. As already mentioned, external genomic DNA contamination was unlikely due to the consistent presence of successful negative controls at all stages of library preparation.

Quite why the SmartSeq kit might fail in this manner remains open to conjecture. It is a widely used kit for handling of small numbers of cells, and its use was recommended by several independent groups. However, all of our contacts had used pre-extracted RNA as a starting material for cDNA synthesis, rather than cells.

7.5.1.1 QC of cDNA production with the HS Screentape D1000 was inadequate

In retrospect, our lack of access to the appropriate kit to QC at the end of cDNA generation (see Section 7.3.1.3.4) almost certainly meant we failed to pick up experimental issues at the earliest possible stage. This situation has now been rectified with the availability of the HS D5000 Screentape, which was used for QC on the repeated experiment (see Section 7.4.1.3.4 and Figure 7-8 and Figure 7-9).

7.5.1.2 Could lengthy sorting contribute to sample apoptosis?

Why feasibly might we have been faced with difficulties in the use of this kit that others have not come up against? In discussions with Clontech, the company conceded that if DNA was present in single stranded form in the reaction mix (for example, when released by apoptotic cells), it is possible that this might be amplified along with mRNA. The lengthy sorting process that is required to separate normal HSCs from AML samples (between one to two hours per sample) is significantly more than most samples would have to undergo. Although HSCs were DAPI negative at the point of sorting, it is possible that a proportion of them will have been damaged enough from the process to have begun to apoptose by the time of cDNA synthesis.

7.5.1.3 Do quiescent HSCs contain relatively less RNA per cell than other cell types?

We know very little about the relative size of transcriptomes between different cell types, or indeed between cells at different stages in the cell cycle. However, from all that we understand about HSC biology, it is reasonable to hypothesise that these metabolically inactive, quiescent cells should have a limited transcriptome. This would result in a relatively low RNA:DNA ratio per cell, which might increase the likelihood of contamination from DNA amplification within the SmartSeq kit (see Section 7.3.2.4.2).

Two observations suggest that the RNA yield per HSC is low in comparison to other cell types. Clontech issue a rough guide to the expected number of PCR cycles to amplify cDNA dependent on the input cell number. When using either cells or RNA as input, the actual number of cycles used (10 and 18 respectively) were significantly in excess of that recommended (see Figure 7-8). Secondarily, when single HSCs from AML samples underwent single cell sorting and cDNA production at the Single Cell Genomics Core Facility in Cambridge, the team there also suggested the cDNA required amplification with an unusually high number of PCR cycles (22) before detection.

It is also interesting that, as shown in Figure 7-7, the relative yields of RNA per cell from HSCs extracted from AML samples were statistically smaller than those extracted from control

samples, which is in keeping with our hypothesis of induced quiescence in the HSCs exposed to AML.

7.5.2 Reasons for failure of validation of selected genes of interest by flow based protein assessment and qPCR based analysis

Despite the concerns regarding the quality of the data generated by the first RNA-Seq experiment, I attempted to see if the observed changes in gene expression in the top target genes could be validated by qPCR analysis and flow based assessment of protein expression.

Of the 8 genes selected for assessment of gene expression validation, only one (CD33) confirmed differential expression between the HSCs exposed to AML and normal BM that was statistically significant (by flow based assessment of protein expression, but not by qPCR based assessment of mRNA levels).

7.5.2.1 Data quality of the first experiment is poor enough to explain lack of gene validation

When we were attempting gene validation, we were aware of apparent significant read contamination with genomic DNA. We assumed this represented internal contamination, due to the failure of the SmartSeq kit to solely select and amplify polyadenylated mRNA. In attempting to analyse the data further, we made a secondary assumption that the amplification of any DNA would occur randomly. Therefore, if a differential number of reads appear to map to a gene between the two sample cohorts, this could be assumed to be due to pre-existing differences in mRNA levels, and therefore suggestive of a true biological difference. Against this argument is the observation that the level of genomic contamination appears to vary between samples (see Table 7-7), which given the small input number of samples, might be a significant factor.

In the light of the dataset subsequently received from the repeat experiment, the lack of validation by independent methods of results from the first experiment seems less surprising. None of the 8 genes chosen for validation appeared differentially expressed in the second dataset. Whilst poor data quality due to genomic DNA contamination might be the primary reason why validation of the first dataset failed, there are several other potential contributing factors, discussed below.

7.5.2.2 Are differences in gene expression between control samples sufficient to prevent validation of the initial RNA-Seq findings?

Even without the concerns over the quality of the initial RNA-Seq data, it is possible that the fact that different control samples than those used in the primary experiment had to be sorted

for RNA extraction, might result in differences between the control cohorts. As shown in Table 7-9, because insufficient further BM sample was available in the Tissue Bank for 4 of the control samples sorted in the original experiment, HSCs were sorted from a further 4 independent samples.

Previous work using primary human samples (see Section 7.1.2.2) had suggested there is considerable variation between the gene expression profiles of control samples. This might be sufficient to prevent the validation of a genuine difference between AML-exposed HSCs and the HSCs derived from one cohort of control samples, using another cohort for validation. However, the obvious counter-argument to this is that such an approach should filter out non-biologically significant differences. Such variability certainly presents a strong argument for the use of larger numbers of both control and AML samples to be studied in each cohort group in subsequent work.

7.5.2.3 HSC-derived RNA was dilute enough to be at the borderline of qPCR detection for many genes of interest

As discussed in brief in Sections 7.3.1.7.1 and 7.3.2.5.4, the concentration of the RNA extracted from the available numbers of HSCs (between 1000 to 5000 per sample) was sufficiently dilute, that the resultant cDNA (despite the optimisation of reverse transcription described in Section 7.3.1.7.1.2) was only detected in the final cycles of qPCR. This led to a number of problems during analysis.

The cDNA concentrations of all samples were so low, that even relatively highly expressed house-keeping genes such as GAPDH were only detected between 28 to 35 qPCR cycles for most samples. Despite selecting genes for validation with apparently high expression levels (as defined by RPKM values from the RNA-Seq data) in anticipation of this issue, for several genes (*S100A10*, *IDH1* and *TRAF4*), expression was not detected by 38 qPCR cycles in most cycles. The quality and reproducibility of data generated for genes detected beyond 35 qPCR cycles is widely considered to be poor, and so interpretation of differences in actual gene expression between samples with low expression is highly challenging.

7.5.2.3.1 Would pre-amplification of low concentration RNA before qPCR-based validation have been helpful?

One solution to the difficulty of working with RNA or cDNA of low concentration is to pre-amplify samples before qPCR based gene expression analysis, using one of a range of commercially available kits. Such kits are designed to amplify the different mRNA transcripts within a sample equally, thus avoiding bias. In retrospect, such a step would not only avoid the

practical difficulties of interpreting gene expression levels of genes detected after 35 cycles of qPCR, but would also produce sufficient cDNA to allow for the validation of many more genes.

However, such an approach is not without its disadvantages. The use of a pre-amplification step brings this qPCR-based assessment of gene expression levels very close to the methodology of RNA-Seq, and therefore subject to the same potential bias of unequal amplification of specific genes. A secondary consideration was cost- current commercial kits are in excess of £1000 for 12 samples, and at this stage in my thesis, and while attempting to validate data of questionable quality, this was deemed to be unjustifiable expenditure.

7.5.3 Using RNA as a starting material leads to better quality data

We made the decision to concentrate on repeating the original experiment using RNA extracted as the source of the starting material, rather than cells. In light of the difficulties described above, we were supplied with free replacement SmartSeq kit by Clontech for cDNA synthesis.

Given our difficulties with using sorted cells as a starting point for the RNA-Seq experiment, considerable care was undertaken to optimise the best technique for extracting quality RNA from these small cell numbers sufficient to use in further processing.

7.5.3.1 What was the best RNA extraction methodology?

As detailed in Section 7.4.2.1, the use of Qiagen columns for RNA extraction appeared to provide a reasonable compromise of quality of material and reproducibility required for RNA extraction from around 1000 cells. However, questions remain over the quality of the RNA extracted by this methodology, which during optimising experiments using 1000 cells was assessed by RIN as being as low as 3 (in one of three samples).

Other laboratories working with small cell numbers (J. Brown, UCL, personal communication) prefer Phenol based extraction methods, which appeared to offer a higher RNA quality as assessed by RIN at low cell number, but poor reproducibility. It was the poor reproducibility of this approach which proved crucial for me in rejecting this methodology, given the limited number of available samples, as well as the cost and time required to sort out stem cells.

If this work were to continue, I would be interested in trying to optimise the Phenol extraction method, to see if repeated attempts could improve its reliability in my hands.

Whilst it appears that the proportion of reads mapping to intronic regions was significantly lower when prior RNA extraction was used, RNA separation from DNA is clearly not perfect, in that between 6 to 33% of reads still mapped to intronic regions (see Table 7-16). This is suggestive of a degree of ongoing DNA contamination at varying levels within the samples, despite the fact that the Qiagen columns did contain an on-column DNA digestion step. In retrospect, I should have included a QC step to check for the presence of DNA contamination after RNA extraction (by qPCR for a gene using primers spanning an intron)

7.5.3.2 Second dataset contained more genes with statistically significant changes in expression levels between the two HSC groups, but only 6 which retained significance when allowing for multiple testing

Using RNA as a starting material for library preparation also appears to have produced a more "uniform" dataset in that many more genes appeared to be differentially expressed between the normal HSCs within AML and control samples.

However, it is interesting that only six genes retained significantly different levels of expression after allowance is made for multiple testing. Of these genes (and their close structural homologues), only three have been reported to have roles in cell cycle control and differentiation. Given this experiment was designed to detect genes which might be responsible for mediating the observed quiescence of normal HSCs from AML samples, I had perhaps naively assumed we might detect more changes of significance between the two groups. I have not been able to find published datasets from similar experiments using the SmartSeq kit to compare with.

In comparison, the experiment described in Section 7.1.2.1, comparing the transcriptomes of HSCs from 7 mice transplanted with AML and matched controls, revealed 54 genes to be differentially expressed between the two groups after allowance for multiple testing. However, this is perhaps not a fair comparison- all 7 mice were transplanted with the same sample of AML, unlike the diversity of AMLs represented in our samples. Given the documented variability of primary human samples already noted in Section 7.1.2.2, the first improvement to this experiment would be to increase substantially the samples of both AML and controls to improve the statistical power of any results.

7.5.4 Further experimental work

7.5.4.1 Exploration of the second RNA-Seq dataset

7.5.4.1.1 Gene expression validation

The RNA-Seq data generated by the second attempt at this experiment were received after my time in the laboratory had concluded. Therefore, I did not have an opportunity to attempt to validate the observed changes in gene expression between HSCs from AML and control samples, as I had done with the first experiment. This is the most pressing outstanding work required in this thesis.

The obvious candidates for validation of differential expression would be the genes highlighted in Section 7.4.2.4.2.2. It would also be useful to perform similar analysis on the genes thought to be differentially regulated within the TNF signalling pathway (see Section 7.4.2.4.3.1.1), before embarking on further functional work in this area.

7.5.4.1.2 Functional analysis of the role of the TNF signalling pathway

If validation experiments did confirm differential expression of genes within the TNF signalling pathway, the next step would be to explore its relevance to the development of bone marrow failure in AML within a functional model. Given that effective biological inhibitors of TNF signalling already exist, one approach would be to revisit the murine transplantation model of AML-induced bone marrow failure detailed in Section 7.1.2.1, and explore if concurrent therapy with TNF inhibitors such as Etanercept or Infliximab delayed the onset of HSC quiescence, or the development of cytopenias, as the AML engrafts.

7.5.4.1.3 Other pathways of potential interest

The results of enrichment analysis of specific biological pathways using the KEGG pathway tool is obviously highly dependant on the criteria used for selection of input genes. As described in Section 7.4.1.4, in our analysis, our input data comprised all genes which exhibited differential expression with an individual p value of <0.05 and fold change of greater than 2. Of these 1013 genes, 810 were recognised by DAVID and used for further analysis.

However, reanalysis using a more discrete dataset (comprising only those genes which show differential expression with an individual p value of ≤0.01 and fold change of greater than 2; a total of 217 genes), highlights enrichment of just two pathways: the cell cycle pathway again (EASE score, p=0.02; after Benjamini-Hochberg correction, p=0.86), and interestingly, the Notch signalling pathway (EASE score, p=0.05; after Benjamini-Hochberg correction, p=0.97)(see Figure 7-14). The Notch signalling pathway has been shown to be a key player in

HSC development, and also potentially the maintenance of quiescence, and as such is an intriguing target for further investigation[367-370].

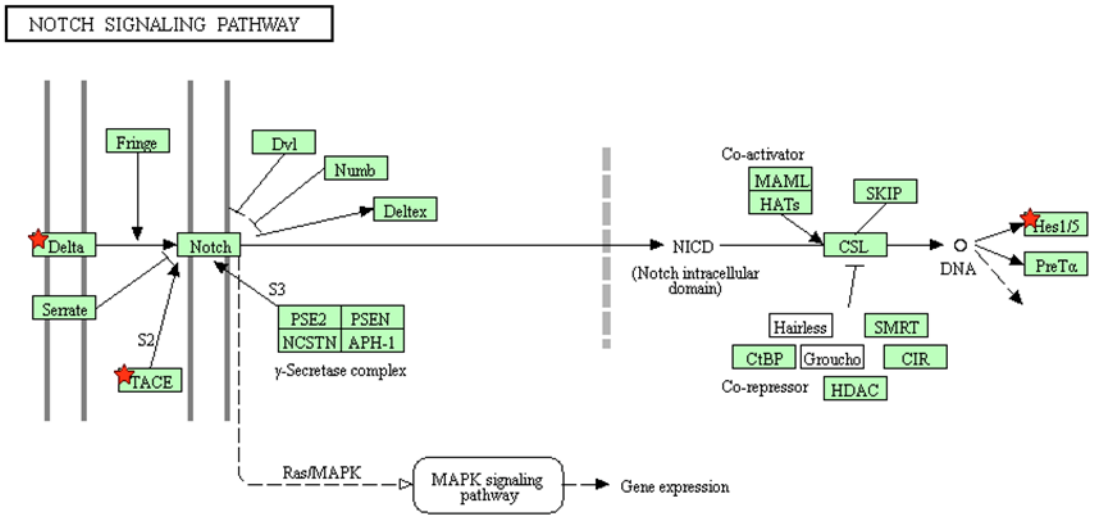

Figure 7-14 KEGG generated figure showing the key genes involved in Notch signalling. Genes differentially expressed between HSCs from AML and control samples (fold change >2, EASE score p value <0.05) are highlighted with a star

7.5.4.2 Are changes in gene expression limited to HSCs from CD34^{low} AML samples, or are they mirrored across AML subtypes?

If money and samples were not limited, I would view it as crucial to repeat this work using another panel of AML samples of an alternative subtype (such as (t(8;21) or inv(16)); sorting the AML and matched control samples, undergoing RNA extraction and then qPCR for the relevant genes of interest to see if such changes in gene expression are found in HSCs across all AML subtypes or simply limited to CD34^{low} samples.

7.5.4.3 Would single cell sorting and analysis of HSC transcriptomes yield more useful data than pooled cells?

One fundamental remaining question is whether this approach of pooling HSCs and studying their transcriptomes en masse can lead to biologically meaningful data. For analysis, we have made the assumption that all the cells within the CD34^{+}38^{-}ALDH^{high}CLL1^{-} subset have homogeneous transcriptomes, but this is highly unlikely to be the case. Cells are likely to be in different states of cycle, may represent different stages of lineage commitment and therefore will most likely have very different transcriptomes. By pooling their transcriptomes together for analysis, some significant changes in gene expression may well be lost.

We remain interested in the possibility of single cell sorting of these populations and the study of transcriptomes on an individual cell level. If, for example, we sorted 100 single HSCs from

AML and control BMs, would the proportion of cells marking out an immature, quiescent phenotype be higher in those HSCs derived from an AML-exposed sample?

We have been fortunate to begin collaboration with the Sanger Institute-EBI Single-Cell Genomics Centre to explore this particular question (see Section 8.2.3.3.1 for further details). However, exploring the effect of AML of HSC behaviour via this approach will have some limitations. The expense of this work will significantly limit the number of samples that can feasibly be analysed. Given the biological variability that we know exists between samples, are any meaningful conclusions likely to be drawn from the comparison of one or two biological replicates? We do not know the biological variability within the HSC subset, and therefore can only guess at the number of cells required to gain adequate representation of the potential diversity within this group. As discussed in Section 7.5.1.2, experts suggest that samples undergoing Single Cell Sorting need to be sorted and the resultant cells frozen within 8 minutes: a time frame which we would be unable to achieve for the majority of samples (both control and AML). Finally, by sorting samples directly into wells, we are unable to confirm the accuracy of the sorting process by a retrospective purity check. Such checks are vital for the validity of any work involving the sorting of a rare population of cells, and in my experience during this thesis, approximately 20% of all samples (both AML and control) fell short of the required standards.

7.5.4.4 Could the presence of pre-leukaemic HSCs alter the transcriptomes of AML-exposed HSCs generated from RNA-Seq?

Recent work exploring the presence of a pre-leukaemic population of HSCs within selected AML samples, which carry early mutations associated with AML development, raises further pertinent queries regarding the work in this chapter. Such cells presumably have altered transcriptome profiles when compared to normal HSCs, as a result of their acquired mutations. If a difference in the transcriptomes between HSCs from AML samples and normal controls had been detected and validated from this RNA-Seq data, it might be difficult to assess if such a difference is related to the reversible effect of the AML on the normal HSCs, or due to the presence of pre-leukaemic cells within the sorted fraction.

However, I would hypothesise this is unlikely to be a major problem for two main reasons. Primarily, the proportion of cells within the HSC population which carry pre-leukaemic mutations is relatively low[107]. Secondarily, the exact pre-leukaemic mutational profile will vary sample to sample, and therefore the effect of any individual mutation should be negated by the use of multiple samples.

7.5.4.4.1 Do the transcriptomes vary between HSCs from AML exposed samples at diagnosis and remission?

One approach which would remove any uncertainty about the effect of pre-leukaemic HSCs on the transcriptomes of AML-exposed stem cells would involve making a comparison between HSCs extracted from diagnosis and remission samples from the same patient. For a patient to have regained remission, their blood counts must have recovered, and therefore HSC quiescence must have been reversed.

Provisional work exploring the viability of such an approach was undertaken. Interestingly the separation of HSCs from AML remission bone marrows proved technically extremely challenging. Whereas we can normally identify 1000 to 2000 normal HSCs per vial from bone marrow samples taken at diagnosis from the tissue bank, in the 2 AML remission samples where separation of HSCs was attempted, only 50 $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ cells were obtained.

This was a non-quantitative experiment, and to date we have not attempted to quantify the concentration of HSCs in remission marrows (via the techniques described in Chapter 5). The scarcity of HSCs in these samples would certainly fit with the often haemodilute appearances of bone marrow aspirates taken in first remission. Nevertheless, despite an apparently very small number of HSCs being present, they are clearly sufficient to have produced normal blood counts. This represents another line of evidence to suggest that the normal HSCs at diagnosis with AML are significantly functionally impaired.

Chapter 8 Discussion

8.1 Summary of main findings

In this thesis, I have attempted to study the interaction between AML and normal haematopoietic stem cells, and how this might result in the haematopoietic failure observed during the development of the disease.

Starting from the group's previous observations that HSC numbers at diagnosis with AML (at least with $CD34^{low}$ AMLs) appear preserved when compared to controls, I developed and validated a method of separating normal HSCs from all primary AML samples, irrespective of karyotype or genetic mutational status.

Using this technique, I then proceeded to assess HSC numbers across a variety of diagnostic primary samples, and found that HSC numbers are relatively preserved in these too compared to controls. Further analysis of the AML samples suggested that there is a subgroup with a particularly low stem cell concentration at diagnosis. Low stem cell number at diagnosis appears to be predictive for a worse overall survival, an observation that may be independent of other prognostic factors such as age or cytogenetics.

During validation work for the panel of antibodies for HSC separation, we also identified different functional behaviour patterns in HSCs dependant on the cytogenetics of the AML they were extracted from. HSCs derived from poor risk cytogenetics samples exhibited different in vitro behaviour, with relatively impaired growth over a 7 week culture period. This finding was not observed in patients with good or intermediate risk AML.

We attempted to examine a cohort of patients who did not present with trilineage bone marrow failure, as these represent a group of patients whose pattern of presentation is difficult to explain with our theory of HSC quiescence. Analysis of a large database of Bart's patients suggested that patients presenting with preservation of counts are rare. We were unable to identify patient factors which might influence presentation platelet count or haemoglobin, such as age, gender or cytogenetics. In all patients examined, mature cells appear to be of normal genetic origin, and not evolved from the leukaemia clone.

As a side-project of potential therapeutic importance, we examined the effect of fluctuating CD33 expression on normal HSCs on exposure to AML, and the potential effect this might have on patients who received GO as therapy. We showed that the normal HSCs appear to have

significantly downregulated CD33 levels at diagnosis with AML, when compared to controls, which rapidly return to normal on remission.

Why these changes occur remains unclear. We had hypothesised that a fluctuating level of CD33 might have therapeutic importance for HSCs in avoiding the toxic effects of targeted chemotherapy with the anti-CD33 antibody GO. However, results from an in vitro model designed to test the cytotoxicity of GO on normal HSCs separated from AML and control samples, suggested that GO did not appear to have a significant cytotoxic effect on either population. Nevertheless, at the time of writing, newer anti-CD33 agents (with different methods of cytotoxicity) are beginning human trials, and it will be interesting to see whether these have a more marked effect on the process of normal haematopoiesis.

We attempted to address the mechanism by which AML functionally inhibits HSC activity. Normal HSCs from human primary diagnostic AML bone marrow samples were sorted, and their transcriptomes compared to HSCs from controls using RNA-Seq. An initial experiment produced a dataset of poor quality, due to the presence of a significance proportion of reads mapping to intronic DNA regions. This was thought to be due to a failure of selection of polyadenylated mRNA for cDNA synthesis during library production. Unsurprisingly, attempts to validate the observed differences in expression levels of six genes between the two HSC subtypes, by qPCR and flow based cytometry, revealed consistent statistically significant changes in CD33 expression only (by flow-based assessment of protein expression).

Repeating the RNA-Seq experiment using column-based RNA extraction of sorted cells before the start of library preparation yielded data of significantly better quality (in terms of the proportion of reads mapping to exonic regions). Six genes of interest appeared to be differentially expressed between the two subgroups after correction was made for multiple testing. Enrichment analysis suggests there is an overepresentation in this dataset of genes involved in the TNF signalling and cell cycle control pathways. Unfortunately, due to time constraints at the end of my PhD, validation of these genes of interest by independent qPCR was not possible.

8.2 Key areas for further experimental effort expanding on the work described within this thesis

Although this work summarised in this thesis answered some of our questions about the interaction between normal HSCs and AMLs, it has also raised many, as yet unanswered queries. Planned experiments relevant to completed work have been described in detail on a

chapter by chapter basis. Below, I have summarised the key areas which I think would warrant further experimental effort in this field.

8.2.1 Does stem cell concentration at diagnosis with AML have an independent prognostic impact?

One of the most interesting findings from this project came from the quantitative analysis of bone marrow samples at AML diagnosis, namely that normal HSC concentration represents an independent prognostic marker for overall survival. However, discussed in Section 5.4.4.2.1, validation of these findings in an independent cohort of samples is required. We are currently in discussions with the Royal Marsden to see if such a collaboration is possible.

Sample analysis for HSC concentration performed as described in Chapter 5.3 is a laborious process. It requires clinical staff to reserve the first aspirate of a diagnostic bone marrow sample for analysis, followed by speedy transport to the laboratory to facilitate analysis within 24 hours. Staining and flow based analysis is time-consuming, requiring three to four hours per sample. As such, even if HSC concentration at diagnosis did carry prognostic significance, one wonders about its utility as an assay in general clinical practise.

One approach currently being considered (D Taussig, personal communication) is the analysis of peripheral blood taken at the time of diagnosis for HSC concentration, rather than bone marrow. Supplies of blood from patients at diagnosis are much more plentiful, and presumably do not suffer from the same issues of haemodilution as bone marrow (assuming they have been taken before transfusion). Normal HSCs are easily identifiable in the peripheral blood of patients with AML[65].

The most obvious flaw of this approach is that there is no evidence at present that the concentration of HSCs in peripheral blood is proportional to that seen in bone marrow. Intuitively, it seems more logical that if normal HSCs are excluded from their niche environment by AML infiltration, they might be more likely to leave the bone marrow and enter the peripheral circulation.

8.2.2 Do poor risk karyotype AMLs exert a different effect on normal HSCs to good or intermediate disease samples, as suggested by their different behaviour in vitro?

One of the most intriguing and unexpected findings of this PhD is that HSCs extracted from different karyotypic subtypes of AML appear to exhibit different behaviour in vitro, with a

failure to grow in 7 week LTC being much more likely in those from poor risk AMLs than others. As discussed at length in Chapter 4, there are a number of factors which might explain this observation, from difficulty in flow-separating HSCs from AML in this setting; to a "normal" HSC population which might be functionally impaired from a heavy pre-leukaemic mutation burden; to poor risk AMLs having a more marked inhibitory effect on normal HSCs by nature of differential cytokine secretion.

This group of samples might prove technically challenging to work with in the laboratory, but they are derived from the very patients (with slow-growing, chemoresistant disease) who might have the most to gain from a therapeutic reversal of inhibition of normal haematopoiesis.

It is clear that a larger cohort of samples need to be investigated to see if these findings are consistently reproduced. It would also be interesting to see if the failure of normal HSC growth in 7 week LTC is recapitulated in the gold-standard test of stem cell function, the mouse xenograft model.

It is hoped that these experimental approaches might be included in a current ongoing multicentre funding application, targeted at improving our understanding of the biology of poor risk karyotype AML samples.

8.2.3 What is the mechanism of BM suppression in AML?

From a clinical perspective, the most important aspect of this project is the attempts to understand how AML induces quiescence in normal HSCs. If this process was understood, and a signalling pathway amenable to therapeutic modification identified, it might provide us with the means of rebooting normal haematopoiesis. However, this process has proven frustratingly elusive.

8.2.3.1 Technical considerations with HSC work

Working with HSCs extracted from primary samples poses significant laboratory challenges, which I have come to fully appreciate over the last three years! The most significant of these is the limited number of cells contained within each biological sample, which precludes the use of the majority of laboratory techniques used in signalling work, including western blotting, ELISAs or even mass spectrometry.

