Characterising early human lympho-myeloid haematopoietic progenitors and their relationship with acute myeloid leukaemia stem cells

Zahra Aboukhalil

Linacre College
University of Oxford

A thesis submitted for the degree of Doctor of Philosophy

Michaelmas 2018

Abstract

Haematopoiesis is a quintessential cellular hierarchy, with the haematopoietic stem cell having the ability to produce all blood cells. Downstream of the stem cell are multiple progenitor populations with increasing lineage-restriction. However, our knowledge of the progenitors that produce the lymphoid and myeloid lineages remains incomplete. Furthermore, in acute myeloid leukaemia there is a differentiation arrest at the lympho-myeloid progenitor stage, with the development of a progenitor-like leukaemia stem cells. A thorough comparison of the progenitor-like leukaemia stem cell with normal stem and progenitor cells is yet to be completed.

Presented here is a comprehensive study of the lympho-myeloid progenitors in both normal and leukaemic haematopoiesis. Firstly, the normal lympho-myeloid progenitors were fully characterised, both functionally and transcriptionally. Extensive single-cell analysis allowed the heterogeneity and continuum underlying lympho-myeloid differentiation to be uncovered. Although the majority of single-cells generated only one mature lineage, rare multi-lineage single-cells were detected.

The lympho-myeloid progenitors were then studied in acute myeloid leukaemia. It was demonstrated that there is often an expansion of the lympho-myeloid progenitors, which develop leukaemia-propagating ability. Transcriptionally, it was seen that the leukaemia stem cells displayed dysregulated expression profiles, whilst retaining some resemblance to the normal progenitors. A set of genes upregulated in the leukaemia stem cell was identified, and it was proposed that six of these genes may play a role in leukaemogenesis.

Studying both normal and leukaemic haematopoiesis is essential in advancing the treatment of acute myeloid leukaemia. Firstly, a more thorough understanding of the process of normal lympho-myeloid differentiation may aid the development of differentiation therapies. Secondly, by identifying functionally relevant markers of the leukaemia stem cell, treatments targeting the apex of the leukaemic hierarchy may be developed. And through the eradication of the leukaemia stem cell, we may truly eliminate the leukaemic disease.

Word Count: 35,100

Characterising early human lympho-myeloid haematopoietic progenitors and their relationship with acute myeloid leukaemia stem cells



Zahra Aboukhalil

Linacre College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas 2018

Acknowledgements

Firstly, I would like to thank Bilyana Stoilova and Dimitris Karamitros, who led the project on normal human progenitors. I would also like to thank them for all their help and support in training me throughout my DPhil.

I would like to acknowledge that all ossicle experiments were performed by Andreas Reinisch at Stanford University, and all bioinformatic analysis of single-cell qRT-PCR and single-cell RNA-sequencing data was done by Fiona Hamey at the University of Cambridge. I would like to thank David Cruz Hernandez for all his support and guidance with CRISPR experiments. I would like to thank Ryan Beveridge for his advice regarding CRISPR experiments and providing lentiviral constructs, and Connor Sweeney for his help with lentiviral experiments. I would like to thank Asger Jakobsen for his work on genotyping $CD19^+$ and $CD33^+$ populations from engrafted samples and for discussions regarding the nature of the $CD19^+$ engraftment. I would like to thank Jessica Doondeea and Batchimeg Usukhbayar for their support throughout my DPhil, including banking of AML samples, training in mouse experimental skills and help with mouse experiments. I would like to acknowledge that all genotyping of AML samples by targeted re-sequencing was done by Alison Kennedy and Marlen Metzner, and I would like to thank them for training in molecular biology. I would like to thank Eshita Sharma at the Wellcome Trust Centre for Human Genetics for preliminary analysis on normal and leukaemic bone marrow RNA-sequencing data, and Georg Otto for analysis on cord blood RNA-sequencing data. I would like to thank David Sims and Charlotte George from the Computational Genomics: Analysis and Training group for all their guidance and training in computational skills.

I would like to thank my thesis committee, Vincenzo Cerundolo and Claus Nerlov, and co-supervisor Adam Mead, for their advice throughout my DPhil. I would like to thank my supervisor Paresh Vyas for scientific guidance, mentorship and support of this work.

I would like to acknowledge the Medical Research Council for funding this work.

I would like to thank my family for their continued support and encouragement throughout my endeavours. And finally, I would like to thank all of the patients who have donated samples for research, without whom none of this work would have been possible.

Abstract

Haematopoiesis is a quintessential cellular hierarchy, with the haematopoietic stem cell having the ability to produce all blood cells. Downstream of the stem cell are multiple progenitor populations with increasing lineage-restriction. However, our knowledge of the progenitors that produce the lymphoid and myeloid lineages remains incomplete. Furthermore, in acute myeloid leukaemia there is a differentiation arrest at the lympho-myeloid progenitor stage, with the development of a progenitor-like leukaemia stem cells. A thorough comparison of the progenitor-like leukaemia stem cell with normal stem and progenitor cells is yet to be completed.

Presented here is a comprehensive study of the lympho-myeloid progenitors in both normal and leukaemic haematopoiesis. Firstly, the normal lympho-myeloid progenitors were fully characterised, both functionally and transcriptionally. Extensive single-cell analysis allowed the heterogeneity and continuum underlying lympho-myeloid differentiation to be uncovered. Although the majority of single-cells generated only one mature lineage, rare multi-lineage single-cells were detected.

The lympho-myeloid progenitors were then studied in acute myeloid leukaemia. It was demonstrated that there is often an expansion of the lympho-myeloid progenitors, which develop leukaemia-propagating ability. Transcriptionally, it was seen that the leukaemia stem cells displayed dysregulated expression profiles, whilst retaining some resemblance to the normal progenitors. A set of genes upregulated in the leukaemia stem cell was identified, and it was proposed that six of these genes may play a role in leukaemogenesis.

Studying both normal and leukaemic haematopoiesis is essential in advancing the treatment of acute myeloid leukaemia. Firstly, a more thorough understanding of the process of normal lympho-myeloid differentiation may aid the development of differentiation therapies. Secondly, by identifying functionally relevant markers of the leukaemia stem cell, treatments targeting the apex of the leukaemic hierarchy may be developed. And through the eradication of the leukaemia stem cell, we may truly eliminate the leukaemic disease.

Word Count: 35,100

Contents

List of Figures xiii

List of Tables xix

List of Abbreviations xxi

1 Introduction 1

1.1 Normal Haematopoiesis . . . 1

1.1.1 The Haematopoietic Stem Cell and Multi-Potent Progenitor 2

1.1.2 The Common Lymphoid and Myeloid Progenitors . . . 3

1.1.3 The Lymphoid Primed Multi-Potent Progenitor . . . 5

1.1.4 The Multi-Lymphoid Progenitor and Alternative LMPPs . . 7

1.2 Leukaemic Haematopoiesis . . . 9

1.2.1 The Leukaemia Stem Cell . . . 9

1.2.2 Pre-Leukaemia . . . 11

1.2.3 The LSC and Relapse . . . 14

1.3 Aims . . . 16

2 Materials and Methods 17

2.1 Processing of Samples from Healthy and Leukaemic Donors . . . 18

2.2 CD34 Enrichment . . . 18

2.3 Flow Cytometry Staining and Sorting of HSPCs . . . 19

2.4 *In Vitro* LM Differentiation Assays . . . 20

2.5 *In Vitro* T Cell Differentiation Assay . . . 21

2.6 *In Vitro* Combined T Cell-LM Differentiation Assay . . . 22

2.7 Colony Forming Unit Assay . . . 23

2.8 Humanised Ossicle Xenotransplantation Assay . . . 24

2.9 Leukaemic Xenotransplantation Assay . . . 25

2.10 Genotyping of Leukaemic Samples . . . 26

2.11 Bulk RNA-Seq of HSPC Populations . . . 26

2.12 Single-Cell RNA-Seq of HSPC Populations . . . 27

2.13 Gene Expression Analysis by Multiplexed qRT-PCR . . . 27

2.14 Bioinformatic Analysis . . . 28

2.14.1 Bulk RNA-Seq Data from Normal CB . . . 28

2.14.2 Single-Cell qRT-PCR Data from Normal CB . . . 29

2.14.3 Single-Cell RNA-Seq Data from Normal CB 30
2.14.4 Bulk RNA-Seq Data from Normal and Leukaemic BM . . . 30
2.15 *In Vitro* Culture of Leukaemic Samples 31
2.16 Design and Cloning of CRISPR-Cas9 sgRNAs 32
2.17 CRISPR-Cas9 Editing by Transfection of K562 Cells 34
2.18 Lentiviral Transduction of Primary Cells 35

3 Characterisation of Normal Human Lympho-Myeloid Progenitors 37
3.1 Introduction to Human LM Progenitors 37
3.2 Immunophenotypic Characterisation of Normal HSPC 40
3.3 *In Vitro* Assays of LM Progenitors 42
3.3.1 Colony Forming Unit Assay of HSPCs 42
3.3.2 Bulk *In Vitro* Cultures of LM Progenitors 44
3.3.3 Limiting Dilution Assay of LM Progenitors 50
3.3.4 Single-Cell *In Vitro* Cultures of LM Progenitors 54
3.4 *In Vivo* Assays of LM Progenitors 64
3.5 Bulk RNA-Seq of Normal HSPCs 69
3.5.1 PCA and ANOVA Analysis of HSPC Populations 71
3.5.2 Transcriptional Distance Between HSPC Populations 77
3.5.3 Expression of Lineage- and LM-Associated Gene Signatures 78
3.6 Single-Cell Transcriptional Analysis of LM Progenitors 87
3.6.1 Single-Cell qRT-PCR of LM Progenitors 87
3.6.2 Single-Cell RNA-Seq of LM Progenitors 90
3.7 Refining the LMPP and GMP Populations 98
3.7.1 Refining the LMPP: $LMPP^{ly}$ and $LMPP^{mix}$ 98
3.7.2 Refining the GMP: GMP $CD38^{hi}$ 108
3.8 Revised Model of Human LM Differentiation 116
3.9 Conclusion on Human LM Progenitors 117

4 Leukaemia-Propagating Potential of Progenitor-Like Leukaemia Stem Cells 119
4.1 Introduction to the Use of Xenotransplantation Models to Study AML 119
4.2 Immunophenotypic Characterisation of $CD34^+$ AML 120
4.3 Xenotransplantation of AML HSPC Populations 126
4.4 Genotyping of Human Engraftment 133
4.5 Kinetics of Human Engraftment 136
4.6 HSPC Immunophenotype of Human Engraftment 143
4.7 Conclusion on the Xenotransplantation of AML 147

5 Transcriptional Analysis of Progenitor-Like Leukaemia Stem Cells 149
5.1 Introduction to the Gene Expression Analysis of AML 149
5.2 Processing and Quality Control of RNA-Seq Data 152
5.3 Analysis of Normal HSPC Populations 158
5.3.1 PCA and DEGs Across Normal HSPC 158
5.3.2 Normal Stem Cell Gene Signature 164
5.4 Analysis of Leukaemic HSPC Populations 167
5.4.1 PCA of Leukaemic HSPC 167
5.4.2 Engraftment and Leukaemic Gene Expression Profiles 170
5.4.3 Genotype and Leukaemic Gene Expression Profiles 171
5.4.4 Top Variant Genes Across Leukaemia 172
5.4.5 Gender and Leukaemic Gene Expression Profiles 180
5.5 Analysis of Normal and Leukaemic HSPC Populations 182
5.5.1 Hierarchical Clustering and PCA of Normal and Leukaemic HSPC 182
5.5.2 PCA Projections of Leukaemic HSPC 187
5.5.3 Expression of Normal DEGs by Leukaemic HSPC 190
5.5.4 DEGs Between Normal and Leukaemic HSPC 191
5.5.5 Expression of Stem Cell Signature by Leukaemic HSPC . . . 199
5.6 Analysis of the Normal and Leukaemic LMPP and GMP 200
5.6.1 Normal LMPP and GMP 200
5.6.2 Leukaemic LMPP and GMP 202
5.6.3 Comparing Normal and Leukaemic LMPP and GMP 204
5.7 Conclusion on the Gene Expression Analysis of AML 209

6 Functional Validation of Putative Leukaemia-Associated Targets 211
6.1 Introduction to the Use of CRISPR on Haematopoietic Cells 211
6.2 Integration of DEG with Genome-Wide CRISPR Screens 213
6.3 CRISPR-Cas9 Editing of K562 Cells 219
6.3.1 Effect of CRISPR Editing on Viability 219
6.3.2 Inference of CRISPR Editing Analysis 226
6.3.3 Relationship Between ICE Scores and GFP Positivity 230
6.3.4 Indels Introduced in K652 by CRISPR Editing 234
6.4 Optimisation of the *In Vitro* Culture of Primary AML 247
6.5 Lentiviral Transduction of Primary HSPCs 260
6.6 Conclusion on the Functional Validation of Putative Leukaemia-Associated Targets 265

7 Discussion **267**
7.1 Normal Human LM Differentiation 267
7.1.1 Identifying Rare Multi-Lineage Progenitors 269
7.1.2 Studying Normal *In Vivo* Haematopoiesis 269
7.1.3 Eosinophil, Mast-Cell, Basophil and Dendritic Cell Lineages 270
7.1.4 Hierarchical Relationships Between LM Progenitors 271
7.1.5 Ontogeny of HSPC Samples and Enriching for LM Progenitors 272
7.1.6 Normal LM Progenitors and Leukaemic Transformation . . . 272
7.2 Leukaemia-Propagating Potential of AML LSCs 273
7.2.1 Immunophenotype Screening of AML 273
7.2.2 Limitations of Xenotransplantation into NSG Mice and Alternative Mouse Models 273
7.2.3 Aberrant CD19 Expression in AML 275
7.2.4 The Effect of the Immunophenotypic LSC Population on Leukaemic Engraftment 276
7.3 Transcriptional Analysis of Normal and Leukaemic HSPC 277
7.3.1 The Normal Counterpart of the LSC 279
7.3.2 Identification of Upregulated Genes in the LSC 280
7.3.3 Loss of the Normal LMPP-GMP Transcriptional Changes . . 281
7.3.4 Epigenetic, Proteomic and Single-Cell Characterisation of the LSC 281
7.4 Functional Validation of Putative Leukaemic-Associated Targets . . 283
7.4.1 Optimising the Culture of Primary AML 283
7.4.2 Lentiviral Infection and Editing of Primary AML 285
7.5 Conclusion 287

Appendices

A Appendices **291**
A.1 Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells 291
A.2 Flow Cytometric Sorting and Analysis 325
A.3 *In Vitro* LM Differentiation Assays 329
A.4 CRISPR-Cas9 Editing of Leukaemic Cells 330
A.5 *In Vitro* Assays of LM Progenitors 333
A.6 Refining the LMPP and GMP Populations: *In Vitro* Assays 340
A.7 Bulk RNA-Seq of Normal CB HSPC 347
A.8 Immunophenotyping of Normal and Leukaemic BM 347
A.9 Genetic and Cytogenetic Profiles of AML Samples 363
A.10 Summary of Xenotransplantation of AML Experiments 367

A.11 Immunophenotype of Human Engraftment 371
A.12 RNA-Seq Analysis of BM HSPCs 389
A.13 CRISPR-Cas9 Editing of K562 Cells 419
A.14 Optimising AML Culture 421

References **423**

List of Figures

1.1 Classical Model of Haematopoiesis 4
1.2 Alternative Model of Haematopoiesis 6
1.3 Conflicting MLP and LMPP Populations 8
1.4 Alternative CD62L^{hi} LMPP 9
1.5 Progression from Pre-Leukaemia to Leukaemia 12
1.6 Progression from Healthy to Leukaemic Haematopoiesis in AML . . 13

3.1 Current Models of Human LM Differentiation 40
3.2 Immunophenotypic separation of human HSPCs 42
3.3 CFU Assay of HSPCs 44
3.4 Bulk SGF15/2 Culture of LM Progenitors 45
3.5 Gene Expression Analysis of Output from Bulk SGF15/2 Cultures . 46
3.6 Bulk SF7a Culture of LM Progenitors 48
3.7 Gene Expression Analysis of Output from Bulk SF7a Cultures . . . 49
3.8 LDA of LM Progenitors 53
3.9 Cloning Efficiency of LM Progenitors in Single-Cell SGF15/2 Cultures 55
3.10 Lineage Output of LM Progenitors in Single-Cell SGF15/2 Cultures 56
3.11 Ly-My Output of LM Progenitors in Single-Cell SGF15/2 Cultures 57
3.12 Cloning Efficiency of LM Progenitors in Single-Cell S7T2GM/G/M Cultures 58
3.13 Lineage Output of LM Progenitors in Single-Cell S7T2GM/G/M Cultures 59
3.14 Ly-My Output of LM Progenitors in Single-Cell S7T2GM/G/M Cultures 60
3.15 Cloning Efficiency of LM Progenitors in Single-Cell SF7b/Dox Cultures 62
3.16 Lineage Output of LM Progenitors in Single-Cell SF7b/Dox Cultures 63
3.17 Ly-My Output of LM Progenitors in Single-Cell SF7b/Dox Cultures 64
3.18 Flow Cytometry Analysis of Output from Ossicle Assays 65
3.19 Number of Cells Injected in Ossicle Assays of LM Progenitors . . . 66
3.20 Human Chimerism and Lineage Output in Ossicle Assays 68
3.21 Number of Injected Cells and My/Ly Ratio in Ossicle Assay 68
3.22 Morphology of Output in Ossicle Assays 69
3.23 Bulk RNA-Seq of Normal HSPCs: Hierarchical Clustering 71
3.24 PCA of Normal CB HSPC 72
3.25 PCA of Normal CB HSPC: Variant Genes 73

3.26 PCA of Normal CB HSPC: Eigenvalues 74
3.27 PCA of Normal CB HSPC: PC1–PC3 75
3.28 PCA of Normal CB HSPC: Loading Plots 75
3.29 Heatmap Showing the Expression of 300 DEGs Genes by HSPC . . 76
3.30 DEG Between HSPC Populations 77
3.31 Gene Signatures of the LM Progenitors 79
3.32 Expression of Lineage-Associated Genes by LM Progenitors 82
3.33 Expression of TFs by the LM Progenitors 85
3.34 Expression of Lineage-Associated Cytokine Genes by LM Progenitors 86
3.35 Single-Cell qRT-PCR of LM Progenitors: Hierarchical Clustering . . 88
3.36 Single-Cell qRT-PCR of LM Progenitors: Diffusion Maps 90
3.37 Single-Cell RNA-Seq of LM Progenitors: Hierarchical Clustering . . 94
3.38 Single-Cell RNA-Seq of LM Progenitors: PCA 97
3.39 Expression of CD10 and CD45RA by Single LMPP Cells 100
3.40 Expression of CD10 and LMPP Functional Output 101
3.41 Refined Sorting Strategy for the $LMPP^{ly}$ and $LMPP^{mix}$ 101
3.42 Cloning Efficiency of $LMPP^{ly}$ and $LMPP^{mix}$ in Single-Cell SGF15/2 Cultures 102
3.43 Lineage Output of $LMPP^{ly}$ and $LMPP^{mix}$ in Single-Cell SGF15/2 Cultures 103
3.44 Ly-My Output of $LMPP^{ly}$ and $LMPP^{mix}$ in Single-Cell SGF15/2 Cultures 104
3.45 Cloning Efficiency of $LMPP^{ly}$ and $LMPP^{mix}$ in Single-Cell SF7b Cultures 105
3.46 Lineage Output of $LMPP^{ly}$ and $LMPP^{mix}$ in Single-Cell SF7b Cultures 106
3.47 Ly-My Output of $LMPP^{ly}$ and $LMPP^{mix}$ in Single-Cell SF7b Cultures 107
3.48 Expression of CD38 by Single GMP Cells 110
3.49 CD38 Correlation with Lineage-Associated Genes in Single GMPs . 110
3.50 Expression of CD38 and GMP Functional Output 111
3.51 Refined Sorting Strategy for the GMP $CD38^{hi}$ 112
3.52 Cloning Efficiency of GMP $CD38^{hi}$ in Single-Cell SGF15/2 Cultures 113
3.53 Lineage Output of GMP $CD38^{hi}$ in Single-Cell SGF15/2 Cultures . 114
3.54 Ly-My Output of GMP $CD38^{hi}$ in Single-Cell SGF15/2 Cultures . . 115
3.55 Revised Model of Human LM differentiation 117

4.1 Immunophenotype of normal and leukaemic BM 122
4.2 Summary of AML Immunophenotype Screening 123
4.3 Rare Expanded MLP AML Immunophenotype 124
4.4 Immunophenotype of AML Samples for Xenotransplantation 125
4.5 Immunophenotype of Human Engraftment 130

4.6 Aberrant CD19 Expression by Original patient Samples 131
4.7 Summary of Human Engraftment in Xenotransplantation Assays . . 132
4.8 Total Human Engraftment Related to Number of Injected Cells . . 137
4.9 Kinetics of Human Engraftment in Xenotransplantation Assays . . 142
4.10 HSPC Immunophenotype of Human Engraftment 147

5.1 Mapping and Pseudoalignment Statistics of Normal and AML HSPCs 155
5.2 Dispersion Estimates of Normal and AML HSPCs 156
5.3 RNA-Seq Count Data Transformation 158
5.4 PCA of Normal BM HSPC 161
5.5 PCA of Normal BM HSPC: PC1–PC3 162
5.6 Heatmap Showing the Expression of 500 DEGs Genes by Normal BM HSPC 163
5.7 PCA of Normal BM HSPC: k-means Clustering 164
5.8 MA Plot Showing DEGs Between HSC-MPP and Progenitor Populations 165
5.9 Heatmap Showing the Expression of Stem Cell-Associated Genes . . 166
5.10 PCA of Leukaemic HSPC 170
5.11 PCA of Leukaemic HSPC Annotated by Engraftment Status 171
5.12 PCA of Leukaemic HSPC Annotated by Genotype 172
5.13 Variant Genes Across Leukaemia 180
5.14 Leukaemic HSPC Separated by Gender 182
5.15 Hierarchical Clustering of Normal and Leukaemic HSPCs 183
5.16 PCA of Normal and Leukaemic HSPC 186
5.17 PCA of Normal and Leukaemic HSPC: PC1–PC3 187
5.18 PCA Projection of Normal and Leukaemic HSPC 190
5.19 Heatmap Showing the Expression of Normal DEGs by Normal and Leukaemic HSPC 191
5.20 MA Plot Showing DEGs Between Normal LMPP-GMP and Leukaemic LMPP-GMP 192
5.21 Heatmap Showing DEGs Between Normal LMPP-GMP and Leukaemic LMPP-GMP 193
5.22 MA Plot Showing DEGs Between Normal HSC and Leukaemic HSPC 196
5.23 Heatmap Showing DEGs Between Normal HSC and Leukaemic HSPC 197
5.24 Overlap Between Two Sets of DEGs Identified Between Normal and Leukaemic HSPCs 198
5.25 Heatmap Showing the Expression of Stem Cell-Associated Genes by Leukaemic HSPC 200
5.26 MA Plot Showing DEGs Between Normal LMPP and GMP 201

5.27 Heatmap Showing the Expression of DEGs Between the Normal LMPP and GMP . . . 202
5.28 MA Plot Showing DEGs Between Leukaemic LMPP and GMP . . . 203
5.29 Heatmap Showing the Expression of DEGs Between the Leukaemic LMPP and GMP . . . 204
5.30 Overlap Between DEGs Between LMPP and GMP from Normal and Leukaemic Haematopoiesis . . . 206
5.31 Heatmap Showing the Expression of DEGs Between the Normal LMPP and GMP by the Leukaemic LMPP and GMP . . . 208

6.1 Integration of Wang et al. CRISPR Screen and Upregulated Leukaemia Genes . . . 215
6.2 Integration of Tzelepis et al. CRISPR Screen and Upregulated Leukaemia Genes . . . 216
6.3 Integration of Steinhart et al. and Park et al. CRISPR Screens and Upregulated Leukaemia Genes . . . 219
6.4 Effect of CRISPR-Cas9 Transfection of K562 Cells on Viability . . . 221
6.5 Summary of Effect of CRISPR-Cas9 Transfection of K562 Cells on Viability . . . 222
6.6 CRISPR-Cas9 Transfection of K562 Cells: Effect of sgRNAs on Viability . . . 225
6.7 CRISPR Editing of K562 Cells: Summary of ICE Scores . . . 229
6.8 Correlation between GFP^+ Cells and ICE Scores . . . 234
6.9 CRISPR Editing of CLP1 in K562 Cells . . . 236
6.10 CRISPR Editing of COPS2 in K562 Cells . . . 238
6.11 CRISPR Editing of FASTKD5 in K562 Cells . . . 240
6.12 CRISPR Editing of PSMD12 in K562 Cells . . . 242
6.13 CRISPR Editing of SBDS in K562 Cells . . . 244
6.14 CRISPR Editing of WDR26 in K562 Cells . . . 246
6.15 Optimising AML Culture: Effect of MSC on Viability . . . 250
6.16 Optimising AML Culture: Summary of Effect of MSC on Viability . 253
6.17 Optimising AML Culture: Effect of SR1 & UM171 in the Absence of MSCs . . . 256
6.18 Optimising AML Culture: CD14 and CD15 Immunophenotype . . . 259
6.19 Lentiviral Transduction of Primary HSPC . . . 262
6.20 Summary of Lentiviral Transduction of Primary HSPC . . . 264

A.1 Immunophenotype of normal BM . . . 352
A.2 Immunophenotype of AML BM . . . 362
A.3 Absence of Human Engraftment in Xenotransplantation Assays . . 371

A.4 Post-Engraftment HSPC Immunophenotype of BIR344 374
A.5 Post-Engraftment HSPC Immunophenotype of STB257 380
A.6 Post-Engraftment HSPC Immunophenotype of BIR352 380
A.7 Post-Engraftment HSPC Immunophenotype of CAR143 383
A.8 Post-Engraftment HSPC Immunophenotype of BIR517 386
A.9 Post-Engraftment HSPC Immunophenotype of OX2116 388

List of Tables

3.1 Immunophenotypic Definitions of HSPC Populations 41
3.2 Frequencies of Lineage Outputs from LDA of LM Progenitors . . . 50
3.3 Summary of Lineage Outputs from Single LMPP, MLP and GMP Cells 61
3.4 Enriched Pathways by the LM Progenitors 79
3.5 Single-Cell qRT-PCR of LM Progenitors: Cell-Type Composition of Clusters 87
3.6 Single-Cell qRT-PCR of LM Progenitors: DEG Between Clusters . 89
3.7 Single-Cell qRT-PCR of LM Progenitors: Expected Numbers in Clusters 89
3.8 Single-Cell RNA-Seq of LM Progenitors: Cell-Type Composition of Clusters 91

4.1 Expansion of Lympho-Myeloid Progenitors in $CD34^+$ AML 126
4.2 Genotype of Human Engraftment 135

5.1 Summary of RNA-Seq of Normal and Leukaemic HSPCs 153
5.2 Upregulated DEGs in HSC and MPP: TF Co-Expression 167
5.3 Leukaemic Samples Replicate Annotation 168
5.4 DEGs Between Normal LMPP-GMP and Leukaemic LMPP-GMP: Pathway Enrichment 195
5.5 Overlapping Genes Between DEGs Between LMPP and GMP from Normal and Leukaemic Haematopoiesis 207
5.6 DEGs Between Leukaemic LMPP and GMP: Pathway Enrichment . 209

A.1 List of antibodies used for flow cytometric sorting and analysis. . . 325
A.2 Summary of *In Vitro* LM Differentiation Assays. 329
A.3 sgRNAs for CRISPR-Cas9 Editing of Leukaemic Cells 330
A.4 Oligonucleotides for the Cloning of sgRNAs into PX458. 331
A.5 Sequencing Primers for sgRNAs. 332
A.6 Lineage Outputs From LDA of LM Progenitors 333
A.7 Lineage Outputs From Single-Cell SGF15/2 Culture of LM Progenitors 335
A.8 Fisher's Test of Single-Cell SGF15/2 Culture of LM Progenitors . . 337
A.9 Lineage Outputs From Single-Cell S7T2GM/G/M Culture of LM Progenitors 338
A.10 Lineage Outputs From Single-Cell SF7b/Dox Culture of LM Progenitors 339

A.11 Lineage Outputs From Single-Cell SGF15/2 Culture of $LMPP^{ly}$ and $LMPP^{mix}$ . . . 341
A.12 Fisher's Test of Single-Cell SGF15/2 Culture of $LMPP^{ly}$ and $LMPP^{mix}$ 342
A.13 Lineage Outputs From Single-Cell SF7b Culture of $LMPP^{ly}$ and $LMPP^{mix}$ . . . 343
A.14 Fisher's Test of Single-Cell SF7b Culture of $LMPP^{ly}$ and $LMPP^{mix}$ 344
A.15 Lineage Outputs From Single-Cell SGF15/2 Culture of GMP $CD38^{hi}$ 345
A.16 Fisher's Test of Single-Cell SGF15/2 Culture of GMP $CD38^{hi}$ . . . 346
A.17 Number of DEGs Between HSPC Populations . . . 347
A.18 Immunophenotype Screening of $CD34^{+}$ AML . . . 348
A.19 Genetic Profile of AML Samples . . . 364
A.20 Cytogenetic Profile of AML Samples . . . 366
A.21 AML Baseline Cytogenetic Assessment Key . . . 367
A.22 Summary of Xenotransplantation of AML Experiments . . . 368
A.23 Normal BM HSPC Populations: 500 LRT Genes . . . 389
A.24 Normal BM HSPC Populations: Upregulated Stem Genes . . . 392
A.25 DEGs Between Normal LMPP-GMP and Leukaemic LMPP-GMP . 396
A.26 DEGs Between Normal HSC and Leukaemic HSPC . . . 404
A.27 DEGs Between Normal LMPP and GMP . . . 409
A.28 DEGs Between Leukaemic LMPP and GMP . . . 417
A.29 Summary of ICE Scores in K562 Cells . . . 419
A.30 Effect of MSC on Annexin-V^{+} t-test Summary . . . 421

List of Abbreviations

AAV . . . . . Adeno-Associated Virus

AML . . . . . Acute Myeloid Leukaemia

ANOVA . . . . Analysis Of Variance

AZA . . . . . Azacitidine

BF . . . . . Bayes Factor

BM . . . . . Bone Marrow

CB . . . . . Cord Blood

CD . . . . . Cluster of Differentiation

CFU . . . . . Colony Forming Unit

CML . . . . . Chronic Myeloid Leukaemia

CMP . . . . . Common Myeloid Progenitor

CNTF . . . . . Ciliary Neurotrophic Growth Factor

CR . . . . . Complete Remission

CRISPR . . . Clustered Regularly Interspaced Short Palindromic Repeats

CS . . . . . Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) Gene Score

DEG . . . . . Differentially Expressed Gene

DMOG . . . . Dimethyloxalylglycine

Ery . . . . . Erythroid

FC . . . . . Fold Change

FDR . . . . . False Discovery Rate

FMO . . . . . Fluorescence Minus One

FSC . . . . . Forward Scatter

G-CSF . . . . Granulocyte-Colony Stimulating Factor

GFP . . . . . Green Fluorescent Protein

GM-CSF . . . Granulocyte Macrophage-Colony Stimulating Factor

GMP . . . . . Granulocyte Macrophage Progenitor

GO . . . . . Gene Ontology

HIF . . . . . Hypoxia-Inducible Factor

HR . . . . . Homologous Recombination

HSC . . . . . Haematopoietic Stem Cell

HSPC . . . . . Haematopoietic Stem and Progenitor Cell

ICE . . . . . Inference of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) Editing

IL . . . . . Interleukin

Indels . . . . . Insertions and Deletions

ITD . . . . . Internal Tandem Duplication

KO . . . . . Knock-Out

LIC . . . . . Leukaemia Initiating Cell

LIF . . . . . Leukemia Inhibitory Factor

Lin . . . . . Lineage

LM . . . . . Lympho-Myeloid

LMPP . . . . . Lymphoid Primed Multi-potent Progenitor

LRT . . . . . Likelihood Ratio Test

LT-HSC . . . . Long-Term Haematopoietic Stem Cell

Ly . . . . . Lymphoid

M-CSF . . . . Macrophage-Colony Stimulating Factor

MA . . . . . Minus Over Average

Meg . . . . . Megakaryocyte

MEP . . . . . Megakaryocyte Erythroid Progenitor

MFI . . . . . Mean Fluorescent Intensity

MLP . . . . . Multi-lymphoid Progenitor

MNC . . . . . Mononuclear Cell

MOI . . . . . Multiplicity Of Infection

MPP . . . . . Multi-potent Progenitor

mRNA . . . . Messenger Ribonucleic Acid (RNA)

MSC . . . . . Mesenchymal Stromal Cell

mTOR . . . . Mammalian Target Of Rapamycin

My . . . . . Myeloid

NSG . . . . . Non-Obese Diabetic (NOD) Severe Combined Immunodeficient (SCID) Gamma

p-adj . . . . . p-Adjusted Value

PB . . . . . Peripheral Blood

PCA . . . . . Principle Components Analysis

PDX . . . . . Patient Derived Xenograft

pHPL . . . . . Pooled Human Platelet Lysate

PKC . . . . . Protein Kinase C

PKR . . . . . Double-Stranded RNA-Activated Protein Kinase

qRT-PCR . . . Quantitive Reverse Transcriptase-Polymerase Chain Reaction (PCR)

rlog . . . . . Regularised logarithm (log)

RNA-Seq . . . Ribonucleic Acid (RNA)-Sequencing

RNP . . . . . Ribonucleoprotein

SCF . . . . . Stem Cell Factor

SCID . . . . . Severe Combined Immunodeficient

SD . . . . . Standard Deviation

sgRNA . . . . Small Guide Ribonucleic Acid (RNA)

SL-IC . . . . . Severe Combined Immunodeficient (SCID) Leukaemia-Initiating Cells

SR1 . . . . . StemRegenin 1

SSC . . . . . Side Scatter

ST-HSC . . . . Short-Term Haematopoietic Stem Cell

t-SNE . . . . . t-Distributed Stochastic Neighbor Embedding

TF . . . . . Transcription Factor

TPO . . . . . Thrombopoietin

VAF . . . . . Variant Allele Frequency

VST . . . . . Variance Stabilising Transformation

WT . . . . . Wild Type

Chapter 1

Introduction

Contents

1.1 Normal Haematopoiesis 1
1.1.1 The Haematopoietic Stem Cell and Multi-Potent Progenitor 2
1.1.2 The Common Lymphoid and Myeloid Progenitors . . . 3
1.1.3 The Lymphoid Primed Multi-Potent Progenitor 5
1.1.4 The Multi-Lymphoid Progenitor and Alternative LMPPs 7
1.2 Leukaemic Haematopoiesis 9
1.2.1 The Leukaemia Stem Cell 9
1.2.2 Pre-Leukaemia 11
1.2.3 The LSC and Relapse 14
1.3 Aims 16

1.1 Normal Haematopoiesis

The haematopoietic stem cell (HSC) has the ability to produce all of the cells in the blood and immune systems through a process termed haematopoiesis. In adults, haematopoiesis produces 10 x 10^9 new, terminally mature cells a day. The majority of this production is red blood cells, short lived myeloid cells and platelets. However, haematopoiesis also replenishes longer lived cells of the acquired immune system, such as lymphoid B and T cells, and NK cells of the innate immune system. The generation of mature blood and immune cells is achieved through the differentiation of the HSC through several progenitor states, with increasing commitment to a mature lineage. This increase in lineage-restriction is associated with a decrease in self-renewal ability.

The mechanisms by which mature blood and immune cells are produced by the HSC are still debated. Questions remain regarding the hierarchical structure of haematopoiesis, the proposed progenitor populations and their inter-progenitor

relationships, lineage-priming of early stem and progenitor cells and heterogeneity at a single-cell level. Here some of the key findings addressing these questions are discussed.

1.1.1 The Haematopoietic Stem Cell and Multi-Potent Progenitor

At the apex of the haematopoietic hierarchy is the HSC and the closely associated multi-potent progenitor (MPP). These populations have the unique ability to produce all haematopoietic lineages, including: erythrocytes, megakaryocytes and platelets, lymphoid and myeloid cells. The HSC is typically defined as being $CD34^{+}CD38^{-}CD45RA^{-}CD90^{+}$, and the MPP as $CD34^{+}CD38^{-}CD45RA^{-}CD90^{-}$ [1–3]. The HSC is characterised by its self-renewal ability, allowing its long-term, multi-lineage engraftment in xenotransplantation assays [1, 2]. Whereas the HSC's self-renewal allows for serial transplantation ability, the MPP is more limited in its ability to engraft in serial transplantation [3]. Following on from this standard definition, it has also been proposed that CD49f marks true HSCs, with single-$CD49f^{+}$ HSCs able to engraft mice [4]. However, it should be noted that only primary xenotransplantations were carried out in this work, therefore not fully validating the self-renewal ability of the single-$CD49f^{+}$ HSCs.

It has long been appreciated that there is heterogeneity within the earliest compartments of haematopoiesis. Through the single-cell transplantation of mouse HSCs, a previous study has uncovered variations in kinetics, self-renewal and lymphoid/myeloid biases within the HSC compartment [5]. Following on from this, it was proposed that by using CD150 and Hoechst dye efflux, sub-populations of the HSC could be identified. These were either myeloid or lymphoid biased, with differing self-renewal associated with each of these sub-populations [6]. Even within the MPP compartment, sub-populations have been identified with differing biases towards the myeloid and lymphoid lineages [7].

It was observed by Månsson et al. that at the gene expression level, lineage priming was seen in the HSC and early progenitors [8]. Enrichment for

a megakaryocyte-erythroid programme was seen in the earliest long-term (LT)-HSC, with downregulation of the programme with progression to later progenitors. A myeloid programme showed sustained enrichment across the HSC and early progenitors, and a lymphoid programme increased with later progenitors [8]. This implies heterogeneity of the early haematopoietic stem and progenitor cell (HSPC) compartments, with sub-populations showing priming and a gradually increasing bias towards mature lineages.

The association of the HSC with the megakaryocytic and erythroid lineages in both mice and humans has now been well characterised. Yamamoto et al. have suggested a bypass model, where megakaryocytic, erythroid and myeloid restricted progenitors were detected within the immunophenotypic HSC compartment [9]. These were seen to have long-term repopulating activity, and it was suggested this may have represented a mechanism utilised in times of stress. In addition to this, it has been demonstrated that the platelet-associated protein von Willebrand factor (vWF) marks a subset of HSCs that are platelet primed [10]. More recently, a hierarchy of lineage-restricted HSCs has been elucidated through the use of single-cell transplantations of mouse HSCs [11]. Here it was shown that a subset of HSCs was seen to contribute only to the megakaryocytic and platelet lineage, while maintaining multi-potency. This was not observed for any other single lineage [11].

In humans, this close relationship between megakaryocytes and the HSC has also been observed [12]. When looking at fetal liver, it was seen that the multi-potent compartment was enriched for megakaryocytic activity. When looking at adult BM, it was seen that (unlike the myeloid and erythroid lineages) the megakaryocyte lineage was the only major lineage being produced by multi-potent stem/progenitor cells [12].

1.1.2 The Common Lymphoid and Myeloid Progenitors

It was initially thought that the first branching point in haematopoiesis was between the lymphoid and myeloid lineages (Figure 1.1). This was supported by the

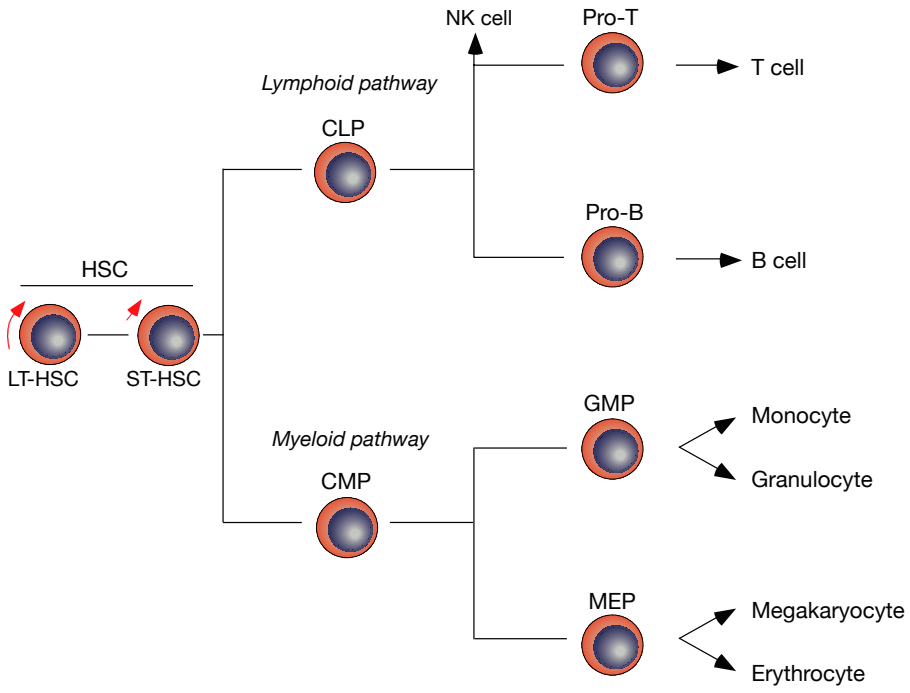


Figure 1.1: Classical Model of Haematopoiesis. A classical model of haematopoiesis demonstrating the segregation between the lymphoid and myeloid lineages by the CLP and CMP progenitors. ST-HSC: Short-Term HSC. Adapted by permission from Macmillan Publishers Ltd: Nature [13], copyright 2000.

identification of both a common myeloid progenitor (CMP) and a common lymphoid progenitor (CLP) ([13, 14] and [15, 16] respectively).

However, the segregation of the lymphoid and myeloid lineages, and the existence of a CMP and CLP population, have since been questioned. Firstly, as discussed above, it was clear that the production of the megakaryocytic and erythroid lineages is closely associated with the earliest stem and progenitor cells. This implies that the megakaryocytic-erythroid lineage is the first segregated lineage. In addition to this, it has been demonstrated that the CMP population is highly heterogenous both in mice and in humans ([17, 18] and [12] respectively). In both species, it was observed that, at a single-cell level, the majority of single CMP cells were already primed towards producing either the megakaryocytic, erythroid or myeloid lineage.

Furthermore, there is evidence for association between the lymphoid and myeloid lineages. In mouse haematopoiesis, it has been shown that during development there

are progenitors with combined lymphoid (B and T cell) and myeloid potential [19]. Here it was suggested that the first branching point in fetal hemopoiesis is between this lympho-myeloid (LM) progenitor and a common myeloerythroid progenitor. Contrary to the classical model of haematopoiesis, these findings suggest that the first branching point may not be between the lymphoid and myeloid lineages. On the other hand, these studies imply that the lymphoid and myeloid lineages may share a common route of differentiation.

1.1.3 The Lymphoid Primed Multi-Potent Progenitor

The existence of a progenitor in adult haematopoiesis with combined lymphoid and myeloid potential, but not megakaryocytic or erythroid, was definitively shown by Adolfsson et al. [20]. In mouse haematopoiesis, the HSC can be found in the Lineage (Lin)$^-$Sca-1$^+$c-kit$^+$ compartment. It was demonstrated that cells in this compartment that also expressed the marker Flt3 had granulocyte, monocyte, B and T cell potential. These Flt3$^+$ cells lacked the potential to produce erythrocytes and megakaryocytes, unlike the Flt3$^-$ HSCs [20]. This led to the proposition of an alternative model of haematopoietic differentiation (Figure 1.2), with a Lin$^-$Sca-1$^+$c-kit$^+$Flt3$^+$ lymphoid primed multi-potent progenitor (LMPP). Here megakaryocytic-erythroid potential is first segregated from an early stem/progenitor population, preceding the LMPP progenitor. A composite model was also proposed where a CMP progenitor is generated by the MPP/ST-HSC. This CMP was proposed to generate both myeloid and megakaryocytic and erythroid output. Down a parallel differentiation trajectory, the LMPP is generated by the MPP/ST-HSC and contributes to the myeloid output by the CMP [20].

In partial agreement with the discovery of the Flt3$^+$ LMPP, the MPP populations have also been segregated based on their expression of Flt3 [21]. Here the MPP was defined as being Lin$^-$Sca-1$^+$c-kit$^+$Thy-1(CD90)$^-$. The MPP was seen to have functional sub-populations based on the expression of both VCAM-1 and Flt3. The Flt3^{hi}VCAM-1$^+$ MPPs were seen to have lymphoid and myeloid (but not megakaryocytic-erythroid) potential [21]. The overlap between the LMPP and

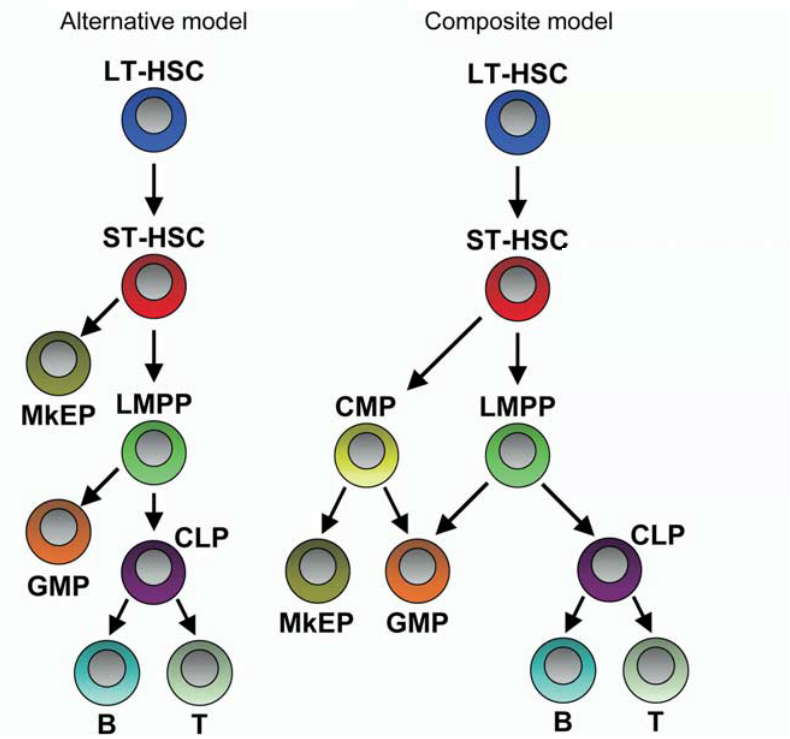


Figure 1.2: Alternative Model of Haematopoiesis. An alternative model of haematopoiesis demonstrating the initial segregation of the megakaryocytic and erythroid lineages. Here the LMPP has joint lymphoid and myeloid potential. The composite model allows for both an LMPP and a CMP progenitor. MkEP: Megakaryocyte and Erythroid Progenitor. Adapted from [20], Copyright (2005), with permission from Elsevier.

this MPP is not entirely clear, due to VCAM-1 and Thy1 having not been used in the original description of the LMPP.

In human haematopoiesis, the existence of a progenitor with combined lymphoid and myeloid potential, but lacking megakaryocytic-erythroid potential, has also been demonstrated [22]. Also termed the LMPP, this progenitor was defined as being Lin^- $CD34^+$ $CD38^-$ $CD90^-$ $CD45RA^+$. Here the LMPP was shown to have granulocyte, monocyte, B and T cell potential. Notably, in this description of the human LMPP it was demonstrated that the LMPP is often expanded and possesses leukaemia stem cell (LSC) activity in AML (see Chapter 1.2.1) [22].

It has also been shown, in mouse haematopoiesis, that there is both heterogeneity and lineage-priming within the LMPP population [23]. Through the transplantation

of barcoded HSC and LMPP cells, it was seen that priming towards a mature lineage was seen at increasing levels from the HSC to the LMPP. Whereas a large proportion of HSCs generated multiple lineages, the majority of LMPP cells were primed towards and generated only one mature lineage. Notably, ~50% of LMPPs were seen to generate only dendritic cells (DCs) [23]. It therefore follows that a similar level of heterogeneity may be seen from the human LMPP.

1.1.4 The Multi-Lymphoid Progenitor and Alternative LMPPs

Accompanying the initial description of the LMPP in humans was the description of an alternative LM progenitor by Doulatov et al. [24]. Termed the multi-lymphoid progenitor (MLP), this progenitor was defined as being Lin^- $CD34^+$ $CD38^-$ $CD90^{neg\text{-}lo}$ $CD45RA^+$ $Flt3^+$ $CD10^+$ $CD7^{-/+}$ [24]. The MLP was demonstrated to have lymphoid (B, NK and T cell) and macrophage potential, but not granulocytic, megakaryocytic or erythroid potential. It was proposed that the MLP had substantial DC potential; however, only monocyte-derived DCs were investigated in these experiments. The ability of the MLP to generate conventional or plasmacytoid DCs was not addressed [24].

In recent work there has been a great deal of confusion between the LMPP and MLP populations. According to their original definitions, the LMPP is $CD10^-$ and the MLP $CD10^+$. A study by Laurenti et al. on early commitment to the lymphoid lineage did maintain the $CD10^+$ definition of the MLP; however, the LMPP was not included in this study [25]. Following on from this, a study by van Galen et al. investigated the effect of lymphoid priming on self-renewal and the expansion of early stem and progenitor cells [26]. Here CD10 was dropped from the immunophenotypic definition of the progenitors and the entire $CD34^+CD38^-CD45RA^+$ compartment was termed the MLP. Finally, CD10 has been re-introduced into immunophenotypic definitions, but with the MLP population being re-defined as either $CD10^+$ or $CD10^-$ (Figure 1.3) [27, 28].

Following on from the initial descriptions of both the LMPP and MLP, an alternatively defined LMPP population was described [29]. Here the LMPP was

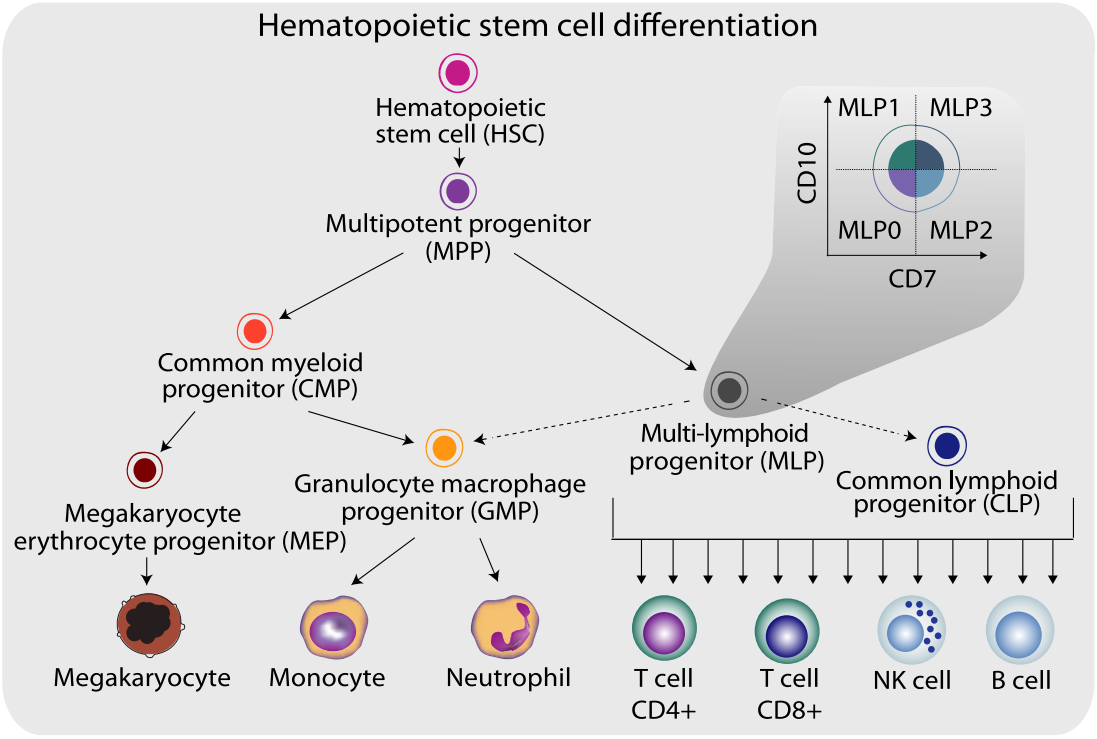


Figure 1.3: Conflicting MLP and LMPP Populations. A model of human haematopoiesis describing conflicting MLP and LMPP populations. Despite its initial definition as being $CD10^{+}$, the MLP is described here as being either $CD10^{-}$ or $CD10^{+}$. These populations are conflicting and overlapping with the previously described $CD10^{-}$ LMPP. Adapted from [27], Copyright (2016), with permission from Elsevier.

defined as being $CD34^{+}Lin^{-}CD10^{-}CD62L^{hi}$. Interestingly, CD38 was not used to define the LMPP, and it was suggested that the LMPP had intermediate levels of CD38 expression. It was postulated that the upregulation of CD38, in comparison to its levels in the HSC, first marked lymphoid priming [29]. It was further postulated that downstream of the $CD62L^{hi}$ was a B cell progenitor (BCP) whose immunophenotypic description overlapped with the MLP, being $CD34^{+}CD10^{+}CD19^{-}$ (Figure 1.4).

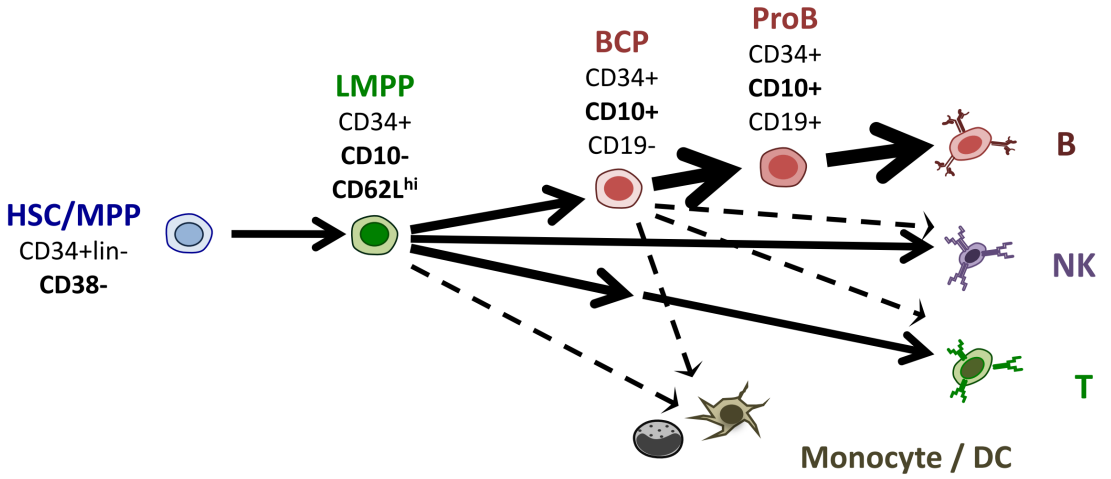


Figure 1.4: Alternative CD62L^{hi} LMPP. A model of human haematopoiesis and lymphoid differentiation describing an alternative CD62L^{hi} LMPP. Adapted by permission from Macmillan Publishers Ltd: Nature Immunology [29], copyright 2012.

1.2 Leukaemic Haematopoiesis

Acute myeloid leukaemia (AML) is the most common adult acute leukaemia. AML is characterised by a differentiation arrest, leading to the accumulation of immature leukaemic blasts in both the bone marrow (BM) and peripheral blood (PB). In the absence of functional, mature immune and blood cells, patients often lack red blood cells and are at risk from infection. The genetic and cytogenetic profiles of AML have been well studied and are closely linked to prognosis and treatment response (described in [30]). Despite many years of research, the prognosis for AML patients remains poor, particularly for elderly patients. Standard treatment for AML continues to be chemotherapy (including azacitidine (AZA) treatment) and/or BM transplantation [30, 31]. In order to improve relapse-free survival rates, it is clear there is an unmet clinical need for novel, targeted treatments for AML.

1.2.1 The Leukaemia Stem Cell

Initially it had been assumed that cancers were diseases composed of homogenous tumours. However, it was soon uncovered that both solid and blood cancers were highly heterogenous in both cell morphology and function (reviewed in [32]).

In AML this heterogeneity was first highlighted by the differing capabilities of leukaemic cells to propagate cultures in a colony forming assay [32, 33]. This small proportion of leukaemic cells with clonogenic ability was termed the AML-colony forming unit (AML-CFU). It was observed that the AML-CFU was highly proliferating in this assay, that it demonstrated properties of self-renewal (being able to re-initiate colonies in serial-replating) and that it was able to differentiate into cells with a partially more mature phenotype [33]. Even in these early descriptions of the hierarchy in AML, it was postulated that the AML-CFU may represent, or be related to, the cell of origin of AML and thus may be an indicator of relapse. Finally, in two-thirds of cases, it was described that the immunophenotype of the AML-CFU resembled CFU-GM progenitors, but immunophenotypes were seen to vary with patients [33].

The concept that AML is organised in a hierarchy with LSCs at the apex was solidified with the identification of rare leukaemia initiating cells *in vivo*. Lapidot et al. demonstrated that rare cells could be identified that had the ability to engraft severe combined immunodeficient (SCID) mice with human leukaemia [34]. It was proposed that these cells represented a stem cell population that had the ability to maintain AML and were more immature than the AML-CFU [34]. The SCID leukaemia-initiating cell (SL-IC) compartment was further studied and characterised by others [2, 35]. The SL-IC was generally seen to be $CD34^{+}CD38^{-}$ leading to the inference that the normal $CD34^{+}CD38^{-}$ HSPCs may be the target for leukaemic transformation [2, 35]. Due to the AML-CFU being found in both the $CD34^{+}CD38^{-}$ and $CD34^{+}CD38^{+}$ compartments and being more frequent than the SL-IC, it was concluded that the SL-IC may be a sub-population within the AML-CFU [2].

Following on from these initial studies, the immunophenotype of the engrafting AML cells has been further refined. The majority of AML cases over-express CD34, and it has been shown that in these cases, leukaemia initiating cells can be detected in both the $CD34^{+}CD38^{-}$ and $CD34^{+}CD38^{+}$ compartments [22, 36]. Additionally, a small proportion of AML cases lack expression of CD34, which results in an alternative hierarchy, genetic structure and $CD34^{-}$ immunophenotype of LSCs [37].

It was originally assumed that the $CD34^{+}CD38^{-}$ LSC population closely resembled, and potentially originated from, the normal HSC. However, Blair et al. demonstrated that, unlike the normal HSC, the LSC lacked expression of CD90 [38]. The progenitor-like immunophenotype of the LSC in $CD34^{+}$ AML was confirmed more recently by Goardon et al. [22]. Here it was shown that the LSC most often resembled the normal LMPP and GMP progenitors and, occasionally, the CMP and MPP progenitors. This observation was seen across genetic and cytogenetic sub-groups of AML. Preliminary work was done on the hierarchical relationships between these populations, but this remains to be definitively demonstrated. The progenitor-like status of the LSC throws into question the cell-type of origin of AML.

It should also be noted that functional heterogeneity within the LSC compartment has been described [39]. Here through the barcoding and serial transplantation of AML, it was seen that three functional classes of LSC were present. Short-term, long-term and quiescent long-term SL-ICs were observed, and these were organised in a hierarchical structure [39]. The immunophenotype of the various SL-IC subsets was not investigated, but it is possible that there may have been immunophenotypic differences between these groups.

1.2.2 Pre-Leukaemia

Taking into account the immunophenotypic definitions of the LSC and its similarity to normal stem and progenitor populations, many groups have questioned the cell of origin of AML. A fundamental discovery, partially addressing this question, was the identification of a pre-leukaemic state pre-existing leukaemic transformation.

First described by Jan et al., pre-leukaemia was proposed to be a state that preceded full leukaemic transformation, with pre-leukaemic cells being the predecessors to leukaemic cells [40]. To conclude this, residual putative normal HSCs were isolated and studied from AML samples. It had been proposed that CD99 and TIM3 were exclusively expressed in AML, and normal HSCs were isolated as being $Lin^{-}CD34^{+}CD38^{-}$ and $TIM3^{-}$ and/or $CD99^{-}$ [40]. It should be noted that this immunophenotypic definition would have included putative normal HSC, MPP,

LMPP and MLP populations, all of which are $CD34^{+}CD38^{-}$. Within this normal population, leukaemic mutations could be detected, indicating that this mutated normal population was the predecessor to the leukaemic cells. It follows that the partially mutated pre-leukaemic cells underwent further round(s) of mutations resulting in fully transformed, leukaemic cells.

In addition to this, the order of mutations in leukaemic transformation has been elucidated by studying the pre-leukaemic cells [40, 41]. Here it was proposed that initial mutations were in 'landscaping' genes, typically involved in epigenetic regulation, and that subsequent mutations were in proliferative genes, involved in activated signalling (Figure 1.5, [41]). Furthermore, it was demonstrated that these pre-leukaemic HSPCs were able to differentiate into mature immune cells, with mature myeloid and lymphoid cells harbouring pre-leukaemic mutations. Finally, when looking at relapse samples, a diverse evolutionary mutational structure was observed, which may have involved the pre-leukaemic clones [41].

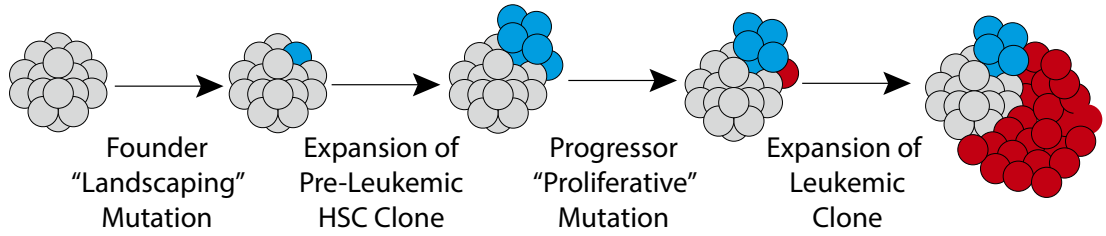


Figure 1.5: Progression from Pre-Leukaemia to Leukaemia. A model demonstrating the order of mutations involved in progressing from a pre-leukaemic state to leukaemia. Adapted from [41], copyright 2014 National Academy of Sciences.

Following on from these initial reports, further work into pre-leukaemic HSCs was described by Shlush et al. [42]. Here more refined immunophenotypic definitions were used for the HSPC populations (it should be noted that the MLP population here was both $CD10^{+}$ and $CD10^{-}$, overlapping with the LMPP). It was seen that, in patients with a *DNMT3A* mutation, pre-leukaemic HSCs harbouring a *DNMT3A* mutation were readily detectable [42]. It was observed that *DNMT3A* mutations

typically preceded mutations in *NPM1* and *FLT3*-ITDs. *DNMT3A* mutant HSCs were seen to have a competitive advantage in transplantation assays, potentially allowing for clonal dominance of the mutant HSCs. This is in agreement with more recent work on the role of *DNMT3A* in age-related clonal haematopoiesis (ARCH) [43, 44]. As with previous work, it was proposed that pre-leukaemic clones were able to survive chemotherapy, due to their presence at remission, and may thus contribute to disease relapse [42].

Figure 1.6 depicts a possible mechanism for leukaemic development from healthy haematopoiesis. Here initiating, pre-leukaemic mutations may occur at either the earliest HSC/MPP stage of haematopoiesis or at an LMPP/GMP progenitor stage. These pre-leukaemic HSPCs are still able to contribute to mature haematopoiesis. Transforming, second hits may occur in either the pre-leukaemic HSC/MPP or LMPP/GMP, transforming these populations into full LSCs. If in the latter populations, these may include reactivation of self-renewal. These LSC populations are no longer able to generate mature output and instead generate leukaemic blasts.

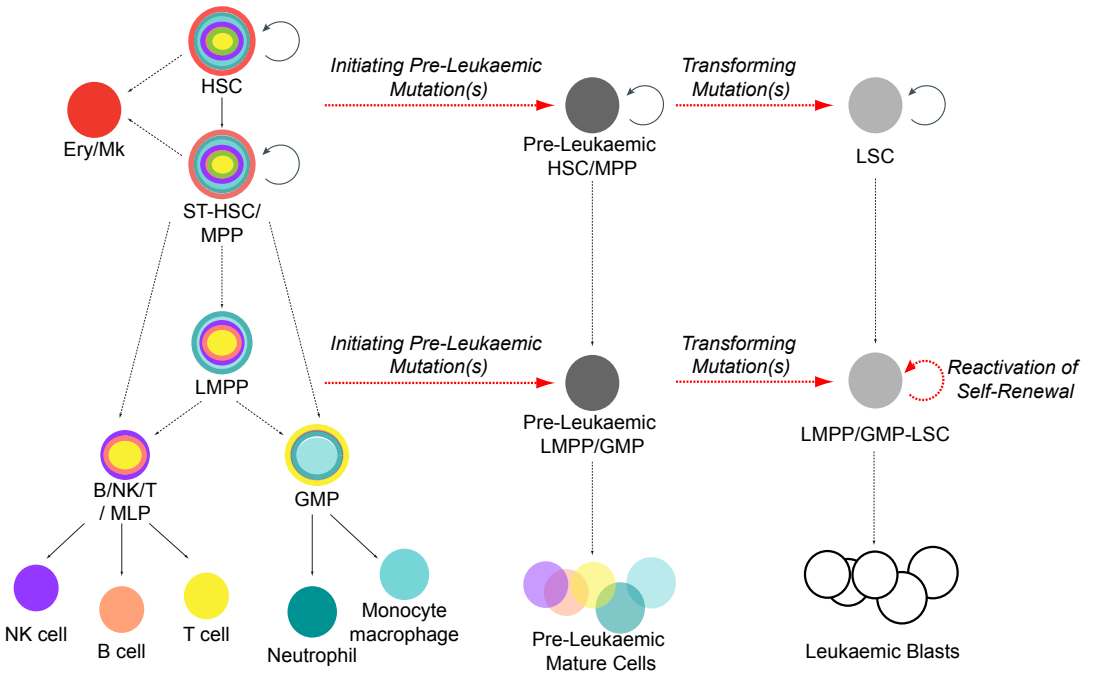


Figure 1.6: Progression from Healthy to Leukaemic Haematopoiesis in AML. A model depicting a possible mechanism of progression from healthy to leukaemic haematopoiesis, passing through a pre-leukaemic state.

1.2.3 The LSC and Relapse

It has long been postulated that the high levels of relapse seen in AML are associated with LSC persistence at remission, potentially through evasion of chemotherapy (reviewed in [32, 45, 46]). In support of this, it has been demonstrated that expression of either an HSC- or LSC-associated gene signature is predictive of patient survival [47]. This agrees with the hypothesis that LSC persistence is an indicator of a poor prognosis.

It has been proposed that interleukin (IL)-3 receptor α (IL-3Rα/CD123) is a potential marker of LSCs [22, 48, 49]. One study has shown that AML patients could be stratified into groups with both high and low IL-3Rα levels [49]. In patients with higher IL-3Rα levels, and presumably higher LSC levels, it was observed that these patients had a lower proportion of complete responders and a higher incidence of relapse [49].

In another study, Craddock et al. characterised the kinetics of the LMPP- and GMP-like LSC populations in both AML and MDS patients treated with either AZA and sodium valproate or induction chemotherapy [50]. Here it was seen that in patients who achieved complete remission (CR) there was typically a reduction of the LSC compartment. However, in patients not responding to treatment, there was no reduction in the LSC compartment. From responsive patients, eradication of the LSC compartment was observed in only three patients, and these patients achieved a durable CR. In the remaining patients, persistence of an LSC compartment was associated with relapse [50]. In a subsequent similar study, the LMPP-like LSC population was characterised in response to AZA treatment [51]. Similarly, here it was seen in non-responders that the LMPP-LSC frequency was unchanged. In patients who achieved some form of CR (CR/CR with incomplete blood count recovery (CRi)/marrow CR (mCR)), reduction in the LMPP-LSC compartment was observed. However, in seven of these patients, despite achieving CR, an increase in the LMPP-LSC frequency was seen to accompany disease relapse [51].

Interestingly, Ho et al. carried out a similar kinetic, studying the LSC at both diagnosis and relapse following chemotherapy treatment [52]. Here all of the

$CD34^{+/-}CD38^{+/-}$ compartments were studied, and their LSC ability and frequency in xenotransplantation assays assessed at both time-points. Firstly, it was observed that at relapse the frequency of the LSC increased 9- to 90-fold [52]. Secondly, it was seen that the LSC ability was diversified at relapse, moving to slightly more mature immunophenotypic compartments that did not previously show LSC ability. Finally, it was also seen that the relapsed-LSCs showed a change in their expression of cell surface markers previously described to be associated with LSCs [52]. On the whole, this suggests a key role the LSC and its plasticity play in relapse.

More recently, Shlush et al. have combined the use of functional xenotransplantation assays and genotyping to demonstrate a link between the LSC and disease relapse [53]. Here diagnostic AML samples were xenotransplanted, and the produced engraftment genotyped and compared to the genotype of both the diagnostic and relapse sample. It was seen that the diagnostic xenografts often contained mutational variants that were detected in the relapse sample [53]. This implied that the relapse clone originated from a rare sub-clone present at diagnosis. In a subset of patients, the origins of the relapse clone could be traced to what was termed an MLP population that was dormant at diagnosis. It should be noted that here the MLP population is both $CD10^{+}$ and $CD10^{-}$, overlapping with the LMPP. In a separate subset of patients, the origin of relapse was traced back to a more mature population, as defined immunophenotypically [53]. Despite their immunophenotype, the latter showed transcriptional similarities to early HSPCs and enrichment for stem-cell programmes [53].

On the whole, a consensus on LSC survival through treatment, persistence at remission and then emergence and expansion at relapse can be concluded. Firstly, this confirms the clinical and functional relevance of the LSC, being not simply an artefact of the assays used to define it. Secondly, this highlights the importance of fully understanding the LSC and targeting it in treatment in order to minimise patient relapse.

1.3 Aims

It is apparent that there are several gaps in our current knowledge of both normal and leukaemic haematopoiesis. Firstly, there is a lack of clarity in the literature regarding the human LM progenitors. The first aim of this work is to characterise the human LM progenitors, focusing on the LMPP, MLP and GMP populations. This will be done both functionally and transcriptionally. In order to investigate the heterogeneity within the LM progenitors, single-cell analyses will be carried out. This will allow the question of multi-potentiality to be addressed and possibly to refine our current understanding of LM differentiation.

This improved knowledge of the LM progenitors and LM differentiation will form a foundation for understanding their leukaemic transformation. Firstly, the transformed LSC populations will be studied functionally in order to assess and characterise their leukaemia-propagating abilities. Then the LSC populations shall be studied transcriptionally. This will be done to gain an insight into both how the LSCs' transcriptional profile relates to normal HSPC and to identify differentially expressed genes (DEGs) in LSCs. Differentially upregulated genes in the LSC will be putative targets involved in leukaemogenesis. Finally, the framework to functionally validate these targets by CRISPR-Cas9 in primary AML samples will be developed.

By identifying functionally relevant markers of the LSC population, treatments specifically targeting the LSC may be developed. And through the eradication of the LSC, it may be possible to both reduce patient relapse and ultimately eliminate the disease.

Chapter 2

Materials and Methods

Contents

2.1 Processing of Samples from Healthy and Leukaemic Donors . . . 18
2.2 CD34 Enrichment . . . 18
2.3 Flow Cytometry Staining and Sorting of HSPCs . . . 19
2.4 *In Vitro* LM Differentiation Assays . . . 20
2.5 *In Vitro* T Cell Differentiation Assay . . . 21
2.6 *In Vitro* Combined T Cell-LM Differentiation Assay . 22
2.7 Colony Forming Unit Assay . . . 23
2.8 Humanised Ossicle Xenotransplantation Assay . . . 24
2.9 Leukaemic Xenotransplantation Assay . . . 25
2.10 Genotyping of Leukaemic Samples . . . 26
2.11 Bulk RNA-Seq of HSPC Populations . . . 26
2.12 Single-Cell RNA-Seq of HSPC Populations . . . 27
2.13 Gene Expression Analysis by Multiplexed qRT-PCR . 27
2.14 Bioinformatic Analysis . . . 28
2.14.1 Bulk RNA-Seq Data from Normal CB . . . 28
2.14.2 Single-Cell qRT-PCR Data from Normal CB . . . 29
2.14.3 Single-Cell RNA-Seq Data from Normal CB . . . 30
2.14.4 Bulk RNA-Seq Data from Normal and Leukaemic BM . 30
2.15 *In Vitro* Culture of Leukaemic Samples . . . 31
2.16 Design and Cloning of CRISPR-Cas9 sgRNAs . . . 32
2.17 CRISPR-Cas9 Editing by Transfection of K562 Cells . 34
2.18 Lentiviral Transduction of Primary Cells . . . 35

Acknowledgements: All experiments done with normal CB were done in conjunction with Bilyana Stoilova and Dimitris Karamitros, please see reference [54]. All ossicle experiments were performed by Andreas Reinisch at Stanford University. All bioinformatic analysis of single-cell qRT-PCR and single-cell RNA-sequencing data was done by Fiona Hamey at the University of Cambridge. All genotyping of AML samples by targeted re-sequencing was done by Alison Kennedy and Marlen Metzner at the University of Oxford.

2.1 Processing of Samples from Healthy and Leukaemic Donors

BM or cord blood (CB) samples from healthy donors were obtained with informed consent (UK protocol MREC 06/ Q1606/ or Administrative Panel on Human Subjects Research Institutional Review Board-approved protocols Stanford IRB no. 18329, no. 6453, and no. 5637). Fresh CB samples were purchased from NHS Cord Blood Bank, UK or from New York Blood Center. Healthy CB and BM samples were processed within 16–34 h after collection. To isolate mononuclear cells (MNCs), a Ficoll gradient separation was carried out. Fresh CB/BM was diluted in a flow cytometry staining buffer (Iscove's Modified Dulbecco's Medium (IMDM) (Thermo Fischer) 10% fetal bovine serum (FBS) + 20 µg/ml DNAse). Diluted CB/BM was layered onto 15 ml of Ficoll, centrifuged at 400 g for 25 min at room temperature with no brake and the interphase of MNCs collected.

MNCs isolated from healthy CB/BM samples were then taken forward for CD34 enrichment (see Chapter 2.2). MNCs isolated from leukaemic BM were stained with anti-human CD34-PerCP and analysed by flow cytometry to assess the CD34 positivity of the sample. Samples with a low CD34 positivity and/or low cell numbers were banked and frozen in liquid nitrogen. Samples with a high CD34 positivity and sufficient cell numbers underwent CD34 enrichment (see Chapter 2.2) and were subsequently banked and frozen in liquid nitrogen.

2.2 CD34 Enrichment

To enrich for $CD34^+$ cells, a CD34 Microbead Kit and MACS separation columns (Miltenyi Biotec, Bisley UK) were used. Briefly, up to $1x10^8$ MNCs were resuspended in 100 µl CD34 microbeads, 100 µl Fc block and 300 µl PBE (phosphate-buffered saline (PBS) + 0.5% bovine serum albumin (BSA) + 2 mM EDTA) and incubated, rolling for 30 min at 4°C. 5 ml PBE was then added, cells spun down at 300 g for 5 min and up to $1x10^8$ MNCs resuspended in 500 µl PBE. 2 ml of PBE was

run through the MACS separation column and cells then loaded onto the column. Columns were washed three times with 3 ml PBE and to elute the $CD34^+$ fraction, the column was removed from the MACS magnet, 5 ml of PBE added and the bound fraction eluted with the plunger.

2.3 Flow Cytometry Staining and Sorting of HSPCs

Antibodies used for flow cytometry staining and sorting are listed in Table A.1. All antibody incubations for flow cytometry were carried out in flow cytometry staining buffer for 20 minutes at 4°C, in the dark. Following antibody incubations, cells were washed with 1 ml flow cytometry staining buffer, centrifuged at 300 g for 5 min and resuspended in flow cytometry staining buffer.

Human CB/BM samples enriched for $CD34^+$ cells were first stained with an antibody cocktail against the following mature lineage (Lin) markers: CD2, CD3, CD4, CD7, CD8a, CD11b, CD14, CD19, CD20, CD56 and CD235a. Cells were then stained with a secondary Qdot 605 conjugated goat F(ab')2 anti-mouse IgG (H + L) antibody. Following this, cells were stained with the following human HSPC panel: CD34-PerCP, CD38-FITC, CD45RA-PE/BV650, CD10-APC, CD123-PE Cy7 and CD90-biotin. Finally, cells were stained with a secondary streptavidin-conjugated APC-eF780 antibody and then resuspended in the Live-Dead stain Hoechst 33258 (Invitrogen, Loughborough UK; final concentration: 1 µg/ml).

Unstained, single-stained and fluorescence-minus-one (FMO) controls were used to determine background staining and compensation in each channel. Single-stained controls used the anti-mouse Ig, κ/negative control compensation particle set and the anti-rat Ig, κ/negative control compensation particles set (BD, Oxford UK). Prior to single-cell sorting, single fluorescent beads were deposited directly into a 96-well plate to establish accuracy of single-cell deposition (> 99%). For index sorting, the following parameters were recorded for all single-cells: FSC, SSC, Hoechst, CD34, CD38, CD45RA, CD10, CD90, CD123 and expression of lineage markers.

Sorting was performed on a BD FACSAria III or BD Fusion, and flow cytometry analysis was done on LSR Fortessa X20. Data analysis was performed using Diva v8.1 or FlowJo v10.0.06 and v10.0.07r2.

2.4 *In Vitro* LM Differentiation Assays

To assess the differentiation of normal HSPC populations into the LM lineages *in vitro* co-culture assays were used. For bulk cultures, 24-well plates were coated with 0.1% gelatin and the mouse stromal cell line MS-5 seeded onto these plates at a density of 2x10^4 cells/well. Cells were cultured in α-MEM medium (Gibco/Thermo Fisher Scientific Loughborough UK) supplemented with 10% FBS (Hyclone, GE Healthcare, SH30070.03 Amersham Hatfield, UK), 1% penicillin-streptomycin, 1% l-glutamine, 0.1 µM DuP-697 (Cayman Chemical, Ann Arbor, MI, USA), 20 ng/ml SCF, 10 ng/ml G-CSF, 10 ng/ml FLT3L, 10 ng/ml IL-15 and 10 ng/ml IL-2 (Peprotech London UK). These culture conditions are termed SGF15/2 (Table A.2). 24 hours following MS-5 plating, 150 HSPCs were deposited in each well. Half of the culture media was changed weekly. Cells were stained and analysed by flow cytometry following 1–3 weeks of culture.

The limiting dilution assay (LDA) was carried out in 96-well plates that had 2,500 MS-5 cells plated per well, in 100 µl media without cytokines. HSPCs were then added to the culture at the following cell doses: 1, 2, 5, 10 and 20 cells and 100 µl of a twice concentrated SGF15/2 culture media added. Cells were stained and analysed by flow cytometry following 2–2.5 weeks of culture. Frequency calculations were performed using L-Calc software (Stem Cell Technologies) and were independently verified by ELDA software (`http://bioinf.wehi.edu.au/software/elda/`). The LDA plots were generated using R, with lines representing the estimates calculated by ELDA software.