8.2.3.2 Considerations with the use of Primary AML diagnostic BM samples

This project has almost entirely focussed on the use of primary human samples of diagnostic AML BM. The work would not have been possible without the excellent storage facilities provided by the Bart's Tissue Bank.

Whilst primary samples represent the only really valid material available for studying the interaction of normal haematopoietic cells with AML, their use does raise some interesting challenges. One key issue is marked inter-sample heterogeneity: not just between AML blasts of different karyotypes, but also within the normal HSC population. This is also the case within the control samples, where the HSC populations vary widely in both concentration and transcriptome analysis. Whilst recognising such biological variability is crucial, for experimental work to reach significant findings, this requires the analysis of a large number of samples, which is both time consuming and costly.

Secondarily, we recognise there is not only heterogeneity between the HSCs of different samples, but also most probably, within the $CD34^{+}CD38^{-}$ population of an individual sample.

8.2.3.3 Transcriptome Analysis

We have utilised transcriptome analysis as a surrogate marker of signalling pathway activation, as current techniques are sufficiently sensitive to allow study of tiny cell populations.

Whole transcriptome analysis such as that provided by RNA-Seq is a useful approach for producing a selection of candidate differentially expressed genes within the HSCs from AML samples. The second RNA-Seq experiment comparing the transcriptomes of normal HSCs sorted from AML and HSC samples has revealed 6 genes of interest are significantly differentially expressed between the groups. KEGG pathway analysis suggested gene enrichment within the cell cycle and TNF signalling pathways.

Validation of these results is required as a first step, but it would be of interest to explore functionally whether these genes have any effects on HSC quiescence or differentiation. As already mentioned in the relevant chapter, it would also be useful to explore whether the same genes are differentially expressed (using targeted qPCR) between HSCs extracted from AML samples of a different subtype (such as t(8;21) or inv(16)).

8.2.3.3.1 Single Cell Transcriptome Analysis

The functional, and presumably transcriptional, heterogeneity of cells within the $CD34^{+}38^{-}$ compartment of bone marrow of both AML and normal BM samples may limit the detection of

genes whose expression are significantly modulated in a subset of cells. However, modern approaches of single cell sorting and subsequent massively parallel single cell RNA-Seq provide a means of addressing this issue[24].

Our group has tentatively begun exploring such approaches. The $CD34^{+}38^{-}ALDH^{high}CLL1^{low}$ cells from one poor risk AML sample (ID 7820) were sorted into two 96 well plates at the BCI Flow laboratory using the BD FACSAria™ Fusion sorter. These cells underwent successful cDNA synthesis and amplification in Cambridge, in collaboration with the Sanger Institute-EBI Single-Cell Genomics Centre. Amplified cDNA suitable for subsequent library preparation was produced in 60% of wells. Library preparation prior to sequencing is still ongoing, but it is hoped that comparative analysis of the transcriptomes of HSCs from AML and control cells might identify different cell populations present within the broader $CD34^{+}38^{-}ALDH^{high}CLL1^{-}$ gate.

Exciting as this technique is, even this first preliminary experiment raised some technical challenges. In order to maintain cell integrity prior to processing, we were advised to sort cells only once from the stained sample. As a result, there was no opportunity to check the performance of the sorter, by checking the purity of the sorted population of cells before processing. This is key when separating out HSCs from leukemic samples, due to their relative rarity when compared to blasts. Secondarily, we were also advised to limit the time of sorting to approximately 8 minutes per 96 well plate, to maintain RNA integrity. This significantly limits the number of samples which could be processed in this fashion; as such a rate of sorting (from an AML sample) is only achievable in one with a high normal HSC concentration.

8.2.3.4 Targeted Cytokine Analysis

The initial findings of Broxmeyer over 30 years ago suggested that AML was capable of causing growth inhibition of HSCs via a diffusible factor[133]. Although hypoxia[60], metabolite availability or even microRNA secretion[148] might be responsible for this effect, an obvious alterative is that cytokines (secreted either directly from AML cells or from altered stromal cells) might be modulating stem cell behaviour. A number of cytokines such as thrombopoietin[58] or angiopoietin[57] have already been shown to have a role in cell cycle regulation and differentiation of HSCs.

To my knowledge, systematic assessment of the cytokine milieu within samples of AML bone marrow in comparison to controls has not been performed, although I note a similar assessment has been made on peripheral blood samples[371].

As there exists a significant number (15 to 30) cytokines of interest to study, and the available volume of bone marrow would be too small to facilitate more than 2 or 3 ELISA assays per sample, I would propose using a technique such as the BD Multiplexed Bead-Based Immunoassay, which enables parallel identification of specifically designed panel of proteins, by bead-tagged antibodies.

For the last two years, I have been collecting the supernatant of bone marrows from newly diagnosed AML and control patients, which have been stored at -196^{o}C (see Section 2.2.4). These would provide an excellent repository of samples to test using the kit described above.

8.2.3.5 Mass Cytometry

Towards the end of this PhD project, Bart's obtained a CyTof II analyser (Fludigm), which gave the possibility of mass cytometry experiments. Antigens (both intra and extracellular) on cells are labelled with antibodies tagged to rare metals. Cells are then nebulised, and the resultant particles ionised by the action of an argon laser, which are subsequently analysed by a time-of-flight mass spectrometer[372]. Although similar in some ways to flow cytometric analysis of cells, the use of rare metals to tag antibodies rather than fluorochromes means that a comparatively huge number of targets (upwards of 30) can be studied simultaneously on the same sample[373]. The constraints of spectral overlap mean vastly fewer targets can be studied on HSCs via standard flow cytometry[374].

This approach has clear applications to this field: if appropriate antibodies were available, given the rarity of the HSC population, CyTof analysis might provide an efficient mechanism of simultaneously validating a large number of protein based targets.

However, there are some limitations. Currently, the range of available labelled antibodies is relatively small (although rapidly expanding). Antibodies are expensive, and the process of sample analysis time consuming. Whilst flow cytometry is well established enough for there to have developed a consensus about where an HSC population might lie, these parameters have yet to be established for analysis with mass cytometry based analysis. Finally, and perhaps most significantly, there is no possibility of follow-on functional work after cell analysis, as by definition during processing, the cells are destroyed.

8.2.3.6 Attempts to develop an in vitro model of the interaction of normal HSCs and AML cells

We have been hampered in our search for the factor(s) responsible for AML-induced HSC quiescence by the lack of a functional in vitro model of the interaction of the two cell types.

As mentioned in Section 7.1.1.1, both our laboratory and Prof Bonnet's at the LRI have developed an in vitro model which involved the culture of AML cells and HSCs together in the presence of a feeder stromal line. After ten days in co-culture, the effect of AML on the HSC population was assessed by proxy of their incorporation of BRDU over the test period.

Although this assay had initial promise, the replicability of results proved poor. During this PhD, I attempted this assay a number of times, modulating conditions such as the source of HSCs (cord and GMPB); the flow based selection criteria of HSCs ($CD34^+$ versus $CD34^+38^-$); the type of feeder cell lines (MS5 versus HS5); normoxia versus hypoxia; degree of contact between cell types (transwell versus no separation), and method of assessing proliferation of HSCs (BRDU incorporation versus Ki67). Unfortunately none of these modifications improved the assay's reliability. For completeness, a summary of the basic method is included in Appendix 2.

In the last few months, I have made some tentative attempts to develop another in vitro model of bone marrow suppression: this time growing sorted HSCs in 7 week LTC, but exposing the cells on a weekly basis to AML-conditioned media to see if this impacted their colony-forming ability. At the time of writing, this had not reproducibly recapitulated an inhibitory effect on normal HSCs. However, there are many possible steps to hone the original protocol (which is given in Appendix 2), and given time, I would consider this worth further exploration.

8.3 New avenues for further research in the broader field of malignancy-associated bone marrow failure

During the writing of this thesis, a number of new avenues for further research into the effect of AML on the normal haematopoietic process presented themselves. I have summarised below the directions I would be most interested in taking in the future.

8.3.1 How does the mechanism of AML–associated bone marrow failure compare with that seen in other malignancies?

This project focussed very specifically on the interaction of AML with normal bone marrow, or more specifically, the HSC population within it. However, failure of haematopoiesis is not simply a phenomenon observed in AML: although clinical experience suggests it is synonymous with AML development, it also occurs with other haematological diseases such as ALL, the chronic leukaemias, myeloma and lymphomas (with significant bone marrow involvement), as well as several forms of metastatic carcinoma.

Historical interest in the BM failure seen in AML dates back over 30 years to the initial observations of Broxmeyer, who noticed an effect on normal haematopoietic colony growth by an apparently diffusible factor secreted by AML cells.

The cause of haematopoietic failure in these other conditions has to date not been perfectly mapped out. In multiple myeloma, roles for hepcidin, BMP2[375], Fas-L[376], TRAIL[377] and TGFβ signalling[298] in inducing cytopenia have been described. In a Science paper, showing the ability of pre-B ALL to alter the stem cell niche in a manner that affected homing behaviour of normal HSCs, Colmone et al also described the observation of similar changes in primary human AML bone marrow samples[378].

It is interesting to hypothesise as to whether the mechanism of bone marrow failure seen in these related conditions is universal, or occurs by disease-specific pathways. Is the concentration of normal HSCs preserved in the same fashion as with AML?

In some ways, investigating the effect of AML on bone marrow is potentially more complex than other malignancies, because of the phenotypic similarities between HSCs and myeloid blasts. However, if a clear (and reversible) mechanism could be identified for the BM failure seen in AML, this might have therapeutic possibilities that extend beyond this patient group.

8.3.2 Does AML just affect HSCs, or does it modulate other levels of normal haematopoiesis?

This project has focused very specifically on the effect of AML on HSCs, defined by flow criteria as either $CD34^{+}CD38^{-}$ or $CD34^{+}CD38^{-}ALDH^{high}CLL1^{-}$ cells. In Section 5.5.2.3, I have discussed that this separation is a simplification, with this group containing cells at various stages of lineage commitment and proliferative potential.

However, it cannot be assumed that AML directly affects normal HSCs only. Might blasts also directly influence the behaviour of committed progenitors, or even mature cells belonging to the granulocytic, erythroid and megakaryocytic lineages? Previous work suggested that the concentration of progenitor cells (defined by $CD34^{+}CD38^{+}$) fell significantly at diagnosis with AML, an observation attributed to a reduction in HSC maturation[1]. Investigating whether AML directly affects these populations is not straightforward: their concentrations in AML BMs are extremely small, and reliable separation from AML blasts might prove challenging (for example, ALDH expression patterns would not help, as normal progenitors are often ALDH intermediate or low expressing cells).

8.3.3 Is the effect of AML on normal HSCs the same irrespective of AML subtype?

AML is extremely heterogeneous disease, characterised by a range of phenotypic appearances and caused by a variety of distinct, recurring cytogenetic and genetic abnormalities. Nevertheless all of these different subtypes of disease appear to have a similar propensity to cause haematopoietic failure (with the exception of APML, which is associated with more marked thrombocytopenia at diagnosis, as discussed in Section 3.4.1.3.4). Similarly, we have shown that HSC concentrations at diagnosis are preserved across disease subtypes. Therefore, can we assume that the mechanism which causes HSC quiescence across all AML subtypes is the same?

At the time of writing, although we lack evidence, I would hypothesise that the basic mechanism responsible for HSC quiescence is preserved across AML subtypes. Despite the suggestion of transcriptome heterogeneity between HSCs extracted from different subgroups (see Chapter 7), such a universal effect seems unlikely to be caused by an acquired mutation or new signalling pathway, as there is no one genetic mutation (at least not to date) found across all AML subtypes[102,196]. I would favour the idea that the normal HSC quiescence seen is a subversion of the normal pathway of homeostasis: either as a result of negative feedback to HSCs to which they are sensitive, out-competition for niche space; relative hypoxia, or possibly the effect of a pro-inflammatory state in the bone marrow microenvironment.

8.3.4 Are the effects of AML limited to normal haematopoiesis precursors, or does it have the ability to modulate the micro environmental stromal environment?

In this project, we have solely investigated the effect of AML on normal haematopoietic stem cells, and have not investigated its effects on the stromal microenvironment. Nevertheless, this interaction might be key: appropriate stromal support is required for HSCs to maintain a state of protected quiescence in the bone marrow, and if AML is able to modulate these pathways, these indirect effects may well prove as important as those exerted directly on stem cells. Key experiments in this area are found summarised in Section 1.5.3.

8.3.5 What does "normal" mean in the context of HSCs anymore anyway?

One of the key developments in the field of AML biology since I started this PhD is the discovery via high-throughput sequencing of the existence of pre-leukaemic stem cells. These cells appear functionally normal, but contain early mutational events that are present within the associated leukaemic clone[107,151]. This has enabled the study of the evolutionary pathway of successive mutations required for leukaemia development, but also raises a key question of just how normal are "normal" HSCs in this context? It also has practical relevance for the work

summarised within this thesis, where the presence of a molecular mutation at low level within a sorted population of cells has been taken as a surrogate marker for leukaemia contamination of normal cells. This might prove to be an oversimplification if the mutation in question is also found as part of the pre-leukaemic clone.

Nevertheless, I would argue strongly that the discovery of cells with early mutational events by no means invalidates the importance of studying normal haematopoiesis in this context. The papers published to date in this area reveal only a small proportion of normal HSCs within the bone marrow to contain mutations. For example, Shlush et al studied the HSC fraction of 12 *DNMT3A* mutated AML samples for the presence of the AML-associated mutation. The observed mutated allele frequency varied between 0 and 69%, with 75% of samples having an observed allele frequency of less than 30%[107]. A significant proportion of patients go on to achieve a life-long remission after chemotherapy, suggesting either that progression of pre-leukaemic mutations is not an inevitable process, or that they are not found in all patients.

8.4 Final comments

This PhD project attracted me when I was looking for research projects for a number of reasons. I was interested in both AML and stem cell biology. I liked the way this project approached AML from what seemed like an entirely new perspective: looking at the effect on normal haematopoietic cells, rather than solely focussing on the biology and behaviour of blasts.

Research progress over the last 30 years has resulted in us understanding vastly more about the genetic structure of AML and its disease evolution. We can use this knowledge to offer patients much more sophisticated prognostics at diagnosis. However, all clinicians are in agreement that clinical treatment has not advanced at the same rate, and that new treatments (both in terms of general approach and specific targeted therapy) are desperately required to improve outcomes.

The idea that in reversing the effect on normal bone marrow might somehow push the evolutionary pathway back a little in favour of normality is very appealing. Reversal (albeit probably temporary) of the cytopenias observed in AML has enormous palliative potential. It is hoped in the future this might be achieved.

Chapter 9 References

1 Miraki-Moud, F. *et al.* Acute myeloid leukemia does not deplete normal hematopoietic stem cells but induces cytopenias by impeding their differentiation. *Proc Natl Acad Sci U S A* **110**, 13576-13581, doi:10.1073/pnas.1301891110 (2013).

2 Coombs, C. C., Tavakkoli, M. & Tallman, M. S. Acute promyelocytic leukemia: where did we start, where are we now, and the future. *Blood cancer journal* **5**, e304, doi:10.1038/bcj.2015.25 (2015).

3 Burnett, A. K. *et al.* Identification of patients with acute myeloblastic leukemia who benefit from the addition of gemtuzumab ozogamicin: results of the MRC AML15 trial. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **29**, 369-377, doi:10.1200/JCO.2010.31.4310 (2011).

4 Burnett, A. K. *et al.* Addition of gemtuzumab ozogamicin to induction chemotherapy improves survival in older patients with acute myeloid leukemia. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **30**, 3924-3931, doi:10.1200/JCO.2012.42.2964 (2012).

5 Rowe, J. M. & Tallman, M. S. How I treat acute myeloid leukemia. *Blood* **116**, 3147-3156, doi:10.1182/blood-2010-05-260117 (2010).

6 Bain, B. *Blood Cells: A Practical Guide*. 4th edn, (Wiley-Blackwell, 2008).

7 Yoder, M. C., Hiatt, K. & Mukherjee, P. In vivo repopulating hematopoietic stem cells are present in the murine yolk sac at day 9.0 postcoitus. *Proc Natl Acad Sci U S A* **94**, 6776-6780 (1997).

8 Tavian, M. & Peault, B. The changing cellular environments of hematopoiesis in human development in utero. *Exp Hematol* **33**, 1062-1069, doi:10.1016/j.exphem.2005.06.025 (2005).

9 Fernandez, K. S. & de Alarcon, P. A. Development of the hematopoietic system and disorders of hematopoiesis that present during infancy and early childhood. *Pediatric clinics of North America* **60**, 1273-1289, doi:10.1016/j.pcl.2013.08.002 (2013).

10 O'Malley, D. P. Benign extramedullary myeloid proliferations. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* **20**, 405-415, doi:10.1038/modpathol.3800768 (2007).

11 Chu, K. A. *et al.* Intrathoracic extramedullary haematopoiesis complicated by massive haemothorax in alpha-thalassaemia. *Thorax* **54**, 466-468 (1999).

12 Malik, M. *et al.* Paraplegia due to extramedullary hematopoiesis in thalassemia treated successfully with radiation therapy. *Haematologica* **92**, e28-30 (2007).

13 Szilvassy, S. J., Humphries, R. K., Lansdorp, P. M., Eaves, A. C. & Eaves, C. J. Quantitative assay for totipotent reconstituting hematopoietic stem cells by a competitive repopulation strategy. *Proc Natl Acad Sci U S A* **87**, 8736-8740 (1990).

14 Boggs, D. R., Boggs, S. S., Saxe, D. F., Gress, L. A. & Canfield, D. R. Hematopoietic stem cells with high proliferative potential. Assay of their concentration in marrow by the frequency and duration of cure of W/Wv mice. *The Journal of clinical investigation* **70**, 242-253 (1982).

15 Wilson, A. & Trumpp, A. Bone-marrow haematopoietic-stem-cell niches. *Nat Rev Immunol* **6**, 93-106, doi:10.1038/nri1779 (2006).

16 Morrison, S. J. & Kimble, J. Asymmetric and symmetric stem-cell divisions in development and cancer. *Nature* **441**, 1068-1074, doi:10.1038/nature04956 (2006).

17 Zavidij, O. *et al.* Hematopoietic activity of human short-term repopulating cells in mobilized peripheral blood cell transplants is restricted to the first 5 months after transplantation. *Blood* **115**, 5023-5025, doi:10.1182/blood-2010-02-271528 (2010).

18 Akashi, K., Traver, D., Miyamoto, T. & Weissman, I. L. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**, 193-197, doi:10.1038/35004599 (2000).

19 Kondo, M., Weissman, I. L. & Akashi, K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91**, 661-672 (1997).

20 Seita, J. & Weissman, I. L. Hematopoietic stem cell: self-renewal versus differentiation. *Wiley interdisciplinary reviews. Systems biology and medicine* **2**, 640-653, doi:10.1002/wsbm.86 (2010).

21 Doulatov, S. *et al.* Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nature immunology* **11**, 585-593, doi:10.1038/ni.1889 (2010).

22 Morrison, S. J. & Weissman, I. L. The long-term repopulating subset of hematopoietic stem cells is deterministic and isolatable by phenotype. *Immunity* **1**, 661-673 (1994).

23 Robb, L. Cytokine receptors and hematopoietic differentiation. *Oncogene* **26**, 6715-6723, doi:10.1038/sj.onc.1210756 (2007).

24 Paul, F. *et al.* Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. *Cell* **163**, 1663-1677, doi:10.1016/j.cell.2015.11.013 (2015).

25 Velten, L. *et al.* Human haematopoietic stem cell lineage commitment is a continuous process. *Nature cell biology* **19**, 271-281, doi:10.1038/ncb3493 (2017).

26 Harker, L. A. *et al.* Effects of megakaryocyte growth and development factor on platelet production, platelet life span, and platelet function in healthy human volunteers. *Blood* **95**, 2514-2522 (2000).

27 Tak, T., Tesselaar, K., Pillay, J., Borghans, J. A. & Koenderman, L. What's your age again? Determination of human neutrophil half-lives revisited. *Journal of leukocyte biology* **94**, 595-601, doi:10.1189/jlb.1112571 (2013).

28 Franco, R. S. The measurement and importance of red cell survival. *American journal of hematology* **84**, 109-114, doi:10.1002/ajh.21298 (2009).

29 Sprent, J. & Tough, D. F. Lymphocyte life-span and memory. *Science* **265**, 1395-1400 (1994).

30 Wu, A. M., Till, J. E., Siminovitch, L. & McCulloch, E. A. A cytological study of the capacity for differentiation of normal hemopoietic colony-forming cells. *Journal of cellular physiology* **69**, 177-184, doi:10.1002/jcp.1040690208 (1967).

31 Harrison, D. E., Zhong, R. K., Jordan, C. T., Lemischka, I. R. & Astle, C. M. Relative to adult marrow, fetal liver repopulates nearly five times more effectively long-term than short-term. *Exp Hematol* **25**, 293-297 (1997).

32 Bowie, M. B. *et al.* Hematopoietic stem cells proliferate until after birth and show a reversible phase-specific engraftment defect. *The Journal of clinical investigation* **116**, 2808-2816, doi:10.1172/JCI28310 (2006).

33 Schofield, R. The relationship between the spleen colony-forming cell and the haemopoietic stem cell. *Blood cells* **4**, 7-25 (1978).

34 Orkin, S. H. & Zon, L. I. Hematopoiesis: an evolving paradigm for stem cell biology. *Cell* **132**, 631-644, doi:10.1016/j.cell.2008.01.025 (2008).

35 Lord, B. I., Testa, N. G. & Hendry, J. H. The relative spatial distributions of CFUs and CFUc in the normal mouse femur. *Blood* **46**, 65-72 (1975).

36 Kiel, M. J. *et al.* SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**, 1109-1121, doi:10.1016/j.cell.2005.05.026 (2005).

37 Zhang, J. *et al.* Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* **425**, 836-841, doi:10.1038/nature02041 (2003).

38 Taichman, R. S. & Emerson, S. G. The role of osteoblasts in the hematopoietic microenvironment. *Stem cells* **16**, 7-15, doi:10.1002/stem.160007 (1998).

39 Calvi, L. M. *et al.* Osteoblastic cells regulate the haematopoietic stem cell niche. *Nature* **425**, 841-846, doi:10.1038/nature02040 (2003).

40 Kopp, H. G., Avecilla, S. T., Hooper, A. T. & Rafii, S. The bone marrow vascular niche: home of HSC differentiation and mobilization. *Physiology* **20**, 349-356, doi:10.1152/physiol.00025.2005 (2005).

41 Avecilla, S. T. *et al.* Chemokine-mediated interaction of hematopoietic progenitors with the bone marrow vascular niche is required for thrombopoiesis. *Nat Med* **10**, 64-71, doi:10.1038/nm973 (2004).

42 Kiel, M. J. & Morrison, S. J. Maintaining hematopoietic stem cells in the vascular niche. *Immunity* **25**, 862-864, doi:10.1016/j.immuni.2006.11.005 (2006).

43 Schweitzer, K. M. *et al.* Constitutive expression of E-selectin and vascular cell adhesion molecule-1 on endothelial cells of hematopoietic tissues. *The American journal of pathology* **148**, 165-175 (1996).

44 Wilson, A. *et al.* Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* **135**, 1118-1129, doi:10.1016/j.cell.2008.10.048 (2008).

45 Eliasson, P. & Jonsson, J. I. The hematopoietic stem cell niche: low in oxygen but a nice place to be. *Journal of cellular physiology* **222**, 17-22, doi:10.1002/jcp.21908 (2010).

46 Blanpain, C., Mohrin, M., Sotiropoulou, P. A. & Passegue, E. DNA-damage response in tissue-specific and cancer stem cells. *Cell Stem Cell* **8**, 16-29, doi:10.1016/j.stem.2010.12.012 (2011).

47 Branzei, D. & Foiani, M. Regulation of DNA repair throughout the cell cycle. *Nature reviews. Molecular cell biology* **9**, 297-308, doi:10.1038/nrm2351 (2008).

48 Harrison, D. E., Astle, C. M. & Delaittre, J. A. Loss of proliferative capacity in immunohemopoietic stem cells caused by serial transplantation rather than aging. *The Journal of experimental medicine* **147**, 1526-1531 (1978).

49 Orford, K. W. & Scadden, D. T. Deconstructing stem cell self-renewal: genetic insights into cell-cycle regulation. *Nature reviews. Genetics* **9**, 115-128, doi:10.1038/nrg2269 (2008).

50 Pietras, E. M., Warr, M. R. & Passegue, E. Cell cycle regulation in hematopoietic stem cells. *The Journal of cell biology* **195**, 709-720, doi:10.1083/jcb.201102131 (2011).

51 el-Deiry, W. S. *et al.* WAF1, a potential mediator of p53 tumor suppression. *Cell* **75**, 817-825 (1993).

52 Hock, H. *et al.* Gfi-1 restricts proliferation and preserves functional integrity of haematopoietic stem cells. *Nature* **431**, 1002-1007, doi:10.1038/nature02994 (2004).

53 Yilmaz, O. H. *et al.* Pten dependence distinguishes haematopoietic stem cells from leukaemia-initiating cells. *Nature* **441**, 475-482, doi:10.1038/nature04703 (2006).

54 Miyamoto, K. *et al.* Foxo3a is essential for maintenance of the hematopoietic stem cell pool. *Cell Stem Cell* **1**, 101-112, doi:10.1016/j.stem.2007.02.001 (2007).

55 Lacorazza, H. D. *et al.* The transcription factor MEF/ELF4 regulates the quiescence of primitive hematopoietic cells. *Cancer cell* **9**, 175-187, doi:10.1016/j.ccr.2006.02.017 (2006).

56 Wilson, A. *et al.* c-Myc controls the balance between hematopoietic stem cell self-renewal and differentiation. *Genes Dev* **18**, 2747-2763, doi:10.1101/gad.313104 (2004).

57 Arai, F. *et al.* Tie2/angiopoietin-1 signaling regulates hematopoietic stem cell quiescence in the bone marrow niche. *Cell* **118**, 149-161, doi:10.1016/j.cell.2004.07.004 (2004).

58 Kimura, S., Roberts, A. W., Metcalf, D. & Alexander, W. S. Hematopoietic stem cell deficiencies in mice lacking c-Mpl, the receptor for thrombopoietin. *Proc Natl Acad Sci U S A* **95**, 1195-1200 (1998).