Single-cell cultures were carried out in the same conditions as the LDA; however, a single HSPC was flow sorted onto the MS-5 stromal layer. Single-cell cultures were also carried out in the following alternative conditions: MS-5 stroma in H5100 medium (StemCell Technologies Cambridge UK) supplemented with 100 ng/ml

SCF, 20 ng/ml IL-7, 50 ng/ml TPO, 10 ng/ml IL-2, 20 ng/ml GM-CSF, 20 ng/ml G-CSF and 10 ng/ml M-CSF (all from Peprotech). These culture conditions are termed S7T2GM/G/M (Table A.2) and have been previously described [24]. In S7T2GM/G/M conditions cells were stained and analysed by flow cytometry following 4 weeks of culture.

For the flow cytometry analysis of *in vitro* LM differentiation assays, cells were harvested by physical disaggregation of wells, filtered through a Corning Falcon Test Tube with Cell Strainer Snap Cap (Thermo Fisher Scientific) and washed with flow cytometry staining buffer. All antibody incubations were carried out in flow cytometry staining buffer for 20 minutes at 4°C, in the dark. Following antibody incubations, cells were washed with 1 ml flow cytometry staining buffer, centrifuged at 300 g for 5 min and resuspended in flow cytometry staining buffer. The following antibodies were used for analysis: anti-human CD15-FITC, anti-human CD14-PE, anti-human CD19-PE Cy7, anti-human CD56-APC or -PE Cy5, anti-human CD45-APC Cy7 and, in some cases, anti-human CD34-BV605 (Table A.1). For single-cell cultures, wells with more than 15 human $CD15^+$, $CD14^+$, $CD56^+$ or $CD19^+$ cells were considered positive for having generated mature cells.

2.5 *In Vitro* T Cell Differentiation Assay

To assess the *in vitro* differentiation of normal HSPC populations into the T cell lineage, a T cell differentiation assay was used. For T cell cultures, the mouse stromal cell line OP9 expressing the human Delta-Like 1 (hDL1) was used [55]. 24-well plates were coated with 0.1% gelatin and OP9-hDL1 cells seeded onto the plates at a density of $2x10^4$ cells/well. Cells were cultured in freshly prepared α-MEM medium (Gibco/Thermo Fisher Scientific, 12000-063) with 20% heat-inactivated FBS (Hyclone, GE Healthcare, SH30070.03 Amersham Hatfield, UK), 1% penicillin-streptomycin and 1% l-glutamine. 24 hours following OP9-hDL1 plating, 150 HSPCs were deposited in each well and cultured in the presence of 10 ng/ml SCF, 5 ng/ml FLT3L and 5 ng/ml IL-7 (Peprotech, London, UK). These culture conditions are termed SF7a (Table A.2). Cells were transferred onto

fresh OP9-hDL1 stroma weekly. To achieve this, cells were harvested by physical disaggregation of wells, filtered through a Corning Falcon Test Tube with Cell Strainer Snap Cap (Thermo Fisher Scientific), washed with flow cytometry staining buffer and transferred to new plates with fresh OP9-hDL1 cells.

For the flow cytometry analysis of the bulk *in vitro* T cell differentiation assay, cells were harvested and stained as described above (see Chapter 2.4). Cells were stained and analysed by flow cytometry following 5 and 7 weeks of culture. The following antibodies were used for analysis: anti-human CD7-FITC, anti-human CD1a-PE, anti-human CD8-PE Cy7, anti-human CD4-APC and anti-human CD45-APC Cy7 (Table A.1).

2.6 *In Vitro* Combined T Cell-LM Differentiation Assay

To assess the differentiation of normal HSPC populations into both the T cell lineage and the LM lineages, a combined *in vitro* T cell-LM differentiation assay was used. For combined T cell-LM cultures, the MS-5 cell line with doxycycline-inducible hDL1 expression was used (MS5-hDL1^{IND100}) [56]. 96-well plates were coated with 0.1% gelatin and MS5-hDL1^{IND100} cells seeded onto the plates at a density of 2,500 cells/well. Cells were cultured in 100 µl of freshly prepared α-MEM medium (Gibco/Thermo Fisher Scientific, Loughborough UK) with 20% FBS (Hyclone, GE Healthcare, SH30070.03HI, Amersham Hatfield, UK), 1% penicillin-streptomycin and 1% l-glutamine. 24 hours following MS5-hDL1^{IND100} plating, a single HSPC was flow sorted onto the stromal layer and cultured in the presence of 20 nM insulin (Sigma-Aldrich, St Louis, MO), 50 ng/ml SCF, 20 ng/ml FLT3L and 10 ng/ml IL-7 (Peprotech London UK). These culture conditions are termed SF7b culture (Table A.2). Half of the culture media was changed weekly.

Cells were harvested as described above (see Chapter 2.4) following 21 days of culture and divided into two groups. Group one was stained and analysed by flow cytometry. Group two was re-seeded onto fresh MS5-hDL1^{IND100} cells and cultured in the presence of 1 µg/ml doxycycline (SF7b/Dox culture conditions, Table A.2). Half of the culture media was changed twice weekly and fresh doxycycline was

added three times weekly. Following 42 days of culture, cells were harvested (as described above), stained and analysed by flow cytometry.

For flow cytometry analysis at day 21, cells were stained with: anti-human CD15-FITC, anti-human CD14-PE, anti-human CD19-PE Cy7, anti-human CD56-APC or -PE Cy5, anti-human CD45-APC Cy7 (Table A.1). Wells with more than 8 human $CD15^+$, $CD14^+$, $CD56^+$ or $CD19^+$ cells were considered positive for having generated mature cells. For flow cytometry analysis at day 42, cells were stained with: anti-human CD7-FITC, anti-human CD1a-PE, anti-human CD8-PE Cy7, anti-human CD4-APC and anti-human CD45-APC Cy7 (Table A.1). Wells with more than 8 $CD7^+$ cells were considered positive for having generated T cells.

2.7 Colony Forming Unit Assay

To assess the differentiation of normal HSPC populations into the myeloid and erythroid lineages, a colony forming unit (CFU) assay was carried out, as previously described [22]. Briefly, 250 HSPCs were plated in 1.1 ml of MethoCult GF+ H4435 (StemCell Technologies) and incubated for 14 days at 37°C. Colony identity was determined to be one of the following based on colony morphology: mixed erythro-myeloid colonies (CFU-mix), monocyte-macrophage colonies (CFU-M), granulocyte colonies (CFU-G), granulocyte and monocyte-macrophage colonies (CFU-GM) or erythroid colonies (burst-forming units-erythroid and CFU-E (E)).

Colony identity was also confirmed by cellular morphology. Individual colonies were collected and cytospun onto a slide on medium acceleration, 800 rpm for 5 min. Cells were then stained with a May-Grünwald-Giemsa stain (Sigma, Poole UK) as follows. Firstly, slides were immersed in 100% methanol for 5 minutes, then in May-Grünwald solution for 5 minutes. Slides were rinsed in distilled water and then transferred to a Giemsa solution, pre-diluted to a concentration of 1/20 with distilled water, for 40 minutes. Slides were rinsed in distilled water and once dry, were mounted with DPX Mountant for histology (Sigma-Aldrich). Cellular morphology was inspected using a Zeiss Axiovert 25/40C light microscope.

Colony identity was also confirmed by flow cytometry, where individual colonies were collected and stained with anti-human CD15-FITC, anti-human CD14-PE and anti-human CD235a-PE Cy5 (Table A.1).

2.8 Humanised Ossicle Xenotransplantation Assay

To assess the *in vivo* differentiation potential of normal HSPC, a humanised ossicle xenotransplantation assay was used, as previously described [57]. Firstly, mesenchymal stromal cells (MSCs) were derived from primary BM samples. BM MNCs were isolated from BM-collection filters by flushing filters with pre-warmed PBS. MNCs were centrifuged at 300 g for 7 min at 4°C, resuspended in α-MEM media (Sigma-Aldrich, St. Louis, MO) with 10% pooled human platelet lysate (pHPL) and cultured. To remove non-adherent cells from the culture, culture vessels were washed with PBS and adherent cells expanded. The adherent cells demonstrated CFU-fibroblast (CFU-F) ability and were cultured to expand the MSC population.

To create a humanised ossicle, MSCs were harvested and resuspended in 60 µl of pHPL and combined with 240 µl of Matrigel-equivalent matrix. This Matrigel-cell mixture was injected subcutaneously into immunodeficient non-obese diabetic (NOD) SCID gamma (NSG) mice. Transplanted ossicles were analysed 8–10 weeks post-injection to assess bone and BM formation. Mice in which successful ossicle niches developed were used for xenotransplantation of primary human BM/CB HSPC populations.

Mice were pre-treated with 200 rads of irradiation 12–24 hours prior to transplantation. HSPC populations were transplanted in a total volume of 20 µl by direct intra-ossicle injection. Experiments were performed in accordance with a protocol approved by Stanford's Administrative Panel on Laboratory Animal Care (no. 22264) and in adherence to the US National Institutes of Health's Guide for the Care and Use of Laboratory Animals. Human engraftment was assessed 1–2 weeks post-transplantations by PB and/or BM sampling. Isolated cells were stained with anti-human CD14-PE or anti-human CD14-APC Cy7, anti- human

CD15-FITC, anti-human HLA-ABC-FITC or anti-human HLA-ABC-PB, anti-human CD19-APC, anti-human CD33-PE and anti-human CD45-V450 (Table A.1). Human engraftment was defined by either $CD33^+$ and/or $CD19^+$ cells amongst human $CD45^+$ HLA-APC^+ cells.

2.9 Leukaemic Xenotransplantation Assay

To assess the leukaemia-propagating potential of leukaemic populations, an NSG xenotransplantation assay was used. All animal experiments were performed in accordance with UK government approved project license PA7C92A40. Firstly, leukaemic MNCs/$CD34^+$ cells were thawed and stained for HSPCs as described above (see Chapter 2.3) using the following reduced antibody cocktail against lineage markers: CD2, CD3, CD4, CD8a, CD19, CD20 and CD235a. Leukaemic samples were then flow sorted for expanded leukaemic populations. Flow sorted cells were resuspended in PBS with 1% FBS, in a maximum volume of 40 µl per injection. Xenotransplantations were carried out using ~8–10 week-old female NSG mice. 24 hours prior to injection, mice were sub-lethally irradiated with two 100 cGy doses, at least 4 hours apart. Following irradiation, mice were provided with drinking water supplemented with Baytril Oral Solution (Hillcrest Animal Hospital). Transplantations were done via intra-osseous injection into the tibia of mice, using a 0.5 ml Insulin Syringe & Needle 30g x 8 mm (BD Biosciences, UK).

Mice were monitored for the development of human leukaemia by both PB and BM sampling. PB samples were taken at ~12 and ~16 weeks post-transplantation and BM samples were taken at ~20 weeks post-transplantation from the non-injected leg and ~24 weeks post-transplantation from the injected leg. Mice were culled following the development of expanded human leukaemia (as assessed by flow cytometry) or following the ~24 weeks post-transplantation BM sample.

PB samples were collected using heparinised capillary tubes (Hawksley, UK). BM samples were collected from the tibia of the mice by aspiration into an insulin syringe. PB/BM samples were transferred to 100 µl PBE post-collection and then washed with an additional 1 ml of PBE. Cells were resuspended in 1 ml of red blood

cell lysis buffer (0.15 M NH_4Cl, 10 mM $KHCO_3$, 0.1 mM EDTA) and incubated at room temperature for 5 min. 1 ml of flow cytometry buffer was then added to cells, and cells spun down at 300 g for 5 min. Cells were then washed in 1 ml of flow cytometry buffer and stained for flow cytometry analysis. Staining was carried out as described above (see Chapter 2.3), however using the following antibodies: anti-mouse CD45 APC, anti-human CD45 APC-eFluor 780, anti-human CD33 PE, anti-human CD19 FITC (Table A.1).

Following culling of engrafted mice, BM was harvested from the tibia, femur, pelvic and back bones of all mice. Collected bones were crushed in flow cytometry buffer and filtered using a 100 µm Corning Sterile Cell Strainers (ThermoFisher Scientific). Cells were spun down at 300 g for 5 min and then red blood cell lysed as described above for PB/BM samples. Bulk BM was analysed by flow cytometry using the above antibodies for PB/BM samples and then frozen in FBS with 10% dimethyl sulfoxide (DMSO).

2.10 Genotyping of Leukaemic Samples

Leukaemic samples were genotyped by targeted re-sequencing using the Fluidigm Access Array (Fluidigm), as previously described [37].

2.11 Bulk RNA-Seq of HSPC Populations

To analyse gene expression profiles of normal and leukaemic HSPCs, bulk RNA-sequencing (RNA-Seq) was used. For normal HSPCs, both fresh CB and thawed BM samples were used. For leukaemic HSPCs, leukaemic samples were thawed from frozen. All samples were processed and stained as described above (see Chapters 2.1, 2.2 and 2.3). ~100 HSPCs were sorted by flow cytometry directly into a lysis buffer containing RNAse inhibitor (Clontech St Germain-en-Laye France) and stored at -80°C until further processing and library preparation. cDNA synthesis was done using Smarter Ultra low input RNA kit v1 (Clonetech) as previously described [58]. Illumina libraries were generated using a Nextera XT DNA sample preparation kit and Index Kit (Illumina Chesterford UK). Library size and quality

were checked using Agilent High-Sensitivity DNA chip with Agilent Bioanalyser (Agilent Technologies Stockport UK). The concentration of indexed libraries was determined using a Qubit High-Sensitivity DNA kit (Invitrogen Loughborough, UK).

For libraries prepared from normal CB HSPCs, libraries were pooled to a final concentration of 5–14 nM and sequenced on an Illumina HiSeq 2000 using single-end 50 bp reads. For libraries prepared from normal and leukaemic BM HSPCs, libraries were pooled to a final concentration of 5–14nM and sequenced on an Illumina HiSeq 4000 using 75 bp paired end reads.

2.12 Single-Cell RNA-Seq of HSPC Populations

To analyse gene expression profiles of normal single HSPCs, single-cell RNA-Seq was used. Fresh CB samples were processed and stained as described above (see Chapters 2.1, 2.2 and 2.3). Single HSPCs were flow sorted directly into a lysis buffer containing RNAse inhibitor (Clontech St Germain-en-Laye France). Single-cell libraries for RNA-Seq were prepared using the Smart-seq2 protocol [59], where 23 cycles were used for the cDNA library pre-amplification. Illumina Nextera XT DNA sample preparation kit and Index Kit (Illumina Chesterford UK) was used for cDNA tagmentation and indexing. ERCC RNA Spike-In Mix (Ambion) was added to the lysis mix at a final dilution of 1:80,000,000. Library size, quality and concentration were assessed as was done for the bulk RNA-Seq. Libraries were pooled to a final concentration of 7–28 nM, and 78–95 single-cell libraries were combined per pool. Sequencing was done on HiSeq4000 using 75 bp paired-end reads. Each pool contained a library generated from an empty well.

2.13 Gene Expression Analysis by Multiplexed qRT-PCR

Single-cell, multiplexed qRT-PCR was carried out on normal HSPCs using the 96.96 Dynamic Array Integrated Fluidic Circuits (IFC) for Gene Expression (Fluidigm). Gene expression analysis was performed as described [58, 60]. TaqMan assays (Applied Biosystems) are listed in [54], Supplementary Tables 2 and 14.

2.14 Bioinformatic Analysis

2.14.1 Bulk RNA-Seq Data from Normal CB

50-bp single-end reads were aligned to the hg38 reference genome using TopHat v2.0.10 [61]. Alignments were processed using Picard tools (`http://picard.sourceforge.net/`). Sequencing reads were filtered for mapq 4, to include only uniquely mapping reads. This gave a range of $15.1\text{x}10^6$ to $56.2\text{x}10^6$ aligned reads. The total number of genes expressed per sample was calculated as reads per kilobase of transcript per million mapped reads (RPKM) > 1. The number of expressed genes ranged from 7,707 to 11,350, with an average of 9,800.

The count matrix was transformed to $\log_2$ counts per million (cpm) scale, and principal-component analysis (PCA) was carried out. One biological replicate of the MPP population was excluded because its global gene expression showed more uniquely expressed genes than in the three remaining MPP biological replicates, and it showed low correlation to the other three replicates. In order to identify DEGs across all HSPC populations, an ANOVA-like test was performed, using the edgeR package for R. The DEGs were ranked by their significance (P value, adjusted for multiple testing), and different numbers of genes were used for the PCA and hierarchical clustering. Eigenvalues from the PCA were calculated as the square of the standard deviation of the principle components. Differential gene expression for one versus one and one versus all comparisons were calculated using edgeR. For gene-signature generation, genes were filtered to have at least a one-fold change in expression and were ranked by P value. Heatmaps and associated hierarchical clustering were generated using GENE-E software (Broad Institute) or using the R packages pvclust and heatmap.2 (gplots). MetaCore Pathway Map enrichment analysis (Thomson Reuters, London UK) was carried out using DEGs by each LM population versus all other populations (one versus all). A P-value threshold of 0.05 was used to identify positively enriched pathway maps.

2.14.2 Single-Cell qRT-PCR Data from Normal CB

Analysis of single-cell qRT-PCR Biomark data was performed in R version 3.3.1 using data exported from the Fluidigm Data Collection software. For quality control, amplification curves with a Quality Score of < 0.65, and any Ct values of > 27 were treated as undetected expression. Any cells in which expression of both housekeeping genes (*B2M* and *GAPDH*) was not detected were removed from further analysis (n = 7 cells). An additional cell was removed as it had a large outlying number of genes detected. The housekeeping gene *ACTB* was also measured in the assay, but unlike *B2M* and *GAPDH*, it did not show robust expression across the majority of cells, and therefore was not used in further analysis. Normalised ΔCt values were calculated by subtraction of the mean of Ct values for *B2M* and *GAPDH* in each cell, as previously described [62]. Housekeeping genes were excluded from further analysis. Genes detected in < 20 cells, with a variance of < 1 across all cells, or expressed in none of the ten-cell control samples, were removed from downstream analysis. Post-quality-control data measured 74 genes in 919 single cells.

Hierarchical clustering was performed on genes and cells by using the hclust function (stats package) with distance measure 1-Spearman's correlation and agglomeration method Ward.D2. The heatmap visualising the clustering was plotted using the heatmap.2 function (gplots package). Cells were divided into three clusters using the cutree function (stats package) on the hierarchical clustering. A gene was classed as expressed differentially by two clusters if it satisfied two criteria. Firstly, the absolute $\log_2$ fold change ($|\ \log_2\text{FC}\ |$) of the mean ΔCt in each cluster needed to be > 1. Secondly, the adjusted P value (Benjamini & Hochberg correction for multiple testing) of two-sided Wilcox test of ΔCt expression values between the two clusters needed to be < 0.01. Diffusion maps [63] were used for dimensionality reduction of the single-cell gene-expression data. This method was implemented using the DiffusionMap function from the destiny R package, using Euclidean distance [64, 65].

2.14.3 Single-Cell RNA-Seq Data from Normal CB

Single-cell RNA-Seq reads were aligned using G-SNAP [66], and mapped reads were assigned to Ensembl genes (release 81, [67]) using HTSeq [68]. The following cells were removed from further analysis: cells with fewer than 500,000 reads mapping to nuclear genes, greater than 20% of mapped reads mapping to mitochondrial genes, or greater than 20% of mapped reads mapping to ERCC spike-ins, or with expression of fewer than 750 different genes with at least ten counts. ERCC spike-in controls identified genes exceeding technical variance [69].

From donors 1 and 2, 163 of 166 cells and 157 of 249 cells, respectively, passed quality control. Single-cell profiles were normalised using the scran R package [70], and variable genes were identified as having variation exceeding technical levels [69]. Data showed batch effects between different donors. The Seurat R package (`https://github.com/satijalab/seurat`) was then used to regress out plate effects from the sequencing data and set more stringent thresholds for variable genes, leading to 1,605 variable genes in donor 1 and 1,273 variable genes in donor 2. PCA was performed using Seurat, and clusters were found using the Seurat::FindClusters function on the first ten principal components. Heatmaps displayed the top genes marking these clusters as identified by the Seurat::FindAllMarkers function and were visualised using the gplots::heatmap.2 function.

2.14.4 Bulk RNA-Seq Data from Normal and Leukaemic BM

The quality of the 75-bp paired-end reads was assessed using FastQC (`https://www.bioinformatics.babraham.ac.uk/projects/fastqc/`) and reads trimmed for Nextera transposase sequences and over-represented sequences using Trimmomatic [71]. Reads were mapped to the hg38 genome using HISAT2 [72] and using Ensembl 86 annotations. Mapped reads were quantified using featureCounts [73]. Reads were also pseudoaligned and quantified using Kallisto [74] and all downstream analysis was using Kallisto processed reads. Differential gene expression was carried out using DESeq2 [75].

Genes with fewer than 5 counts across all samples were discarded, and dispersion factors were estimated using a local fit. Gene counts were normalised using both a variance stabilising transformation (VST) and rlog transformation, and VST normalised data was used for further analysis. To identify DEGs across the normal BM HSPC populations, a likelihood ratio test (LRT) was carried out. Plots were generated using ggplot2 [76] and Scatterplot3d [77] and heatmaps were generated using pheatmap (`https://cran.r-project.org/package=pheatmap`). PCA plots and heatmaps were generated using all expressed genes, genes filtered on statistical variance and genes filtered based on the LRT. Hierarchical clustering was carried out using correlation for clustering distance and ward.D2 method in pheatmap. DEGs were identified between normal and leukaemic samples using the Wald test, where a p-adjusted value (Benjamini-Hochberg correction for multiple testing) of 0.05 was used for significance.

To identify DEGs with a potentially functional role in leukaemogenesis, the above DEGs were integrated with published datasets. Two genome-wide CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) screens in leukaemic cell lines were used [78, 79]. Genome-wide screens in non-leukaemic cell lines were used as a control [80, 81]. Published CRISPR gene scores (CS) [78], CRISPR dropout false discovery rates (FDRs) [79] and Bayes factors (BF) [81–83] were subset for DEGs identified between normal and leukaemic samples (p-adjusted<0.05 and $\mid \log_2 FC \mid > 2$). CS/FDR/BF values were displayed using a heatmap and hierarchically clustered to identify clusters of genes that had the most significant CS/FDR/BF. This identified six genes that were both differentially expressed in the RNA-Seq data and functionally essential in the genome-wide CRISPR-Cas9 screens.

2.15 *In Vitro* Culture of Leukaemic Samples

Preliminary optimisation of the *in vitro* culture of AML samples was carried out. AML cells were cultured both in the presence and absence of the immortalised human BM derived MSC cell line (HTERT-MSC [84]). HTERT-MSCs were maintained

in the following media: Roswell Park Memorial Institute (RPMI), 1% penicillin-streptomycin, 1% l-glutamine, 10% heat inactivated FBS and 1.0 µM hydrocortisone. AML cultures were carried out in the following cytokine supplemented media: IMDM, 15% BIT (bovine serum albumin, insulin, transferrin; Stem Cell Technologies 09500), 100 ng/ml SCF, 50 ng/ml FLT3L, 20 ng/ml IL-3, 20 ng/ml G-CSF (Peprotech London UK), 10^{-4}M β-mercaptoethanol and 1% penicillin-streptomycin (adapted from [85]). AML cells were cultured both in the presence and absence of SR1 and UM729/UM171 and/or dimethyloxaloylglycine (DMOG) as previously described ([85] and [86] respectively).

24 h prior to the addition of AML cells, 50,000 HTERT-MSCs were plated in 96-well plates, in HTERT-MSC maintenance media. Then either 10,000 or 50,000 AML cells were plated in the above cytokine supplemented media in the presence or absence of: SR1 and UM729/UM171 and/or DMOG. AML cells were analysed by flow cytometry prior to culture and then at multiple time-points, ranging from 4 to 15 days, post-culture.

The harvesting of cells for flow cytometry analysis was done as described above (see Chapter 2.4). Staining was carried out as described above (see Chapter 2.3) however using combinations of the following antibodies: anti-human CD34 PerCP, anti-human CD38 FITC/BV421, anti-human CD45RA PE, anti-human CD33 APC Cy7, anti-human CD45 BV650, anti-human Annexin-V APC, anti-human CD11b APC Cy7, anti-human CD14 PE, anti-human CD15 eFluor 450, and anti-mouse CD45 BV605 (Table A.1). HTERT-MSCs were distinguished by their expression of green fluorescent protein (GFP).

2.16 Design and Cloning of CRISPR-Cas9 sgRNAs

To functionally validate the six genes identified above (see Chapter 2.14.4), CRISPR-Cas9 was used. Benchling software was used to design small guide (sg) RNAs (`https://benchling.com/crispr`), which uses on-target (efficiency) and off-target (specificity) scores to rank sgRNAs [87, 88]. sgRNAs were designed with the aim of introducing insertions and/or deletions (indels) into a coding exon of

each gene, with the goal of knocking-out the functional protein. sgRNAs were chosen based on their on-target (efficiency) score and five sgRNAs were tested for each gene to assess editing (Table A.3).

sgRNAs were cloned into the PX458 plasmid (Addgene plasmid 48138) as previously described (`https://www.addgene.org/crispr/zhang/`). Briefly, oligonucleotides were designed by Benchling for cloning into PX458 and synthesised (ThermoFisher Scientific) (Table A.4). 1 µg of PX458 plasmid was digested in a total reaction volume of 20 µl using 1 µl of *Bbs*I restriction enzyme (Fermentas, Thermo Fisher), 1 µl of FastAP Thermosensitive Alkaline Phosphatase (Fermentas, Thermo Fisher) and 2 µl of 10X FastDigest Buffer (Fermentas, Thermo Fisher) at 37°C for 30 min. Oligonucleotides were then phosphorylated and annealed using 1 µl of 100 µM of each oligonucleotide (forward and reverse), 1 µl of 10X T4 Ligation Buffer (New England Biolabs), 6.5 µl nuclease-free water (Thermo Fisher) and 0.5 µl T4 Polynucleotide Kinase (New England Biolabs). Phosphorylation and annealing was done at 37°C for 30 min, 95°C for 5 min and then cooled down to 25°C at the rate of 5°C per min. Phosphorylated oligonucleotides were then ligated into the digested plasmid using 50 ng of digested plasmid, 1 µl of 1:200 diluted oligonucleotide duplex, 5 µl 2X Quickligation Buffer (New England Biolabs), 1 µl Quick Ligase (New England Biolabs) in a total reaction volume of 11 µl at room temperature for 10 min.

Competent *Escherichia coli (E. coli)* were heat-shock transformed with PX458 plasmids containing sgRNA as follows. Bacteria and plasmid were incubated on ice for 30 min and then heat-shocked at 42°C for 45 s. The bacteria-plasmid mix was then returned to ice for 2 min before spreading onto agar plates with ampicillin. Transformed bacteria were grown overnight at 37°C and colonies picked and grown overnight at 37°C in lysogeny broth (LB). DNA was purified using a QIAprep Spin Miniprep Kit (Qiagen) and sequenced to identify clones containing the sgRNA (see Table A.5 for sequencing primers).

2.17 CRISPR-Cas9 Editing by Transfection of K562 Cells

The K562 leukaemic cell line was maintained in a maintenance media containing RPMI, 1% penicillin-streptomycin, 1% l-glutamine and 10% heat inactivated FBS. To transfect K562 cells the following solutions were used (as previously described [89]). 2 g of ATP · disodium salt and 1.2 g of $MgCl_2 \cdot 6H_2O$ were dissolved in nuclease free water (Thermo Fisher) and 20 µl aliquots stored at -20°C (termed 'Solution I'). Additionally, 6 g of KH_2PO_4, 0.6 g of $NaHCO_3$ and 0.2 g of glucose were dissolved in 500 ml of nuclease free water (Thermo Fisher) at a pH 7.4 and 1 ml aliquots stored at 4°C (termed 'Solution II'). For transfection, 20 µl of Solution I was mixed with 1 ml of Solution II (termed 'Solution I-II'). Per transfection reaction, $2.5x10^5$ cells were resuspended in 20 µl of Solution I-II and added to 1 µg of PX458 plasmid containing sgRNAs. Cells were transfected using the FF-120 programme on a 4D-Nucleofector System (Lonza) and then transferred to pre-warmed K562 maintenance media. 72 hours post-transfection, cells were analysed for GFP expression by flow cytometry and/or pelleted for the purification of DNA.

DNA was harvested using either QuickExtract DNA Extraction Solution (Lucigen) or DNeasy Blood & Tissue Kit (Qiagen). QuickExtract DNA Extraction Solution was used as follows: cell pellets were resuspended in 50 µl of Extraction Solution and incubated for 10 min at 65°C and then 5 min at 98°C. Putatively edited regions were amplified using 0.5 µl-1 µl of purified DNA, 1 µl of forward and 1 µl of reverse primer (Table A.5), and 12.5 µl or 25 µl of DreamTaq Green PCR Master Mix (2X) (Thermo Fisher Scientific) in a total reaction volume of either 25 µl or 50 µl respectively. Desired PCR products were isolated by gel extraction using a QIAquick Gel Extraction Kit (Qiagen). PCR products were sequenced, and putatively edited sequences were compared to a control, non-edited sequence using Inference of CRISPR Edits (ICE [90]) software to assess indel frequency.

2.18 Lentiviral Transduction of Primary Cells

With the aim to edit primary cells by CRISPR-Cas9, the lentiviral transduction of primary HSPCs was tested. The pL-CRISPR.SFFV.GFP plasmid (Addgene, plasmid 57827, [91]) was used and lentivirus was titred in HEK 2937 cells to a titre of $4x10^7$ TU/ml. AML and normal CB cells were plated in 96-well plates to a density of 10,000 cells per well. AML cells were cultured in conditions defined above (see Chapter 2.15), and normal CB was cultured in the following conditions: StemSpan Serum Free Expansion Media (Stem Cell Technologies), 100 ng/ml SCF, 100 ng/ml FLT3L and 100 ng/ml TPO. 24 hours following plating of primary cells, lentivirus was added to cells at a multiplicity of infection (MOI) of 0–100, cells spinfected at 300 g for 1 hour at 37°C and then returned to culture. 72 hours post-transduction with lentivirus, cells were harvested and analysed by flow cytometry. Flow cytometry staining was done as described above (see Chapter 2.3) using the following anti-human antibodies: CD34 APC, CD11b APC Cy7 and CD45 BV650, with GFP positivity indicating the extent of transduction.

Chapter 3

Characterisation of Normal Human Lympho-Myeloid Progenitors

Contents

3.1 Introduction to Human LM Progenitors 37
3.2 Immunophenotypic Characterisation of Normal HSPC 40
3.3 *In Vitro* Assays of LM Progenitors 42
3.3.1 Colony Forming Unit Assay of HSPCs 42
3.3.2 Bulk *In Vitro* Cultures of LM Progenitors 44
3.3.3 Limiting Dilution Assay of LM Progenitors 50
3.3.4 Single-Cell *In Vitro* Cultures of LM Progenitors 54
3.4 *In Vivo* Assays of LM Progenitors 64
3.5 Bulk RNA-Seq of Normal HSPCs 69
3.5.1 PCA and ANOVA Analysis of HSPC Populations 71
3.5.2 Transcriptional Distance Between HSPC Populations . . 77
3.5.3 Expression of Lineage- and LM-Associated Gene Signatures 78
3.6 Single-Cell Transcriptional Analysis of LM Progenitors 87
3.6.1 Single-Cell qRT-PCR of LM Progenitors 87
3.6.2 Single-Cell RNA-Seq of LM Progenitors 90
3.7 Refining the LMPP and GMP Populations 98
3.7.1 Refining the LMPP: $LMPP^{ly}$ and $LMPP^{mix}$ 98
3.7.2 Refining the GMP: GMP $CD38^{hi}$ 108
3.8 Revised Model of Human LM Differentiation 116
3.9 Conclusion on Human LM Progenitors 117

All work in Chapter 3 has been adapted from its original publication by Nature Immunology, see Appendix A.1 and [54]. All work in Chapter 3 was done in conjunction with Bilyana Stoilova and Dimitris Karamitros: BS and DK are joint first authors for this work and I am second author for this work.

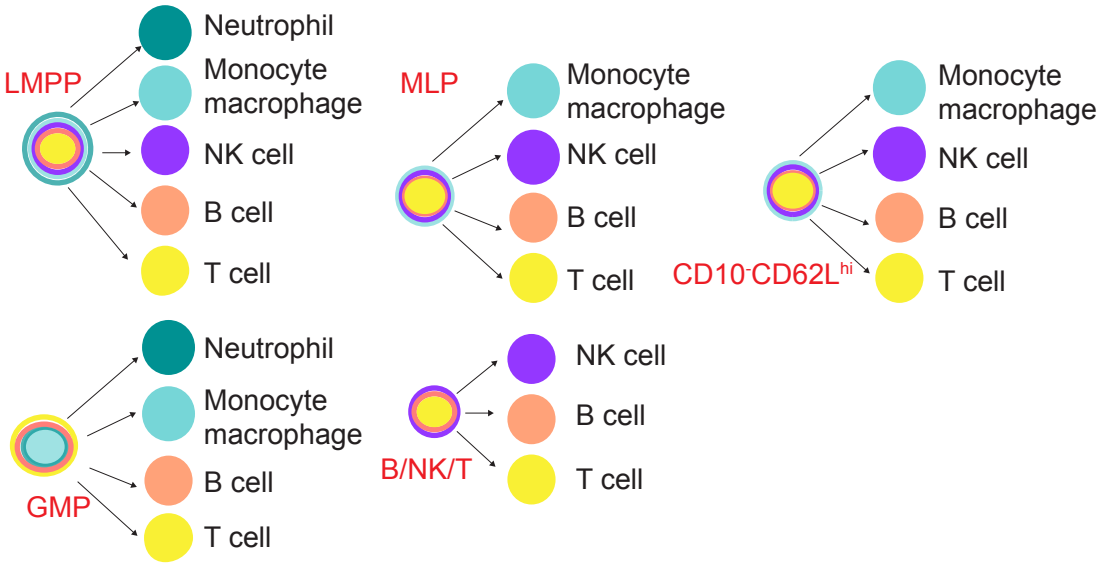
3.1 Introduction to Human LM Progenitors

At the apex of the haematopoietic hierarchy lies the HSC and MPP populations. These populations have the ability to give rise to all the cells of the blood and immune system. Downstream of the HSC and MPP, progenitors with more

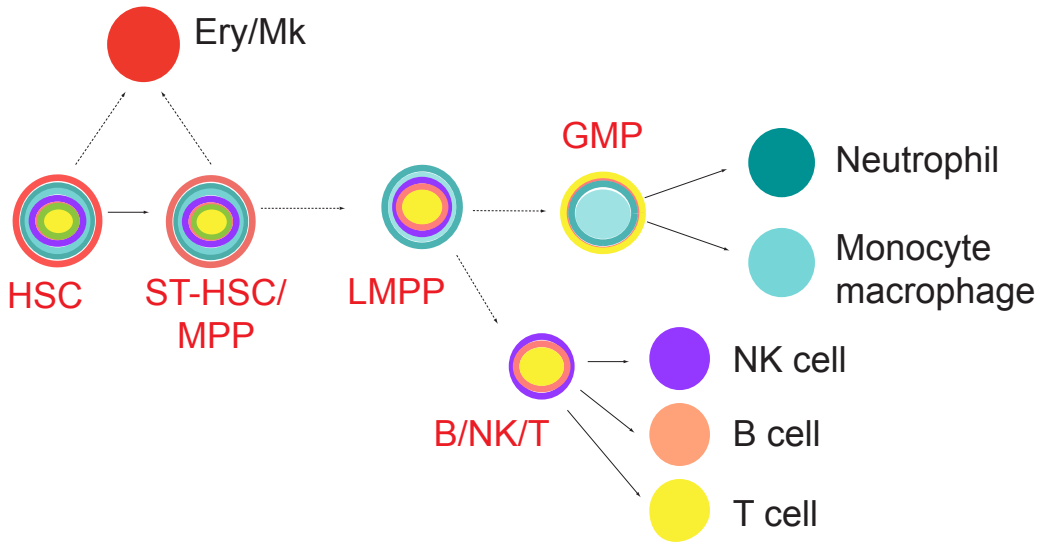
restricted fate towards the lymphoid and myeloid lineages can be found. Firstly, the $Lin^-CD34^+CD38^-CD90^{neg\text{-}lo}CD45RA^+CD10^-$ LMPP has previously been described to have granulocytic, monocytic, B cell and T cell potential, but is unable to generate erythrocytes or megakaryocytes [22]. Additionally, the MLP has been reported initially as Lin^- $CD34^+$ $CD38^-$ $CD90^{neg\text{-}lo}$ $CD45RA^+$ $Flt3^+$ $CD10^+$ $CD7^{-/+}$ cells and with lymphoid (B cell, T cell and NK cell), monocytic and DC potential but not granulocytic [24].

Within the $CD38^+$ compartment, LM progenitors have also been described. Kohn et al. have described a $CD62L^{hi}CD10^-$ LMPP with lymphoid, monocytic and DC potential [29]. Additionally, there is the well-studied granulocyte-monocyte progenitor (GMP) ($Lin^-CD34^+CD38^+CD45RA^+CD123^+$). The GMP is described as primarily a myeloid progenitor, but with residual lymphoid ability [22, 92]. Finally, there is a $Lin^-CD34^+CD38^+CD45RA^+CD10^+$ lymphoid-restricted progenitor with T cell, B cell, NK cell and DC potential [93]. Figure 3.1a shows a summary of the reported LM progenitors and their described lineage output.

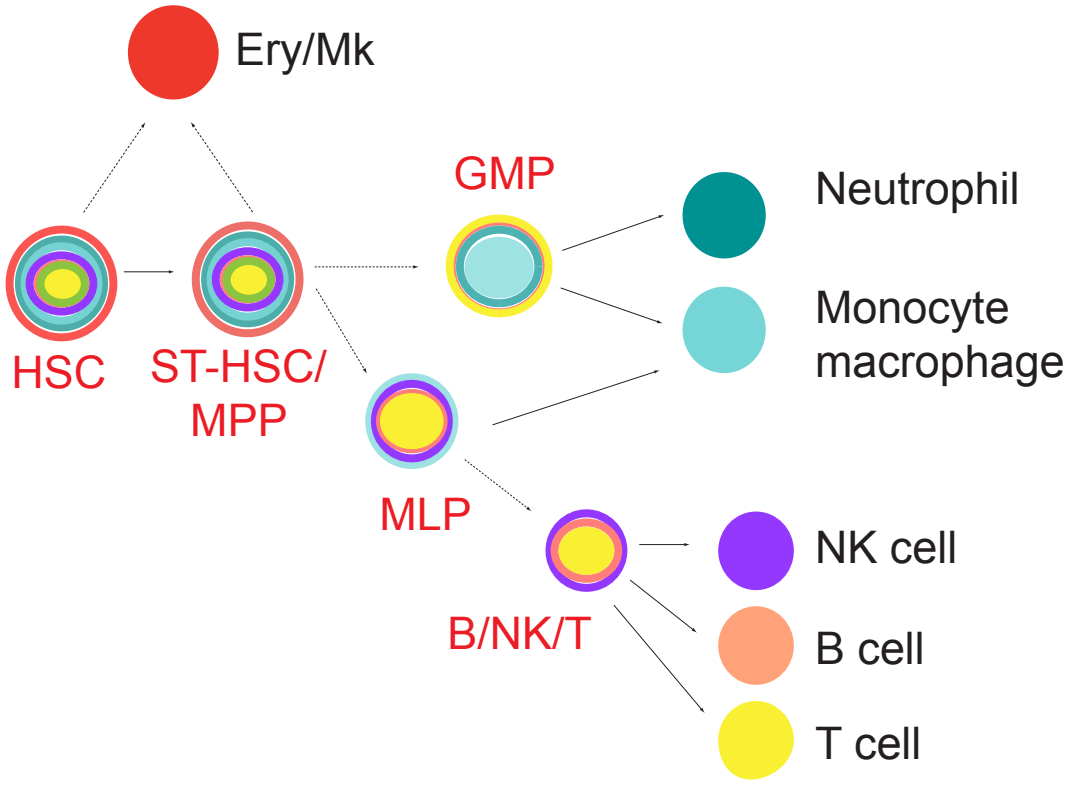
However, recently it has become apparent that clarity is lacking in the characterisation of the LM progenitors. For example, MLP populations which overlap with the LMPP population have been described [26–28]. Additionally, Laurenti et al. have carried out a characterisation of lymphoid commitment; however, only the MLP was included in this analysis [25]. Figures 3.1b and 3.1c describe two potential models of LM differentiation that could be concluded from the literature. Therefore, we aimed to fully characterise the LM progenitors found in human haematopoiesis. Firstly, the functional output of the LM progenitors was analysed, both *in vitro* and *in vivo*. This was done both at the population and the single-cell level, in order to address the heterogeneity of these populations. Following on from this, the gene expression profiles of the LM progenitors were characterised, again at both the population and single-cell level. This allowed for the refinement of the immunophenotypic definitions of the LM progenitors, purifying for functional potential.



(a) Lineage Output LM Progenitors



(b) Proposed Model LM Differentiation 1



(c) Proposed Model LM Differentiation 2

Figure 3.1: Current Models of Human LM Differentiation. (a) Lineage output of lympho-myeloid progenitors. (b–c) Models of human haematopoietic differentiation and lineage diversification. Dashed arrows - hierarchical relationships not experimentally validated.

3.2 Immunophenotypic Characterisation of Normal HSPC

Firstly, the immunophenotype of the eight HSPC populations was investigated. These populations were defined as in Table 3.1, and it was seen by flow cytometry analysis that they were separable and distinct (Figure 3.2). These populations accounted for 82% of cells and the remaining cells did not compose a distinct population.

In both the $CD38^-$ and $CD38^+$ compartments, $CD10^-$ and $CD10^+$ populations could be seen. In the $CD38^-$ compartment these were the $CD10^+$ MLP and $CD10^-$ LMPP. In the $CD38^+$ these were the $CD10^+$ B-NK progenitor and $CD10^-$ GMP.

The LMPP and MLP populations were rare, being only 0.2% of the BM

Lin$^-$CD34$^+$ compartment (Table 3.1). This is the equivalent of ~2 cells among 1 x 10^5 BM MNCs. These populations were marginally more frequent in CB, being found at the frequency of ~1 cell among 1 x 10^4 CB MNCs. On the other hand, GMPs were 20-fold more abundant in CB and 100-fold more abundant in BM than the LMPP or MLP (~1.5–2 cells among 1 x 10^3 MNCs). Due to the frequency of the LM progenitors, fresh CB was used for all downstream experiments, with cells' being double sorted to high purity (of > 99%, except CMPs, with a purity of 97%).

Table 3.1: Immunophenotypic Definitions of HSPC Populations. HSPC definitions for human CB and BM, as in [54]. Shown is the mean frequency (± SD) of Lin$^-$CD34$^+$ CB/BM. CB N=44 and BM N=7.

Population	Immunophenotype	CB: Analysis Gates	BM: Analysis Gates	CB: Sorting Gates
Stem and early progenitors	Lin$^-$ CD34$^+$ CD38$^-$	15.4±5.9	16.1±10.3	8.9±3.5
Stem and early progenitors	Lin$^-$ CD34$^+$ CD38$^-$ CD10$^-$	12.8±5.2	15.8±10.4	7.0±3.0
HSC	Lin$^-$ CD34$^+$ CD38$^-$ CD90$^+$ CD45RA$^-$ CD10$^-$	5.3±2.7	11.4±8.5	2.0±1.2
MPP	Lin$^-$ CD34$^+$ CD38$^-$ CD90$^-$ CD45RA$^-$ CD10$^-$	4.0±2.3	3.7±3.4	1.0±1.0
LMPP	Lin$^-$ CD34$^+$ CD38$^-$ CD90$^{-/lo}$ CD45RA$^+$ CD10$^-$	1.4±1.0	0.2±0.2	0.7±0.7
MLP	Lin$^-$ CD34$^+$ CD38$^-$ CD90$^{-/lo}$ CD45RA$^+$ CD10$^+$	1.1±0.8	0.2±0.1	0.6±0.6
Late Progenitors	Lin$^-$ CD34$^+$ CD38$^+$	84.6±5.9	83.9±10.3	41.4±11.4
Late Progenitors	Lin$^-$ CD34$^+$ CD38$^+$ CD10$^-$	78.7±7.0	78.4±8.5	38.0±11.1
CMP	Lin$^-$ CD34$^+$ CD38$^+$ CD123$^+$ CD45RA$^-$ CD10$^-$	37.7±11.3	34.7±6.9	9.1±6.1
GMP	Lin$^-$ CD34$^+$ CD38$^+$ CD123$^+$ CD45RA$^+$ CD10$^-$	16.8±6.1	17.8±6.1	6.2±3.1
MEP	Lin$^-$ CD34$^+$ CD38$^+$ CD123$^-$ CD45RA$^-$ CD10$^-$	13.9±7.5	14.6±4.9	4.1±2.6
B/NK	Lin$^-$ CD34$^+$ CD38$^+$ CD90$^-$ CD45RA$^+$ CD10$^+$	2.0±1.7	5.1±7.5	0.9±1.2

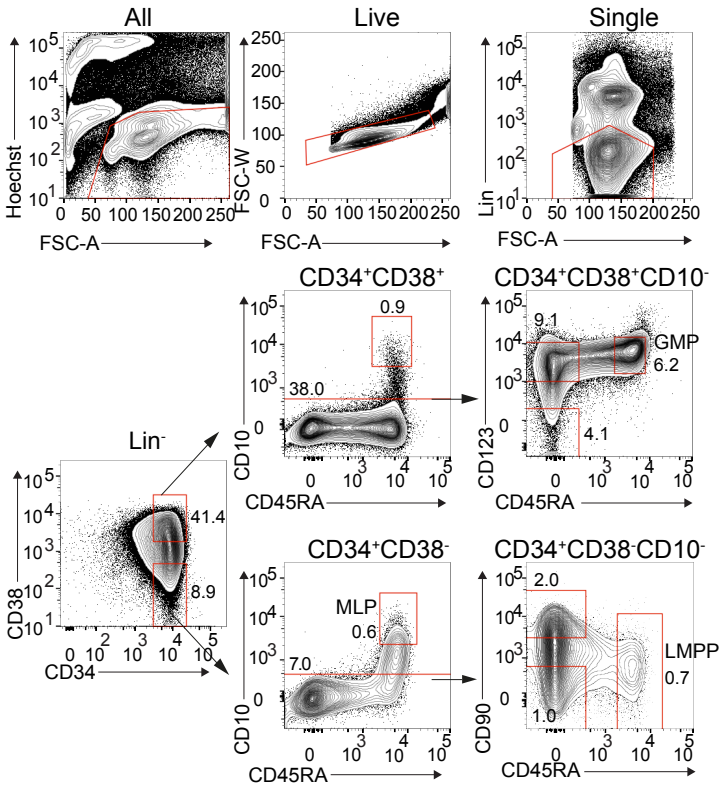
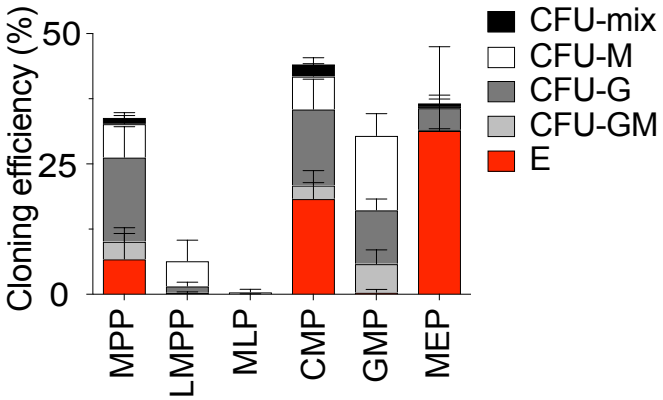


Figure 3.2: Immunophenotypic separation of human HSPCs. Representative plot of flow sorting of human CB HSPCs. Population frequencies shown are the mean from 44 CB donors calculated as a percentage of Lin$^-$CD34$^+$ compartment.

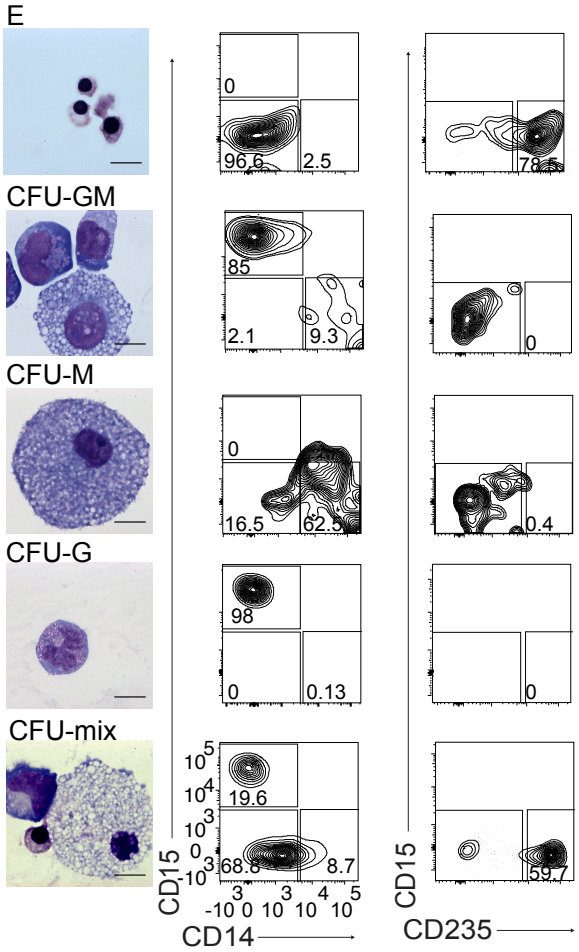
3.3 *In Vitro* Assays of LM Progenitors

3.3.1 Colony Forming Unit Assay of HSPCs

Firstly, the erythroid and myeloid potential of the LM progenitors was investigated in a methylcellulose-based CFU assay. It was seen that the LMPP and MLP had lower myeloid clonogenic potential (6% LMPPs and $<$ 1% MLPs) than that of GMPs (31%) (Figure 3.3a). The GMP and LMPP generated granulocyte colonies, monocyte-macrophage colonies and granulocyte and monocyte-macrophage colonies with no or minimal erythrocyte potential ($<$ 0.5%). The MLP had the ability to generate only a few monocytic colonies (Figure 3.3a). Colony identity was further confirmed by flow cytometry analysis and morphology analysis of cytospun colonies stained with May-Grünwald-Giemsa (Figure 3.3b).



(a) CFU Assay Output



(b) Morphology and Flow Cytometry Analysis of CFU Assay Output

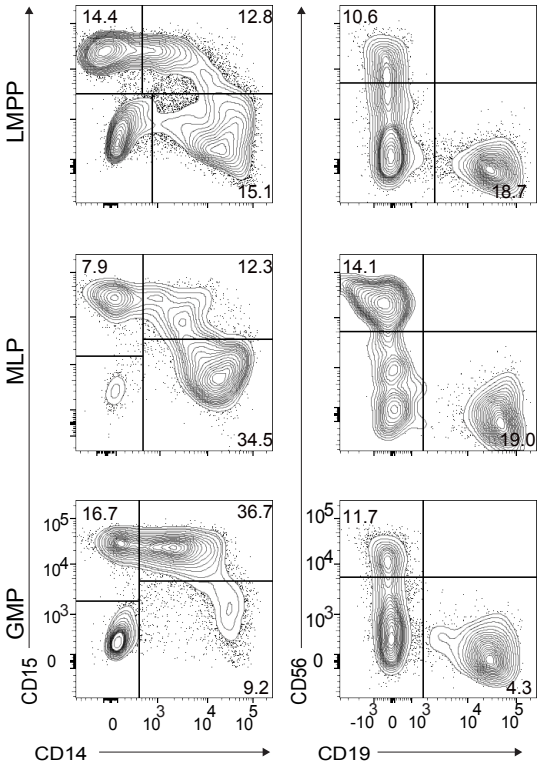
Figure 3.3: CFU Assay of HSPCs. CFU assay of the cloning efficiency and myelo-erythroid lineage output of CB HSPCs. 150 HSPC cells plated in methylcellulose and output measured after 2 weeks incubation. (a) Shown is the frequency of mixed erythro-myeloid colonies (CFU-mix), monocyte-macrophage colonies (CFU-M), granulocyte colonies (CFU-G), granulocyte and monocyte-macrophage colonies (CFU-GM) and erythroid colonies (burst-forming units-erythroid and CFU-E (E)). (b) Microscopy showing the morphology of cells from the CFU assay in (a), stained with May-Grünwald-Giemsa. On the right is shown the flow cytometry analysis of cells harvested from the colonies. Scale bar: 10 µm. Numbers indicate percent $CD15^{+}CD14^{-}$ neutrophil granulocytes, $CD15^{-}CD14^{+}$ monocyte-macrophages, $CD15^{-}CD14^{-}$ cells, or $CD15^{-}CD14^{-}CD235^{+}$ erythroid cells.

3.3.2 Bulk *In Vitro* Cultures of LM Progenitors

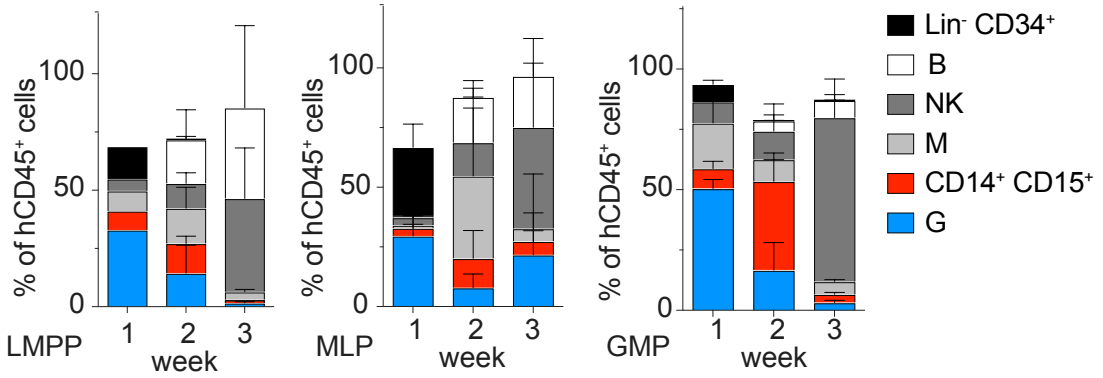
Next, the lymphoid and myeloid differentiation potential of the LM progenitors was investigated. Firstly, this was done at the population level, where 150 LMPPs, MLPs and GMPs were cultured in a novel *in vitro* liquid culture on MS-5 mouse stromal cells. The differentiation into the LM lineages was promoted by the addition of the following cytokines: SCF, G-CSF, FLT3L, IL-2 and IL-15, and the prostaglandin-synthesis inhibitor DUP-697. These conditions were termed "SGF15/2 culture" here, see Chapter 2.4 and Table A.2 for more information. Lineage output was detected by flow cytometry (Figure 3.4a).

To determine the optimal time-point for analysis, a kinetic analysis was carried out across 1–3 weeks of culture (Figure 3.4b). It was seen that at 2 weeks of culture, human $CD45^{+}$ $CD15^{+}$ ($hCD45^{+}$ $CD15^{+}$) neutrophil-granulocytes (Gs), $hCD45^{+}$ $CD14^{+}$ monocyte-macrophages (Ms), $hCD45^{+}$ $CD19^{+}$ B cells and $hCD45^{+}$ $CD56^{+}$ NK cells could all be detected. Furthermore, it was confirmed that the *in vitro* generated Gs, Ms, B cells and NK cells, purified by flow cytometry, expressed the appropriate lineage-associated genes (Figure 3.5). Therefore, all further cultures in SGF15/2 conditions were analysed following 2 weeks of culture.

The SGF15/2 culture conditions did not allow for the development of lymphoid T cells. Therefore, the T cell production by the LM progenitors was investigated utilising a different *in vitro* culture on OP9 stromal cells expressing the human Notch receptor ligand DL1 (OP9-hDL1 cells) and in the presence of SCF, FLT3L and IL-7. These conditions were termed "SF7a culture" here, see Chapter 2.4 and



(a) Flow Cytometric Analysis of Bulk SGF15/2 Culture



(b) Kinetics of Bulk SGF15/2 Culture

Figure 3.4: Bulk SGF15/2 Culture of LM Progenitors. Lineage output of LM progenitors in SGF15/2 culture conditions (see Chapter 2.4 and Table A.2 for culture conditions). 150 LMPPs, MLPs and GMPs were plated in SGF15/2 conditions and analysed by flow cytometry after 1, 2 or 3 weeks of culture. (a) Representative flow cytometric analysis of the B cell, NK, monocytic and granulocytic output following 2 weeks of culture. (b) Kinetics of lineage output of LM progenitors in SGF15/2 conditions at 1, 2, and 3 weeks of culture. Frequencies are the mean from 3 CB donors and calculated as a percentage of human $CD45^+$ cells.

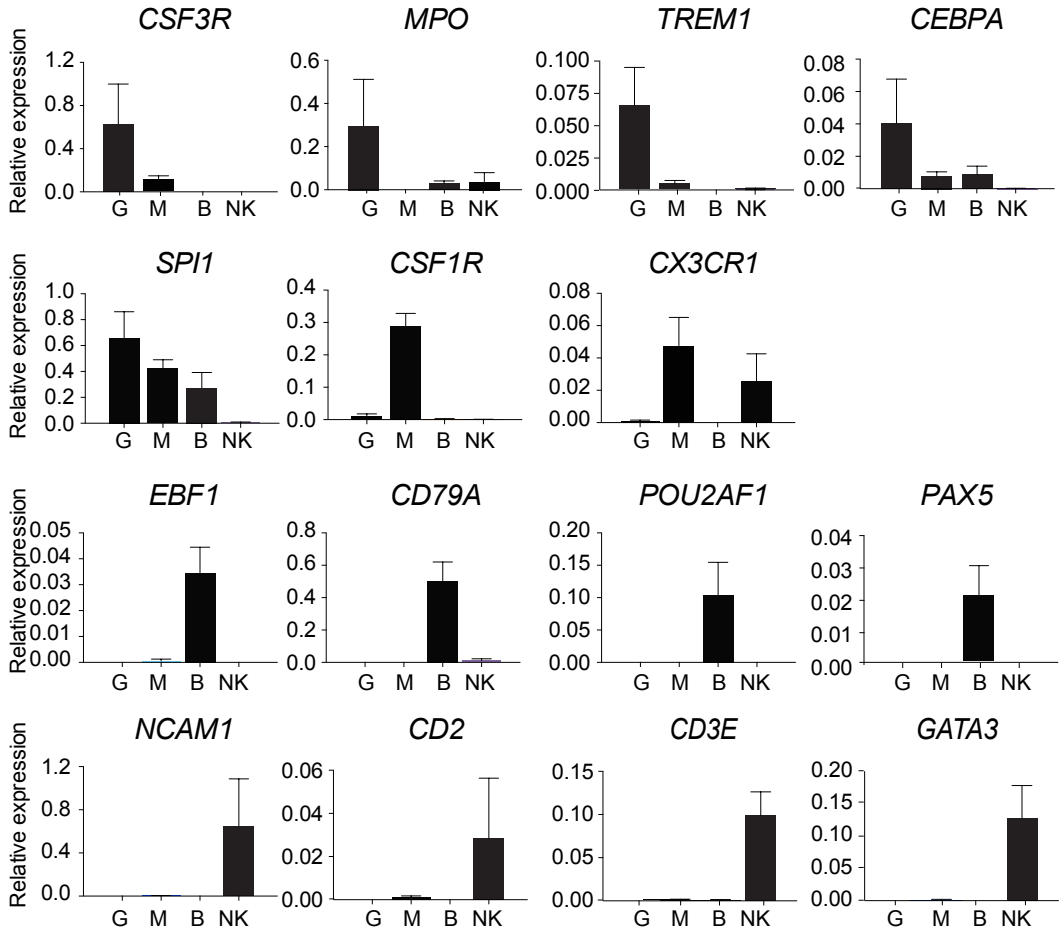
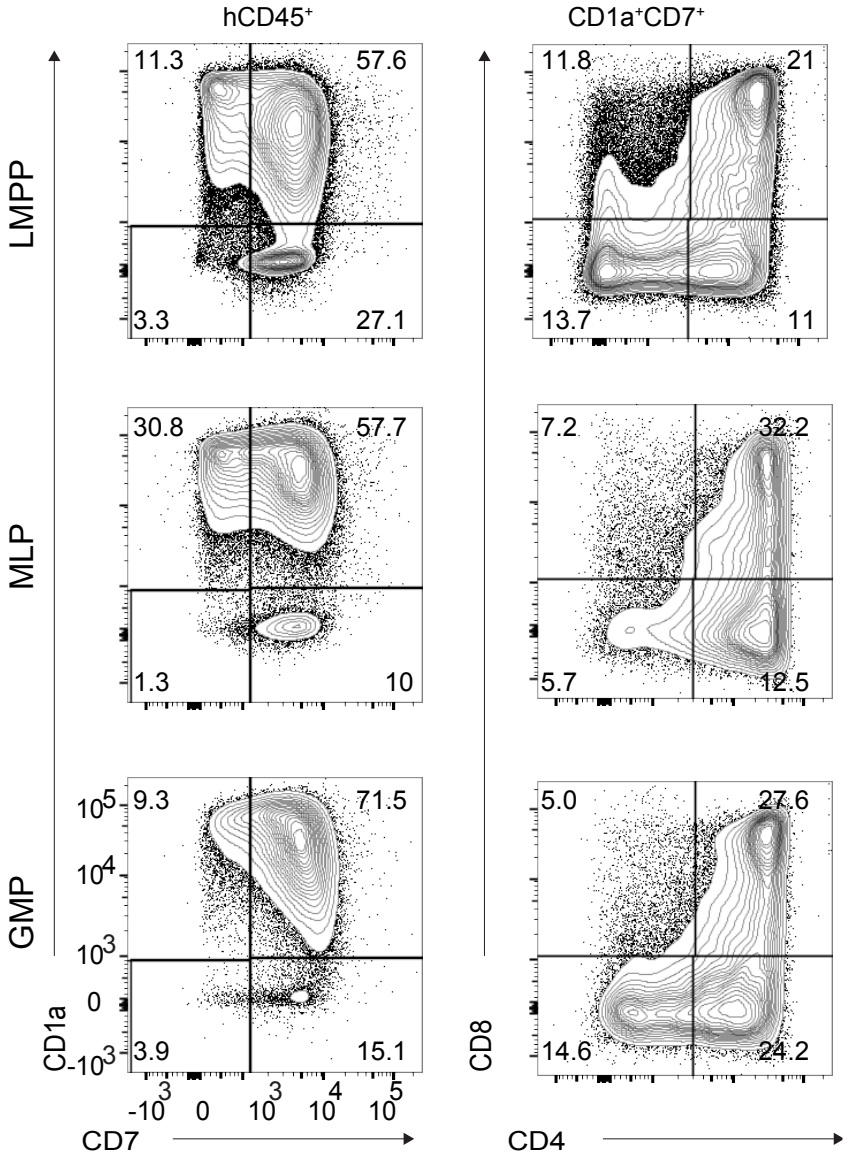


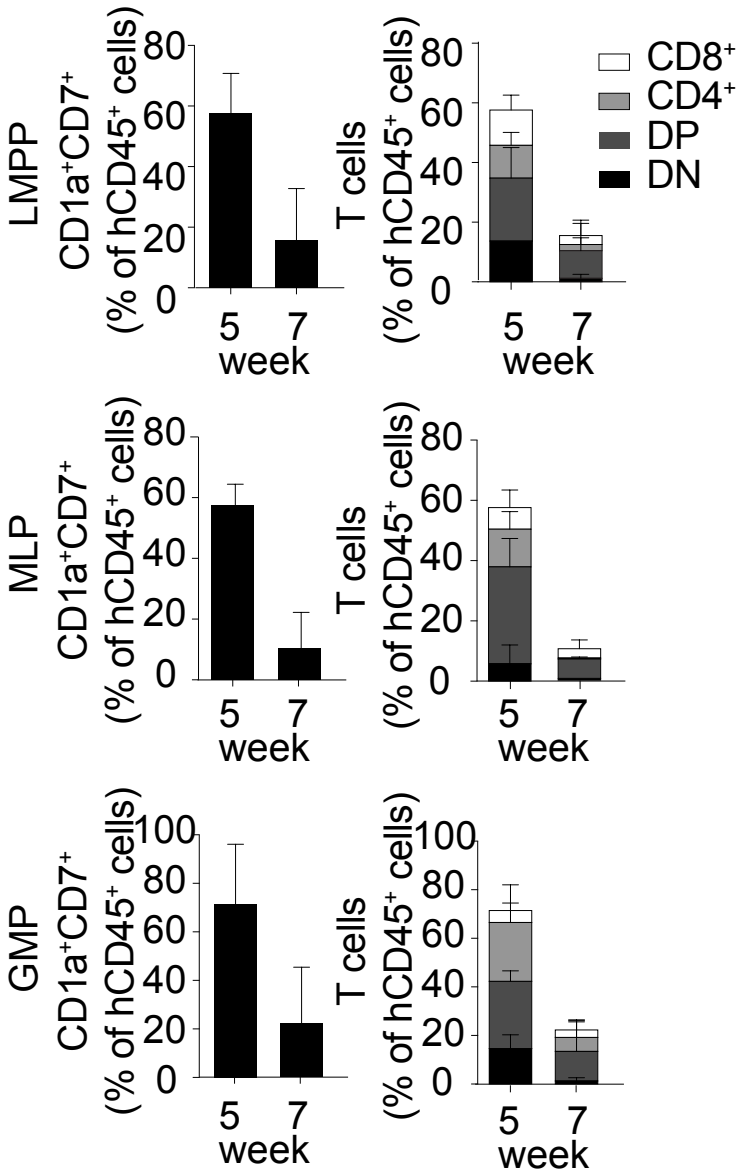
Figure 3.5: Gene Expression Analysis of Output from Bulk SGF15/2 Cultures. 150 LM progenitor cells were cultured in SGF15/2 culture conditions (see Chapter 2.4 and Table A.2 for culture conditions). Lineage output was analysed by flow cytometry after 2 weeks of culture and shown is the gene expression by output cells, purified by flow cytometry. Results are presented relative to those of the control genes *GAPDH* and *ACTB*. See [54], Supplementary Table 2 for TaqMan probes.

Table A.2 for more information. T cell output was analysed at 5 and 7 weeks of culture by flow cytometry (Figure 3.6).

All of the LM progenitors were seen to generate hCD7$^+$CD1a$^+$ cells of the T cell lineage. This included hCD7$^+$CD1a$^+$ CD4$^+$CD8$^+$ (double-positive) T cells and hCD7$^+$CD1a$^+$ CD4$^-$CD8$^+$ or CD4$^+$CD8$^-$ (single-positive) T cells (Figure 3.6). Finally, it was confirmed that T cell populations purified by flow cytometry expressed the appropriate lineage-associated genes (Figure 3.7).



(a) Flow Cytometric Analysis of Bulk SF7a Culture



(b) Kinetics of Bulk SF7a Culture

Figure 3.6: Bulk SF7a Culture of LM Progenitors. Lineage output of LM progenitors in SF7a culture conditions (see Chapter 2.4 and Table A.2 for culture conditions). 150 LMPPs, MLPs and GMPs were plated in SF7a conditions and analysed by flow cytometry after 5 or 7 weeks of culture. (a) Representative flow cytometric analysis of the T cell output following 5 weeks of culture. (b) Kinetics of lineage output of LM progenitors in SF7a conditions at 5 and 7 weeks of culture. Shown is the frequency of $CD7^+CD1a^+$ cells of the T cell lineage out of $hCD45^+$. The $CD7^+CD1a^+$ $CD4^-CD8^+$cell (CD8), $CD7^+CD1a^+$ $CD4^+CD8^-$ cell (CD4), $CD7^+CD1a^+$ $CD4^+CD8^+$ cell (DP) and $CD7^+CD1a^+$ $CD4^-CD8^-$cell (DN) composition of the $CD7^+CD1a^+$ cells is indicated. Frequencies are the mean from 5 CB donors and calculated as a percentage of human $CD45^+$ cells.

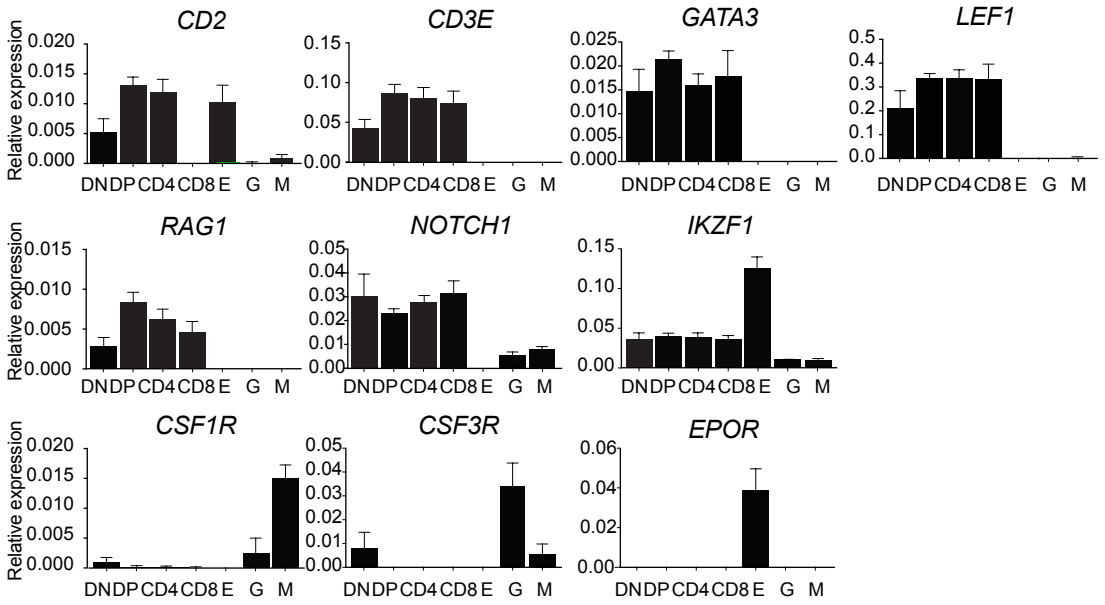


Figure 3.7: Gene Expression Analysis of Output from Bulk SF7a Cultures. 150 LM progenitor cells were cultured in SF7a culture conditions (see Chapter 2.4 and Table A.2 for culture conditions). Lineage output was analysed by flow cytometry after 5 or 7 weeks of culture and shown is the gene expression by output cells, purified by flow cytometry. As a control, the expression from mature non-T cells obtained from the sorting of cells from E, G and M colonies from Figure 3.3 is also shown. Results are presented relative to those of the control genes *GAPDH* and *ACTB*. See [54], Supplementary Table 2 for TaqMan probes.

3.3.3 Limiting Dilution Assay of LM Progenitors

Following on from the population analysis of LM functional potential *in vitro*, the clonal potential of these progenitors was examined. Initially this was done using an LDA in SGF15/2 conditions. Lineage output was detected by flow cytometry analysis following 2–2.5 weeks of culture (Figure 3.8a).

1, 2, 5, 10 and 20 LMPP/MLP/GMP cells were plated and the frequency with which a population generated a lineage was estimated using Poisson statistics. 1 in 2 LMPPs produced B cells and 1 in 3 LMPPs produced NK cells. But only 1 in 5 LMPPs produced monocytes and 1 in 10 LMPPs produced granulocytes (Figure 3.8b and Table 3.2).

GMPs produced myeloid cells with a higher frequency than the LMPP, with 1 in 2 GMPs having generated monocytes and 1 in 4 GMPs having generated granulocytes. But the GMP generated lymphoid cells at a lower frequency than the LMPP, with 1 in 22 GMPs having produced B cells and 1 in 8 GMPs having produced NK cells (Figure 3.8b and Table 3.2).

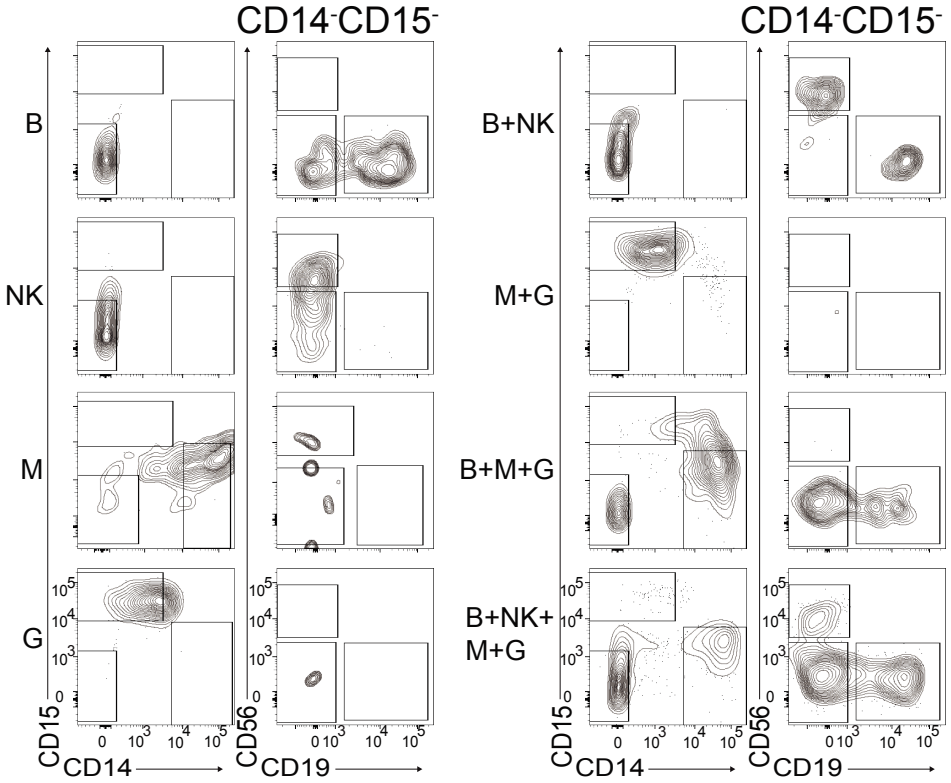
The MLP was primarily a lymphoid progenitor in this assay and 1 in 11 MLPs produced B cells and 1 in 18 produced NK cells. Myeloid output from the MLP was very rare (Figure 3.8b and Table 3.2).

Bi-lineage and multi-lineage cells were detected from all 3 LM progenitors, but at a lower frequency than the uni-lineage output, ranging from 1 in 6 to 1 in 789 (Table 3.2 and A.6).

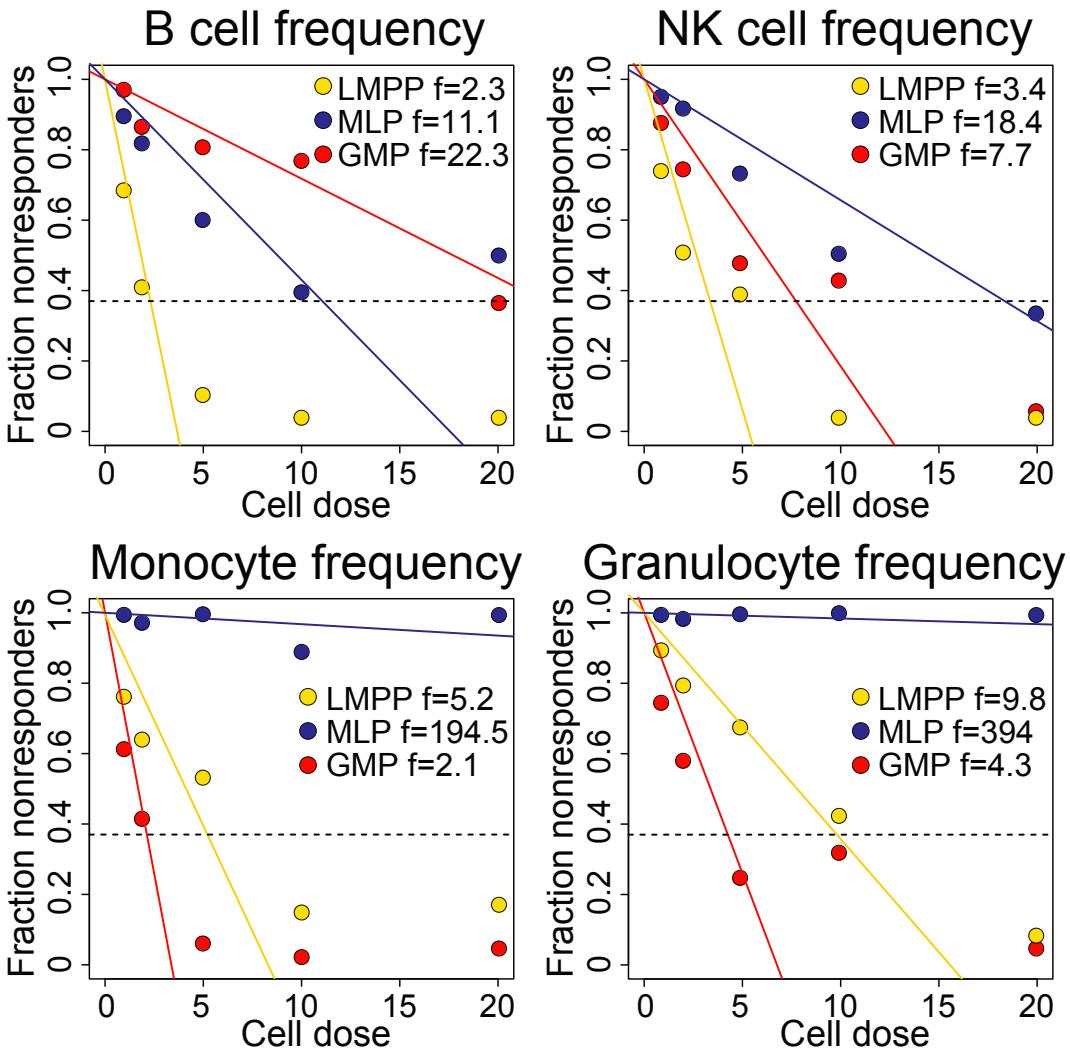
Table 3.2: Frequencies of Lineage Outputs from LDA of LM Progenitors. An LDA was carried out on the LM progenitors. 1, 2, 5, 10 and 20 LMPP/MLP/GMP cells were plated in SGF15/2 culture conditions and culture output analysed by flow cytometry after 2–2.5 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). Shown are LDA frequencies as "1 in X cells can give rise to". ND- not detected. Frequencies calculated using L-Calc software (Stem Cell Technologies). See Table A.6 for absolute numbers of lineage outputs.