59 Solar, G. P. *et al.* Role of c-mpl in early hematopoiesis. *Blood* **92**, 4-10 (1998).

60 Simsek, T. *et al.* The distinct metabolic profile of hematopoietic stem cells reflects their location in a hypoxic niche. *Cell Stem Cell* **7**, 380-390, doi:10.1016/j.stem.2010.07.011 (2010).

61 Takubo, K. *et al.* Regulation of the HIF-1alpha level is essential for hematopoietic stem cells. *Cell Stem Cell* **7**, 391-402, doi:10.1016/j.stem.2010.06.020 (2010).

62 Ito, K. *et al.* A PML-PPAR-delta pathway for fatty acid oxidation regulates hematopoietic stem cell maintenance. *Nat Med* **18**, 1350-1358, doi:10.1038/nm.2882 (2012).

63 Lucas, D., Battista, M., Shi, P. A., Isola, L. & Frenette, P. S. Mobilized hematopoietic stem cell yield depends on species-specific circadian timing. *Cell Stem Cell* **3**, 364-366, doi:10.1016/j.stem.2008.09.004 (2008).

64 Heidt, T. *et al.* Chronic variable stress activates hematopoietic stem cells. *Nat Med* **20**, 754-758, doi:10.1038/nm.3589 (2014).

65 Ailles, L. E., Gerhard, B. & Hogge, D. E. Detection and characterization of primitive malignant and normal progenitors in patients with acute myelogenous leukemia using long-term coculture with supportive feeder layers and cytokines. *Blood* **90**, 2555-2564 (1997).

66 Kent, D. *et al.* Regulation of hematopoietic stem cells by the steel factor/KIT signaling pathway. *Clinical cancer research : an official journal of the American Association for Cancer Research* **14**, 1926-1930, doi:10.1158/1078-0432.CCR-07-5134 (2008).

67 Broudy, V. C. Stem cell factor and hematopoiesis. *Blood* **90**, 1345-1364 (1997).

68 Calvi, L. M. & Link, D. C. The hematopoietic stem cell niche in homeostasis and disease. *Blood* **126**, 2443-2451, doi:10.1182/blood-2015-07-533588 (2015).

69 Christopher, M. J., Liu, F., Hilton, M. J., Long, F. & Link, D. C. Suppression of CXCL12 production by bone marrow osteoblasts is a common and critical pathway for cytokine-induced mobilization. *Blood* **114**, 1331-1339, doi:10.1182/blood-2008-10-184754 (2009).

70 Calandra, G. *et al.* AMD3100 plus G-CSF can successfully mobilize CD34+ cells from non-Hodgkin's lymphoma, Hodgkin's disease and multiple myeloma patients previously failing mobilization with chemotherapy and/or cytokine treatment: compassionate use data. *Bone Marrow Transplant* **41**, 331-338, doi:10.1038/sj.bmt.1705908 (2008).

71 Notta, F. *et al.* Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment. *Science* **333**, 218-221, doi:10.1126/science.1201219 (2011).

72 Majeti, R., Park, C. Y. & Weissman, I. L. Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood. *Cell Stem Cell* **1**, 635-645, doi:10.1016/j.stem.2007.10.001 (2007).

73 Challen, G. A., Boles, N., Lin, K. K. & Goodell, M. A. Mouse hematopoietic stem cell identification and analysis. *Cytometry. Part A : the journal of the International Society for Analytical Cytology* **75**, 14-24, doi:10.1002/cyto.a.20674 (2009).

74 Hogan, C. J., Shpall, E. J. & Keller, G. Differential long-term and multilineage engraftment potential from subfractions of human CD34+ cord blood cells transplanted into NOD/SCID mice. *Proc Natl Acad Sci U S A* **99**, 413-418, doi:10.1073/pnas.012336799 (2002).

75 Anjos-Afonso, F. *et al.* CD34(-) cells at the apex of the human hematopoietic stem cell hierarchy have distinctive cellular and molecular signatures. *Cell Stem Cell* **13**, 161-174, doi:10.1016/j.stem.2013.05.025 (2013).

76 Coulombel, L. Identification of hematopoietic stem/progenitor cells: strength and drawbacks of functional assays. *Oncogene* **23**, 7210-7222, doi:10.1038/sj.onc.1207941 (2004).

77 Broxmeyer, H. E. Colony assays of hematopoietic progenitor cells and correlations to clinical situations. *Critical reviews in oncology/hematology* **1**, 227-257 (1984).

78 Eaves, C. J. & Eaves, A. C. Erythropoietin (Ep) dose-response curves for three classes of erythroid progenitors in normal human marrow and in patients with polycythemia vera. *Blood* **52**, 1196-1210 (1978).

79 Dexter, T. M., Wright, E. G., Krizsa, F. & Lajtha, L. G. Regulation of haemopoietic stem cell proliferation in long term bone marrow cultures. *Biomedicine / [publiee pour l'A.A.I.C.I.G.]* **27**, 344-349 (1977).

80 Sutherland, H. J., Eaves, C. J., Lansdorp, P. M., Thacker, J. D. & Hogge, D. E. Differential regulation of primitive human hematopoietic cells in long-term cultures maintained on genetically engineered murine stromal cells. *Blood* **78**, 666-672 (1991).

81 Sutherland, H. J., Eaves, C. J., Eaves, A. C., Dragowska, W. & Lansdorp, P. M. Characterization and partial purification of human marrow cells capable of initiating long-term hematopoiesis in vitro. *Blood* **74**, 1563-1570 (1989).

82 McCulloch, E. A. & Till, J. E. Proliferation of Hemopoietic Colony-Forming Cells Transplanted into Irradiated Mice. *Radiation research* **22**, 383-397 (1964).

83 Hodgson, G. S. & Bradley, T. R. Properties of haematopoietic stem cells surviving 5-fluorouracil treatment: evidence for a pre-CFU-S cell? *Nature* **281**, 381-382 (1979).

84 Kamel-Reid, S. & Dick, J. E. Engraftment of immune-deficient mice with human hematopoietic stem cells. *Science* **242**, 1706-1709 (1988).

85 Kollet, O. *et al.* beta2 microglobulin-deficient (B2m(null)) NOD/SCID mice are excellent recipients for studying human stem cell function. *Blood* **95**, 3102-3105 (2000).

86 Vardiman, J. W., Harris, N. L. & Brunning, R. D. The World Health Organization (WHO) classification of the myeloid neoplasms. *Blood* **100**, 2292-2302, doi:10.1182/blood-2002-04-1199 (2002).

87 ONS. *Cancer Registration Statistics* <http://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases/bulletins/cancerregistrationstatisticsengland/> (2016).

88 Bhayat, F., Das-Gupta, E., Smith, C., McKeever, T. & Hubbard, R. The incidence of and mortality from leukaemias in the UK: a general population-based study. *BMC cancer* **9**, 252, doi:10.1186/1471-2407-9-252 (2009).

89 Juliusson, G. *et al.* Acute myeloid leukemia in the real world: why population-based registries are needed. *Blood* **119**, 3890-3899, doi:10.1182/blood-2011-12-379008 (2012).

90 Le Beau, M. M. *et al.* Clinical and cytogenetic correlations in 63 patients with therapy-related myelodysplastic syndromes and acute nonlymphocytic leukemia: further evidence for characteristic abnormalities of chromosomes no. 5 and 7. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **4**, 325-345 (1986).

91 Bizzozero, O. J., Jr., Johnson, K. G. & Ciocco, A. Radiation-related leukemia in Hiroshima and Nagasaki, 1946-1964. I. Distribution, incidence and appearance time. *N Engl J Med* **274**, 1095-1101, doi:10.1056/NEJM196605192742001 (1966).

92 Owen, C., Barnett, M. & Fitzgibbon, J. Familial myelodysplasia and acute myeloid leukaemia--a review. *Br J Haematol* **140**, 123-132, doi:10.1111/j.1365-2141.2007.06909.x (2008).

93 Evans, D. I. & Steward, J. K. Down's syndrome and leukaemia. *Lancet* **2**, 1322 (1972).

94 Swerdlow, S. H., Campo, E. & Harris, N. L. *WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues*. (IARC, 2008).

95 Craig, F. E. & Foon, K. A. Flow cytometric immunophenotyping for hematologic neoplasms. *Blood* **111**, 3941-3967, doi:10.1182/blood-2007-11-120535 (2008).

96 Grimwade, D. & Freeman, S. D. Defining minimal residual disease in acute myeloid leukemia: which platforms are ready for "prime time"? *Blood* **124**, 3345-3355, doi:10.1182/blood-2014-05-577593 (2014).

97 Dohner, H. *et al.* Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood* **115**, 453-474, doi:10.1182/blood-2009-07-235358 (2010).

98 Sanz, M. A. *et al.* Management of acute promyelocytic leukemia: recommendations from an expert panel on behalf of the European LeukemiaNet. *Blood* **113**, 1875-1891, doi:10.1182/blood-2008-04-150250 (2009).

99 Mrozek, K., Heerema, N. A. & Bloomfield, C. D. Cytogenetics in acute leukemia. *Blood Rev* **18**, 115-136, doi:10.1016/S0268-960X(03)00040-7 (2004).

100 Mrozek, K. Cytogenetic, molecular genetic, and clinical characteristics of acute myeloid leukemia with a complex karyotype. *Semin Oncol* **35**, 365-377, doi:10.1053/j.seminoncol.2008.04.007 (2008).

101 Grimwade, D. *et al.* Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood* **116**, 354-365, doi:10.1182/blood-2009-11-254441 (2010).

102 Cancer Genome Atlas Research, N. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* **368**, 2059-2074, doi:10.1056/NEJMoa1301689 (2013).

103 Thiede, C. *et al.* Prevalence and prognostic impact of NPM1 mutations in 1485 adult patients with acute myeloid leukemia (AML). *Blood* **107**, 4011-4020, doi:10.1182/blood-2005-08-3167 (2006).

104 Kottaridis, P. D. *et al.* The presence of a FLT3 internal tandem duplication in patients with acute myeloid leukemia (AML) adds important prognostic information to cytogenetic risk group and response to the first cycle of chemotherapy: analysis of 854 patients from the United Kingdom Medical Research Council AML 10 and 12 trials. *Blood* **98**, 1752-1759 (2001).

105 Falini, B. *et al.* Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *N Engl J Med* **352**, 254-266, doi:10.1056/NEJMoa041974 (2005).

106 Falini, B., Nicoletti, I., Martelli, M. F. & Mecucci, C. Acute myeloid leukemia carrying cytoplasmic/mutated nucleophosmin (NPMc+ AML): biologic and clinical features. *Blood* **109**, 874-885, doi:10.1182/blood-2006-07-012252 (2007).

107 Shlush, L. I. *et al.* Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* **506**, 328-333, doi:10.1038/nature13038 (2014).

108 Jaiswal, S. *et al.* Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med* **371**, 2488-2498, doi:10.1056/NEJMoa1408617 (2014).

109 Chou, W. C. *et al.* Nucleophosmin mutations in de novo acute myeloid leukemia: the age-dependent incidences and the stability during disease evolution. *Cancer Res* **66**, 3310-3316, doi:10.1158/0008-5472.CAN-05-4316 (2006).

110 Meloni, G. *et al.* Late relapse of acute myeloid leukemia with mutated NPM1 after eight years: evidence of NPM1 mutation stability. *Haematologica* **94**, 298-300, doi:10.3324/haematol.2008.000059 (2009).

111 Ivey, A. *et al.* Assessment of Minimal Residual Disease in Standard-Risk AML. *N Engl J Med* **374**, 422-433, doi:10.1056/NEJMoa1507471 (2016).

112 Vardiman, J. W. *et al.* The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* **114**, 937-951, doi:10.1182/blood-2009-03-209262 (2009).

113 Arber, D. A. *et al.* The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* **127**, 2391-2405, doi:10.1182/blood-2016-03-643544 (2016).

114 Levis, M. & Small, D. FLT3: ITDoes matter in leukemia. *Leukemia* **17**, 1738-1752, doi:10.1038/sj.leu.2403099 (2003).

115 Levis, M. FLT3 mutations in acute myeloid leukemia: what is the best approach in 2013? *Hematology / the Education Program of the American Society of Hematology. American Society of Hematology. Education Program* **2013**, 220-226, doi:10.1182/asheducation-2013.1.220 (2013).

116 Rombouts, W. J., Blokland, I., Lowenberg, B. & Ploemacher, R. E. Biological characteristics and prognosis of adult acute myeloid leukemia with internal tandem duplications in the Flt3 gene. *Leukemia* **14**, 675-683 (2000).

117 Mead, A. J. *et al.* FLT3 tyrosine kinase domain mutations are biologically distinct from and have a significantly more favorable prognosis than FLT3 internal tandem duplications in patients with acute myeloid leukemia. *Blood* **110**, 1262-1270, doi:10.1182/blood-2006-04-015826 (2007).

118 Corces-Zimmerman, M. R., Hong, W. J., Weissman, I. L., Medeiros, B. C. & Majeti, R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators

and persist in remission. *Proc Natl Acad Sci U S A* **111**, 2548-2553, doi:10.1073/pnas.1324297111 (2014).

119 Warren, M. *et al.* Clinical impact of change of FLT3 mutation status in acute myeloid leukemia patients. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* **25**, 1405-1412, doi:10.1038/modpathol.2012.88 (2012).

120 Wander, S. A., Levis, M. J. & Fathi, A. T. The evolving role of FLT3 inhibitors in acute myeloid leukemia: quizartinib and beyond. *Therapeutic advances in hematology* **5**, 65-77, doi:10.1177/2040620714532123 (2014).

121 *AML 17 Trial Protocol*, <https://trials.cardiff.ac.uk/aml//files/aml17_protocolv2.pdf> (

122 *AML 18 Trial Design*, <http://medicine.cf.ac.uk/HCTU/our-trials/aml-18/> (

123 *AML 19 Pilot Trial* <http://medicine.cf.ac.uk/HCTU/our-trials/aml-19-pilot/> (

124 Bennett, J. M. *et al.* Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *Br J Haematol* **33**, 451-458 (1976).

125 Sanz, M. A. *et al.* Risk-adapted treatment of acute promyelocytic leukemia with all-trans-retinoic acid and anthracycline monochemotherapy: a multicenter study by the PETHEMA group. *Blood* **103**, 1237-1243, doi:10.1182/blood-2003-07-2462 (2004).

126 *AML 19 Trial Design*, <http://medicine.cf.ac.uk/HCTU/our-trials/aml-19/> (

127 Bonnet, D. & Dick, J. E. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nat Med* **3**, 730-737 (1997).

128 Jamieson, C. H. *et al.* Granulocyte-macrophage progenitors as candidate leukemic stem cells in blast-crisis CML. *N Engl J Med* **351**, 657-667, doi:10.1056/NEJMoa040258 (2004).

129 Vargaftig, J. *et al.* Frequency of leukemic initiating cells does not depend on the xenotransplantation model used. *Leukemia* **26**, 858-860, doi:10.1038/leu.2011.250 (2012).

130 Taussig, D. C. *et al.* Leukemia-initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34(-) fraction. *Blood* **115**, 1976-1984, doi:10.1182/blood-2009-02-206565 (2010).

131 Majeti, R. *et al.* CD47 is an adverse prognostic factor and therapeutic antibody target on human acute myeloid leukemia stem cells. *Cell* **138**, 286-299, doi:10.1016/j.cell.2009.05.045 (2009).

132 Jin, L., Hope, K. J., Zhai, Q., Smadja-Joffe, F. & Dick, J. E. Targeting of CD44 eradicates human acute myeloid leukemic stem cells. *Nat Med* **12**, 1167-1174, doi:10.1038/nm1483 (2006).

133 Broxmeyer, H. E., Jacobsen, N., Kurland, J., Mendelsohn, N. & Moore, A. S. In vitro suppression of normal granulocytic stem cells by inhibitory activity derived from human leukemia cells. *J Natl Cancer Inst* **60**, 497-511 (1978).

134 Miller, A. M., Marmor, J. B., Page, P. L., Russell, J. L. & Robinson, S. H. Unregulated growth of murine leukemic cells and suppression of normal granulocyte growth in diffusion chamber cultures. *Blood* **47**, 737-745 (1976).

135 van der Pol, M. A. *et al.* Assessment of the normal or leukemic nature of CD34+ cells in acute myeloid leukemia with low percentages of CD34 cells. *Haematologica* **88**, 983-993 (2003).

136 Cheng, H. *et al.* Leukemic marrow infiltration reveals a novel role for Egr3 as a potent inhibitor of normal hematopoietic stem cell proliferation. *Blood* **126**, 1302-1313, doi:10.1182/blood-2015-01-623645 (2015).

137 Min, I. M. *et al.* The transcription factor EGR1 controls both the proliferation and localization of hematopoietic stem cells. *Cell Stem Cell* **2**, 380-391, doi:10.1016/j.stem.2008.01.015 (2008).

138 Li, S. *et al.* The transcription factors Egr2 and Egr3 are essential for the control of inflammation and antigen-induced proliferation of B and T cells. *Immunity* **37**, 685-696, doi:10.1016/j.immuni.2012.08.001 (2012).

139 Safford, M. *et al.* Egr-2 and Egr-3 are negative regulators of T cell activation. *Nature immunology* **6**, 472-480, doi:10.1038/ni1193 (2005).

140 Geyh, S. *et al.* Functional inhibition of mesenchymal stromal cells in acute myeloid leukemia. *Leukemia* **30**, 683-691, doi:10.1038/leu.2015.325 (2016).

141 Kim, J. A. *et al.* Microenvironmental remodeling as a parameter and prognostic factor of heterogeneous leukemogenesis in acute myelogenous leukemia. *Cancer Res* **75**, 2222-2231, doi:10.1158/0008-5472.CAN-14-3379 (2015).

142 Hanoun, M. *et al.* Acute myelogenous leukemia-induced sympathetic neuropathy promotes malignancy in an altered hematopoietic stem cell niche. *Cell Stem Cell* **15**, 365-375, doi:10.1016/j.stem.2014.06.020 (2014).

143 Frisch, B. J. *et al.* Functional inhibition of osteoblastic cells in an in vivo mouse model of myeloid leukemia. *Blood* **119**, 540-550, doi:10.1182/blood-2011-04-348151 (2012).

144 Raaijmakers, M. H. *et al.* Bone progenitor dysfunction induces myelodysplasia and secondary leukaemia. *Nature* **464**, 852-857, doi:10.1038/nature08851 (2010).

145 Blau, O. *et al.* Mesenchymal stromal cells of myelodysplastic syndrome and acute myeloid leukemia patients have distinct genetic abnormalities compared with leukemic blasts. *Blood* **118**, 5583-5592, doi:10.1182/blood-2011-03-343467 (2011).

146 Johnstone, R. M., Adam, M., Hammond, J. R., Orr, L. & Turbide, C. Vesicle formation during reticulocyte maturation. Association of plasma membrane activities with released vesicles (exosomes). *J Biol Chem* **262**, 9412-9420 (1987).

147 van der Pol, E., Boing, A. N., Harrison, P., Sturk, A. & Nieuwland, R. Classification, functions, and clinical relevance of extracellular vesicles. *Pharmacological reviews* **64**, 676-705, doi:10.1124/pr.112.005983 (2012).

148 Huan, J. *et al.* Coordinate regulation of residual bone marrow function by paracrine trafficking of AML exosomes. *Leukemia* **29**, 2285-2295, doi:10.1038/leu.2015.163 (2015).

149 Horiguchi, H. *et al.* Extracellular vesicle miR-7977 is involved in hematopoietic dysfunction of mesenchymal stromal cells via poly(rC) binding protein 1 reduction in myeloid neoplasms. *Haematologica* **101**, 437-447, doi:10.3324/haematol.2015.134932 (2016).

150 Miyamoto, T., Weissman, I. L. & Akashi, K. AML1/ETO-expressing nonleukemic stem cells in acute myelogenous leukemia with 8;21 chromosomal translocation. *Proc Natl Acad Sci U S A* **97**, 7521-7526 (2000).

151 Jan, M. *et al.* Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. *Sci Transl Med* **4**, 149ra118, doi:10.1126/scitranslmed.3004315 (2012).

152 Busque, L. *et al.* Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nat Genet* **44**, 1179-1181, doi:10.1038/ng.2413 (2012).

153 Steensma, D. P. *et al.* Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood* **126**, 9-16, doi:10.1182/blood-2015-03-631747 (2015).

154 Killick, S. B. *et al.* Guidelines for the diagnosis and management of adult aplastic anaemia. *Br J Haematol* **172**, 187-207, doi:10.1111/bjh.13853 (2016).

155 Boultwood, J., Pellagatti, A., McKenzie, A. N. & Wainscoat, J. S. Advances in the 5q- syndrome. *Blood* **116**, 5803-5811, doi:10.1182/blood-2010-04-273771 (2010).

156 Jaju, R. J. *et al.* Combined immunophenotyping and FISH identifies the involvement of B-cells in 5q- syndrome. *Genes, chromosomes & cancer* **29**, 276-280 (2000).

157 Vannucchi, A. M. *et al.* Philadelphia chromosome-negative chronic myeloproliferative neoplasms: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of oncology : official journal of the European Society for Medical Oncology / ESMO* **26 Suppl 5**, v85-99, doi:10.1093/annonc/mdv203 (2015).

158 Kennedy, J. A. *et al.* Treatment outcomes following leukemic transformation in Philadelphia-negative myeloproliferative neoplasms. *Blood* **121**, 2725-2733, doi:10.1182/blood-2012-10-464248 (2013).

159 Pulsoni, A. *et al.* M4 acute myeloid leukemia: the role of eosinophilia and cytogenetics in treatment response and survival. The GIMEMA experience. *Haematologica* **93**, 1025-1032, doi:10.3324/haematol.11889 (2008).

160 Lugthart, S. *et al.* Clinical, molecular, and prognostic significance of WHO type inv(3)(q21q26.2)/t(3;3)(q21;q26.2) and various other 3q abnormalities in acute myeloid leukemia. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **28**, 3890-3898, doi:10.1200/JCO.2010.29.2771 (2010).

161 Keinanen, M., Griffin, J. D., Bloomfield, C. D., Machnicki, J. & de la Chapelle, A. Clonal chromosomal abnormalities showing multiple-cell-lineage involvement in acute myeloid leukemia. *N Engl J Med* **318**, 1153-1158, doi:10.1056/NEJM198805053181803 (1988).

162 van Lom, K., Hagemeijer, A., Vandekerckhove, F., Smit, E. M. & Lowenberg, B. Cytogenetic clonality analysis: typical patterns in myelodysplastic syndrome and acute myeloid leukaemia. *Br J Haematol* **93**, 594-600 (1996).

163 Knuutila, S., Majander, P. & Ruutu, T. 8;21 and 15;17 translocations: abnormalities in a single cell lineage in acute myeloid leukemia. *Acta haematologica* **92**, 88-90 (1994).

164 Haferlach, T. *et al.* The abnormal eosinophils are part of the leukemic cell population in acute myelomonocytic leukemia with abnormal eosinophils (AML M4Eo) and carry the pericentric inversion 16: a combination of May-Grunwald-Giemsa staining and fluorescence in situ hybridization. *Blood* **87**, 2459-2463 (1996).

165 Shi, G., Weh, H. J., Duhrsen, U., Zeller, W. & Hossfeld, D. K. Chromosomal abnormality inv(3)(q21q26) associated with multilineage hematopoietic progenitor cells in hematopoietic malignancies. *Cancer genetics and cytogenetics* **96**, 58-63 (1997).

166 Pasqualucci, L. *et al.* Mutated nucleophosmin detects clonal multilineage involvement in acute myeloid leukemia: Impact on WHO classification. *Blood* **108**, 4146-4155, doi:10.1182/blood-2006-06-026716 (2006).

167 Marsee, D. K., Pinkus, G. S. & Yu, H. CD71 (transferrin receptor): an effective marker for erythroid precursors in bone marrow biopsy specimens. *Am J Clin Pathol* **134**, 429-435, doi:10.1309/AJCPCRK3MOAOJ6AT (2010).

168 Olivier, E. N., Qiu, C., Velho, M., Hirsch, R. E. & Bouhassira, E. E. Large-scale production of embryonic red blood cells from human embryonic stem cells. *Exp Hematol* **34**, 1635-1642, doi:10.1016/j.exphem.2006.07.003 (2006).

169 Loken, M. R., Shah, V. O., Dattilio, K. L. & Civin, C. I. Flow cytometric analysis of human bone marrow: I. Normal erythroid development. *Blood* **69**, 255-263 (1987).

170 Hattangadi, S. M., Wong, P., Zhang, L., Flygare, J. & Lodish, H. F. From stem cell to red cell: regulation of erythropoiesis at multiple levels by multiple proteins, RNAs, and chromatin modifications. *Blood* **118**, 6258-6268, doi:10.1182/blood-2011-07-356006 (2011).

171 Wu, H., Liu, X., Jaenisch, R. & Lodish, H. F. Generation of committed erythroid BFU-E and CFU-E progenitors does not require erythropoietin or the erythropoietin receptor. *Cell* **83**, 59-67 (1995).

172 McCulloch, E. A., Siminovitch, L. & Till, J. E. Spleen-Colony Formation in Anemic Mice of Genotype Ww. *Science* **144**, 844-846 (1964).

173 Bauer, A. *et al.* The glucocorticoid receptor is required for stress erythropoiesis. *Genes Dev* **13**, 2996-3002 (1999).

174 Broudy, V. C., Lin, N. L., Priestley, G. V., Nocka, K. & Wolf, N. S. Interaction of stem cell factor and its receptor c-kit mediates lodgment and acute expansion of hematopoietic cells in the murine spleen. *Blood* **88**, 75-81 (1996).

175 Singbrant, S. *et al.* Canonical BMP signaling is dispensable for hematopoietic stem cell function in both adult and fetal liver hematopoiesis, but essential to preserve colon architecture. *Blood* **115**, 4689-4698, doi:10.1182/blood-2009-05-220988 (2010).

176 Sabri, S. *et al.* Differential regulation of actin stress fiber assembly and proplatelet formation by alpha2beta1 integrin and GPVI in human megakaryocytes. *Blood* **104**, 3117-3125, doi:10.1182/blood-2003-12-4398 (2004).

177 Arai, F. & Suda, T. Regulation of hematopoietic stem cells in the osteoblastic niche. *Advances in experimental medicine and biology* **602**, 61-67 (2007).

178 Ebbe, S. Biology of megakaryocytes. *Progress in hemostasis and thrombosis* **3**, 211-229 (1976).

179 Machlus, K. R. & Italiano, J. E., Jr. The incredible journey: From megakaryocyte development to platelet formation. *The Journal of cell biology* **201**, 785-796, doi:10.1083/jcb.201304054 (2013).

180 Schwer, H. D. *et al.* A lineage-restricted and divergent beta-tubulin isoform is essential for the biogenesis, structure and function of blood platelets. *Current biology : CB* **11**, 579-586 (2001).

181 Kelley, M. J., Jawien, W., Ortel, T. L. & Korczak, J. F. Mutation of MYH9, encoding non-muscle myosin heavy chain A, in May-Hegglin anomaly. *Nat Genet* **26**, 106-108, doi:10.1038/79069 (2000).

182 Italiano, J. E., Jr., Lecine, P., Shivdasani, R. A. & Hartwig, J. H. Blood platelets are assembled principally at the ends of proplatelet processes produced by differentiated megakaryocytes. *The Journal of cell biology* **147**, 1299-1312 (1999).

183 Ebbe, S. & Stohlman, F., Jr. Megakaryocytopoiesis in the Rat. *Blood* **26**, 20-35 (1965).

184 Tomer, A. Human marrow megakaryocyte differentiation: multiparameter correlative analysis identifies von Willebrand factor as a sensitive and distinctive marker for early (2N and 4N) megakaryocytes. *Blood* **104**, 2722-2727, doi:10.1182/blood-2004-02-0769 (2004).

185 Tomer, A., Harker, L. A. & Burstein, S. A. Purification of human megakaryocytes by fluorescence-activated cell sorting. *Blood* **70**, 1735-1742 (1987).

186 Mitjavila-Garcia, M. T. *et al.* Expression of CD41 on hematopoietic progenitors derived from embryonic hematopoietic cells. *Development* **129**, 2003-2013 (2002).

187 Bartley, T. D. *et al.* Identification and cloning of a megakaryocyte growth and development factor that is a ligand for the cytokine receptor Mpl. *Cell* **77**, 1117-1124 (1994).

188 Bussel, J. B. *et al.* A randomized, double-blind study of romiplostim to determine its safety and efficacy in children with immune thrombocytopenia. *Blood* **118**, 28-36, doi:10.1182/blood-2010-10-313908 (2011).

189 Bussel, J. B. *et al.* Eltrombopag for the treatment of chronic idiopathic thrombocytopenic purpura. *N Engl J Med* **357**, 2237-2247, doi:10.1056/NEJMoa073275 (2007).

190 Ito, T., Ishida, Y., Kashiwagi, R. & Kuriya, S. Recombinant human c-Mpl ligand is not a direct stimulator of proplatelet formation in mature human megakaryocytes. *Br J Haematol* **94**, 387-390 (1996).

191 Zuo, Z., Polski, J. M., Kasyan, A. & Medeiros, L. J. Acute erythroid leukemia. *Archives of pathology & laboratory medicine* **134**, 1261-1270, doi:10.1043/2009-0350-RA.1 (2010).

192 Deschler, B. & Lubbert, M. Acute myeloid leukemia: epidemiology and etiology. *Cancer* **107**, 2099-2107, doi:10.1002/cncr.22233 (2006).

193 Appelbaum, F. R. *et al.* Age and acute myeloid leukemia. *Blood* **107**, 3481-3485, doi:10.1182/blood-2005-09-3724 (2006).

194 Charrin, C. *et al.* Structural rearrangements of chromosome 3 in 57 patients with acute myeloid leukemia: clinical, hematological and cytogenetic features. *The hematology journal : the official journal of the European Haematology Association* **3**, 21-31, doi:10.1038/sj/thj/6200143 (2002).

195 Betz, S. A., Foucar, K., Head, D. R., Chen, I. M. & Willman, C. L. False-positive flow cytometric platelet glycoprotein IIb/IIIa expression in myeloid leukemias secondary to platelet adherence to blasts. *Blood* **79**, 2399-2403 (1992).

196 Papaemmanuil, E. *et al.* Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med* **374**, 2209-2221, doi:10.1056/NEJMoa1516192 (2016).

197 Randall, T. D. & Weissman, I. L. Phenotypic and functional changes induced at the clonal level in hematopoietic stem cells after 5-fluorouracil treatment. *Blood* **89**, 3596-3606 (1997).

198 Kikushige, Y. *et al.* TIM-3 is a promising target to selectively kill acute myeloid leukemia stem cells. *Cell Stem Cell* **7**, 708-717, doi:10.1016/j.stem.2010.11.014 (2010).

199 Zhao, X. *et al.* Targeting C-type lectin-like molecule-1 for antibody-mediated immunotherapy in acute myeloid leukemia. *Haematologica* **95**, 71-78, doi:10.3324/haematol.2009.009811 (2010).

200 Chung, S. S., Tavakkoli, M. B. S., Devlin, S. M. & Park, C. Y. in *Blood* Vol. 122 2891-2891 (Blood, 2012).

201 Bakker, A. B. *et al.* C-type lectin-like molecule-1: a novel myeloid cell surface marker associated with acute myeloid leukemia. *Cancer Res* **64**, 8443-8450, doi:10.1158/0008-5472.CAN-04-1659 (2004).

202 Hosen, N. *et al.* CD96 is a leukemic stem cell-specific marker in human acute myeloid leukemia. *Proc Natl Acad Sci U S A* **104**, 11008-11013, doi:10.1073/pnas.0704271104 (2007).

203 Gordon, M. Y., Goldman, J. M. & Gordon-Smith, E. C. 4-Hydroperoxycyclophosphamide inhibits proliferation by human granulocyte-macrophage colony-forming cells (GM-CFC) but spares more primitive progenitor cells. *Leuk Res* **9**, 1017-1021 (1985).

204 Storms, R. W. *et al.* Isolation of primitive human hematopoietic progenitors on the basis of aldehyde dehydrogenase activity. *Proc Natl Acad Sci U S A* **96**, 9118-9123 (1999).

205 Schuurhuis, G. J. *et al.* Normal hematopoietic stem cells within the AML bone marrow have a distinct and higher ALDH activity level than co-existing leukemic stem cells. *Plos One* **8**, e78897, doi:10.1371/journal.pone.0078897 (2013).

206 Pearce, D. J. *et al.* Characterization of cells with a high aldehyde dehydrogenase activity from cord blood and acute myeloid leukemia samples. *Stem cells* **23**, 752-760, doi:10.1634/stemcells.2004-0292 (2005).

207 Gerber, J. M. *et al.* A clinically relevant population of leukemic CD34(+)CD38(-) cells in acute myeloid leukemia. *Blood* **119**, 3571-3577, doi:10.1182/blood-2011-06-364182 (2012).

208 Cheung, A. M. *et al.* Aldehyde dehydrogenase activity in leukemic blasts defines a subgroup of acute myeloid leukemia with adverse prognosis and superior NOD/SCID engrafting potential. *Leukemia* **21**, 1423-1430, doi:10.1038/sj.leu.2404721 (2007).

209 Ran, D. *et al.* Aldehyde dehydrogenase activity among primary leukemia cells is associated with stem cell features and correlates with adverse clinical outcomes. *Exp Hematol* **37**, 1423-1434, doi:10.1016/j.exphem.2009.10.001 (2009).

210 Paul, S. P., Taylor, L. S., Stansbury, E. K. & McVicar, D. W. Myeloid specific human CD33 is an inhibitory receptor with differential ITIM function in recruiting the phosphatases SHP-1 and SHP-2. *Blood* **96**, 483-490 (2000).

211 Dinndorf, P. A. *et al.* Expression of normal myeloid-associated antigens by acute leukemia cells. *Blood* **67**, 1048-1053 (1986).

212 Griffin, J. D., Linch, D., Sabbath, K., Larcom, P. & Schlossman, S. F. A monoclonal antibody reactive with normal and leukemic human myeloid progenitor cells. *Leuk Res* **8**, 521-534 (1984).

213 Vitale, C. *et al.* Surface expression and function of p75/AIRM-1 or CD33 in acute myeloid leukemias: engagement of CD33 induces apoptosis of leukemic cells. *Proc Natl Acad Sci U S A* **98**, 5764-5769, doi:10.1073/pnas.091097198 (2001).

214 Taussig, D. C. *et al.* Hematopoietic stem cells express multiple myeloid markers: implications for the origin and targeted therapy of acute myeloid leukemia. *Blood* **106**, 4086-4092, doi:10.1182/blood-2005-03-1072 (2005).

215 Jaiswal, S. *et al.* CD47 is upregulated on circulating hematopoietic stem cells and leukemia cells to avoid phagocytosis. *Cell* **138**, 271-285, doi:10.1016/j.cell.2009.05.046 (2009).

216 Zhang, P. J. *et al.* Immunoreactivity of MIC2 (CD99) in acute myelogenous leukemia and related diseases. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* **13**, 452-458, doi:10.1038/modpathol.3880077 (2000).

217 Sui, L. *et al.* IgSF13, a novel human inhibitory receptor of the immunoglobulin superfamily, is preferentially expressed in dendritic cells and monocytes. *Biochem Biophys Res Commun* **319**, 920-928, doi:10.1016/j.bbrc.2004.05.065 (2004).

218 Korver, W. *et al.* Monoclonal antibodies against IREM-1: potential for targeted therapy of AML. *Leukemia* **23**, 1587-1597, doi:10.1038/leu.2009.99 (2009).

219 van Rhenen, A. *et al.* The novel AML stem cell associated antigen CLL-1 aids in discrimination between normal and leukemic stem cells. *Blood* **110**, 2659-2666, doi:10.1182/blood-2007-03-083048 (2007).

220 Larsen, H. O., Roug, A. S., Just, T., Brown, G. D. & Hokland, P. Expression of the hMICL in acute myeloid leukemia-a highly reliable disease marker at diagnosis and during follow-up. *Cytometry B Clin Cytom* **82**, 3-8, doi:10.1002/cyto.b.20614 (2012).

221 Monney, L. *et al.* Th1-specific cell surface protein Tim-3 regulates macrophage activation and severity of an autoimmune disease. *Nature* **415**, 536-541, doi:10.1038/415536a (2002).

222 Jan, M. *et al.* Prospective separation of normal and leukemic stem cells based on differential expression of TIM3, a human acute myeloid leukemia stem cell marker. *Proc Natl Acad Sci U S A* **108**, 5009-5014, doi:10.1073/pnas.1100551108 (2011).

223 Huang, Q. *et al.* A rapid, one step assay for simultaneous detection of FLT3/ITD and NPM1 mutations in AML with normal cytogenetics. *Br J Haematol* **142**, 489-492, doi:10.1111/j.1365-2141.2008.07205.x (2008).

224 Taussig, D. C. *et al.* Anti-CD38 antibody-mediated clearance of human repopulating cells masks the heterogeneity of leukemia-initiating cells. *Blood* **112**, 568-575, doi:10.1182/blood-2007-10-118331 (2008).

225 Mrozek, K. *et al.* Prognostic significance of the European LeukemiaNet standardized system for reporting cytogenetic and molecular alterations in adults with acute myeloid leukemia. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **30**, 4515-4523, doi:10.1200/JCO.2012.43.4738 (2012).

226 Grimwade, D. *et al.* The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. *Blood* **92**, 2322-2333 (1998).

227 Gale, R. E. *et al.* The impact of FLT3 internal tandem duplication mutant level, number, size, and interaction with NPM1 mutations in a large cohort of young adult patients with acute myeloid leukemia. *Blood* **111**, 2776-2784, doi:10.1182/blood-2007-08-109090 (2008).

228 Goardon, N. *et al.* Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer cell* **19**, 138-152, doi:10.1016/j.ccr.2010.12.012 (2011).

229 Pearce, D. J. *et al.* AML engraftment in the NOD/SCID assay reflects the outcome of AML: implications for our understanding of the heterogeneity of AML. *Blood* **107**, 1166-1173, doi:10.1182/blood-2005-06-2325 (2006).

230 Burnett, A. & Russell, N. H. *AML 17 Trial Protocol Version 8.0*, <https://trials.cardiff.ac.uk/aml//files/new5/AML%2017%20Protocol%20V8.0%20October%202012.pdf> (2012).

231 Walter, R. B. *et al.* Effect of complete remission and responses less than complete remission on survival in acute myeloid leukemia: a combined Eastern Cooperative Oncology Group, Southwest Oncology Group, and M. D. Anderson Cancer Center Study. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **28**, 1766-1771, doi:10.1200/JCO.2009.25.1066 (2010).

232 Hou, H. A. *et al.* TP53 mutations in de novo acute myeloid leukemia patients: longitudinal follow-ups show the mutation is stable during disease evolution. *Blood cancer journal* **5**, e331, doi:10.1038/bcj.2015.59 (2015).

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

234 Falini, B. *et al.* NPM1 mutations and cytoplasmic nucleophosmin are mutually exclusive of recurrent genetic abnormalities: a comparative analysis of 2562 patients with acute myeloid leukemia. *Haematologica* **93**, 439-442, doi:10.3324/haematol.12153 (2008).

235 Alcalay, M. *et al.* Acute myeloid leukemia bearing cytoplasmic nucleophosmin (NPMc+ AML) shows a distinct gene expression profile characterized by up-regulation of genes involved in stem-cell maintenance. *Blood* **106**, 899-902, doi:10.1182/blood-2005-02-0560 (2005).

236 Haferlach, C. *et al.* AML with mutated NPM1 carrying a normal or aberrant karyotype show overlapping biologic, pathologic, immunophenotypic, and prognostic features. *Blood* **114**, 3024-3032, doi:10.1182/blood-2009-01-197871 (2009).

237 Henckaerts, E., Langer, J. C. & Snoeck, H. W. Quantitative genetic variation in the hematopoietic stem cell and progenitor cell compartment and in lifespan are closely linked at multiple loci in BXD recombinant inbred mice. *Blood* **104**, 374-379, doi:10.1182/blood-2003-12-4304 (2004).

238 Ogawa, T., Kitagawa, M. & Hirokawa, K. Age-related changes of human bone marrow: a histometric estimation of proliferative cells, apoptotic cells, T cells, B cells and macrophages. *Mechanisms of ageing and development* **117**, 57-68 (2000).

239 Lichtman, M. A. & Rowe, J. M. The relationship of patient age to the pathobiology of the clonal myeloid diseases. *Semin Oncol* **31**, 185-197 (2004).

240 Lanino, E. *et al.* Strategies of the donor search for children with second CR ALL lacking a matched sibling donor. *Bone Marrow Transplant* **41 Suppl 2**, S75-79, doi:10.1038/bmt.2008.59 (2008).

241 Pang, W. W. *et al.* Human bone marrow hematopoietic stem cells are increased in frequency and myeloid-biased with age. *Proc Natl Acad Sci U S A* **108**, 20012-20017, doi:10.1073/pnas.1116110108 (2011).

242 Pearce, D. J., Anjos-Afonso, F., Ridler, C. M., Eddaoudi, A. & Bonnet, D. Age-dependent increase in side population distribution within hematopoiesis: implications for our understanding of the mechanism of aging. *Stem cells* **25**, 828-835, doi:10.1634/stemcells.2006-0405 (2007).

243 Mendez-Ferrer, S., Chow, A., Merad, M. & Frenette, P. S. Circadian rhythms influence hematopoietic stem cells. *Current opinion in hematology* **16**, 235-242, doi:10.1097/MOH.0b013e32832bd0f5 (2009).

244 Takagi, S., Tanaka, O. & Miura, Y. Magnetic resonance imaging of femoral marrow in patients with myelodysplastic syndromes or leukemia. *Blood* **86**, 316-322 (1995).

245 Brierley, C. K. *et al.* An extreme example of focal bone marrow involvement in acute myeloid leukemia. *American journal of hematology* **88**, 335-336, doi:10.1002/ajh.23350 (2013).

246 Pichler, H. *et al.* No impact of total or myeloid Cd34+ cell numbers on neutrophil engraftment and transplantation-related mortality after allogeneic pediatric bone marrow transplantation. *Biology of blood and marrow transplantation : journal of the American Society for Blood and Marrow Transplantation* **20**, 676-683, doi:10.1016/j.bbmt.2014.01.026 (2014).

247 Crocker, P. R., Paulson, J. C. & Varki, A. Siglecs and their roles in the immune system. *Nat Rev Immunol* **7**, 255-266, doi:10.1038/nri2056 (2007).

248 Nguyen, D. H., Ball, E. D. & Varki, A. Myeloid precursors and acute myeloid leukemia cells express multiple CD33-related Siglecs. *Exp Hematol* **34**, 728-735, doi:10.1016/j.exphem.2006.03.003 (2006).

249 Erickson-Miller, C. L. *et al.* Characterization of Siglec-5 (CD170) expression and functional activity of anti-Siglec-5 antibodies on human phagocytes. *Exp Hematol* **31**, 382-388 (2003).

250 Vitale, C. *et al.* Engagement of p75/AIRM1 or CD33 inhibits the proliferation of normal or leukemic myeloid cells. *Proc Natl Acad Sci U S A* **96**, 15091-15096 (1999).

251 von Gunten, S. *et al.* Siglec-9 transduces apoptotic and nonapoptotic death signals into neutrophils depending on the proinflammatory cytokine environment. *Blood* **106**, 1423-1431, doi:10.1182/blood-2004-10-4112 (2005).

252 Lajaunias, F., Dayer, J. M. & Chizzolini, C. Constitutive repressor activity of CD33 on human monocytes requires sialic acid recognition and phosphoinositide 3-kinase-mediated intracellular signaling. *European journal of immunology* **35**, 243-251, doi:10.1002/eji.200425273 (2005).

253 Hauswirth, A. W. *et al.* Expression of the target receptor CD33 in CD34+/CD38-/CD123+ AML stem cells. *European journal of clinical investigation* **37**, 73-82, doi:10.1111/j.1365-2362.2007.01746.x (2007).

254 Andrews, R. G., Singer, J. W. & Bernstein, I. D. Precursors of colony-forming cells in humans can be distinguished from colony-forming cells by expression of the CD33 and CD34 antigens and light scatter properties. *The Journal of experimental medicine* **169**, 1721-1731 (1989).

255 Andrews, R. G., Torok-Storb, B. & Bernstein, I. D. Myeloid-associated differentiation antigens on stem cells and their progeny identified by monoclonal antibodies. *Blood* **62**, 124-132 (1983).

256 Olweus, J., Lund-Johansen, F. & Terstappen, L. W. Expression of cell surface markers during differentiation of CD34+, CD38-/lo fetal and adult bone marrow cells. *ImmunoMethods* **5**, 179-188 (1994).

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

258 Liu, H. & Verfaillie, C. M. Myeloid-lymphoid initiating cells (ML-IC) are highly enriched in the rhodamine-c-kit(+)CD33(-)CD38(-) fraction of umbilical cord CD34(+) cells. *Exp Hematol* **30**, 582-589 (2002).

259 Pollard, J. A. *et al.* Correlation of CD33 expression level with disease characteristics and response to gemtuzumab ozogamicin containing chemotherapy in childhood AML. *Blood* **119**, 3705-3711, doi:10.1182/blood-2011-12-398370 (2012).

260 Jilani, I. *et al.* Differences in CD33 intensity between various myeloid neoplasms. *Am J Clin Pathol* **118**, 560-566, doi:10.1309/1WMW-CMXX-4WN4-T55U (2002).

261 Guglielmi, C. *et al.* Immunophenotype of adult and childhood acute promyelocytic leukaemia: correlation with morphology, type of PML gene breakpoint and clinical outcome. A cooperative Italian study on 196 cases. *Br J Haematol* **102**, 1035-1041 (1998).

262 De Propris, M. S. *et al.* High CD33 expression levels in acute myeloid leukemia cells carrying the nucleophosmin (NPM1) mutation. *Haematologica* **96**, 1548-1551, doi:10.3324/haematol.2011.043786 (2011).

263 Agency, E. M. Refusal Assessment Report for Mylotarg. (European Medicines Agency London, 2008).

264 Sullivan, N. & Lyne, L. Sensitivity of fibroblasts derived from ataxia-telangiectasia patients to calicheamicin gamma 1I. *Mutation research* **245**, 171-175 (1990).

265 Linenberger, M. L. CD33-directed therapy with gemtuzumab ozogamicin in acute myeloid leukemia: progress in understanding cytotoxicity and potential mechanisms of drug resistance. *Leukemia* **19**, 176-182, doi:10.1038/sj.leu.2403598 (2005).

266 Taksin, A. L. *et al.* High efficacy and safety profile of fractionated doses of Mylotarg as induction therapy in patients with relapsed acute myeloblastic leukemia: a prospective study of the alfa group. *Leukemia* **21**, 66-71, doi:10.1038/sj.leu.2404434 (2007).

267 Walter, R. B. *et al.* CD33 expression and P-glycoprotein-mediated drug efflux inversely correlate and predict clinical outcome in patients with acute myeloid leukemia treated with gemtuzumab ozogamicin monotherapy. *Blood* **109**, 4168-4170, doi:10.1182/blood-2006-09-047399 (2007).

268 Dean, M., Fojo, T. & Bates, S. Tumour stem cells and drug resistance. *Nature reviews. Cancer* **5**, 275-284, doi:10.1038/nrc1590 (2005).

269 Larson, R. A. *et al.* Antibody-targeted chemotherapy of older patients with acute myeloid leukemia in first relapse using Mylotarg (gemtuzumab ozogamicin). *Leukemia* **16**, 1627-1636, doi:10.1038/sj.leu.2402677 (2002).

270 Sievers, E. L. *et al.* Efficacy and safety of gemtuzumab ozogamicin in patients with CD33-positive acute myeloid leukemia in first relapse. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **19**, 3244-3254 (2001).

271 Kell, W. J. *et al.* A feasibility study of simultaneous administration of gemtuzumab ozogamicin with intensive chemotherapy in induction and consolidation in younger patients with acute myeloid leukemia. *Blood* **102**, 4277-4283, doi:10.1182/blood-2003-05-1620 (2003).

272 Petersdorf, S. H. *et al.* A phase 3 study of gemtuzumab ozogamicin during induction and postconsolidation therapy in younger patients with acute myeloid leukemia. *Blood* **121**, 4854-4860, doi:10.1182/blood-2013-01-466706 (2013).

273 Rowe, J. M. & Lowenberg, B. Gemtuzumab ozogamicin in acute myeloid leukemia: a remarkable saga about an active drug. *Blood* **121**, 4838-4841, doi:10.1182/blood-2013-03-490482 (2013).

274 Stein, E. M. *et al.* in *56th ASH Annual Meeting and Exposition* (San Francisco, 2014).

275 Kennedy, D. A. *et al.* in *AACR 106th Annual Meeting 2015* (Philadelphia, 2015).

276 Baeuerle, P. A., Kufer, P. & Bargou, R. BiTE: Teaching antibodies to engage T-cells for cancer therapy. *Current opinion in molecular therapeutics* **11**, 22-30 (2009).

277 Laszlo, G. S. *et al.* Cellular determinants for preclinical activity of a novel CD33/CD3 bispecific T-cell engager (BiTE) antibody, AMG 330, against human AML. *Blood* **123**, 554-561, doi:10.1182/blood-2013-09-527044 (2014).

278 Hu, Y. <http://bioinf.wehi.edu.au/software/elda/> (2014).

279 Hu, Y. & Smyth, G. K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *Journal of immunological methods* **347**, 70-78, doi:10.1016/j.jim.2009.06.008 (2009).

280 Khan, N. K. *et al.* in *ASH* (Orlando, USA, 2015).

281 Jedema, I. *et al.* Internalization and cell cycle-dependent killing of leukemic cells by Gemtuzumab Ozogamicin: rationale for efficacy in CD33-negative malignancies with endocytic capacity. *Leukemia* **18**, 316-325, doi:10.1038/sj.leu.2403205 (2004).

282 Mohrin, M. *et al.* Hematopoietic stem cell quiescence promotes error-prone DNA repair and mutagenesis. *Cell Stem Cell* **7**, 174-185, doi:10.1016/j.stem.2010.06.014 (2010).

283 Milyavsky, M. *et al.* A distinctive DNA damage response in human hematopoietic stem cells reveals an apoptosis-independent role for p53 in self-renewal. *Cell Stem Cell* **7**, 186-197, doi:10.1016/j.stem.2010.05.016 (2010).

284 Scharenberg, C. W., Harkey, M. A. & Torok-Storb, B. The ABCG2 transporter is an efficient Hoechst 33342 efflux pump and is preferentially expressed by immature human hematopoietic progenitors. *Blood* **99**, 507-512 (2002).

285 Bunting, K. D. ABC transporters as phenotypic markers and functional regulators of stem cells. *Stem cells* **20**, 11-20, doi:10.1634/stemcells.20-3-274 (2002).

286 Herrmann, H. *et al.* CD34(+)/CD38(-) stem cells in chronic myeloid leukemia express Siglec-3 (CD33) and are responsive to the CD33-targeting drug gemtuzumab/ozogamicin. *Haematologica* **97**, 219-226, doi:10.3324/haematol.2010.035006 (2012).

287 Jawad, M. *et al.* Analysis of factors that affect in vitro chemosensitivity of leukaemic stem and progenitor cells to gemtuzumab ozogamicin (Mylotarg) in acute myeloid leukaemia. *Leukemia* **24**, 74-80, doi:10.1038/leu.2009.199 (2010).

288 Hills, R. K. *et al.* Addition of gemtuzumab ozogamicin to induction chemotherapy in adult patients with acute myeloid leukaemia: a meta-analysis of individual patient data

from randomised controlled trials. *The Lancet. Oncology* **15**, 986-996, doi:10.1016/S1470-2045(14)70281-5 (2014).

289 Hasle, H. *et al.* Gemtuzumab ozogamicin as postconsolidation therapy does not prevent relapse in children with AML: results from NOPHO-AML 2004. *Blood* **120**, 978-984, doi:10.1182/blood-2012-03-416701 (2012).

290 Broxmeyer, H. E., Jacobsen, N., Kurland, J., Mendelsohn, N. & Moore, A. S. In vitro suppression of normal granulocytic stem cells by inhibitory activity derived from human leukemia cells. *J Natl Cancer Inst* **60**, 497-511 (1978).

291 Quesenberry, P. J. *et al.* Inhibition of normal murine hematopoiesis by leukemic cells. *N Engl J Med* **299**, 71-75, doi:10.1056/NEJM197807132990204 (1978).

292 Berger, J. & Moller, D. E. The mechanisms of action of PPARs. *Annual review of medicine* **53**, 409-435, doi:10.1146/annurev.med.53.082901.104018 (2002).