Frequency	LMPP	MLP	GMP
B	2	11	22
NK	3	18	8
M	5	194	2
G	10	394	4
B_NK	6	38	31

Continuation of Table 3.2			
Frequency	LMPP	MLP	GMP
M_G	16	ND	7
B_M	10	392	32
B_G	13	789	38
M_NK	9	392	11
NK_G	15	ND	17
B_NK_M	16	392	44
B_NK_G	18	ND	49
B_M_G	22	ND	43
NK_M_G	23	ND	20
B_NK_M_G	29	ND	56



(a) Flow Cytometric Analysis of LDA of LM Progenitors



(b) Frequencies of Lineage Output Generated in LDA by LM Progenitors

Figure 3.8: LDA of LM Progenitors. An LDA was carried out on the LM progenitors. 1, 2, 5, 10 and 20 LMPP/MLP/GMP cells were plated in SGF15/2 culture conditions and culture output analysed by flow cytometry after 2–2.5 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). (a) Representative flow cytometric profiles of the culture outputs. Plots are gated on human $CD45^+$ cells. (b) LDA plots showing the frequency (f: 1 in X cells can give rise to) of LMPP, MLP and GMP cells with B cell, NK cell, monocytic and granulocytic potential. Lines represent the estimates calculated by ELDA software (`http://bioinf.wehi.edu.au/software/elda/`).

3.3.4 Single-Cell *In Vitro* Cultures of LM Progenitors

The LDA does not definitively allow the deduction of the frequency of multi-lineage functional potential at a single-cell level. This is due to multiple cells being plated in a single well which is read out for lineage potential. It cannot be distinguished whether all cells within a well had all detected lineage potentials, or whether the well contained a heterogenous mixture of single-cells, with different lineages being produced by different single-cells. Therefore, to determine whether the LM progenitors have multi-lineage functional potential, single-cell assays must be used.

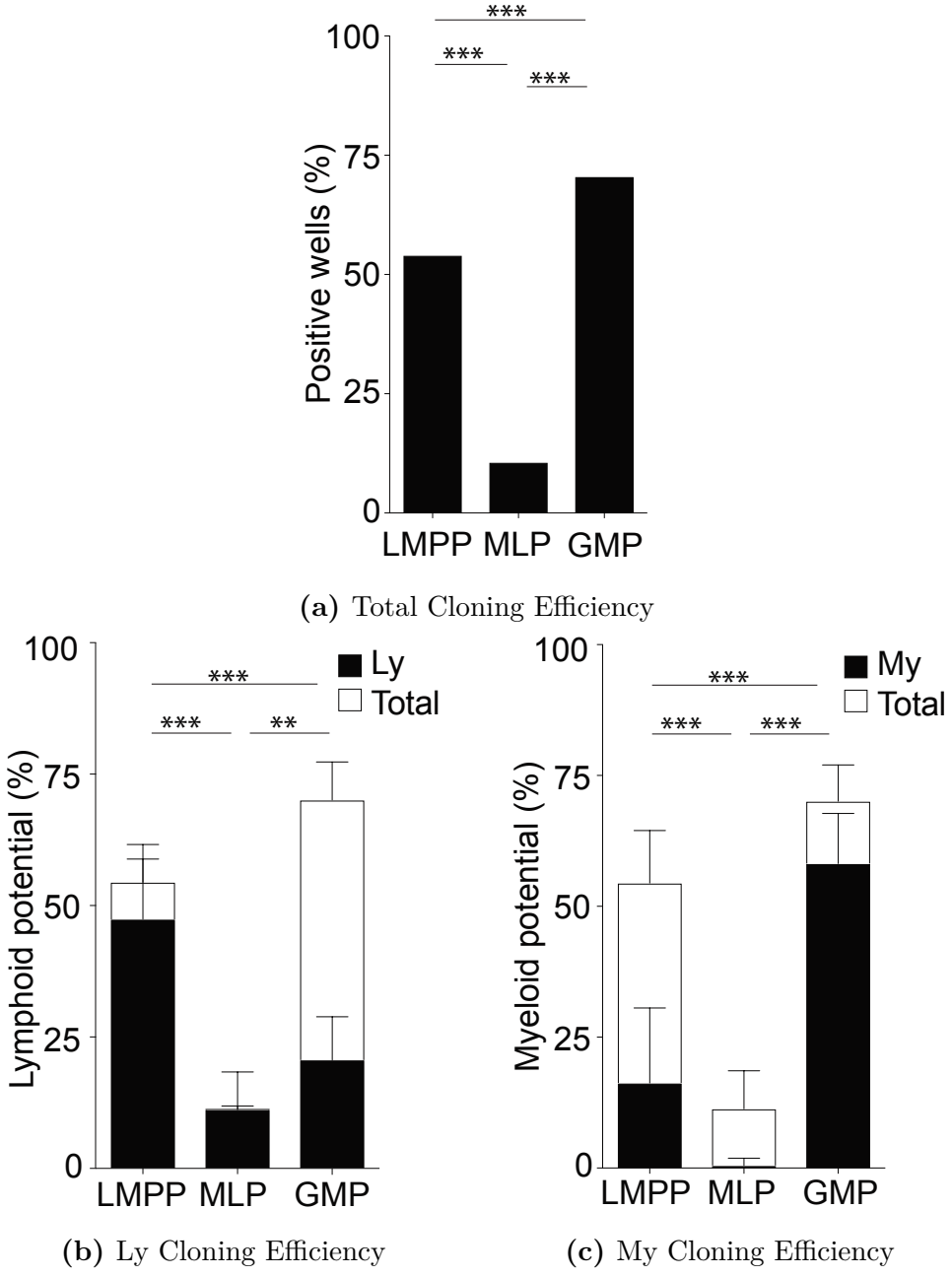
Single-cell *in vitro* cultures were carried out on 1,136 LMPPs, 710 MLPs and 1,622 GMPs, isolated from 22 CB donors (a total of 6.3 x 10^9 MNCs). This provided robust quantitative data, with the ability to detect rare functional potentials (Table A.7).

First it was observed that the LMPP and GMP had a higher cloning efficiency (54% and 71%, respectively) than that of MLPs (11%) (Figure 3.9a). The LMPP and MLP were lymphoid-biased progenitors, whereas the GMP was predominantly a myeloid progenitor (Figures 3.9b and 3.9c).

When focusing only on the wells that produced any type of mature output (productive/positive wells) it was seen that the majority of single cells were seen to generate only one lineage. 69% of LMPPs, 88% of MLPs and 63% of GMPs produced a uni-lineage output (Figure 3.10a). Within the uni-lineage output, 92% of LMPPs had a lymphoid output (B or NK) and 8% had myeloid output (G or M). The MLP was almost exclusively a lymphoid progenitor, with very low myeloid output (3%) (Figure 3.10a). The GMP was primarily a myeloid progenitor (79% uni-lineage output) with 21% lymphoid output (Figure 3.10a).

Bi-lineage output was detected in 24% of LMPPs, 12% of MLPs and 33% of GMPs (Figure 3.10b). The production of three or more lineages was detected but rare (6% of LMPPs, 0% of MLPs and 3% of GMPs) (Figure 3.10c).

Only 8% of all assayed LMPPs, 7% of GMPs and hardly any MLPs (0.3%) displayed combined LM potential (Figure 3.11). The LM output from LMPPs was significantly higher than that of GMPs (P = 0.0125) or MLPs (P = 0.0019, Table A.8).

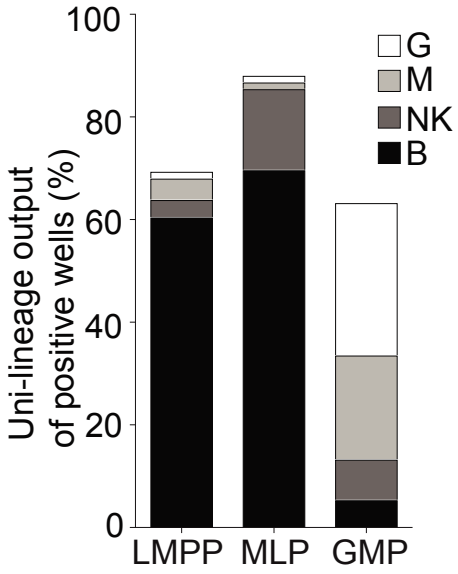


(a) Total Cloning Efficiency

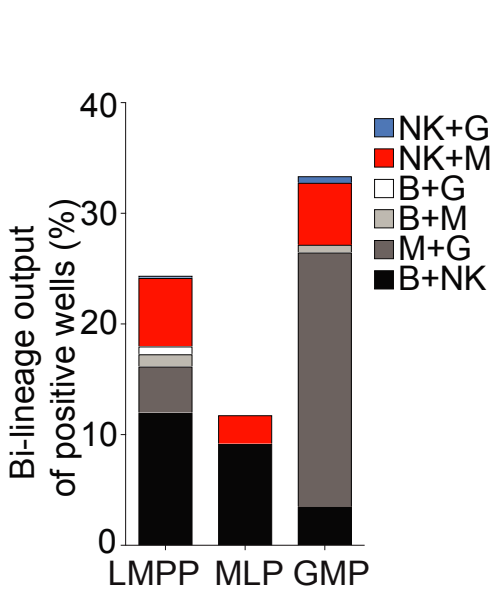
(b) Ly Cloning Efficiency

(c) My Cloning Efficiency

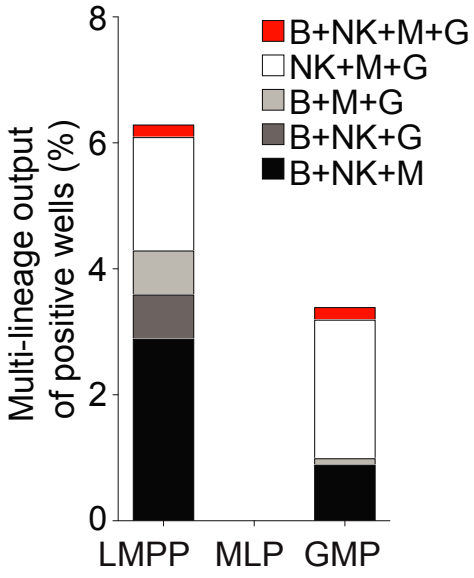
Figure 3.9: Cloning Efficiency of LM Progenitors in Single-Cell SGF15/2 Cultures. Single-cell SGF15/2 cultures were carried out on the LM progenitors. 1,136 LMPPs, 710 MLPs and 1,622 GMPs as single cells, isolated from 22 CB donors were cultured and culture output analysed by flow cytometry after 2–2.5 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is (a) the total cloning efficiency of single LMPPs (615 of 1,136 cells), MLPs (76 of 710 cells) and GMPs (1,145 of 1,622 cells), (b) cloning efficiency of lymphoid lineages and (c) myeloid lineages. Ly: lymphoid plus mixed output, My: myeloid plus mixed clones.



(a) Uni-lineage Output



(b) Bi-lineage Output



(c) Multi-lineage Output

Figure 3.10: Lineage Output of LM Progenitors in Single-Cell SGF15/2 Cultures. Single-cell SGF15/2 cultures were carried out on the LM progenitors. 1,136 LMPPs, 710 MLPs and 1,622 GMPs as single cells, isolated from 22 CB donors were cultured and culture output analysed by flow cytometry after 2–2.5 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the (a) uni-lineage, (b) bi-lineage and (c) multi-lineage output. Presented as the proportion of cells of various lineages in wells that produced cells (positive wells).

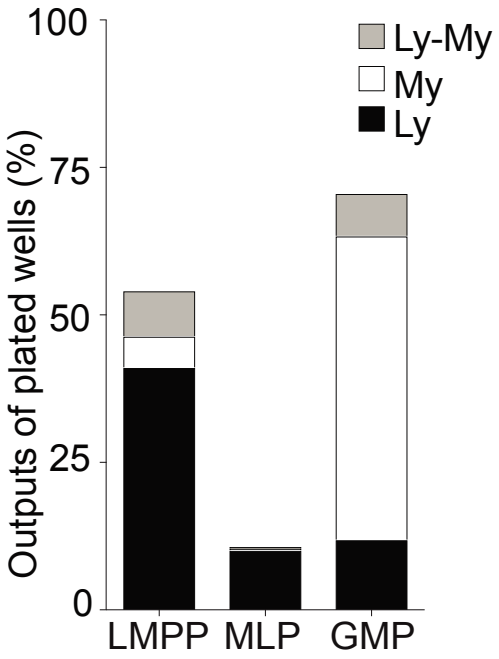
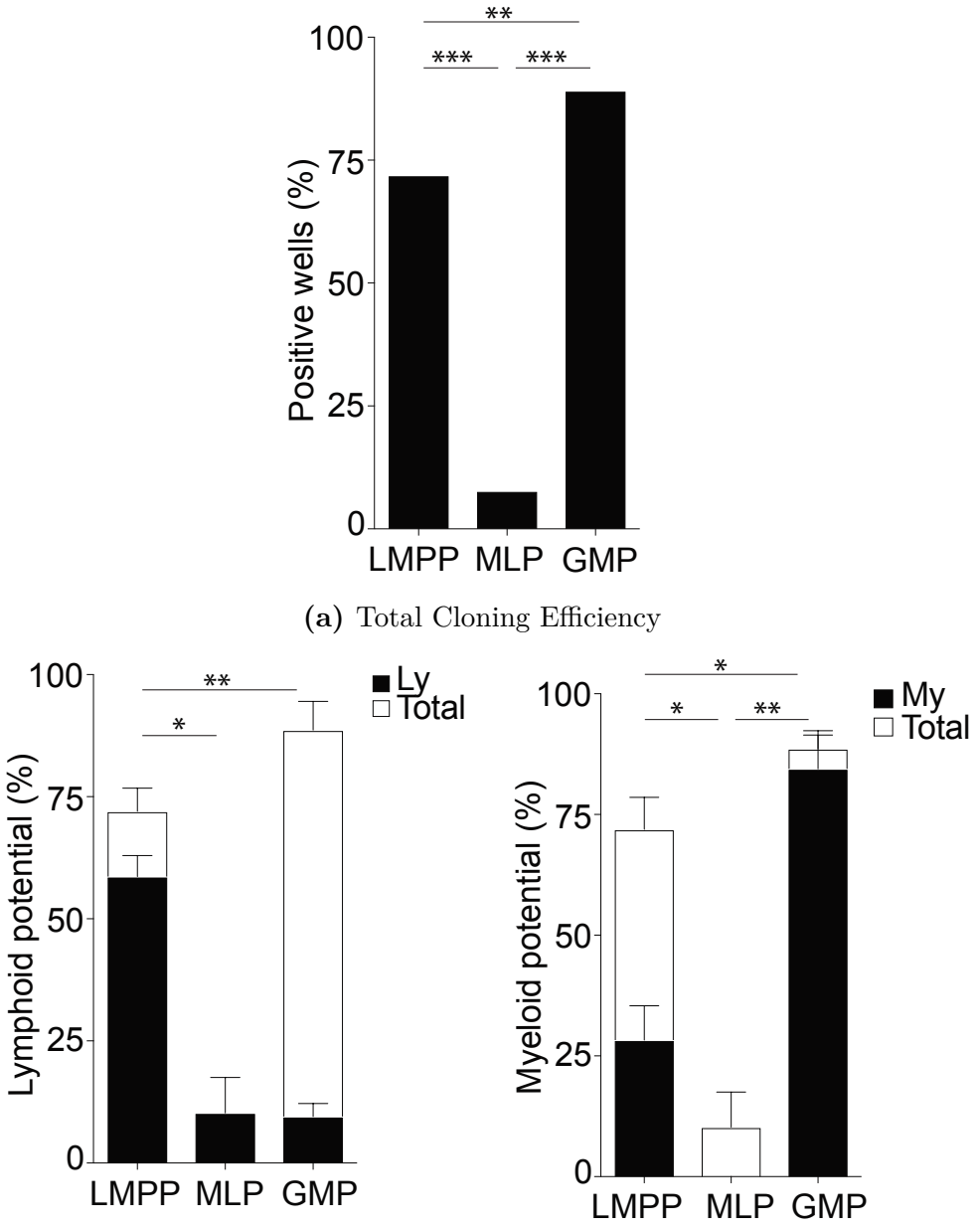


Figure 3.11: Ly-My Output of LM Progenitors in Single-Cell SGF15/2 Cultures. Single-cell SGF15/2 cultures were carried out on the LM progenitors. 1,136 LMPPs, 710 MLPs and 1,622 GMPs as single cells, isolated from 22 CB donors were cultured and culture output analysed by flow cytometry after 2–2.5 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the lymphoid, myeloid and lympho-myeloid (Ly-My) output, presented as the frequency among all plated cells.

In addition to the single-cell assays in SGF15/2 conditions, the LM progenitors' lineage potential was also assayed in another single-cell culture system. Previously used to define the MLP population, this culture was with MS-5 stroma in the presence of the cytokines SCF, IL-7, thrombopoietin, IL-2, GM-CSF, G-CSF and M-CSF [24]. These conditions were termed "S7T2GM/G/M culture" here, see Chapter 2.4 and Table A.2 for more information.

96 LMPPs, 52 MLPs and 110 GMPs as single cells were cultured in S7T2GM/G/M conditions and similar results to those of SGF15/2 culture were seen (Figure 3.12, 3.13 and 3.14). It was observed that S7T2GM/G/M cultures were less permissive for granulocytic output. Most of the output from LMPPs was uni-lineage, with rarer bi-lineage and less frequent multi-lineage output (Table A.9). MLPs exhibited only lymphoid uni-lineage output (Figure 3.13a and Table A.9).

Finally, we analysed the ability of LM progenitors to produce all of the lymphoid lineages, including T cells, in combination with myeloid cells. This utilised another

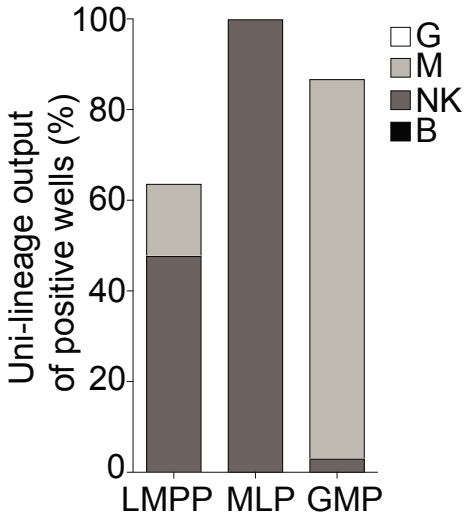


(a) Total Cloning Efficiency

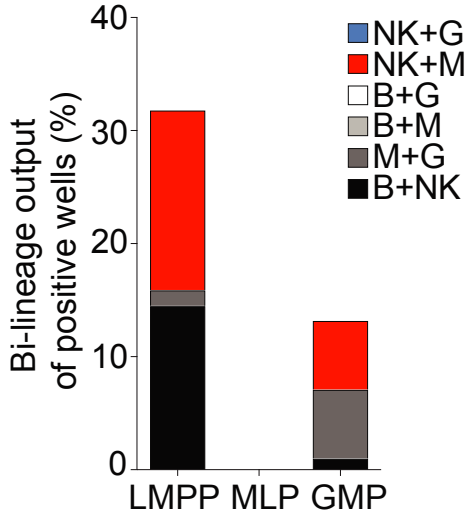
(b) Ly Cloning Efficiency

(c) My Cloning Efficiency

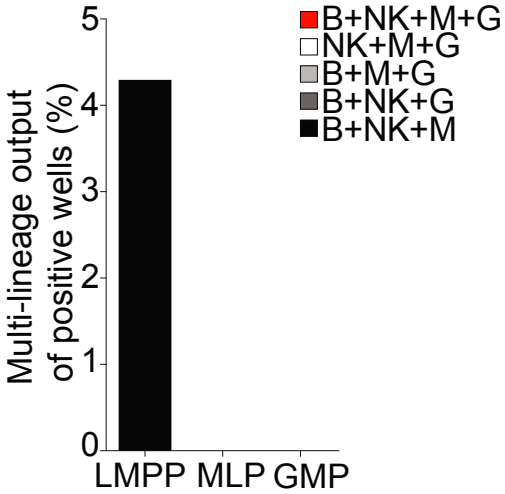
Figure 3.12: Cloning Efficiency of LM Progenitors in Single-Cell S7T2GM/G/M Cultures. Single-cell S7T2GM/G/M cultures were carried out on the LM progenitors. 96 LMPPs, 52 MLPs and 110 GMPs as single cells, isolated from 2 CB donors were cultured and culture output analysed by flow cytometry after 2–2.5 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is (a) the total cloning efficiency of single LMPPs (69 of 96 cells), MLPs (4 of 52 cells) and GMPs (98 of 110 cells), (b) cloning efficiency of lymphoid lineages and (c) myeloid lineages. Ly: lymphoid plus mixed output, My: myeloid plus mixed clones.



(a) Uni-lineage Output



(b) Bi-lineage Output



(c) Multi-lineage Output

Figure 3.13: Lineage Output of LM Progenitors in Single-Cell S7T2GM/G/M Cultures. Single-cell S7T2GM/G/M cultures were carried out on the LM progenitors. 96 LMPPs, 52 MLPs and 110 GMPs as single cells, isolated from 2 CB donors were cultured and culture output analysed by flow cytometry after 2–2.5 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the (a) uni-lineage, (b) bi-lineage and (c) multi-lineage output. Presented as the proportion of cells of various lineages in wells that produced cells (positive wells).

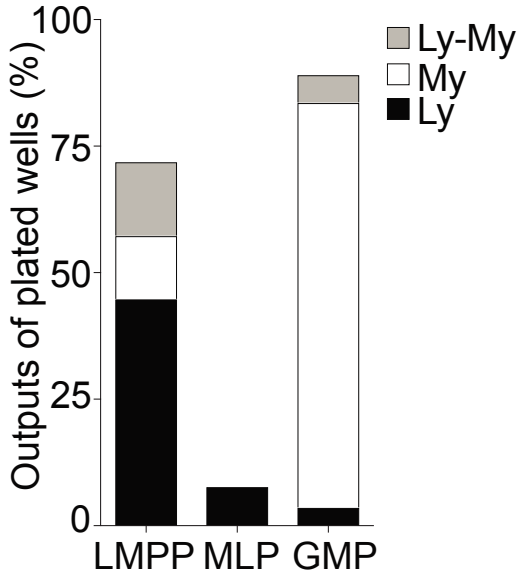


Figure 3.14: Ly-My Output of LM Progenitors in Single-Cell S7T2GM/G/M Cultures. Single-cell S7T2GM/G/M cultures were carried out on the LM progenitors. 96 LMPPs, 52 MLPs and 110 GMPs as single cells, isolated from 2 CB donors were cultured and culture output analysed by flow cytometry after 2–2.5 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the lymphoid, myeloid and lympho-myeloid (Ly-My) output, presented as the frequency among all plated cells.

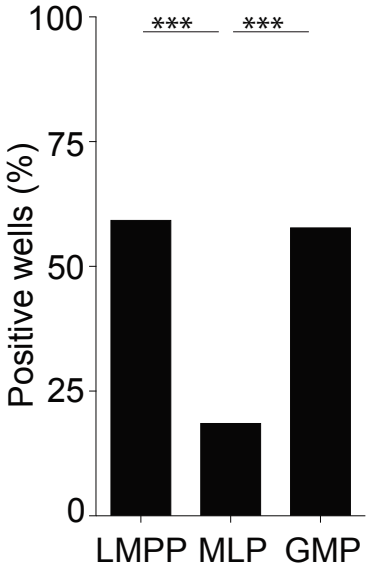
independent culture condition optimised for the detection of B, NK, T cell and myeloid potential. Single LMPPs, MLPs and GMPs were cultured in the presence of SCF, FLT3L and IL-7 on an MS-5 stromal cell line that could be induced to express hDL1 by doxycycline treatment (MS-5- hDL1$^{\mathrm{IND}}$ stroma). The B-NK-M-G output was analysed at 3 weeks and the T cell output was analysed at 6 weeks of culture. These conditions were termed "SF7b/Dox culture" here, see Chapter 2.4 and Table A.2 for more information.

215 LMPPs, 197 MLPs and 219 GMPs were cultured in SF7b/Dox conditions (Figures 3.15, 3.16 and 3.17 and Table A.10). Uni-lineage T cell output was detected in LMPP and MLP populations (3% of productive wells) but was almost completely absent in GMP populations ($< 0.1\%$) (Figure 3.16a). T cell output combined with other lymphoid output was detected for 1–5% of LMPPs and MLPs and rarely for GMPs (Figure 3.16a). LM output was detected only for LMPPs (14%) (Figure 3.16c).

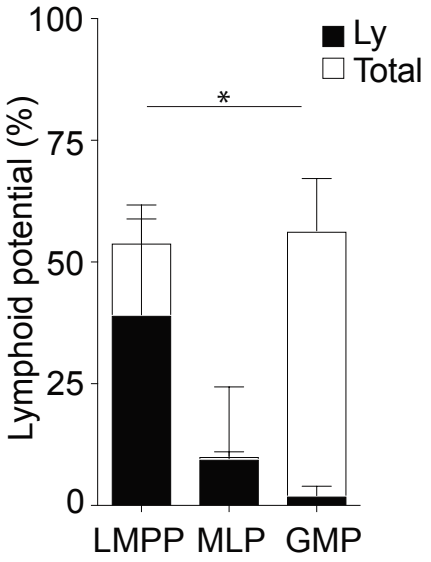
Overall, 24 functionally different single-cell progenitor types were identified across all of the *in vitro* clonal assays. All 24 progenitor types were observed for LMPPs, and only subsets of those were observed for MLPs and GMPs (Table 3.3).

Table 3.3: Summary of Lineage Outputs from Single LMPP, MLP and GMP Cells. Summary of the lineage outputs detected from single LMPP, MLP and GMP cells in Chapter 3.3.4, see Figures 3.10, 3.13 and 3.16.

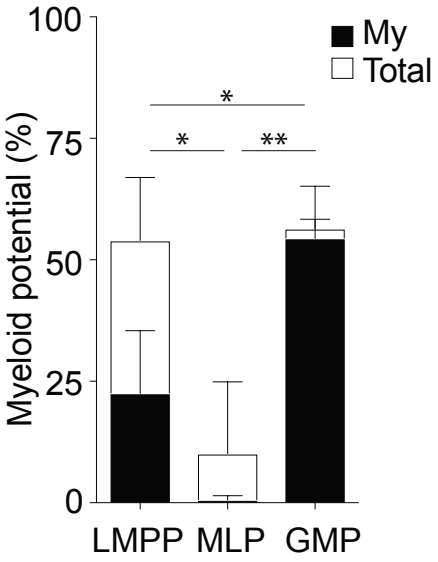
Potentials	LMPP	MLP	GMP
Lymphoid	B, NK, T, B+NK, NK+T, B+NK+T	B, NK, T, B+NK, B+NK+T	B, NK, B+NK, NK+T
Myeloid	M, G, M+G	M, G	M, G, M+G
Ly-My	B+M, B+G, G+T, NK+M, NK+G, B+NK+M, B+NK+G, B+M+G, NK+M+G, B+NK+M+G, B+NK+M+T, NK+M+G+T, B+M+G+T, B+NK+G+T, B+NK+M+G+T	NK+M	B+M, NK+M, NK+G, B+NK+M, B+M+G, NK+M+G, B+NK+M+G
Total no.	24	8	14



(a) Total Cloning Efficiency

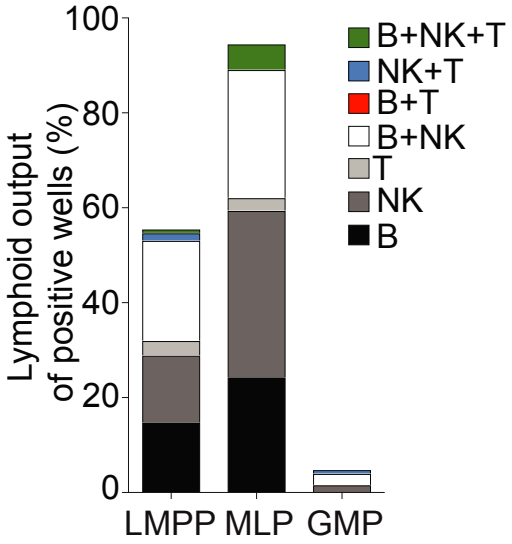


(b) Ly Cloning Efficiency

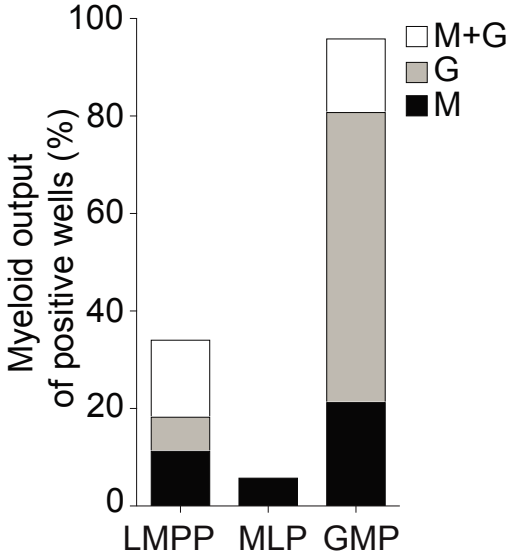


(c) My Cloning Efficiency

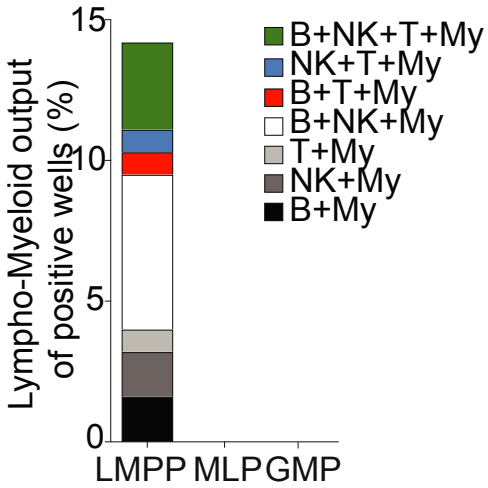
Figure 3.15: Cloning Efficiency of LM Progenitors in Single-Cell SF7b/Dox Cultures. Single-cell SF7b/Dox cultures were carried out on the LM progenitors. 215 LMPPs, 197 MLPs and 219 GMPs as single cells, isolated from 3 CB donors were cultured on an MS-5 stromal cell line that can be induced to express hDL1 by doxycycline treatment. The B-NK-M-G output was analysed at 3 weeks and the T cell output was analysed at 6 weeks. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is (a) the total cloning efficiency of single LMPPs (128 of 215 cells), MLPs (37 of 197 cells) and GMPs (127 of 219 cells), (b) cloning efficiency of lymphoid lineages and (c) myeloid lineages. Ly: lymphoid plus mixed output, My: myeloid plus mixed clones.



(a) Lymphoid Output



(b) Myeloid Output



(c) Lympho-Myeloid Output

Figure 3.16: Lineage Output of LM Progenitors in Single-Cell SF7b/Dox Cultures. Single-cell SF7b/Dox cultures were carried out on the LM progenitors. 215 LMPPs, 197 MLPs and 219 GMPs as single cells, isolated from 3 CB donors were cultured on an MS-5 stromal cell line that can be induced to express hDL1 by doxycycline treatment. The B-NK-M-G output was analysed at 3 weeks and the T cell output was analysed at 6 weeks. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the (a) lymphoid, (b) myeloid and (c) lympho-myeloid output. Presented as the proportion of cells of various lineages in wells that produced cells (positive wells).

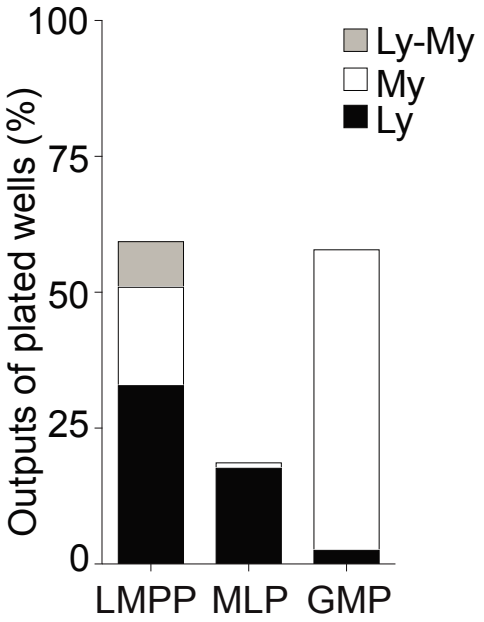


Figure 3.17: Ly-My Output of LM Progenitors in Single-Cell SF7b/Dox Cultures. Single-cell SF7b/Dox cultures were carried out on the LM progenitors. 215 LMPPs, 197 MLPs and 219 GMPs as single cells, isolated from 3 CB donors were cultured on an MS-5 stromal cell line that can be induced to express hDL1 by doxycycline treatment. The B-NK-M-G output was analysed at 3 weeks and the T cell output was analysed at 6 weeks. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the lymphoid, myeloid and lympho-myeloid (Ly-My) output, presented as the frequency among all plated cells.

3.4 *In Vivo* Assays of LM Progenitors

Although it has been suggested that single HSCs may be xenotransplanted and generate long-term, multi-lineage engraftment [4], the xenotransplantation of single human progenitor cells has not yet been successful. Furthermore, the xenotransplantation of bulk human progenitors into immunodeficient mice has been observed to generate only low levels of engraftment (<0.1%) [3, 22, 24, 28]. Therefore, to assess the *in vivo* potential of the LM progenitors, a novel humanised ossicle mouse model was used [57].

Human BM-derived mesenchymal stromal cells were subcutaneously injected into immunodeficient mice, where they formed a humanised ossicle over 8 weeks.

LMPPs, MLPs and GMPs were injected into the ossicle, and lineage output was analysed 1 and 2 weeks after injection (see Chapter 2.8). Flow cytometry analysis following xenotransplantation was able to detect hCD45$^+$hCD33$^+$ hCD14$^+$ cells (M), hCD45$^+$hCD33$^+$ hCD15$^+$ cells (G) and hCD45$^+$hCD33$^-$hCD19$^+$ cells (B) (Figure 3.18).

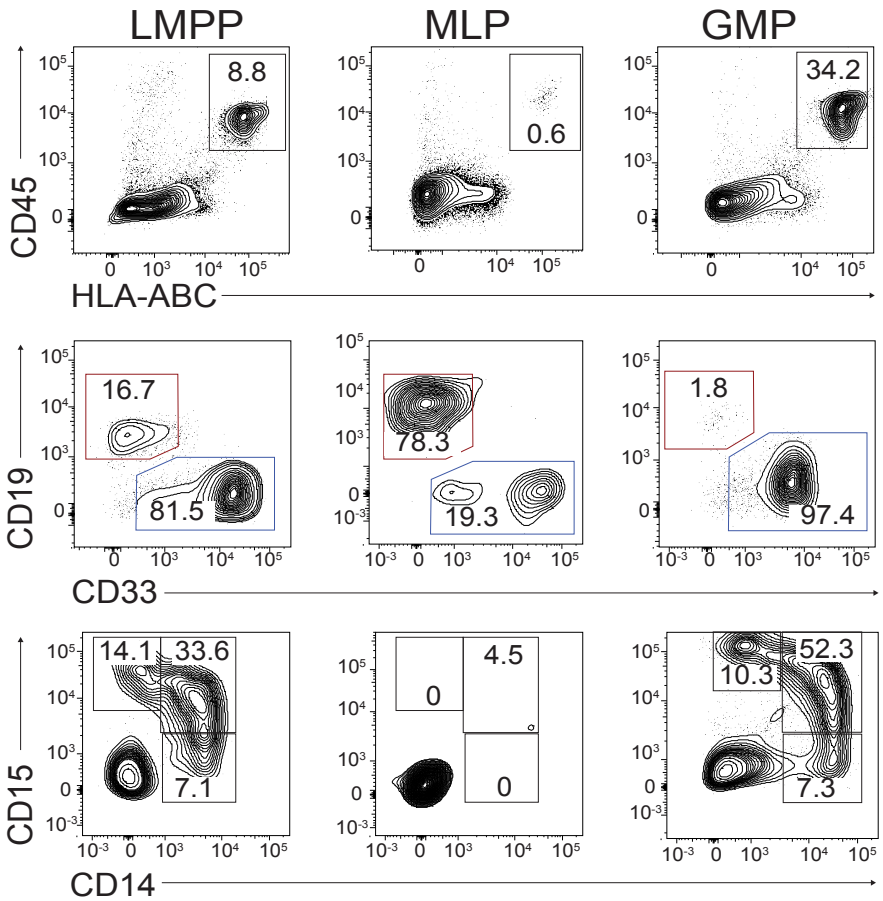


Figure 3.18: Flow Cytometry Analysis of Output from Ossicle Assays. The *in vivo* potential of LM progenitors was assayed using a humanised ossicle mouse model (for details of the ossicle assays see Chapter 2.8 and [57]). Shown is the flow cytometry analysis of engraftment 2 weeks post-transplantation. Numbers in outlined areas indicate percent human CD45$^+$HLA-ABC$^+$ cells (human engraftment) and within human engraftment the CD19$^+$ (B) cells, and CD33$^+$ (myeloid) cells. Within the myeloid cells the CD15$^+$CD14$^-$ granulocytes, CD15$^+$CD14$^+$ immature myeloid cells and CD15$^-$CD14$^+$ monocytes are indicated.

Engraftment was observed at both 1 and 2 weeks post-xenotransplantation, with greater engraftment seen at week 2 (data not shown). All downstream analyses were done at 2 weeks following xenotransplantation. The number of cells injected

varied (~300–60,000 cells, depending on the LM progenitor, Figure 3.19), therefore the mean cell engraftment per 1,000 transplanted cells is shown.

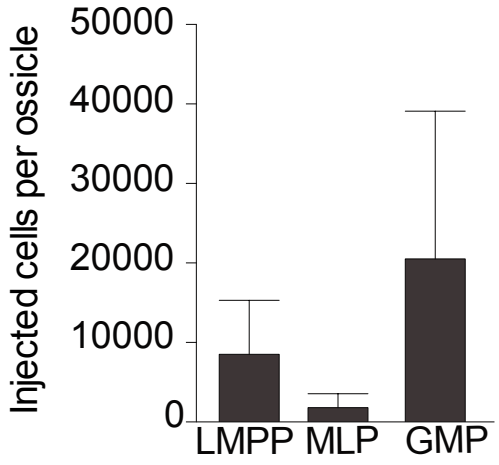
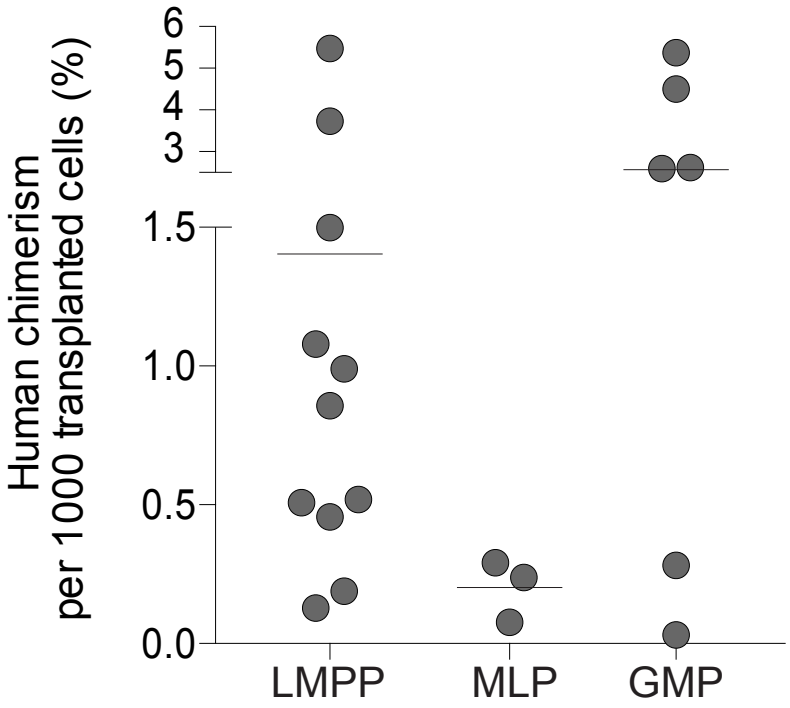


Figure 3.19: Number of Cells Injected in Ossicle Assays of LM Progenitors. The *in vivo* potential of LM progenitors was assayed using a humanised ossicle mouse model (for details of the ossicle assays see Chapter 2.8 and [57]). Shown is the number of LMPP, MLP and GMP cells transplanted in the ossicle assay.

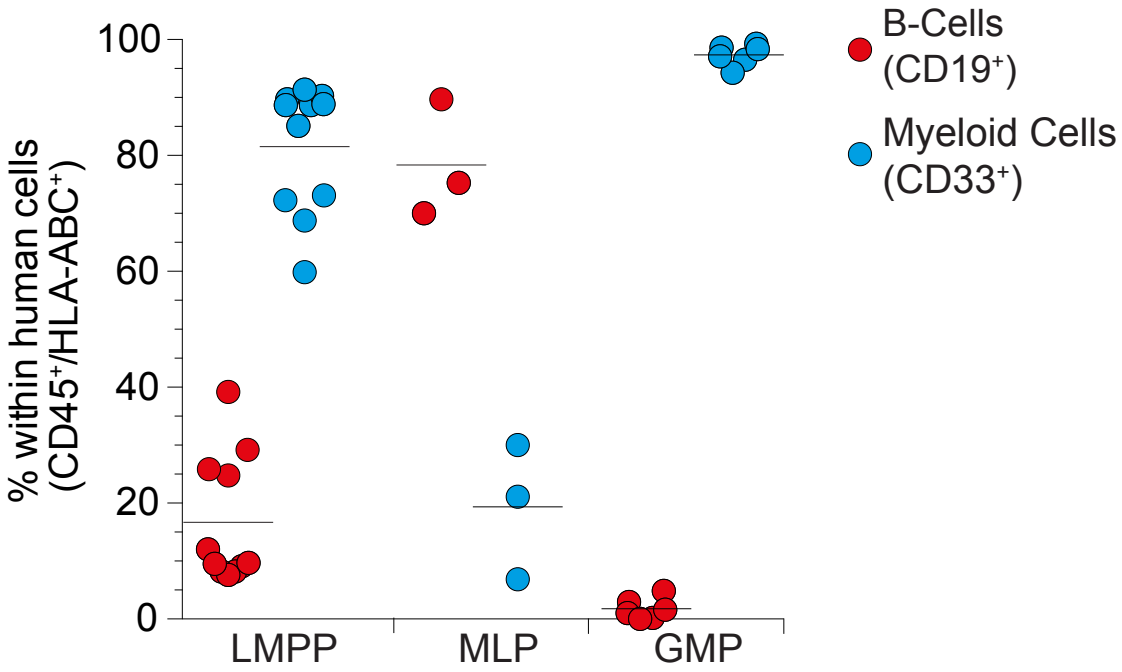
GMPs had the highest mean engraftment (2.6%), followed by LMPPs (1.4%), with the MLP having the lowest engraftment (0.2%) (Figure 3.20a). LMPPs produced more $CD33^+$ myeloid cells (82%) than $CD19^+$ B cells (17%) (Figure 3.20b). MLPs generated more lymphoid B cells (78%) than myeloid cells (19%) (Figure 3.20b). There was no correlation between the number of transplanted cells and the ratio of lymphoid cells to myeloid cells (Figure 3.21).

GMPs predominantly generated myeloid cells (97%) (Figure 3.20b). Myeloid cells generated from LMPPs and GMPs expressed markers of monocytes (CD14) and granulocytes (CD15) (Figure 3.18). No $CD14^+$ or $CD15^+$ cells were detected as being produced from MLPs (Figure 3.18).

Morphological analysis of the engrafted cells stained with May-Grünwald-Giemsa confirmed that $CD15^+$ cells were granulocytic and $CD14^+$ cells were monocytic



(a) Human Chimerism



(b) Lineage Output

Figure 3.20: Human Chimerism and Lineage Output in Ossicle Assays. The *in vivo* potential of LM progenitors was assayed using a humanised ossicle mouse model (for details of the ossicle assay see Chapter 2.8 and [57]). Shown is (a) the human chimerism and (b) lymphoid and myeloid output 2 weeks post-transplantation.

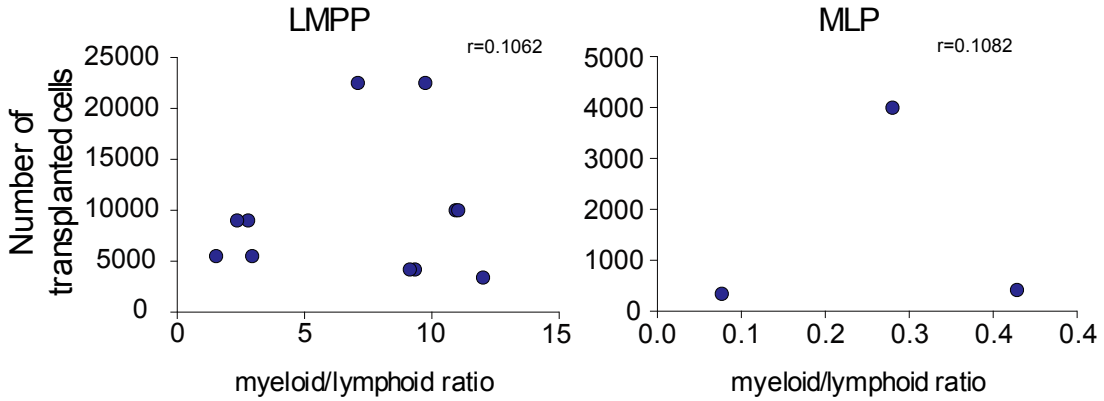


Figure 3.21: Number of Injected Cells and My/Ly Ratio in Ossicle Assay. The *in vivo* potential of LM progenitors was assayed using a humanised ossicle mouse model (for details of the ossicle assay see Chapter 2.8 and [57]). Shown is the correlation between the number of transplanted cells and the My/Ly ratio of the output cells.

(Figure 3.22). $CD14^+$ $CD15^+$ cells, generated by LMPPs and GMPs were immature myeloid cells by morphology (Figure 3.22).

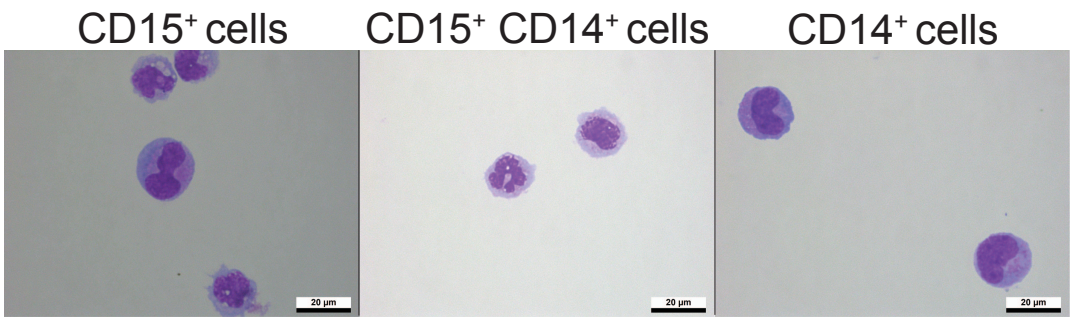
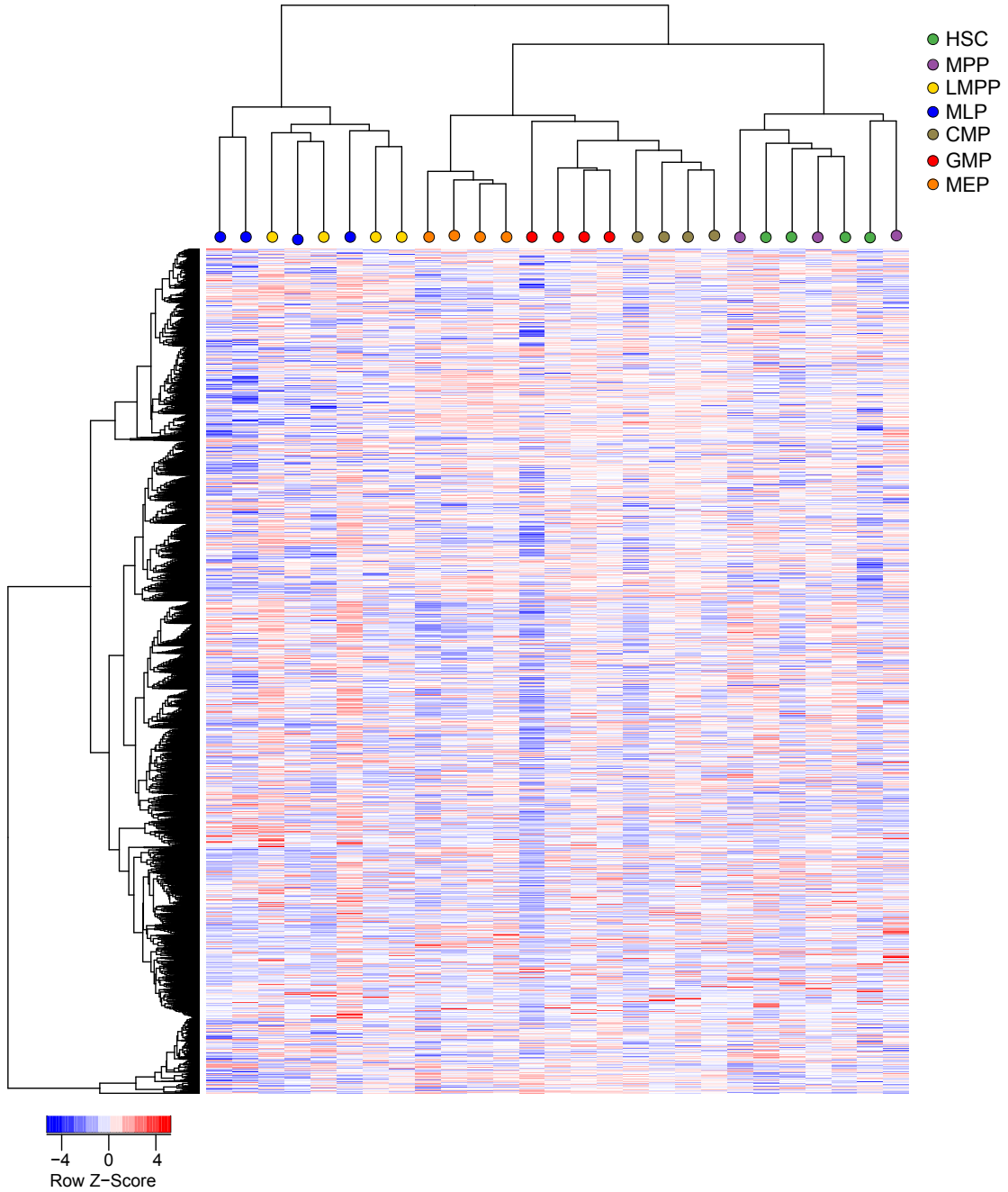


Figure 3.22: Morphology of Output in Ossicle Assays. The *in vivo* potential of LM progenitors was assayed using a humanised ossicle mouse model (for details of the ossicle assay see Chapter 2.8 and [57]). Shown is microscopy of $CD15^{+}CD14^{-}$, $CD15^{+}CD14^{+}$ and $CD15^{-}CD14^{+}$ myeloid cells generated by LMPPs at 2 weeks after transplantation stained with May-Grünwald-Giemsa. Scale bars: 20 µm.

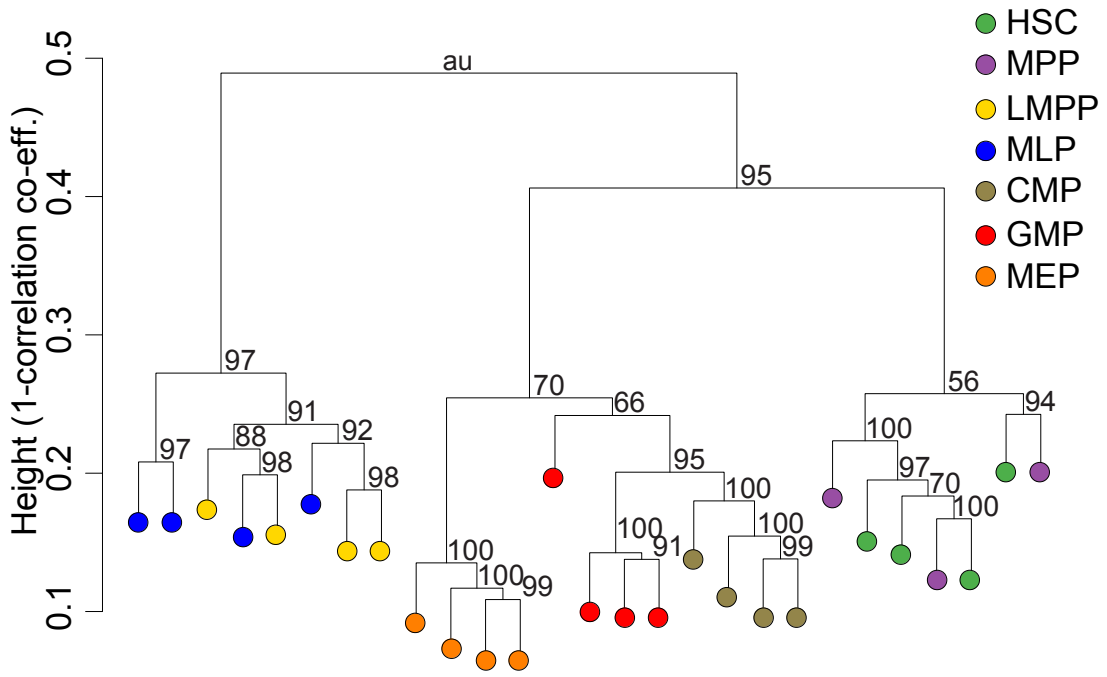
3.5 Bulk RNA-Seq of Normal HSPCs

Having now characterised both the *in vitro* and *in vivo* functional potential of the LM progenitors, the gene expression profiles were then examined. Initially this was done at the population level and bulk RNA-Seq of the seven human CB HSPC populations (HSC, MPP, LMPP, MLP, CMP, GMP and MEP) was carried out. For details on initial data processing see Chapter 2.14.1.

Firstly, an unbiased approach was taken to examine the global gene expression profiles and hierarchical clustering was carried out using all expressed genes (Figure 3.23). The LMPP and MLP were seen to cluster away from the other HSPCs, and HSCs and MPPs clustered away from the more mature progenitors. The clustering was also able to separate out the CMP, GMP and MEP populations (Figure 3.23).



(a) Heatmap of Hierarchical Clustering



(b) Bootstrap Permutation of Hierarchical Clustering

Figure 3.23: Bulk RNA-Seq of Normal HSPCs: Hierarchical Clustering. Bulk RNA-Seq was carried out on normal HSPC CB populations. (a) Heatmap showing hierarchical clustering and the expression of all expressed genes by HSPC populations. Rows represent genes and columns represent a biological replicate of an HSPC population. Colour scale is normalised per row. (b) Hierarchical clustering of HSPC populations using all expressed genes and 1,000 bootstrap permutation analyses, with height calculated as (1 - [correlation co-efficient]). Numbers above dendrogram bars indicate approximate unbiased P values (au).

3.5.1 PCA and ANOVA Analysis of HSPC Populations

In another unbiased approach, a PCA was carried out on the HSPC populations using all expressed genes. It was seen that PC1, representing the majority of the variance across the populations, separated out the MEP from the lymphoid populations (LMPP and MLP) (Figure 3.24). PC2 was seen to separate out one biological replicate of the GMP, potentially due to the heterogeneity of this population.

To focus on the biological variability across the HSPC populations an analysis of variance (ANOVA)-like test in edgeR was used to identify DEGs across the HSPC populations (see [54], Supplementary Table 4). The PCA was then run using

300–10,000 of the most significantly DEG (ranked by p-adjusted value) (Figure 3.24). The clearest separation of the HSPC populations was seen when the 300 most significant DEGs were used. Here it was seen that PC2 separated out the more immature HSC and MPP from the remaining HSPC populations. As with all genes, PC1 separated out the megakaryocytic-erythroid populations from the more lymphoid populations (Figure 3.24).

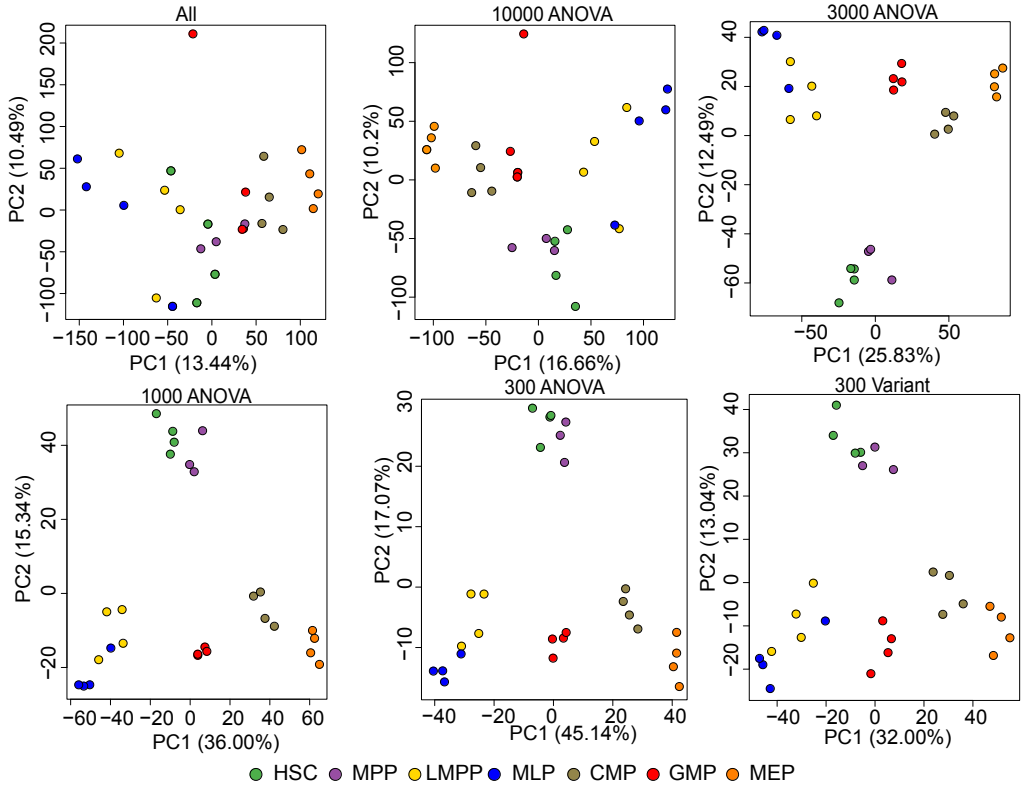


Figure 3.24: PCA of Normal CB HSPC. PCA of normal CB HSPC populations using: all expressed genes, 300–10,000 of the most DEGs across normal HSPC populations (as determined by an ANOVA-like test, see [54], Supplementary Table 4) or the 300 most variable genes, as indicated. The percentage variance contributed by each PC is indicated. Each point represents a biological replicate.

A PCA was also carried out using an unbiased selection of genes with the highest variance across all HSPC populations, which does not specify population identity *a priori* (Figures 3.24 and 3.25). Using 300–10,000 of the genes with the highest variance, these PCAs were seen to give similar results as to when the ANOVA gene sets were used.

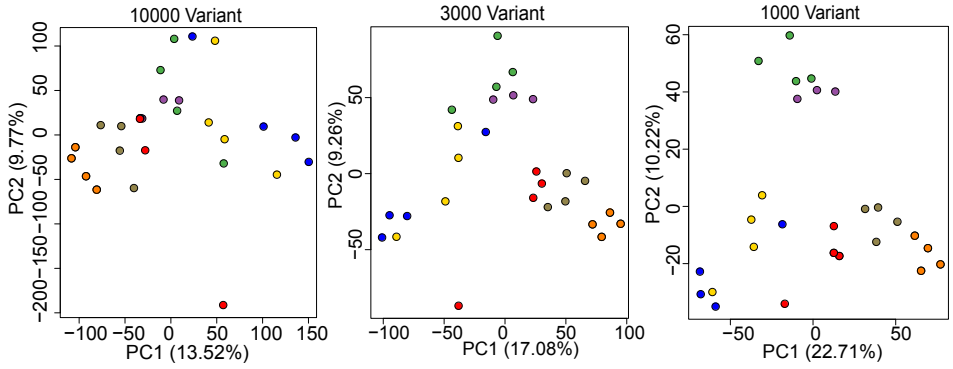


Figure 3.25: PCA of Normal CB HSPC: Variant Genes. PCA of normal CB HSPC populations using 1,000–10,000 of the most variable genes, as indicated. The percentage variance contributed by each PC is indicated. Each point represents a biological replicate.

To identify the PCs which represented meaningful biological variance, the eigenvalues of the PCs were investigated. Figure 3.26 shows the eigenvalues for the PCA run with the 300 most significant DEGs and those from a randomised dataset. It was seen that PC1–PC3 captured most of the biological variation, whereas PC4 showed similar levels of variation as those from the random dataset (Figure 3.26). When plotting PC3 it was seen that this represented an axis of myeloid differentiation that was not seen when looking only at PC1 and PC2 (Figures 3.24 and 3.27).

To confirm the biology underlying the PCs identified, the loading plots are shown in Figure 3.28. Here the genes with the most extreme loading scores, responsible for driving the PC, are annotated. For the PCA using the 300 most significant DEGs identified by the ANOVA-like test, genes associated with stem cells (*HLF*, *MECOM* and *NFIB*), lymphoid cells (*IGJ*, *IRF8* and *MME*) and erythroid-megakaryocytic cells (*HBD* and *HPGDS*) could all be seen (Figure 3.28, highlighted in red).

Hierarchical clustering was then carried out using the 300 most significant DEGs identified by the ANOVA (Figure 3.29). The HSPC populations were clearly separated, with clusters of genes showing distinct expression across the HSPC populations. Within each of these clusters, genes related to the biology and lineage of the HSPC populations could be identified. For example, GMPs had the highest expression of *ELANE*, *MPO* and *PRTN3*, whereas LMPPs and MLPs shared

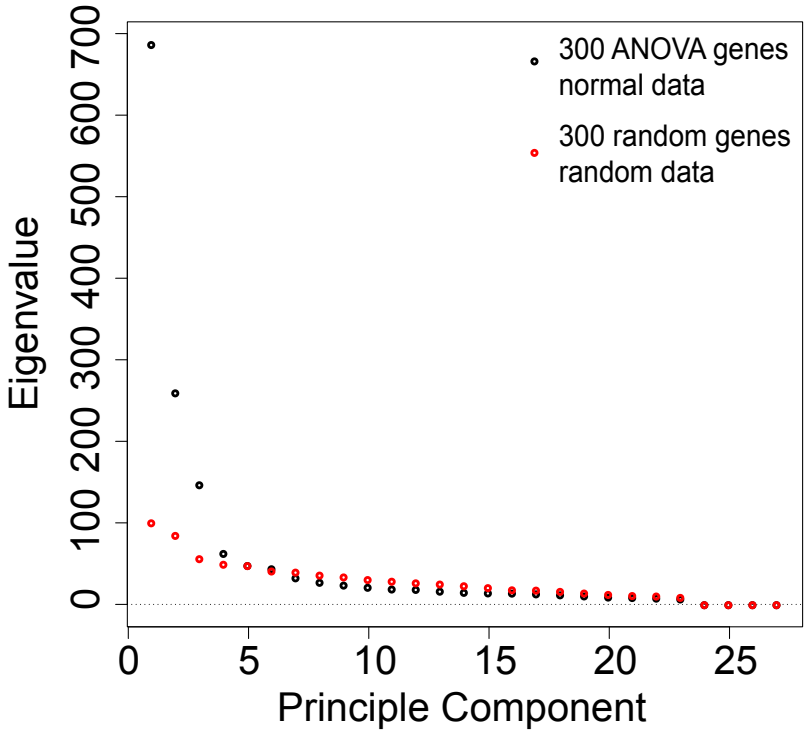


Figure 3.26: PCA of Normal CB HSPC: Eigenvalues. Plot comparing the eigenvalues for each PC for the PCA of HSPC using the top 300 ANOVA genes (black) and 300 randomly selected genes from a randomised expression matrix (red). See Figures 3.24 and 3.27 for PCA plots.

expression of many lymphoid-cell-associated genes (for example, *IL-7R*, *LCK*, *SYK*, *ADA*, *HLX*, *LST1* and *ITGAL*) (Figure 3.29).

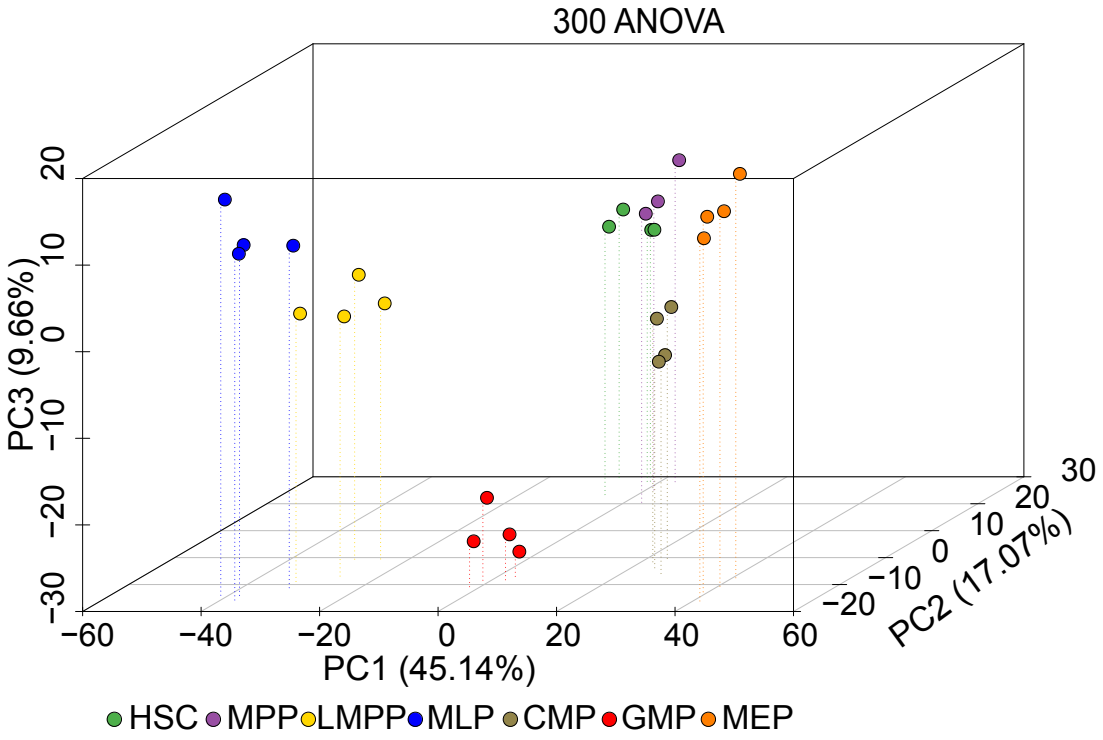


Figure 3.27: PCA of Normal CB HSPC: PC1–PC3. PCA of normal CB HSPC populations using the 300 most DEG across normal HSPC populations, as determined by an ANOVA-like test (see [54], Supplementary Table 4). The percentage variance contributed by each PC is indicated and PCs 1–3 are shown. Each point represents a biological replicate.

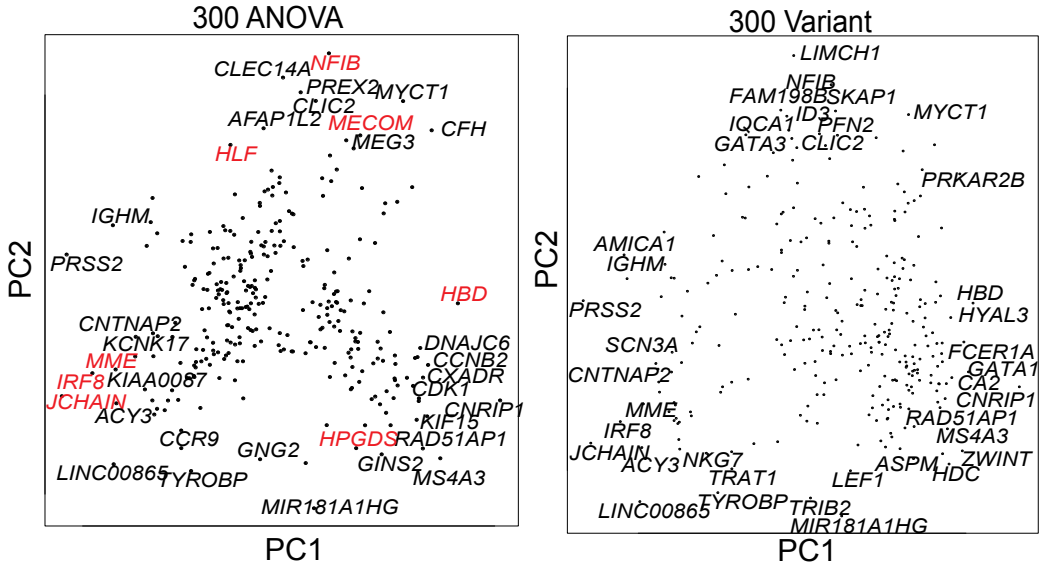


Figure 3.28: PCA of Normal CB HSPC: Loading Plots. Loading plot showing the genes with the most extreme loadings scores for the PCA run with the top 300 ANOVA genes (left) or the 300 genes with the greatest variance in expression (right).

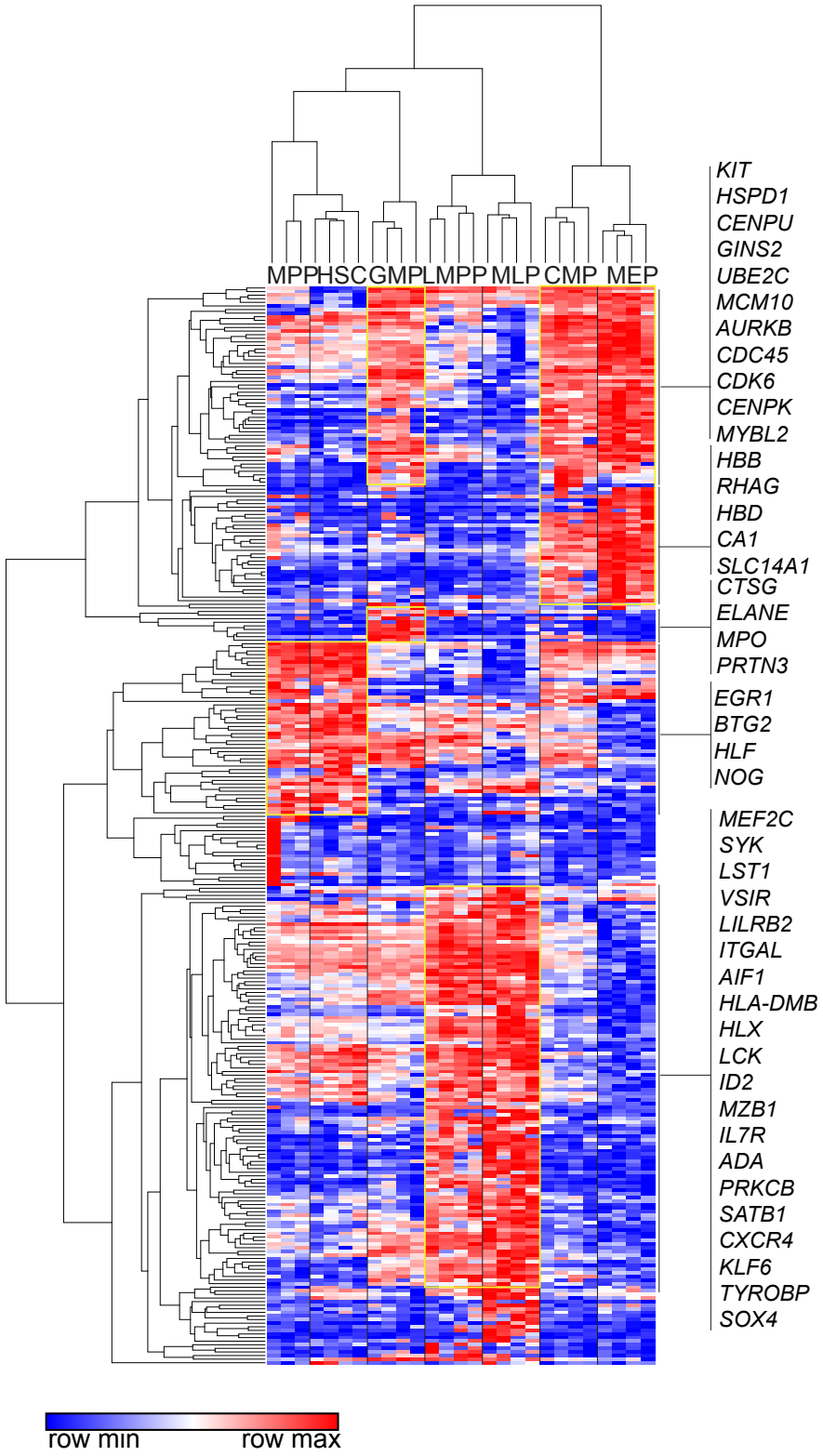


Figure 3.29: Heatmap Showing the Expression of 300 DEGs Genes by HSPC. Heatmap showing the expression of the 300 most DEGs across normal HSPC populations, as determined by an ANOVA-like test (see [54], Supplementary Table 4). Rows represent genes and columns represent a biological replicate of an HSPC population. Colour scale is normalised per row.

3.5.2 Transcriptional Distance Between HSPC Populations

The transcriptional distance between two populations may be represented by the number of DEGs between them. In order to gain insight into this transcriptional distance, pairwise comparisons were drawn between all HSPC populations. This allowed the relationship between populations to be represented without the assumption of any hierarchical relationship. The most closely related populations were HSCs and MPPs (only 13 DEGs separated these populations), while LMPPs and MLPs were also closely related (85 DEGs) (Figure 3.30 and Tables A.17). GMPs were most closely related to CMPs (40 DEGs) but retained similarity to LMPPs (183 DEGs) (Figure 3.30 and Tables A.17). For the DEGs identified from all pairwise comparisons see [54], Supplementary Tables 5-11.

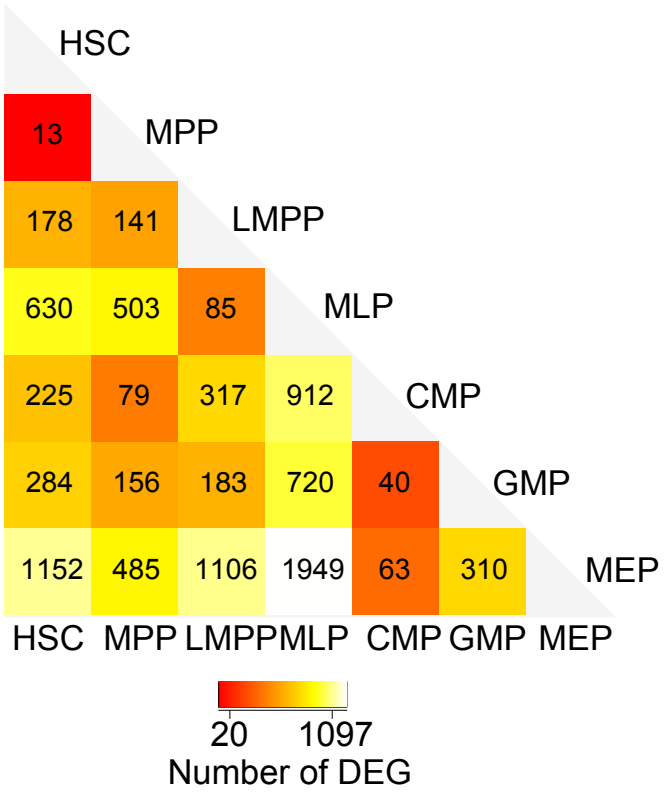


Figure 3.30: DEG Between HSPC Populations. DEG were identified between HSPC populations and shown in a heatmap where rows and columns represent an HSPC population. Colour scale represents the number of DEG. See Table A.17.

3.5.3 Expression of Lineage- and LM-Associated Gene Signatures

Following on from the global analysis of the HSPC populations, analysis was then focussed on the LM progenitors. Gene signatures for the LMPP, MLP and GMP were generated using a comparison of each population with all other HSPC populations. These genes were then filtered for genes unique amongst the LM progenitors and ranked by p-adjusted value (see [54], Supplementary Table 12).

It was seen that these gene signatures had distinct elevated expression in their respective populations (Figure 3.31). The GMP signature included many myeloid-associated genes, and the MLP signature included many lymphoid-associated genes. In contrast, the LMPP signature included both lymphoid-associated genes (*ETS1*, *EBF1* and *CYTIP*) and myeloid-associated genes (*TRPM2*, *S100A8*, *PADI4* and *ALOX15B*) (Figure 3.31).

To confirm the lineage-association of the LM Progenitors, the expression of previously identified lineage-associated genes by the LM progenitors was investigated [28]. As expected, GMPs expressed genes associated with immature myeloid cells and neutrophils. On the other hand, LMPPs and MLPs expressed genes associated with B cells, monocytes and DCs but not those associated with neutrophils (with the exception of *FOSB*) Figure 3.32).

Furthermore, a MetaCore enrichment analysis was carried out on the DEG identified to be upregulated in each of the LM progenitors when compared to the remaining HSPC populations. In this analysis the GMP showed enrichment for pathways associated with myeloid maturation (for example, granulocyte development (false-discovery rate (FDR) = 0.0136)), whereas MLPs showed enrichment for lymphopoiesis pathways (for example, Notch signalling (FDR $<$ 0.001)) (Table 3.4). LMPPs showed more balanced enrichment for both lymphoid pathways and myeloid pathways (for example, M-CSF signalling (FDR $<$ 0.001) and BCR signalling (FDR = 0.049)) (Table 3.4).

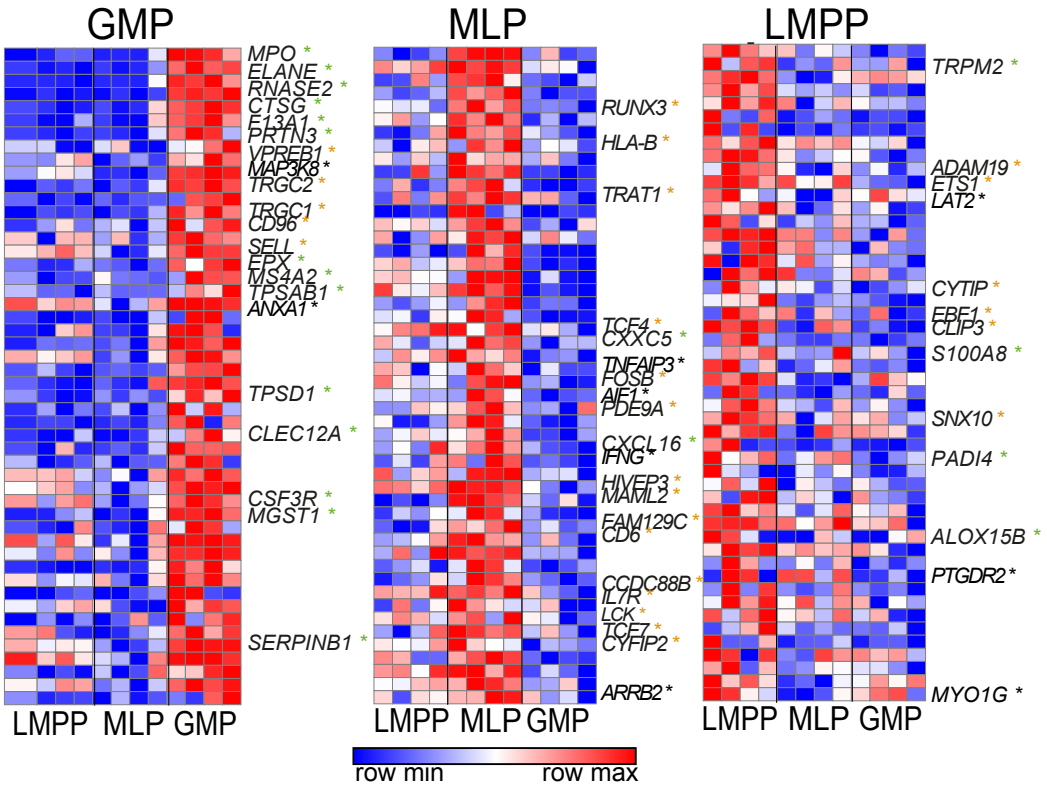


Figure 3.31: Gene Signatures of the LM Progenitors. Gene signatures were derived for the LMPP, MLP and GMP populations. DEGs were identified between each population and the remaining 6 HSPC populations, and these genes were filtered for genes unique across the LMPP, MLP and GMP populations. Shown are the 50 most significantly DEGs. Asterisk colours indicate genes associated with the lymphoid lineage (orange), myeloid lineage (green) or genes encoding products associated with immunological function (black). Rows represent genes and columns represent a biological replicate of an HSPC population. Colour scale is normalised per row.

Table 3.4: Enriched Pathways by the LM Progenitors. DEGs were identified between each of the LM Progenitors and the remaining HSPC populations. Metacore enrichment analysis was carried out on these DEGs, and the 10 most significant pathways are shown.

Metacore Pathway	FDR	Genes in Pathway
LMPP Upregulated Genes		
Immune response: M-CSF-receptor signalling pathway	6.9e-04	SYK, TYROBP, PRKCB, TCF7L2, GRB2, MSR1, ETS1, FOS
Immune response: Antigen presentation by MHC class II	1.1e-03	CD3D, CD3E, LGMN, CD74, HLA-DMB
Aberrant production of IL-2 and IL-17 in SLE T cells	3.8e-03	SYK, CD3D, CD3E, GRB2, ETS1, FOS

Continuation of Table 3.4		
Metacore Pathway	FDR	Genes in Pathway
Immune response: KLRK1 (NKG2D) signalling pathway	1.2e-02	SYK, TYROBP, GRB2, FOS, HCST
Immune response: TREM1 signalling pathway	2.9e-02	SYK, TYROBP, LAT2, GRB2, FOS
HBV signalling via protein kinases leading to HCC	3.5e-02	PRKCB, GRB2, FOS
Immune response: CD16 signalling in NK cells	4.3e-02	SYK, CYTH1, CYTIP, GRB2, FCRL3, FOS
Immune response: B cell antigen receptor (BCR) pathway	4.9e-02	SYK, PRKCB, GRB2, ETS1, FOS
Cell adhesion: Integrin inside-out signalling in neutrophils	5.3e-02	SYK, TYROBP, CD44, LTB4R, CYTH1, CYTIP
Immune response: MIF-induced cell adhesion, migration and angiogenesis	5.3e-02	CD44, CXCR4, CD74, FOS
MLP Upregulated Genes		
Reproduction: Gonadotropin-releasing hormone (GnRH) signalling	6.0e-06	JUNB, FOSB, CACNA1C, DUSP4, MEF2D, DNM1, FOS, FOSL1, JUND, PER1, ELK1, PRKACA, PRKACB, RPS6KA2, GNAS, DUSP1
Cell adhesion: Integrin inside-out signalling in T cells	6.0e-06	RASSF5, ITGB2, STK4, ICAM1, PRKCA, RAPGEF1, PRKACA, ITGAL, ITGB2, CYTH1, PIK3R1, GNAI1, GNAI2, GNAI3, LCK, WASF2, CXCR4, CCR7
Development: Notch signalling Pathway	2.7e-05	NOTCH1, UBA52, FURIN, HES1, CIR1, NFKBIA, HIST1H4C, HIST1H4F, HIST1H4I, HIST4H4, H3F3A, H3F3B
Immune response: MIF-induced cell adhesion, migration and angiogenesis	4.3e-05	ICAM1, ITGAL, ITGB2, PIK3R1, FOS, FOSB, FOSL1, JUNB, JUND, MMP9, GNAI1, GNAI2, GNAI3, CXCR4, CD74, DUSP1, PAK1, FOS
Immune response: MIF-mediated glucocorticoid regulation	5.7e-05	ICAM1, NR3C1, NFKBIA, ZYX, NFKBIA, IFNG, CD74, FOS
Transcription: CREB signalling pathway	5.7e-05	CACNA1C, PRKCA, IGF1R, CAMK4, MAP3K11, CCDC88B, PRKACA, PRKCB, PRKACB, RPS6KA2, GNAS, PIK3R1

Continuation of Table 3.4		
Metacore Pathway	FDR	Genes in Pathway
Development: Thromboxane A2 signalling pathway	6.1e-05	LEF1, TCF7, TCF3, TCF4, TCF7L2, CCDC88B, PIK3R1, PRKCB, PRKACA, GNA12, GNAI1, GNAI2, GNAI3, PRKACA, PRKACB, GNAS, PRKACA, PRKCB, PRKD2, PIK3R1, ARHGEF1
Immune response: ICOS signalling pathway in T-helper cell	8.5e-05	STK4, NFATC2, BCL6, IL10, PIK3R1, MAF, FOS, FOSB, FOSL1, JUNB, JUND, PIK3R1, IFNG, PAK1, FOS
Development: Role of IL-8 in angiogenesis	1.1e-04	UBA52, ADD1, NFKBIA, PIK3R1, ELK1, GNAI1, GNAI2, GNAI3, FOS, CAV1
Immune response: KLRK1 (NKG2D) signalling pathway	1.9e-04	SYK, ULBP1, FOS, FOSB, FOSL1, JUNB, JUND, TYROBP, LCK, IFNG, PAK1, PIK3R1, FOS, HCST
GMP Upregulated Genes		
Immune response: Antigen presentation by MHC class I: cross-presentation	8.8e-03	HSP90B1, TYROBP, TLR4, CLEC12A, CALR, FCGR2A
Immune response: Antigen presentation by MHC class I, classical pathway	1.4e-02	HSP90B1, CD99, CALR, CALNX, ERAP2
Immune response: TREM1 signalling pathway	1.4e-02	TYROBP, TLR4, LAT2, HGF
Development: Transcription regulation of granulocyte development	1.4e-02	CSF3R, ELANE, PRTN3, MPO
Transport: Macropinocytosis regulation by growth factors	1.4e-02	HGF, CSF1R, PDE3B
SLE genetic marker-specific pathways in B cells	1.4e-02	CD38, TLR4, IRF8, TNFSF13B, HGF, FCGR2B
Renal tubulointerstitial injury in Lupus Nephritis	1.4e-02	CSF1R, CD44, VIM, HGF, TNFSF13B
Cell adhesion: Cell-matrix glycoconjugates	1.6e-02	ELANE, MUC1, CD44, SELL
Development: Role of proteases in hematopoietic stem cell mobilization	1.8e-02	ELANE, CD44, CTSG
Protein folding and maturation: Angiotensin system maturation Human version	1.8e-02	ELANE, CPA3, PRTN3, CTSG

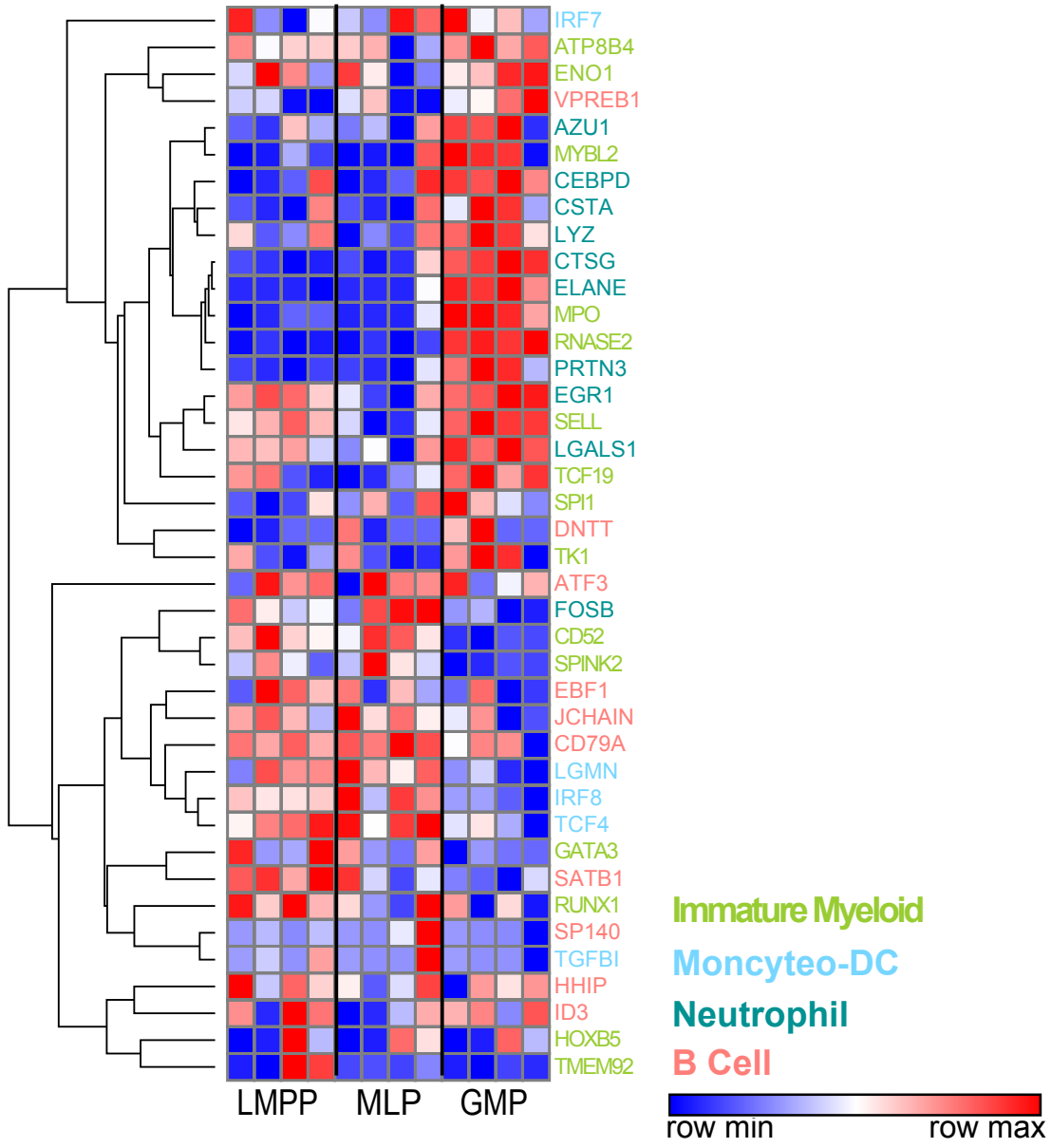
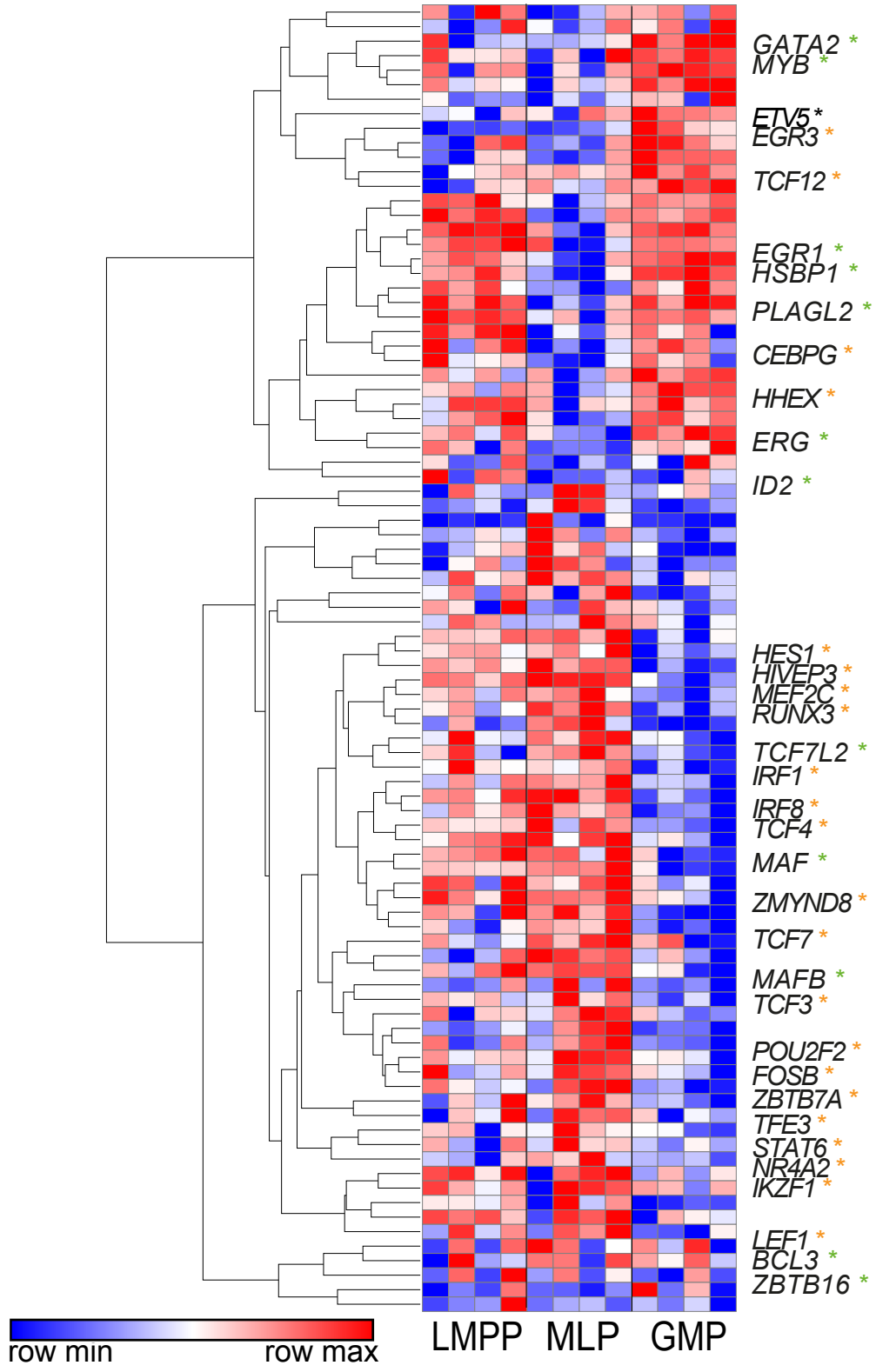


Figure 3.32: Expression of Lineage-Associated Genes by LM Progenitors. The expression of lineage-associate genes (identified by Velten et al. [28]) by the LM progenitors is shown. Genes are colour-coded according to their classification. Expression values are normalised per gene. Rows represent genes and columns represent a biological replicate of an HSPC population. Colour scale is normalised per row.

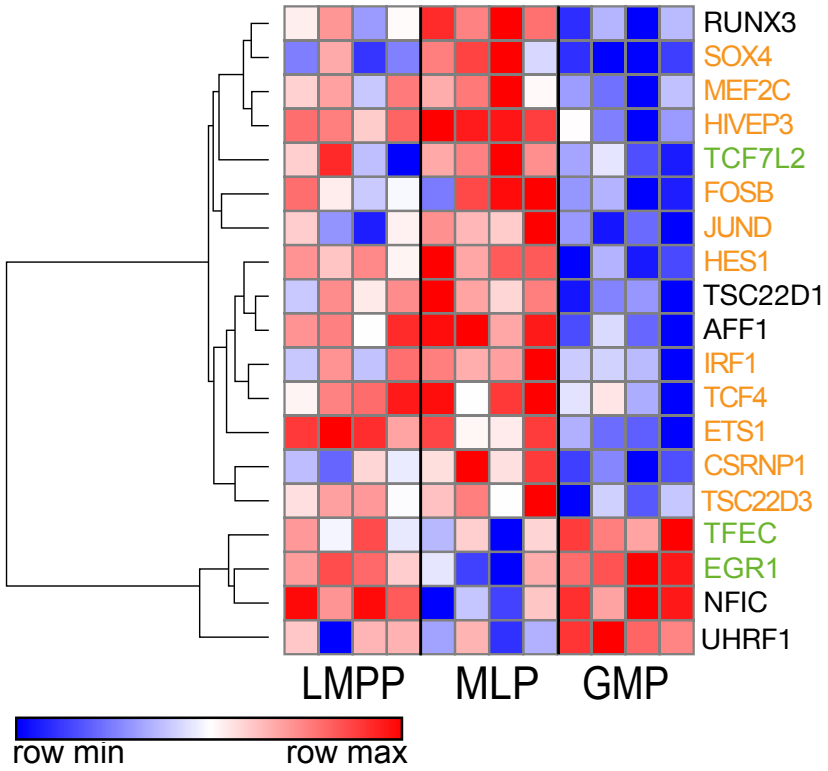
Two approaches were then used to identify transcription factors (TFs) driving the lineage-associated gene expression patterns in the LM progenitors. First, the expression of TFs previously identified to play a role in haematopoiesis was assessed [94] (Figure 3.33a). These TFs were refined for those that showed distinct patterns of expression across the LM progenitors. Secondly, the MLP and GMP were taken as the most lymphoid and myeloid populations, respectively, and differentially expressed TFs between the two populations identified (Figure 3.33b).

In both analyses, GMPs expressed mainly genes encoding myeloid-associated TFs (for example, *ERG*, *GATA2*, *MYB* and *EGR1*), while lymphoid-associated TFs (for example, *HES1*, *RUNX3*, *POU2F2*, *LEF1*, *IKZF1*, *IRF8*, *TCF4*) had their highest expression in MLPs (Figure 3.33). LMPPs showed expression of both lymphoid-associated TFs and myeloid-associated TFs (Figure 3.33).

Additionally, a similar trend was seen for genes encoding cytokine and chemokine receptors (Figure 3.34).



(a) Haematopoietic-Related TF [94]



(b) Differentially Expressed TFs Between the MLP and GMP

Figure 3.33: Expression of TFs by the LM Progenitors. Two approaches were carried out to investigate the expression of lineage-associated TFs by the LM progenitors. (a) The expression of TFs previously identified were investigated [94] and (b) TFs differentially expressed between the MLP and GMP were identified. Asterisk/gene name colours indicate genes associated with the lymphoid lineage (orange), myeloid lineage (green) or genes encoding products associated with immunological function (black). Rows represent genes and columns represent a biological replicate of an HSPC population. Colour scale is normalised per row.

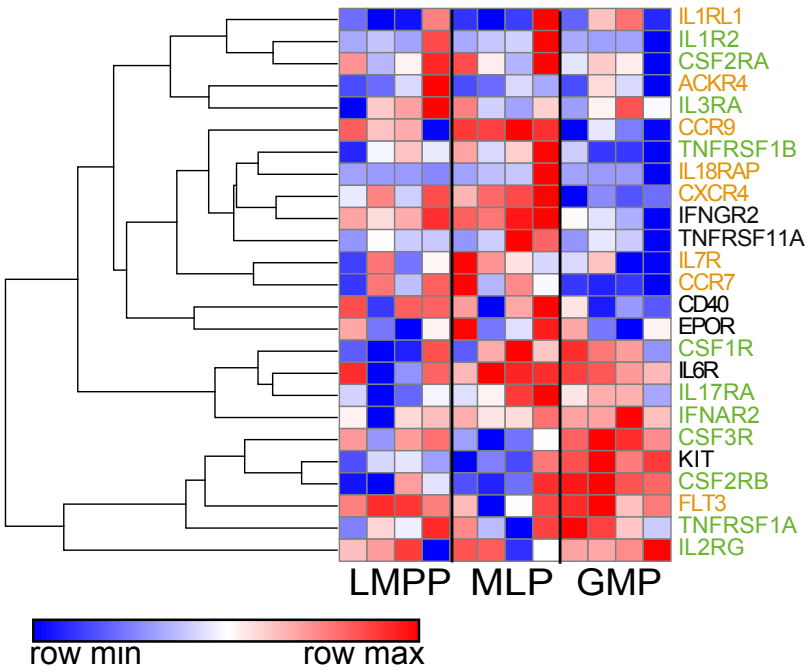


Figure 3.34: Expression of Lineage-Associated Cytokine Genes by LM Progenitors. The expression of lineage-associated cytokine genes by the LM progenitors is shown. gene name colours indicate genes associated with the lymphoid lineage (orange), myeloid lineage (green) or genes encoding products associated with immunological function (black). Expression values are normalised per gene. Rows represent genes and columns represent a biological replicate of an HSPC population. Colour scale is normalised per row.

3.6 Single-Cell Transcriptional Analysis of LM Progenitors

After characterising the LM progenitors transcriptionally at the population level, the transcriptional heterogeneity within these populations was investigated by single-cell analysis. Single-cell transcriptional analysis was carried out by both single-cell qRT-PCR and single-cell RNA-Seq. Single-cells were sorted by index flow cytometry for both of these single-cell transcriptional assays. Additionally, the single-cells previously sorted for functional assays were also index sorted (see Chapter 3.3.4). The expression of cell surface markers recorded by the index flow sorting data was used as a commonality between the single-cell functional and single-cell transcriptional data. This commonality allowed the correlation of the single-cell functional and transcriptional data, as done previously [95].

3.6.1 Single-Cell qRT-PCR of LM Progenitors

Initially, the expression of 96 lineage-associated genes by 919 single LMPPs, MLPs and GMPs was profiled ([54], Supplementary Table 14). Genes with low variance and levels of detection were excluded, leaving 74 genes that were taken forward for analysis.

Hierarchical clustering assigned GMPs, LMPPs and MLPs to three clusters: cluster 1 (543 cells) was mainly MLPs and LMPPs, cluster 2 (150 cells) was a mixture of GMPs, LMPPs and MLPs, and cluster 3 (226 cells) was mainly GMPs (Figure 3.35 and Table 3.5).

Table 3.5: Single-Cell qRT-PCR of LM Progenitors: Cell-Type Composition of Clusters. The expression of 74 lineage-associated genes by 919 single LMPPs, MLPs and GMPs using qRT-PCR was investigated. Cell type composition of the 3 clusters identified in the dendrogram in Figure 3.35 is shown.

Cell Type	Cluster 1	Cluster 2	Cluster 3
GMP	25	80	203
LMPP	230	57	19
MLP	288	13	4

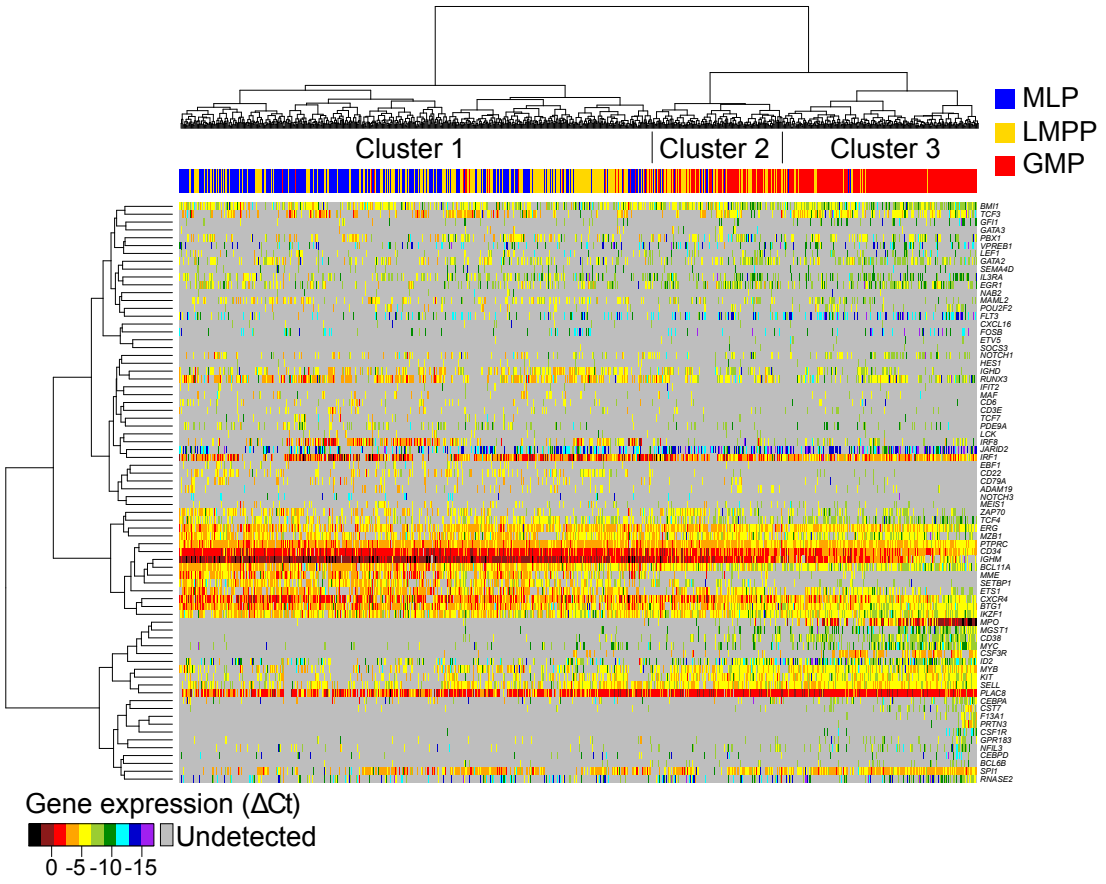


Figure 3.35: Single-Cell qRT-PCR of LM Progenitors: Hierarchical Clustering. The expression of 74 lineage-associated genes by 919 single LMPPs, MLPs and GMPs using qRT-PCR was investigated. Heatmap showing the hierarchical clustering based on the expression of these genes, identifying 3 major clusters of cells. Rows indicate genes and columns indicate a single cell.

To assess the lineage-association of the clusters, DEGs between pairs of the clusters were identified. Cluster 1 showed higher expression of lymphoid-associated genes than that of the other clusters, and cluster 3 showed higher expression of myeloid-associated genes than that of the other clusters (Table 3.6). Cluster 2 had a mixed LM expression profile (Table 3.6).

Table 3.6: Single-Cell qRT-PCR of LM Progenitors: DEG Between Clusters. The expression of 74 lineage-associated genes by 919 single LMPPs, MLPs and GMPs using qRT-PCR was investigated. DEG between cluster pairs of the 3 clusters identified in the dendrogram in Figure 3.35 are shown. Orange: lymphoid-associated genes. Green: myeloid-associated genes.

A	B	High A vs B	High B vs A
Cluster 1	Cluster 3	CD22, PDE9A, ADAM19, MAF, CD3E, TCF7, CD6, IRF8, SETBP1, CD79A, MME, NOTCH3, EBF1	MYC, CD38, CSF1R, NFIL3, CSF3R, ID2, POU2F2, BCL6B, CST7, MGST1, PRTN3, SOCS3, CEBPA, ETV5, GFI1, KIT, MPO, F13A1, GPR183, MYB
Cluster 1	Cluster 2	CD22, PDE9A, MAF, CD6, IRF8, MME, EBF1	CD38, BCL6B, MGST1, GATA3, ETV5, KIT, MPO
Cluster 2	Cluster 3	SETBP1, GATA3, MME	MYC, TCF3, CD38, CSF1R, CSF3R, CST7, MGST1, PRTN3, CEBPA, GFI1, MPO, F13A1, GPR183

Table 3.6 suggests that cluster 1 represented a more lymphoid cluster, cluster 2 an LM cluster and cluster 3 a more myeloid cluster. When looking at the number of LM progenitors with a particular functional output, it was seen that the proportions of LM progenitors with Ly, Ly-My and My output mirrored the number of LM progenitors in each of the clusters (Table 3.7).

Table 3.7: Single-Cell qRT-PCR of LM Progenitors: Expected Numbers in Clusters. The expression of 74 lineage-associated genes by 919 single LMPPs, MLPs and GMPs using qRT-PCR was investigated. The expected number of cells per cluster, based on the functional assays, in the 3 clusters identified in the dendrogram in Figure 3.35 are shown.

Cell Type	Ly	Ly-My	My
GMP	49	27	232
LMPP	229	46	30
MLP	285	10	10

Next, a diffusion map method of dimensionality reduction of gene-expression data, adapted for single-cell data, was utilised [64, 96]. Progenitor identity was indicated on the diffusion map and it was observed that MLP, LMPP and GMP cells formed a continuum (Figure 3.36). This was in agreement with the hierarchical clustering

(Figure 3.35). When the diffusion map was annotated by cluster assignment, it could be seen that cluster 2 was positioned between clusters 1 and 3, in agreement with its mixed LM transcriptional signature (Figure 3.35).

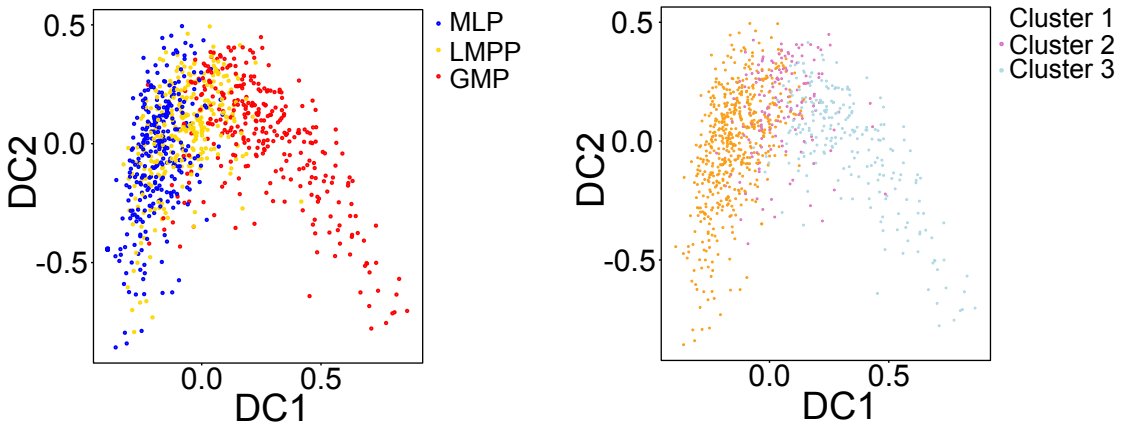


Figure 3.36: Single-Cell qRT-PCR of LM Progenitors: Diffusion Maps. The expression of 74 lineage-associated genes by 919 single LMPPs, MLPs and GMPs using qRT-PCR was investigated. Diffusion map dimensionality reductions coloured in by cell type (left) and cluster membership (right), based on qRT-PCR gene expression analysis.

3.6.2 Single-Cell RNA-Seq of LM Progenitors

Although informative, the single-cell qRT-PCR was limited by gene-selection. To overcome this, single-cell RNA-Seq was then carried out on 91 LMPPs, 110 MLPs and 119 GMPs from two different donors (157 cells and 163 cells from each donor). Genes were filtered to those above the level of technical variance as calculated from ERCC controls. Variable genes were then identified from both of the donors and the union of these two variable gene sets determined. Clustering of both donors using this union of variable genes identified 3 clusters (Figure 3.37).

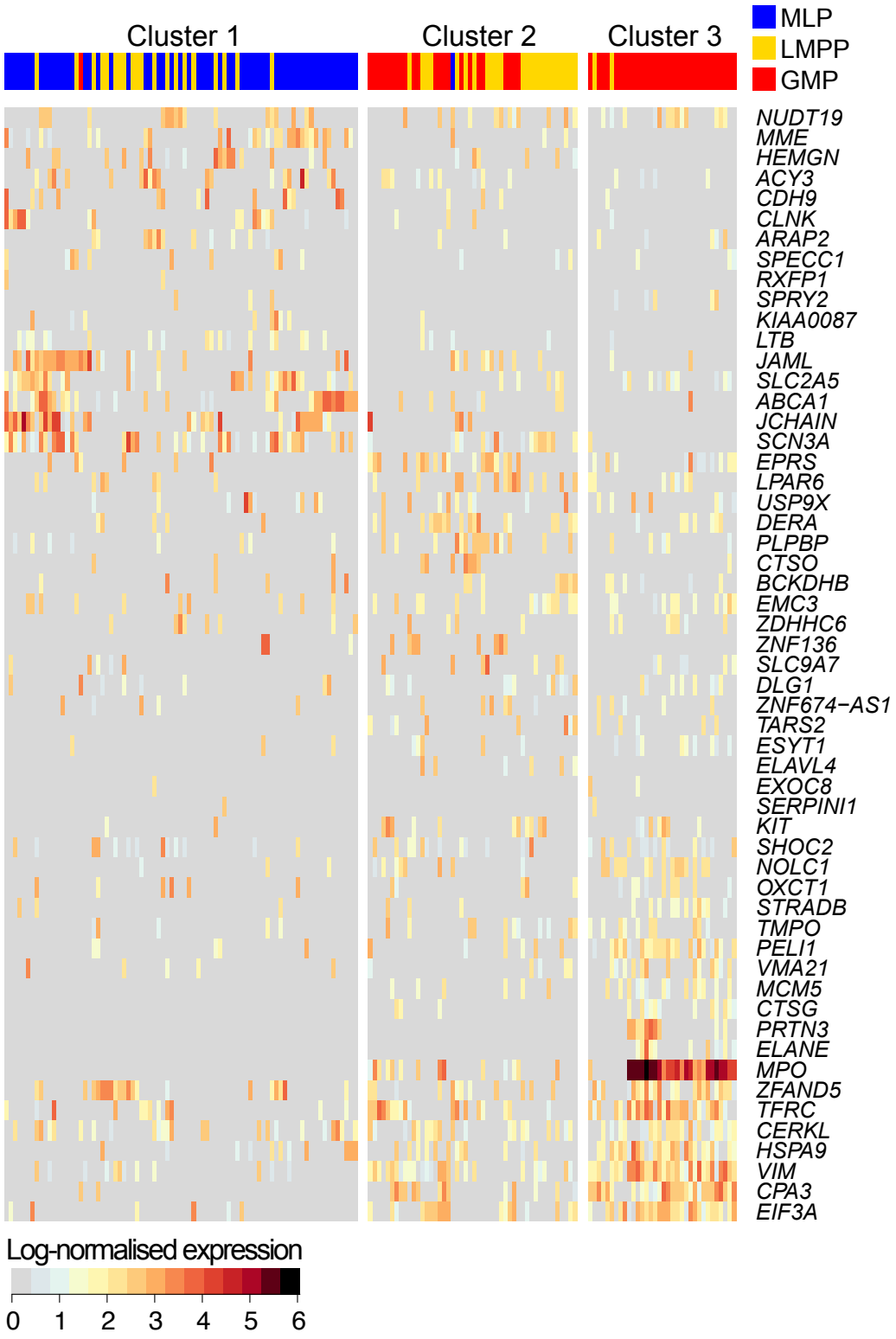
Most cells in cluster 1 were MLPs and most cells in cluster 3 were GMPs, while cluster 2 comprised LMPPs and GMPs (Figure 3.37 and Table 3.8).

Cluster 1 showed high expression of lymphoid-associated genes (for example, *MME*, *JCHAIN* and *ABCA1*) (Figure 3.37). Cluster 3 showed higher expression of myeloid genes (for example, *CPA3*, *MPO* and *VIM*) than that of the other clusters. Cluster 2 showed a mixed transcriptional signature and higher expression of the immature-associated gene *KIT* than that of the other clusters (Figure 3.37).

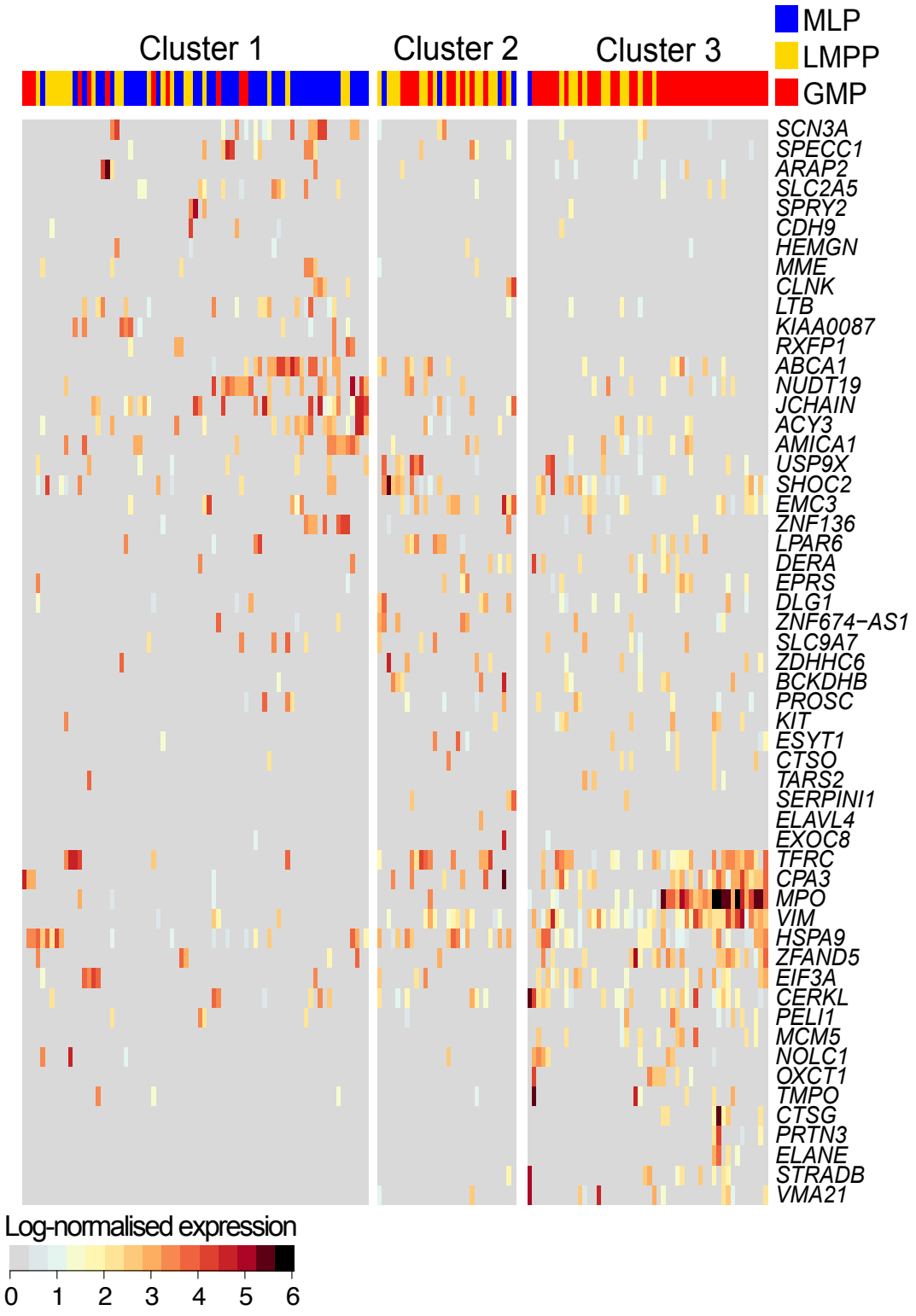
Table 3.8: Single-Cell RNA-Seq of LM Progenitors: Cell-Type Composition of Clusters. Single-cell RNA-Seq was carried out on 91 LMPPs, 110 MLPs and 119 GMPs from two different donors (157 cells and 163 cells from each donor). Cell type composition of the 3 clusters identified in Figure 3.37 is shown.

Cell Type	Cluster 1	Cluster 2	Cluster 3
Donor 1			
GMP	1	23	32
LMPP	20	24	2
MLP	60	1	0
Donor 2			
GMP	11	11	41
LMPP	20	15	10
MLP	44	4	1

PCA was then carried out on the single-cell RNA-Seq profiles of the LM progenitors. This revealed a transcriptional continuum of LMPP, MLP and GMP populations (Figure 3.38). This also reflected the results seen by the diffusion map analysis of the single-cell qRT-PCR data (Figure 3.36).

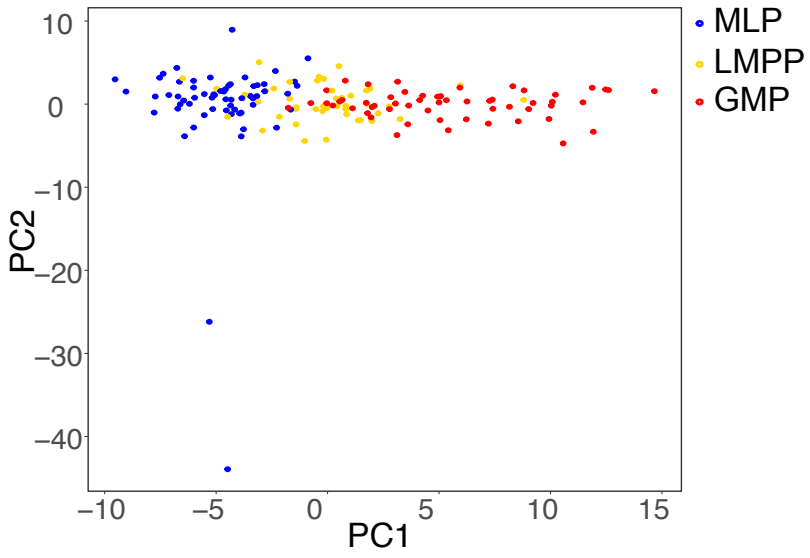

Cluster 1
Cluster 2
Cluster 3
MLP
LMPP
GMP
NUDT19
MME
HEMGN
ACY3
CDH9
CLNK
ARAP2
SPECC1
RXFP1
SPRY2
KIAA0087
LTB
JAML
SLC2A5
ABCA1
JCHAIN
SCN3A
EPRS
LPAR6
USP9X
DERA
PLPBP
CTSO
BCKDHB
EMC3
ZDHHC6
ZNF136
SLC9A7
DLG1
ZNF674−AS1
TARS2
ESYT1
ELAVL4
EXOC8
SERPINI1
KIT
SHOC2
NOLC1
OXCT1
STRADB
TMPO
PELI1
VMA21
MCM5
CTSG
PRTN3
ELANE
MPO
ZFAND5
TFRC
CERKL
HSPA9
VIM
CPA3
EIF3A
Log-normalised expression
0 1 2 3 4 5 6

(a) Donor 1

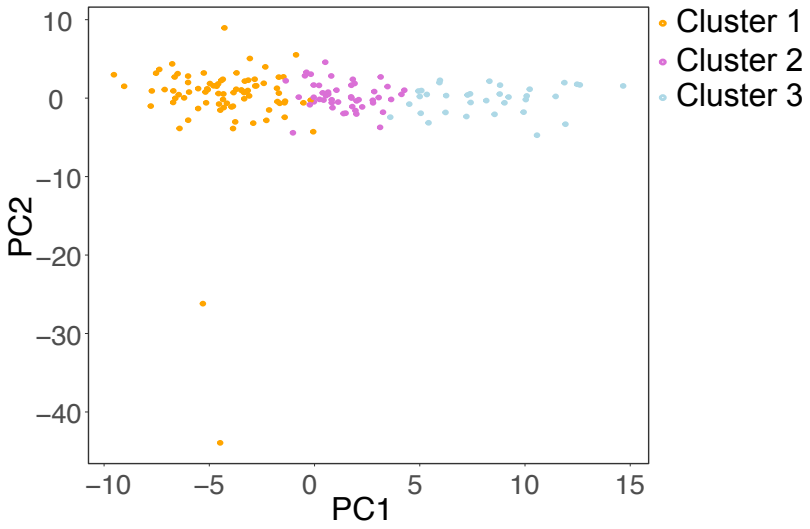

Cluster 1
Cluster 2
Cluster 3
MLP
LMPP
GMP
SCN3A
SPECC1
ARAP2
SLC2A5
SPRY2
CDH9
HEMGN
MME
CLNK
LTB
KIAA0087
RXFP1
ABCA1
NUDT19
JCHAIN
ACY3
AMICA1
USP9X
SHOC2
EMC3
ZNF136
LPAR6
DERA
EPRS
DLG1
ZNF674−AS1
SLC9A7
ZDHHC6
BCKDHB
PROSC
KIT
ESYT1
CTSO
TARS2
SERPINI1
ELAVL4
EXOC8
TFRC
CPA3
MPO
VIM
HSPA9
ZFAND5
EIF3A
CERKL
PELI1
MCM5
NOLC1
OXCT1
TMPO
CTSG
PRTN3
ELANE
STRADB
VMA21
Log-normalised expression
0 1 2 3 4 5 6

(b) Donor 2

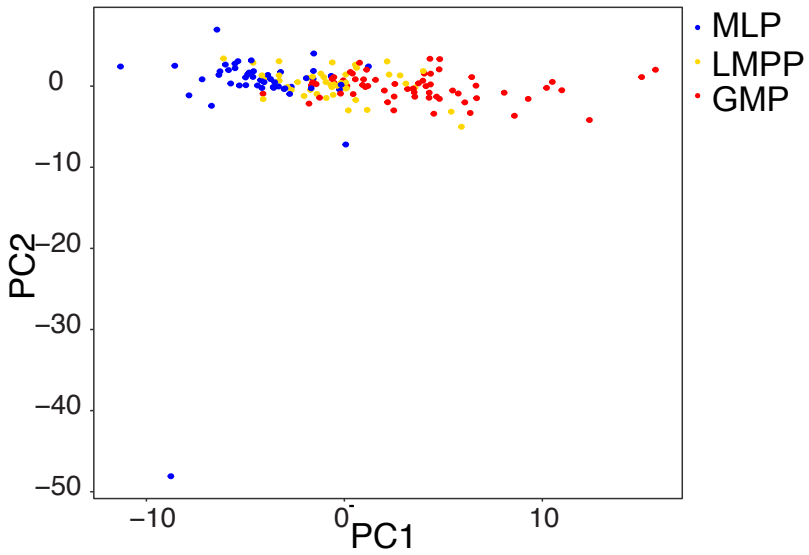
Figure 3.37: Single-Cell RNA-Seq of LM Progenitors: Hierarchical Clustering. Single-cell RNA-Seq was carried out on 91 LMPPs, 110 MLPs and 119 GMPs from two different donors (157 cells and 163 cells from each donor). Heatmaps show the hierarchical clustering based on the expression of 55 genes with the highest and most variable expression between clusters from (a) donor 1 and (b) donor 2. Rows indicate genes and columns indicate a single cell.



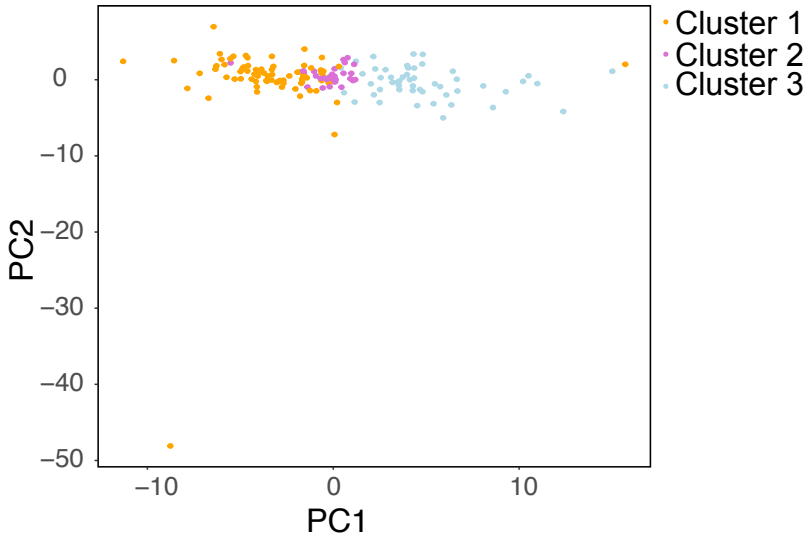
(a) Donor 1: Cell Type



(b) Donor 1: Cluster Identity



(c) Donor 2: Cell Type



(d) Donor 2: Cluster Identity

Figure 3.38: Single-Cell RNA-Seq of LM Progenitors: PCA. Single-cell RNA-Seq was carried out on 91 LMPPs, 110 MLPs and 119 GMPs from two different donors (157 cells and 163 cells from each donor). PCA on (a–b) Donor 1 and (c–d) Donor 2 using 5,388 genes with highly variable expression (above the level of technical variance calculated from external control RNA). Cell type and cluster identity of cells is indicated.

3.7 Refining the LMPP and GMP Populations

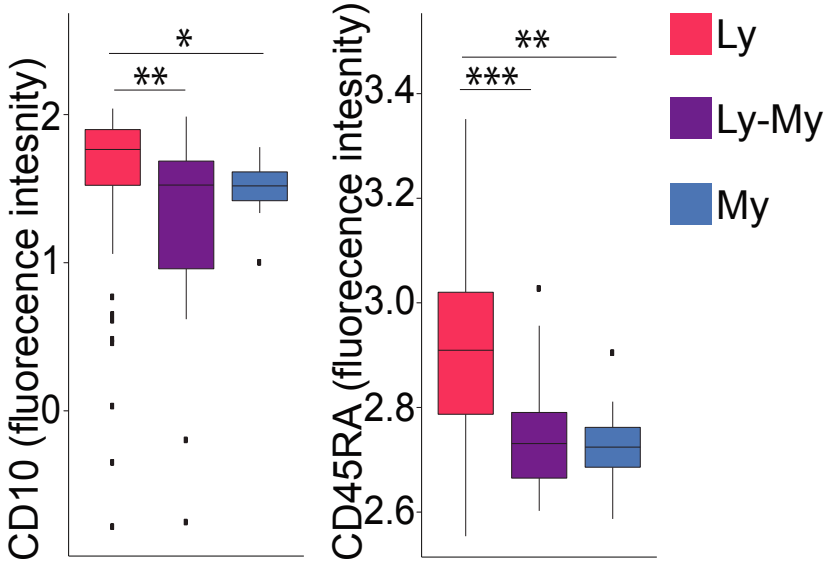
It was apparent from both the single-cell functional and transcriptional data that the current immunophenotypic definitions of the LM progenitors do not purify functionally pure populations. The MLP was predominantly a lymphoid progenitor, representing the most functionally pure LM progenitor. However, the GMP retained considerable lymphoid potential and the LMPP was the most heterogenous LM progenitor. Using the cell-surface marker data from index sorting, functional potential was correlated with immunophenotype with the aim of refining the LMPP and GMP to more functionally pure progenitors.

3.7.1 Refining the LMPP: LMPP^{ly} and LMPP^{mix}

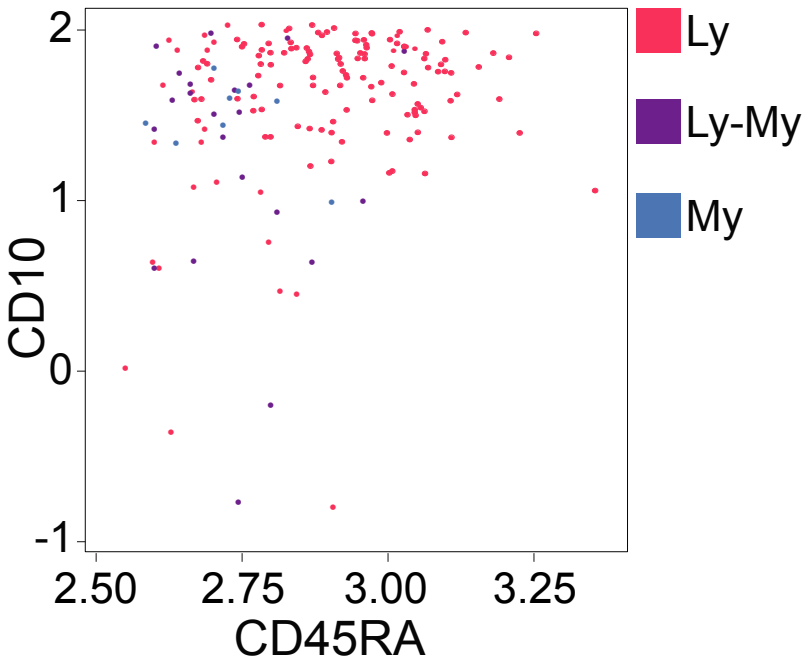
Flow cytometry index data demonstrated that single LMPP cells with lymphoid output had significantly higher expression of CD10 and CD45RA, when compared to LMPPs with myeloid and LM output (Figure 3.39). This was further confirmed by the higher expression of CD10 in single LMPPs in the lymphoid-biased cluster 1 than in single LMPPs in the myeloid-biased cluster 3 (Figure 3.39c).

Therefore, a new flow sorting strategy was developed for the LMPPs to purify two new sub-populations within the CD10^{-}CD45RA^{+} gates. These were the CD10^{hi}CD45RA^{hi} LMPPs (called "LMPP^{ly} cells" here) and CD10^{lo}CD45RA^{lo} LMPPs (called "LMPP^{mix} cells" here) (Figure 3.40). The aim of this refinement was to maximise the lymphoid-only potential and the mixed myeloid and LM potential in the LMPP^{ly} and LMPP^{mix}, respectively.

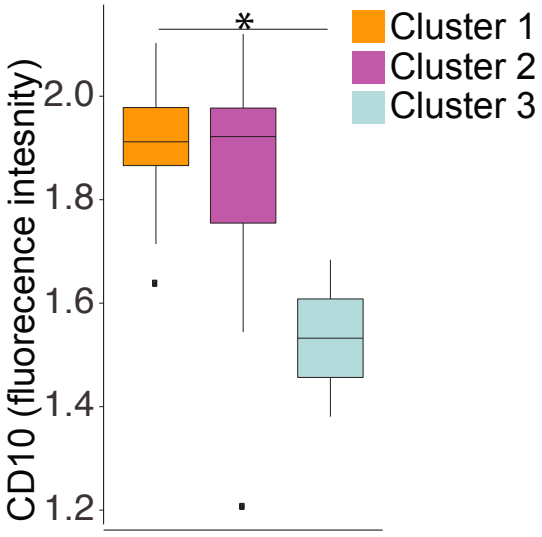
It was seen that 26% of total LMPPs were LMPP^{ly} cells and 27% were LMPP^{mix} cells (Figure 3.41). When cultured in SGF15/2 conditions and analysed after 2 weeks, LMPP^{ly} cells had significantly lower cloning efficiency than that of LMPPs and LMPP^{mix} cells, but significantly higher cloning efficiency than that of MLPs (Figure 3.42 and Table A.12; $P < 0.0001$, for all comparisons (Fisher's exact test)).



(a) CD10 and CD45RA Expression: Boxplots



(b) CD10 and CD45RA Expression: Scatter Plot



(c) CD10 Expression: Cluster Identity

Figure 3.39: Expression of CD10 and CD45RA by Single LMPP Cells. (a) Expression of the surface markers CD10 and CD45RA on LMPPs, grouped by functional output as lymphoid (composed of uni-lymphoid output (B or NK) and bi-lymphoid output (B and NK)), myeloid (composed of uni-myeloid output (M or G) and bi-myeloid output (M and G)) or LM. CD10, P = 0.0052, lymphoid versus LM, and P = 0.027, lymphoid versus myeloid; CD45RA, P = 4.8 0^{-6}, lymphoid versus LM, and P = 0.0027, lymphoid versus myeloid (Wilcoxon rank sum test). (b) Expression of CD10 and CD45RA on CB LMPPs, measured by flow cytometry. Colours indicate output from functional assays. (c) CD10 fluorescence intensity in LMPPs in the 3 clusters identified from clustering in Figure 3.37. Wilcoxon rank sum test is used to define significance. Expression values calculated with the "flowCore::logicleTransform" function in R software.

LMPP^{ly} cells were lymphoid progenitors with almost no myeloid potential and significantly lower myeloid potential than that of LMPPs or LMPP^{mix} cells (Figures 3.42 and 3.43 and Table A.12; P = 0.0496 and P = 0.0280, respectively (Fisher's exact test)). LMPP^{ly} cells had very limited residual LM potential (1.6%) (Figures 3.43 and 3.44).

LMPP^{mix} cells retained almost all of the myeloid potential and most of the LM potential from the LMPP (Figures 3.42 and 3.43 and Table A.12). See Table A.11 for details on all output from cultures in SGF15/2 conditions.

This suggested that functionally, LMPP^{ly} cells were an intermediate between LMPPs and MLPs. That was confirmed by a second *in vitro* culture (SF7b)

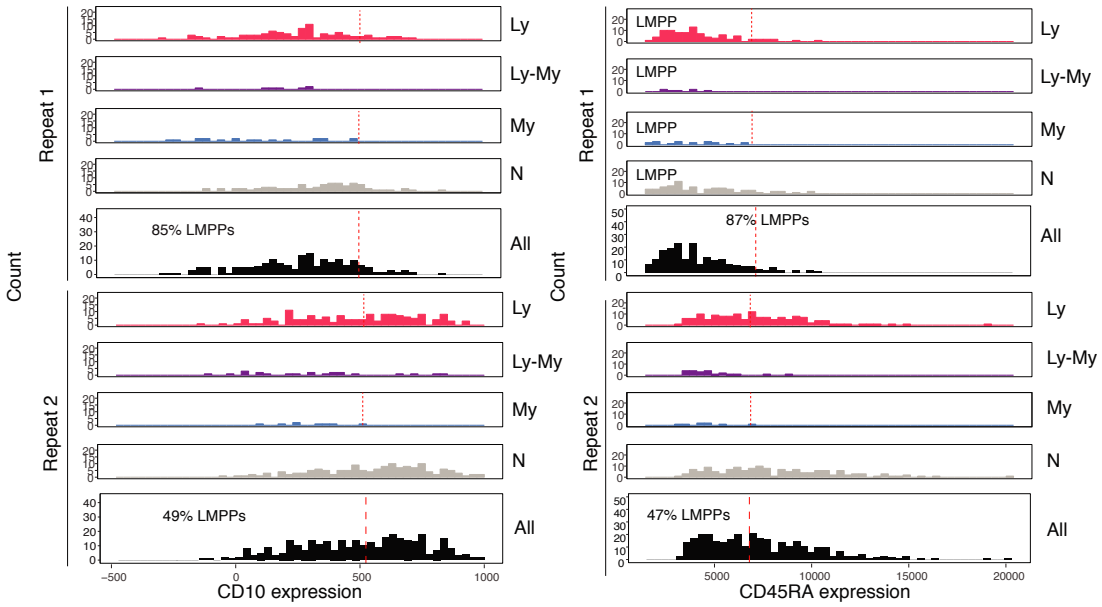


Figure 3.40: Expression of CD10 and LMPP Functional Output. Histograms showing expression of CD10 and CD45RA for LMPP with different functional output. The percentage cut-offs used to define the new sorting strategy for LMPP^{ly} and LMPP^{mix} are shown. Thresholds were defined based on maximum CD10 and CD45RA expression of LMPPs with myeloid output.

(Figures 3.45, 3.46 and 3.47 and Tables A.13 and A.14).

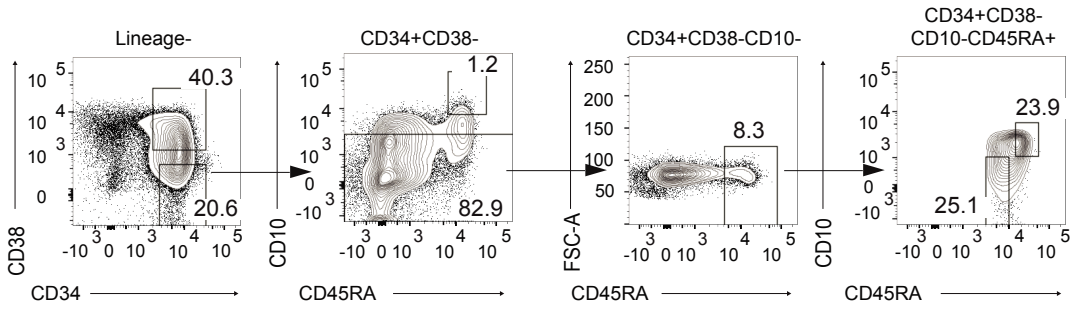
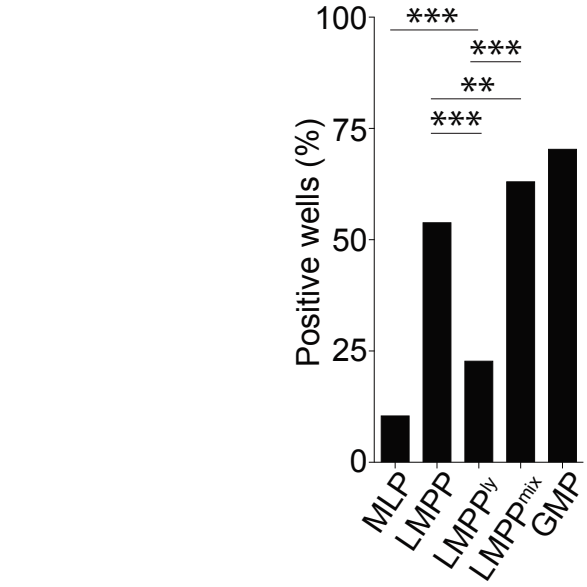
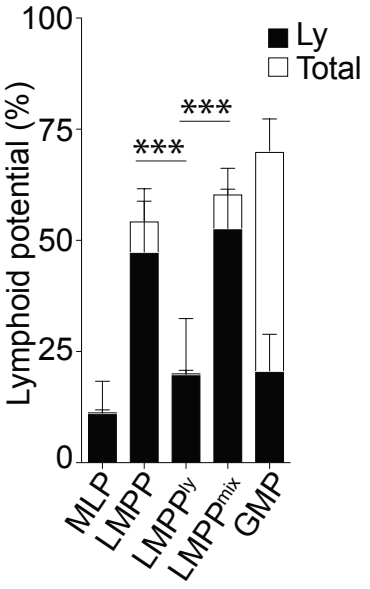


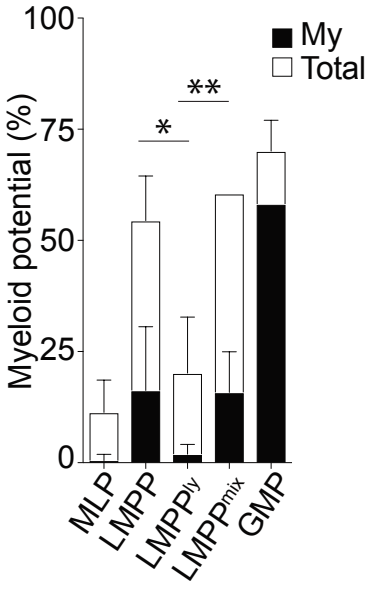
Figure 3.41: Refined Sorting Strategy for the LMPP^{ly} and LMPP^{mix}. Revised sorting strategy for LMPP cells based on expression of CD10 and CD45RA defined in Figures 3.39 and 3.40 to isolate the putative LMPP^{ly} and LMPP^{mix}. Numbers indicate percentage cells from the parental gate.



(a) Total Cloning Efficiency

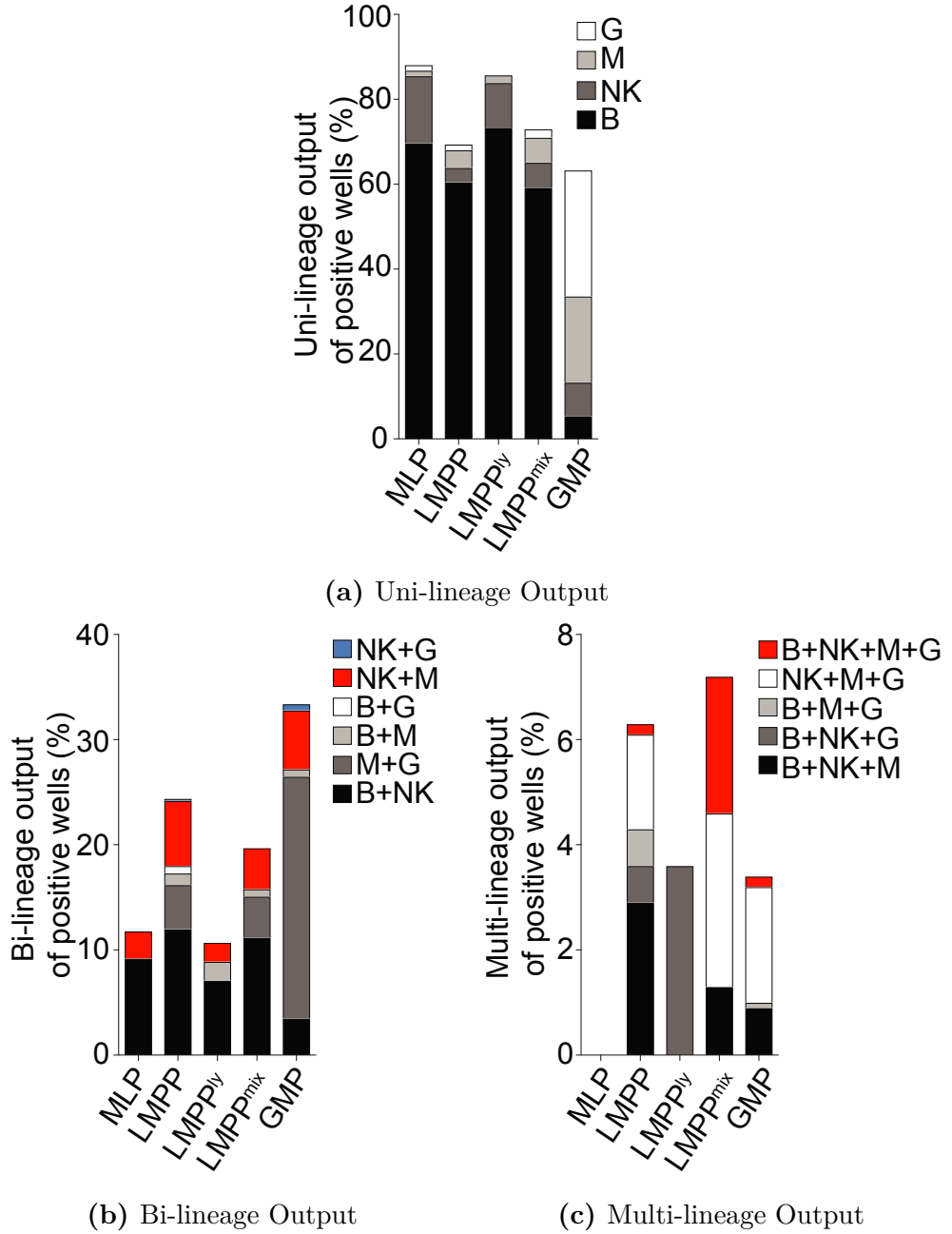


(b) Ly Cloning Efficiency



(c) My Cloning Efficiency

Figure 3.42: Cloning Efficiency of LMPP^{ly} and LMPP^{mix} in Single-Cell SGF15/2 Cultures. Single-cell SGF15/2 cultures were carried out on the LMPP^{ly} and LMPP^{mix}. Single CB MLPs, LMPPs, LMPP^{ly} cells (56 of 244 cells), LMPP^{mix} cells (152 of 240 cells) and GMPs as single cells were cultured and culture output analysed by flow cytometry after 2–2.5 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is (a) the total cloning efficiency (b) cloning efficiency of lymphoid lineages and (c) myeloid lineages. Ly: lymphoid plus mixed output, My: myeloid plus mixed clones.



(a) Uni-lineage Output

(b) Bi-lineage Output

(c) Multi-lineage Output

Figure 3.43: Lineage Output of LMPP^{ly} and LMPP^{mix} in Single-Cell SGF15/2 Cultures. Single-cell SGF15/2 cultures were carried out on the LMPP^{ly} and LMPP^{mix}. Single CB MLPs, LMPPs, LMPP^{ly} cells (56 of 244 cells), LMPP^{mix} cells (152 of 240 cells) and GMPs as single cells were cultured and culture output analysed by flow cytometry after 2–2.5 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the (a) uni-lineage, (b) bi-lineage and (c) multi-lineage output. Presented as the proportion of cells of various lineages in wells that produced cells (positive wells).

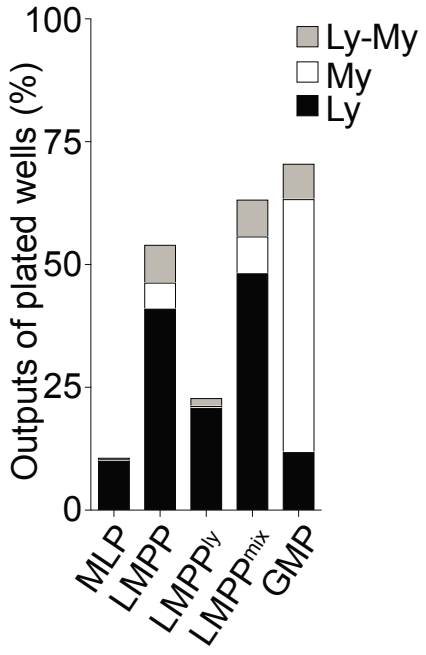
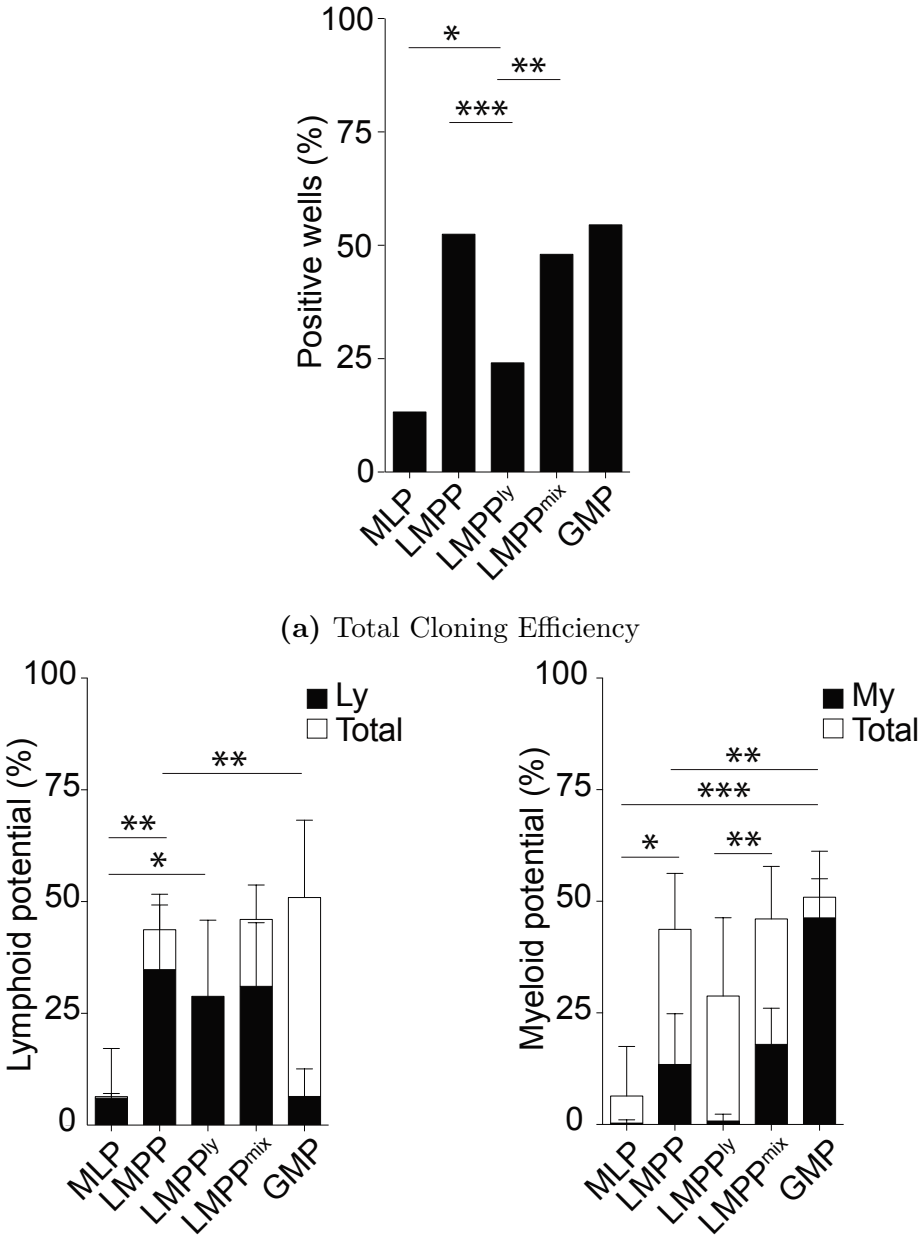


Figure 3.44: Ly-My Output of LMPP^{ly} and LMPP^{mix} in Single-Cell SGF15/2 Cultures. Single-cell SGF15/2 cultures were carried out on the LMPP^{ly} and LMPP^{mix}. Single CB MLPs, LMPPs, LMPP^{ly} cells (56 of 244 cells), LMPP^{mix} cells (152 of 240 cells) and GMPs as single cells were cultured and culture output analysed by flow cytometry after 2–2.5 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the lymphoid, myeloid and lympho-myeloid (Ly-My) output, presented as the frequency among all plated cells.

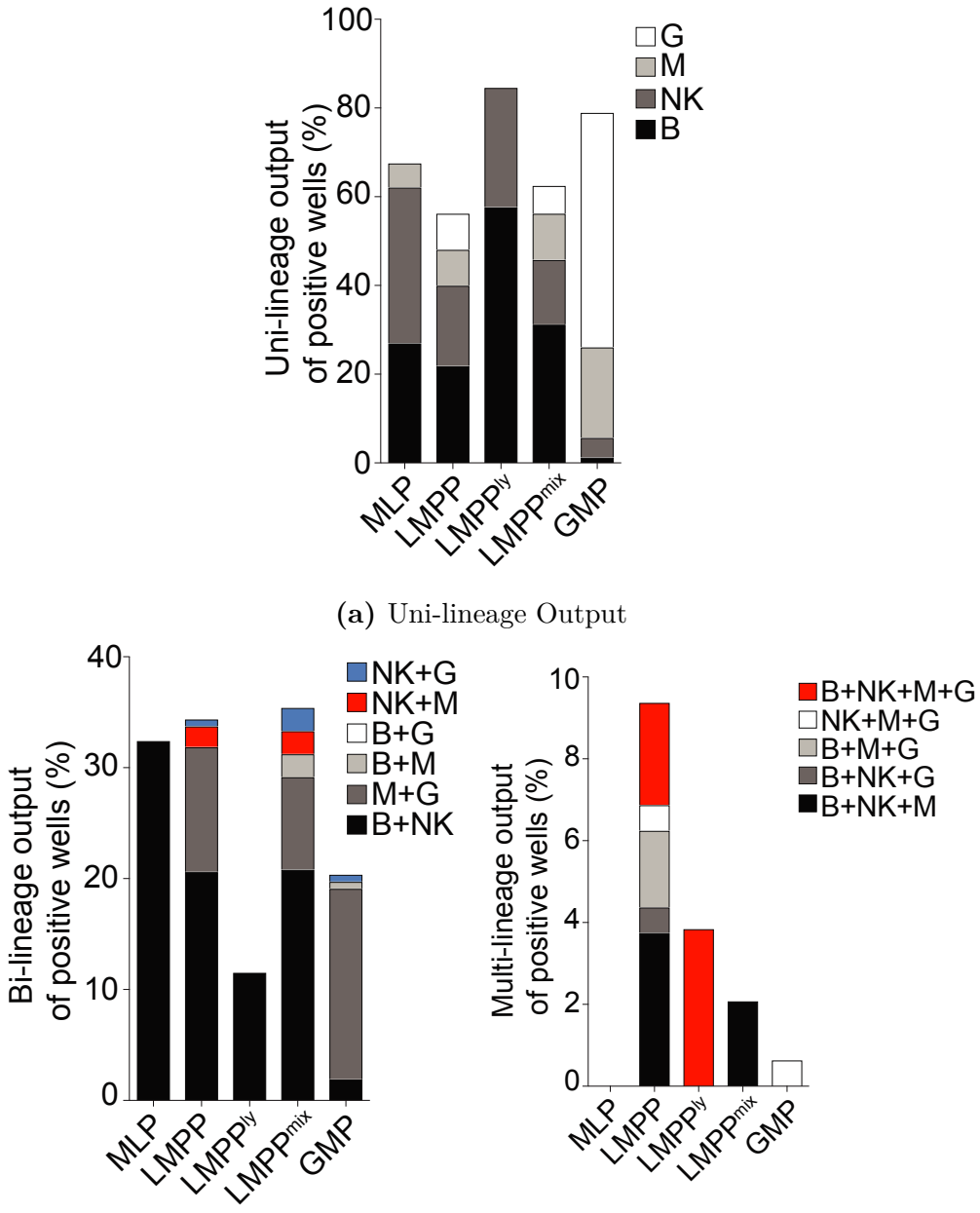


(a) Total Cloning Efficiency

(b) Ly Cloning Efficiency

(c) My Cloning Efficiency

Figure 3.45: Cloning Efficiency of LMPP^{ly} and LMPP^{mix} in Single-Cell SF7b Cultures. Single-cell SF7b cultures were carried out on the LMPP^{ly} and LMPP^{mix}. Single CB MLPs, LMPPs, LMPP^{ly} cells (26 of 108 cells), LMPP^{mix} cells (48 of 100 cells) and GMPs as single cells were cultured and culture output analysed by flow cytometry after 3 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is (a) the total cloning efficiency (b) cloning efficiency of lymphoid lineages and (c) myeloid lineages. Ly: lymphoid plus mixed output, My: myeloid plus mixed clones.