293 Yu, S. & Reddy, J. K. Transcription coactivators for peroxisome proliferator-activated receptors. *Biochim Biophys Acta* **1771**, 936-951, doi:10.1016/j.bbalip.2007.01.008 (2007).

294 Meirhaeghe, A. & Amouyel, P. Impact of genetic variation of PPARgamma in humans. *Molecular genetics and metabolism* **83**, 93-102, doi:10.1016/j.ymgme.2004.08.014 (2004).

295 Blobe, G. C., Schiemann, W. P. & Lodish, H. F. Role of transforming growth factor beta in human disease. *N Engl J Med* **342**, 1350-1358, doi:10.1056/NEJM200005043421807 (2000).

296 Ravitz, M. J. & Wenner, C. E. Cyclin-dependent kinase regulation during G1 phase and cell cycle regulation by TGF-beta. *Advances in cancer research* **71**, 165-207 (1997).

297 Dong, M. & Blobe, G. C. Role of transforming growth factor-beta in hematologic malignancies. *Blood* **107**, 4589-4596, doi:10.1182/blood-2005-10-4169 (2006).

298 Bruns, I. *et al.* Multiple myeloma-related deregulation of bone marrow-derived CD34(+) hematopoietic stem and progenitor cells. *Blood* **120**, 2620-2630, doi:10.1182/blood-2011-04-347484 (2012).

299 Connolly, E. C., Freimuth, J. & Akhurst, R. J. Complexities of TGF-beta targeted cancer therapy. *International journal of biological sciences* **8**, 964-978, doi:10.7150/ijbs.4564 (2012).

300 Jaganathan, B. G., Anjos-Afonso, F., Kumar, A. & Bonnet, D. Active RHOA favors retention of human hematopoietic stem/progenitor cells in their niche. *J Biomed Sci* **20**, 66, doi:10.1186/1423-0127-20-66 (2013).

301 Pritchard, C. C., Cheng, H. H. & Tewari, M. MicroRNA profiling: approaches and considerations. *Nature reviews. Genetics* **13**, 358-369, doi:10.1038/nrg3198 (2012).

302 Lee, Y. S., Shibata, Y., Malhotra, A. & Dutta, A. A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes Dev* **23**, 2639-2649, doi:10.1101/gad.1837609 (2009).

303 Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K. & Liu, X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *Plos One* **9**, e78644, doi:10.1371/journal.pone.0078644 (2014).

304 Chenchik, A. *et al. RT-PCR Methods for Gene Cloning and Analysis*. 309-319 (BioTechniques Books, 1998).

305 Clontech. <http://www.clontech.com/GB/Products/cDNA_Synthesis_and_Library_Construction/Next_Gen_Sequencing_Kits/Single_cell_RNA_Seq_Kits_for_mRNA_seq/Single_Cell_RNA_Seq_v4> (

306 van Dijk, E. L., Jaszczyszyn, Y. & Thermes, C. Library preparation methods for next-generation sequencing: tone down the bias. *Experimental cell research* **322**, 12-20, doi:10.1016/j.yexcr.2014.01.008 (2014).

307 Illumina. <http://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_nextera/nextera-xt/nextera-xt-library-prep-guide-15031942-01.pdf> (

308 Metzker, M. L. Sequencing technologies - the next generation. *Nature reviews. Genetics* **11**, 31-46, doi:10.1038/nrg2626 (2010).

309 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169, doi:10.1093/bioinformatics/btu638 (2015).

310 <http://www.bioconductor.org> (

311 *Primer Bank: PCR Primers for Gene Expression Detection and Quantification*, <https://pga.mgh.harvard.edu/primerbank/> (

312 *BLAST: Basic Local Alignment Search Tool*, <http://blast.ncbi.nlm.nih.gov/Blast.cgi> (

313 Hirayasu, K. Microbially cleaved immunoglobulins are sensed by the innate immune receptor LILRA2. *Nat. Microbiol.* (2016).

314 Zheng, J. *et al.* Inhibitory receptors bind ANGPTLs and support blood stem cells and leukaemia development. *Nature* **485**, 656-660, doi:10.1038/nature11095 (2012).

315 Yan, H. *et al.* IDH1 and IDH2 mutations in gliomas. *N Engl J Med* **360**, 765-773, doi:10.1056/NEJMoa0808710 (2009).

316 Mardis, E. R. *et al.* Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med* **361**, 1058-1066, doi:10.1056/NEJMoa0903840 (2009).

317 Levis, M. Targeting IDH: the next big thing in AML. *Blood* **122**, 2770-2771, doi:10.1182/blood-2013-09-522441 (2013).

318 Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer cell* **18**, 553-567, doi:10.1016/j.ccr.2010.11.015 (2010).

319 Sasaki, M. *et al.* IDH1(R132H) mutation increases murine haematopoietic progenitors and alters epigenetics. *Nature* **488**, 656-659, doi:10.1038/nature11323 (2012).

320 Shi, X. *et al.* Functions of idh1 and its mutation in the regulation of developmental hematopoiesis in zebrafish. *Blood* **125**, 2974-2984, doi:10.1182/blood-2014-09-601187 (2015).

321 Yoshida, K. *et al.* Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* **478**, 64-69, doi:10.1038/nature10496 (2011).

322 Shirai, C. L. *et al.* Mutant U2AF1 Expression Alters Hematopoiesis and Pre-mRNA Splicing In Vivo. *Cancer cell* **27**, 631-643, doi:10.1016/j.ccell.2015.04.008 (2015).

323 Graubert, T. A. *et al.* Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nat Genet* **44**, 53-57, doi:10.1038/ng.1031 (2012).

324 Dvinge, H., Kim, E., Abdel-Wahab, O. & Bradley, R. K. RNA splicing factors as oncoproteins and tumour suppressors. *Nature reviews. Cancer* **16**, 413-430, doi:10.1038/nrc.2016.51 (2016).

325 Ilagan, J. O. *et al.* U2AF1 mutations alter splice site recognition in hematological malignancies. *Genome research* **25**, 14-26, doi:10.1101/gr.181016.114 (2015).

326 Kayne, P. S. *et al.* Extremely conserved histone H4 N terminus is dispensable for growth but essential for repressing the silent mating loci in yeast. *Cell* **55**, 27-39 (1988).

327 Rai, T. S. *et al.* HIRA orchestrates a dynamic chromatin landscape in senescence and is required for suppression of neoplasia. *Genes Dev* **28**, 2712-2725, doi:10.1101/gad.247528.114 (2014).

328 Warner-Schmidt, J. L. *et al.* Role of p11 in cellular and behavioral effects of 5-HT4 receptor stimulation. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **29**, 1937-1946, doi:10.1523/JNEUROSCI.5343-08.2009 (2009).

329 Kwon, M., MacLeod, T. J., Zhang, Y. & Waisman, D. M. S100A10, annexin A2, and annexin a2 heterotetramer as candidate plasminogen receptors. *Front Biosci* **10**, 300-325 (2005).

330 Choi, K. S., Fogg, D. K., Yoon, C. S. & Waisman, D. M. p11 regulates extracellular plasmin production and invasiveness of HT1080 fibrosarcoma cells. *Faseb J* **17**, 235-246, doi:10.1096/fj.02-0697com (2003).

331 Wiley, S. R. *et al.* A novel TNF receptor family member binds TWEAK and is implicated in angiogenesis. *Immunity* **15**, 837-846 (2001).

332 Feng, S. L. *et al.* The Fn14 immediate-early response gene is induced during liver regeneration and highly expressed in both human and murine hepatocellular carcinomas. *The American journal of pathology* **156**, 1253-1261, doi:10.1016/S0002-9440(10)64996-6 (2000).

333 Meighan-Mantha, R. L. *et al.* The mitogen-inducible Fn14 gene encodes a type I transmembrane protein that modulates fibroblast adhesion and migration. *J Biol Chem* **274**, 33166-33176 (1999).

334 Han, H. *et al.* Identification of differentially expressed genes in pancreatic cancer cells using cDNA microarray. *Cancer Res* **62**, 2890-2896 (2002).

335 Lo, M. C. *et al.* Combined gene expression and DNA occupancy profiling identifies potential therapeutic targets of t(8;21) AML. *Blood* **120**, 1473-1484, doi:10.1182/blood-2011-12-395335 (2012).

336 Kaplan, M. J. *et al.* The apoptotic ligands TRAIL, TWEAK, and Fas ligand mediate monocyte death induced by autologous lupus T cells. *Journal of immunology* **169**, 6020-6029 (2002).

337 Lynch, C. N. *et al.* TWEAK induces angiogenesis and proliferation of endothelial cells. *J Biol Chem* **274**, 8455-8459 (1999).

338 Wiley, S. R. & Winkles, J. A. TWEAK, a member of the TNF superfamily, is a multifunctional cytokine that binds the TweakR/Fn14 receptor. *Cytokine & growth factor reviews* **14**, 241-249 (2003).

339 Brown, S. A., Richards, C. M., Hanscom, H. N., Feng, S. L. & Winkles, J. A. The Fn14 cytoplasmic tail binds tumour-necrosis-factor-receptor-associated factors 1, 2, 3 and 5 and mediates nuclear factor-kappaB activation. *The Biochemical journal* **371**, 395-403, doi:10.1042/BJ20021730 (2003).

340 Trebing, J. *et al.* A novel llama antibody targeting Fn14 exhibits anti-metastatic activity in vivo. *mAbs* **6**, 297-308, doi:10.4161/mabs.26709 (2014).

341 Chopra, M. *et al.* Blocking TWEAK-Fn14 interaction inhibits hematopoietic stem cell transplantation-induced intestinal cell death and reduces GVHD. *Blood* **126**, 437-444, doi:10.1182/blood-2015-01-620583 (2015).

342 Kedinger, V. & Rio, M. C. TRAF4, the unique family member. *Advances in experimental medicine and biology* **597**, 60-71, doi:10.1007/978-0-387-70630-6_5 (2007).

343 Ye, X. *et al.* TRAF family proteins interact with the common neurotrophin receptor and modulate apoptosis induction. *J Biol Chem* **274**, 30202-30208 (1999).

344 Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* **4**, 44-57, doi:10.1038/nprot.2008.211 (2009).

345 Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13, doi:10.1093/nar/gkn923 (2009).

346 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**, 27-30 (2000).

347 *Clontech: TECH NOTE*

Single-Cell Transcriptome Studies: A Powerful Way to Highlight Subtle Differences Between Cells That May Be Hidden in a Population, <http://www.clontech.com/US/Products/cDNA_Synthesis_and_Library_Construction/NGS_Learning_Resources/Technical_Notes/v4_Single-Cell_mRNA-Seq?sitex=10020:22372:US> (

348 Arreola, R., Valderrama, B., Morante, M. L. & Horjales, E. Two mammalian glucosamine-6-phosphate deaminases: a structural and genetic study. *FEBS letters* **551**, 63-70 (2003).

349 Aust, G., Zhu, D., Van Meir, E. G. & Xu, L. Adhesion GPCRs in Tumorigenesis. *Handbook of experimental pharmacology* **234**, 369-396, doi:10.1007/978-3-319-41523-9_17 (2016).

350 Saito, Y. *et al.* Maintenance of the hematopoietic stem cell pool in bone marrow niches by EVI1-regulated GPR56. *Leukemia* **27**, 1637-1649, doi:10.1038/leu.2013.75 (2013).

351 Ishii, N. *et al.* Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction. *Journal of human genetics* **51**, 1087-1099, doi:10.1007/s10038-006-0070-9 (2006).

352 Zhang, X. Q. *et al.* A long non-coding RNA signature in glioblastoma multiforme predicts survival. *Neurobiology of disease* **58**, 123-131, doi:10.1016/j.nbd.2013.05.011 (2013).

353 Sattari, A. *et al.* Upregulation of long noncoding RNA MIAT in aggressive form of chronic lymphocytic leukemias. *Oncotarget* **7**, 54174-54182, doi:10.18632/oncotarget.11099 (2016).

354 NIAID, N. *David Bioinformatics Resources 6.8*, <https://david.ncifcrf.gov> (

355 Kriegler, M., Perez, C., DeFay, K., Albert, I. & Lu, S. D. A novel form of TNF/cachectin is a cell surface cytotoxic transmembrane protein: ramifications for the complex physiology of TNF. *Cell* **53**, 45-53 (1988).

356 Locksley, R. M., Killeen, N. & Lenardo, M. J. The TNF and TNF receptor superfamilies: integrating mammalian biology. *Cell* **104**, 487-501 (2001).

357 Wajant, H., Pfizenmaier, K. & Scheurich, P. Tumor necrosis factor signaling. *Cell death and differentiation* **10**, 45-65, doi:10.1038/sj.cdd.4401189 (2003).

358 Tsimberidou, A. M. *et al.* The prognostic significance of cytokine levels in newly diagnosed acute myeloid leukemia and high-risk myelodysplastic syndromes. *Cancer* **113**, 1605-1613, doi:10.1002/cncr.23785 (2008).

359 Dybedal, I., Bryder, D., Fossum, A., Rusten, L. S. & Jacobsen, S. E. Tumor necrosis factor (TNF)-mediated activation of the p55 TNF receptor negatively regulates maintenance of cycling reconstituting human hematopoietic stem cells. *Blood* **98**, 1782-1791 (2001).

360 Dufour, C. *et al.* TNF-alpha and IFN-gamma are overexpressed in the bone marrow of Fanconi anemia patients and TNF-alpha suppresses erythropoiesis in vitro. *Blood* **102**, 2053-2059, doi:10.1182/blood-2003-01-0114 (2003).

361 Pronk, C. J., Veiby, O. P., Bryder, D. & Jacobsen, S. E. Tumor necrosis factor restricts hematopoietic stem cell activity in mice: involvement of two distinct receptors. *The Journal of experimental medicine* **208**, 1563-1570, doi:10.1084/jem.20110752 (2011).

362 Goselink, H. M., Willemze, R. & Falkenburg, J. H. F. Tumor necrosis factor alpha production by acute myeloid leukemia blasts results in impaired production of normal hematopoietic progenitor cells. *Exp Hematol* **28**, 69-70 (2000).

363 Grana, X., Garriga, J. & Mayol, X. Role of the retinoblastoma protein family, pRB, p107 and p130 in the negative control of cell growth. *Oncogene* **17**, 3365-3383, doi:10.1038/sj.onc.1202575 (1998).

364 Eber, S. W. *et al.* Ankyrin-1 mutations are a major cause of dominant and recessive hereditary spherocytosis. *Nat Genet* **13**, 214-218, doi:10.1038/ng0696-214 (1996).

365 Love, P. E., Warzecha, C. & Li, L. Ldb1 complexes: the new master regulators of erythroid gene transcription. *Trends in genetics : TIG* **30**, 1-9, doi:10.1016/j.tig.2013.10.001 (2014).

366 Whitman, S. P. *et al.* FLT3 internal tandem duplication associates with adverse outcome and gene- and microRNA-expression signatures in patients 60 years of age or older with primary cytogenetically normal acute myeloid leukemia: a Cancer and Leukemia Group B study. *Blood* **116**, 3622-3626, doi:10.1182/blood-2010-05-283648 (2010).

367 Butko, E., Pouget, C. & Traver, D. Complex regulation of HSC emergence by the Notch signaling pathway. *Dev Biol* **409**, 129-138, doi:10.1016/j.ydbio.2015.11.008 (2016).

368 Jacobsen, S. E. Defining 'stemness': Notch and Wnt join forces? *Nature immunology* **6**, 234-236, doi:10.1038/ni0305-234 (2005).

369 Wang, W. *et al.* Notch Receptor-Ligand Engagement Maintains Hematopoietic Stem Cell Quiescence and Niche Retention. *Stem cells* **33**, 2280-2293, doi:10.1002/stem.2031 (2015).

370 Cheung, T. H. & Rando, T. A. Molecular regulation of stem cell quiescence. *Nature reviews. Molecular cell biology* **14**, 329-340, doi:10.1038/nrm3591 (2013).

371 Kornblau, S. M. *et al.* Recurrent expression signatures of cytokines and chemokines are present and are independently prognostic in acute myelogenous leukemia and myelodysplasia. *Blood* **116**, 4251-4261, doi:10.1182/blood-2010-01-262071 (2010).

372 Bandura, D. R. *et al.* Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Analytical chemistry* **81**, 6813-6822, doi:10.1021/ac901049w (2009).

373 Bendall, S. C. *et al.* Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332**, 687-696, doi:10.1126/science.1198704 (2011).

374 Perfetto, S. P., Chattopadhyay, P. K. & Roederer, M. Seventeen-colour flow cytometry: unravelling the immune system. *Nat Rev Immunol* **4**, 648-655, doi:10.1038/nri1416 (2004).

375 Maes, K. *et al.* In anemia of multiple myeloma, hepcidin is induced by increased bone morphogenetic protein 2. *Blood* **116**, 3635-3644, doi:10.1182/blood-2010-03-274571 (2010).

376 Silvestris, F., Tucci, M., Cafforio, P. & Dammacco, F. Fas-L up-regulation by highly malignant myeloma plasma cells: role in the pathogenesis of anemia and disease progression. *Blood* **97**, 1155-1164 (2001).

377 Silvestris, F., Cafforio, P., Tucci, M. & Dammacco, F. Negative regulation of erythroblast maturation by Fas-L(+)/TRAIL(+) highly malignant plasma cells: a major pathogenetic mechanism of anemia in multiple myeloma. *Blood* **99**, 1305-1313 (2002).

378 Colmone, A. *et al.* Leukemic cells create bone marrow niches that disrupt the behavior of normal hematopoietic progenitor cells. *Science* **322**, 1861-1865, doi:10.1126/science.1164390 (2008).

Appendix 1: Stock Solutions

Manufacturers' Details

4', 6-Diamidino-2-phenyindole (DAPI)	Sigma, Cat: D9564
Dulbecco's Modified Eagle's Medium (DMEM)	Sigma, Cat: D6046
Dulbecco's Phosphate Buffered Solution (PBS)	Sigma, Cat: D8537
Ethanol Absolute, 200 Proof	Sigma, Cat: E7023-500
Fetal Bovine Serum (FCS)	Gibco, Cat: 10500-064
Iscove's Modified Dulbecco's Medium (IMDM)	Gibco, Cat: Z1980-032
Lymphoprep™	Fresnius, Cat: 2015-06
MethoCult H4435	Stem Cell Technologies, Cat: 04435
MyeloCult H5100	Stem Cell Technologies, Cat: 05100
Streptomycin/Penicillin	Sigma, Cat: P4333
Trypsin-EDTA	Sigma, Cat: T4299

Stock Solution Preparation

Activated Aldefluor™ Reagent: 50µg dry inactive Aldefluor™ Reagent (Stem Cell Technologies, Cat: 01703) was dissolved in 25µl DMSO (Stem Cell Technologies, Cat: 01706) at room temperature. After 1 minute, 25µl 2N Hydrochloric acid (Stem Cell Technologies, Cat: 01704) was added and the mixture left at room temperature for 15 minutes. 360µl Aldefluor™ Assay Buffer was added to the vial and mixed prior to storage at -20°C.

ALDH/DNase/DAPI: To 10ml Aldefluor™ Assay Buffer, 100µl DNase and 5µl DAPI were added.

DAPI Stock 500µg/ml: To 10mg DAPI (Sigma, Cat D9564), 400µl of water was added to make a primary stock of 25mg/ml. A secondary 1/50 dilution was made by adding 100 µl of primary stock to 4.9ml of water, to make the used stock of 500µg/ml. 500µl aliquots were stored at -20°C.

DNase: One vial of DNase (Sigma, Cat: D4513) was suspended with 11ml of PBS, mixed well and left to dissolve for 20 minutes. Resultant 500mcl aliquots were frozen at -20°C.

2% HAG: 1g Human γ-globulins (Sigma, Cat: G4386) was mixed with 50ml PBS, and left to dissolve for 30 minutes at room temperature. The resultant solution was filtered through a 0.22µm Millex GP Filter Unit to remove undissolved matter, and aliquoted into 500mcl aliquots. These were stored at -20°C.

2% HAG dissolved in ALDH Buffer: As above, but 50ml PBS was replaced by 50ml Aldefluor™ Assay Buffer (Stem Cell Technologies, Cat: 01702).

2% PBS: One 500ml bottle of PBS (Sigma, Cat: D8537) was mixed with 10ml FBS (Gibco, Cat: D6046) and 5ml Streptomycin/Penicillin (Sigma, Cat: P4333).

2% PBS/DNase/DAPI: To 10ml 2% PBS, add 100µl DNase and 5µl DAPI.

10mM Tris HCL: 0.788g Trizma-HCl (Sigma, Cat: T5941) diluted in 500ml Millipore water. Ph adjusted to 8.0 using buffers.

Appendix 2: Methodology for the AML-Stem Cell Co-Culture Assays

App 2.1 The 10 day in vitro co-culture assay

This is a functional assay designed to model the interaction between stem cells and AML in vitro. This assay was originally developed by Dr D. Taussig and Dr F. Anjos-Afonso at the LRI. Stem cells are grown in co-culture with AML cells, in the presence of a feeder layer of irradiated stem cells. After 10 days in culture, the divisional status of the stem cells (as measured by their incorporation of the BRDU dye) is determined, and compared to controls. No results from this assay are included in this thesis, but for completeness the protocol is included here.

App 2.1.1 Feeder Cell Layer Preparation

MS5 cells were plated into a 24 well plate in IMDM media, at a concentration of 40,000 cells per well in a total volume of 600µl. Cells were incubated at 37^{o}C; 5% CO_2. After 48 hours in culture, the plate was irradiated at 7.5Gy. Two hours post irradiation, the well media was changed to H5100 Myelocult^{TM} (Stem Cell Technologies).

App 2.1.2 Addition of AML cells

The following day, primary AML samples (normally samples obtained from leucophoresis of AML patients at presentation, rather than BM) were thawed, washed and resuspended to a concentration of 12x10^6 cells/ml. 720,000 cells were added to each well, to give an AML cell concentration of 1.2x10^6 cells). Care was taken that the volume of the added AML cells did not exceed 10% of the total well volume.

App 2.1.3 Normal Stem Cells

App 2.1.3.1 Selection of Stem cells

24 hours later, stem cells were added to the system. Previously lineage-depleted GMPB samples were thawed, washed and stained with CD34 as previously described. CD34$^+$ cells were sorted as previously described under sterile conditions using the BD Aria Cell Sorter. The resultant cells were then stained using the CellVue Maroon cell labelling kit (eBiosciences 880870), in a technique modified for small cell numbers described below.

App 2.1.3.2 Stem cell Labelling

Sorted CD34$^+$ cells were spun (1500 rpm, 10 minutes), supernatant aspirated off and cells resuspended in 3ml RPMI, and spun (1500 rpm, 6 minutes). Meanwhile, a 2x working dye

solution was prepared by adding 1µl of dye to 250µl Diluent. The supernatant was aspirated, and cells resuspended in 250µl Diluent C. The cells were then added to the 2x dye solution, and immediately mixed by uniform pipetting. The cells were then incubated for 2 to 5 minutes at room temperature. Staining was stopped by the addition of 500µl FCS, and samples rested for 1 minute. 3ml of RPMI with 10% FCS was then added, and the sample centrifuged (6 minutes, 1500rpm). Sample was washed 3 times with 4ml 10% RPMI. The cell were then resuspended in 500µl of 10% RPMI, and cell numbers counted using a haemocytometer as previously described.

BRDU (BD, Cat 559619) was diluted from stock to a concentration of 1mM per ml (31µl stock to 1ml PBS). 6µl was added to each well to produce a working concentration of 10µM.

Cells were then added to the wells (to a maximum of 5000 cells per well) in a small (approximately 10µl) volume. The plate was returned to the incubator for a further 6 days at 37^{o}C.

App 2.1.4 Processing of Cells prior to Analysis

On Day 11 of the assay, wells were visually inspected under the microscope to assess cellular proliferation. The contents of each well were pipetted into a new labelled test tube. Each well was washed with 700µl of PBS, and the contents added to the collection tube. Adherent cells were removed by the addition of 700µl trypsin, followed by incubation at 37^{o}C for 15 to 20 minutes. The contents were transferred into the appropriate tube. The wells were then washed again with PBS as described.

After centrifugation, cells were resuspended in 1ml Fixed Viability Dye (Fixable Viability Dye eFluor 780, eBioscience 65-0865-14, APCCy7) (the dye having been prediluted to a ratio of 1µl dye: 1ml PBS), and incubated at 4^{o}C for 30 minutes. Cells were washed in 3ml 2% PBS, centrifuged (1300 rpm, 6 minutes), and the supernatant discarded. 5µl CD45-PE (BD, 555483) and 1µl Sca-1-Pecy7 (BD, 558162, Clone D7) were added to each tube, and cells incubated at room temperature for 15 minutes. Cells were washed in 4ml 2% PBS, centrifuged (1300rpm, 6 minutes), and supernatant discarded.

Cells were then fixed as follows. Cells were resuspended with 100µl BD Cytofix/Cytoperm buffer (BD, 51-2090KZ), and incubated at room temperature for 15 minutes. Cells were then washed in 1ml PermWash buffer (BD 51-2091KZ), and centrifuged (1400rpm, 6 minutes). Cells were then resuspended in 100µl BD Cytoperm Plus buffer (BD, 51-2356KC), and incubated on

ice for 10 minutes. 1ml PermWash buffer was then added, and cells centrifuged (1400 rpm, 6 minutes). Cells were then resuspended in 100µl BD Cytofix/Cytoperm buffer, and incubated at room temperature for 5 minutes. 1ml PermWash buffer was then added, and cells centrifuged (1400rpm, 6 min). Cells were then resuspended with 100µl dilute DNase (made of 300µl STOCK DNase and 700 µl PBS), and incubated at 37^{o}C for 1 hour. 1ml PermWash buffer was then added, and cells centrifuged (1400rpm, 6 minutes). 15µl of anti-BRDU FITC antibody (BD, 556028) was added, and the tubes incubated at room temperature for 20 minutes. 1ml PermWash buffer was then added, and cells centrifuged (1400rpm, 6 minutes). Cells were resuspended in 300µl of a 1µg/ml DAPI/2% PBS solution. Cells were then incubated at 4^{o}C for at least 20 minutes prior to FACS analysis.

App 2.2. The 7 week in vitro conditioned media assay

App 2.2.1 Introduction

This was an attempt to develop an in vitro method of the interaction between normal HSCs and AML, with a modification of the 7 week culture assay. At the time of design, I hoped it might reflect the effect of AML on the long-term repopulating HSCs, rather than the 10 day co-culture assay described above, which must largely reflect the activity of the short term repopulating cells.

Rather than co-culture of AML and HSCs, this assay tests the effect of media conditioned by AML on the growth of HSCs over 7 weeks. This approach is based on the work of Broxmeyer, who reported that normal HSCs were inhibited by the apparent secretion of a diffusible factor from AML cells, and that this interaction did not require cell-cell contact. If AML cells and HSCs were truly co-cultured together in a 7 week culture assay, it would be impossible to assess the source (normal or malignant) of any visible colonies at the end of this period. Secondarily, as discussed at length in Chapter 4, different AMLs grow very differently in a 7 week culture assay (with good risk disease patients appearing to grow less well), and therefore any inhibition effect might fluctuate significantly over the 7 week culture period.

The significance of the media being conditioned by AML (rather than the same numbers of non-malignant cells) was tested for by the use of a control media conditioned by GMPB cells plated at the same concentration.

App 2.2.2 Plate set up

A 96 well flat bottom cell culture plate (Costar, Cat: 3598) was coated with collagen using the following method. 50µl of Collagen Solution (Stem Cell Technologies, Cat: 04902) was added to

each well, and left to dry for 2 minutes before the excess collagen was removed. The plate was left open under a hood to dry for at least an hour. Subsequently, each well was washed with 100µl PBS.

App 2.2.3 MS5 preparation

MS-5 cells were prepared at a concentration of $1x10^6$ cells/ml in IMDM/10% FCS, and placed in a 50ml Falcon Tube. The cells were irradiated at a dose of 70Gy (RadSource 2000, X-ray emitter). Cell viability one hour post-irradiation was assessed using a Beckmann Coulter Vicell XR Cell Viability Counter.

App 2.2.3.1 MS5 plating for LTC

Cells were pelleted, and re-suspended in MyeloCult™ H5100 (Stem Cell Technologies, Cat: 05100) to a concentration of 125,000 cells/ml. 100µl of cell/H5100 suspension was added to 60 wells per plate, having filled the outside wells with 100µl water.

App 2.2.3.2 MS5 plating for AML & GMPB Culture

At the same time, after MS5 irradiation, MS5 cells were resuspended in Myelocult™ H5100 to a concentration of 66,666 cells/ml. 2.4 ml of cell/H5100 suspension was added to each well of a 6 well plate (Costar, 6 well cell culture plate, cat 3506).

App 2.2.4 AML and GMPB cells plated for preparation of conditioned media

Primary cells from an AML known to grow in culture were taken and GMPB were thawed, washed and counted as previously described. Cells were plated into a 6 well plate, previously coated with MS5 cells as detailed above. Cells were added to a concentration of $1.5x10^6$ cells per well.

Plates were then placed in a long term incubator at 37°C, 5% CO_2 for 3 days. At the end of this period, cell growth was confirmed by visual examination of the wells. Media was removed by pipette and filtered through a 0.22µm diameter pore membrane to remove cells (Millex GP, Cat SLGV033RS). Media was snap frozen at -80°C .

App 2.2.5 HSC sorting and plating

The following day, $CD34^+38^-$ HSCs were sorted as previously described from a GMPB sample. Cells were pelleted (1500rpm, 10 minutes), diluted and then plated in serial dilutions of 40, 20, 10, 5 and 2 cells per well. A maximum additional fluid volume of 10µl per well was added. 20 replicates for each cell dilution were added.

The entire experiment was set up in triplicate to allow for comparative testing of the effects of AML-conditioned media, GMPB-conditioned media, and unconditioned media (positive control) on HSC growth.

App 2.2.6 Plate Management weeks 2 to 5

The plates were then placed in a long term incubator at 37^{o}C, 5% CO_2. Weekly, from Week one to four, the media was changed as followed. Taking care to not disturb the bottom of each well, 65µl of media was removed from the top of each well using a multichannel pipette. This was replaced with 75µl of MyeloCult^{TM} H5100, which was pre-prepared as a mastermix of 40µl fresh Myelocult^{TM} 5100 and 35µl thawed conditioned (either by AML or GMPB) Myelocult^{TM}. In the third set of plates, media changes were made with 75µl of unconditioned media.

At the end of week five of culture, the media was removed from each well, using a Gilson Safe Aspiration Station, having taken care not to disturb the 10 to 20µl at the bottom of each well. Using a Gilson Distriman pipette, 100µl of MethoCult H4435 enriched Methylcellulose medium was added to each well, and the plates returned to the incubator. After a further two weeks culture at 37^{o}C, each individual well was examined for the presence of myeloid and erythroid colonies and scored as positive or negative. An initial stem cell concentration was calculated using the ELDA limiting dilution assay analysis for each condition.

Appendix 3: Genes with significantly altered expression between AML-exposed and Control HSCs, as detected by RNA-Seq Attempt 1

Table App 3-1 Genes upregulated in HSCs from Controls versus AML (Fold change >2; p<0.05)

ID	Gene	Description	Log Fold Change	P.Value	adj.P. Val
ENSG00000090104	RGS1	regulator of G-protein signalling 1	3.23	0.0179	1.00
ENSG00000105383	CD33	CD33 molecule	2.98	0.0046	1.00
ENSG00000163564	PYHIN1	pyrin and HIN domain family, member 1	2.91	0.0280	1.00
ENSG00000187626	ZKSCAN4	zinc finger with KRAB and SCAN domains 4	2.90	0.0026	1.00
ENSG00000239998	LILRA2	leukocyte immunoglobulin-like receptor, subfamily A (with TM domain), member 2	2.87	0.0027	1.00
ENSG00000167920	TMEM99	transmembrane protein 99	2.86	0.0119	1.00
ENSG00000188000	OR7D2	olfactory receptor, family 7, subfamily D, member 2	2.85	0.0256	1.00
ENSG00000169908	TM4SF1	transmembrane 4 L six family member 1	2.85	0.0178	1.00
ENSG00000197128	ZNF772	zinc finger protein 772	2.79	0.0090	1.00
ENSG00000133574	GIMAP4	GTPase, IMAP family member 4	2.76	0.0069	1.00
ENSG00000186603	HPDL	4-hydroxyphenylpyruvate dioxygenase-like	2.67	0.0005	1.00
ENSG00000169515	CCDC8	coiled-coil domain containing 8	2.59	0.0407	1.00
ENSG00000173114	LRRN3	leucine rich repeat neuronal 3	2.59	0.0010	1.00
ENSG00000157542	KCNJ6	potassium channel, inwardly rectifying subfamily J, member 6	2.44	0.0311	1.00
ENSG00000109163	GNRHR	gonadotropin-releasing hormone receptor	2.42	0.0103	1.00
ENSG00000180938	ZNF572	zinc finger protein 572	2.41	0.0046	1.00
ENSG00000143248	RGS5	regulator of G-protein signalling 5	2.39	0.0027	1.00
ENSG00000105374	NKG7	natural killer cell granule protein 7	2.39	0.0412	1.00
ENSG00000133636	NTS	Neurotensin	2.38	0.0076	1.00
ENSG00000021826	CPS1	carbamoyl-phosphate synthase 1, mitochondrial	2.37	0.0239	1.00
ENSG00000179933	C14orf119	chromosome 14 open reading frame 119	2.37	0.0006	1.00
ENSG00000117228	GBP1	guanylate binding protein 1, interferon-inducible	2.37	0.0062	1.00
ENSG00000164120	HPGD	hydroxyprostaglandin dehydrogenase 15-(NAD)	2.34	0.0377	1.00
ENSG00000144339	TMEFF2	transmembrane protein with EGF-like and two follistatin-like domains 2	2.32	0.0160	1.00
ENSG00000173452	TMEM196	transmembrane protein 196	2.29	0.0165	1.00
ENSG00000147206	NXF3	nuclear RNA export factor 3	2.28	0.0074	1.00
ENSG00000159640	ACE	angiotensin I converting enzyme	2.20	0.0002	1.00
ENSG00000145723	GIN1	gypsy retrotransposon integrase 1	2.20	0.0287	1.00

ENSG00000184730	APOBR	apolipoprotein B receptor	2.18	0.0489	1.00
ENSG00000099330	OCEL1	occludin/ELL domain containing 1	2.18	0.0146	1.00
ENSG00000189057	FAM111B	family with sequence similarity 111, member B	2.17	0.0132	1.00
ENSG00000238227	C9orf69	chromosome 9 open reading frame 69	2.15	0.0385	1.00
ENSG00000119328	FAM206A	family with sequence similarity 206, member A	2.14	0.0254	1.00
ENSG00000228623	ZNF883	zinc finger protein 883	2.13	0.0186	1.00
ENSG00000185112	FAM43A	family with sequence similarity 43, member A	2.11	0.0079	1.00
ENSG00000106013	ANKRD7	ankyrin repeat domain 7	2.10	0.0256	1.00
ENSG00000108679	LGALS3BP	lectin, galactoside-binding, soluble, 3 binding protein	2.09	0.0043	1.00
ENSG00000196290	NIF3L1	NIF3 NGG1 interacting factor 3-like 1	2.07	0.0145	1.00
ENSG00000152782	PANK1	pantothenate kinase 1	2.06	0.0232	1.00
ENSG00000137860	SLC28A2	solute carrier family 28 (concentrative nucleoside transporter), member 2	2.06	0.0464	1.00
ENSG00000198128	OR2L3	olfactory receptor, family 2, subfamily L, member 3	2.06	0.0404	1.00
ENSG00000169896	ITGAM	integrin, alpha M (complement component 3 receptor 3 subunit)	2.06	0.0201	1.00
ENSG00000135346	CGA	glycoprotein hormones, alpha polypeptide	2.03	0.0384	1.00
ENSG00000061492	WNT8A	wingless-type MMTV integration site family, member 8A	2.02	0.0277	1.00
ENSG00000060709	RIMBP2	RIMS binding protein 2	2.02	0.0113	1.00
ENSG00000152766	ANKRD22	ankyrin repeat domain 22	2.01	0.0312	1.00
ENSG00000115525	ST3GAL5	ST3 beta-galactoside alpha-2,3-sialyltransferase 5	2.00	0.0422	1.00
ENSG00000166510	CCDC68	coiled-coil domain containing 68	2.00	0.0461	1.00
ENSG00000204644	ZFP57	ZFP57 zinc finger protein	2.00	0.0132	1.00
ENSG00000187713	TMEM203	transmembrane protein 203	1.99	0.0029	1.00
ENSG00000171700	RGS19	regulator of G-protein signaling 19	1.97	0.0321	1.00
ENSG00000197568	HHLA3	HERV-H LTR-associating 3	1.97	0.0451	1.00
ENSG00000139914	FITM1	fat storage-inducing transmembrane protein 1	1.96	0.0045	1.00
ENSG00000166347	CYB5A	cytochrome b5 type A (microsomal)	1.96	0.0138	1.00
ENSG00000103343	ZNF174	zinc finger protein 174	1.95	0.0398	1.00
ENSG00000174007	CEP19	centrosomal protein 19kDa	1.94	0.0153	1.00
ENSG00000180257	ZNF816	zinc finger protein 816	1.92	0.0493	1.00
ENSG00000090339	ICAM1	intercellular adhesion molecule 1	1.90	0.0122	1.00
ENSG00000060140	STYK1	serine/threonine/tyrosine kinase 1	1.88	0.0048	1.00
ENSG00000003056	M6PR	mannose-6-phosphate receptor (cation dependent)	1.87	0.0195	1.00
ENSG00000049883	PTCD2	pentatricopeptide repeat domain 2	1.86	0.0355	1.00
ENSG00000143178	TBX19	T-box 19	1.85	0.0205	1.00

ENSG00000156140	ADAMTS3	ADAM metallopeptidase with thrombospondin type 1 motif, 3	1.85	0.0316	1.00
ENSG00000150540	HNMT	histamine N-methyltransferase	1.83	0.0487	1.00
ENSG00000204442	FAM155A	family with sequence similarity 155, member A	1.83	0.0053	1.00
ENSG00000133665	DYDC2	DPY30 domain containing 2	1.83	0.0187	1.00
ENSG00000127903	ZNF835	zinc finger protein 835	1.83	0.0179	1.00
ENSG00000196242	OR2C3	olfactory receptor, family 2, subfamily C, member 3	1.81	0.0381	1.00
ENSG00000184276	DEFB108B	defensin, beta 108B	1.80	0.0291	1.00
ENSG00000125895	TMEM74B	transmembrane protein 74B	1.79	0.0212	1.00
ENSG00000116791	CRYZ	crystallin, zeta (quinone reductase)	1.79	0.0113	1.00
ENSG00000149257	SERPINH1	serpin peptidase inhibitor, clade H (heat shock protein 47), member 1, (collagen binding protein 1)	1.78	0.0050	1.00
ENSG00000227507	LTB	lymphotoxin beta (TNF superfamily, member 3)	1.77	0.0268	1.00
ENSG00000167333	TRIM68	tripartite motif containing 68	1.77	0.0285	1.00
ENSG00000186160	CYP4Z1	cytochrome P450, family 4, subfamily Z, polypeptide 1	1.76	0.0080	1.00
ENSG00000106078	COBL	cordon-bleu WH2 repeat protein	1.76	0.0497	1.00
ENSG00000010165	METTL13	methyltransferase like 13	1.75	0.0138	1.00
ENSG00000132763	MMACHC	methylmalonic aciduria (cobalamin deficiency) cblC type, with homocystinuria	1.74	0.0399	1.00
ENSG00000150656	CNDP1	carnosine dipeptidase 1 (metallopeptidase M20 family)	1.72	0.0417	1.00
ENSG00000093072	CECR1	cat eye syndrome chromosome region, candidate 1	1.71	0.0083	1.00
ENSG00000057935	MTA3	metastasis associated 1 family, member 3	1.71	0.0326	1.00
ENSG00000185880	TRIM69	tripartite motif containing 69	1.71	0.0104	1.00
ENSG00000049089	COL9A2	collagen, type IX, alpha 2	1.71	0.0297	1.00
ENSG00000197847	SLC22A20	solute carrier family 22, member 20	1.70	0.0368	1.00
ENSG00000111261	MANSC1	MANSC domain containing 1	1.69	0.0270	1.00
ENSG00000125630	POLR1B	polymerase (RNA) I polypeptide B, 128kDa	1.69	0.0149	1.00
ENSG00000128944	KNSTRN	kinetochore-localized astrin/SPAG5 binding protein	1.69	0.0213	1.00
ENSG00000181392	SYNE4	spectrin repeat containing, nuclear envelope family member 4	1.69	0.0450	1.00
ENSG00000171806	METTL18	methyltransferase like 18	1.69	0.0401	1.00
ENSG00000107104	KANK1	KN motif and ankyrin repeat domains 1	1.68	0.0063	1.00
ENSG00000183625	CCR3	chemokine (C-C motif) receptor 3	1.68	0.0495	1.00
ENSG00000108797	CNTNAP1	contactin associated protein 1	1.67	0.0219	1.00
ENSG00000114993	RTKN	rhotekin	1.67	0.0043	1.00
ENSG00000142583	SLC2A5	solute carrier family 2 (facilitated glucose/fructose transporter), member 5	1.67	0.0023	1.00
ENSG00000076641	PAG1	phosphoprotein membrane anchor with glycosphingolipid microdomains 1	1.66	0.0495	1.00

ENSG00000113140	SPARC	secreted protein, acidic, cysteine-rich (osteonectin)	1.66	0.0383	1.00
ENSG00000254685	FPGT	fucose-1-phosphate guanylyltransferase	1.65	0.0267	1.00
ENSG00000105369	CD79A	CD79a molecule, immunoglobulin-associated alpha	1.65	0.0144	1.00
ENSG00000198743	SLC5A3	solute carrier family 5 (sodium/myo-inositol cotransporter), member 3	1.64	0.0210	1.00
ENSG00000048028	USP28	ubiquitin specific peptidase 28	1.64	0.0326	1.00
ENSG00000109684	CLNK	cytokine-dependent hematopoietic cell linker	1.60	0.0256	1.00
ENSG00000180245	RRH	retinal pigment epithelium-derived rhodopsin homolog	1.59	0.0336	1.00
ENSG00000184343	SRPK3	SRSF protein kinase 3	1.59	0.0223	1.00
ENSG00000213699	SLC35F6	solute carrier family 35, member F6	1.59	0.0380	1.00
ENSG00000116096	SPR	sepiapterin reductase (7,8-dihydrobiopterin:NADP+ oxidoreductase)	1.57	0.0259	1.00
ENSG00000174177	CTU2	cytosolic thiouridylase subunit 2 homolog (S. pombe)	1.57	0.0265	1.00
ENSG00000143457	GOLPH3L	golgi phosphoprotein 3-like	1.57	0.0236	1.00
ENSG00000151320	AKAP6	A kinase (PRKA) anchor protein 6	1.57	0.0361	1.00
ENSG00000171314	PGAM1	phosphoglycerate mutase 1 (brain)	1.57	0.0244	1.00
ENSG00000223501	VPS52	vacuolar protein sorting 52 homolog (S. cerevisiae)	1.56	0.0497	1.00
ENSG00000101076	HNF4A	hepatocyte nuclear factor 4, alpha	1.56	0.0447	1.00
ENSG00000081853	PCDHGA2	protocadherin gamma subfamily A, 2	1.56	0.0074	1.00
ENSG00000138413	IDH1	isocitrate dehydrogenase 1 (NADP+), soluble	1.56	0.0072	1.00
ENSG00000213937	CLDN9	claudin 9	1.55	0.0492	1.00
ENSG00000183850	ZNF730	zinc finger protein 730	1.55	0.0261	1.00
ENSG00000125484	GTF3C4	general transcription factor IIIC, polypeptide 4, 90kDa	1.55	0.0069	1.00
ENSG00000120658	ENOX1	ecto-NOX disulfide-thiol exchanger 1	1.55	0.0391	1.00
ENSG00000185477	GPRIN3	GPRIN family member 3	1.54	0.0377	1.00
ENSG00000214782	MS4A18	membrane-spanning 4-domains, subfamily A, member 18	1.54	0.0226	1.00
ENSG00000166971	AKTIP	AKT interacting protein	1.54	0.0288	1.00
ENSG00000198610	AKR1C4	aldo-keto reductase family 1, member C4	1.53	0.0390	1.00
ENSG00000184162	NR2C2AP	nuclear receptor 2C2-associated protein	1.53	0.0474	1.00
ENSG00000159618	ADGRG5	adhesion G protein-coupled receptor G5	1.53	0.0414	1.00
ENSG00000221986	MYBPHL	myosin binding protein H-like	1.53	0.0420	1.00
ENSG00000139344	AMDHD1	amidohydrolase domain containing 1	1.53	0.0140	1.00
ENSG00000018869	ZNF582	zinc finger protein 582	1.52	0.0055	1.00
ENSG00000178038	ALS2CL	ALS2 C-terminal like	1.51	0.0433	1.00
ENSG00000001460	STPG1	sperm-tail PG-rich repeat containing 1	1.51	0.0139	1.00
ENSG00000188868	ZNF563	zinc finger protein 563	1.51	0.0342	1.00

ENSG00000173930	SLCO4C1	solute carrier organic anion transporter family, member 4C1	1.51	0.0426	1.00
ENSG00000238243	OR2W3	olfactory receptor, family 2, subfamily W, member 3	1.50	0.0490	1.00
ENSG00000179855	GIPC3	GIPC PDZ domain containing family, member 3	1.50	0.0419	1.00
ENSG00000197312	DDI2	DNA-damage inducible 1 homolog 2	1.50	0.0365	1.00
ENSG00000175544	CABP4	calcium binding protein 4	1.50	0.0487	1.00
ENSG00000176222	ZNF404	zinc finger protein 404	1.49	0.0398	1.00
ENSG00000144504	ANKMY1	ankyrin repeat and MYND domain containing 1	1.48	0.0235	1.00
ENSG00000137124	ALDH1B1	aldehyde dehydrogenase 1 family, member B1	1.48	0.0285	1.00
ENSG00000188321	ZNF559	zinc finger protein 559	1.47	0.0375	1.00
ENSG00000173088	C10orf131	chromosome 10 open reading frame 131	1.47	0.0272	1.00
ENSG00000244242	IFITM10	interferon induced transmembrane protein 10	1.46	0.0266	1.00
ENSG00000138459	SLC35A5	solute carrier family 35, member A5	1.46	0.0259	1.00
ENSG00000114771	AADAC	arylacetamide deacetylase	1.45	0.0489	1.00
ENSG00000164318	EGFLAM	EGF-like, fibronectin type III and laminin G domains	1.45	0.0450	1.00
ENSG00000160446	ZDHHC12	zinc finger, DHHC-type containing 12	1.45	0.0426	1.00
ENSG00000100726	TELO2	telomere maintenance 2	1.44	0.0379	1.00
ENSG00000167625	ZNF526	zinc finger protein 526	1.43	0.0406	1.00
ENSG00000245680	ZNF585B	zinc finger protein 585B	1.42	0.0081	1.00
ENSG00000113966	ARL6	ADP-ribosylation factor-like 6	1.42	0.0460	1.00
ENSG00000118894	EEF2KMT	eukaryotic elongation factor 2 lysine methyltransferase	1.41	0.0407	1.00
ENSG00000124786	SLC35B3	solute carrier family 35 (adenosine 3'-phospho 5'-phosphosulfate transporter), member B3	1.40	0.0187	1.00
ENSG00000112578	BYSL	bystin-like	1.40	0.0204	1.00
ENSG00000168778	TCTN2	tectonic family member 2	1.40	0.0466	1.00
ENSG00000149716	ORAOV1	oral cancer overexpressed 1	1.39	0.0197	1.00
ENSG00000134864	GGACT	gamma-glutamylamine cyclotransferase	1.39	0.0278	1.00
ENSG00000101265	RASSF2	Ras association (RalGDS/AF-6) domain family member 2	1.39	0.0208	1.00
ENSG00000168675	LDLRAD4	low density lipoprotein receptor class A domain containing 4	1.38	0.0199	1.00
ENSG00000166908	PIP4K2C	phosphatidylinositol-5-phosphate 4-kinase, type II, gamma	1.38	0.0188	1.00
ENSG00000106852	LHX6	LIM homeobox 6	1.37	0.0203	1.00
ENSG00000080839	RBL1	retinoblastoma-like 1	1.37	0.0421	1.00
ENSG00000133111	RFXAP	regulatory factor X-associated protein	1.36	0.0316	1.00
ENSG00000185252	ZNF74	zinc finger protein 74	1.36	0.0269	1.00
ENSG00000178921	PFAS	phosphoribosylformylglycinamidine synthase	1.33	0.0143	1.00

ENSG00000111728	ST8SIA1	ST8 alpha-N-acetyl-neuraminide alpha-2,8-sialyltransferase 1	1.31	0.0289	1.00
ENSG00000215915	ATAD3C	ATPase family, AAA domain containing 3C	1.31	0.0219	1.00
ENSG00000169242	EFNA1	ephrin-A1	1.31	0.0470	1.00
ENSG00000127423	AUNIP	aurora kinase A and ninein interacting protein	1.31	0.0306	1.00
ENSG00000272602	ZNF595	zinc finger protein 595	1.30	0.0464	1.00
ENSG00000197134	ZNF257	zinc finger protein 257	1.30	0.0266	1.00
ENSG00000154114	TBCEL	tubulin folding cofactor E-like	1.30	0.0167	1.00
ENSG00000112983	BRD8	bromodomain containing 8	1.30	0.0332	1.00
ENSG00000105409	ATP1A3	ATPase, Na+/K+ transporting, alpha 3 polypeptide	1.29	0.0414	1.00
ENSG00000167384	ZNF180	zinc finger protein 180	1.29	0.0292	1.00
ENSG00000196418	ZNF124	zinc finger protein 124	1.28	0.0359	1.00
ENSG00000103356	EARS2	glutamyl-tRNA synthetase 2, mitochondrial	1.28	0.0231	1.00
ENSG00000196437	ZNF569	zinc finger protein 569	1.28	0.0147	1.00
ENSG00000140025	EFCAB11	EF-hand calcium binding domain 11	1.28	0.0318	1.00
ENSG00000120694	HSPH1	heat shock 105kDa/110kDa protein 1	1.27	0.0493	1.00
ENSG00000154133	ROBO4	roundabout guidance receptor 4	1.27	0.0422	1.00
ENSG00000121075	TBX4	T-box 4	1.27	0.0258	1.00
ENSG00000134057	CCNB1	cyclin B1	1.27	0.0482	1.00
ENSG00000065054	SLC9A3R2	solute carrier family 9, subfamily A (NHE3, cation proton antiporter 3), member 3 regulator 2	1.26	0.0474	1.00
ENSG00000262209	PCDHGB3	protocadherin gamma subfamily B, 3	1.26	0.0393	1.00
ENSG00000097046	CDC7	cell division cycle 7	1.26	0.0285	1.00
ENSG00000169398	PTK2	protein tyrosine kinase 2	1.26	0.0076	1.00
ENSG00000163959	SLC51A	solute carrier family 51, alpha subunit	1.26	0.0337	1.00
ENSG00000167315	ACAA2	acetyl-CoA acyltransferase 2	1.26	0.0283	1.00
ENSG00000204428	LY6G5C	lymphocyte antigen 6 complex, locus G5C	1.26	0.0388	1.00
ENSG00000150433	TMEM218	transmembrane protein 218	1.25	0.0270	1.00
ENSG00000146021	KLHL3	kelch-like family member 3	1.25	0.0156	1.00
ENSG00000230510	PPP5D1	PPP5 tetratricopeptide repeat domain containing 1	1.25	0.0397	1.00
ENSG00000178773	CPNE7	copine VII	1.24	0.0391	1.00
ENSG00000147789	ZNF7	zinc finger protein 7	1.24	0.0313	1.00
ENSG00000180104	EXOC3	exocyst complex component 3	1.23	0.0334	1.00
ENSG00000134285	FKBP11	FK506 binding protein 11, 19 kDa	1.23	0.0118	1.00
ENSG00000133805	AMPD3	adenosine monophosphate deaminase 3	1.23	0.0124	1.00
ENSG00000130635	COL5A1	collagen, type V, alpha 1	1.22	0.0232	1.00

ENSG00000168564	CDKN2AIP	CDKN2A interacting protein	1.21	0.0282	1.00
ENSG00000004468	CD38	CD38 molecule	1.20	0.0393	1.00
ENSG00000134255	CEPT1	choline/ethanolamine phosphotransferase 1	1.20	0.0147	1.00
ENSG00000140263	SORD	sorbitol dehydrogenase	1.20	0.0228	1.00