(a) Uni-lineage Output

(b) Bi-lineage Output

(c) Multi-lineage Output

Figure 3.46: Lineage Output of LMPP^{ly} and LMPP^{mix} in Single-Cell SF7b Cultures. Single-cell SF7b cultures were carried out on the LMPP^{ly} and LMPP^{mix}. Single CB MLPs, LMPPs, LMPP^{ly} cells (26 of 108 cells), LMPP^{mix} cells (48 of 100 cells) and GMPs as single cells were cultured and culture output analysed by flow cytometry after 3 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the (a) uni-lineage, (b) bi-lineage and (c) multi-lineage output. Presented as the proportion of cells of various lineages in wells that produced cells (positive wells).

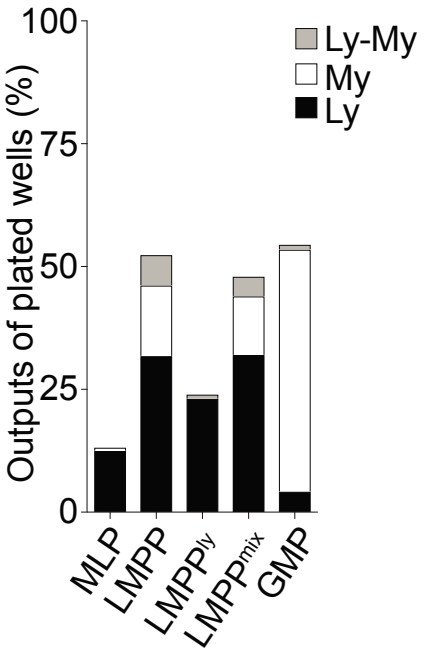


Figure 3.47: Ly-My Output of LMPP^{ly} and LMPP^{mix} in Single-Cell SF7b Cultures. Single-cell SF7b cultures were carried out on the LMPP^{ly} and LMPP^{mix}. Single CB MLPs, LMPPs, LMPP^{ly} cells (26 of 108 cells), LMPP^{mix} cells (48 of 100 cells) and GMPs as single cells were cultured and culture output analysed by flow cytometry after 3 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the lymphoid, myeloid and lympho-myeloid (Ly-My) output, presented as the frequency among all plated cells.

3.7.2 Refining the GMP: GMP CD38^{hi}

On the basis of flow cytometry indexing data, GMPs with myeloid-only output had significantly higher expression of the cell-surface marker CD38 than that of GMPs with LM or lymphoid output (Figure 3.48). In addition to this, GMPs in the more myeloid cluster 3 had significantly higher CD38 expression than that of GMPs in the more lymphoid cluster 1 or the LM cluster 2 (Figure 3.48c).

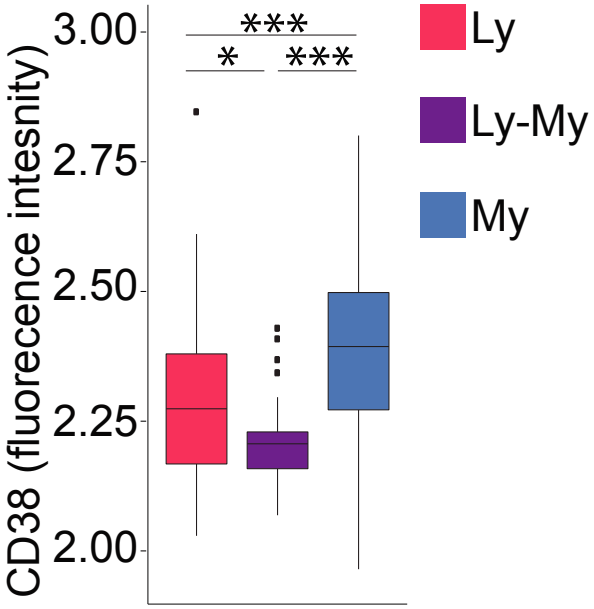
To further confirm the potentially myeloid nature of GMPs with a higher CD38 level, the correlation between CD38 and lineage-associated genes was assessed. It was shown that there was a significant positive correlation between expression of CD38 and expression of a myeloid-associated gene (*MPO*), and a negative correlation between expression of CD38 and expression of lymphoid-associated genes (*MME* and *SELL*), as assessed by single cell qRT-PCR (P = 2.2 x 10-16 and ρ = 0.53 (*MPO*), P = 7.1 x 10^{-5} and ρ = -0.22 (*MME*), and P = 1.3 x 10^{-5} and ρ = -0.24 (*SELL*) (Spearman's rank correlation coefficient); Figure 3.49).

To purify a GMP sub-population without lymphoid potential on the basis of CD38 expression, the entire Lin^{-}CD34^{+} population was divided into CD38^{hi} cells (44% of CD38^{+} cells), CD38^{lo} cells (15% of CD38^{-} cells) and CD38^{mid} cells (an area between the two new gates) (Figures 3.50 and 3.51).

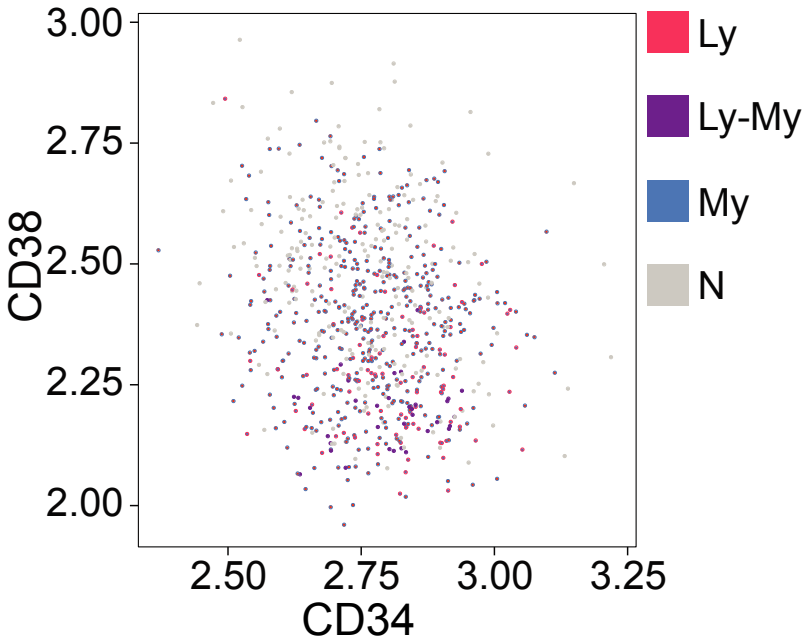
CD38^{hi}, CD38^{mid} and CD38^{lo} cells were then further purified to isolate CD38^{hi} GMPs, CD38^{mid} (CD38^{mid} CD45RA^{+}CD10^{-}) cells and CD38^{lo} LMPPs (Figure 3.51). CD38^{lo} LMPPs were rare (1 in 1 x 10^{8} MNCs), and no conclusions could be reached about their functional potential (data not shown).

The *in vitro* lineage potential of single CD38^{hi} GMPs (279 cells) and CD38^{mid} cells (693 cells) with that of conventionally purified LMPPs (1,136 cells) and GMPs (1,622) was assessed using SGF15/2 culture conditions. Whereas the CD38^{hi} GMPs and LMPPs had a similar cloning efficiency of ~55%, the GMPs and CD38^{mid} cells had a slightly higher cloning efficiency of ~70% (Figure 3.52).

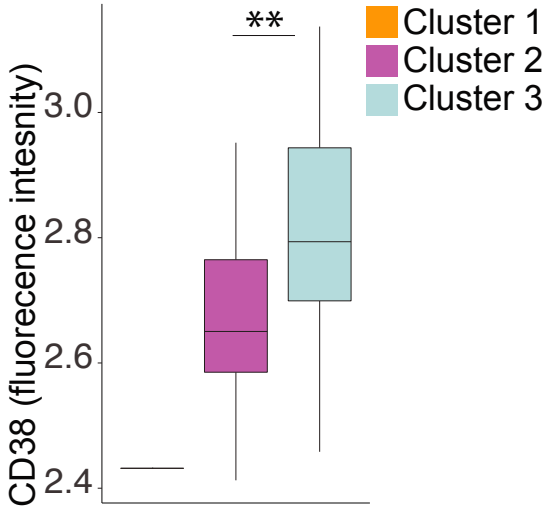
All four populations produced principally uni-lineage output (63- 72%) (Figure 3.53). CD38^{hi} GMPs had much lower lymphoid potential (P < 0.0001 (Fisher's exact test)) and LM potential (P = 0.0115 (Fisher's exact test)) than that of



(a) CD38 Expression: Boxplots



(b) CD38 Expression: Scatter Plot



(c) CD10 Expression: Cluster Identity

Figure 3.48: Expression of CD38 by Single GMP Cells. (a) Expression of the surface marker CD38 on CB GMPs, grouped by functional output. P = 1.57 0^{-11} (myeloid versus lymphoid) and P = 1.6 0^{-8} (myeloid versus lympho-myeloid) (Wilcoxon rank-sum test). (b) Expression of CD38 and CD34 by CB GMPs; colours indicate output from functional assays (c) CD38 fluorescence intensity in GMPs in the 3 clusters identified from clustering in Figure 3.37. Wilcoxon rank sum test is used to define significance. Expression values calculated with the "flowCore::logicleTransform" function in R software.

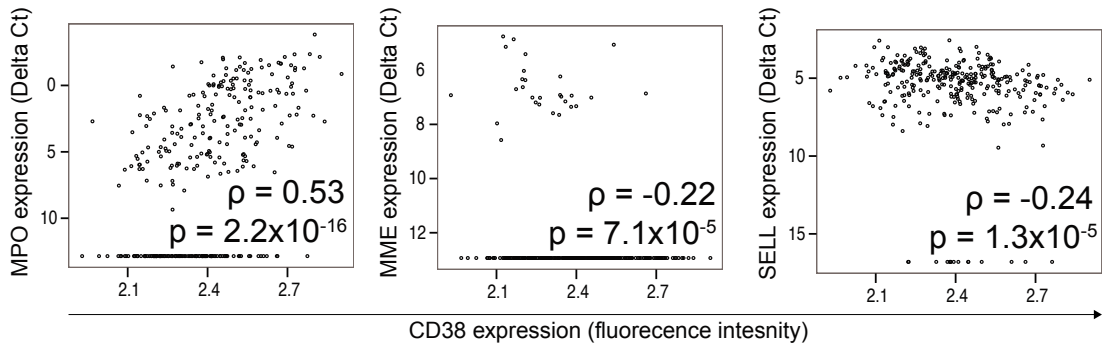


Figure 3.49: CD38 Correlation with Lineage-Associated Genes in Single GMPs. Correlation between expression of selected lineage-associated genes and CD38 cell surface marker expression in GMPs. ρ indicates Spearman's rank correlation coefficient values and p is corresponding p-value from cor.test function in R.

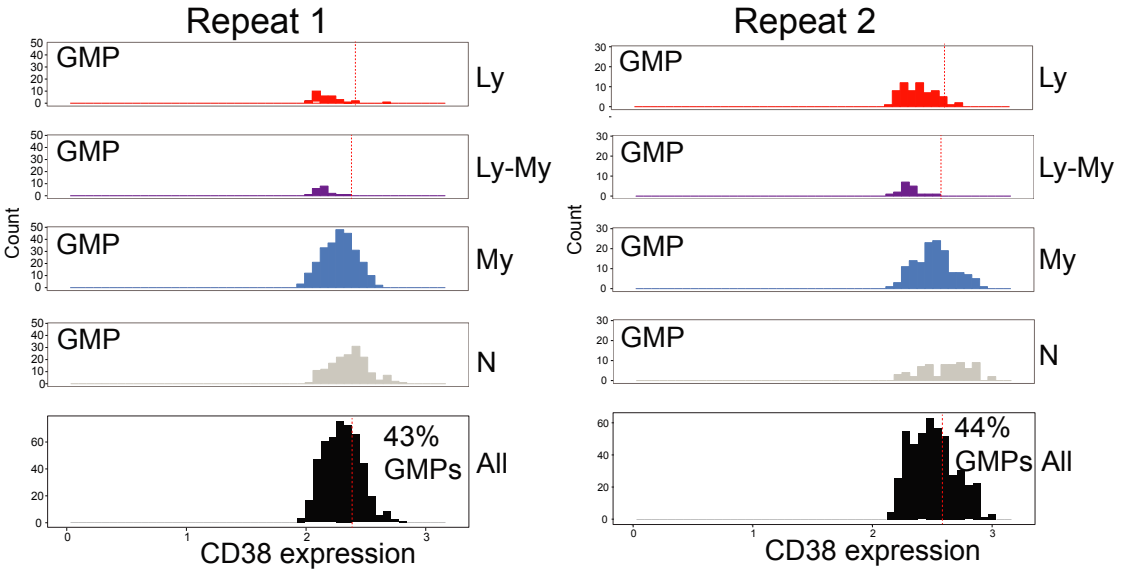


Figure 3.50: Expression of CD38 and GMP Functional Output. Histograms showing expression of CD38 for GMP with different functional output. The percentage cut-off used to define the new sorting strategy for GMP CD38^{hi} and CD38^{mid} are shown.

conventionally purified GMPs (Figures 3.53 and 3.54 and Tables A.15 and A.16), indicative of a functionally purer population.

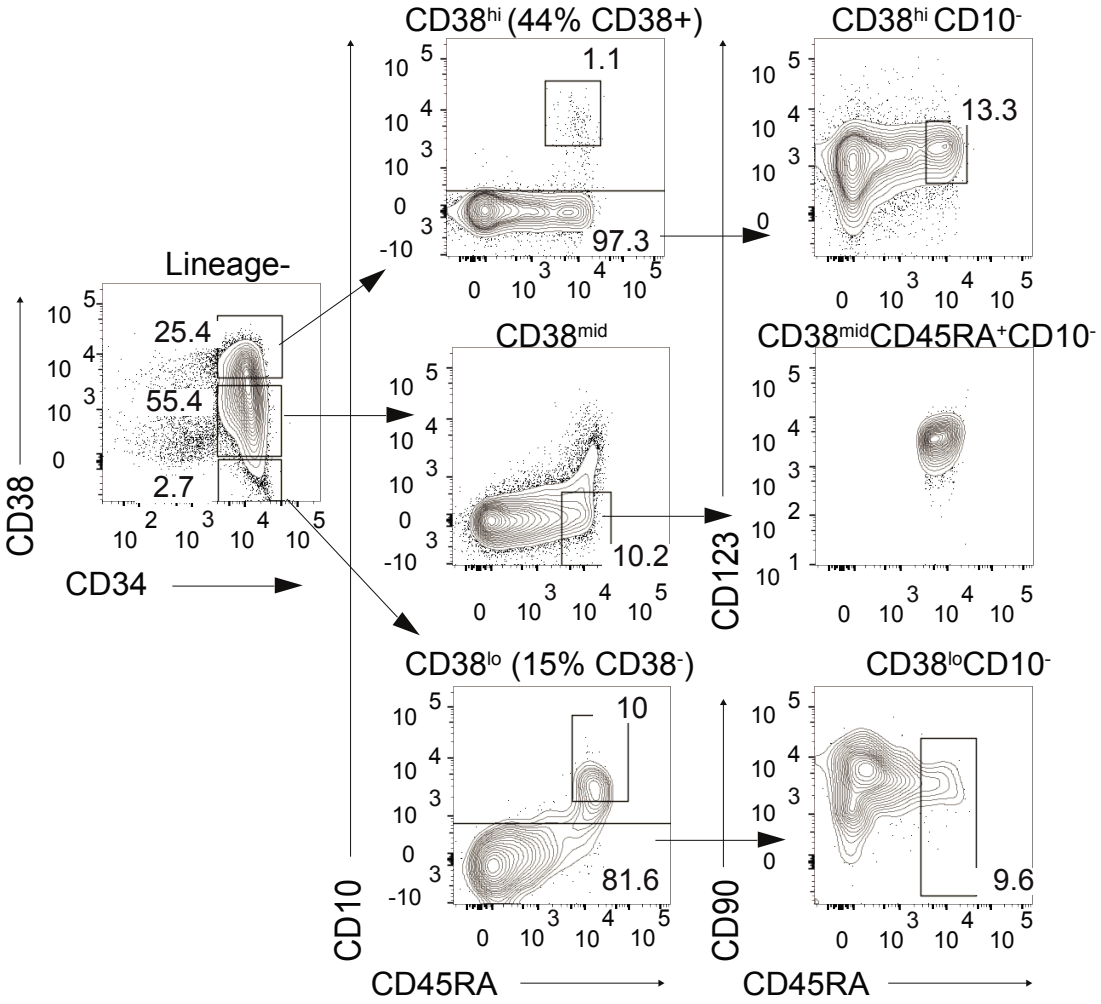
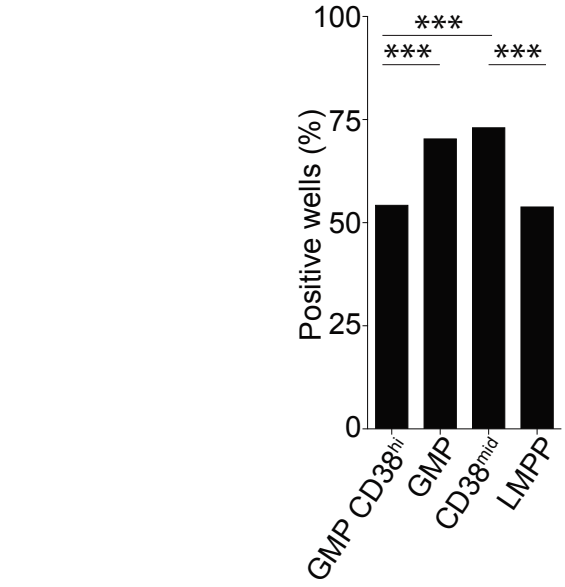
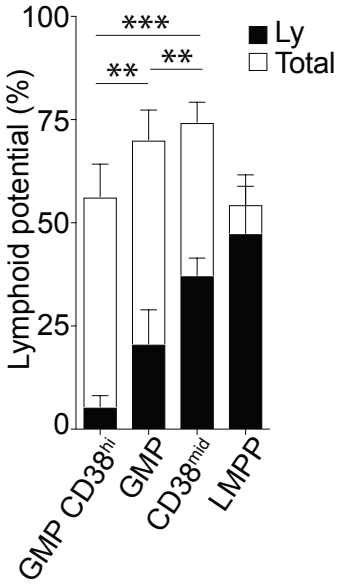


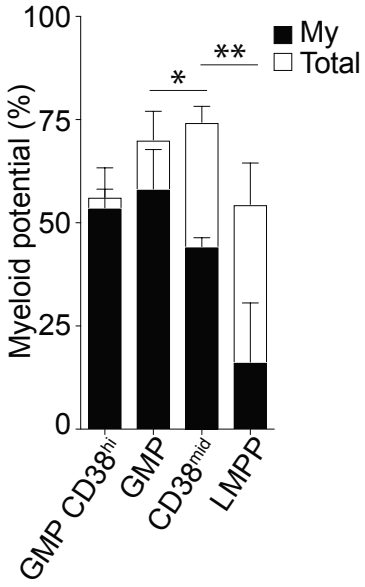
Figure 3.51: Refined Sorting Strategy for the GMP CD38^{hi}. Revised sorting strategy for GMP cells based on expression of CD38 defined in Figures 3.49 and 3.50 to isolate the putative GMP CD38^{hi}. Numbers indicate percentage cells from the parental gate.



(a) Total Cloning Efficiency

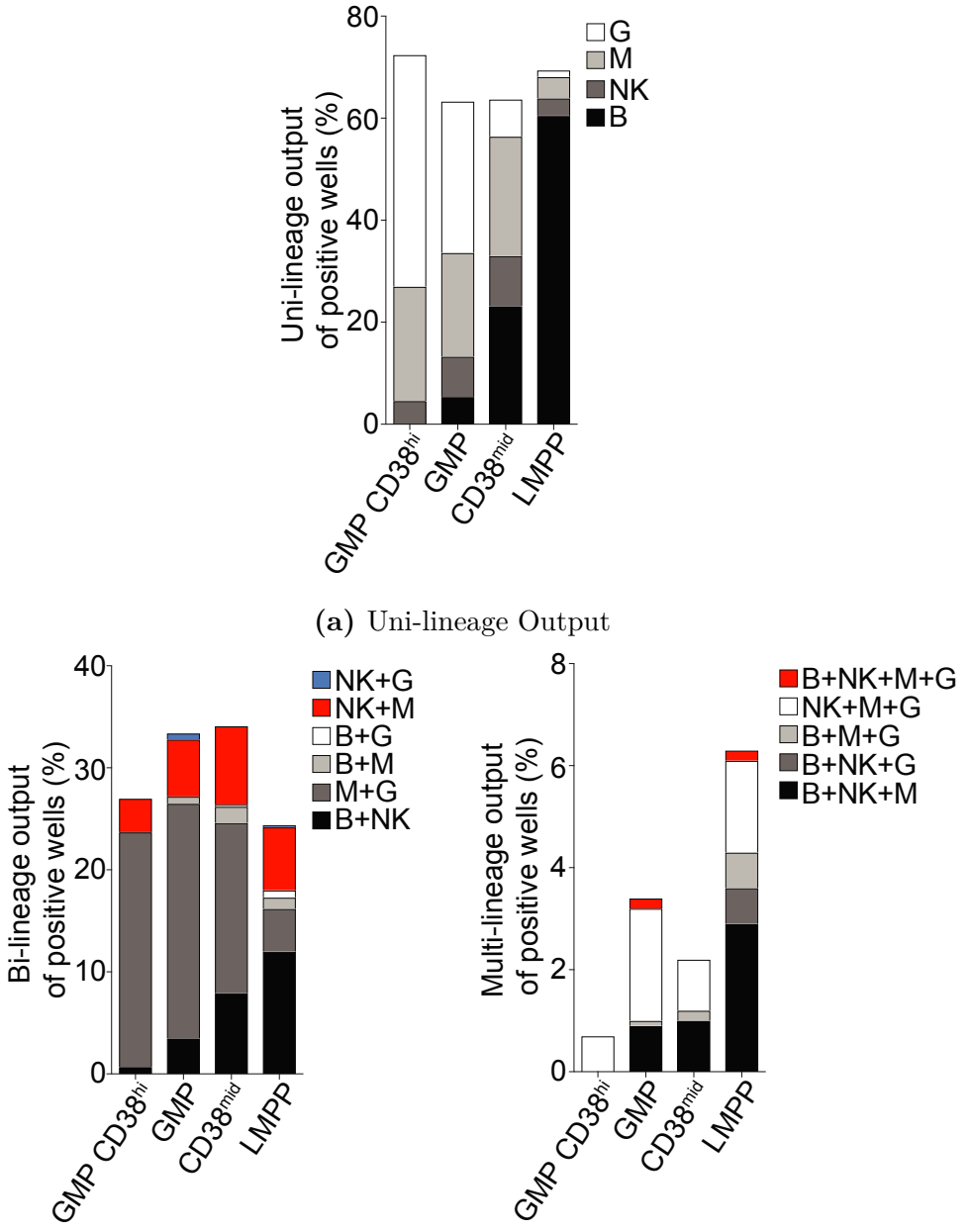


(b) Ly Cloning Efficiency



(c) My Cloning Efficiency

Figure 3.52: Cloning Efficiency of GMP CD38^{hi} in Single-Cell SGF15/2 Cultures. Single-cell SGF15/2 cultures were carried out on the CB CD38^{hi} GMPs (152 of 279 cells), CD38^{mid} cells (508 of 693 cells) and LMPPs and culture output analysed by flow cytometry after 2–2.5 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is (a) the total cloning efficiency (b) cloning efficiency of lymphoid lineages and (c) myeloid lineages. Ly: lymphoid plus mixed output, My: myeloid plus mixed clones.



(a) Uni-lineage Output

(b) Bi-lineage Output

(c) Multi-lineage Output

Figure 3.53: Lineage Output of GMP CD38^{hi} in Single-Cell SGF15/2 Cultures. Single-cell SGF15/2 cultures were carried out on the CB CD38^{hi} GMPs (152 of 279 cells), CD38^{mid} cells (508 of 693 cells) and LMPPs and culture output analysed by flow cytometry after 2–2.5 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the (a) uni-lineage, (b) bi-lineage and (c) multi-lineage output. Presented as the proportion of cells of various lineages in wells that produced cells (positive wells).

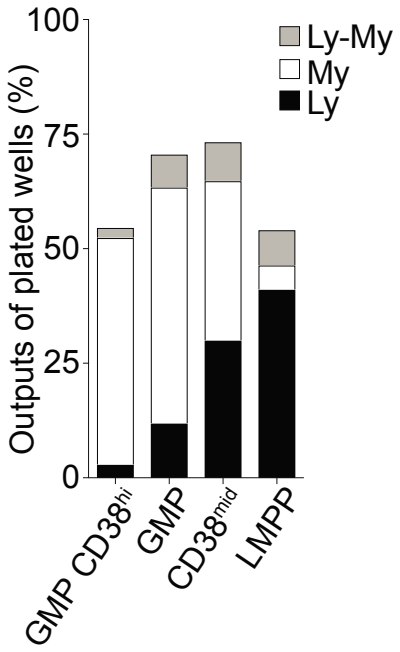


Figure 3.54: Ly-My Output of GMP CD38^{hi} in Single-Cell SGF15/2 Cultures. Single-cell SGF15/2 cultures were carried out on the CB CD38^{hi} GMPs (152 of 279 cells), CD38^{mid} cells (508 of 693 cells) and LMPPs and culture output analysed by flow cytometry after 2–2.5 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the lymphoid, myeloid and lympho-myeloid (Ly-My) output, presented as the frequency among all plated cells.

3.8 Revised Model of Human LM Differentiation

Collectively, the single-cell analyses suggested that the progenitor hierarchy downstream of the HSC may be more complex than previous models have suggested (Figure 3.55). Multiple, rare, functionally distinct bi- and multi-lineage LM progenitors (LMP) could differentiate either directly from the HSC or MPP. The term LMP, rather than LMPP has been used to describe these LM progenitors, as not all LMP will be lymphoid biased.

Multi-lineage LMPs were rare in the single-cell analyses. Lymphoid-only and myeloid-only progenitors are shown in Figure 3.55. Bi-lineage progenitors were more frequent and uni-potent progenitors were most common. The hierarchical relationships between LMP and lymphoid and myeloid only bi- and uni-lineage progenitors remain to be determined. This model also leaves open the question of when commitment to either lymphoid or myeloid fate occurs. Erythro-megakaryocytic differentiation leading to an MEP, erythroid (BFU-E) and megakaryocytic progenitor (MkP) could occur directly from either the HSC or MPP and has not been studied here.

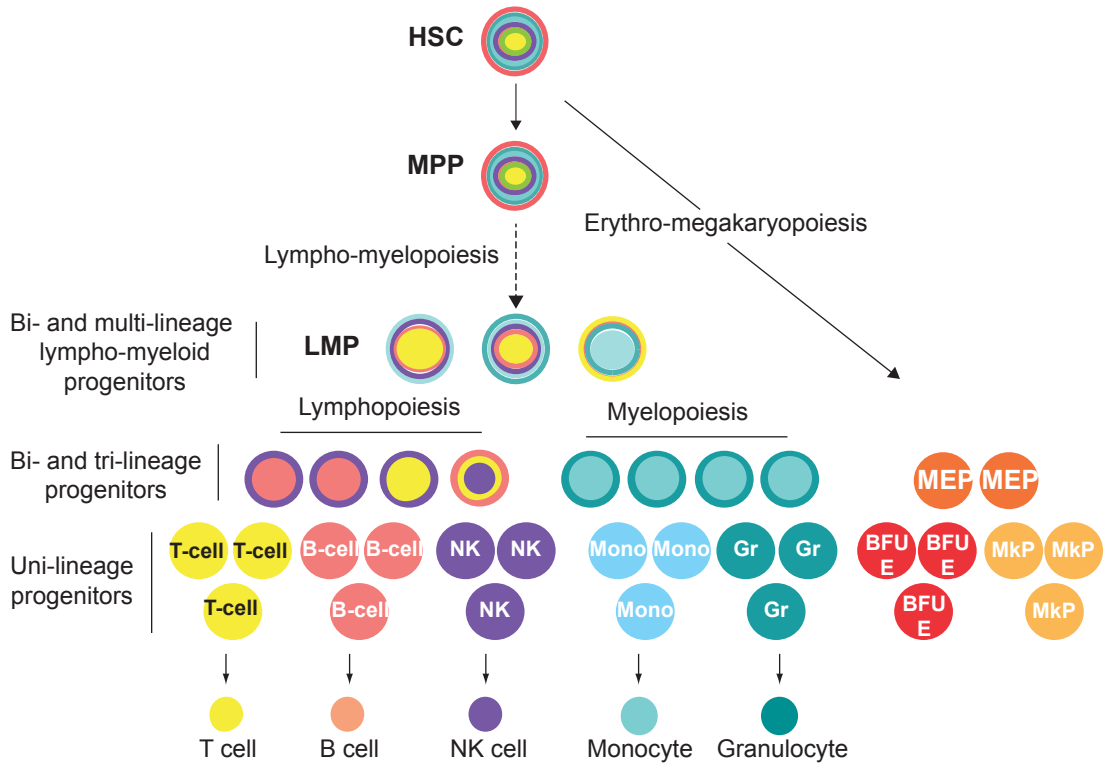


Figure 3.55: Revised Model of Human LM differentiation. Proposed novel model of human LM differentiation based on single-cell functional and transcriptional analysis of the LM progenitors.

3.9 Conclusion on Human LM Progenitors

It has been demonstrated that the three LM progenitors form distinct and separable populations, as defined immunophenotypically. The bulk functional output from these progenitors was characterised, and it was confirmed that all three populations had the ability to generate both lymphoid and myeloid cells. The bulk transcriptional profiles of the progenitors were investigated, and it was seen that these were distinct and reflected their functional output. At the single-cell level, the functional output from these progenitors revealed the intra-population heterogeneity of the LM progenitors. The MLP was seen to be the most functionally pure population, being predominantly a B cell progenitor. The GMP had primarily a myeloid output, but retained lymphoid potential. The LMPP represented the

most balanced LM progenitor. Although lymphoid biased, it still generated myeloid cells. Across all progenitors, the majority of single-cells were uni-lineage, producing only one mature cell type. Bi- and multi-lineage progenitors were also detected but with increasing rarity. Single-cells with the ability to generate both lymphoid and myeloid cells were detected, but were extremely rare. These were most commonly found within the LMPP. Finally, the GMP and LMPP populations were re-defined to more functionally pure populations. From the GMP, the GMP^{hi} almost exclusively produced myeloid cells. From the LMPP, the $LMPP^{ly}$ retained almost all of the lymphoid potential and the $LMPP^{mix}$ retained the myeloid and LM ability. This characterisation of the normal LM progenitors may aid the study of the transformed LM progenitors in AML.

Chapter 4

Leukaemia-Propagating Potential of Progenitor-Like Leukaemia Stem Cells

Contents

4.1 Introduction to the Use of Xenotransplantation Models to Study AML 119
4.2 Immunophenotypic Characterisation of CD34$^+$ AML . 120
4.3 Xenotransplantation of AML HSPC Populations . . . 126
4.4 Genotyping of Human Engraftment 133
4.5 Kinetics of Human Engraftment 136
4.6 HSPC Immunophenotype of Human Engraftment . . . 143
4.7 Conclusion on the Xenotransplantation of AML 147

4.1 Introduction to the Use of Xenotransplantation Models to Study AML

It has long been postulated that cancer is a disease maintained by a population of cells who possess stem-cell-like abilities, termed cancer stem cells (reviewed here [32]). It had been proposed that leukaemia, like normal haematopoiesis, is organised in a hierarchy with LSCs maintaining the disease [32, 97–99]. The existence of leukaemia-initiating cells was first shown functionally in AML through the use of xenotransplantation assays in mice [34]. The CD34$^+$CD38$^-$ population of leukaemic cells was able to engraft NOD/SCID mice and was termed SCID leukaemia-initiating cells (SL-ICs).

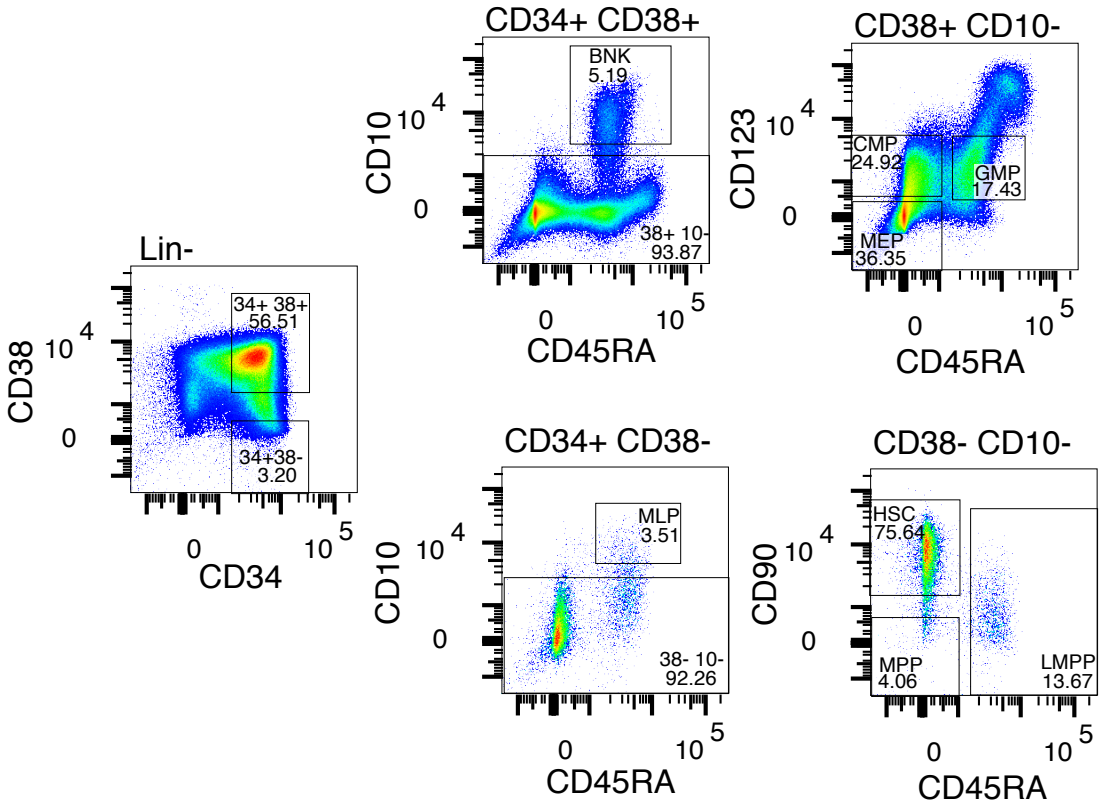
The majority of AML cases are characterised by the expression of the early HSPC cell surface marker CD34, with typically >80% of cells being CD34$^+$ [22]. In ~20–40% of cases >80% of cells lack CD34 expression and this is classified as CD34$^-$ AML [22, 37, 100]. CD34$^-$ AML is characterised by a mutational profile and hierarchical structure that is distinct to CD34$^+$ AML [37] and is not studied here.

Initially, SL-ICs were seen exclusively in the CD34$^+$CD38$^-$ compartment [2, 34, 35]. However, more recently it has been demonstrated that the AML LSC can be found also in the CD34$^+$CD38$^+$ compartment [22, 36]. It has been demonstrated that in ~85% of CD34$^+$ AML cases, the CD34$^+$CD38$^-$CD45RA$^+$ LMPP and CD34$^+$CD38$^+$CD45RA$^+$ GMP populations were expanded and contained LSC activity. In the remaining ~15% of cases, the CD34$^+$CD38$^-$CD45RA$^-$ MPP and CD34$^+$CD38$^+$CD45RA$^+$ CMP populations were expanded and had LSC ability [22]. This is of particular interest because it demonstrates that the LSC populations in AML have a progenitor-like, rather than stem-like, immunophenotype.

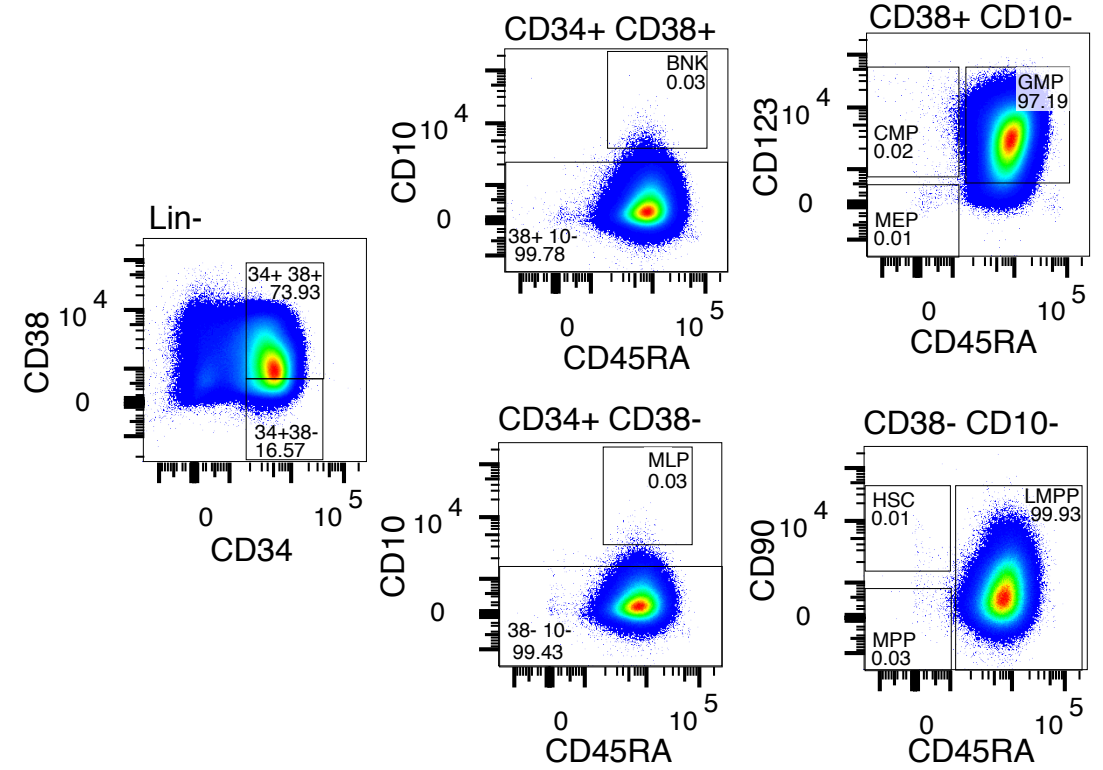
I aimed to further characterise the sub-set of CD34$^+$ AML with an expanded LMPP and GMP immunophenotype. This was to be done in the context of the CD34$^+$ CD38$^-$ CD45RA$^+$ CD10$^+$ MLP population [24], which had not previously been studied in AML. Firstly, I carried out immunophenotype screening of AML samples to identify samples with an expanded LMPP and GMP immunophenotype. These samples spanned a range of genetic and cytogenetic groups. I then aimed to investigate the leukaemia-propagating ability of the HSPC populations within these samples, characterising the immunophenotype and kinetics of engraftment. Finally, the gene expression profiles of these samples were also characterised by RNA-Seq (see Chapter 5).

4.2 Immunophenotypic Characterisation of CD34$^+$ AML

The HSPC populations in BM were defined as in CB (Table 3.1). When looking at normal BM, all eight of the normal HSPC populations (as defined in Table 3.1) were seen. However, in a typical CD34$^+$ AML sample, a differentiation arrest resulted in an expanded LMPP and GMP immunophenotype. In normal BM the frequency of the LMPP is very rare, being typically 0.2% of CD34$^+$ cells (Table 3.1); however, in AML this population was expanded by several hundred-fold (Tables A.18 and 4.1). This LMPP and GMP expansion was usually accompanied by a severe reduction in the remaining HSPC populations, with little to no CD45RA$^-$ compartment remaining (Figure 4.1b).



(a) Normal Sample OX700



(b) AML Sample STB291

Figure 4.1: Immunophenotype of normal and leukaemic BM. HSPC immunophenotype of (a) normal BM samples showing expected HSPC populations and (b) of AML sample STB291 showing expanded LM populations. Frequencies indicate the frequency of the parent gate.

47 AML samples, previously identified as being CD34$^+$, were screened for their HSPC immunophenotype (Figure 4.2 and Table A.18). Firstly, it could be seen that 37/47 AML samples had a CD38$^+$CD45RA$^+$ GMP-like immunophenotype. In 13/37 cases, this was accompanied by an expansion in the CD38$^-$CD45RA$^+$ LMPP-like population, where in some cases the LMPP-like population composed up to ~65% of CD34$^+$ cells. In 3/47 cases, the LMPP-like population was expanded without a large GMP-like population, and in these cases the LMPP composed up to ~90% of CD34$^+$ cells (Figure 4.2 and Table A.18).

Consistent with previous findings [22], in some cases an expansion in the CD38$^{+/-}$CD45RA$^-$ population(s) was seen (Figure 4.2 and Table A.18). This was not typically seen with a large LMPP-like population; however, it was occasionally seen with mild expansion of the GMP-like compartment (Figure 4.2 and Table A.18).

The possible expansion of the CD10$^+$ MLP [24] population in AML had not been previously characterised. This possible expansion was investigated here, and it was seen that the MLP was mostly absent in AML. The MLP was only rarely expanded, notably in sample STB257. This sample showed an 87.59-fold expansion of the CD10$^+$CD45RA$^+$ MLP and a 128.79-fold expansion of the CD10$^-$CD45RA$^+$ LMPP (Figures 4.2, 4.3 and Table A.18).

From the 47 AML samples screened for immunophenotype, 19 samples were taken forward for further analysis in xenotransplantation assays (Figure 4.4 and Table 4.1). These samples were chosen primarily for the presence of an expanded LMPP and GMP population and additionally based on sample viability and availability. Full immunophenotypes of these samples can be seen in Figure A.2.

There was no selection for samples on the basis of mutational or cytogenetic profile; however, these samples were genotyped by targeted re-sequencing using the Fluidigm Access Array and available clinical cytogenetic information was

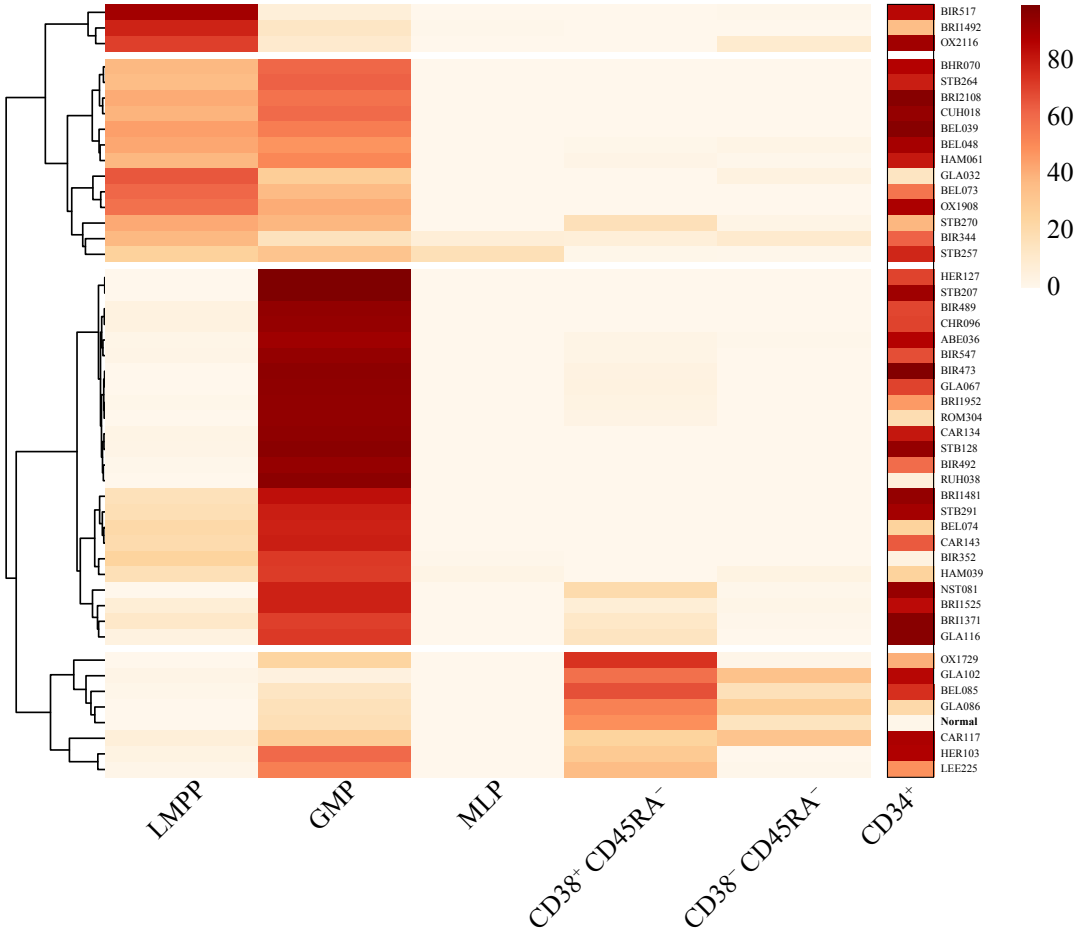


Figure 4.2: Summary of AML Immunophenotype Screening. Relative size of HSPC populations in 47 $CD34^+$ AML samples screened for immunophenotype. $CD34^+$ compartment shown as a percentage of the Lin^- compartment. All other HSPC populations shown as a percentage of the $CD34^+$ compartment. Normal size of HSPC populations in normal human BM from Table 3.1 and [54].

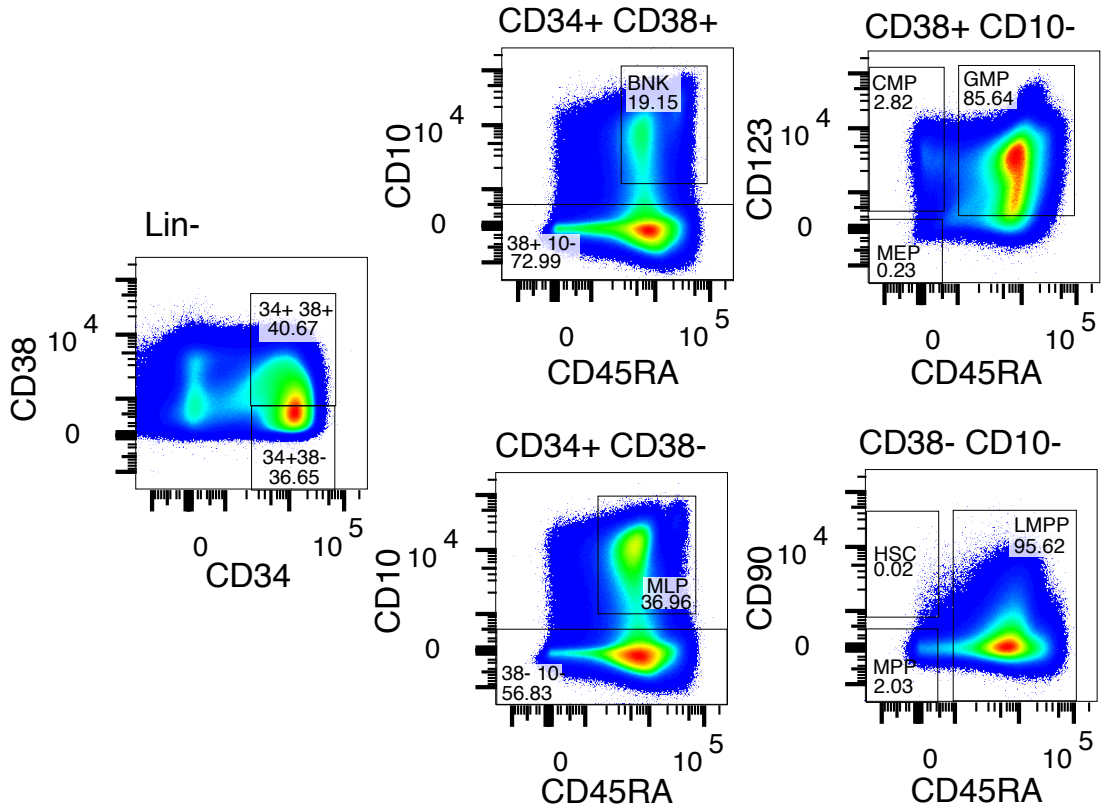


Figure 4.3: Rare Expanded MLP AML Immunophenotype. HSPC immunophenotype of AML sample STB257 showing the rare immunophenotype of an expanded MLP population. Frequencies indicate the frequency of the parent gate.

collected (Tables A.19, A.20). Genotyping was carried out for genes known to be commonly mutated in AML [101] and these included: *ASXL1*, *BCOR*, *CBL*, *CDKN2A*, *ETV6*, *FLT3*, *IDH1*, *JAK2*, *KIT*, *KRAS*, *MAU2*, *NIPBL*, *NRAS*, *PDS5B*, *PHF6*, *RAD21*, *RUNX1*, *SMC3*, *SRSF2*, *STAG1*, *STAG2*, *TET2*, *TP53*, *U2AF1* and *WT1* (Table A.19).

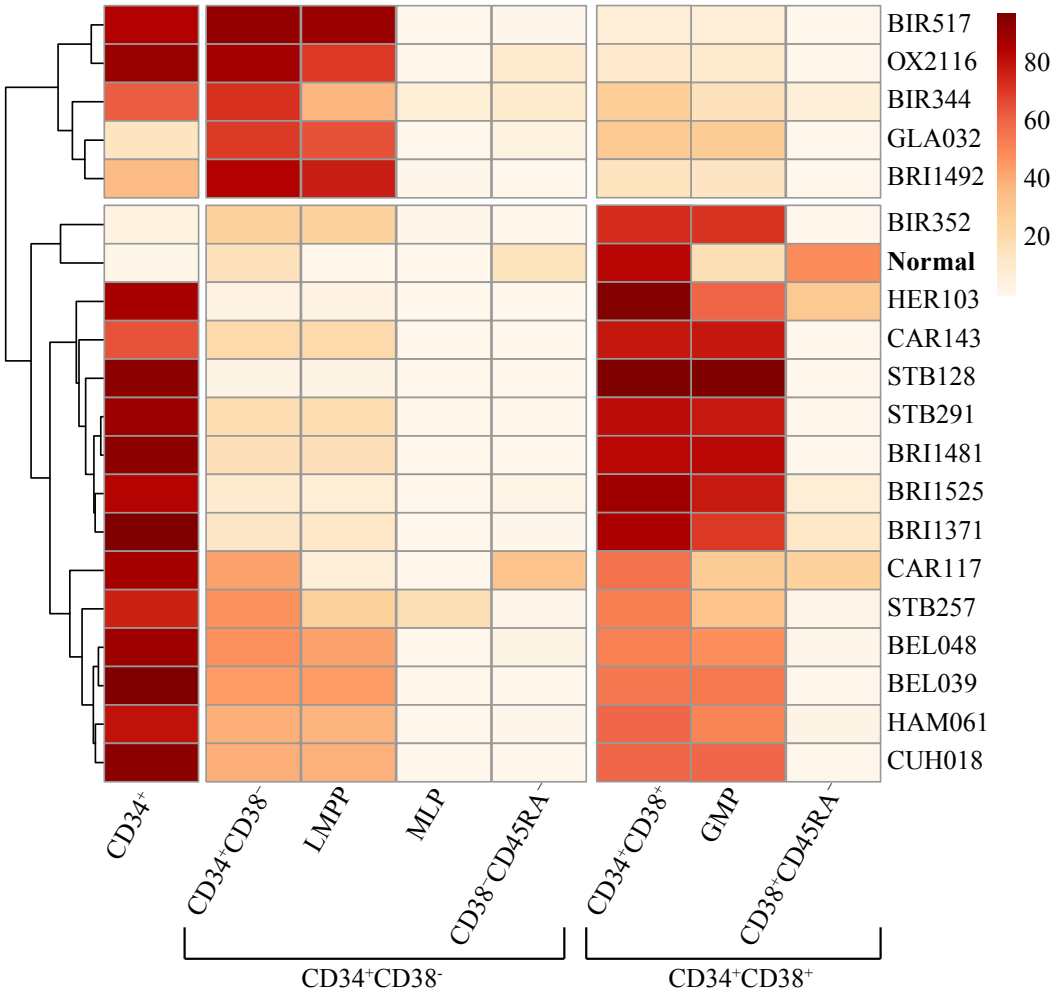


Figure 4.4: Immunophenotype of AML Samples for Xenotransplantation. Relative size of HSPC populations. $CD34^+$ compartment shown as a percentage of the Lin^- compartment. All other HSPC populations shown as a percentage of the $CD34^+$ compartment. Normal size of HSPC populations in normal human BM from Table 3.1 and [54].

Table 4.1: Expansion of Lympho-Myeloid Progenitors in CD34$^+$ AML. Relative size of HSPC populations as a percentage of either the Lin$^-$ or CD34$^+$ compartments. Fold change is relative to the average size of the HSPC population in normal human BM from Table 3.1 and [54].

Sample	CD34$^+$	CD34$^+$ CD38$^-$	CD34$^+$ CD38$^+$	LMPP		GMP		MLP	
	% Lin$^-$	% CD34$^+$	% CD34$^+$	% CD34$^+$	Fold Change	% CD34$^+$	Fold Change	% CD34$^+$	Fold Change
STB128	94.11	2.30	97.70	2.17	10.84	97.03	5.45	0.00	0.00
HAM061	80.79	39.81	60.19	38.03	190.14	51.41	2.89	0.19	0.94
BIR517	85.23	92.99	7.01	90.97	454.84	6.62	0.37	0.13	0.66
GLA032	14.13	70.72	29.28	65.44	327.19	27.71	1.56	0.05	0.26
OX2116	91.41	88.95	11.05	71.23	356.17	10.57	0.59	0.29	1.43
STB291	90.50	18.31	81.69	18.20	90.99	79.22	4.45	0.00	0.02
BEL048	90.31	47.56	52.44	43.02	215.09	47.99	2.70	0.00	0.00
BEL039	97.58	45.18	54.82	44.89	224.43	54.55	3.06	0.00	0.02
CUH018	94.10	39.50	60.50	39.46	197.28	60.36	3.39	0.00	0.01
STB257	77.32	47.40	52.60	25.76	128.79	32.88	1.85	17.52	87.59
BIR352	3.75	25.83	74.17	24.82	124.08	72.60	4.08	0.63	3.15
BRI1481	94.57	16.61	83.39	16.59	82.94	83.30	4.68	0.00	0.00
BRI1525	84.35	9.82	90.18	7.63	38.17	78.36	4.40	0.00	0.02
BRI1371	97.16	13.20	86.80	12.06	60.32	71.08	3.99	0.00	0.00
HER103	87.85	3.35	96.65	3.18	15.92	60.86	3.42	0.00	0.00
CAR143	65.00	20.40	79.60	20.35	101.77	79.35	4.46	0.01	0.03
BRI1492	35.72	85.38	14.62	77.95	389.73	13.67	0.77	0.61	3.03
CAR117	88.81	43.29	56.71	6.28	31.38	27.73	1.56	0.00	0.01
BIR344	62.72	73.12	26.88	37.74	188.70	15.82	0.89	7.43	37.17
Normal	1.00	16.10	83.90	0.20	1.00	17.80	1.00	0.20	1.00

4.3 Xenotransplantation of AML HSPC Populations

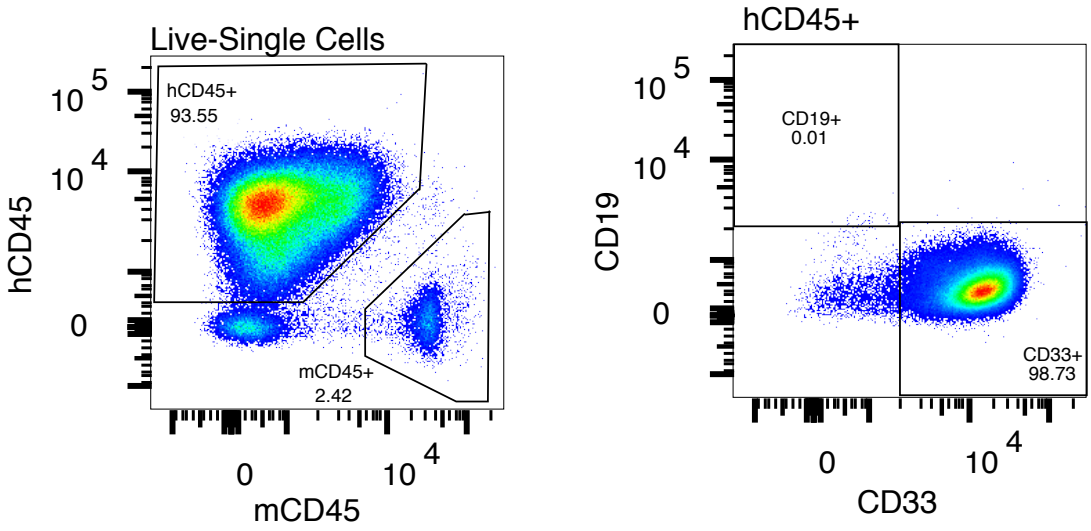
The 19 AML samples identified from the immunophenotype screening were then further analysed in xenotransplantation assays, as described previously [22]. Briefly, expanded HSPC populations were flow sorted and xenotransplanted by intra-osseous injection into sub-lethally irradiated NSG mice. Mice were then monitored for the development of human leukaemia. For full experimental details of xenotransplantation assays, see Chapter 2.9. For a summary of all xenotransplantation experiments carried out, see Chapter A.22. All expanded HSPC populations seen at immunophenotype screening were sorted and if sufficient

numbers obtained, xenotransplanted. Typically, 10^4–10^6 cells were injected per xenotransplantation (Table A.22).

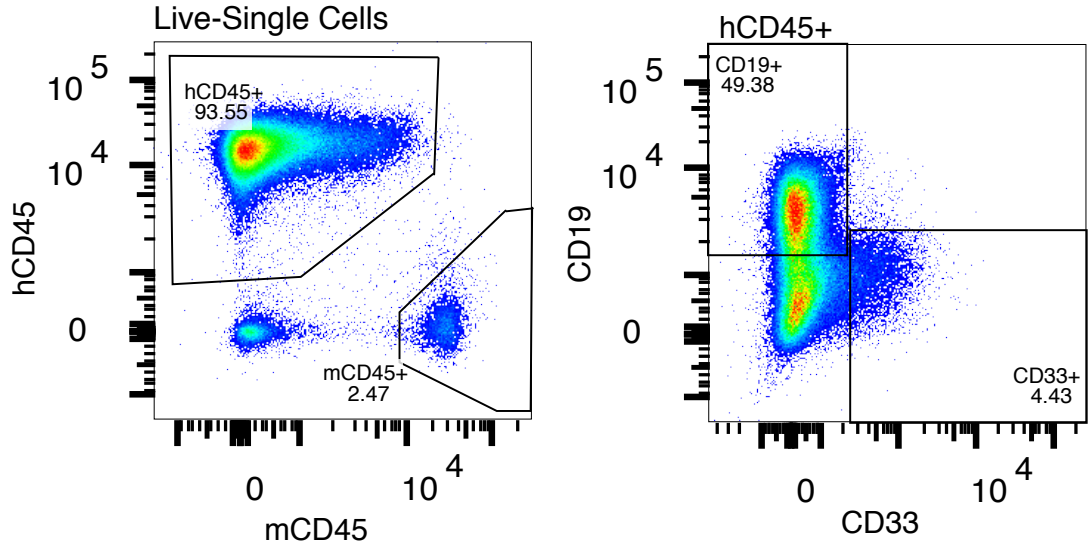
To monitor mice for the development of human leukaemia, PB was sampled ~12 and ~16 weeks post-transplantation, and BM was sampled ~20 and ~24 weeks post-transplantation (from the non-injected tibia and injected tibia respectively). PB/BM MNCs were stained to identify human cells (using the pan-haematopoietic cell marker CD45) of both the myeloid (CD33) and lymphoid (CD19) lineages. It is a known limitation of the xenotransplantation assay that not all human AML samples are able to adapt to the mouse microenvironment, with only up to two thirds of samples engrafting in optimised assays [102–104]. In this study, 8/19 AML samples engrafted NSG mice and the remaining samples did not give rise to any detectable human haematopoietic cells (Figures 4.7, A.3).

4/8 engrafting AML samples gave rise to myeloid ($CD33^+$) engraftment (Figures 4.7, 4.5a). The remaining four engrafting samples gave rise to varying degrees of lymphoid ($CD19^+$) and myeloid engraftment (Figures 4.7, 4.5b, 4.5c, 4.5d, 4.5e). For one of these samples, OX2116, it was noted at diagnosis that lymphoid blasts with B and T cell markers were present. Furthermore, previous studies have demonstrated that the B cell marker CD19 may be aberrantly expressed in myeloid leukaemia [105–110].

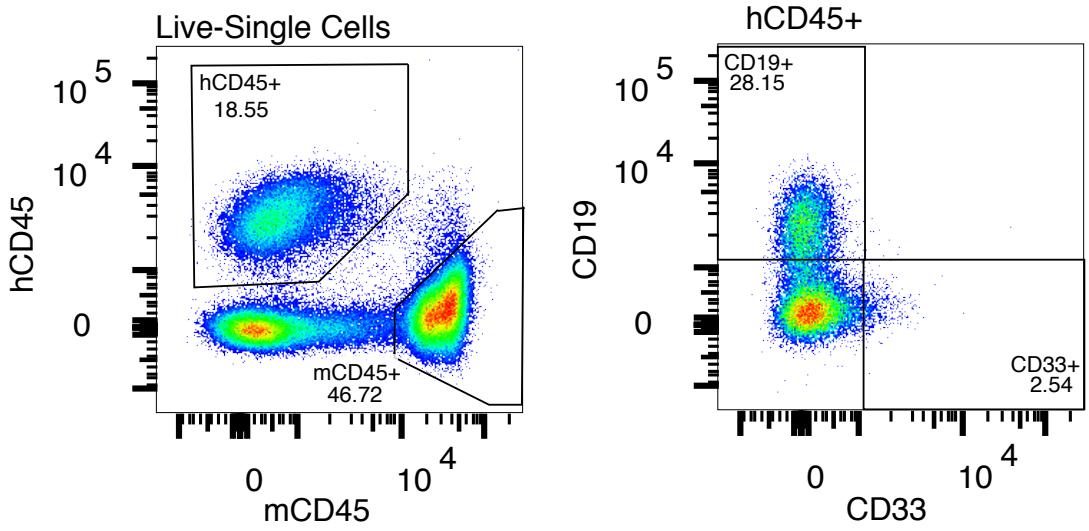
In two of the engrafting samples that showed $CD19^+$ engraftment, the original patient sample was immunophenotyped for the expression of CD19. It was seen in both of these cases that CD19 was aberrantly expressed with CD34, indicative of leukaemic blasts rather than normal B cells (Figure 4.6). Furthermore, in patient STB257 abnormal $CD34^+CD33^+CD19^+$ cells were seen (Figure 4.6). In both cases the immunophenotypic profile of the original patient sample was mimicked in the engrafted material.



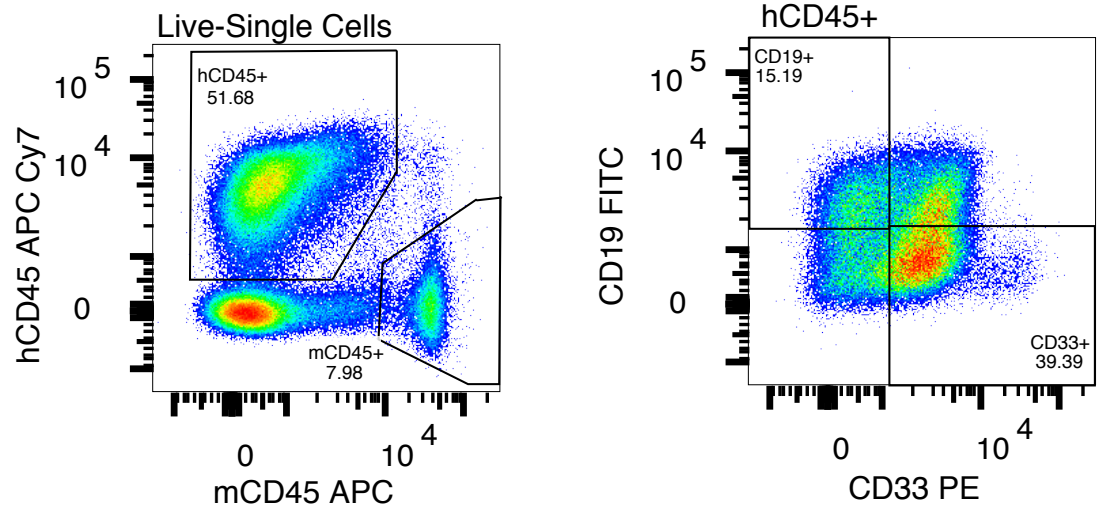
(a) Sample BIR517: Myeloid Engraftment



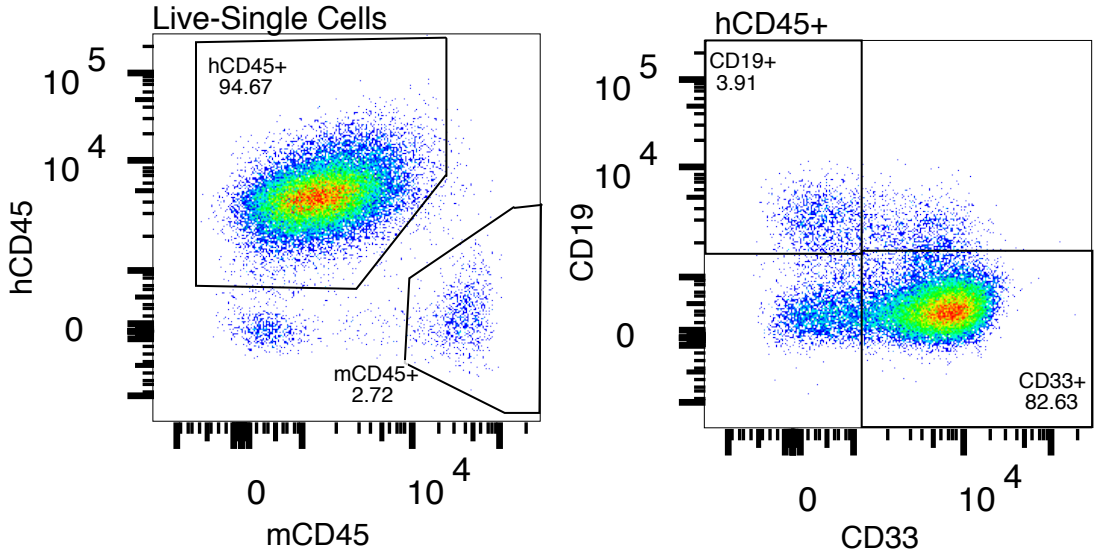
(b) Sample OX2116: LM Engraftment



(c) Sample BEL039: LM Engraftment

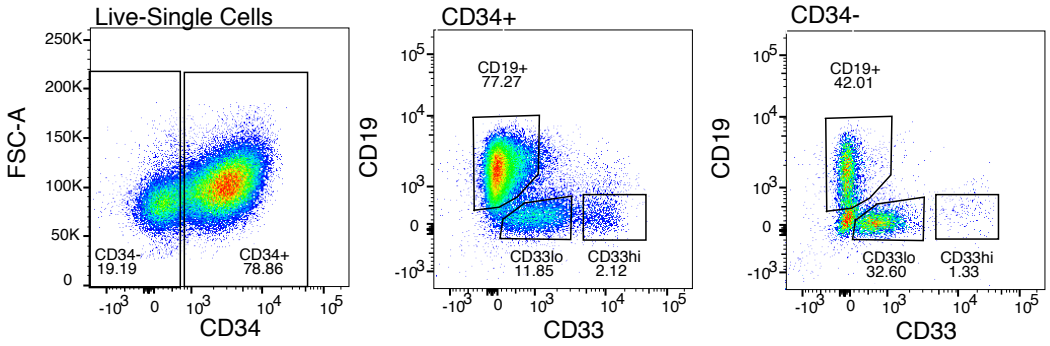


(d) Sample STB257: LM Engraftment

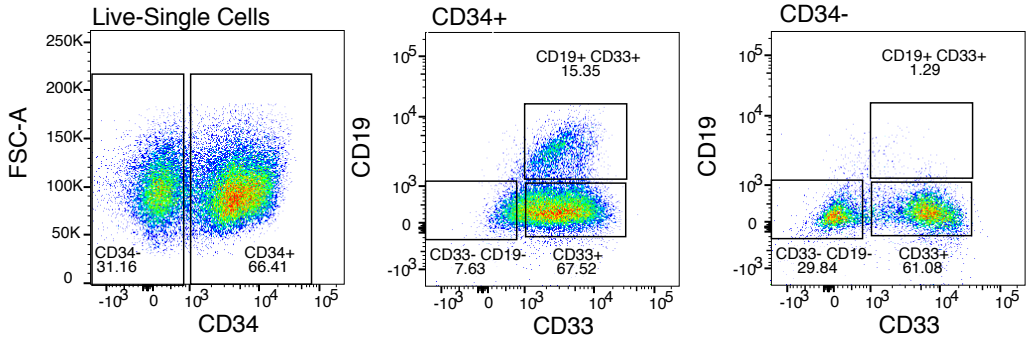


(e) Sample BIR344: LM Engraftment

Figure 4.5: Immunophenotype of Human Engraftment. Representative flow cytometry analyses of the BM from mice transplanted with AML samples (a) BIR517 ~16 weeks post-transplantation, (b) OX2116 ~24 weeks post-transplantation, (c) BEL039 ~24 weeks post-transplantation, (d) STB257 ~24 weeks post-transplantation and (e) BIR344 ~16 weeks post-transplantation. Engraftment shows both a (a) myeloid and (b–e) LM immunophenotype. Frequencies indicate the frequency of the parent gate.



(a) Sample BEL039



(b) Sample STB257

Figure 4.6: Aberrant CD19 Expression by Original patient Samples. Immunophenotyping of original patient samples for the expression of CD19, CD33 and CD34. Samples (a) BEL039 and (b) STB257 are shown. Frequencies indicate the frequency of the parent gate.

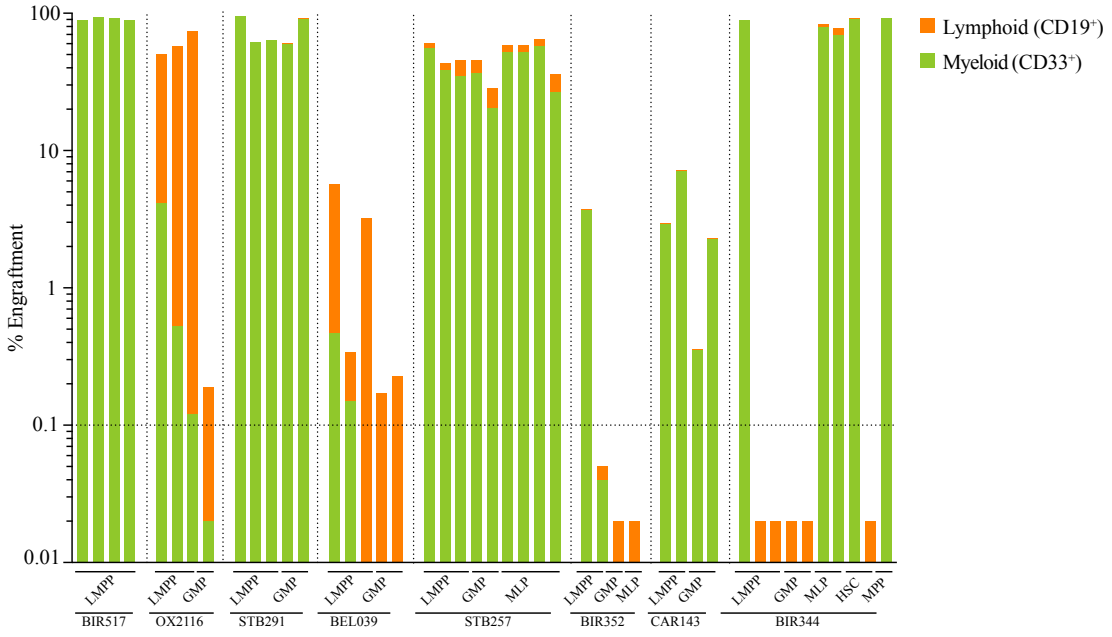


Figure 4.7: Summary of Human Engraftment in Xenotransplantation Assays. Human engraftment in xenotransplantation assays across engrafting AML samples. Percentage (%) engraftment indicates the proportion of live events that hCD45$^+$CD33/CD19$^+$ events composed at terminal engraftment. The injected HSPC population and AML sample are indicated below. Green: myeloid (CD33$^+$) engraftment. Orange: lymphoid (CD19$^+$) engraftment. To be considered positive for human engraftment, a cut-off of 0.1% engraftment was used.

4.4 Genotyping of Human Engraftment

To confirm the leukaemic nature of the $CD19^+$ engraftment, bulk BM harvested from engrafted mice was genotyped for the mutations detected at diagnosis (Table 4.2). 7/8 of the engrafting AML samples were genotyped. Sample BIR352 was not genotyped due to it not showing any mutations in the genes screened at diagnosis.

It was seen that across all genotyped engraftment, the presence/absence of mutations and variant allele frequencies (VAFs) of mutations were consistent across the various HSPC populations injected from a sample.

From the 3 myeloid engrafters, all 3 showed the presence of leukaemic mutations. Both the *ETV6* and *JAK2* mutations were detected at a VAF of ~50% in BIR517 post-engraftment, suggesting heterozygous mutations in both genes. A potentially small clone mutated in the *BCOR* gene was not present post-engraftment (Table 4.2). In sample STB291, a *FLT3*-internal tandem duplication (ITD) was present both before and after engraftment. Additionally, a small clone mutated in *NRAS* at diagnosis expanded to a VAF of ~50%, suggesting clonal expansion to a dominant heterozygous clone (Table 4.2). Finally, sample CAR143 had shown multiple mutations at diagnosis at varying VAFs, but post-engraftment only a mutation in *NRAS* was detected. Again, this was seen at ~50%, implying clonal selection and loss of the other clone(s) (Table 4.2).

All four LM engrafters were genotyped to confirm the human engraftment observed was not due to residual normal HSPC populations giving rise to normal progeny. Sample OX2116 maintained the presence of mutations in the genes *CBL*, *TP53*, and *TET2*, with the *CBL* mutation likely to be homozygous and the *TP53* mutation present in all cells. Mutations detected at lower levels in *WT1* and *STAG1* were lost post-engraftment (Table 4.2). In sample BEL039, post-engraftment mutations in *U2AF1*, *IDH1* and *TET2* were all detected at ~50%, implying heterozygous clones for these. Two other mutations detected in *TET2* at ~50% pre-engraftment were not detected post-engraftment (Table 4.2). Sample

STB257 showed the presence of a *FLT3*-ITD both pre- and post-engraftment (Table 4.2). Finally, only one mutation was detected in sample BIR344 at diagnosis and this was not present post-engraftment (Table 4.2). It is possible other mutations were present that were not captured by the initial genotyping.

Furthermore, the $CD19^+$ and $CD33^+$ populations were purified and genotyped, to confirm both populations contained the leukaemic mutations. It was observed that all of the mutations found in the bulk genotyped cells were present in both populations (data not shown).

Table 4.2: Genotype of Human Engraftment. Genotyping of bulk BM harvested from engrafted mice to confirm the presence of leukaemic mutations detected at diagnosis. Mutated genes were amplified and sequenced to allow the detection of known mutations. FLT3-ITDs were detected by PCR. Original VAFs were determined by targeted re-sequencing.

Sample	**Mutation VAF**				
BIR517	**ETV6 T1193G**	**JAK2 G1849T**	**BCOR 3489_3490insGAAC**		
Original VAF	52%	51%	6%		
LMPP 1	42%	47%	0%		
LMPP 2	44%	47%	0%		
LMPP 3	46%	49%	0%		
LMPP 4	47%	48%	0%		
OX2116	**WT1 C1091A**	**CBL C1033G**	**TP53 T91C**	**TET2 3452delT**	**STAG1 G3456A**
Original VAF	6%	82%	93%	32%	13%
LMPP 1	0%	82%	100%	43%	0%
LMPP 2	0%	82%	100%	36%	0%
STB291	**FLT3 ITD**	**NRAS C181A**			
Original VAF	Present	14%			
GMP 1	Present	45%			
GMP 2	Present	48%			
LMPP 1	Present	47%			
LMPP 2	Present	49%			
LMPP 3	Present	49%			
BEL039	**U2AF1 C101T**	**IDH1 C394T**	**TET2 2474delC**	**TET2 2849_2850 insA**	**TET2 T2847A**
Original VAF	36%	39%	42%	48%	48%
LMPP 1	49%	42%	50%	0%	0%
LMPP 2	43%	55%	49%	0%	0%
STB257	**FLT3 ITD**				
Original VAF	Present				
GMP 1	Present				
GMP 2	Present				
LMPP 1	Present				
LMPP 2	Present				
LMPP 3	Present				
MLP 1	Present				
MLP 2	Present				
MLP 3	Present				
MLP 4	Present				
CAR143	**JAK2 G1849T**	**RUNX1 C1193T**	**NRAS T190A**	**RUNX1 930dupC**	**ASXL1 2040dupC**
Original VAF	9%	31%	32%	36%	36%

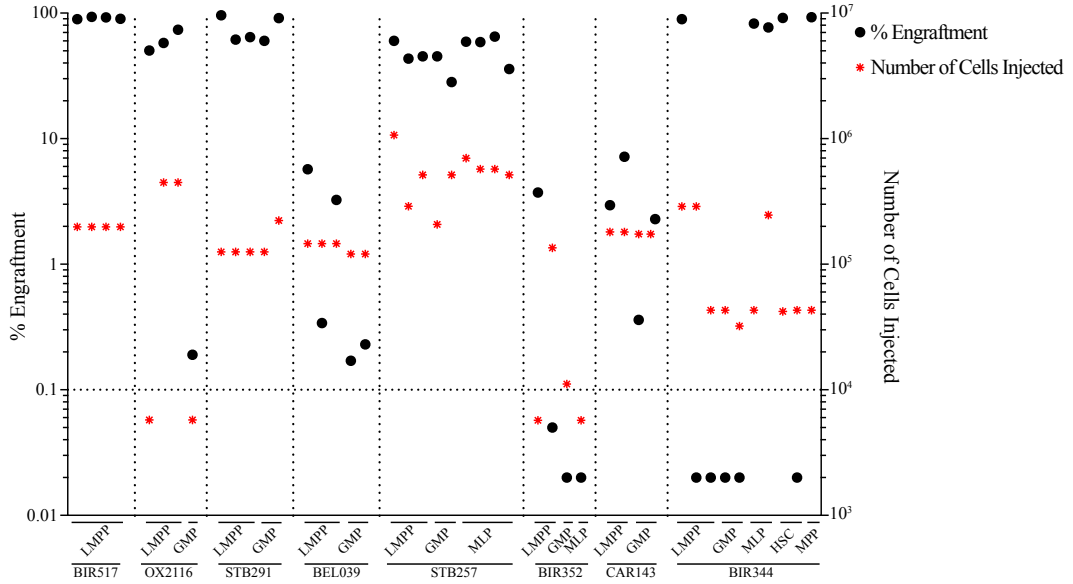
Continuation of Table 4.2					
Sample	**Mutation VAF**				
GMP 1	0%	0%	45%	0%	0%
GMP 2	0%	0%	51%	0%	0%
LMPP 1	0%	0%	50%	0%	0%
LMPP 2	0%	0%	46%	0%	0%
BIR344	**NRAS G35A**				
Original VAF	22%				
HSC 2	0%				
LMPP 3	0%				
MLP 1	0%				
MLP 2	0%				
MPP 1	0%				

4.5 Kinetics of Human Engraftment

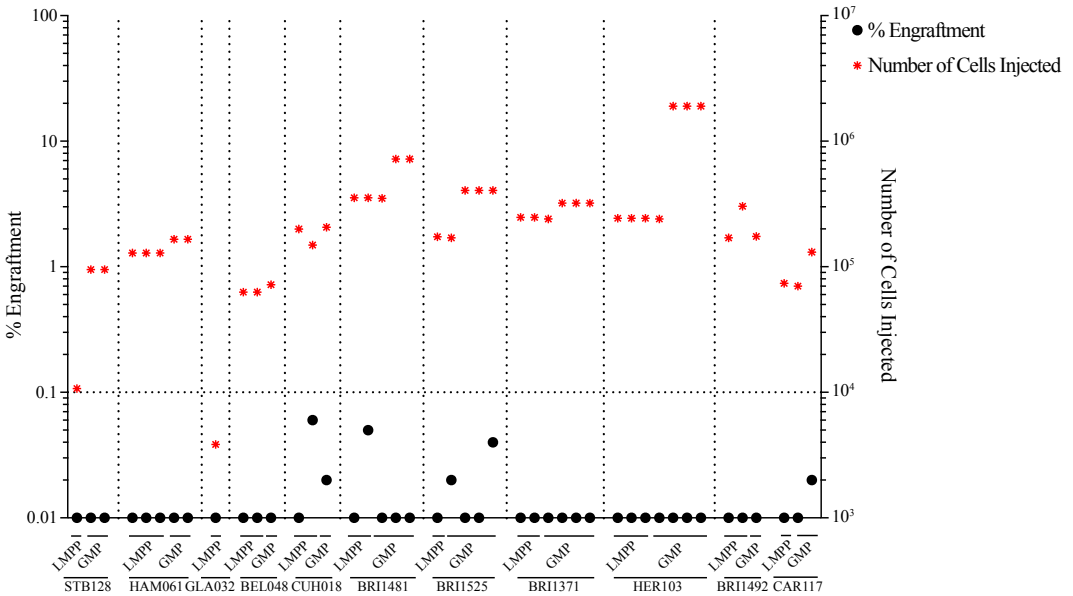
Across all AML samples a range of 3,845 to 1,902,206 cells were injected, with an average of ~300,000 cells injected per transplantation (Figure 4.8). It was seen that there was no relationship between the number of cells transplanted and the presence or absence of human engraftment (Figure 4.8).

When looking at the kinetics of human engraftment, if human engraftment was detected in PB at ~12/16 weeks, then there was usually an increase in engraftment levels in the BM at ~20/24 weeks. Terminal BM samples, harvested from the femur, tibia, pelvic and back bones, usually showed engraftment levels similar to or higher than those seen in PB/BM.

Samples BIR517 and STB291 showed very similar patterns of myeloid engraftment that increased with time in PB and BM (Figures 4.9a, 4.9c). Sample BIR344 showed a similar pattern, with increasing myeloid engraftment, however with a minor $CD19^+$ population (Figure 4.9d). Additionally, not all xenotransplantations of the various HSPC populations from BIR344 gave rise to human engraftment. Sample CAR143 gave rise to predominantly myeloid engraftment, however at lower levels to BIR517 and ST291 (Figure 4.9h). Sample OX2116 also showed increasing levels of human engraftment; however, this was predominantly $CD19^+$, and the GMP xenotransplantation did not engraft (Figure 4.9b). Sample BEL039 produced



(a) Engrafting AML Samples

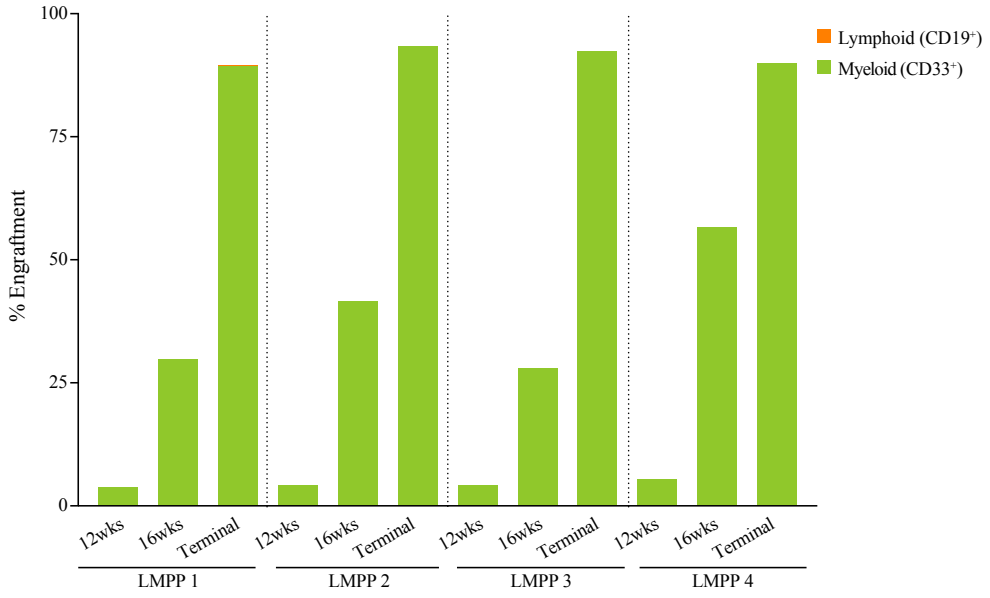


(b) Non-Engrafting AML Samples

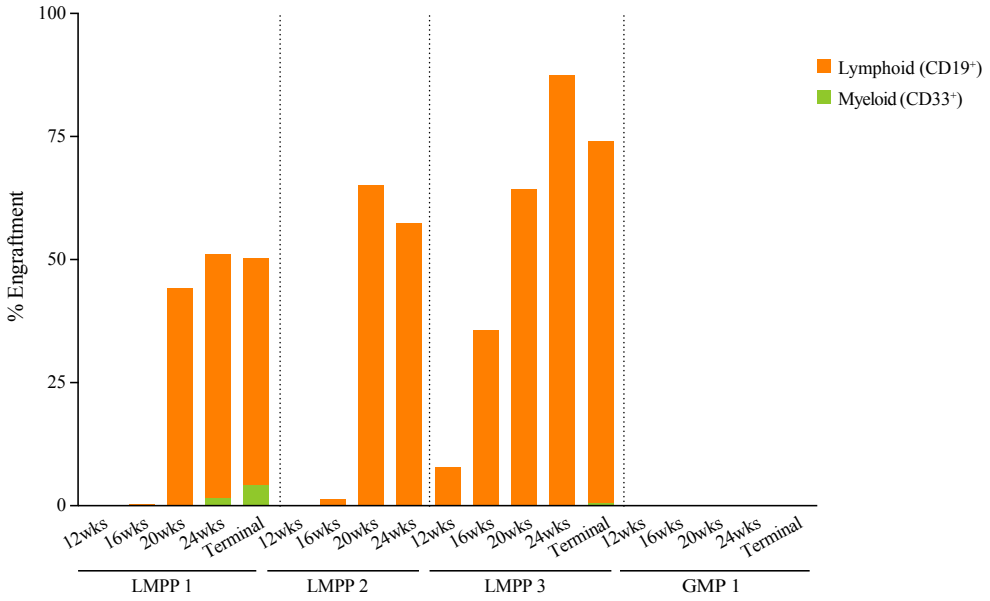
Figure 4.8: Total Human Engraftment Related to Number of Injected Cells. Human engraftment in xenotransplantation assays across (a) engrafting and (b) non-engrafting AML samples. Percentage (%) engraftment indicates the proportion of live events that hCD45$^+$CD33/CD19$^+$ events composed at terminal engraftment (black dot). The injected HSPC population and AML sample are indicated below. The number of cells injected per xenotransplantation is indicated (red star). To be considered positive for human engraftment, a cut-off of 0.1% engraftment was used.

mixed LM engraftment and the levels of engraftment were seen to be higher following LMPP, rather than GMP, xenotransplantation. The two GMP xenotransplantations only peaked above the positive threshold of 0.1% for human engraftment at the terminal BM stage (Figure 4.9e). STB257 also showed a mixed LM engraftment but to higher levels than BEL039, and this was consistent across all injected populations and increased with time (Figure 4.9g).

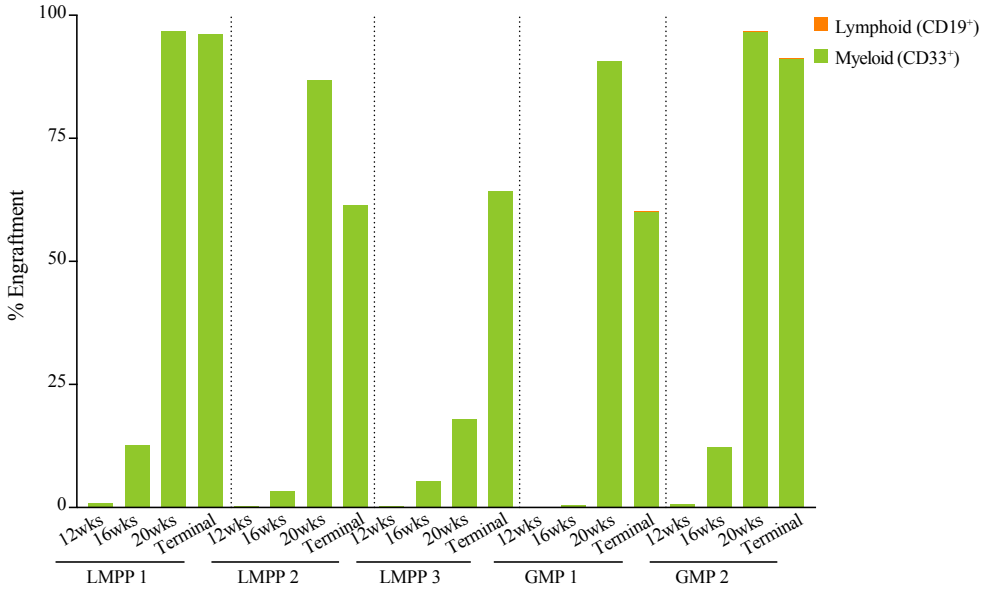
Interestingly, only one xenotransplantation of the LMPP from BIR352 engrafted successfully, and human engraftment was not detectable until the injected tibia BM was sampled at ~24 weeks. The terminal harvested BM showed lower levels of human engraftment, most likely due to the engraftment being localised to the site of injection (Figure 4.9f).



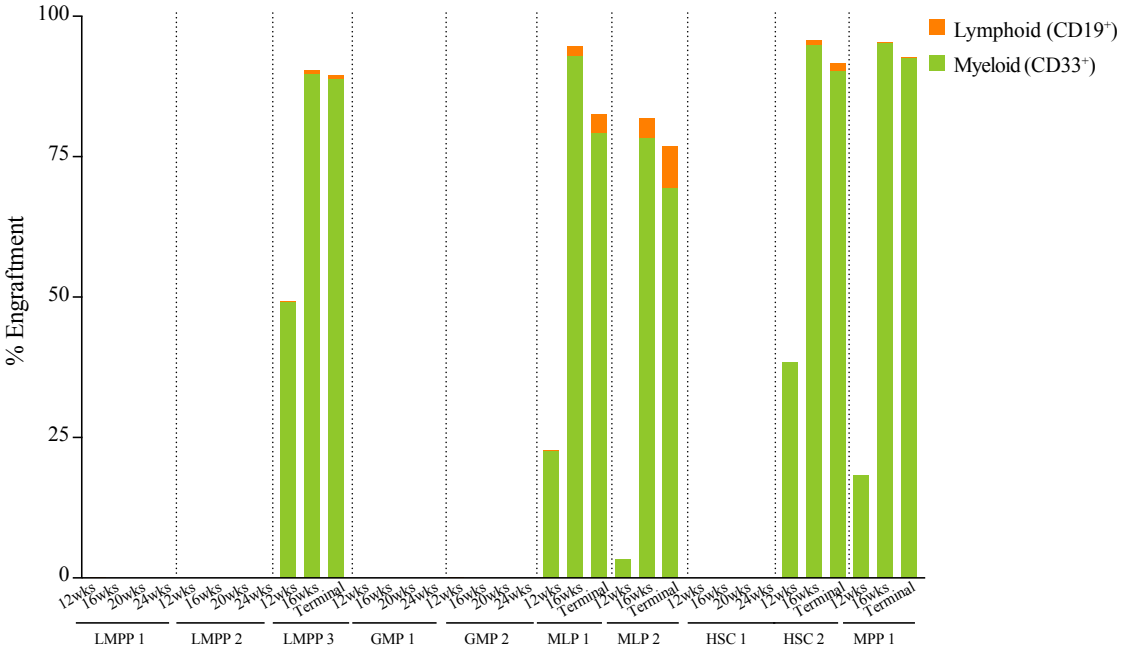
(a) Sample BIR517: Engraftment Kinetics



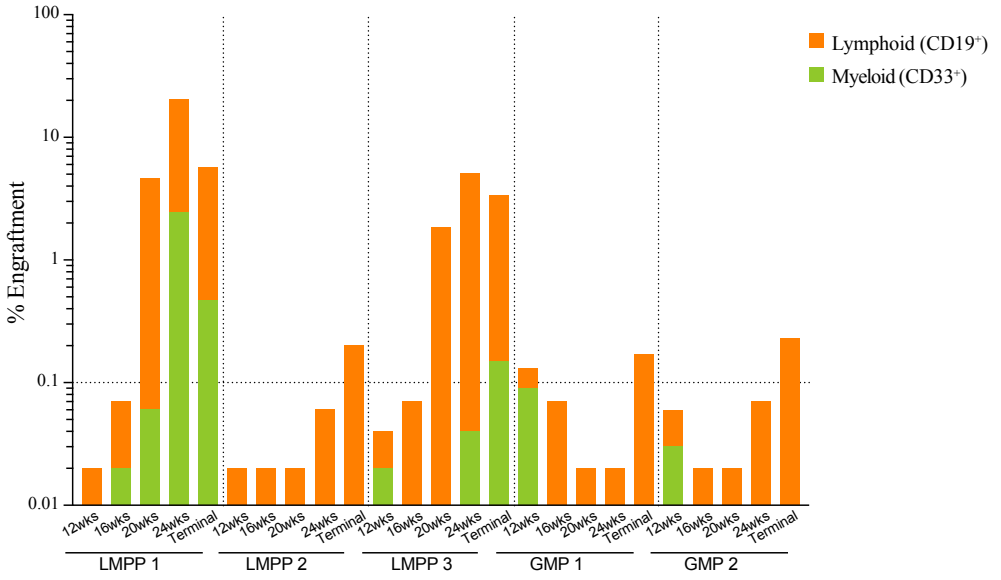
(b) Sample OX2116: Engraftment Kinetics



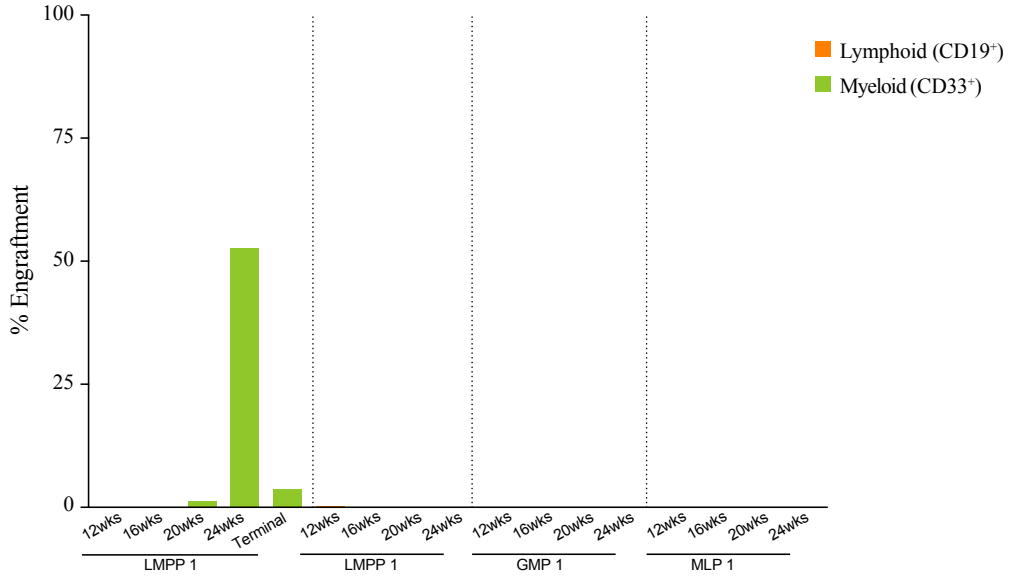
(c) Sample STB291: Engraftment Kinetics



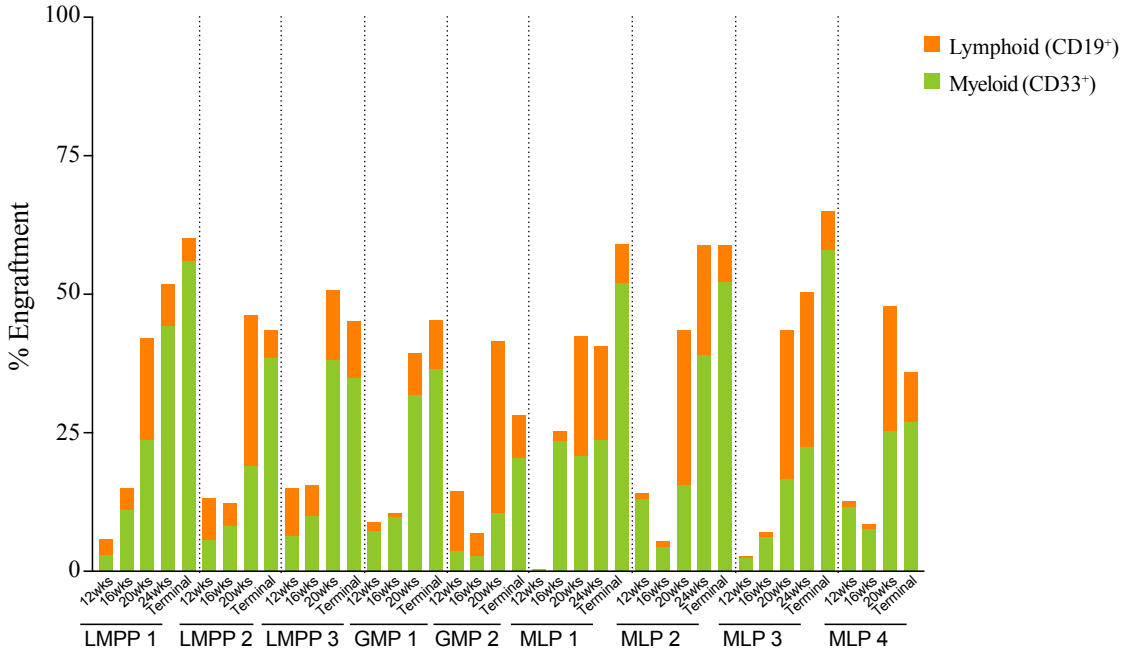
(d) Sample BIR344: Engraftment Kinetics



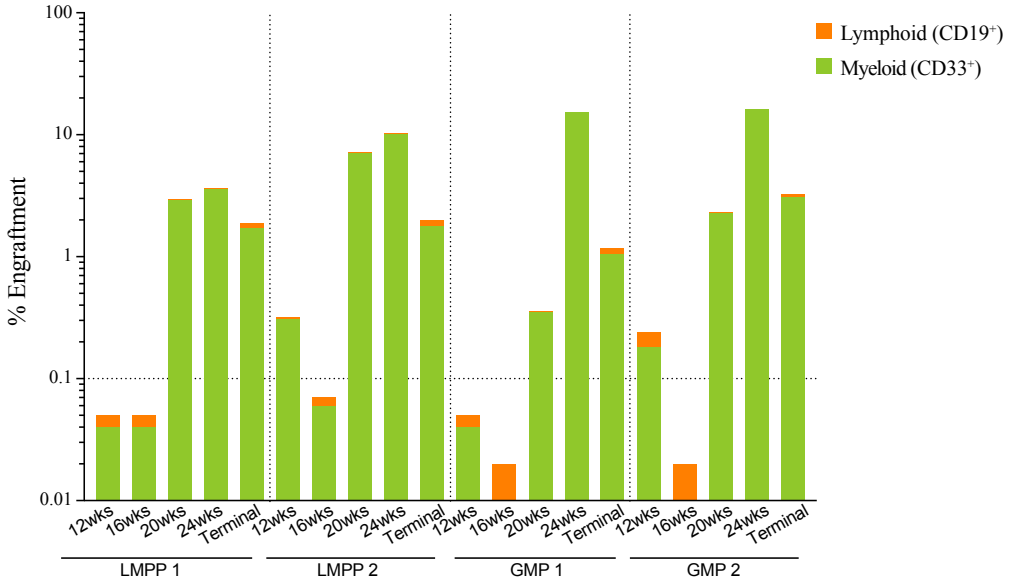
(e) Sample BEL039: Engraftment Kinetics



(f) Sample BIR352: Engraftment Kinetics



(g) Sample STB257: Engraftment Kinetics



(h) Sample CAR143: Engraftment Kinetics

Figure 4.9: Kinetics of Human Engraftment in Xenotransplantation Assays. Kinetics of human engraftment in xenotransplantation assays across engrafting AML samples. Percentage (%) engraftment indicates the proportion of live events that hCD45$^+$CD33/CD19$^+$ events composed. The injected HSPC population and time-point are indicated below. Green: myeloid (CD33$^+$) engraftment. Orange: lymphoid (CD19$^+$) engraftment. 12 and 16 week time-points represent PB samples and 20, 24 and terminal time-points represent BM samples.

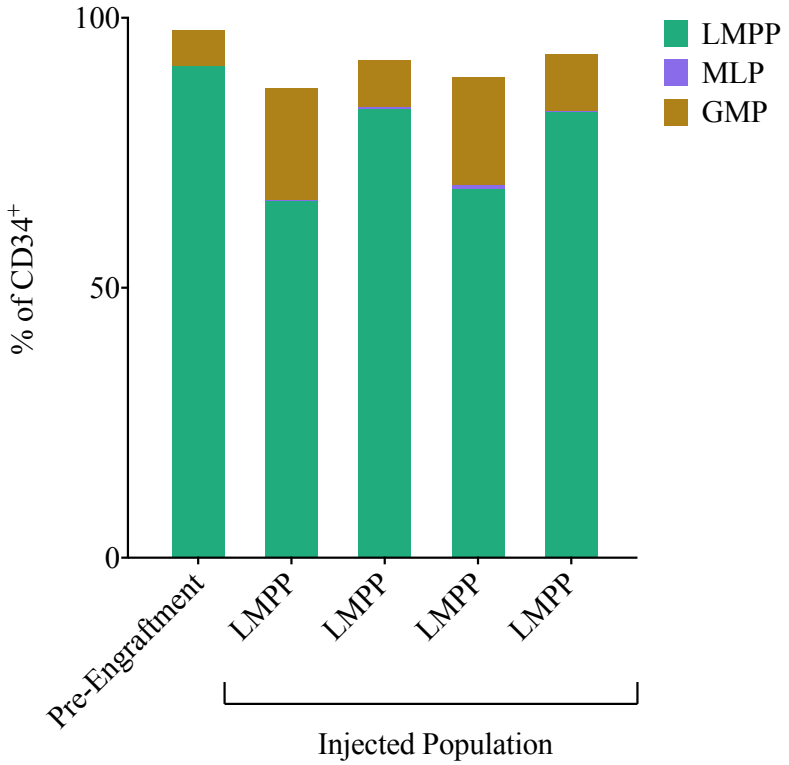
4.6 HSPC Immunophenotype of Human Engraftment

In addition to characterising the expression of myeloid and lymphoid markers by the human engraftment, the HSPC immunophenotype was also characterised in 6/8 engrafting samples. Samples BIR517, OX2116 and BIR352 all showed an LMPP/LMPP-GMP immunophenotype pre-engraftment. Post-injection of the LMPP, samples BIR517 and OX2116 showed predominantly LMPP immunophenotypes, with smaller GMP populations (Figures 4.10a, 4.10b, 4.10d, A.6, A.8, A.9). Sample BIR352 showed a larger GMP population with a smaller LMPP population, reflecting its pre-engraftment immunophenotype (Figure 4.10d, A.6).

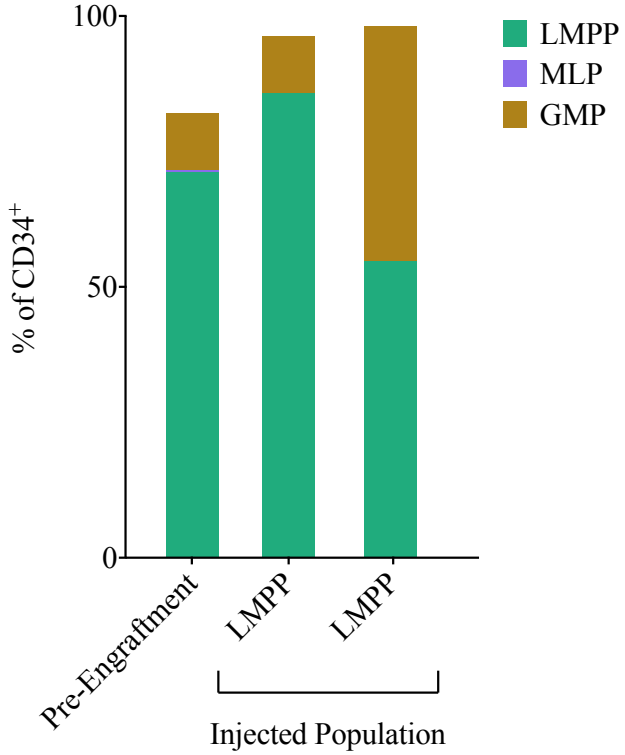
Sample CAR143 showed a dominant GMP population pre-engraftment, with a minor LMPP population. This immunophenotype was maintained post-engraftment, irrespective of the injected population (Figure 4.10f, A.7).

Sample BIR344 showed a more complex immunophenotype pre-engraftment, with minor CD45RA$^-$ HSC and MPP populations and a minor CD10$^+$ MLP. Post-engraftment, BIR344 showed a more LMPP-GMP dominant immunophenotype (Figure 4.10c, A.4). The minor MLP population was still present; however, the CD45RA$^-$ HSC and MPP populations were greatly reduced.

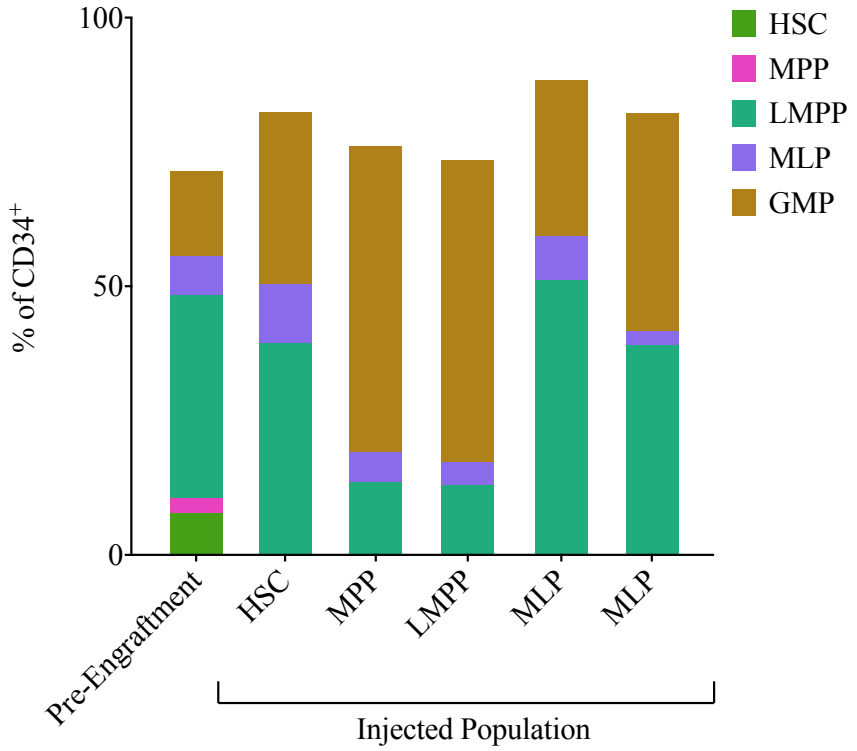
Sample STB257 had a combined LMPP-GMP-MLP immunophenotype prior to xenotransplantation. However following xenotransplantation, both the LMPP and GMP populations were reduced in size (Figure 4.10e, A.5). The MLP population was slightly reduced, with the majority of cells up-regulating both CD10 and CD38 to resemble the BNK progenitor.



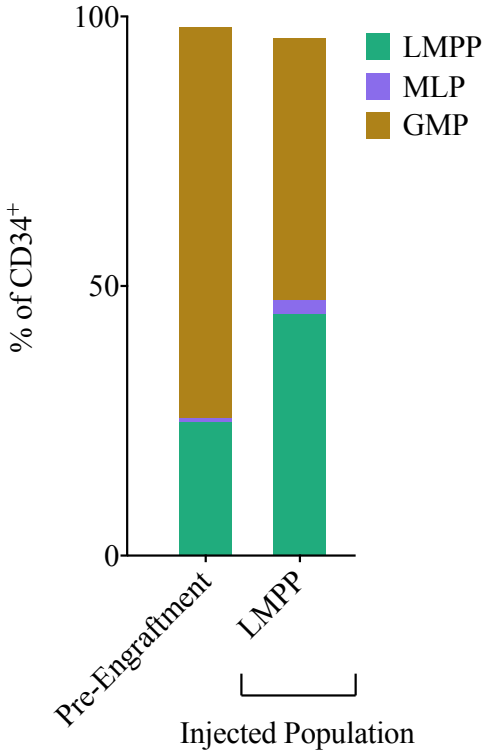
(a) Sample BIR517



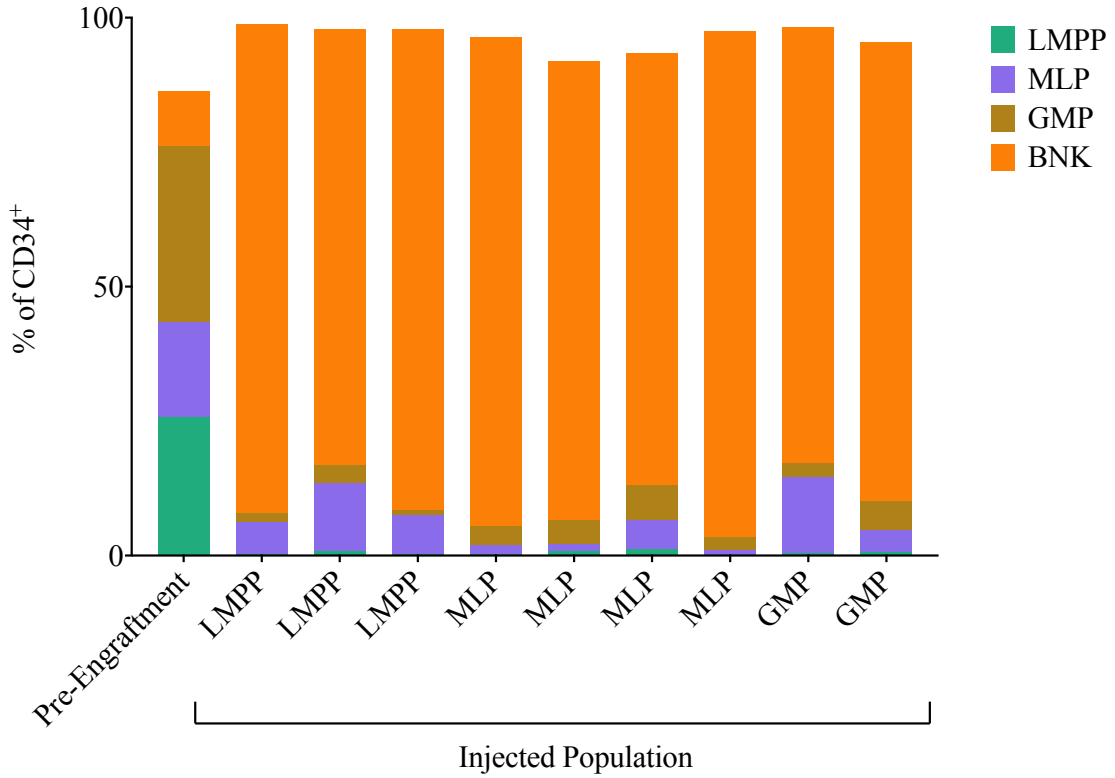
(b) Sample OX2116



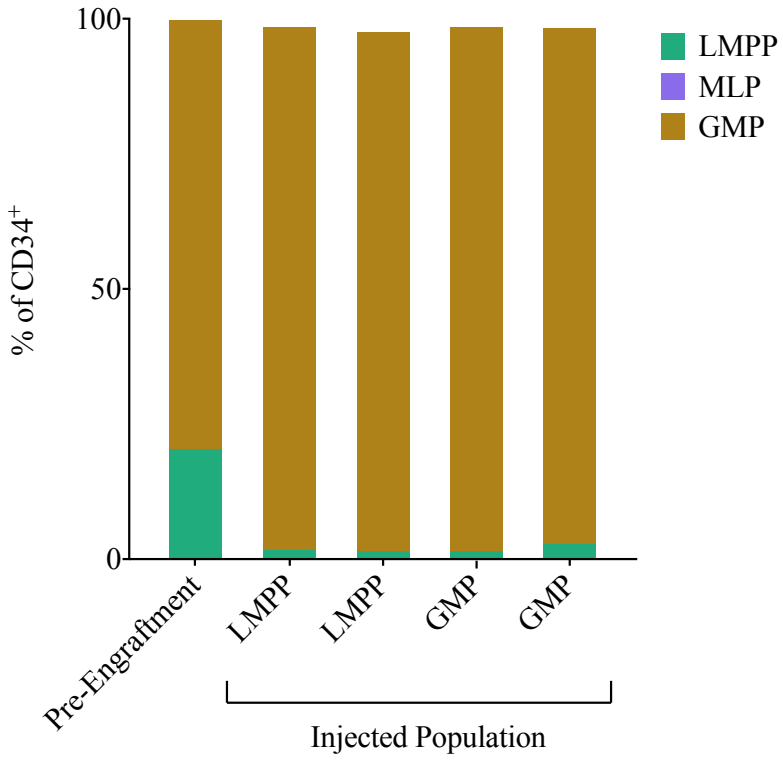
(c) Sample BIR344



(d) Sample BIR352



(e) Sample STB257



(f) Sample CAR143

Figure 4.10: HSPC Immunophenotype of Human Engraftment. Post-engraftment HSPC immunophenotype of human engraftment in xenotransplantation assays across engrafting AML samples. Percentage (%) of $CD34^+$ indicated on the y-axis indicates the proportion of the $CD34^+$ compartment that each HSPC population composed post-engraftment. The observed HSPC population post-engraftment is shown as indicated by the key. The injected HSPC population is indicated on the x-axis. HSPC immunophenotyping was carried out on terminal BM samples.

4.7 Conclusion on the Xenotransplantation of AML

The immunophenotype of ~50 AML patient samples was screened, and it was observed that the majority of AML cases showed an expanded progenitor-like profile. In the majority of cases this was a GMP-expanded immunophenotype, often with an expanded LMPP-like population. The MLP population, previously unstudied in AML, was only very rarely observed to be expanded. In xenotransplantation assays it was seen that eight samples had the ability to engraft mice with leukaemia. Four of these engrafting samples gave a myeloid-only engraftment and four of these samples gave an engraftment with aberrant expression of CD19. Across all the engrafting samples it could be proposed, based on the genotype of the engraftment, that typically only one clone from the original patient sample engrafted. However, the same clone was seen to engraft regardless of the immunophenotypic population injected. Similarly, the HSPC immunophenotype of engraftment was seen to be consistent across all injections of a particular sample, and this was mostly seen to mimic the original sample immunophenotype. Therefore, progenitor-like leukaemic populations have leukaemia-propagating ability and, in primary xenografts, are functionally consistent within a sample.

Chapter 5

Transcriptional Analysis of Progenitor-Like Leukaemia Stem Cells

Contents

5.1 Introduction to the Gene Expression Analysis of AML 149
5.2 Processing and Quality Control of RNA-Seq Data . . 152
5.3 Analysis of Normal HSPC Populations 158
5.3.1 PCA and DEGs Across Normal HSPC 158
5.3.2 Normal Stem Cell Gene Signature 164
5.4 Analysis of Leukaemic HSPC Populations 167
5.4.1 PCA of Leukaemic HSPC 167
5.4.2 Engraftment and Leukaemic Gene Expression Profiles . 170
5.4.3 Genotype and Leukaemic Gene Expression Profiles . . . 171
5.4.4 Top Variant Genes Across Leukaemia 172
5.4.5 Gender and Leukaemic Gene Expression Profiles 180
5.5 Analysis of Normal and Leukaemic HSPC Populations 182
5.5.1 Hierarchical Clustering and PCA of Normal and Leukaemic HSPC 182
5.5.2 PCA Projections of Leukaemic HSPC 187
5.5.3 Expression of Normal DEGs by Leukaemic HSPC . . . 190
5.5.4 DEGs Between Normal and Leukaemic HSPC 191
5.5.5 Expression of Stem Cell Signature by Leukaemic HSPC 199
5.6 Analysis of the Normal and Leukaemic LMPP and GMP 200
5.6.1 Normal LMPP and GMP 200
5.6.2 Leukaemic LMPP and GMP 202
5.6.3 Comparing Normal and Leukaemic LMPP and GMP . . 204
5.7 Conclusion on the Gene Expression Analysis of AML 209

5.1 Introduction to the Gene Expression Analysis of AML

In order to better understand the underlying biology of the LSCs and leukaemogenesis, the gene expression profiles of LSCs may be analysed. Gene expression profiling of primary human AML has previously been carried out in order to gain insight into the disease. This has been done both by microarray analysis [22, 47, 111–125] and RNA-Seq analysis [124, 126–129]. However traditionally, due to the

large number of cells needed for these methods, this profiling was done on bulk cells that were unrefined for LSC populations. Most often bulk BM/PB cells [111–113, 115–119, 121–127, 129] were used for analysis.

In a few cases, gene expression profiling has been carried out on refined populations. Eppert et al. carried out microarray analysis on the following immunophenotypic populations: $CD34^-CD38^-$, $CD34^+CD38^-$, $CD34^-CD38^+$ and $CD34^+CD38^+$ [47]. Others have looked at the $CD34^+CD38^-$ and $CD34^+CD38^+$ populations [120]. CD90 has been included in some studies as a marker to distinguish the normal HSCs from LSCs, with Gentles et al. defining the LSCs as $Lin^-CD34^+CD38^-CD90^-$, the leukaemic progenitor cells as $Lin^-CD34^+CD38^+$ and leukaemic blasts as Lin^-CD34^- [114]. Majeti et al. have incorporated CD123 in addition to CD90, defining the LSCs as $Lin^-CD34^+CD38^-CD123^+/CD90^-$ [128]. Finally, some analysis has been done on the LMPP- and GMP-like LSCs, defined as $Lin^-CD34^+CD38^-CD90^-CD45RA^+$ and $Lin^-CD34^+CD38^+CD123^{+/lo}CD110^-CD45RA^+$ respectively [22].

The majority of studies have used leukaemic gene expression profiles to generate gene signatures that may be associated with patient prognosis. In some cases this has been related to sub-types of leukaemia, and/or aimed to classify AML samples [111–113, 116–118, 121–127]. Others have compared leukaemic stem/progenitor populations within AML [22, 47, 114, 120]. For example, Eppert et al. compared immunophenotypic populations enriched for LSC ability with populations without detectable LSC potential. This was then compared to a normal HSC signature that had been generated [47]. On the other hand, AML samples with varying LSC frequencies have been compared, as determined by xenotransplantation LDA [129]. In order to consolidate these efforts, meta-analyses have been carried out ([130, 131]) on the published gene expression profiles and refined the reported prognostic gene signatures.

However, only a few groups have compared AML with normal haematopoiesis in attempt to further understand leukaemogenesis. Stirewalt et al. were the first to carry out this analysis, comparing bulk AML samples to normal bulk BM, bulk PB, $CD34^+$ BM and $CD34^+$ PB [119]. This analysis identified genes proposed to have

expression changes specific to AML. However, this was limited by looking at bulk cells, where the majority of cells would have been leukaemic blasts rather than LSCs. More recently, a similar comparison of bulk AML cells with either normal bulk BM or PB was carried out, identifying up- and down-regulated genes in leukaemia [115].

Following on from the work done by Stirewalt et al., Majeti et al. refined this analysis by comparing LSC populations with normal HSCs. Here both the definition of the LSCs ($Lin^-CD34^+CD38^-CD123^+/CD90^-$) and its proposed normal counterpart the HSCs ($Lin^-CD34^+CD38^-CD90^{+/lo}$) were refined [128]. Several pathways and gene networks were identified and suggested to be dysregulated when comparing the LSCs to the normal HSCs [128]. However here, only the $CD38^-$ LSCs were investigated and it was shown in following years that the $CD38^+$ compartment also contains LSC ability [22, 36].

More recently, Rapin et al. carried out a comparison of leukaemic and normal gene expression profiles, using previously published AML data sets [132] (AML data from: [111, 121–125, 130]). Although these previously published datasets on AML were limited by being generated from bulk cells, as a normal comparison highly purified HSPC populations were used. Most importantly, the stem and progenitor populations were defined as follows: HSC: $Lin^-CD34^+CD38^-CD90^+CD45RA^-$, MPP: $Lin^-CD34^+CD38^-CD90^-CD45RA^-$, CMP: $Lin^-CD34^+CD38^+CD45RA^-CD123^+$ and GMP: $Lin^-CD34^+CD38^+CD45RA^+CD123^+$. Additionally, myeloid precursors and mature myeloid cells were also purified and studied, as defined previously [132]. Firstly, the normal counterpart of individual leukaemic patients was identified, and this was seen to vary across leukaemia sub-types. Secondly, DEGs were identified between each patient sample and its normal counterpart. This enabled the identification of transcriptional pathways proposed to be involved in leukaemic transformation and maintenance [132].

I aimed to fully characterise the gene expression profile of highly purified LSC populations in AML. To achieve this, RNA-Seq was carried out on the immunophenotypic LMPP- and GMP- like populations and, if present, the MLP-like population from AML samples. As a normal comparison, RNA-Seq was carried out

on all 7 normal BM HSPC populations, as defined in Chapter 3. Firstly, the normal HSPC populations were analysed to confirm the gene expression data recapitulated the haematopoietic hierarchy. The LSC populations were then analysed in isolation, where the variability across AML samples was investigated. Next, the leukaemic populations were analysed in the context of the normal HSPCs, allowing these expression profiles to be compared and contrasted. This analysis then informed the direct comparison of LSC populations with their normal counterparts, identifying genes differentially expressed in leukaemia. Finally, the gene expression changes between the LMPP and GMP populations in normal haematopoiesis were identified, and it was investigated how this is dysregulated in leukaemia.

5.2 Processing and Quality Control of RNA-Seq Data

RNA-Seq libraries were prepared from all 7 HSPC populations from 3 healthy BM controls and from the LMPP- and GMP-like LSC populations from 17 AML patients (Table 5.1). All 17 of these patients had previously been analysed for their leukaemia propagating activity in xenotransplantation assays (see Chapter 4). As previously described in Chapter 4, sample STB257 had an immunophenotypic MLP population, and this was also analysed by RNA-Seq (Table 5.1).

Table 5.1: Summary of RNA-Seq of Normal and Leukaemic HSPCs.

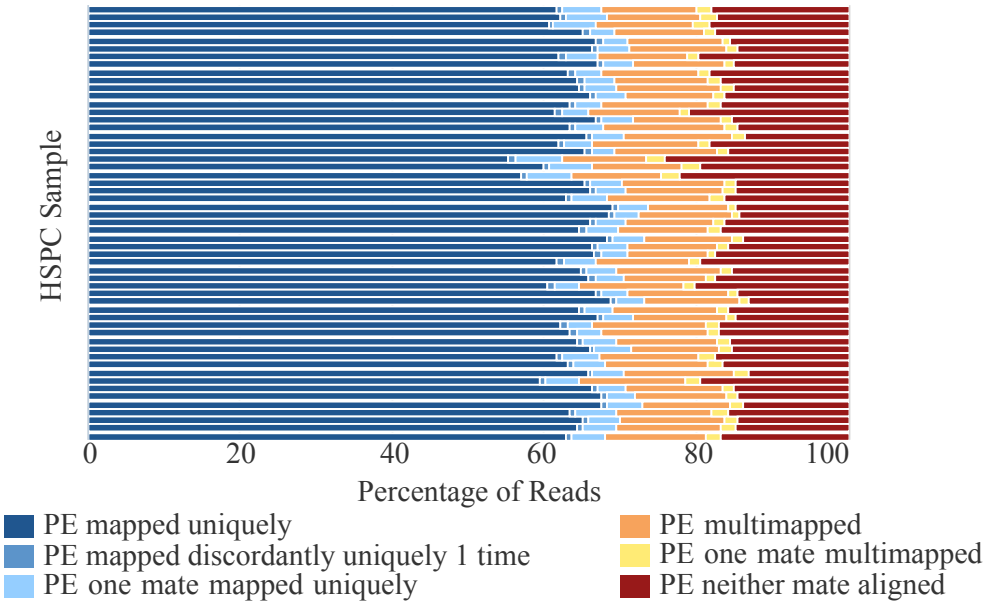
Condition	patient	Population
Normal	OX671	HSC MPP LMPP MLP CMP GMP MEP
	OX700	HSC MPP LMPP MLP CMP GMP MEP
	OX701	HSC MPP LMPP MLP CMP GMP MEP
Leukaemic	BEL039	LMPP GMP
	BEL048	LMPP GMP
	BIR1525	LMPP GMP
	BIR517	LMPP GMP
	BRI1371	LMPP GMP
	BRI1481	LMPP GMP
	BRI1492	LMPP GMP
	CAR117	LMPP GMP
	CAR143	LMPP GMP
	CUH018	LMPP GMP
	GLA032	LMPP

Continuation of Table 5.1		
Condition	patient	Population
	HAM061	LMPP GMP
	HER103	LMPP GMP
	OX2116	LMPP GMP
	STB128	LMPP GMP
	STB257	LMPP MLP GMP
	STB291	LMPP GMP

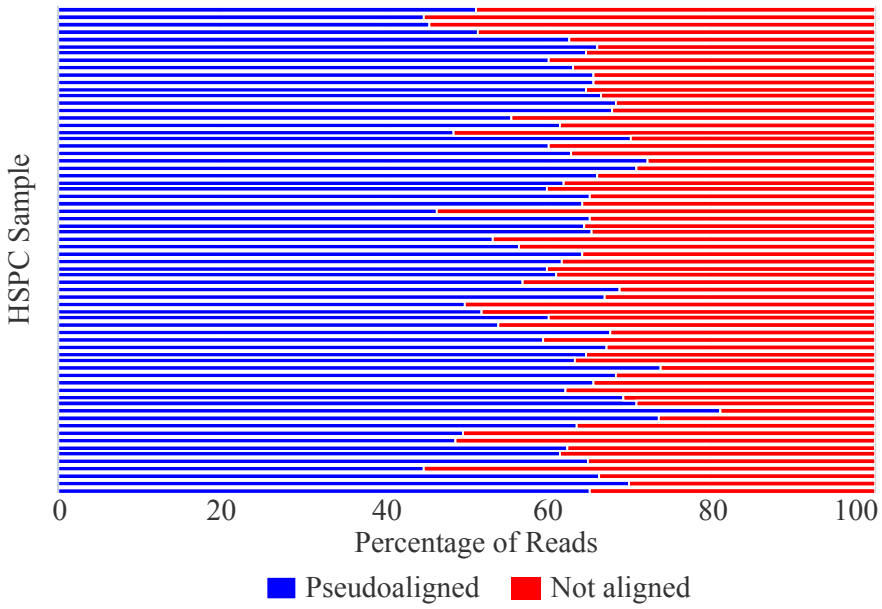
RNA-Seq data was processed and analysed as described in Chapter 2.14.4. Trimmed reads were processed traditionally by mapping to the human genome and alternatively pseudoaligned and quantified using Kallisto [74] (Figure 5.1). Kallisto pseudoaligns reads to a reference, without the alignment of individual bases, and is therefore a faster method when compared to traditional mapping [74].

Mapping showed a high proportion of reads that mapped uniquely to the genome, typically ~60–70% of reads. Only ~20% of reads did not align, and the remainder showed multiple-mapping to the genome (Figure 5.1a). On the other hand, pseudoalignment with Kallisto gave a similar proportion of reads being pseudoaligned as were uniquely mapped (Figure 5.1b). It has been suggested that pseudoalignment may be more accurate than traditional mapping methods, and Kallisto pseudoaligned results were taken forward for further analysis [133, 134]. Preliminary analysis was also carried out on reads mapped and quantified with featureCounts [73] and similar results were seen from both methods (data not shown).

Gene dispersion estimates were then calculated (Figure 5.2). As expected, genes with lower average expression values typically displayed a higher variation in their expression. The estimated individual gene dispersions are shown in Figure 5.2, grey. Then by sharing information across all gene dispersions, a fitted curve of expected variation in expression was calculated (Figure 5.2, red). Finally, individual gene



(a) Mapping Statistics



(b) Pseudoalignment Statistics

Figure 5.1: Mapping and Pseudoalignment Statistics of Normal and AML HSPCs. A summary of (a) mapping and (b) pseudoalignment statistics of bulk RNA-seq of normal and leukaemic HSPC populations. Mapping performed using HISAT2 and pseudoalignment done using Kallisto. Plots generated using MultiQC [135]. Each row represents one RNA-seq sample. PE: paired-end read.

dispersions were shrunk towards the fitted line (Figure 5.2, blue dots), with the exception of outliers (Figure 5.2, encircled in blue), as previously described [75].

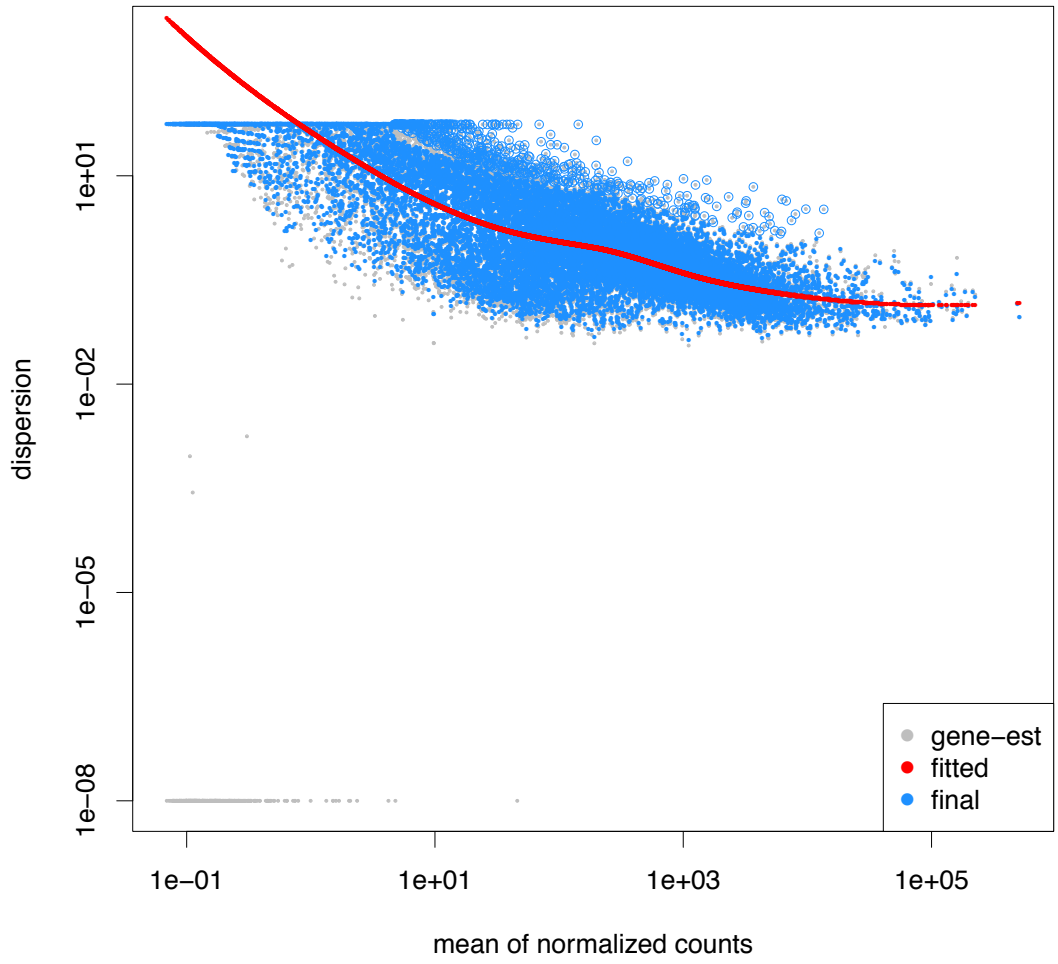


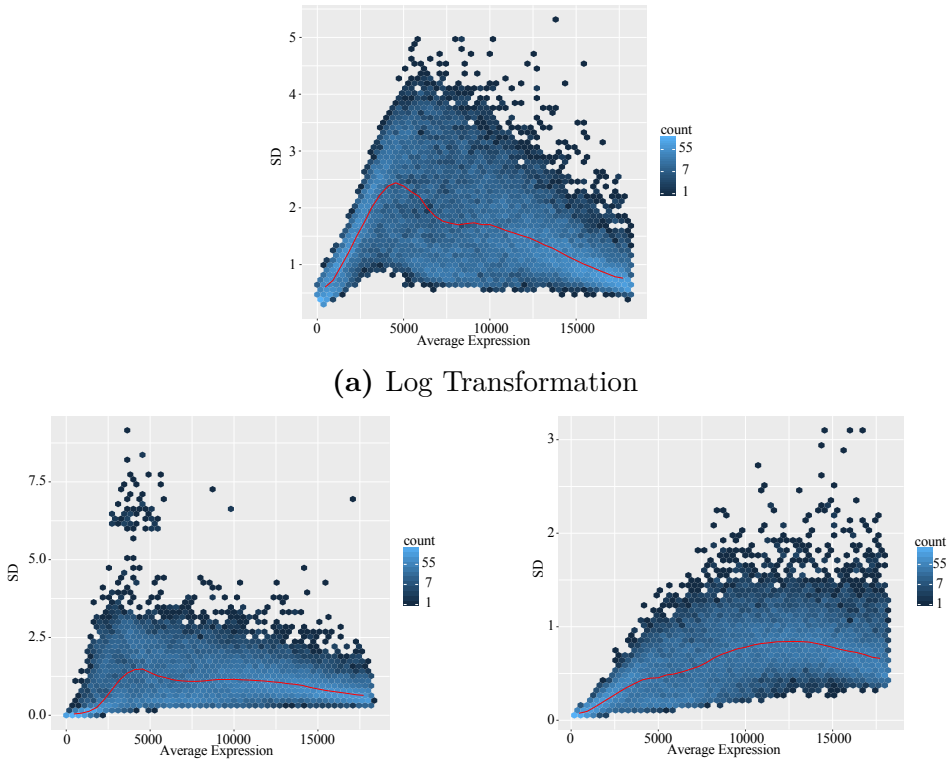
Figure 5.2: Dispersion Estimates of Normal and AML HSPCs. Dispersion estimates from RNA-Seq data from normal and AML HSPC using a local fit type and estimated by DESeq2. gene-est: gene-wise dispersion estimates

For many downstream analyses of RNA-Seq data, it is preferable to have data that is homoskedastic (i.e. data that shows homogenous variance). Therefore, counts were then transformed and normalised in DESeq2, by the following approaches. Firstly, the effect of a simple log transformation on counts can be seen in Figure 5.3a. As expected, genes with a lower average expression showed a spreading of their variance (indicated by standard deviation) with the maximum in the average

standard deviation being ~2.5 (as indicated by the red line of best fit, Figure 5.3a). This is not ideal for downstream analyses, as it would result in the logFC values for lowly expressed, highly variant genes being exaggerated [75].

Therefore, the regularised log (rlog) transformation has previously been developed to not only log transform data, but to also shrink the variation of lowly expressed genes, which is often due to technical noise [75]. The effect of this transformation was seen to lower the average standard deviation to ~1.25 (as indicated by the red line of best fit, Figure 5.3b). The average standard deviation was also seen to be more constant across genes of varying expression levels. However, this transformation was not effective on some genes, which showed a markedly higher standard deviation of ~6.25–7.5. This effect was seen to be exaggerated when reads mapping to non-coding exons were included (data not shown).

In the original version of DESeq, instead of an rlog transformation, a variance stabilising transformation (VST) was used [136]. The VST is also effective at producing more homoskedastic data, but can be limited by certain factors, for example if the data contains very large variations in sequencing depth [75]. It can be seen that the VST was successful at reducing the variation across genes of all expression levels (Figure 5.3c). Here the average standard deviation was never higher than 1, and the genes that were seen to have abnormally high variation, as a result of rlog transformation, were corrected (Figure 5.3c). Therefore, VST normalised data was taken forward for downstream analysis.



(a) Log Transformation

(b) rlog Transformation

(c) VST

Figure 5.3: RNA-Seq Count Data Transformation. Transformation of raw RNA-Seq count data from normal and AML HSPC using either (a) log transformation (b) rlog transformation or (c) VST (as previously described [75]).

5.3 Analysis of Normal HSPC Populations

5.3.1 PCA and DEGs Across Normal HSPC

Firstly, the normal HSPC populations were analysed in isolation to confirm that the gene expression data recapitulated the haematopoietic hierarchy. As an unbiased approach to investigate the global gene expression profile, a PCA was carried out on the normal HSPC populations using all expressed genes (Figure 5.4a). It was seen that the PCA was able to recapitulate the haematopoietic hierarchy and separate out the HSPC populations. PC2 was an approximate axis of differentiation, with the earliest stem and progenitor HSC and MPP populations sitting at the bottom of this axis. However, it could be seen that there was some variability across the MPP populations, possibly reflecting the heterogenous nature and poor definition

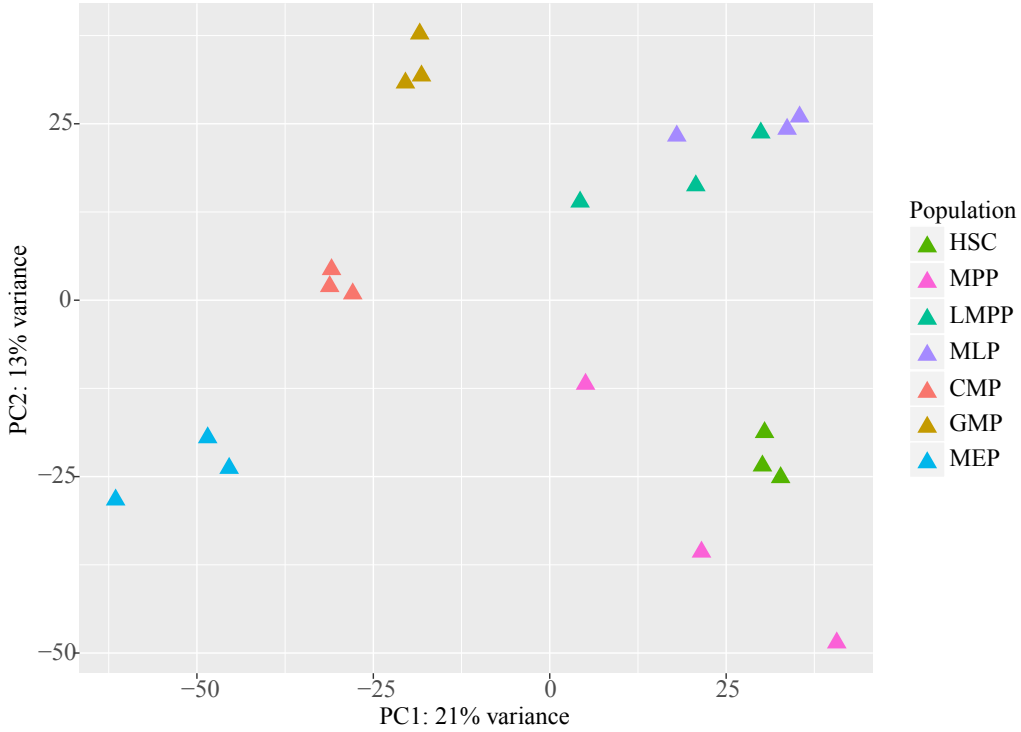
of the MPPs/ST-HSCs in humans [3, 12, 137]. PC1 separated out the MEP from the more lymphoid populations (LMPP and MLP). Using these genes, there was not a clear separation of the LMPP and MLP populations (Figure 5.4a).

A PCA was also carried out using an unbiased selection of the 500 genes with the highest variance across all HSPC populations, which does not specify population identity *a priori*. Here the HSC and MPP populations were more clearly segregated from the slightly more mature progenitors on PC2 and there was some separation of the LMPP and MLP populations (Figure 5.4c).

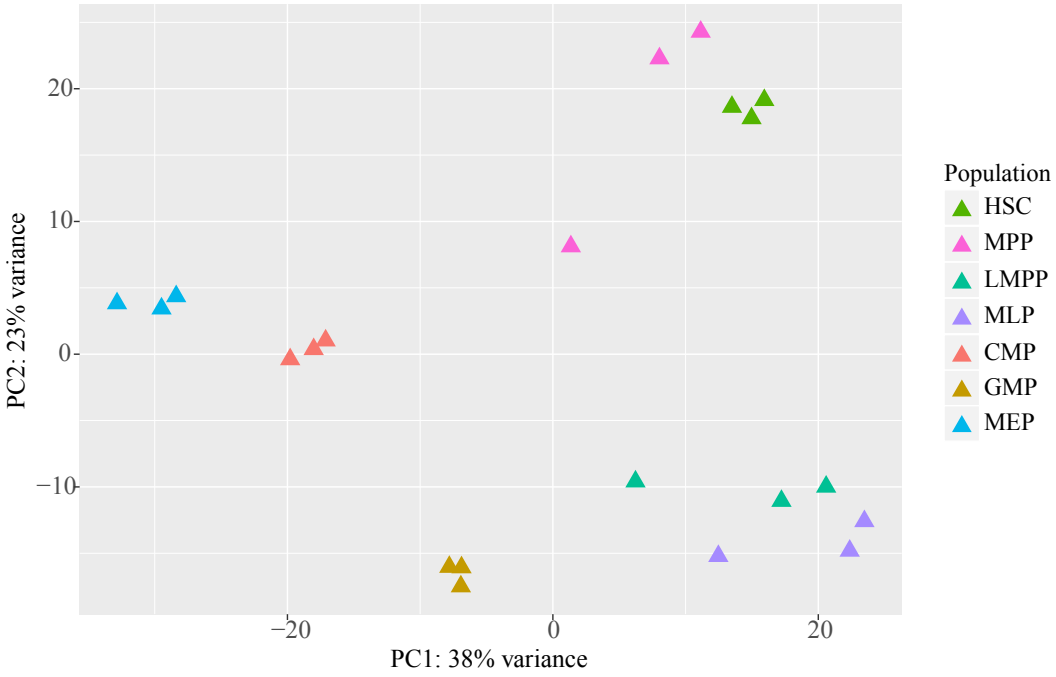
DEGs were identified across the HSPC populations using an LRT in DESeq2, which is comparable to an ANOVA test but for data that is not normally distributed. This analysis does use population identity to inform its identification of DEGs. Using the 500 most significantly DEGs, a PCA was run and shown in Figure 5.4c (see Table A.23 for DEGs). Similar to the 500 most variant genes, a more clear and distinct separation of the HSPC populations was seen.

When a PCA of normal HSPCs was carried out using the 500 most significantly DEGs, PC1 segregated the MEP from the more lymphoid populations (LMPP and MLP) (Figure 5.4c). When looking at the first three components of the PCA (Figure 5.5), it was seen that PC3 represented an axis of myeloid differentiation. The GMP was well separated with the highest values on PC3, representing its myeloid potential. The CMP sat below the GMP and, reflecting its myeloid potential, the LMPP was seen to have a higher PC3 value than the more lymphoid MLP (see Chapter 3 and Figure 5.5).

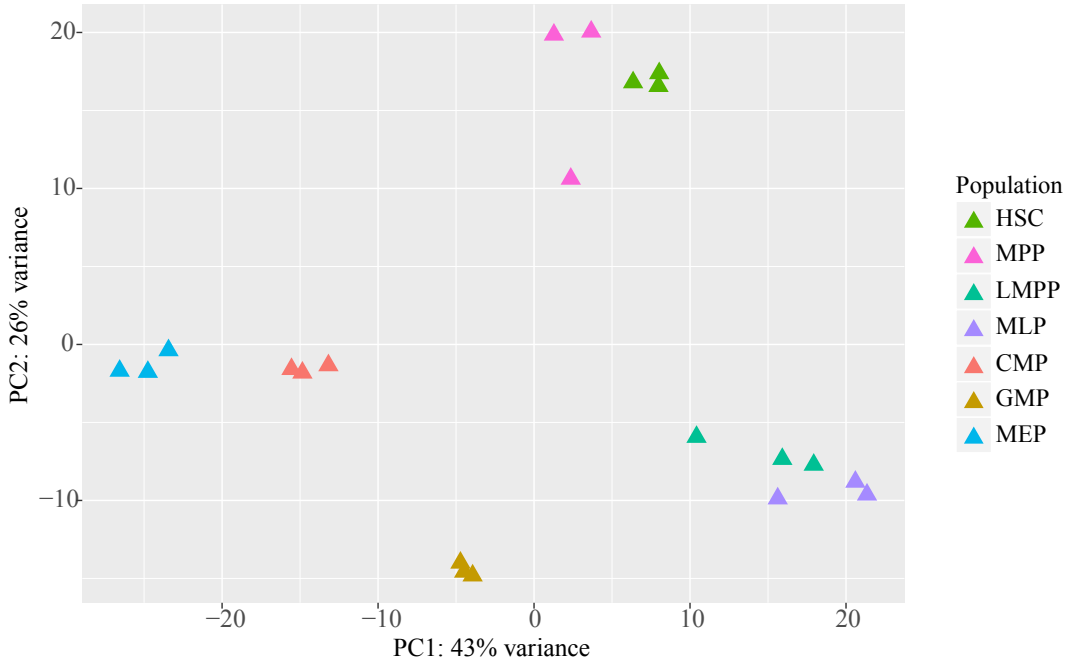
When focussing on the 500 most significantly DEGs across the normal HSPC populations, it was seen that these reflected the function and lineage output of the HSPC populations (see Table A.23 for DEGs). Figure 5.6 shows the expression of these genes by the HSPC populations, where both genes and HSPC populations are hierarchically clustered. It can be seen that the hierarchical clustering was able to cluster not only the HSPC populations, but also the genes into clusters that reflected the function and lineage output of the HSPC populations.



(a) All Genes



(b) 500 Variant Genes



(c) 500 LRT Genes

Figure 5.4: PCA of Normal BM HSPC. PCA of normal BM HSPC populations using (a) all expressed genes (b) 500 most variable genes and (c) 500 most differentially expressed genes across normal HSPC populations, as determined by an LRT (see Table A.23 for DEGs). The percentage variance contributed by each PC is indicated. Each point represents a biological replicate.

Firstly, a cluster of genes can be seen with a high level of expression in the MEP (Figure 5.6). The genes in this cluster include the key haematopoietic TF *TAL1* and the erythroid-megakaryocytic TF *GFI1B*. They also showed an enrichment for the Human Gene Atlas category $CD71^{+}$_EarlyErythroid (p-adjusted value = 2.56 x 10^{-08}, as determined by Enrichr analysis [138]).

Another cluster consisted of myeloid-associated genes showing higher expression in the CMP and GMP, with some overlap with the MEP. This also contained a sub-cluster of genes that showed markedly higher expression in the GMP (Figure 5.6). The myeloid cluster of genes included the myeloid TFs *CEBPA*, *CEBPD*, and *SPI1* and granule associated genes *AZU1*, *RNASE3*, *LYZ*, and *ELANE*.

A cluster of genes showed increased expression in the earliest HSC and MPP

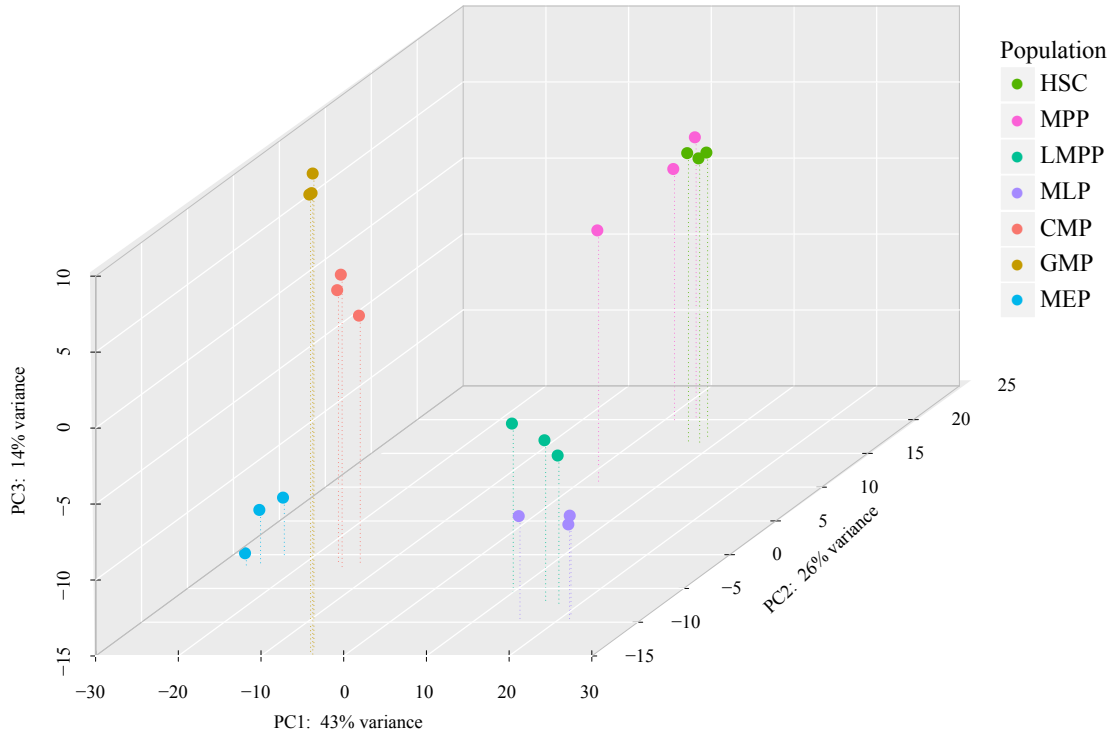


Figure 5.5: PCA of Normal BM HSPC: PC1–PC3. PCA of normal BM HSPC populations using the 500 most differentially expressed genes across normal HSPC populations, as determined by an LRT (see Table A.23 for DEGs). The percentage variance contributed by each PC is indicated and PCs 1-3 are shown. Each point represents a biological replicate.

populations (Figure 5.6). These genes included *MEIS1*, *HLF* and *ERG*, all of which have previously been implicated with stem cell maintenance [139–141]. They also showed enrichment for the ARCHS4 tissue category $CD34^+$ cell (p-adjusted value = 1.39 x 10^{-08}, as determined by Enrichr analysis [138]).

Across the LMPP, MLP, HSC and MPP there was a cluster of genes which showed intermediate expression in HSCs-MPPs and slightly higher expression in LMPPs and MLPs (Figure 5.6). The genes in this cluster showed a mixed lineage association and included the early stem associated TF *HOXB3* and stem/progenitor gene *CD34*, antigen presentation associated genes *HLA-DR*, *HLA-DPB1*, *HLA-DMA* and *HLA-A*, lymphoid gene *SELL* and myeloid gene *CSF3R*. They also showed enrichment for the Human Gene Atlas category $CD33^+$_Myeloid, and the ARCHS4 tissue categories $CD34^+$ cell, $CD19^+$ B cells and macrophage (p-

adjusted values = 0.01183, 1.39 x 10^{-12}, 0.005638 and 0.002705 respectively, as determined by Enrichr analysis [138]).

Lastly there was a cluster of genes that showed increased expression in the LMPP and MLP (Figure 5.6). These genes included the early stem TF *HOXA9*, antigen presentation associated gene *HLA-B*, key haematopoietic signalling gene *FLT3* and lymphoid associated genes *MME*, *JCHAIN* and *IRF8*.

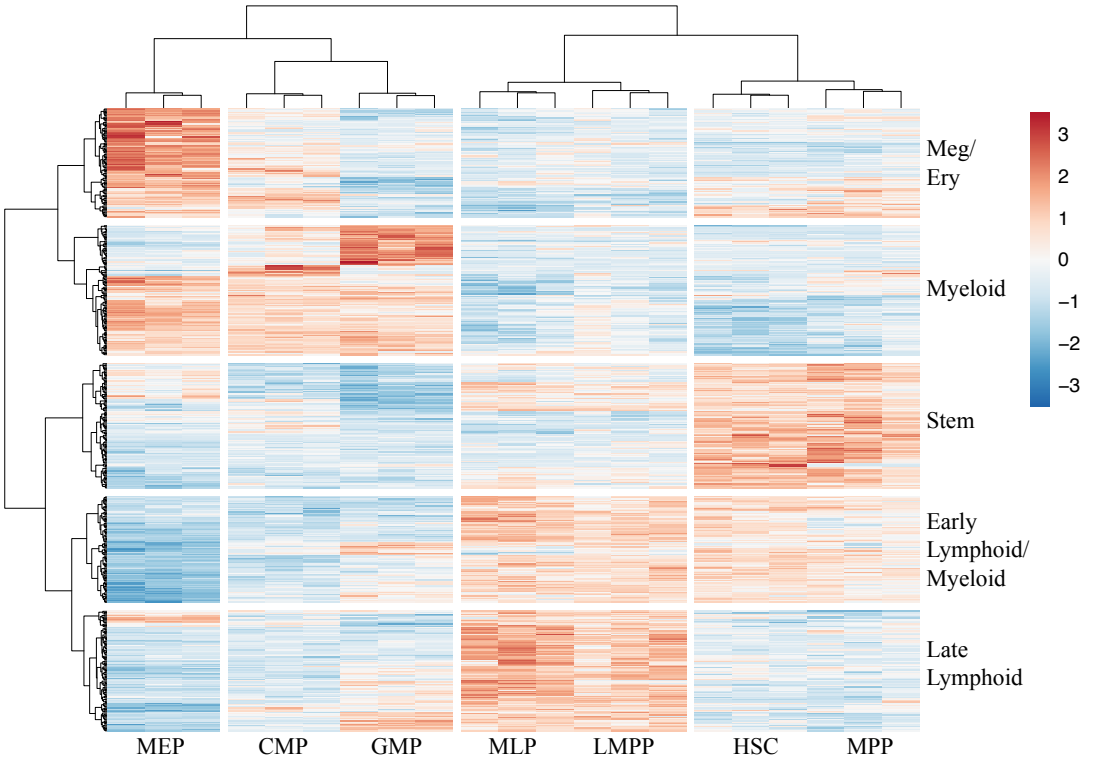


Figure 5.6: Heatmap Showing the Expression of 500 DEGs Genes by Normal BM HSPC. Heatmap showing the expression of the 500 most DEGs across normal HSPC populations, as determined by an LRT. Rows represent genes and columns represent a biological replicate of an HSPC population. Hierarchical clustering performed using the wardD2 method. Colour scale is normalised per row.

In order to cluster the normal HSPC according to their gene expression profiles, k-means clustering was used. With the aim of an unbiased approach, k-means clustering was carried out using all expressed genes to assign four clusters. The cluster identity was then displayed on the PCA plot of normal HSPCs, generated with the 500 most variant genes across normal HSPCs (Figure 5.7, see also Figure 5.4b). It can be seen that the 4 clusters as determined by k-means clustering reflected

the function and lineage output of the HSPC populations. Cluster 1 identified the earliest stem/progenitor HSC and MPP populations. Cluster 4 identified the more erythroid-megakaryocytic MEP and CMP populations. Cluster 3 identified the slightly more mature myeloid GMP, and Cluster 2 identified the more lymphoid LMPP and MLP populations (Figure 5.7).

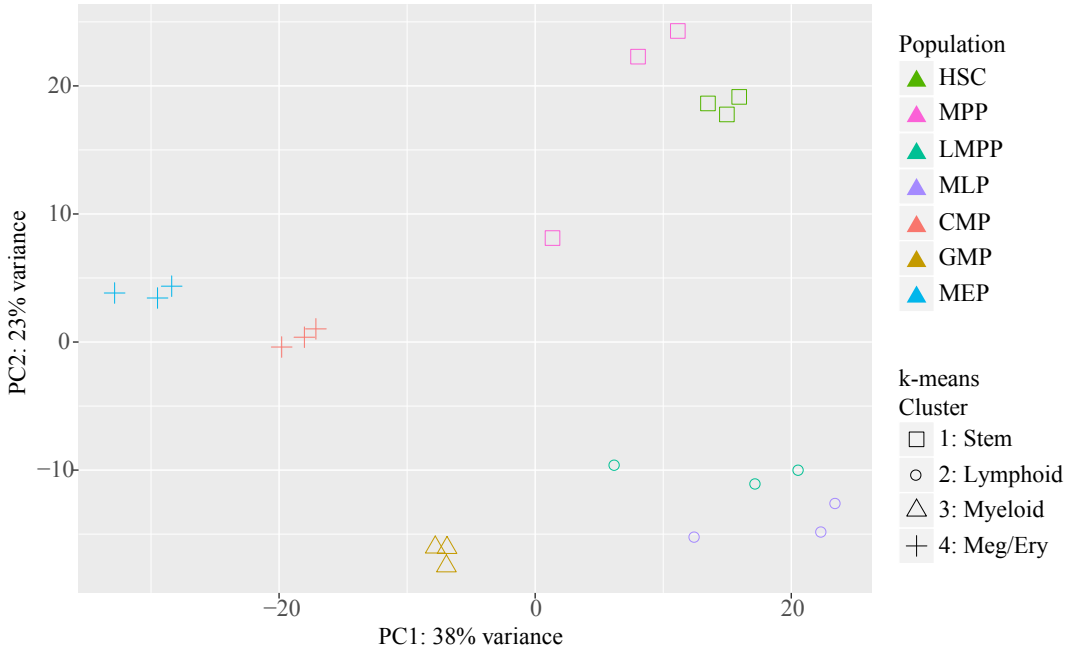


Figure 5.7: PCA of Normal BM HSPC: k-means Clustering. PCA of normal BM HSPC populations using the 500 most variant genes across normal HSPC populations. k-means clustering was carried out on normal HSPC populations using a k=4 and all expressed genes. The cluster assignment of each HSPC population is indicated by shape, as shown in the key. The percentage variance contributed by each PC is indicated. Each point represents a biological replicate.