Table App 3-2 Genes upregulated in HSCs from AML samples versus controls (Fold change >2; p<0.05)

ID	Gene	Description	Log Fold Change	P.Value	adj.P. Val
ENSG00000160201	U2AF1	U2 small nuclear RNA auxiliary factor 1	3.42	0.0018	1.00
ENSG00000180596	HIST1H2BC	histone cluster 1, H2bc	3.37	0.0122	1.00
ENSG00000272196	HIST2H2AA4	histone cluster 2, H2aa4	3.14	0.0271	1.00
ENSG00000276903	HIST1H2AL	histone cluster 1, H2al	3.02	0.0063	1.00
ENSG00000088826	SMOX	spermine oxidase	2.99	0.0033	1.00
ENSG00000235961	PNMA6A	paraneoplastic Ma antigen family member 6A	2.94	0.0037	1.00
ENSG00000180611	MB21D2	Mab-21 domain containing 2	2.88	0.0004	1.00
ENSG00000274997	HIST1H2AH	histone cluster 1, H2ah	2.69	0.0061	1.00
ENSG00000122824	NUDT10	nudix (nucleoside diphosphate linked moiety X)-type motif 10	2.64	0.0188	1.00
ENSG00000182195	LDOC1	leucine zipper, down-regulated in cancer 1	2.62	0.0424	1.00
ENSG00000197747	S100A10	S100 calcium binding protein A10	2.61	0.0015	1.00
ENSG00000276410	HIST1H2BB	histone cluster 1, H2bb	2.48	0.0469	1.00
ENSG00000119138	KLF9	Kruppel-like factor 9	2.43	0.0160	1.00
ENSG00000165030	NFIL3	nuclear factor, interleukin 3 regulated	2.43	0.0057	1.00
ENSG00000146592	CREB5	cAMP responsive element binding protein 5	2.43	0.0028	1.00
ENSG00000134594	RAB33A	RAB33A, member RAS oncogene family	2.43	0.0294	1.00
ENSG00000165730	STOX1	storkhead box 1	2.40	0.0063	1.00
ENSG00000184206	GOLGA6L4	golgin A6 family-like 4	2.39	0.0147	1.00
ENSG00000178127	NDUFV2	NADH dehydrogenase (ubiquinone) flavoprotein 2, 24kDa	2.33	0.0070	1.00
ENSG00000111405	ENDOU	endonuclease, polyU-specific	2.32	0.0114	1.00
ENSG00000187837	HIST1H1C	histone cluster 1, H1c	2.32	0.0041	1.00
ENSG00000213977	TAX1BP3	Tax1 (human T-cell leukemia virus type I) binding protein 3	2.29	0.0180	1.00
ENSG00000180573	HIST1H2AC	histone cluster 1, H2a	2.27	0.0400	1.00
ENSG00000170379	TCAF2	TRPM8 channel-associated factor 2	2.27	0.0002	1.00
ENSG00000172733	PURG	purine-rich element binding protein G	2.26	0.0330	1.00
ENSG00000259511	UBE2Q2L	ubiquitin-conjugating enzyme E2Q family member 2-like	2.26	0.0081	1.00
ENSG00000168298	HIST1H1E	histone cluster 1, H1e	2.21	0.0139	1.00
ENSG00000278588	HIST1H2BI	histone cluster 1, H2bi	2.20	0.0368	1.00
ENSG00000169429	CXCL8	chemokine (C-X-C motif) ligand 8	2.18	0.0248	1.00
ENSG00000124575	HIST1H1D	histone cluster 1, H1d	2.18	0.0199	1.00
ENSG00000006327	TNFRSF12A	tumor necrosis factor receptor superfamily, member 12A	2.18	0.0232	1.00
ENSG00000196368	NUDT11	nudix (nucleoside diphosphate linked moiety X)-type motif 11	2.17	0.0226	1.00

ENSG00000006042	TMEM98	transmembrane protein 98	2.17	0.0494	1.00
ENSG00000076604	TRAF4	TNF receptor-associated factor 4	2.13	0.0168	1.00
ENSG00000022267	FHL1	four and a half LIM domains 1	2.11	0.0381	1.00
ENSG00000123095	BHLHE41	basic helix-loop-helix family, member e41	2.09	0.0246	1.00
ENSG00000273802	HIST1H2BG	histone cluster 1, H2bg	2.09	0.0091	1.00
ENSG00000143367	TUFT1	tuftelin 1	2.08	0.0155	1.00
ENSG00000131981	LGALS3	lectin, galactoside-binding, soluble, 3	2.07	0.0486	1.00
ENSG00000060558	GNA15	guanine nucleotide binding protein (G protein), alpha 15 (Gq class)	2.06	0.0056	1.00
ENSG00000196110	ZNF699	zinc finger protein 699	2.06	0.0098	1.00
ENSG00000254122	PCDHGB7	protocadherin gamma subfamily B, 7	2.05	0.0009	1.00
ENSG00000165389	SPTSSA	serine palmitoyltransferase, small subunit A	2.05	0.0162	1.00
ENSG00000109686	SH3D19	SH3 domain containing 19	2.05	0.0115	1.00
ENSG00000256188	TAS2R30	taste receptor, type 2, member 30	2.04	0.0183	1.00
ENSG00000000003	TSPAN6	tetraspanin 6	2.03	0.0197	1.00
ENSG00000143839	REN	renin	2.03	0.0422	1.00
ENSG00000085514	PILRA	paired immunoglobin-like type 2 receptor alpha	2.02	0.0019	1.00
ENSG00000197329	PELI1	pellino E3 ubiquitin protein ligase 1	2.02	0.0056	1.00
ENSG00000140478	GOLGA6D	golgin A6 family, member D	2.00	0.0027	1.00
ENSG00000172005	MAL	mal, T-cell differentiation protein	2.00	0.0354	1.00
ENSG00000197238	HIST1H4J	histone cluster 1, H4j	1.99	0.0038	1.00
ENSG00000133661	SFTPD	surfactant protein D	1.99	0.0335	1.00
ENSG00000185775	SPATA31A6	SPATA31 subfamily A, member 6	1.98	0.0187	1.00
ENSG00000115828	QPCT	glutaminyl-peptide cyclotransferase	1.97	0.0043	1.00
ENSG00000102359	SRPX2	sushi-repeat containing protein, X-linked 2	1.97	0.0352	1.00
ENSG00000164674	SYTL3	synaptotagmin-like 3	1.95	0.0085	1.00
ENSG00000146072	TNFRSF21	tumor necrosis factor receptor superfamily, member 21	1.94	0.0218	1.00
ENSG00000231824	C18orf42	chromosome 18 open reading frame 42	1.94	0.0078	1.00
ENSG00000203907	OOEP	oocyte expressed protein	1.93	0.0088	1.00
ENSG00000173334	TRIB1	tribbles pseudokinase 1	1.92	0.0021	1.00
ENSG00000133131	MORC4	MORC family CW-type zinc finger 4	1.92	0.0041	1.00
ENSG00000135226	UGT2B28	UDP glucuronosyltransferase 2 family, polypeptide B28	1.91	0.0167	1.00
ENSG00000260691	ANKRD20A1	ankyrin repeat domain 20 family, member A1	1.90	0.0425	1.00
ENSG00000179021	C3orf38	chromosome 3 open reading frame 38	1.89	0.0245	1.00
ENSG00000184205	TSPYL2	TSPY-like 2	1.89	0.0202	1.00

ENSG00000115540	MOB4	MOB family member 4, phocein	1.88	0.0146	1.00
ENSG00000123091	RNF11	ring finger protein 11	1.87	0.0192	1.00
ENSG00000105708	ZNF14	zinc finger protein 14	1.84	0.0080	1.00
ENSG00000213185	FAM24B	family with sequence similarity 24, member B	1.84	0.0151	1.00
ENSG00000198553	KCNRG	potassium channel regulator	1.84	0.0353	1.00
ENSG00000172058	SERF1A	small EDRK-rich factor 1A (telomeric)	1.83	0.0387	1.00
ENSG00000277443	MARCKS	myristoylated alanine-rich protein kinase C substrate	1.83	0.0045	1.00
ENSG00000147180	ZNF711	zinc finger protein 711	1.82	0.0362	1.00
ENSG00000166171	DPCD	deleted in primary ciliary dyskinesia homolog (mouse)	1.82	0.0367	1.00
ENSG00000204538	PSORS1C2	psoriasis susceptibility 1 candidate 2	1.81	0.0015	1.00
ENSG00000198947	DMD	dystrophin	1.81	0.0052	1.00
ENSG00000108187	PBLD	phenazine biosynthesis-like protein domain containing	1.81	0.0323	1.00
ENSG00000174945	AMZ1	archaelysin family metallopeptidase 1	1.79	0.0350	1.00
ENSG00000196866	HIST1H2AD	histone cluster 1, H2ad	1.79	0.0372	1.00
ENSG00000179833	SERTAD2	SERTA domain containing 2	1.78	0.0151	1.00
ENSG00000112245	PTP4A1	protein tyrosine phosphatase type IVA, member 1	1.78	0.0037	1.00
ENSG00000127666	TICAM1	toll-like receptor adaptor molecule 1	1.78	0.0392	1.00
ENSG00000163734	CXCL3	chemokine (C-X-C motif) ligand 3	1.77	0.0494	1.00
ENSG00000072133	RPS6KA6	ribosomal protein S6 kinase, 90kDa, polypeptide 6	1.77	0.0175	1.00
ENSG00000189060	H1F0	H1 histone family, member 0	1.77	0.0147	1.00
ENSG00000127252	HRASLS	HRAS-like suppressor	1.77	0.0473	1.00
ENSG00000184897	H1FX	H1 histone family, member X	1.76	0.0093	1.00
ENSG00000133818	RRAS2	related RAS viral (r-ras) oncogene homolog 2	1.74	0.0258	1.00
ENSG00000101470	TNNC2	troponin C type 2 (fast)	1.74	0.0412	1.00
ENSG00000163661	PTX3	pentraxin 3, long	1.74	0.0293	1.00
ENSG00000214274	ANG	angiogenin, ribonuclease, RNase A family, 5	1.71	0.0484	1.00
ENSG00000154734	ADAMTS1	ADAM metallopeptidase with thrombospondin type 1 motif, 1	1.70	0.0194	1.00
ENSG00000159212	CLIC6	chloride intracellular channel 6	1.70	0.0311	1.00
ENSG00000171044	XKR6	XK, Kell blood group complex subunit-related family, member 6	1.69	0.0340	1.00
ENSG00000105894	PTN	pleiotrophin	1.69	0.0497	1.00
ENSG00000185272	RBM11	RNA binding motif protein 11	1.68	0.0419	1.00
ENSG00000162599	NFIA	nuclear factor I/A	1.68	0.0400	1.00
ENSG00000187627	RGPD1	RANBP2-like and GRIP domain containing 1	1.67	0.0137	1.00
ENSG00000205078	SYCE1L	synaptonemal complex central element protein 1-like	1.66	0.0249	1.00

ENSG00000183463	URAD	ureidoimidazoline (2-oxo-4-hydroxy-4-carboxy-5-) decarboxylase	1.66	0.0368	1.00
ENSG00000256646	PSMA2	Uncharacterized protein	1.65	0.0042	1.00
ENSG00000196470	SIAH1	siah E3 ubiquitin protein ligase 1	1.65	0.0255	1.00
ENSG00000274641	HIST1H2BO	histone cluster 1, H2bo	1.65	0.0475	1.00
ENSG00000198523	PLN	phospholamban	1.65	0.0429	1.00
ENSG00000075618	FSCN1	fascin actin-bundling protein 1	1.65	0.0210	1.00
ENSG00000134463	ECHDC3	enoyl CoA hydratase domain containing 3	1.64	0.0353	1.00
ENSG00000189129	PLAC9	placenta-specific 9	1.64	0.0379	1.00
ENSG00000144935	TRPC1	transient receptor potential cation channel, subfamily C, member 1	1.64	0.0081	1.00
ENSG00000069956	MAPK6	mitogen-activated protein kinase 6	1.63	0.0100	1.00
ENSG00000047634	SCML1	sex comb on midleg-like 1 (Drosophila)	1.63	0.0172	1.00
ENSG00000117407	ARTN	artemin	1.63	0.0449	1.00
ENSG00000251655	PRB1	proline-rich protein BstNI subfamily 1	1.63	0.0413	1.00
ENSG00000141574	SECTM1	secreted and transmembrane 1	1.63	0.0138	1.00
ENSG00000105327	BBC3	BCL2 binding component 3	1.63	0.0459	1.00
ENSG00000132141	CCT6B	chaperonin containing TCP1, subunit 6B (zeta 2)	1.62	0.0171	1.00
ENSG00000131724	IL13RA1	interleukin 13 receptor, alpha 1	1.62	0.0436	1.00
ENSG00000178934	LGALS7B	lectin, galactoside-binding, soluble, 7B	1.61	0.0158	1.00
ENSG00000068354	TBC1D25	TBC1 domain family, member 25	1.61	0.0145	1.00
ENSG00000136840	ST6GALNAC4	ST6 (alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3)-N-acetylgalactosaminide alpha-2,6-sialyltransferase 4	1.61	0.0196	1.00
ENSG00000181588	MEX3D	mex-3 RNA binding family member D	1.60	0.0180	1.00
ENSG00000175352	NRIP3	nuclear receptor interacting protein 3	1.59	0.0454	1.00
ENSG00000029993	HMGB3	high mobility group box 3	1.59	0.0142	1.00
ENSG00000117479	SLC19A2	solute carrier family 19 (thiamine transporter), member 2	1.58	0.0174	1.00
ENSG00000143845	ETNK2	ethanolamine kinase 2	1.58	0.0025	1.00
ENSG00000104147	OIP5	Opa interacting protein 5	1.58	0.0182	1.00
ENSG00000224186	C5orf66	chromosome 5 open reading frame 66	1.58	0.0254	1.00
ENSG00000206549	PRSS50	protease, serine, 50	1.57	0.0051	1.00
ENSG00000142549	IGLON5	IgLON family member 5	1.57	0.0198	1.00
ENSG00000151364	KCTD14	potassium channel tetramerization domain containing 14	1.57	0.0089	1.00
ENSG00000099194	SCD	stearoyl-CoA desaturase (delta-9-desaturase)	1.57	0.0110	1.00
ENSG00000196369	SRGAP2B	SLIT-ROBO Rho GTPase activating protein 2B	1.56	0.0154	1.00
ENSG00000139725	RHOF	ras homolog family member F (in filopodia)	1.56	0.0240	1.00
ENSG00000125249	RAP2A	RAP2A, member of RAS oncogene family	1.56	0.0305	1.00

ENSG00000184785	SMIM10	small integral membrane protein 10	1.55	0.0177	1.00
ENSG00000173166	RAPH1	Ras association (RalGDS/AF-6) and pleckstrin homology domains 1	1.55	0.0102	1.00
ENSG00000110108	TMEM109	transmembrane protein 109	1.55	0.0481	1.00
ENSG00000124103	FAM209A	family with sequence similarity 209, member A	1.55	0.0385	1.00
ENSG00000101400	SNTA1	syntrophin, alpha 1	1.54	0.0118	1.00
ENSG00000135272	MDFIC	MyoD family inhibitor domain containing	1.54	0.0470	1.00
ENSG00000147059	SPIN2A	spindlin family, member 2A	1.54	0.0429	1.00
ENSG00000196476	C20orf96	chromosome 20 open reading frame 96	1.54	0.0193	1.00
ENSG00000175592	FOSL1	FOS-like antigen 1	1.54	0.0222	1.00
ENSG00000078246	TULP3	tubby like protein 3	1.54	0.0105	1.00
ENSG00000169957	ZNF768	zinc finger protein 768	1.53	0.0180	1.00
ENSG00000213085	CFAP45	cilia and flagella associated protein 45	1.53	0.0457	1.00
ENSG00000108960	MMD	monocyte to macrophage differentiation-associated	1.52	0.0074	1.00
ENSG00000169991	IFFO2	intermediate filament family orphan 2	1.51	0.0156	1.00
ENSG00000186272	ZNF17	zinc finger protein 17	1.51	0.0274	1.00
ENSG00000268447	SSX2B	synovial sarcoma, X breakpoint 2B	1.51	0.0120	1.00
ENSG00000228741	SPATA13	spermatogenesis associated 13	1.50	0.0141	1.00
ENSG00000232653	GOLGA8N	golgin A8 family, member N	1.50	0.0370	1.00
ENSG00000196363	WDR5	WD repeat domain 5	1.50	0.0353	1.00
ENSG00000139433	GLTP	glycolipid transfer protein	1.49	0.0060	1.00
ENSG00000139112	GABARAPL1	GABA(A) receptor-associated protein like 1	1.49	0.0149	1.00
ENSG00000173610	UGT2A1	UDP glucuronosyltransferase 2 family, polypeptide A1, complex locus	1.49	0.0109	1.00
ENSG00000170899	GSTA4	glutathione S-transferase alpha 4	1.49	0.0344	1.00
ENSG00000143971	ETAA1	Ewing tumor-associated antigen 1	1.48	0.0225	1.00
ENSG00000276203	ANKRD20A3	ankyrin repeat domain 20 family, member A3	1.47	0.0368	1.00
ENSG00000170262	MRAP	melanocortin 2 receptor accessory protein	1.47	0.0213	1.00
ENSG00000120833	SOCS2	suppressor of cytokine signaling 2	1.47	0.0180	1.00
ENSG00000205572	SERF1B	small EDRK-rich factor 1B (centromeric)	1.47	0.0365	1.00
ENSG00000132879	FBXO44	F-box protein 44	1.46	0.0122	1.00
ENSG00000079739	PGM1	phosphoglucomutase 1	1.45	0.0365	1.00
ENSG00000102109	PCSK1N	proprotein convertase subtilisin/kexin type 1 inhibitor	1.45	0.0341	1.00
ENSG00000166526	ZNF3	zinc finger protein 3	1.45	0.0102	1.00
ENSG00000170340	B3GNT2	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 2	1.44	0.0406	1.00
ENSG00000168811	IL12A	interleukin 12A	1.43	0.0431	1.00

ENSG00000258643	BCL2L2-PABPN1	BCL2L2-PABPN1 readthrough	1.42	0.0203	1.00
ENSG00000164543	STK17A	serine/threonine kinase 17a	1.42	0.0112	1.00
ENSG00000215790	SLC35E2	solute carrier family 35, member E2	1.42	0.0461	1.00
ENSG00000119280	C1orf198	chromosome 1 open reading frame 198	1.41	0.0264	1.00
ENSG00000155918	RAET1L	retinoic acid early transcript 1L	1.41	0.0172	1.00
ENSG00000108001	EBF3	early B-cell factor 3	1.41	0.0236	1.00
ENSG00000180354	MTURN	maturin, neural progenitor differentiation regulator homolog (Xenopus)	1.41	0.0135	1.00
ENSG00000116035	VAX2	ventral anterior homeobox 2	1.40	0.0211	1.00
ENSG00000271303	SRXN1	sulfiredoxin 1	1.40	0.0248	1.00
ENSG00000179562	GCC1	GRIP and coiled-coil domain containing 1	1.40	0.0308	1.00
ENSG00000157557	ETS2	v-ets avian erythroblastosis virus E26 oncogene homolog 2	1.40	0.0277	1.00
ENSG00000197050	ZNF420	zinc finger protein 420	1.40	0.0204	1.00
ENSG00000273899	NOL12	nucleolar protein 12	1.39	0.0438	1.00
ENSG00000122042	UBL3	ubiquitin-like 3	1.39	0.0448	1.00
ENSG00000151468	CCDC3	coiled-coil domain containing 3	1.38	0.0313	1.00
ENSG00000163659	TIPARP	TCDD-inducible poly(ADP-ribose) polymerase	1.37	0.0361	1.00
ENSG00000248333	CDK11B	cyclin-dependent kinase 11B	1.37	0.0382	1.00
ENSG00000160683	CXCR5	chemokine (C-X-C motif) receptor 5	1.37	0.0093	1.00
ENSG00000180228	PRKRA	protein kinase, interferon-inducible double stranded RNA dependent activator	1.37	0.0264	1.00
ENSG00000204314	PRRT1	proline-rich transmembrane protein 1	1.36	0.0434	1.00
ENSG00000059769	DNAJC25	DnaJ (Hsp40) homolog, subfamily C , member 25	1.36	0.0487	1.00
ENSG00000148459	PDSS1	prenyl (decaprenyl) diphosphate synthase, subunit 1	1.35	0.0129	1.00
ENSG00000132122	SPATA6	spermatogenesis associated 6	1.34	0.0274	1.00
ENSG00000206418	RAB12	RAB12, member RAS oncogene family	1.33	0.0452	1.00
ENSG00000111266	DUSP16	dual specificity phosphatase 16	1.33	0.0093	1.00
ENSG00000162496	DHRS3	dehydrogenase/reductase (SDR family) member 3	1.33	0.0427	1.00
ENSG00000184792	OSBP2	oxysterol binding protein 2	1.33	0.0050	1.00
ENSG00000146535	GNA12	guanine nucleotide binding protein (G protein) alpha 12	1.33	0.0363	1.00
ENSG00000175325	PROP1	PROP paired-like homeobox 1	1.33	0.0494	1.00
ENSG00000145365	TIFA	TRAF-interacting protein with forkhead-associated domain	1.33	0.0247	1.00
ENSG00000169297	NR0B1	nuclear receptor subfamily 0, group B, member 1	1.33	0.0486	1.00
ENSG00000138032	PPM1B	protein phosphatase, Mg2+/Mn2+ dependent, 1B	1.32	0.0198	1.00

Appendix 4: Genes with significantly different expression between AML-exposed and Control HSCs, as detected by RNA-Seq Attempt 2

Table App 4 -1 Genes upregulated in HSCs from Controls versus AML (Fold change >2, p<0.01)

ID	Gene	Description	logFC	P.Value	adj.P. Val
ENSG00000169877	AHSP	alpha hemoglobin stabilizing protein	8.47	0.0002	0.17
ENSG00000055118	KCNH2	potassium channel, voltage gated eag related subfamily H, member 2	8.20	0.0002	0.19
ENSG00000133742	CA1	carbonic anhydrase I	8.14	0.0005	0.22
ENSG00000158578	ALAS2	5'-aminolevulinate synthase 2	7.52	0.0007	0.26
ENSG00000155961	RAB39B	RAB39B, member RAS oncogene family	7.07	0.0001	0.11
ENSG00000136153	LMO7	LIM domain 7	6.93	0.0001	0.12
ENSG00000101057	MYBL2	v-myb avian myeloblastosis viral oncogene homolog-like 2	6.93	0.0008	0.27
ENSG00000029534	ANK1	ankyrin 1	6.93	0.0008	0.27
ENSG00000093009	CDC45	cell division cycle 45	6.90	0.0001	0.12
ENSG00000004939	SLC4A1	solute carrier family 4 (anion exchanger), member 1 (Diego blood group)	6.57	0.0015	0.33
ENSG00000010539	ZNF200	zinc finger protein 200	6.54	0.0003	0.20
ENSG00000090932	DLL3	delta-like 3 (Drosophila)	6.51	0.0011	0.30
ENSG00000216775	RP1-152L7.5		6.48	0.0013	0.33
ENSG00000211898	IGHD	immunoglobulin heavy constant delta	6.47	0.0004	0.22
ENSG00000273014	RP11-225B17.2		6.45	0.0009	0.29
ENSG00000113552	GNPDA1	glucosamine-6-phosphate deaminase 1	6.43	0.0000	0.01
ENSG00000131747	TOP2A	topoisomerase (DNA) II alpha	6.42	0.0005	0.22
ENSG00000119599	DCAF4	DDB1 and CUL4 associated factor 4	6.36	0.0005	0.22
ENSG00000244734	HBB	hemoglobin subunit beta	6.34	0.0091	0.55
ENSG00000156876	SASS6	SAS-6 centriolar assembly protein	6.27	0.0003	0.19
ENSG00000168754	FAM178B	family with sequence similarity 178 member B	6.22	0.0086	0.55
ENSG00000183814	LIN9	lin-9 DREAM MuvB core complex component	6.22	0.0005	0.22
ENSG00000178999	AURKB	aurora kinase B	6.13	0.0009	0.29
ENSG00000134690	CDCA8	cell division cycle associated 8	6.12	0.0008	0.27
ENSG00000101412	E2F1	E2F transcription factor 1	6.12	0.0034	0.43
ENSG00000204644	ZFP57	ZFP57 zinc finger protein	6.06	0.0083	0.55
ENSG00000114988	LMAN2L	lectin, mannose-binding 2-like	6.00	0.0068	0.53
ENSG00000170558	CDH2	cadherin 2, type 1, N-cadherin (neuronal)	5.95	0.0034	0.43
ENSG00000172247	C1QTNF4	C1q and tumor necrosis factor related protein 4	5.94	0.0008	0.28
ENSG00000137700	SLC37A4	solute carrier family 37 (glucose-6-phosphate transporter), member 4	5.93	0.0017	0.35
ENSG00000080819	CPOX	coproporphyrinogen oxidase	5.92	0.0032	0.43
ENSG00000243406	MRPS31P5	mitochondrial ribosomal protein S31 pseudogene 5	5.90	0.0001	0.12