5.3.2 Normal Stem Cell Gene Signature

Next, a gene signature representing the earliest stem and progenitor HSC and MPP populations was generated. This was done by comparing the HSC and MPP populations to the remaining progenitor populations (LMPP, MLP, CMP, GMP and MEP) using a Wald test in DESeq2. As a 'stem' signature, 1151 genes were identified to be upregulated in the HSC and MPP (p-adjusted value < 0.05) and

145 genes were upregulated with an absolute log fold-change (| logFC |) of > 2. These results are shown as a MA (minus over average) plot in Figure 5.8. Here, average normalised gene expression is shown against logFC, where a positive fold change indicates upregulation in stem populations.

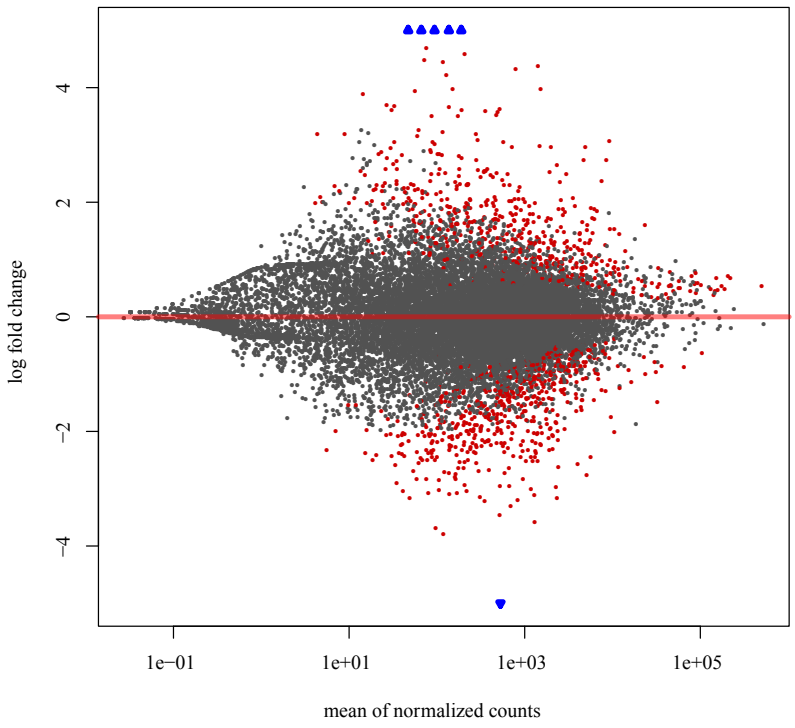


Figure 5.8: MA Plot Showing DEGs Between HSC-MPP and Progenitor Populations. DEGs were identified between early stem (HSC and MPP) and progenitor (LMPP, MLP, CMP, GMP, MEP) populations using a Wald test and shown in an MA plot [75]. A positive logFC indicates upregulation in stem populations, and a negative logFC indicates downregulation in stem populations. Mean of normalised counts represents the mean expression of each gene. Points in red represent genes with a p-adjusted value $<$ 0.05. Blue triangles indicate points which are off-scale on the plot. For gene names, logFC and p-adjusted values see A.24.

When looking at the expression of the 500 most significantly upregulated DEGs in the stem populations, it can be seen that these were able to segregate the HSC and MPP from the remaining progenitors (Figure 5.9). A clear upregulation of these genes was seen in both the HSC and MPP populations. These DEGs included key genes previously implicated in haematopoietic stem cell activity, including *ERG*,

ETV6, *HLF*, *MEIS1*, *HOXB3*, *BMI1* ([139–144], respectively). Additionally, Enrichr analysis was carried out on the significantly upregulated genes (p-adjusted value < 0.05) [138]. The 10 most significantly co-expressed TF are shown, as predicted by the ARCHS4 meta-analysis of RNA-Seq data [145]. All 10 TF have previously been associated with HSPC activity (Table 5.2).

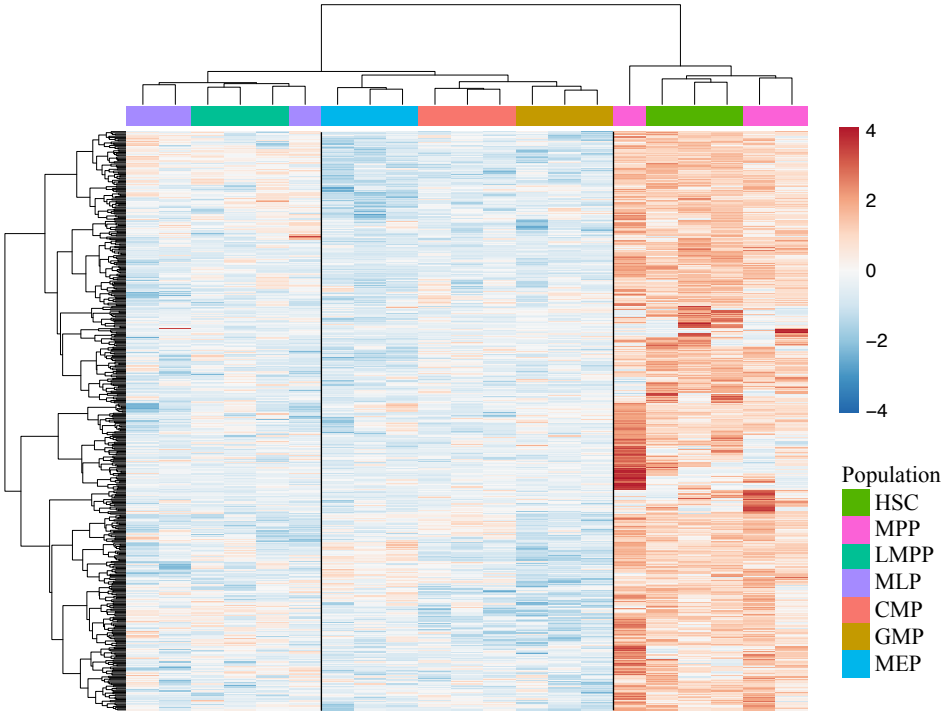


Figure 5.9: Heatmap Showing the Expression of Stem Cell-Associated Genes. DEGs were identified between early stem (HSC and MPP) and progenitor (LMPP, MLP, CMP, GMP, MEP) populations using a Wald test and the expression of the 500 most significant DEGs are shown across all HSPC. Rows represent genes and columns represent a biological replicate of an HSPC population. Hierarchical clustering performed using the wardD2 method. Colour scale is normalised per row. Populations are annotated by the colour bar above columns, as indicated by the key. For gene names, logFC and p-adjusted values see A.24.

Table 5.2: Upregulated DEGs in HSC and MPP: TF Co-Expression. DEGs were identified between early stem (HSC and MPP) and progenitor (LMPP, MLP, CMP, GMP, MEP) populations using a Wald test. Enrichr analysis was carried out on the significantly upregulated genes (p-adjusted value < 0.05) [138], and the 10 most significantly co-expressed ARCHS4 TF are shown. All 10 TF have previously been associated with HSPC activity, see cited references.

ARCHS4 Co-Expressed TF	p-adj	Reference
ERG	7.63×10^{-11}	[140]
HHEX	1.96×10^{-10}	[86]
HOXA3	3.12×10^{-09}	[146, 147]
HOXA9	4.89×10^{-08}	[147–149]
MYB	7.07×10^{-07}	[150]
TSC22D1	2.03×10^{-06}	[151]
ARNT	2.03×10^{-06}	[152]
ELK3	6.88×10^{-06}	[153]
KLF10	2.05×10^{-05}	[154]
ETV6	2.05×10^{-05}	[142]

5.4 Analysis of Leukaemic HSPC Populations

5.4.1 PCA of Leukaemic HSPC

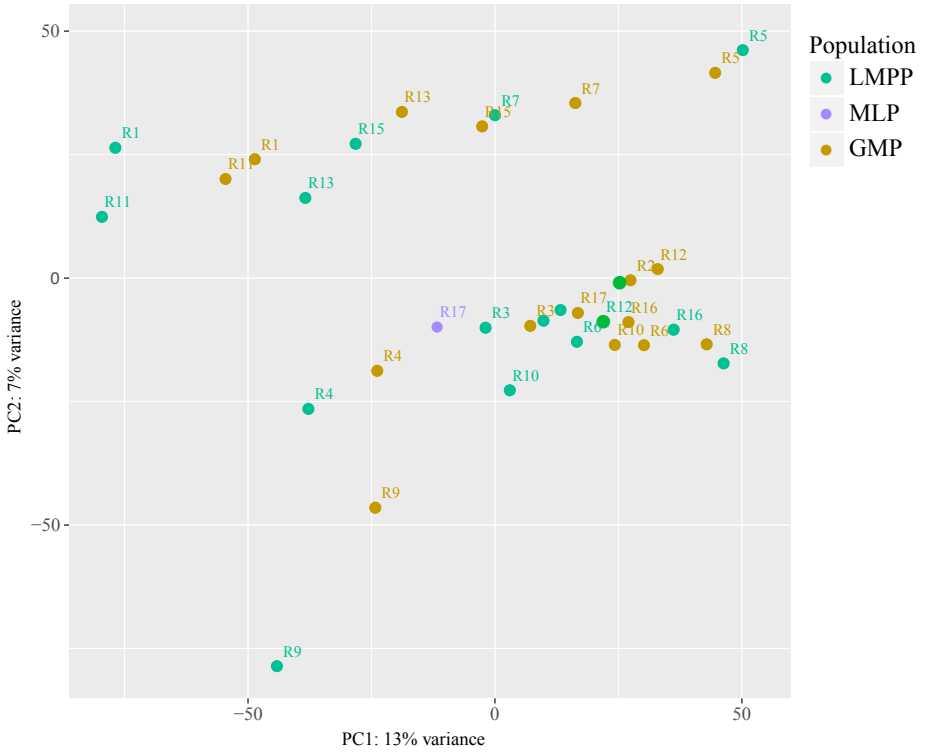
The gene expression profiles of the 17 AML patient samples were then analysed in isolation in order to investigate the variation across leukaemic samples. For this analysis, the leukaemic biological replicates are annotated as indicated in Table 5.3. When a PCA was carried out on all expressed genes, it was seen that there was no separation based on the immunophenotype of the leukaemic populations (Figure 5.10a). LSC populations from the same patient typically sat close together, with some samples from the same patient showing some minor separation on either PC1 or PC2. However, it was seen that both PC1 and PC2 represented small proportions of the variance (13% and 7% respectively) suggesting that with no gene selection, the variation was spread over many PCs (Figure 5.10a).

A PCA was then run using the 500 most variant genes across the leukaemic samples. Here it was seen that the variance attributed to PC1 and PC2 has only marginally increased (15% and 12% respectively), suggesting the variation across leukaemia is quite complex (Figure 5.10b). Here it was seen that the

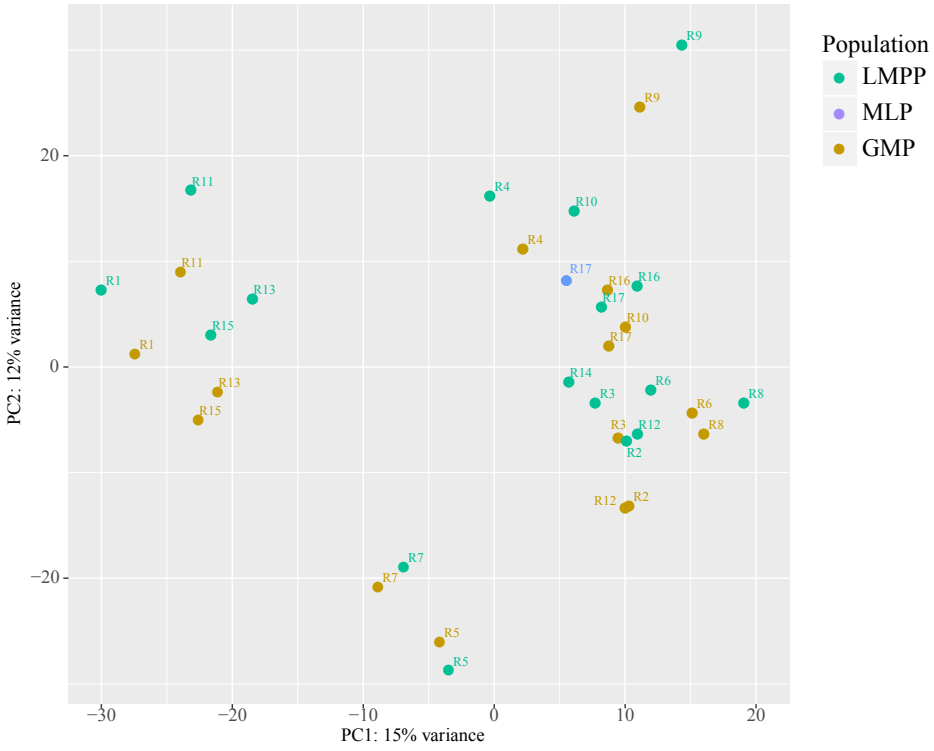
patients clustered closer together, and there was still no separation based on immunophenotype (Figure 5.10b).

Table 5.3: Leukaemic Samples Replicate Annotation. Leukaemic samples are annotated as indicated, with each patient denoted as Replicate 1 (R1), Replicate 2 (R2) etc.

Patient	Annotation
STB128	R01
HAM061	R02
BIR0517	R03
BIR01525	R04
OX2116	R05
STB291	R06
BEL039	R07
BEL048	R08
CUH018	R09
CAR0143	R10
BRI1481	R11
CAR0117	R12
HER0103	R13
GLA032	R14
BRI1371	R15
BRI1492	R16
STB257	R17



(a) All Genes



(b) 500 Variant Genes

Figure 5.10: PCA of Leukaemic HSPC. PCA of Leukaemic HSPC populations using (a) all expressed genes and (b) 500 most variable genes. The percentage variance contributed by each PC is indicated. Each point represents a biological replicate, as described in Table 5.3.

5.4.2 Engraftment and Leukaemic Gene Expression Profiles

All 17 of the AML samples used for gene expression analysis had previously been studied in xenotransplantation assays for leukaemia-propagating ability (Chapter 4). It was then investigated whether the separation of the AML samples, based on their gene expression profiles, was related to their engraftment status (Figure 5.11). When the PCA was carried out on AML samples using the 500 most variant genes and annotated with their engraftment status, it was seen that there was no separation based on engraftment (Figure 5.11). Although the engrafters R5 and R7 showed the most negative PC2 values, engrafters such as R17 and R10 had high PC2 values. And although several non-engrafters had the most negative PC1 values, both engrafters and non-engrafters had positive PC1 values (Figure 5.11). Therefore, it is likely that the engraftment of AML samples in this case was not related to their intrinsic gene expression profiles, but more likely due to the limitations of the xenotransplantation assay (see Chapter 4).

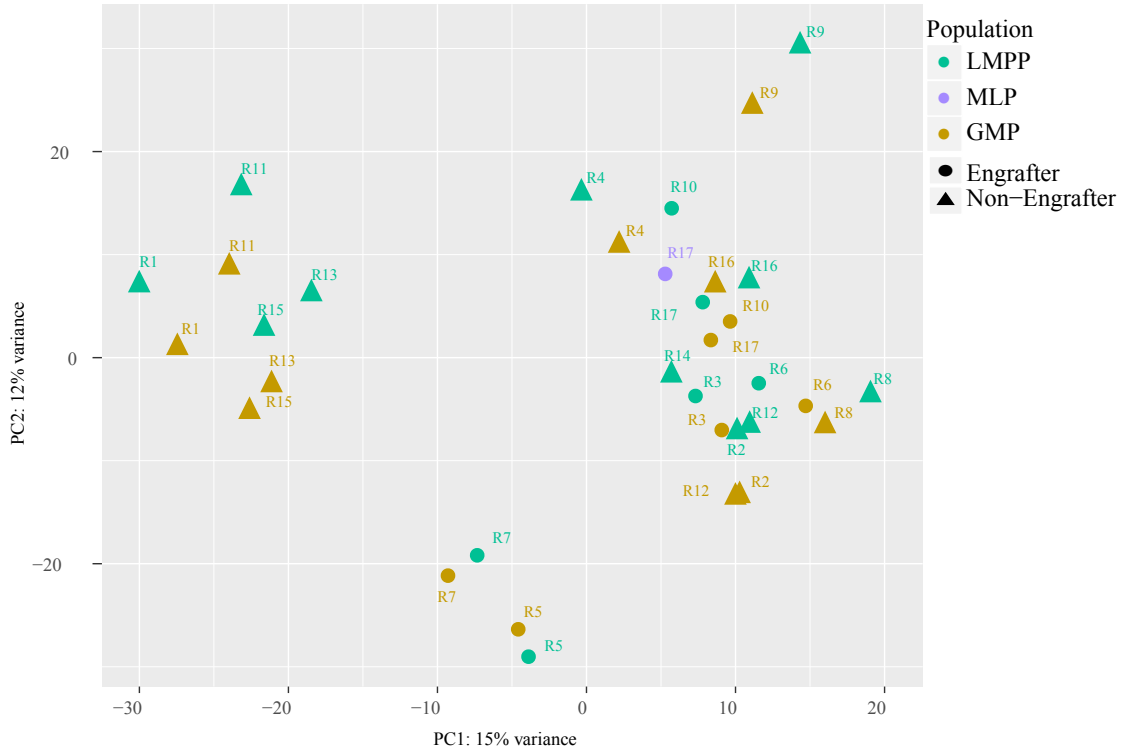


Figure 5.11: PCA of Leukaemic HSPC Annotated by Engraftment Status. PCA of Leukaemic HSPC populations using the 500 most variable genes. The percentage variance contributed by each PC is indicated. Each point represents a biological replicate, as described in Table 5.3. Engraftment status is indicated by shape, as described in Chapter 4.

5.4.3 Genotype and Leukaemic Gene Expression Profiles

Next it was investigated by PCA whether the genotype of the AML samples may be driving the samples' separation. All 17 AML samples had previously been genotyped (Chapter 4), and Figure 5.12 shows the PCA using the 500 most variant genes across AML samples now annotated by genotype. The annotation indicates the genes in which mutations were detected for an AML sample (Figure 5.12). As with engraftment, there was no clear separation of AML samples by genotype (Figure 5.12). Although it was seen that the samples with the lowest values on PC2 have mutations in the *TET2* gene, the samples with the highest values on PC2 also had mutations in *TET2*. Similarly, across PC1, *FLT3*, *RUNX1*, *U2AF1* and *JAK2* mutants were seen across this axis.

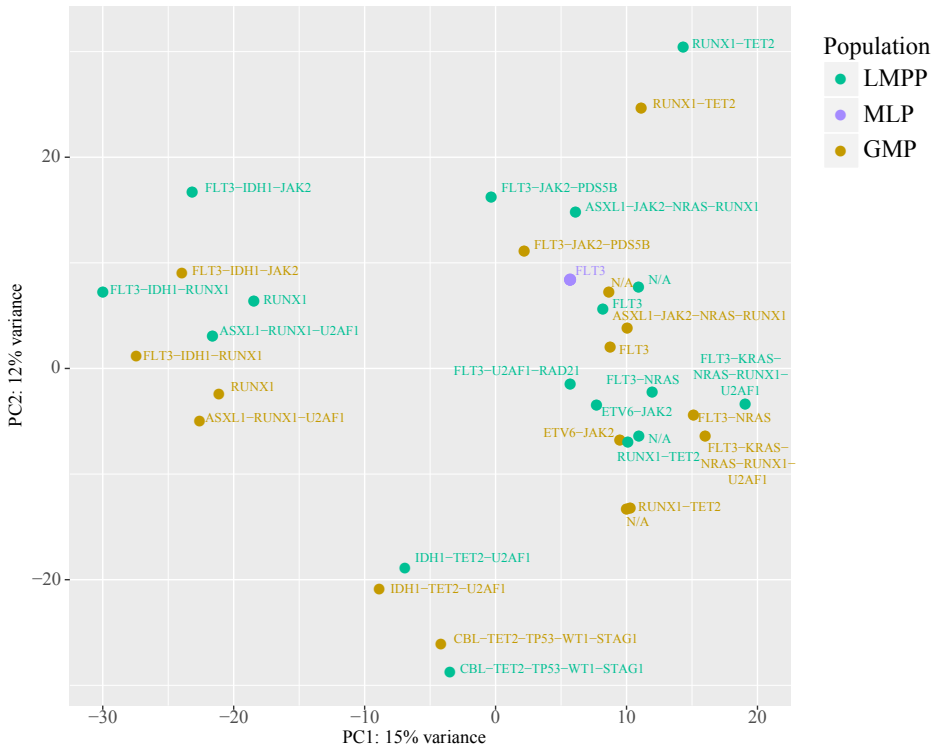


Figure 5.12: PCA of Leukaemic HSPC Annotated by Genotype. PCA of Leukaemic HSPC populations using the 500 most variable genes. The percentage variance contributed by each PC is indicated. Each point represents a biological replicate, as described in Table 5.3. Genotype is annotated, as described in Chapter 4. N/A: Not applicable i.e. no mutations were detected in this sample.

5.4.4 Top Variant Genes Across Leukaemia

In order to gain further insight into the factors driving the variation across these leukaemic samples, the most variable genes across the AML samples were identified (Figure 5.13). Figure 5.13 shows the 10 most variable genes across the 17 AML samples, with the genotype, biological replicate and immunophenotype annotated.

The 2 most variable genes *DDX3Y* and *RPS4Y1* are both located on the Y chromosome, with *DDX3Y* encoding an RNA helicase and *RPS4Y1* a ribosomal protein (Figures 5.13a, 5.13b). This suggests that the gender of these patients underlay some of the variation seen in gene expression and is further investigated below (Figure 5.14).

The next most variable gene identified was the immunoglobulin gene *JCHAIN*, typically associated with B cell differentiation (Figure 5.13c). As may have been

expected, the immunophenotypic MLP population, which is primarily a B cell progenitor (Chapter 3) showed high expression of this gene. However, in this patient sample, both the LMPP and more myeloid GMP also showed high expression of this gene. Across most samples, the LMPP and GMP showed similar levels of *JCHAIN* expression; however, there were some cases where the LMPP showed elevated *JCHAIN* expression compared to the GMP (for example R8, Figure 5.13c).

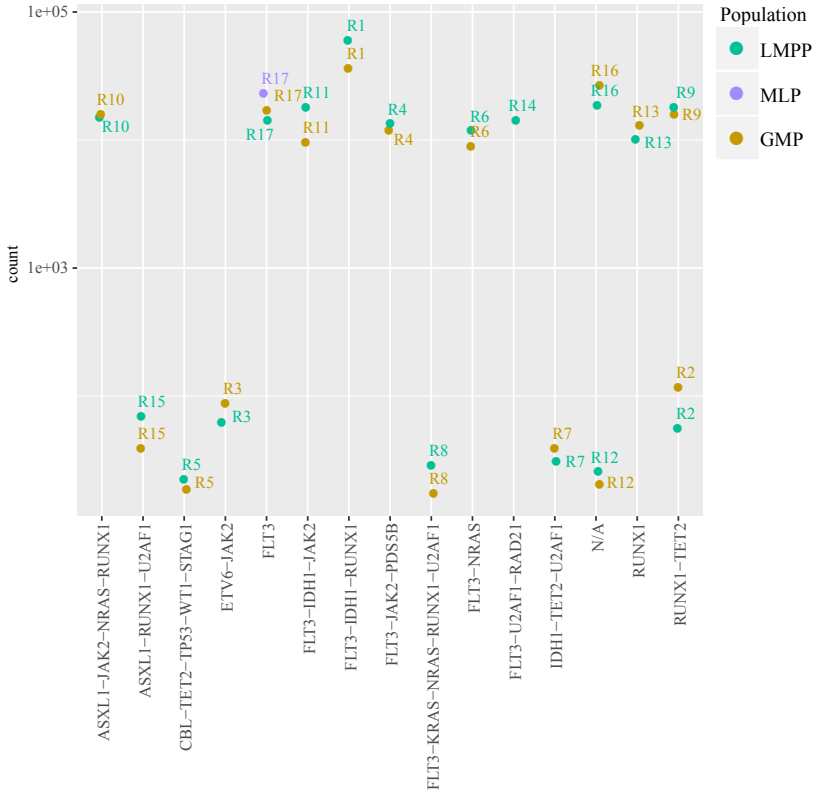
The fourth most variable gene, *CPA3* is a known marker of mast cells and basophils [155]. It has been previously suggested that *CPA3* may be over-expressed specifically in AML mutated in *FLT3* or *NPM1* [115]. However, this was not seen in this dataset, where samples with a *FLT3* mutation had variable *CPA3* levels, and no samples were mutated in *NPM1*.

Next in this set of variable genes was the lymphoid-associated gene *DNTT*, which is a DNA polymerase involved in the development of B and T cells [156]. On the other hand, the myeloid-associated gene *MPO*, which encodes the granular enzyme myeloperoxidase, was also identified as being highly variable (Figure 5.13f). The gene *F13A1* is known to show elevated expression in sub-sets of macrophages [157], and was seen to be variable across the AML samples.

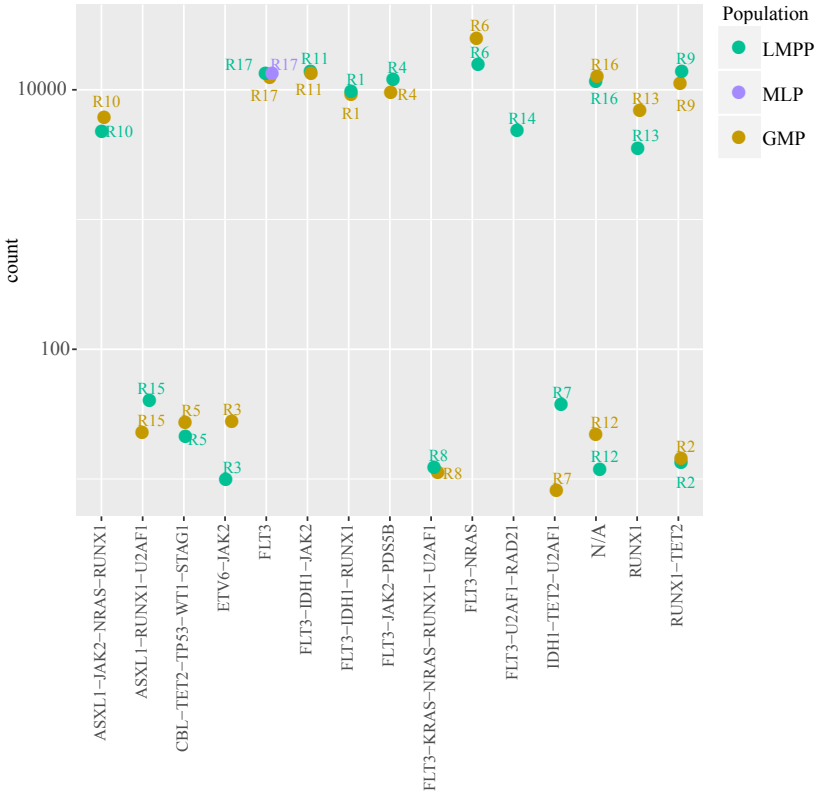
A highly variable gene of particular interest was *MARCKS* (Figure 5.13h). *MARCKS* encodes a substrate of protein kinase C (PKC) and has been implicated in several cellular functions including cell adhesion, cell growth and apoptosis [158]. Increased PKC signalling is associated with several types of cancer and the association of MARKCS with PKC is suggested to be involved in drug resistance in several cancers, including multiple myeloma [158].

Similarly, another highly variable gene previously identified to play a role in cancer was *S100A16* (Figure 5.13i). S100A16 is a calcium binding protein and S100 family members in general play a role in many processes, including cell growth, differentiation and motility [159]. S100A16 expression has been identified in several solid tumours, including breast and colon cancer, and is associated with a poor prognosis [159, 160].

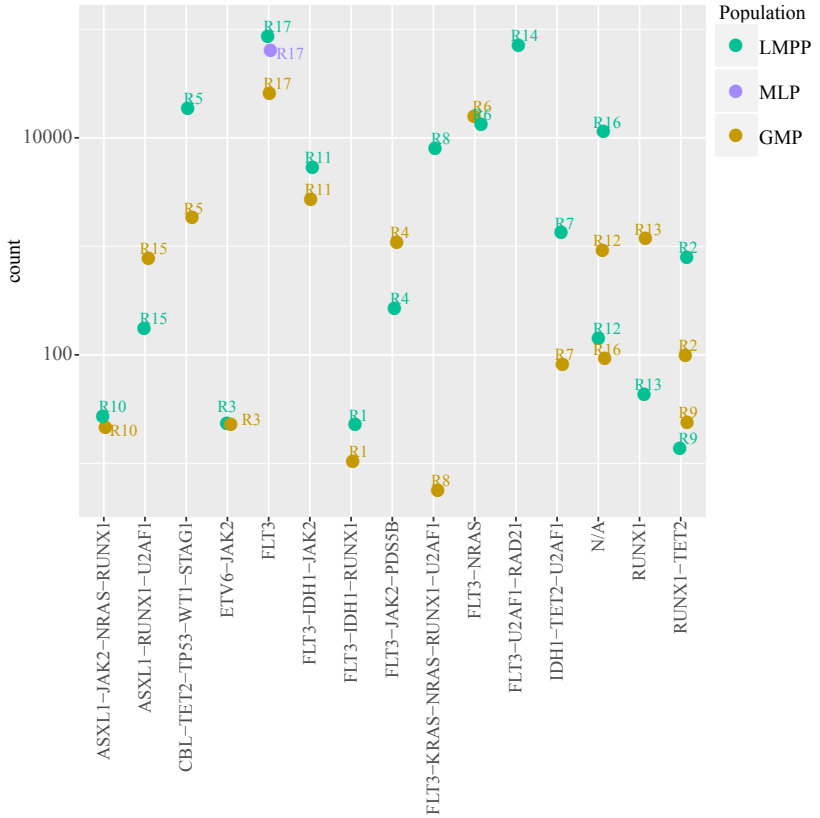
Finally, *RXFP1* was identified to be variable across the AML samples (Figure 5.13j). *RXFP1* encodes for a G protein-coupled receptor for the protein relaxin, which plays a role in several processes, including apoptosis, inflammation and angiogenesis [161]. It has been identified to be potentially involved in the progression of solid cancers, including prostate cancer, and more recently was identified as being differentially upregulated in chronic myeloid leukaemia (CML) LSCs when compared to HSCs [161].



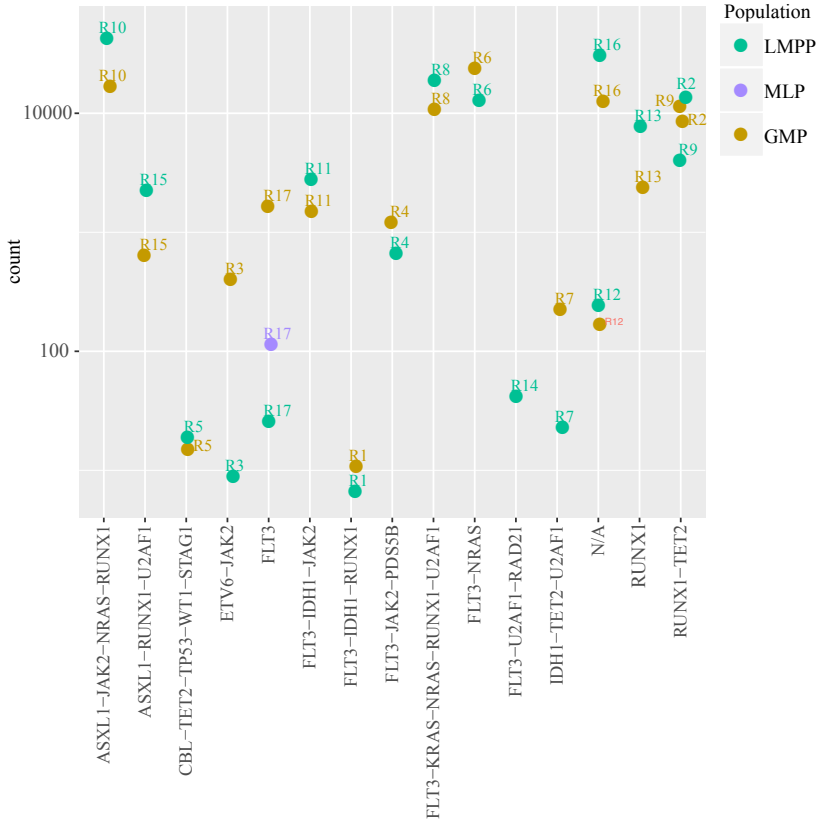
(a) DDX3Y



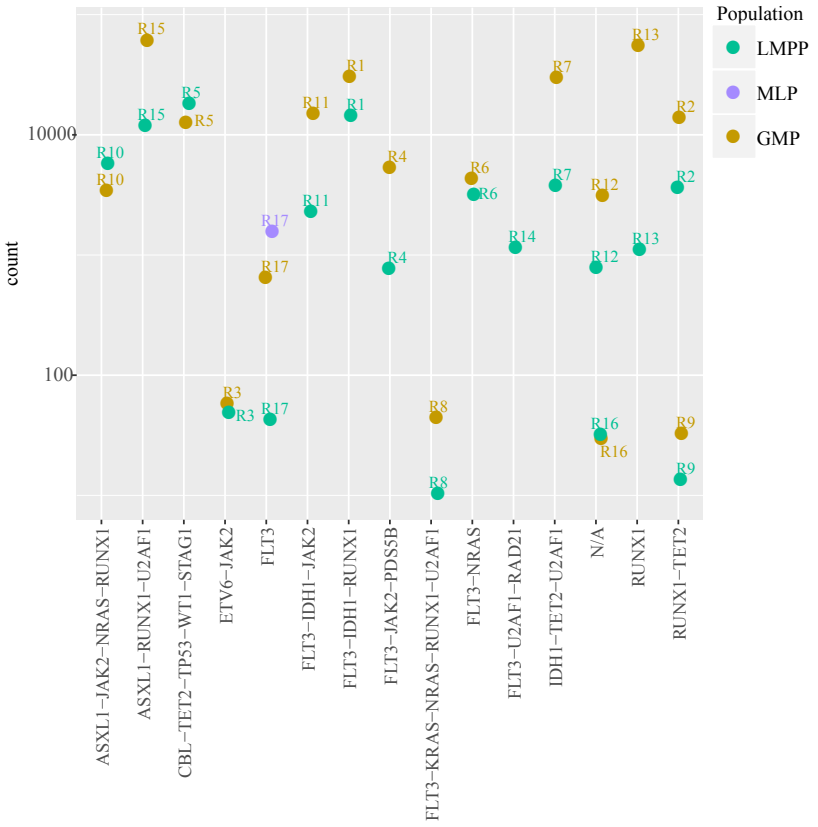
(b) RPS4Y1



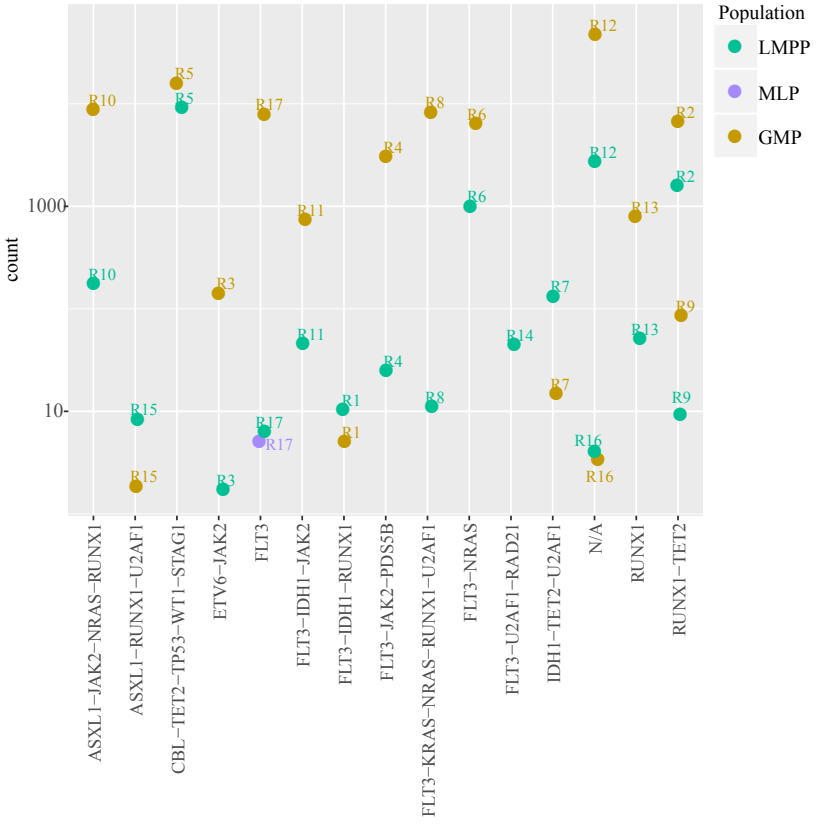
(c) JCHAIN



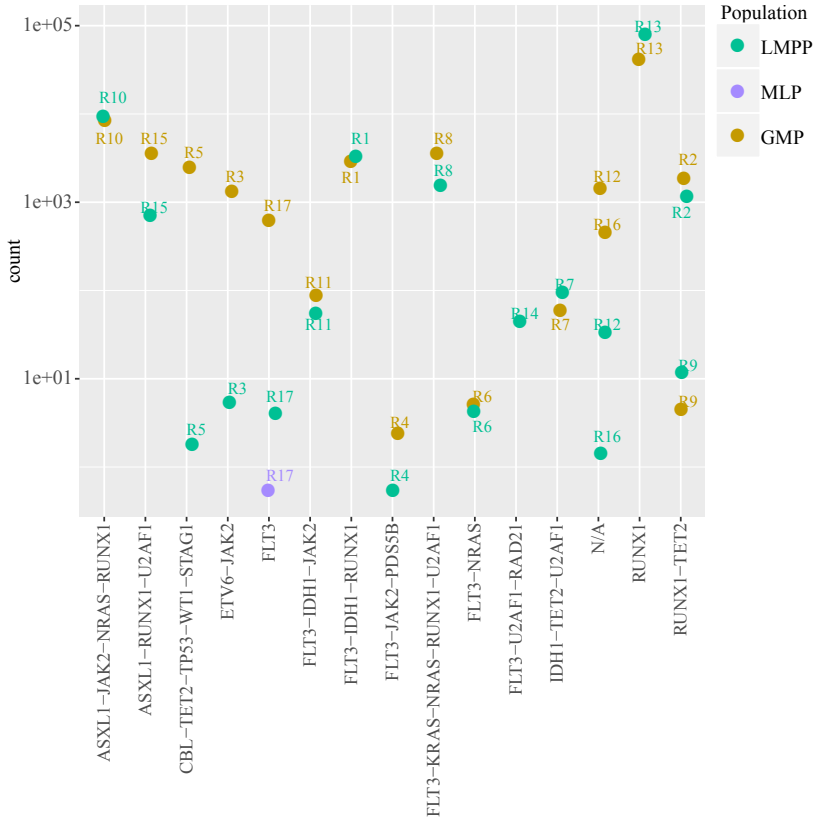
(d) CPA3


Population
LMPP
MLP
GMP
count
10000
100
ASXL1−JAK2−NRAS−RUNX1
ASXL1−RUNX1−U2AF1
CBL−TET2−TP53−WT1−STAG1
ETV6−JAK2
FLT3
FLT3−IDH1−JAK2
FLT3−IDH1−RUNX1
FLT3−JAK2−PDS5B
FLT3−KRAS−NRAS−RUNX1−U2AF1
FLT3−NRAS
FLT3−U2AF1−RAD21
IDH1−TET2−U2AF1
N/A
RUNX1
RUNX1−TET2

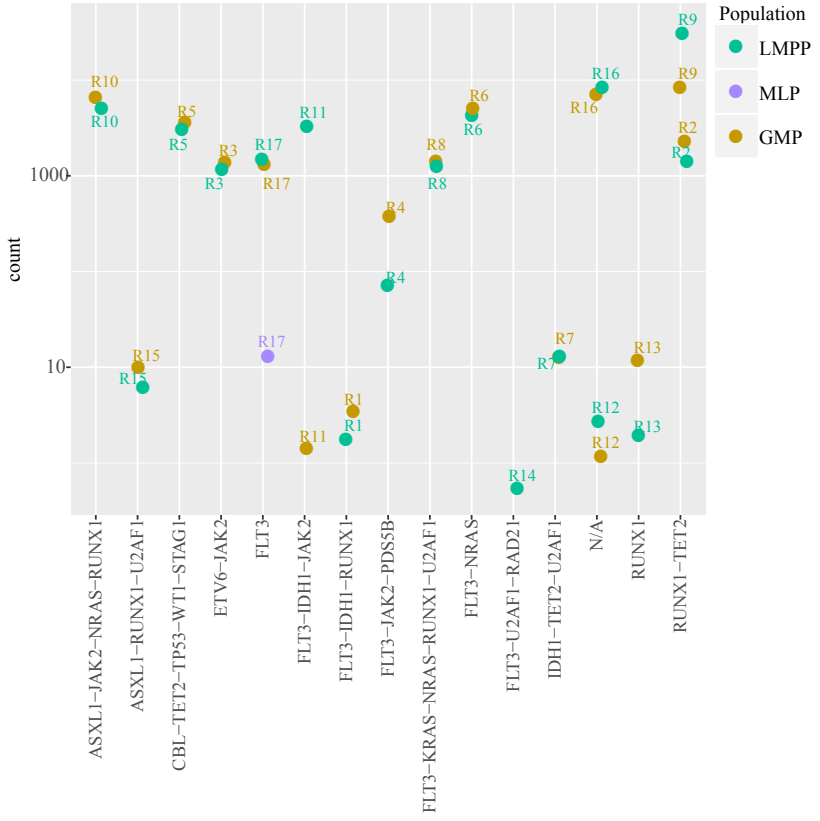
(e) DNTT


Population
LMPP
MLP
GMP
count
1000
10
ASXL1−JAK2−NRAS−RUNX1
ASXL1−RUNX1−U2AF1
CBL−TET2−TP53−WT1−STAG1
ETV6−JAK2
FLT3
FLT3−IDH1−JAK2
FLT3−IDH1−RUNX1
FLT3−JAK2−PDS5B
FLT3−KRAS−NRAS−RUNX1−U2AF1
FLT3−NRAS
FLT3−U2AF1−RAD21
IDH1−TET2−U2AF1
N/A
RUNX1
RUNX1−TET2

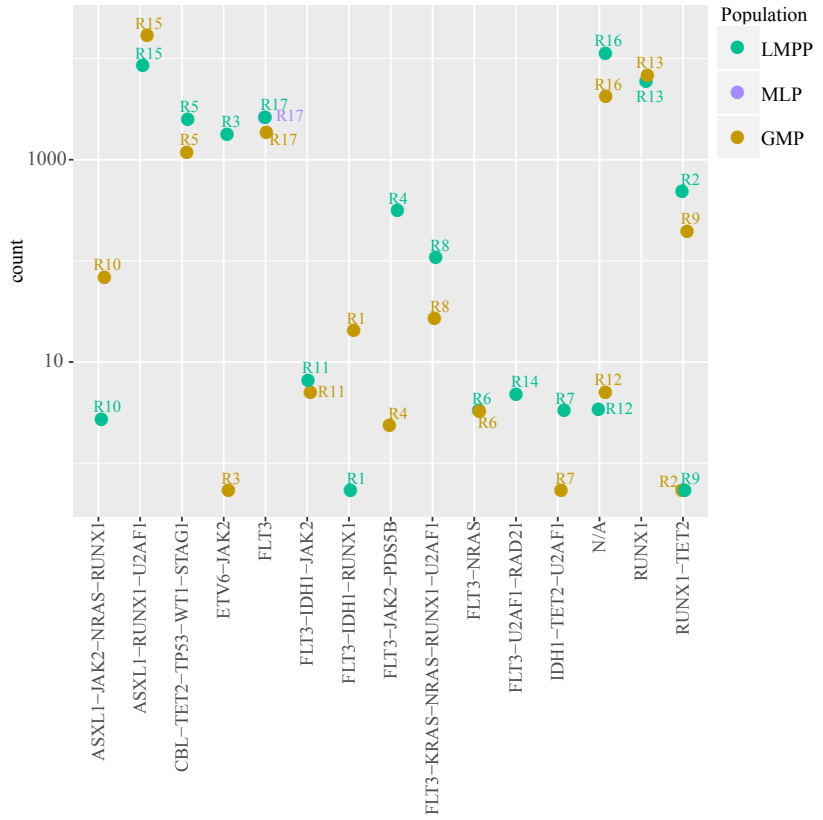
(f) MPO



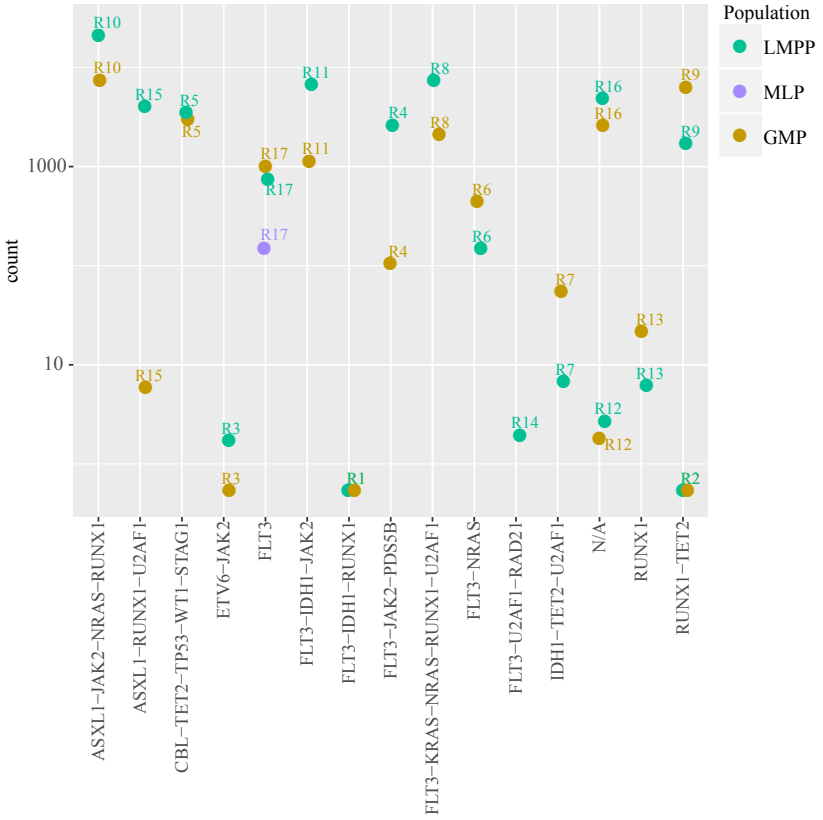
(g) F13A1



(h) MARCKS



(i) S100A16



(j) RXFP1

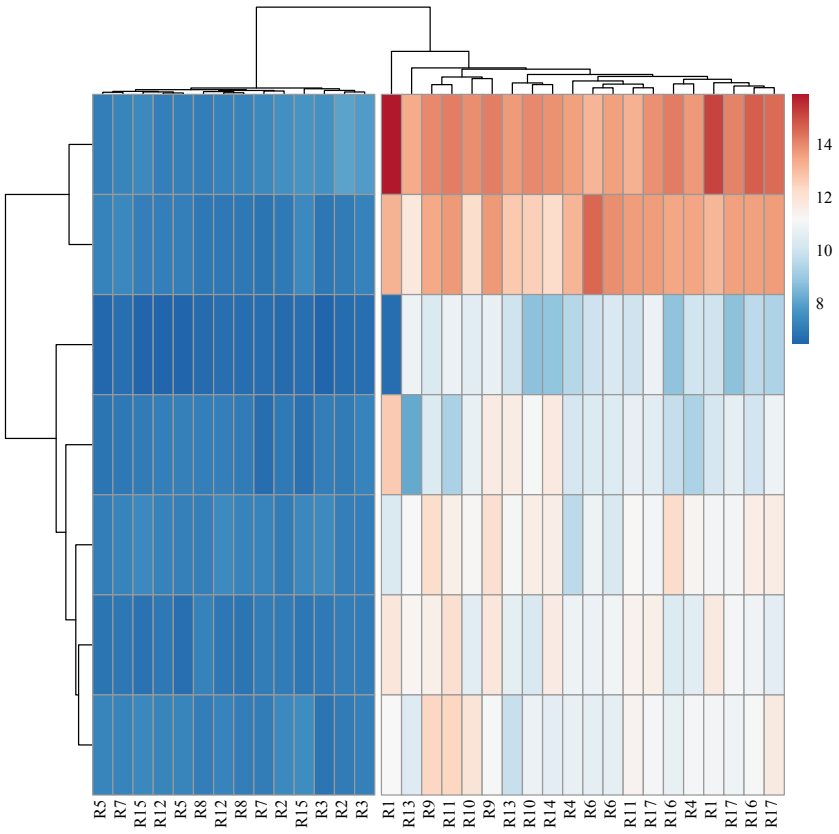
Figure 5.13: Variant Genes Across Leukaemia. Expression of the 10 most variant genes across all leukaemic HSPC populations, shown as DESeq2 normalised counts. Mutated genes in each sample are indicated below the plot and patient replicate is annotated. Colour shows population as indicated by key. Each point represents a biological replicate, as described in Table 5.3.

5.4.5 Gender and Leukaemic Gene Expression Profiles

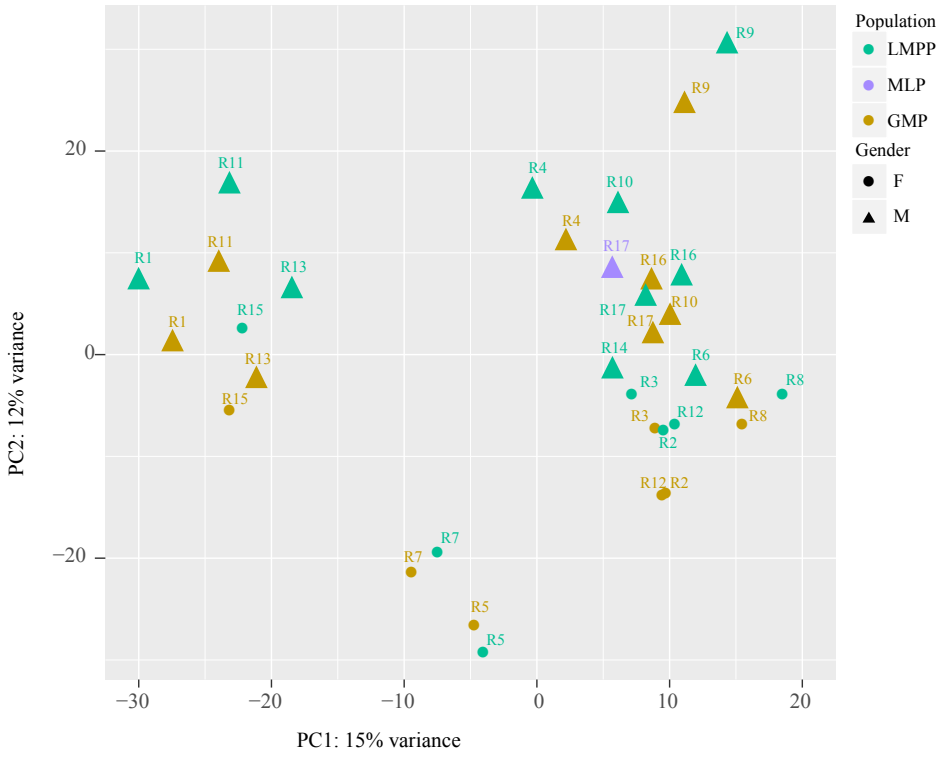
The two most variable genes across AML samples were both Y-linked genes (Figures 5.13a, 5.13b), implying that the gender of the AML patients may be driving some of the variability seen in gene expression. When looking at all Y-linked genes that showed a robust expression across this dataset, a clear separation in samples was seen (Figure 5.14a). It was deduced that the biological replicates in the cluster that showed increased expression of Y-linked genes were from male patients, and the remainder were female patients.

When looking at the PCA carried out on AML samples using the 500 most variable genes and then labelling samples by gender (as identified by Figure 5.14a), it was seen that gender did separate out the patient samples (Figure 5.14b). It was seen that the majority of the samples with a positive PC2 value were identified as male, and that the samples with more negative PC2 values were female. This was with the exception of the female LMPP R15 sample, which showed a PC2 value equivalent to several male samples (Figure 5.14b).

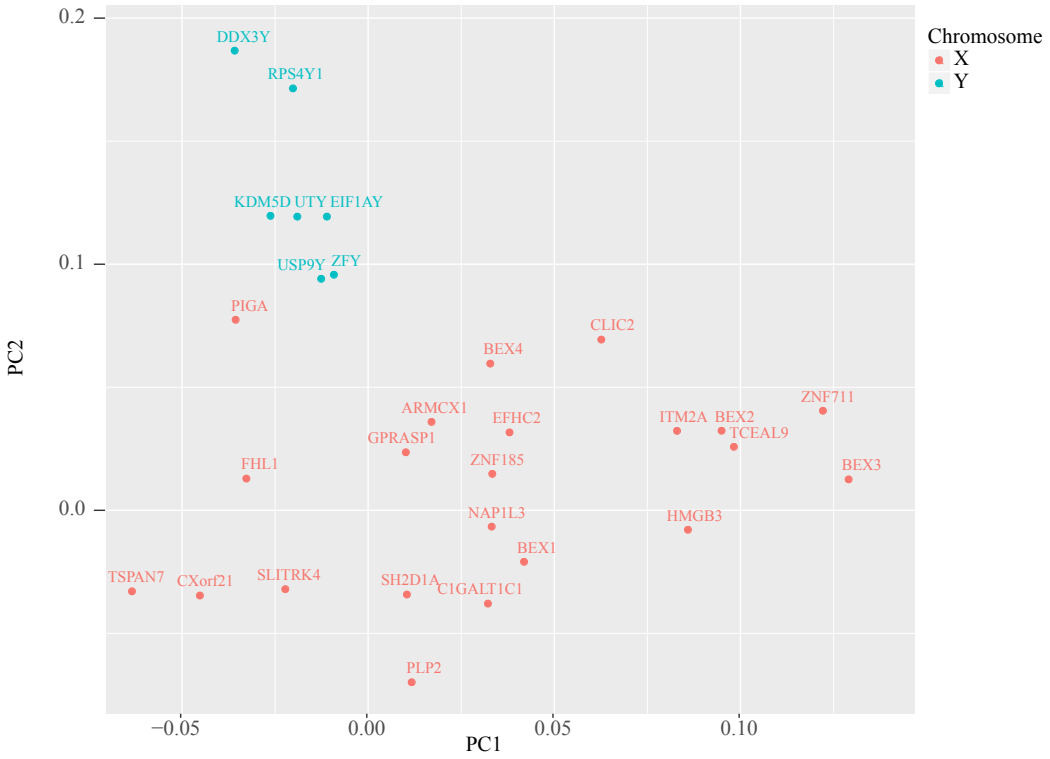
The separation of leukaemic samples based on gender was confirmed when the loadings of the PCA were investigated. The loading value for each gene is an indication of its contribution to each PC. When the loading values for X- and Y-linked genes were plotted, a clear separation of these genes was seen along the PC2 axis (Figure 5.14c).



(a) Heatmap of Y-Linked Genes



(b) Leukaemic HSPC: Gender



(c) Loading Plot of X- and Y-Linked Genes

Figure 5.14: Leukaemic HSPC Separated by Gender. (a) Y-linked genes with robust expression in HSPCs were identified and their expression in leukaemic HSPC shown. Rows represent genes and columns represent a biological replicate of an HSPC population (labelled below plot). Colour scale is normalised per row. (b) PCA of leukaemic HSPC populations using 500 most variable genes. The percentage variance contributed by each PC is indicated. Gender is shown by shape, as indicated by the key. Each point represents a biological replicate, as described in Table 5.3. (c) Loading plot showing the loading values of X- and Y-linked genes from the PCA shown in (b). M: male, F: female.

5.5 Analysis of Normal and Leukaemic HSPC Populations

5.5.1 Hierarchical Clustering and PCA of Normal and Leukaemic HSPC

In order to investigate the relationship between normal and leukaemic HSPC, the leukaemic gene expression profiles were analysed in the context of normal gene expression. First, as an unbiased clustering approach, hierarchical clustering was carried out on all leukaemic and normal HSPC using all expressed genes (Figure 5.15). It was seen that all the leukaemic populations clustered separately to the

normal HSPC, confirming their dysregulated gene expression profiles compared to normal haematopoiesis. Within the normal HSPC, most normal populations were seen to cluster together, with clear clusters for the CMP, GMP and MEP populations. The more stem populations (HSC and MPP) and lymphoid populations (LMPP and MLP) showed some intermingling in their clustering (Figure 5.15). Within the leukaemic populations, however, clustering was not based on immunophenotype. Instead, clustering was according to patient sample, as indicated by the numbering below the dendrogram (as described in Table 5.3). All populations from the same patient clustered together, with the exception of R9, whose LMPP- and GMP-like populations clustered separately.

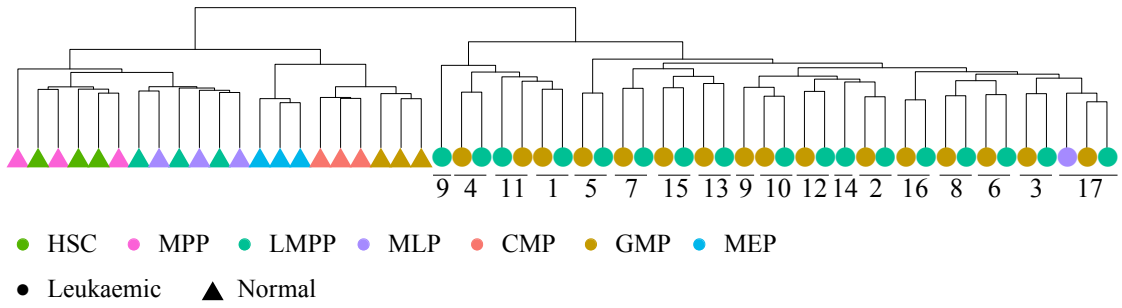


Figure 5.15: Hierarchical Clustering of Normal and Leukaemic HSPCs. Hierarchical clustering of normal and leukaemic HSPCs using all expressed genes. Populations and normal/leukaemic status are as indicated by the key. Each point represents a biological replicate, as described in Table 5.3.

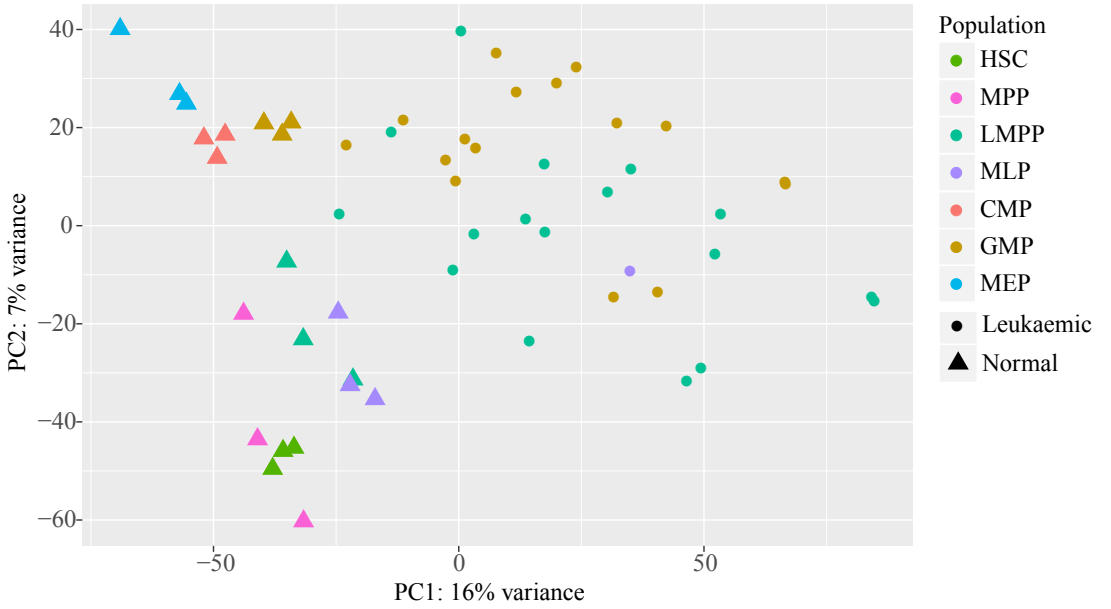
Following on from this, another unbiased approach was taken, to study where the leukaemic profiles sat relative to normal populations, and a PCA was carried out using all expressed genes (Figure 5.16a). It was seen that PC1 clearly separated out the normal from the leukaemic populations, unsurprisingly confirming this was the largest source of variation in the dataset. PC2 was seen to separate out the normal populations, with the earliest HSC and MPP populations having the most negative PC2 values, and the MEP the most positive PC2 values (Figure 5.16a). The myeloid GMP was seen to have higher PC2 values when compared to the lymphoid LMPP and MLP. When looking at the leukaemic populations,

the GMP-like populations had significantly higher PC2 values than the LMPP-like populations (Figure 5.16a, Student's *t*-test: P= 0.00382).

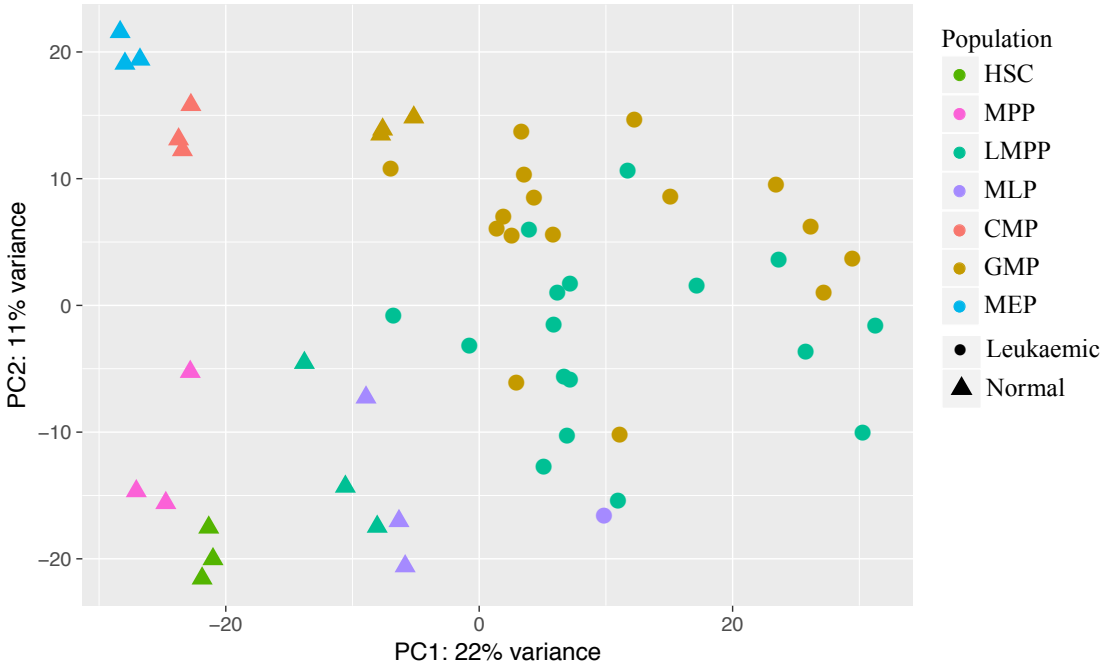
To focus on the variation across normal and leukaemic HSPCs, the 500 most variant genes across normal and leukaemic HSPCs were then used for a PCA and are shown in Figure 5.16b. Similar to the PCA using all expressed genes, PC1 here clearly separated the normal from leukaemic HSPC. In addition to this, PC1 captured more of the variance (22% compared to 16%). Using a more refined selection of genes, there was clearer separation of the normal populations on PC2, again with the HSC and MPP showing the lowest values on PC2 and the MEP the highest (Figure 5.16b). Here the LMPP and MLP showed similar PC2 values to the HSC and MPP, and the GMP showed similar values to the CMP. The leukaemic populations spanned the PC2 axis and as with the PCA with all genes, there was a trend of the LMPP-like populations to have lower PC2 values than the GMP-like populations.

It was clear that the main source of variation in this dataset was the normal or leukaemic origin of samples. Therefore, to focus on the normal gene expression differences seen across normal HSPCs, a PCA was carried out on normal and leukaemic HSPCs using the 500 most significantly DEGs across normal HSPCs (Figure 5.16c, see also Figures 5.4c, 5.6 and Table A.23). As was seen with only normal HSPCs (Figure 5.4c), PC1 separated out the more lymphoid LMPP and MLP from the MEP (Figure 5.16c). There was still some separation of the stem HSC and MPP populations by PC2. Furthermore, the LSC populations sat closer to the progenitors, rather than the HSC populations (Figure 5.16c), as previously suggested [22]. The one immunophenotypic MLP-like LSC was seen to sit closer towards the more lymphoid normal LMPP and MLP populations. The LMPP- and GMP-like LSC populations sat mostly amongst the normal GMP populations, but with a trend of the LMPP-like populations to have slightly higher PC2 values, moving them closer to the normal LMPPs (Figure 5.16c).

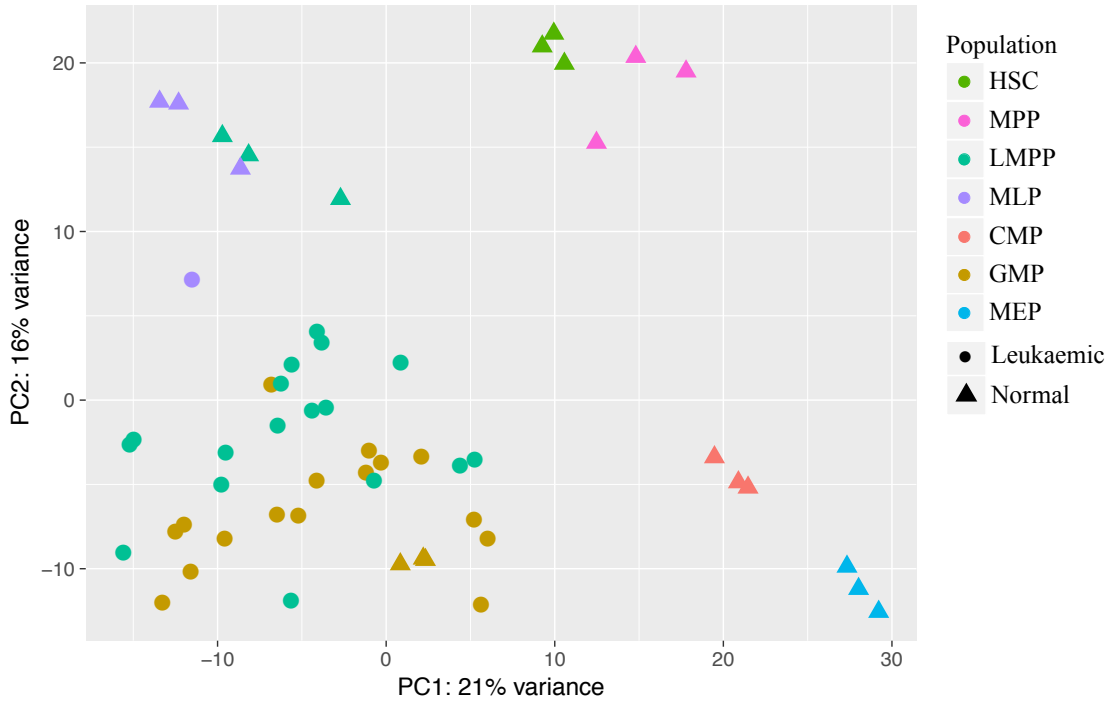
It was seen with normal HSPCs that when a PCA was carried out using the 500 most significantly DEGs across normal HSPCs that PC3 represented an axis of myeloid differentiation (Figure 5.5). Therefore, it was investigated whether



(a) All Genes



(b) 500 Variant Genes



(c) 500 LRT Genes

Figure 5.16: PCA of Normal and Leukaemic HSPC. PCA of Normal and Leukaemic HSPC populations using (a) all expressed genes (b) 500 most variable genes across normal and leukaemic HSPCs and (c) 500 most differentially expressed genes across normal HSPC populations, as determined by an LRT (see Figures 5.4, 5.6 and Table A.23). The percentage variance contributed by each PC is indicated. Each point represents a biological replicate.

this held true when a PCA using the same genes, but on normal and leukaemic HSPC, was carried out. It was seen that in this case, PC3 did not separate out the normal GMP and therefore did not represent an axis of myeloid differentiation (Figure 5.17). All three of the LM populations showed similar high PC3 values, with the HSC and MPP having the lowest values. Here this may represent a combined axis of LM differentiation.

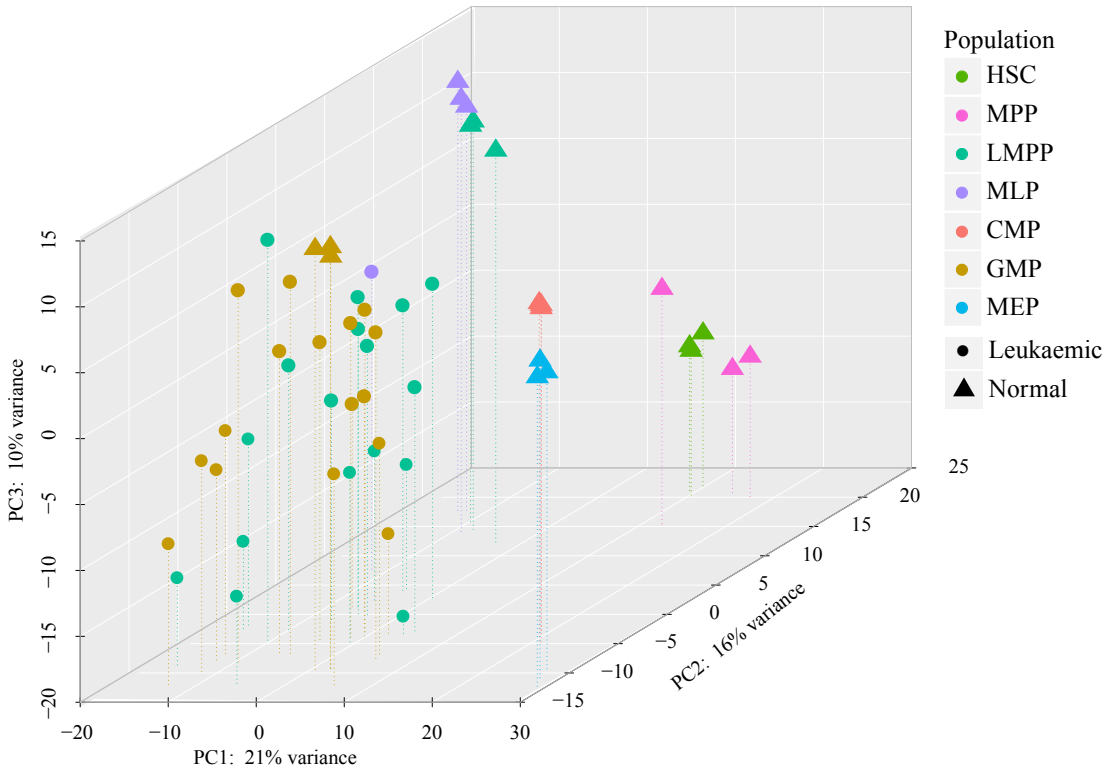


Figure 5.17: PCA of Normal and Leukaemic HSPC: PC1–PC3. PCA of normal and leukaemic HSPC populations using the 500 most DEGs across normal HSPC populations, as determined by an LRT (see Figures 5.4, 5.6 and Table A.23). The percentage variance contributed by each PC is indicated and PCs 1-3 are shown. Each point represents a biological replicate.

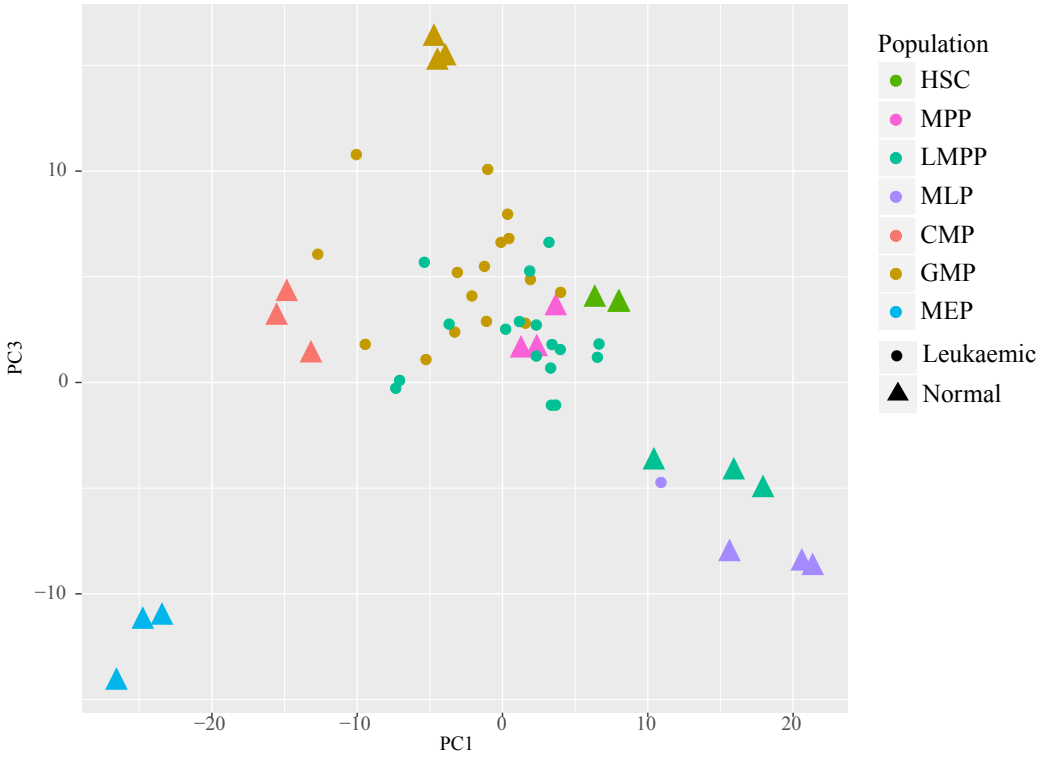
5.5.2 PCA Projections of Leukaemic HSPC

To investigate where the LSC populations sat on a myeloid axis of differentiation, PCA projections were used. Here the PCA was run using the 500 most significantly DEGs across normal HSPC, and the PCA plot generated using only the normal HSPC populations. Using this PCA, the position of the leukaemic samples was predicted, and the leukaemic samples projected onto the original PCA, an approach previously used by [132].

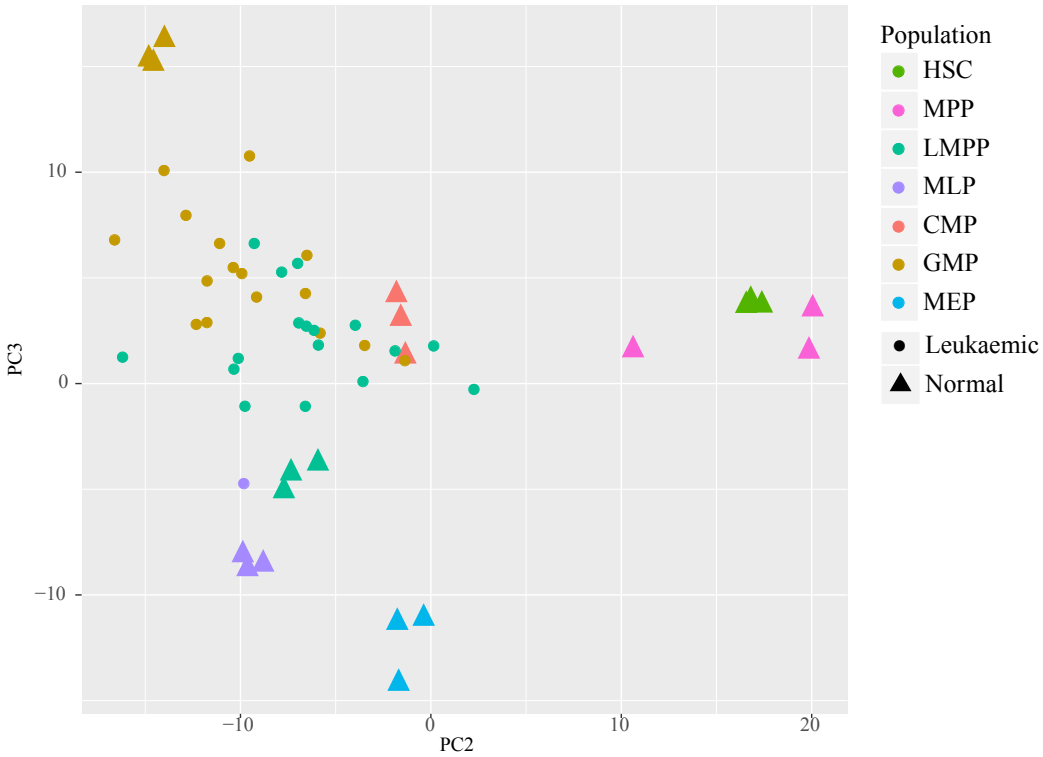
PC1 and PC3 of this approach are shown in Figure 5.18a. As in the normal-only PCA, PC1 separated out the lymphoid from the megakaryocytic-erythroid populations, and PC3 pulled out the myeloid GMPs. The HSCs, MPPs and more immature and heterogenous CMPs sat in the middle of this plot (Figure 5.18a). The one immunophenotypic MLP-like LSC was seen to sit closer to the normal

LMPPs and MLPs. However, the remaining LMPP- and GMP-like LSC populations sat amongst the HSC, MPP and CMP populations, indicating their differentiation arrest, characteristic of the disease (Figure 5.18a).