ENSG00000149257	SERPINH1	serpin peptidase inhibitor, clade H (heat shock protein 47), member 1, (collagen binding protein 1)	5.89	0.0014	0.33
ENSG00000143627	PKLR	pyruvate kinase, liver and RBC	5.88	0.0022	0.42
ENSG00000119514	GALNT12	polypeptide N-acetylgalactosaminyltransferase 12	5.84	0.0002	0.19
ENSG00000138600	SPPL2A	signal peptide peptidase like 2A	5.83	0.0016	0.34
ENSG00000114374	USP9Y	ubiquitin specific peptidase 9, Y-linked	5.83	0.0072	0.54
ENSG00000144647	POMGNT2	protein O-linked mannose N-acetylglucosaminyltransferase 2 (beta 1,4-)	5.77	0.0027	0.42
ENSG00000133657	ATP13A3	ATPase type 13A3	5.77	0.0026	0.42
ENSG00000070182	SPTB	spectrin, beta, erythrocytic	5.75	0.0014	0.33
ENSG00000106351	AGFG2	ArfGAP with FG repeats 2	5.71	0.0003	0.20
ENSG00000160172	FAM86C2P	family with sequence similarity 86 member C2, pseudogene	5.70	0.0037	0.44
ENSG00000114315	HES1	hes family bHLH transcription factor 1	5.69	0.0029	0.42
ENSG00000097046	CDC7	cell division cycle 7	5.67	0.0063	0.52
ENSG00000273352	RP11-61L19.3		5.66	0.0025	0.42
ENSG00000215022	RP1-257A7.4		5.65	0.0062	0.52
ENSG00000144749	LRIG1	leucine-rich repeats and immunoglobulin-like domains 1	5.64	0.0025	0.42
ENSG00000143479	DYRK3	dual specificity tyrosine-(Y)-phosphorylation regulated kinase 3	5.64	0.0099	0.56
ENSG00000151413	NUBPL	nucleotide binding protein-like	5.64	0.0027	0.42
ENSG00000181450	ZNF678	zinc finger protein 678	5.62	0.0015	0.33
ENSG00000206172	HBA1	hemoglobin subunit alpha 1	5.60	0.0041	0.46
ENSG00000254837	AP001372.2		5.58	0.0051	0.49
ENSG00000113621	TXNDC15	thioredoxin domain containing 15	5.54	0.0014	0.33
ENSG00000124279	FASTKD3	FAST kinase domains 3	5.53	0.0025	0.42
ENSG00000196436	NPIPB15	nuclear pore complex interacting protein family member B15	5.46	0.0009	0.29
ENSG00000114739	ACVR2B	activin A receptor type IIB	5.45	0.0023	0.42
ENSG00000207475	SNORA80E	small nucleolar RNA, H/ACA box 80E	5.45	0.0002	0.19
ENSG00000169575	VPREB1	pre-B lymphocyte 1	5.44	0.0019	0.38
ENSG00000260018	RP11-505K9.1		5.43	0.0017	0.35
ENSG00000270885	RASL10B	RAS-like family 10 member B	5.42	0.0051	0.49
ENSG00000136514	RTP4	receptor (chemosensory) transporter protein 4	5.40	0.0076	0.54
ENSG00000182986	ZNF320	zinc finger protein 320	5.39	0.0003	0.20
ENSG00000148773	MKI67	marker of proliferation Ki-67	5.38	0.0068	0.53
ENSG00000272143	FGF14-AS2	FGF14 antisense RNA 2	5.37	0.0093	0.55
ENSG00000229666	MAST4-AS1	MAST4 antisense RNA 1	5.33	0.0030	0.42
ENSG00000172840	PDP2	pyruvate dehyrogenase phosphatase catalytic subunit 2	5.31	0.0027	0.42
ENSG00000135451	TROAP	trophinin associated protein	5.25	0.0051	0.49
ENSG00000080839	RBL1	retinoblastoma-like 1	5.24	0.0057	0.50
ENSG00000182118	FAM89A	family with sequence similarity 89 member A	5.24	0.0080	0.55
ENSG00000112077	RHAG	Rh-associated glycoprotein	5.23	0.0062	0.52
ENSG00000167513	CDT1	chromatin licensing and DNA replication factor 1	5.23	0.0066	0.53

ENSG00000228175	GEMIN8P4	gem nuclear organelle associated protein 8 pseudogene 4	5.23	0.0093	0.55
ENSG00000122824	NUDT10	nudix hydrolase 10	5.16	0.0030	0.42
ENSG00000159055	MIS18A	MIS18 kinetochore protein A	5.14	0.0030	0.42
ENSG00000101400	SNTA1	syntrophin, alpha 1	5.11	0.0089	0.55
ENSG00000245105	A2M-AS1	A2M antisense RNA 1 (head to head)	5.02	0.0074	0.54
ENSG00000252658	RNU6-786P	RNA, U6 small nuclear 786, pseudogene	5.00	0.0019	0.38
ENSG00000233974	RP11-823P9.3		4.99	0.0003	0.19
ENSG00000086506	HBQ1	hemoglobin subunit theta 1	4.97	0.0052	0.49
ENSG00000185010	F8	coagulation factor VIII, procoagulant component	4.92	0.0087	0.55
ENSG00000104221	BRF2	BRF2, RNA polymerase III transcription initiation factor 50 kDa subunit	4.92	0.0096	0.56
ENSG00000272654	RP11-422P24.11		4.91	0.0097	0.56
ENSG00000152382	TADA1	transcriptional adaptor 1	4.86	0.0042	0.46
ENSG00000213742	ZNF337-AS1	ZNF337 antisense RNA 1	4.85	0.0011	0.30
ENSG00000273226	RP11-513M16.8		4.85	0.0075	0.54
ENSG00000172167	MTBP	MDM2 binding protein	4.85	0.0076	0.54
ENSG00000137054	POLR1E	polymerase (RNA) I polypeptide E	4.84	0.0059	0.50
ENSG00000227632	AC018804.6		4.78	0.0034	0.43
ENSG00000145569	FAM105A	family with sequence similarity 105 member A	4.72	0.0049	0.49
ENSG00000163808	KIF15	kinesin family member 15	4.62	0.0088	0.55
ENSG00000135749	PCNXL2	pecanex-like 2 (Drosophila)	4.56	0.0031	0.43
ENSG00000165832	TRUB1	TruB pseudouridine (psi) synthase family member 1	4.53	0.0032	0.43
ENSG00000082516	GEMIN5	gem nuclear organelle associated protein 5	4.49	0.0054	0.49
ENSG00000280106	CTC-523E23.3		4.47	0.0049	0.49
ENSG00000166886	NAB2	NGFI-A binding protein 2 (EGR1 binding protein 2)	4.43	0.0091	0.55
ENSG00000066651	TRMT11	tRNA methyltransferase 11 homolog	4.42	0.0054	0.49
ENSG00000213222	AC093724.2		4.31	0.0006	0.25
ENSG00000227345	PARG	poly(ADP-ribose) glycohydrolase	4.28	0.0092	0.55
ENSG00000132964	CDK8	cyclin-dependent kinase 8	4.24	0.0085	0.55
ENSG00000147586	MRPS28	mitochondrial ribosomal protein S28	4.18	0.0080	0.55
ENSG00000129480	DTD2	D-tyrosyl-tRNA deacylase 2 (putative)	4.16	0.0093	0.55
ENSG00000177051	FBXO46	F-box protein 46	4.10	0.0078	0.54
ENSG00000065054	SLC9A3R2	solute carrier family 9, subfamily A (NHE3, cation proton antiporter 3), member 3 regulator 2	4.05	0.0046	0.48
ENSG00000270953	RP11-2E11.9		4.04	0.0041	0.46
ENSG00000267751	AC009005.2		3.93	0.0083	0.55
ENSG00000235036	RP5-1099D15.1		3.80	0.0005	0.22
ENSG00000185386	MAPK11	mitogen-activated protein kinase 11	3.59	0.0070	0.54
ENSG00000244921	MTCYBP18	MT-CYB pseudogene 18	3.54	0.0089	0.55
ENSG00000235065	RPL24P2	ribosomal protein L24 pseudogene 2	2.91	0.0098	0.56
ENSG00000233585	AC115617.2		2.89	0.0048	0.49

ENSG00000210196	MT-TP	mitochondrially encoded tRNA proline	2.80	0.0067	0.53
ENSG00000177683	THAP5	THAP domain containing 5	2.78	0.0085	0.55
ENSG00000140848	CPNE2	copine II	2.40	0.0095	0.55
ENSG00000253785	CTC-308K20.3		2.31	0.0030	0.42
ENSG00000197535	MYO5A	myosin VA	1.87	0.0076	0.54
ENSG00000240616	RPS6P25	ribosomal protein S6 pseudogene 25	1.80	0.0084	0.55
ENSG00000233225	AC004987.9		1.78	0.0068	0.53
ENSG00000277105	FP236383.10		1.73	0.0072	0.54
ENSG00000184220	CMSS1	cms1 ribosomal small subunit homolog (yeast)	1.70	0.0088	0.55
ENSG00000151353	TMEM18	transmembrane protein 18	1.54	0.0080	0.55

Table App 4-2 Genes upregulated in HSCs from AML samples versus controls (Fold change >2, p<0.01)

ID	Gene	Description	logFC	P.Value	adj.P. Val
ENSG00000182885	ADGRG3	adhesion G protein-coupled receptor G3	8.46	0.0000	0.00
ENSG00000225783	MIAT	myocardial infarction associated transcript (non-protein coding)	7.3	0.0000	0.02
ENSG00000148225	WDR31	WD repeat domain 31	7.04	0.0000	0.01
ENSG00000144026	ZNF514	zinc finger protein 514	6.94	0.0005	0.22
ENSG00000241014	RP11-244H3.1		6.91	0.0000	0.04
ENSG00000183722	LHFP	lipoma HMGIC fusion partner	6.73	0.0002	0.19
ENSG00000171509	RXFP1	relaxin/insulin-like family peptide receptor 1	6.47	0.0000	0.01
ENSG00000185278	ZBTB37	zinc finger and BTB domain containing 37	6.41	0.0001	0.12
ENSG00000086288	NME8	NME/NM23 family member 8	6.31	0.0006	0.24
ENSG00000115255	REEP6	receptor accessory protein 6	6.28	0.0027	0.42
ENSG00000177943	MAMDC4	MAM domain containing 4	6.2	0.0007	0.25
ENSG00000273329	RP11-448A19.1		6.16	0.0003	0.19
ENSG00000122861	PLAU	plasminogen activator, urokinase	6.16	0.0001	0.16
ENSG00000119917	IFIT3	interferon induced protein with tetratricopeptide repeats 3	6.11	0.0020	0.38
ENSG00000147118	ZNF182	zinc finger protein 182	6.01	0.0050	0.49
ENSG00000164050	PLXNB1	plexin B1	6.01	0.0012	0.31
ENSG00000228506	RP11-98I9.4		5.97	0.0053	0.49
ENSG00000107551	RASSF4	Ras association (RalGDS/AF-6) domain family member 4	5.93	0.0040	0.45
ENSG00000118785	SPP1	secreted phosphoprotein 1	5.92	0.0037	0.44
ENSG00000204116	CHIC1	cysteine rich hydrophobic domain 1	5.91	0.0007	0.26
ENSG00000198736	MSRB1	methionine sulfoxide reductase B1	5.85	0.0010	0.30
ENSG00000165633	VSTM4	V-set and transmembrane domain containing 4	5.82	0.0011	0.30
ENSG00000130956	HABP4	hyaluronan binding protein 4	5.81	0.0055	0.49
ENSG00000186615	KTN1-AS1	KTN1 antisense RNA 1	5.77	0.0029	0.42
ENSG00000077984	CST7	cystatin F	5.75	0.0069	0.53
ENSG00000232907	DLGAP4-AS1	DLGAP4 antisense RNA 1	5.74	0.0003	0.19
ENSG00000166173	LARP6	La ribonucleoprotein domain family member 6	5.68	0.0038	0.44
ENSG00000128000	ZNF780B	zinc finger protein 780B	5.65	0.0027	0.42
ENSG00000137571	SLCO5A1	solute carrier organic anion transporter family member 5A1	5.65	0.0004	0.22
ENSG00000005238	FAM214B	family with sequence similarity 214 member B	5.63	0.0020	0.38
ENSG00000235750	KIAA0040	KIAA0040	5.61	0.0037	0.44
ENSG00000113161	HMGCR	3-hydroxy-3-methylglutaryl-CoA reductase	5.6	0.0044	0.47
ENSG00000125388	GRK4	G protein-coupled receptor kinase 4	5.6	0.0001	0.12
ENSG00000229124	VIM-AS1	VIM antisense RNA 1	5.57	0.0025	0.42
ENSG00000064999	ANKS1A	ankyrin repeat and sterile alpha motif domain containing 1A	5.57	0.0044	0.47
ENSG00000121577	POPDC2	popeye domain containing 2	5.53	0.0039	0.44
ENSG00000236939	BAALC-AS2	BAALC antisense RNA 2	5.46	0.0011	0.30

ENSG00000101255	TRIB3	tribbles pseudokinase 3	5.46	0.0010	0.30
ENSG00000116285	ERRFI1	ERBB receptor feedback inhibitor 1	5.44	0.0058	0.50
ENSG00000165959	CLMN	calmin (calponin-like, transmembrane)	5.43	0.0011	0.30
ENSG00000166793	YPEL4	yippee like 4	5.42	0.0004	0.20
ENSG00000068097	HEATR6	HEAT repeat containing 6	5.4	0.0048	0.49
ENSG00000155636	RBM45	RNA binding motif protein 45	5.35	0.0014	0.33
ENSG00000157483	MYO1E	myosin IE	5.33	0.0036	0.44
ENSG00000235109	ZSCAN31	zinc finger and SCAN domain containing 31	5.32	0.0068	0.53
ENSG00000120784	ZFP30	ZFP30 zinc finger protein	5.31	0.0025	0.42
ENSG00000159882	ZNF230	zinc finger protein 230	5.25	0.0011	0.30
ENSG00000163472	TMEM79	transmembrane protein 79	5.2	0.0027	0.42
ENSG00000179476	C14orf28	chromosome 14 open reading frame 28	5.2	0.0050	0.49
ENSG00000106012	IQCE	IQ motif containing E	5.19	0.0057	0.50
ENSG00000225855	RUSC1-AS1	RUSC1 antisense RNA 1	5.18	0.0012	0.31
ENSG00000100097	LGALS1	lectin, galactoside-binding, soluble, 1	5.17	0.0074	0.54
ENSG00000186827	TNFRSF4	tumor necrosis factor receptor superfamily member 4	5.16	0.0032	0.43
ENSG00000150961	SEC24D	SEC24 homolog D, COPII coat complex component	5.16	0.0064	0.52
ENSG00000162894	FCMR	Fc fragment of IgM receptor	5.15	0.0031	0.43
ENSG00000173890	GPR160	G protein-coupled receptor 160	5.14	0.0087	0.55
ENSG00000134243	SORT1	sortilin 1	5.14	0.0014	0.33
ENSG00000260822	GS1-358P8.4		5.13	0.0092	0.55
ENSG00000004864	SLC25A13	solute carrier family 25 (aspartate/glutamate carrier), member 13	5.12	0.0014	0.33
ENSG00000121207	LRAT	lecithin retinol acyltransferase (phosphatidylcholine--retinol O-acyltransferase)	5.11	0.0045	0.47
ENSG00000146856	AGBL3	ATP/GTP binding protein-like 3	5.11	0.0019	0.38
ENSG00000099985	OSM	oncostatin M	5.09	0.0071	0.54
ENSG00000198720	ANKRD13B	ankyrin repeat domain 13B	5.03	0.0028	0.42
ENSG00000132881	RSG1	REM2 and RAB-like small GTPase 1	5.01	0.0036	0.44
ENSG00000153832	FBXO36	F-box protein 36	4.99	0.0075	0.54
ENSG00000176399	DMRTA1	DMRT-like family A1	4.98	0.004	0.45
ENSG00000229619	MBNL1-AS1	MBNL1 antisense RNA 1	4.98	0.0032	0.43
ENSG00000172171	TEFM	transcription elongation factor, mitochondrial	4.97	0.0027	0.42
ENSG00000172824	CES4A	carboxylesterase 4A	4.95	0.0078	0.55
ENSG00000206417	H1FX-AS1	H1FX antisense RNA 1	4.94	0.0004	0.22
ENSG00000170175	CHRNB1	cholinergic receptor, nicotinic beta 1	4.92	0.0053	0.49
ENSG00000130702	LAMA5	laminin subunit alpha 5	4.88	0.0090	0.55
ENSG00000136541	ERMN	ermin	4.87	0.0037	0.44
ENSG00000213398	LCAT	lecithin-cholesterol acyltransferase	4.85	0.0092	0.55
ENSG00000184602	SNN	stannin	4.83	0.0053	0.49
ENSG00000251002	AE000661.37		4.81	0.0082	0.55
ENSG00000250510	GPR162	G protein-coupled receptor 162	4.81	0.0025	0.42
ENSG00000206344	HCG27	HLA complex group 27 (non-protein coding)	4.77	0.0015	0.33

ENSG00000196083	IL1RAP	interleukin 1 receptor accessory protein	4.77	0.0044	0.47
ENSG00000177570	SAMD12	sterile alpha motif domain containing 12	4.75	0.0089	0.55
ENSG00000162650	ATXN7L2	ataxin 7-like 2	4.74	0.0072	0.54
ENSG00000186174	BCL9L	B-cell CLL/lymphoma 9-like	4.74	0.0036	0.44
ENSG00000187037	GPR141	G protein-coupled receptor 141	4.73	0.0040	0.45
ENSG00000248508	SRP14-AS1	SRP14 antisense RNA1 (head to head)	4.65	0.0058	0.50
ENSG00000188206	HNRNPU-AS1		4.6	0.0098	0.56
ENSG00000143748	NVL	nuclear VCP-like	4.59	0.0063	0.52
ENSG00000245060	LINC00847	long intergenic non-protein coding RNA 847	4.58	0.0049	0.49
ENSG00000215045	GRID2IP	glutamate receptor, ionotropic, delta 2 (Grid2) interacting protein	4.55	0.0065	0.52
ENSG00000231312	AC007246.3		4.54	0.0095	0.55
ENSG00000258682	CTD-2002H8.2		4.52	0.0054	0.49
ENSG00000244280	ECEL1P2	endothelin converting enzyme-like 1, pseudogene 2	4.51	0.0094	0.55
ENSG00000137752	CASP1	caspase 1	4.5	0.0046	0.48
ENSG00000241112	RPL29P14	ribosomal protein L29 pseudogene 14	4.49	0.0062	0.52
ENSG00000166912	MTMR10	myotubularin related protein 10	4.48	0.0077	0.54
ENSG00000175787	ZNF169	zinc finger protein 169	4.47	0.0077	0.54
ENSG00000260027	HOXB7	homeobox B7	4.45	0.0086	0.55
ENSG00000144445	KANSL1L	KAT8 regulatory NSL complex subunit 1 like	4.45	0.0020	0.38
ENSG00000177669	MBOAT4	membrane bound O-acyltransferase domain containing 4	4.44	0.0068	0.53
ENSG00000164120	HPGD	hydroxyprostaglandin dehydrogenase 15-(NAD)	4.41	0.0035	0.44
ENSG00000279140	RP11-477I4.4		4.4	0.0012	0.31
ENSG00000240225	ZNF542P	zinc finger protein 542, pseudogene	4.39	0.0023	0.42
ENSG00000237298	TTN-AS1	TTN antisense RNA 1	4.38	0.0025	0.42
ENSG00000065717	TLE2	transducin like enhancer of split 2	4.31	0.0044	0.47
ENSG00000106086	PLEKHA8	pleckstrin homology domain containing A8	4.29	0.0043	0.47
ENSG00000138101	DTNB	dystrobrevin beta	4.29	0.0074	0.54
ENSG00000259630	CTD-2262B20.1		4.29	0.0091	0.55
ENSG00000146592	CREB5	cAMP responsive element binding protein 5	4.26	0.0033	0.43
ENSG00000159885	ZNF222	zinc finger protein 222	4.23	0.0069	0.53
ENSG00000183484	GPR132	G protein-coupled receptor 132	4.19	0.0082	0.55
ENSG00000174004	NRROS	negative regulator of reactive oxygen species	4.18	0.0073	0.54
ENSG00000247287	RP11-902B17.1		4.03	0.0056	0.50
ENSG00000100196	KDELR3	KDEL endoplasmic reticulum protein retention receptor 3	3.99	0.0082	0.55
ENSG00000184313	MROH7	maestro heat-like repeat family member 7	3.97	0.0060	0.51
ENSG00000244701	RP5-894A10.2		3.97	0.0087	0.55
ENSG00000151694	ADAM17	ADAM metallopeptidase domain 17	3.93	0.0087	0.55
ENSG00000176593	CTD-2368P22.1		3.87	0.0039	0.44
ENSG00000197329	PELI1	pellino E3 ubiquitin protein ligase 1	3.8	0.003	0.42

ENSG00000183696	UPP1	uridine phosphorylase 1	3.73	0.0057	0.50
ENSG00000169992	NLGN2	neuroligin 2	3.65	0.0028	0.42
ENSG00000270673	YTHDF3-AS1	YTHDF3 antisense RNA 1 (head to head)	3.61	0.0072	0.54
ENSG00000262075	DKFZP434A062	uncharacterized LOC26102	3.48	0.0085	0.55
ENSG00000218418	RP11-296E7.1		3.32	0.0048	0.49
ENSG00000173846	PLK3	polo-like kinase 3	3.15	0.0088	0.55
ENSG00000129355	CDKN2D	cyclin-dependent kinase inhibitor 2D (p19, inhibits CDK4)	2.97	0.0025	0.42
ENSG00000196214	ZNF766	zinc finger protein 766	2.45	0.0072	0.54
ENSG00000180573	HIST1H2AC	histone cluster 1, H2ac	2.28	0.0084	0.55
ENSG00000179743	FLJ37453	uncharacterized LOC729614	2.2	0.0064	0.52
ENSG00000256235	SMIM3	small integral membrane protein 3	2.16	0.0091	0.55
ENSG00000102362	SYTL4	synaptotagmin like 4	2.05	0.0037	0.44
ENSG00000154144	TBRG1	transforming growth factor beta regulator 1	1.95	0.0052	0.49
ENSG00000138744	NAAA	N-acylethanolamine acid amidase	1.82	0.0055	0.50
ENSG00000120833	SOCS2	suppressor of cytokine signaling 2	1.72	0.0089	0.55
ENSG00000106829	TLE4	transducin like enhancer of split 4	1.52	0.0051	0.49
ENSG00000125844	RRBP1	ribosome binding protein 1	1.48	0.0060	0.51
ENSG00000160190	SLC37A1	solute carrier family 37 (glucose-6-phosphate transporter), member 1	1.43	0.0074	0.54
ENSG00000064601	CTSA	cathepsin A	1.34	0.0086	0.55

Appendix 5: Abbreviations

ALDH	Aldehyde dehydrogenases
ALL	Acute lymphoblastic leukaemia
AML	Acute myeloid leukaemia
ANOVA	Analysis of variance (between group means)
APML	Acute Promyelocytic Leukaemia
ARAC	Cytosine arabinoside
ARIA II	High speed cell FACS based cell sorter manufactured by BD.
ATRA	All-trans retinoic acid
BCI	Barts' Cancer Institute
BM	Bone marrow
BMP4	Bone morphogenetic protein 4
BRDU	Bromodeoxyuridine (5-bromo-2'-deoxyuridine)
CBF	Core binding factor
CD	Cluster Differentiation
cDNA	complementary deoxyribonucleic acid
CEBPA	CCAAT/enhancer-binding protein alpha
CISH	Chromogenic in situ hybridization
cKit	Proto-oncogene cKit or CD117
CLL	Chronic Lymphocytic leukaemia
CLL-1	Human C-type lectin-like molecule-1
CML	Chronic myeloid leukaemia
CR	Complete remission
CT	Cycle Threshold value
DA	Daunorubicin
DAPI	4'-6-diamidino-2-phenylindole
DIC	Disseminated intravascular coagulation
DMSO	Dimethylsulphoxide
DNMT3A	DNA (cytosine-5)-methyltransferase 3A
EDTA	Ethylenediamine tetra-acetate
ELDA	Extreme limiting dilution analysis
ELISA	Enzyme linked immunoabsorbent assay
FACS	Fluorescence-activated cell sorting
FCS	Fetal Calf Serum
FDA	Food and Drug Administration (US)

FET	Fisher's Exact test
FISH	Fluorescence in situ hybridization
FITC	Fluorescein isothiocyanate
FLAG	Fludarabine, high dose cytosine arabinose and G-CSF in combination
FLT3	FMS-like tyrosine kinase 3 or CD135
FMO	Fluorescence minus one
FSC	Forward scatter
5-FU	5-fluorouracil
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
G-CSF	Granulocyte-colony stimulating factor
GMPBSC	G-CSF-mobilized peripheral blood stem cells
GO	Gemtuzumab Ozomycin
Gy	Gray
H&E	Haematoxylin and eosin stain
Hb	Haemoglobin
HS5	Human bone marrow derived stromal cell line
HSC	Haematopoietic Stem Cells
IDH1/2	Isocitrate dehydrogenase 1
IL	Interleukin
IREM-1	Immune receptor expressed on myeloid cells 1 or CD300
ITIM	Immunoreceptor tyrosine-based inhibitory motifs
IQR	Interquartile range
Ki67	Marker of proliferation KI-67, encoded by the *MKI67* gene
Lin	Lineage (in reference to lineage specific markers expressed on maturing cells)
LRI	London Research Institute
LSC	Leukaemia stem cell
LTC-IC	Long-term culture Initiating cell (assay)
MDS	Myelodysplastic syndrome
MFI	Median fluorescence intensity
MGG	May-Grünwald-Giemsa stain
MIC2	Also known as CD99
MLL	Mixed lineage leukaemia
MPN	Myeloproliferative neoplasm
MRC	Medical Research Council
MRD	Minimal residual disease
MS5	Murine bone marrow derived stromal cell line

NOD/SCID	Non-obese diabetic mice with severe combined immunodeficiency disease
NPM	Nucleophosmin
N-RAS	Neuroblastoma RAS viral (v-ras) oncogene homolog
NSG	Non-obese diabetic/severe combined immunodeficiency/interleukin-2 receptor γ-chain null
OS	Overall survival
PCR	Polymerase chain reaction
PE	Phycoerythrin
PeCy7	Phycoerythrin coupled to the indotricarbocyanine dye Cy7
PerCP	Peridinin-chlorophyll-protein
PPAR	Peroxisome proliferator-activated receptors
qPCR	Quantitative polymerase chain reaction, also known as real time PCR
RFS	Relapse free survival
RIN	RNA integrity number
RhoA	Ras homolog gene family, member A
RhoB	Ras homolog gene family, member B
RNA	Ribose nucleic acid
RUNX1	Runt-related transcription factor 1
SD	Standard deviation
SRC	Proto-oncogene tyrosine-protein kinase Src
Siglec 3	Sialic acid binding Ig-like lectin 3
SH2	SRC-homolog 2
SHP1	SH2 domain containing protein tyrosine phosphatase 1
SOCS3	Suppressor of Cytokine signalling 3 protein
SCT	Stem cell transplant
SSC	Side scatter
SYBR	Asymmetrical cyanine dye, used as a nucleic acid stain
TGFβ	Transforming growth factor-β
TNF	Tumor necrosis factor
Th1	Type 1 T helper cells
TIM3	T-cell immunoglobulin and mucin-domain containing
WCC	White cell count
WHO	World Health Organisation
WT1	Wilms tumor protein