When looking at PC2 and PC3, the progenitor-like gene expression profile of the LSCs was analysed (Figure 5.18b). On the PC2 axis the LSCs sat closer to the progenitors than the normal HSCs and MPPs. PC3 represented the same myeloid axis as shown in Figure 5.18a. It was seen that the leukaemic populations spanned the distance below the GMPs to the normal CMPs. There was a trend for the LMPP-like LSC populations to have lower PC3 values and the GMP-like LSC populations to have higher PC3 values (Figure 5.18b).



(a) PC1 and PC3



(b) PC2 and PC3

Figure 5.18: PCA Projection of Normal and Leukaemic HSPC. PCA of Normal and Leukaemic HSPC populations using 500 most DEGs across normal HSPC populations, as determined by an LRT (see Figures 5.4, 5.6 and Table A.23). The PCA was run on normal HSPC and the position of the leukaemic HSPC projected onto this PCA. Either (a) PC1 and PC3 or (b) PC2 and PC3 are shown. Each point represents a biological replicate.

5.5.3 Expression of Normal DEGs by Leukaemic HSPC

DEGs were identified across the normal HSPCs, and these genes showed a clear pattern of expression across the normal populations. This pattern could be related to the HSPCs' function and lineage affiliation (Figure 5.6). However, when looking at the expression of these genes by leukaemic samples, it was seen that they no longer resembled normal HSPCs (Figure 5.19). The clear patterns of expression of these genes were still seen by the normal HSPCs; however, the leukaemic samples showed a dysregulated pattern of expression. Interestingly, here the leukaemic samples were clustered most closely with the normal LMPP and MLP populations. This most likely reflected their combined early LM gene expression profile. On the other hand, CMP, GMP and MEP populations showed distinct gene expression that clustered away from leukaemic samples (Figure 5.19).

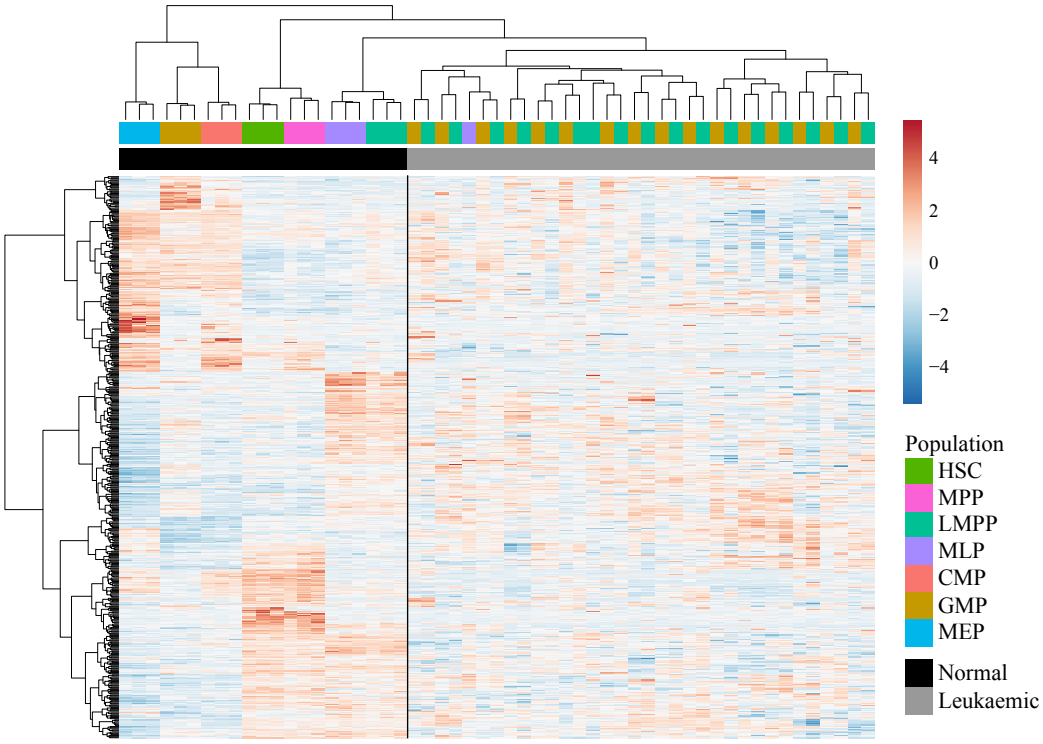


Figure 5.19: Heatmap Showing the Expression of Normal DEGs by Normal and Leukaemic HSPC. Heatmap showing the expression of the 500 most DEGs across normal HSPC populations, as determined by an LRT (see Figures 5.4, 5.6 and Table A.23) by normal and leukaemic HSPCs. Rows represent genes and columns represent a biological replicate of an HSPC population. Hierarchical clustering performed using the wardD2 method. Colour scale is normalised per row. Populations and normal/leukaemic status are annotated by the colour bar above columns, as indicated by the key.

5.5.4 DEGs Between Normal and Leukaemic HSPC

The ultimate goal of this analysis was to identify genes that were differentially expressed between normal and leukaemic HSPC that may be involved in leukaemogenesis and putative therapeutic targets. It has been seen, firstly, that the leukaemic populations have a progenitor-like (not stem cell-like) gene expression profile. Secondly, it has been seen that these progenitors have an immature myeloid gene expression profile, with a trend towards the GMP-like LSCs lying closer to the normal GMP and the LMPP-like LSCs with the normal LMPP (Chapters 5.5.1 and 5.5.2). Therefore, in order to identify these targets, the leukaemic LMPP and GMP were compared with the normal LMPP and GMP. DEGs were identified in DESeq2 using a Wald test and a design formula of ~population + condition was

used (where condition indicates normal/leukaemic status). This should take into account population differences but identify genes differentially expressed between normal and leukaemic HSPC.

Figure 5.20 is an MA plot showing the DEGs identified between the normal and leukaemic LMPP and GMP. 1316 genes were identified to be differentially expressed with a p-adjusted value < 0.05 and 191 genes with a p-adjusted value < 0.05 and $|\text{logFC}| > 2$. Of the latter, 129 DEGs were upregulated and 62 DEGs were downregulated in leukaemia. For DEG names, logFC and p-adjusted values see Table A.25.

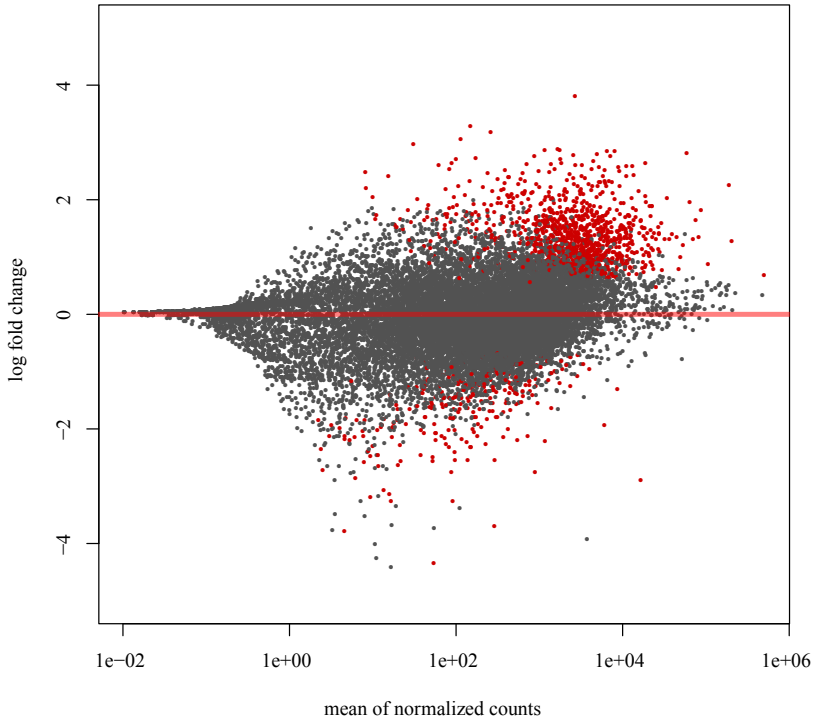


Figure 5.20: MA Plot Showing DEGs Between Normal LMPP-GMP and Leukaemic LMPP-GMP. DEGs were identified between normal LMPP and GMP and leukaemic LMPP and GMP, using a Wald test in DESeq2. A design formula of ~population + condition was used (where condition indicates normal/leukaemic status) and DEGs are shown in an MA plot [75]. A positive logFC indicates upregulation in leukaemic populations and a negative logFC indicates downregulation in leukaemic populations. Mean of normalised counts represents the mean expression of each gene. Points in red represent genes with a p-adjusted value < 0.05.

The expression of the DEGs between normal and leukaemic LMPPs and GMPs with a p-adjusted value < 0.05 can be seen in Figure 5.21. It was seen that these genes clearly separated the leukaemic populations from the normal LMPPs and GMPs. The first cluster of genes showed a clear downregulation in leukaemia compared to their normal counterparts. The second cluster of genes showed an overall upregulation in leukaemia. Although there was some variability in the expression of these genes by the leukaemic LMPP and GMP, it is clear these genes were expressed at much reduced levels by the normal LMPP and GMP (Figure 5.21).

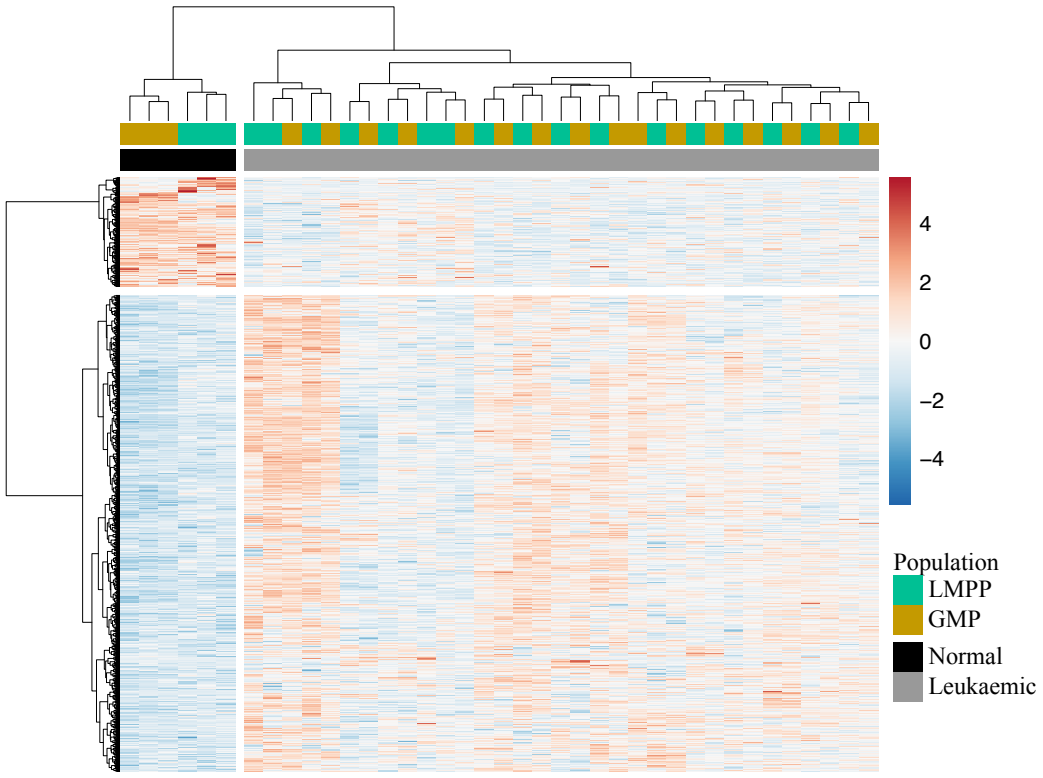


Figure 5.21: Heatmap Showing DEGs Between Normal LMPP-GMP and Leukaemic LMPP-GMP. DEGs were identified between normal LMPP and GMP and leukaemic LMPP and GMP, using a Wald test in DESeq2. A design formula of ~population + condition was used (where condition indicates normal/leukaemic status) and the expression of DEGs are shown in a heatmap. DEGs with a p-adjusted value $<$ 0.05 are shown. Colour scale is normalised per row. Populations and normal/leukaemic status are annotated by the colour bar above columns, as indicated by the key. For gene names, logFC and p-adjusted values see Table A.25.

When reviewing the DEGs between normal and leukaemic LMPP and GMP, it was seen that these genes were enriched for pathways previously associated with

cancer and haematological malignancies (Table 5.4). Several pathways involving sumoylation were identified, and it has recently been proposed that reactive oxygen species and sumoylation play a role in the chemoresistance of AML [162]. Enrichment for a pathway involving the double-stranded RNA-activated protein kinase (PKR) was identified, and elevated levels of PKR have previously been associated with poor survival in AML [163]. A pathway involving the cell cycle proteins Ras and Rho was identified, and these proteins have previously been shown to play an important role in both solid cancers and leukaemia (reviewed in [164, 165]).

Following on from this, a pathway involving Substance P was identified. It has been suggested that Substance P and its signalling play a role in cancer progression. Targeting of Substance P and its receptor neurokinin-1 receptor (NK1R) has been proposed as a therapeutic strategy [166, 167]. An epigenetic associated pathway was also enriched in AML. Epigenetics is known to play a major role in AML, with mutations in epigenetic regulators commonly occurring in AML and dysregulated epigenetic profiles having been observed (reviewed here [168]).

The cytokines leukaemia inhibitory factor (LIF) and ciliary neurotrophic growth factor (CNTF) were also identified by this analysis. LIF has been shown to play a complex role in the differentiation and proliferation of myeloid cells and is involved in maintaining self-renewal in embryonic stem-cells [169]. CNTF, a cytokine family member of LIF, has been shown to be expressed by breast cancer cell lines and the CNTF receptor-α is potentially a marker for glioma initiating cells [170, 171].

A mammalian target of rapamycin (mTOR) signalling pathway was seen to be enriched and mTOR signalling has been widely implicated in AML [172–174]. Furthermore, the mTOR pathway is regulated by Ras signalling, which itself was identified to be enriched in AML. Finally, a pathway involving positive regulation of the transcriptional coactivators YAP/TAZ, that are found in the Hippo signalling pathway, was seen to be enriched. It has been proposed that YAP/TAZ are fundamental to cancer initiation and proliferation, and are active in cancer stem cells (reviewed here [175]).

Table 5.4: DEGs Between Leukaemic LMPP and GMP: Pathway Enrichment. DEGs were identified between normal LMPP and GMP and leukaemic LMPP and GMP, using a Wald test in DESeq2. A design formula of ~population + condition was used (where condition indicates normal/leukaemic status). Metacore enrichment analysis was carried out on these DEGs and the 10 most significant pathways are shown. All the shown pathways show an upregulation in AML.

Pathway	FDR
Proteolysis: Putative SUMO-1 pathway	1.95×10^{-06}
Apoptosis and survival: Role of PKR in stress-induced apoptosis	2.99×10^{-06}
Cell cycle: Influence of Ras and Rho proteins on G1/S Transition	2.99×10^{-06}
Cytoskeleton remodelling: Substance P mediated membrane blebbing	4.00×10^{-05}
DNA damage: Role of SUMO in p53 regulation	5.79×10^{-05}
Proteolysis: Role of Parkin in the Ubiquitin-Proteasomal Pathway	9.27×10^{-05}
Transcription: Epigenetic regulation of gene expression	0.000187
Development: Role of CNTF and LIF in regulation of oligodendrocyte development	0.000310
Signal transduction: mTORC1 downstream signalling	0.000329
Development: Negative regulation of STK3/4 (Hippo) pathway and positive regulation of YAP/TAZ function	0.000349

It is known that the LSCs shares multiple characteristics with the HSCs, regarding their self-renewal and stem cell ability [22, 47]. Therefore, as an additional comparison, the leukaemic populations were compared to the normal HSCs. A Wald test was carried out in DESeq2 and a design formula of ~condition was used (where condition indicates normal/leukaemic status). Figure 5.22 shows an MA plot for these DEGs. 772 genes were identified to be differentially expressed with a p-adjusted value < 0.05, and 348 genes with a p-adjusted value < 0.05 and $|\ \text{logFC}\ | > 2$. Of the latter, 218 DEGs were upregulated and 130 DEGs were downregulated in leukaemia. For DEG names, logFC and p-adjusted values see Table A.26.

The expression of the DEGs between normal HSCs and leukaemic HSPCs with a p-adjusted value < 0.05 can be seen in Figure 5.23. Again, the separation between normal and leukaemic populations was clearly seen. First, a cluster of genes that showed markedly higher expression in the normal HSCs was seen. However, there was some sporadic expression of these genes across the leukaemic samples (Figure 5.23). Similar to the DEGs above, the genes upregulated in leukaemia did show

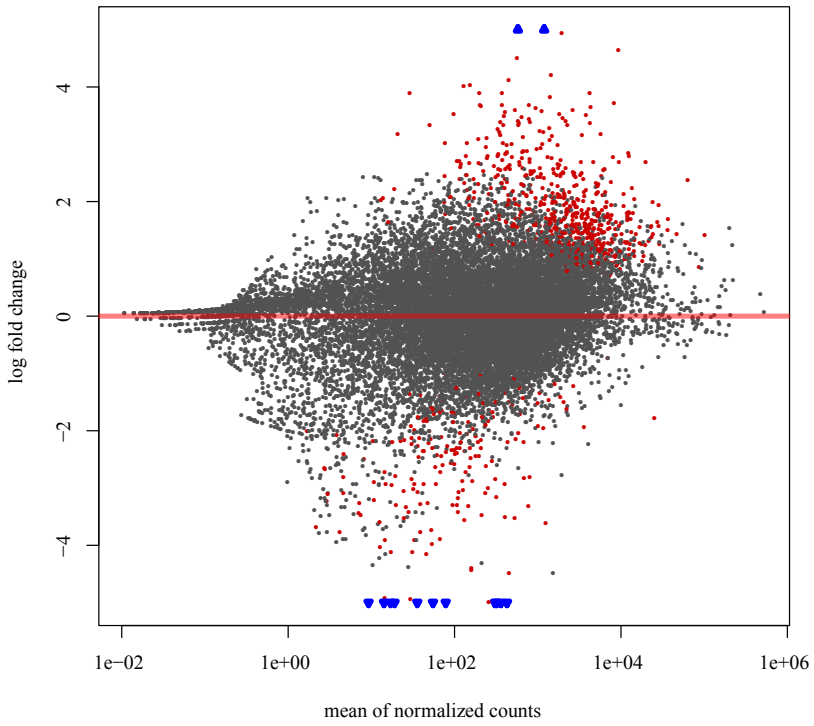


Figure 5.22: MA Plot Showing DEGs Between Normal HSC and Leukaemic HSPC. DEGs were identified between normal HSC and leukaemic HSPC, using a Wald test in DESeq2. A design formula of ~condition was used (where condition indicates normal/leukaemic status) and DEGs are shown in an MA plot [75]. A positive logFC indicates upregulation in leukaemic populations, and a negative logFC indicates downregulation in leukaemic populations. Mean of normalised counts represents the mean expression of each gene. Points in red represent genes with a p-adjusted value < 0.05. Blue triangles indicate points which are off-scale on the plot.

some variability in expression across the leukaemic samples, but much lower levels of expression in the normal HSCs.

Two sets of DEGs were generated: normal LMPP and GMP compared to the leukaemic LMPP and GMP, and normal HSC compared to the leukaemic populations. Out of a total of 1316 DEGs (p-adjusted value < 0.05), the former shared 348 DEGs with the latter comparison (i.e. 26%) (Figure 5.24). These shared 348 DEGs represented 45% of the normal HSC, compared to leukaemic HSPCs comparison.

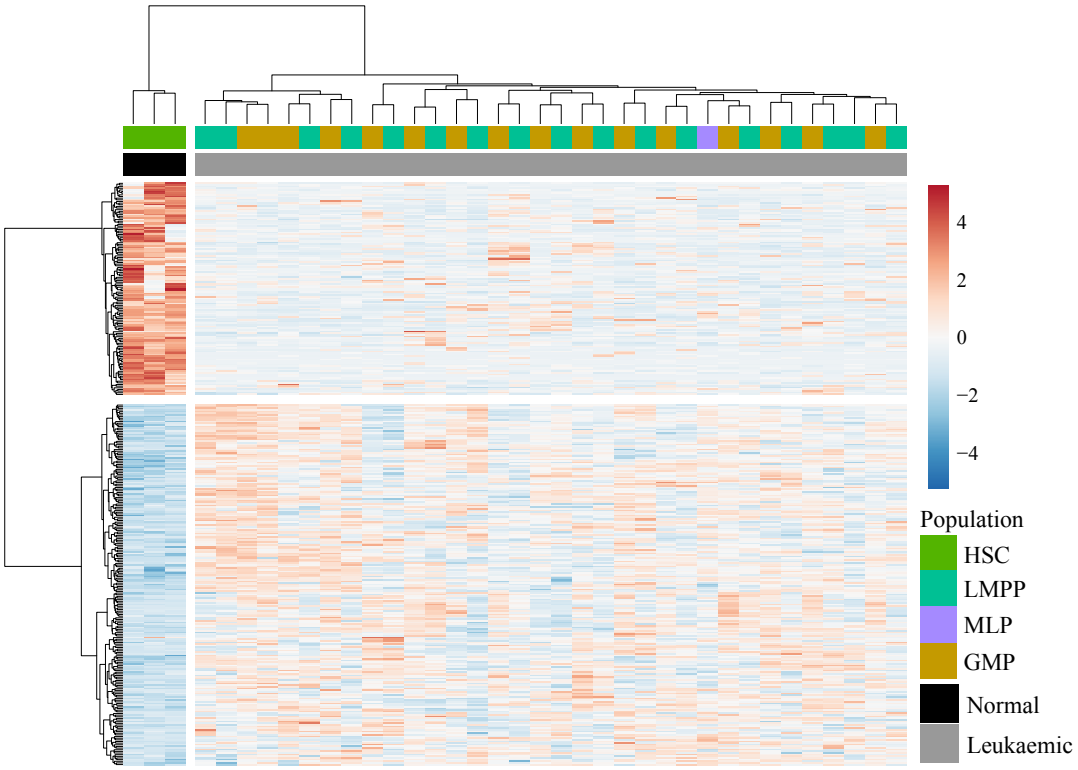


Figure 5.23: Heatmap Showing DEGs Between Normal HSC and Leukaemic HSPC. DEGs were identified between normal HSC and leukaemic HSPC, using a Wald test in DESeq2. A design formula of ~condition was used (where condition indicates normal/leukaemic status) and the expression of DEG are shown in a heatmap. DEGs with a p-adjusted value < 0.05 are shown. Colour scale is normalised per row. Populations and normal/leukaemic status are annotated by the colour bar above columns, as indicated by the key. For gene names, logFC and p-adjusted values see Table A.26.

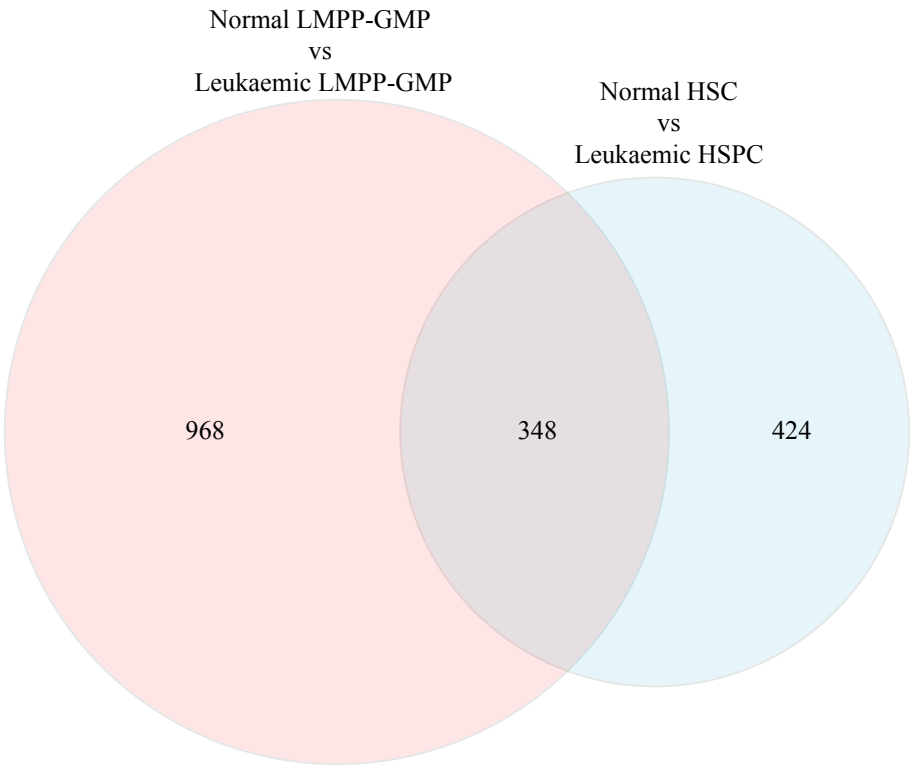


Figure 5.24: Overlap Between Two Sets of DEGs Identified Between Normal and Leukaemic HSPCs. DEGs were identified between normal HSC and leukaemic populations, using a Wald test in DESeq2. A design formula of ~condition was used (where condition indicates normal/leukaemic status). DEGs were also identified between normal LMPP and GMPP and leukaemic LMPP and GMP, using a Wald test in DESeq2. Here a design formula of ~population + condition was used. The overlap between these two lists of DEGs is shown, when considering genes with a p-adjusted value < 0.05.

5.5.5 Expression of Stem Cell Signature by Leukaemic HSPC

When analysing the normal HSPCs, DEGs upregulated in the early HSC and MPP populations were identified (Figure 5.9). It has been suggested that although the immunophenotype and gene expression profile of the LSCs resemble progenitors, the LSCs re-activate a stem-cell programme [22]. Therefore, the LSCs' expression of the 200 most significantly upregulated genes by the HSCs and MPPs was investigated (Figure 5.25). As seen before, these genes showed distinct upregulation in the HSC and MPP populations compared to normal progenitors (Figures 5.9, 5.25). In contrast, the LSCs did not show uniform expression of these DEGs. It was seen that the LSCs variably upregulated stem-associated genes, with different samples upregulating different clusters of stem-associated genes (Figure 5.25).

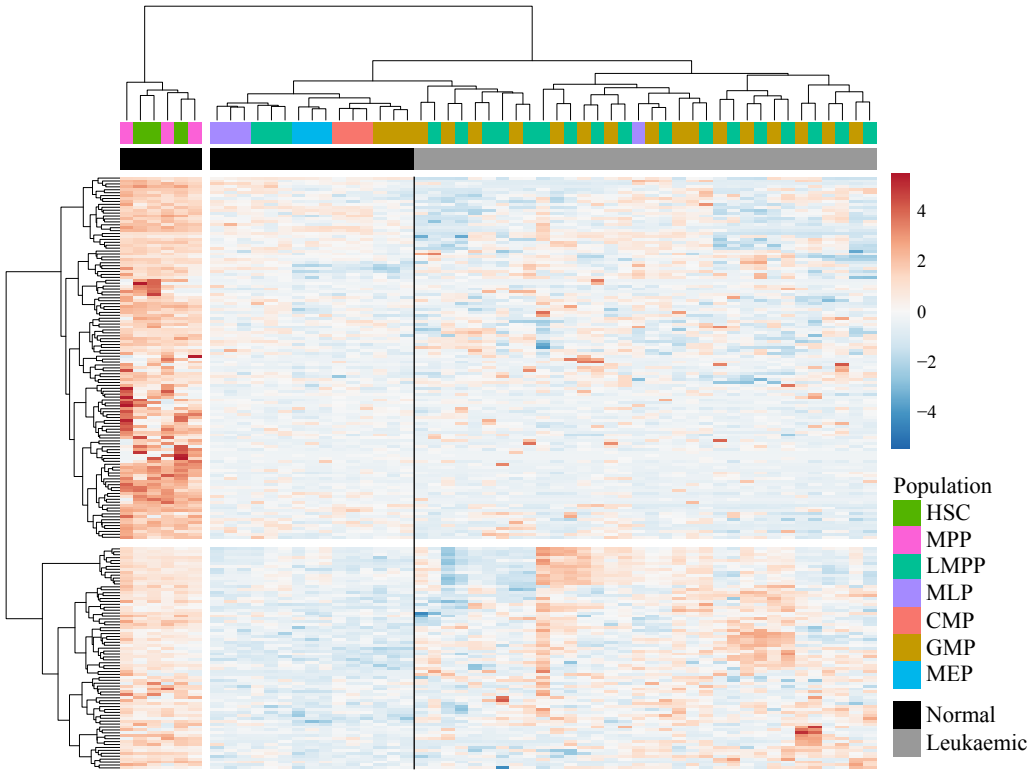


Figure 5.25: Heatmap Showing the Expression of Stem Cell-Associated Genes by Leukaemic HSPC. DEGs were identified between early stem (HSC and MPP) and progenitor (LMPP, MLP, CMP, GMP, MEP) populations using a Wald test, and the expression of the 200 most significant DEGs are shown across normal and leukaemic HSPC. Rows represent genes and columns represent a biological replicate of an HSPC population. Hierarchical clustering performed using the wardD2 method. Colour scale is normalised per row. Populations and normal/leukaemic status are annotated by the colour bar above columns, as indicated by the key. For gene names, logFC and p-adjusted values see A.24.

5.6 Analysis of the Normal and Leukaemic LMPP and GMP

5.6.1 Normal LMPP and GMP

Finally, the analysis focussed on the gene expression changes between the LMPP and GMP populations in normal haematopoiesis and how this is dysregulated in leukaemia. It has been shown that functionally the normal LMPP and GMP are very different (Chapter 3). Whereas the GMP was primarily a myeloid progenitor with only residual lymphoid ability, the LMPP was a lymphoid-biased LM progenitor

(Chapter 3). Therefore, it was expected for there to be considerable gene expression differences between the normal LMPP and GMP.

DEGs were identified between the normal LMPPs and GMPs, using a Wald test in DESeq2. Figure 5.26 shows an MA plot for these DEGs. 1280 genes were identified to be differentially expressed with a p-adjusted value < 0.05, and 544 genes with a p-adjusted value < 0.05 and $|\ \text{logFC}\ | > 2$. Of the latter, 288 DEGs were upregulated and 256 DEGs were downregulated in the LMPP. For DEG names, logFC and p-adjusted values see Table A.27.

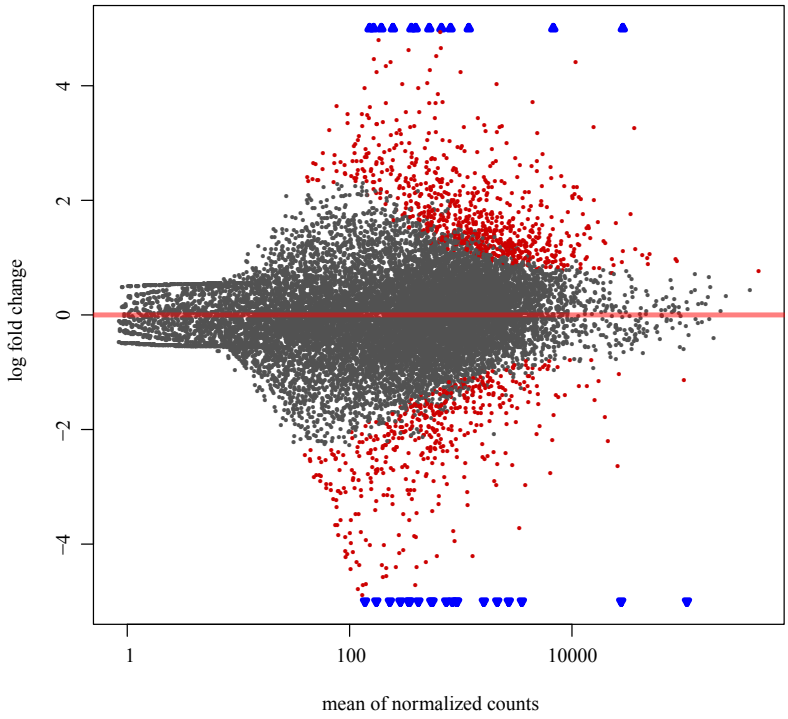


Figure 5.26: MA Plot Showing DEGs Between Normal LMPP and GMP. DEGs were identified between normal LMPP and GMP populations using a Wald test and shown in an MA plot [75]. A positive logFC indicates upregulation in the LMPP, and a negative logFC indicates downregulation in the LMPP. Mean of normalised counts represents the mean expression of each gene. Points in red represent genes with a p-adjusted value < 0.05. Blue triangles indicate points which are off-scale on the plot. For genes see A.27.

As expected, when looking at the expression of the DEGs identified with a p-adjusted value < 0.05, clear separation between the normal LMPP and GMP

was seen (Figure 5.27). It was seen that the two clusters of genes showed distinct opposite regulation in their respective populations.

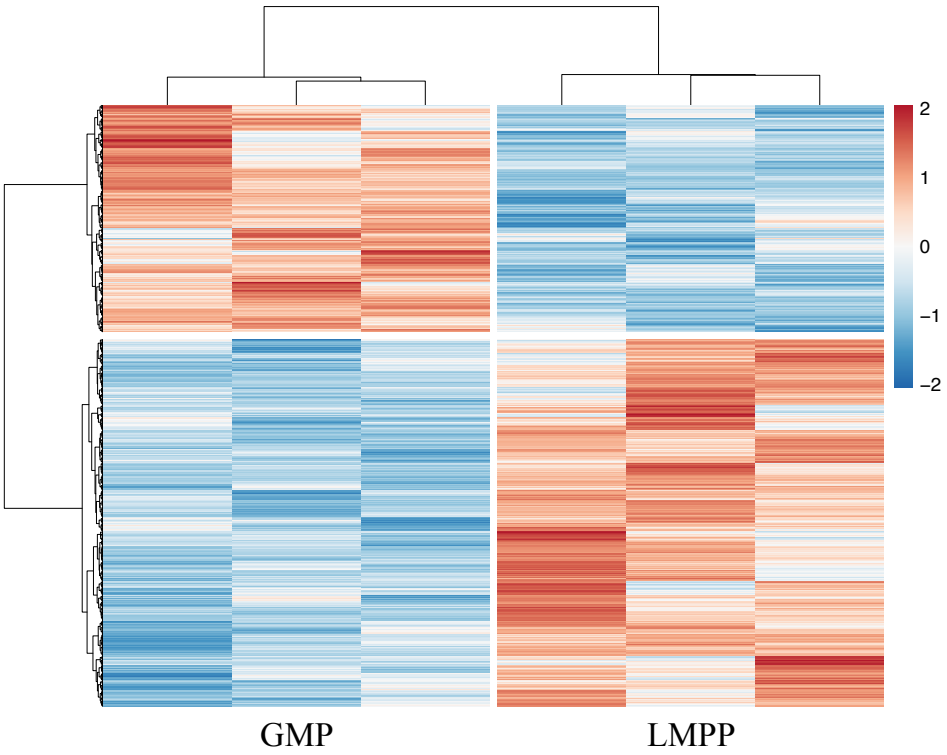


Figure 5.27: Heatmap Showing the Expression of DEGs Between the Normal LMPP and GMP. DEGs were identified between normal LMPP and GMP populations using a Wald test and the expression of DEGs with a p-adjusted value < 0.05 is shown. Rows represent genes and columns represent a biological replicate of an HSPC population. Hierarchical clustering performed using the wardD2 method. Colour scale is normalised per row. Populations are annotated below the plot. For gene names, logFC and p-adjusted values see A.27.

5.6.2 Leukaemic LMPP and GMP

The same analysis was then repeated for the leukaemic LMPP- and GMP-like populations, to see if these gene expression differences were maintained. DEGs were identified between the leukaemic LMPP and GMP, using a Wald test in DESeq2. Figure 5.28 shows an MA plot for these DEGs. 242 genes were identified to be differentially expressed with a p-adjusted value < 0.05 and 4 genes with a p-adjusted value < 0.05 and $|\ \text{logFC}\ | > 2$. All 4 DEGs were downregulated in the LMPP. For DEG names, logFC and p-adjusted values see Table A.28.

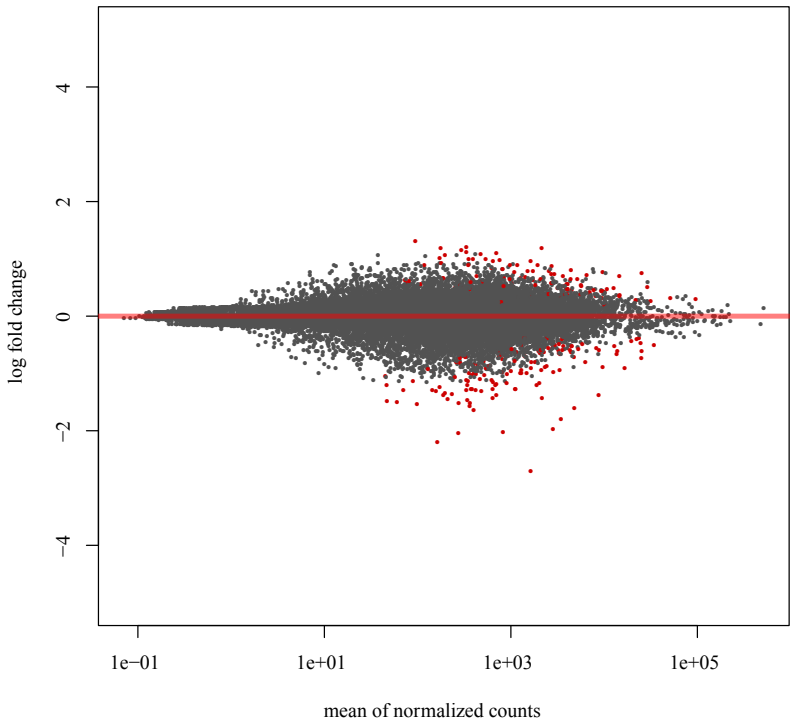


Figure 5.28: MA Plot Showing DEGs Between Leukaemic LMPP and GMP. DEGs were identified between leukaemic LMPP and GMP populations using a Wald test and shown in an MA plot [75]. A positive logFC indicates upregulation in the LMPP and a negative logFC indicates downregulation in the LMPP. Mean of normalised counts represents the mean expression of each gene. Points in red represent genes with a p-adjusted value < 0.05. Blue triangles indicate points which are off-scale on the plot. For genes see A.28.

The expression of these DEGs with a p-adjusted value < 0.05 is shown in Figure 5.29. Although these genes were mostly able to separate the leukaemic LMPP from the GMP, their pattern of expression was much more variable. There was neither consistent upregulation of each cluster of genes in its respective population, nor consistent downregulation in the other population (Figure 5.29).

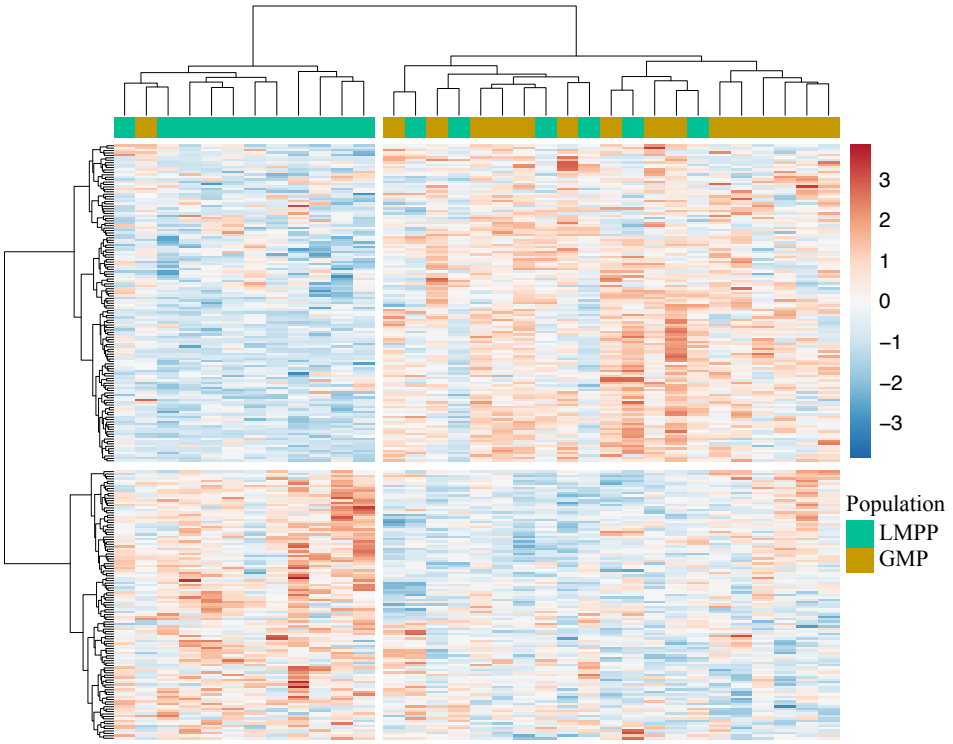


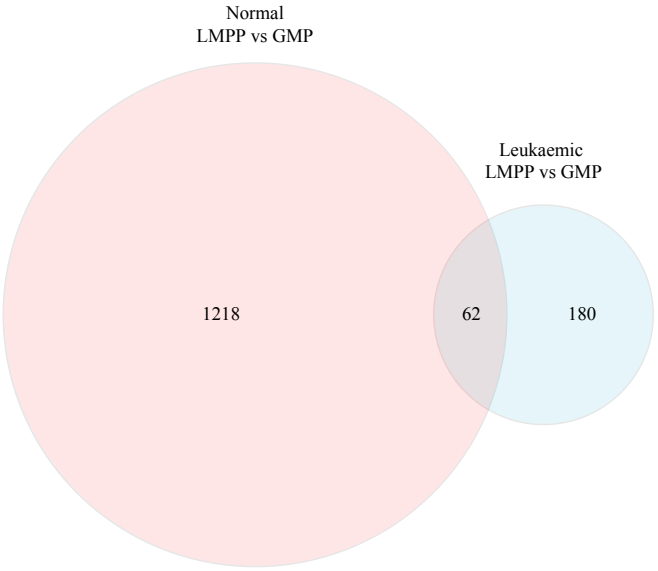
Figure 5.29: Heatmap Showing the Expression of DEGs Between the Leukaemic LMPP and GMP. DEGs were identified between leukaemic LMPP and GMP populations using a Wald test and the expression of DEGs with a p-adjusted value < 0.05 is shown. Rows represent genes and columns represent a biological replicate of an HSPC population. Hierarchical clustering performed using the wardD2 method. Colour scale is normalised per row. Populations are annotated by the colour bar above columns, as indicated by the key. For gene names, logFC and p-adjusted values see A.28.

5.6.3 Comparing Normal and Leukaemic LMPP and GMP

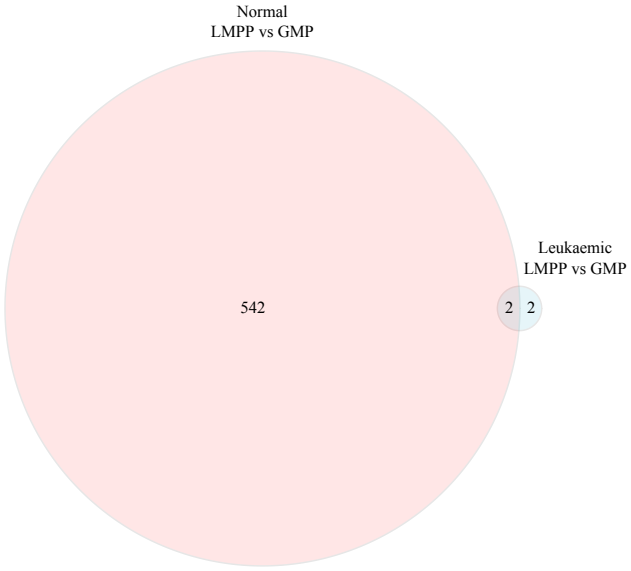
If the gene expression differences seen between the normal LMPP and GMP were maintained in leukaemia, a high proportion of the DEGs identified in both comparisons should be overlapping. When looking at the DEGs with a p-adjusted value < 0.05, only 62 out of a total of 1280 (5%) DEGs identified between normal LMPP and GMP were shared with the DEGs identified between leukaemic LMPP and GMP (Figure 5.30a and Table 5.5). These 62 genes represented 62 out of a total of 242 (26%) of the DEGs identified between leukaemic LMPP and GMP.

However, when looking at genes with a p-adjusted value < 0.05 and $|\ \text{logFC}\ | > 2$, arguably the genes more likely to have a biologically relevant function, it was seen that very few genes from the leukaemic comparison passed this threshold (Figure

5.30b). Only 4 genes were differentially expressed between the leukaemic LMPP and GMP, and only 2 of these were shared with the DEGs between normal LMPP and GMP (Table 5.5). One of these was *CD38*, the cell surface marker segregating these populations, with the LMPP being $CD38^-$ and the GMP $CD38^+$. The second, *MELK* is a serine/threonine-protein kinase with a broad role in many processes including cell cycle regulation, self-renewal regulation and apoptosis. *MELK* has previously been shown to be expressed by HSCs and is upregulated in both solid cancers and AML [176, 177]. These 2 shared genes represented 2 out of a total of 544 (0.4%) of the DEGs identified between normal LMPP and GMP (Figure 5.30b).



(a) p-adjusted value < 0.05



(b) p-adjusted value < 0.05 & $|$ logFC $| > 2$

Figure 5.30: Overlap Between DEGs Between LMPP and GMP from Normal and Leukaemic Haematopoiesis. DEGs were identified between normal/leukaemic LMPP and GMP populations using a Wald test. The overlap between the DEGs between normal LMPP and GMP and leukaemic LMPP and GMP is shown when considering DEGs with (a) a p-adjusted value < 0.05 and (b) a p-adjusted value < 0.05 and an $|$ logFC $| > 2$.

Table 5.5: Overlapping Genes Between DEGs Between LMPP and GMP from Normal and Leukaemic Haematopoiesis. DEGs were identified between normal/leukaemic LMPP and GMP populations using a Wald test. The overlapping genes between the DEGs between normal LMPP and GMP and leukaemic LMPP and GMP is shown when considering DEGs with either a p-adjusted value < 0.05 or a p-adjusted value < 0.05 and an $|\text{logFC}| > 2$. p-adj: p-adjusted value.

Criteria	Gene			
p-adj < 0.05	LYZ	HAL	MPO	ETNK1
	CDC20	MEIS1	TSC22D1	SRGN
	CSF3R	ANGPT1	APLP2	FYB
	SRSF6	TSPOAP1	PHC3	STON2
	KLF7	TROVE2	IGFBP7	CD38
	C6orf62	DNTT	N4BP2L2	DNMT1
	ZNF521	C1QBP	BNIP3L	CLSPN
	PLAC8	CRBN	C16orf72	TSC22D2
	CNST	BUB1B	POLQ	UQCRQ
	CD47	DUT	DENND3	USP9X
	ASH1L	MCM4	TSC22D3	KIF15
	DHFR	ZNF706	H2AFZ	NRIP1
	MELK	NHSL2	FOXP1	UHRF1
	CENPM	H2AFY	PTTG1	USP34
	MTURN	SERINC5	CACUL1	HMGA1
	SPC24	SBDS		
p-adj < 0.05 and $\|\text{logFC}\| > 2$	CD38	MELK		

The moderate overlap observed between the DEGs identified between the LMPP and GMP from leukaemic and normal conditions, implies that the gene expression changes normally seen between the LMPP and GMP were not maintained in leukaemia. To confirm this, the expression of the DEGs between the normal LMPP and GMP by the leukaemic LMPP and GMP was investigated (Figure 5.31). When looking at the 544 DEGs with a p-adjusted value < 0.05 and $|\text{logFC}| > 2$, it can be seen that these genes did not segregate the leukaemic LMPP and GMP (Figure 5.31). There was a clear dysregulation in gene expression profiles and leukaemic samples were seen to cluster by patient.

Since this dataset suggested that the LMPP and GMP in leukaemia do not maintain the gene expression differences seen in normal haematopoiesis, the biology underlying their transcriptional differences was investigated. A Metacore enrichment

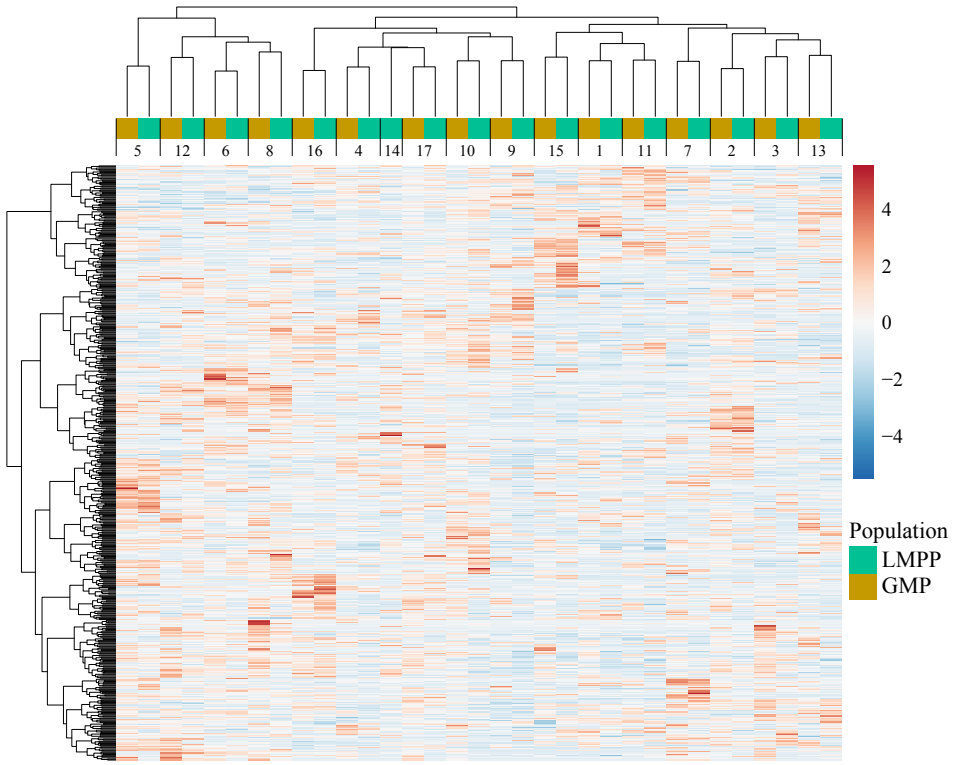


Figure 5.31: Heatmap Showing the Expression of DEGs Between the Normal LMPP and GMP by the Leukaemic LMPP and GMP. DEGs were identified between normal LMPP and GMP populations using a Wald test. The expression of DEGs with a p-adjusted value < 0.05 and $|\ \text{logFC}\ | > 2$ by leukaemic LMPP and GMP is shown. Rows represent genes and columns represent a biological replicate of an HSPC population. Replicates annotated above the plot, as indicated in 5.3. Hierarchical clustering performed using the wardD2 method. Colour scale is normalised per row. Populations are annotated by the colour bar above columns, as indicated by the key. For gene names, logFC and p-adjusted values see A.27.

analysis was carried out on the DEGs identified between the leukaemic LMPP and GMP (Table 5.6). When looking at the 5 most significantly enriched Metacore pathways and GO processes, it can be seen that these all involved the cell cycle and were upregulated in the GMP. This suggests that the GMP-like LSC may be more actively cycling than the LMPP-like LSC, which is in agreement with previous work comparing the $CD34^{+}CD38^{-}$ and $CD34^{+}CD38^{+}$ LSCs [120].

Table 5.6: DEGs Between Leukaemic LMPP and GMP: Pathway Enrichment. DEGs were identified between leukaemic LMPP and GMP populations using a Wald test. Metacore enrichment analysis was carried out on these DEGs and the 5 most significant pathways and 5 most significant gene ontology (GO) processes are shown. All the shown pathways show a downregulation in the LMPP and upregulation in the GMP.

Pathway/Process	FDR
Metacore Pathways	
Cell cycle: Role of APC in cell cycle regulation	3.527×10^{-13}
Cell cycle: Spindle assembly and chromosome separation	2.482×10^{-08}
Cell cycle: The metaphase checkpoint	2.966×10^{-05}
Inhibition of remyelination in multiple sclerosis: regulation of cytoskeleton proteins	1.596×10^{-03}
Cell cycle: Role of Nek in cell cycle regulation	3.085×10^{-03}
GO Processes	
Cell Cycle	5.165×10^{-17}
Mitotic Cell Cycle	1.132×10^{-16}
Mitotic Cell Cycle Process	9.834×10^{-14}
Cell Cycle Process	1.464×10^{-13}
Mitotic Cell Cycle Phase Transition	4.510×10^{-12}

5.7 Conclusion on the Gene Expression Analysis of AML

The bulk RNA-Seq analysis of normal HSPC populations has revealed that these were distinct populations, with gene expression profiles that reflected their functional abilities. When looking at AML samples in isolation, it was seen that variation across leukaemia was complex. Genes associated with LM differentiation and previously reported to play a role in cancer were identified. These may reflect the underlying mechanisms of leukaemogenesis in the samples. This is in agreement with the strong patient effect observed throughout the analysis. When the AML samples were studied with normal HSPCs, it could be seen that the leukaemic gene expression profile was progenitor-like, with partial re-activation of stem cell associated genes. The myeloid differentiation arrest was clear, with the leukaemic progenitors displaying an immature myeloid profile. However, the AML samples were highly dysregulated and no longer resembling normal HSPCs. This allowed DEGs to be identified between normal HSPC and leukaemic HSPC, by two methods. Finally, when comparing the leukaemic LMPP and GMP, it was apparent that the

normal biological differences between these populations was lost. In AML it was seen that the GMP-like population potentially represented a more actively cycling LSC.

Chapter 6

Functional Validation of Putative Leukaemia-Associated Targets

Contents

6.1 Introduction to the Use of CRISPR on Haematopoietic Cells 211
6.2 Integration of DEG with Genome-Wide CRISPR Screens 213
6.3 CRISPR-Cas9 Editing of K562 Cells 219
6.3.1 Effect of CRISPR Editing on Viability 219
6.3.2 Inference of CRISPR Editing Analysis 226
6.3.3 Relationship Between ICE Scores and GFP Positivity . 230
6.3.4 Indels Introduced in K652 by CRISPR Editing 234
6.4 Optimisation of the *In Vitro* Culture of Primary AML 247
6.5 Lentiviral Transduction of Primary HSPCs 260
6.6 Conclusion on the Functional Validation of Putative Leukaemia-Associated Targets 265

6.1 Introduction to the Use of CRISPR on Haematopoietic Cells

Bacteria and archaea utilise immune defence systems to destroy invading nucleic acids, a method termed 'Clustered Regularly Interspaced Short Palindromic Repeats' (CRISPR)/CRISPR-associated (Cas) systems. The adaptation of these systems to be used as a research tool has revolutionised the genome editing field [178–181]. CRISPR-Cas9 editing has been widely used on haematopoietic cells, with optimised protocols for mouse HSPC having been published [182]. Additionally, human haematopoietic cell lines have been edited using CRISPR-Cas9, including genome-wide CRISPR-Cas9 screens in leukaemic cell lines [78, 79].

The editing of primary human HSPCs has proven to be more challenging, especially when targeting the HSC. The more traditional delivery method of the CRISPR-Cas9 machinery is by lentiviral transduction. Many have studied the

lentiviral transduction of primary HSPC, primarily in the context of clinical gene therapy [183–185]. The challenges associated with the lentiviral transduction of HSPC are linked to the difficulty of maintaining stem cell activity *in vitro*, the quiescent nature of HSPCs and potentially low transduction efficiencies [183, 184].

Alternative delivery methods of the CRISPR-Cas9 machinery into primary cells have now been developed. Firstly, direct electroporation of plasmid(s) containing the Cas9 and guide RNAs has been used to successfully edit primary $CD34^+$ HSPC [186–189]. This can be limited by increased levels of cell death following electroporation. Alternatively, the guide RNA may be electroporated with the mRNA encoding the Cas9 protein [186, 190]. Secondly, the electroporation of the pre-assembled ribonucleoprotein (RNP) complex, containing the Cas9 protein and guide RNA, has also been shown to successfully edit primary $CD34^+$ HSPC [89, 182, 190, 191].

Whereas the non-lentiviral methods have the advantage of potentially higher editing efficiencies, these are all transient methods with the guide RNA and Cas9 eventually being degraded. For incorporation of the CRISPR machinery into the host's genome, lentiviral methods are more appropriate. These also have the advantage of allowing a fluorescent reporter to be incorporated and used for selection of transduced cells. There is the possibility to incorporate a fluorescent reporter using an adeno-associated virus (AAV) in conjunction with RNP, but this relies on the less efficient homologous recombination (HR) method of DNA repair [89]. Additionally, for a traditional drop-out screen the presence of the guide RNA in the host's genome is tracked, and this is only possible when using lentiviral integration [78, 79]. It has more recently been shown that edited cells may be tracked by the presence of genomic scarring as a result of DNA repair [188].

Although the editing of human leukaemia cell lines is well published, the editing of primary human AML remains a challenge. A primary limiting factor in this process is the *in vitro* culture of AML cells, which typically tend to differentiate and/or die *in vitro*. However, progress has been made in the culture of AML cells, with the identification of culture conditions that may better support LSC

maintenance *in vitro* [85]. There is very little published work on the CRISPR-Cas9 editing of primary human AML. Brabetz et al. targeted the *IDH2* gene in primary AML blasts by HR, using lentiviral transduction. However, no downstream work was carried out on the edited cells [192]. Hou et al. knocked-out (KO) the *SPRY3* gene in primary AML blasts, by transfection of the guide RNA and Cas9 mRNA. Although an *in vitro* functional assay was carried out, no *in vivo* work was done on edited cells [193].

I aimed to functionally validate the DEG identified in Chapter 5.5.4. Firstly, to identify genes with a potential functional role in leukaemogenesis, I integrated my DEG with genome-wide CRISPR screens in leukaemic cell lines [78, 79]. Secondly, a CRISPR-Cas9 strategy was designed to target these genes of interest, and gene editing was tested and confirmed in a leukaemic cell line. The ultimate goal was to target these genes in primary human AML LSCs. To enable the editing of primary AML, the *in vitro* culture of AML cells was optimised and preliminary experiments on the lentiviral transduction of AML were carried out.

6.2 Integration of DEG with Genome-Wide CRISPR Screens

In Chapter 5.5.4 it was shown that genes upregulated in the LSC when compared to normal HSPC could be identified. When comparing the leukaemic LMPP- and GMP-like LSCs to the normal LMPP and GMP, 129 genes were identified to be upregulated in leukaemia (p-adjusted value < 0.05 and logFC > 2, Table A.25). In order to identify which of these may play a functional role in leukaemogenesis, these DEG were integrated with publicly available genome-wide CRISPR-Cas9 screens.

Two major genome-wide CRISPR-Cas9 screens have been carried out in AML. These were both drop-out screens, where genome-wide targeting of genes by CRISPR aims to identify genes that are essential to the cells' survival. Lentiviral transduction of cells with the guide RNA allows incorporation of the guide sequence into the host's genome. If a gene is essential to the survival of a cell, then its KO by CRISPR would result in cell death. Post-transduction sequencing would

reveal the absence/reduction in this guide sequence and is an indication of the essentiality of the gene.

Tzelepis et al. carried out a CRISPR screen across 5 AML cell lines (MOLM-13, MV4-11, HL-60, OCI-AML2 and OCI-AML3) and 2 non-AML cell lines (HT29C and HT-1080) [79]. This enabled the identification of previously known therapeutic targets and novel targets of interest, including *KAT2A*. Following on from this, Wang et al. carried out a CRISPR screen across 14 AML cell lines (NB4, PL-21, SKM-1, OCI-AML3, MV4 11, EOL-1, MonoMac1, OCI- AML2, P31/FUJ, Molm 13, TF-1, THP-1, HEL and OCI-AML5) [78]. Here lethal partners of oncogenic Ras were uncovered and *PREX1* was identified to activate mitogen-activated protein kinase (MAPK) signalling specifically in AML [78].

Firstly, the DEG identified in Chapter 5.5.4 were integrated with the dataset from Wang et al. In this screen, a CRISPR gene score (CS) was calculated for each gene, where the more negative the CS the more severe the drop-out of a gene and its proposed essentiality [78]. The CS for the 129 upregulated genes in leukaemia were identified and are shown in Figure 6.1. It was clearly seen that the hierarchical clustering was able to identify a cluster of genes which had increased levels of drop-out across the AML cell lines. These 6 genes were: *CLP1*, *SBDS*, *PSMD12*, *WDR26*, *COPS2* and *FASTKD5* (Figure 6.1).

Turning to the CRISPR screen carried out by Tzelepis et al., the 129 DEG were also integrated with this dataset. Here, MAGeCK software was used to assess gene drop-out, and each gene was assigned a false discovery rate (FDR) associated with its drop-out, and proposed essentiality [79, 194]. The FDRs of drop-out for the 129 upregulated genes were identified and are shown in Figure 6.2. Similarly, it was seen that the hierarchical clustering was able to identify a cluster of genes with a highly significant drop-out. These included the 6 genes identified in Figure 6.1 and additionally *ITGB5*, *FAM91A1*, *SYT15*, *KPNA2* and *GOLT1B*. It was seen that these genes also showed varying levels of drop-out in the non-AML cell lines. However, since both HT29C and HT-1080 are cancerous cell lines (being a colon adenocarcinoma and fibrocarcinoma cell line respectively),

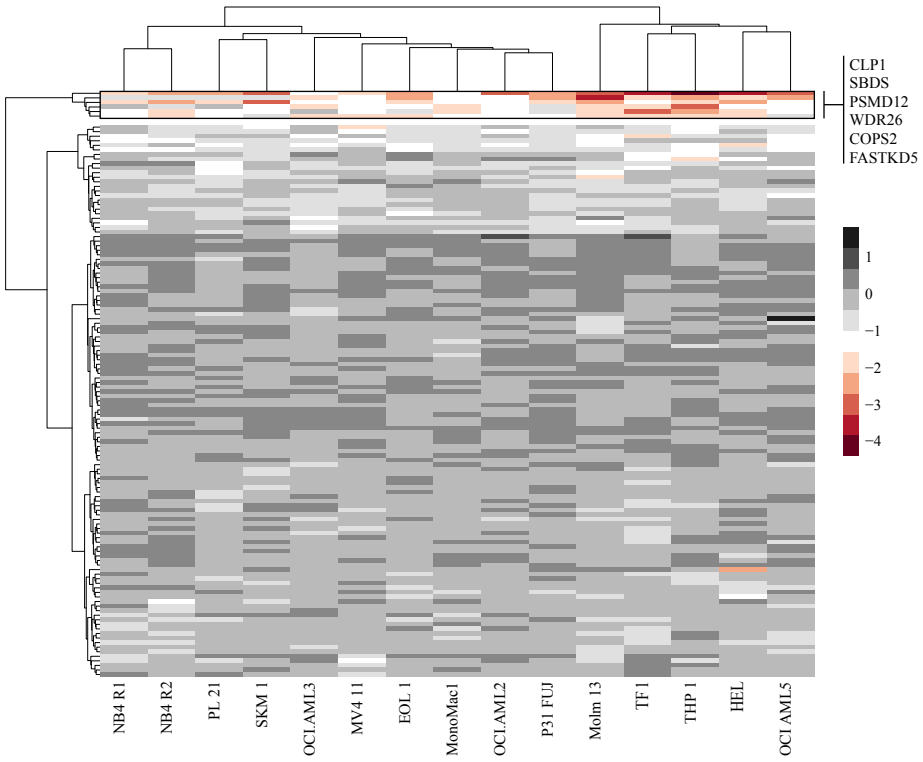


Figure 6.1: Integration of Wang et al. CRISPR Screen and Upregulated Leukaemia Genes. Wang et al. conducted a genome-wide CRISPR-Cas9 screen across 14 human AML cell lines [78]. A CS was calculated for each gene, where the more negative the CS the more severe the drop-out of a gene and its proposed essentiality [78]. DEGs were identified between normal LMPP and GMP and leukaemic LMPP and GMP (see Chapter 5.5.4). These datasets were integrated to identify the CS for DEG with a p-adjusted value < 0.05 and logFC > 2 (i.e. up-regulated genes in leukaemia). The CS for these genes are shown in a heatmap where the colour scale indicates the CS. Hierarchical clustering done using the wardD2 method. The AML cell line is indicated below heatmap. R1/2: Replicate 1/2 for NB4 cell line.

if these genes were involved in a mechanism that spanned cancer types, this is potentially biologically meaningful.

The 6 genes that were identified to be both upregulated in leukaemia and drop-out in both of the genome-wide CRISPR screens [78, 79] were then focussed on as putative targets of interest. The essentiality of these genes in non-AML genome-wide CRISPR drop-out screens was then investigated. Firstly, a genome-wide screen carried out by Steinhart et al. was interrogated. Steinhart et al. carried out a screen in pancreatic ductal adenocarcinoma (PDAC) cell lines, in comparison with non-PDAC cell lines previously screened [83]. BAGEL software was used to assess

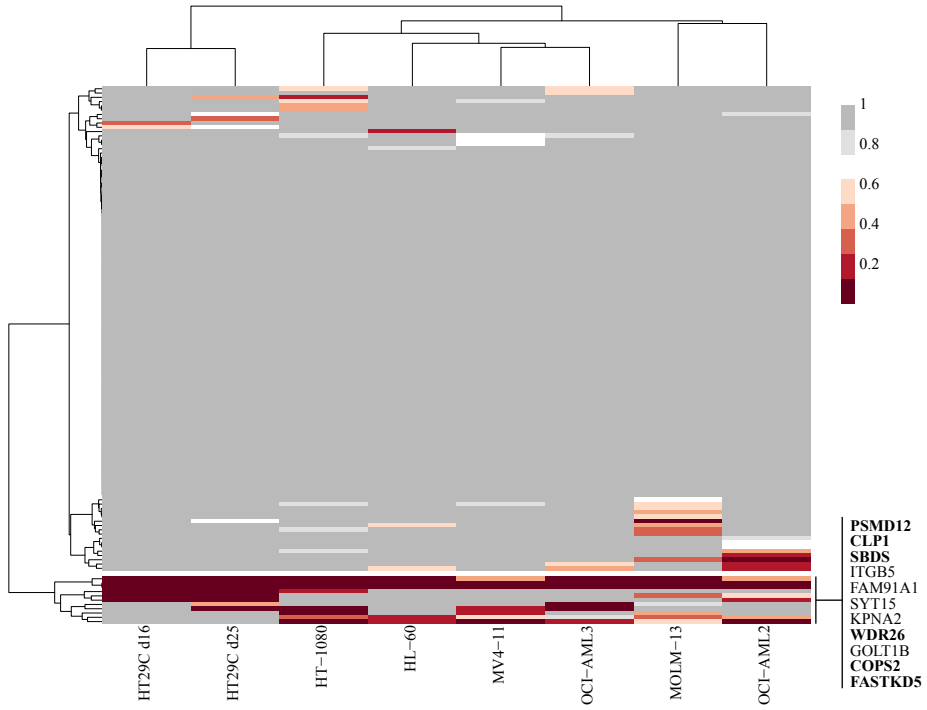


Figure 6.2: Integration of Tzelepis et al. CRISPR Screen and Upregulated Leukaemia Genes. Tzelepis et al. conducted a genome-wide CRISPR-Cas9 screen across 5 human AML cell lines and 2 non-AML cell lines (HT29C and HT-1080) [79]. MAGeCK software was used to assess gene drop-out, and each gene was assigned an FDR associated with its drop-out and proposed essentiality [79, 194]. DEGs were identified between normal LMPP and GMP and leukaemic LMPP and GMP (see Chapter 5.5.4). These datasets were integrated to identify the FDR for DEG with a p-adjusted value $<$ 0.05 and logFC $>$ 2 (i.e. up-regulated genes in leukaemia). The FDR for these genes are shown in a heatmap where the colour scale indicates the FDR. The genes also identified in Figure 6.1 are indicated in bold. Hierarchical clustering done using the wardD2 method. The cell line is indicated below heatmap. d16/25: day 16/25, multiple time-points were used to assess drop-out in HT29C cells.

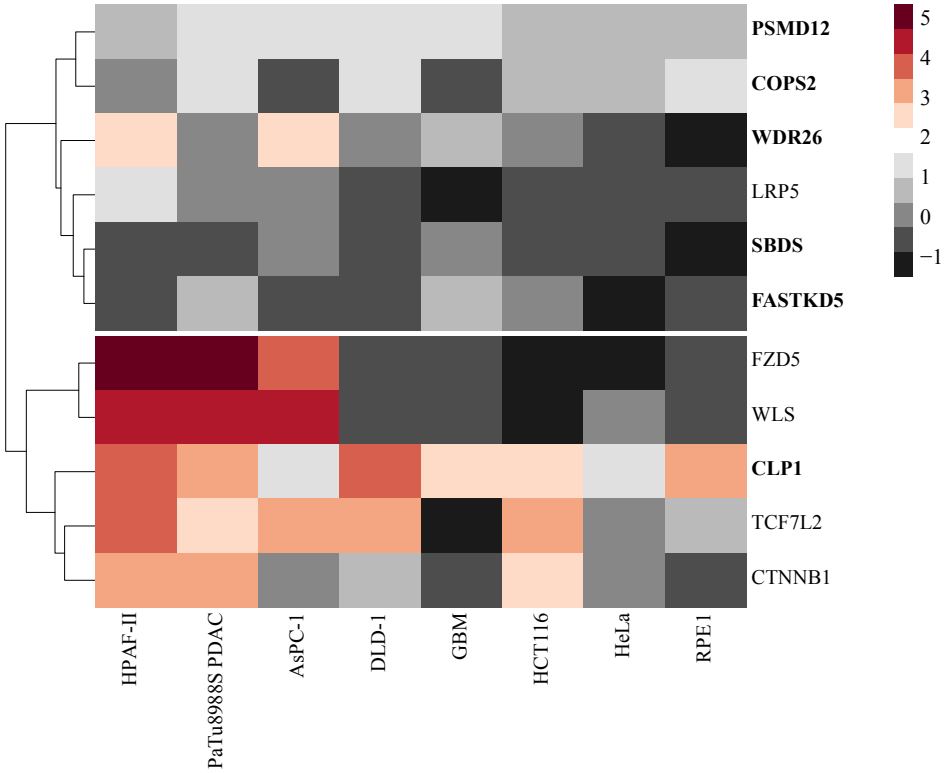
gene drop-out and each gene was assigned a Bayes factor (BF) associated with its proposed essentiality, where the higher the BF the more essential the gene [81, 82].

Figure 6.3a shows the BF for the 6 genes of interest (Figure 6.3a, indicated in bold) and the 5 genes that showed the highest BF in PDAC cell lines (HPAF-II, PaTu8988S PDAC and AsPC-1) as compared to the non-PDAC cell lines (DLD-1, GBM, HCT116, HeLa and RPE1). It was seen that the hierarchical clustering was able to segregate these genes into 2 clusters. One cluster contained 5 out of 6 genes of interest and showed lower BF across all the cell lines. This cluster

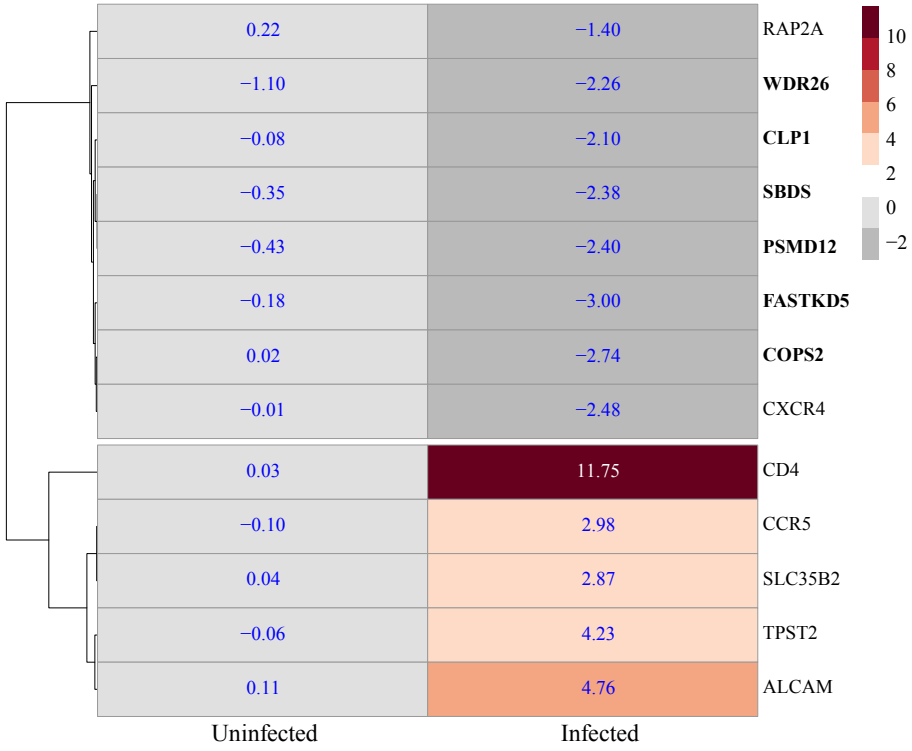
also contained the gene *LRP5* which did have a BF higher in PDAC compared to non-PDAC cell lines, however was the lowest out of the top 5 genes identified by Steinhart et al. (Figure 6.3a).

In the second cluster of genes, the remaining 4 genes identified to have higher BF in PDAC cell lines and *CLP1* were seen. *CLP1* appeared to have moderate–high BFs for all cell lines. Importantly, the cell line RPE1 is a non-cancerous cell line, immortalised by expression of *TERT*. All genes of interest (with the exception of *CLP1*) showed low BF in RPE1, indicating they were not fundamentally essential genes. This suggested that it is possible that *CLP1* plays a more fundamental role in cellular biology.

Park et al. have carried out a different style of CRISPR screen [80]. Here a $CD4^+$ T cell line was generated and a genome-wide CRISPR screen was used to identify genes that, when lost, gave protection from a human immunodeficiency virus (HIV) infection. Here the abundance of guide RNA was compared before and after an HIV infection, with uninfected cells as a control. Guide RNAs that were enriched post-infection must have inactivated a gene that was necessary for HIV-mediated killing, with the most enriched gene being the HIV co-receptor *CD4* [80]. Here the logFC in the abundance of the guide RNA was calculated, where the higher the logFC the more enriched the gene in the edited cells [80]. The logFC values were identified for the 6 genes of interest (Figure 6.3b, indicated in bold), the 5 most enriched genes post-HIV infection (*CD4*, *CCR5*, *SLC35B2*, *TPST2* and *ALCAM*) and 2 infection control genes (*RAP2A* and *CXCR4*) (Figure 6.3b). It can be seen that the hierarchical clustering separated out the genes enriched post-HIV infection. On the other hand, the 6 genes of interest clustered with the control genes, not expected to change upon infection. Notably, in the absence of HIV, none of the 6 genes of interest showed a large logFC (Figure 6.3b).



(a) Steinhart et al. Genome-Wide CRISPR-Cas9 Screen



(b) Park et al. Genome-Wide CRISPR-Cas9 Screen

Figure 6.3: Integration of Steinhart et al. and Park et al. CRISPR Screens and Upregulated Leukaemia Genes. Steinhart et al. and Park et al. conducted genome-wide CRISPR-Cas9 screens in non-leukaemic cells [80, 81]. (a) Steinhart et al. carried out the screen in PDAC and non-PDAC cell lines. BAGEL software was used to assess gene drop-out and each gene was assigned a BF associated with its proposed essentiality, where the higher the BF the more essential the gene [81–83]. The cell line is indicated below the heatmap. (b) Park et al. carried out the screen in a T cell line that was infected with HIV. The logFC in the abundance of the sgRNA was calculated, where the higher the logFC the more enriched the gene in the edited cells [80]. The infected/uninfected status of the T cell line is indicated below the heatmap. The logFC is indicated in blue on the respective cells. The (a) BF and (b) logFC of the genes identified in Figure 6.1 are shown in a heatmap where the colour scale indicates the BF/logFC. The BF/logFC of genes identified to be of interest or control genes from the original screens are also shown [80, 81]. Hierarchical clustering done using the wardD2 method.

6.3 CRISPR-Cas9 Editing of K562 Cells

6.3.1 Effect of CRISPR Editing on Viability

The six genes that were identified to be both upregulated in leukaemia and drop-out in both of the genome-wide CRISPR screens [78, 79] were focussed on as putative targets of interest. In order to target these genes, a CRISPR-Cas9 strategy was designed and optimised in the leukaemic K562 cell line, with the ultimate goal of targeting primary AML cells. Firstly, small guide (sg) RNAs were designed against all six genes, with the aim of introducing insertions and deletions (indels) into a coding exon of each gene and knocking-out the functional protein. 5 sgRNAs were selected for testing of editing of K562 cells *in vitro* (Table A.3 and see Chapter 2.16 for full details).

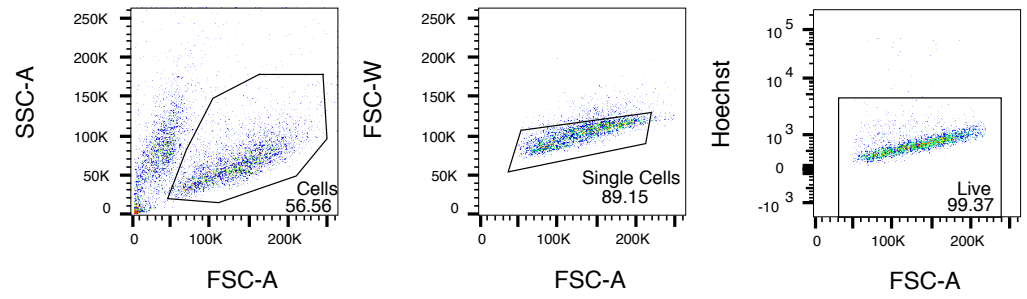
Each sgRNA was cloned into the PX458 plasmid containing the Cas9 protein, and K562 cells were transfected by electroporation with the plasmid encoding the sgRNA and Cas9. Results are shown for sgRNAs that could be successfully cloned into PX458 and transfected into K562 cells. Firstly, the potential effect of transfection with sgRNA on cell death was investigated. If highly efficient editing of a gene essential to cellular biology were to occur, major cell death would be expected. Figure 6.4 shows representative flow cytometry analysis of viability of K562 cells

post-transfection. It was seen that transfection with a plasmid (either PX458 or the control pMax plasmid encoding GFP) caused a slight decrease in morphologically live cells as compared with a mock transfection with no DNA (Figure 6.4).

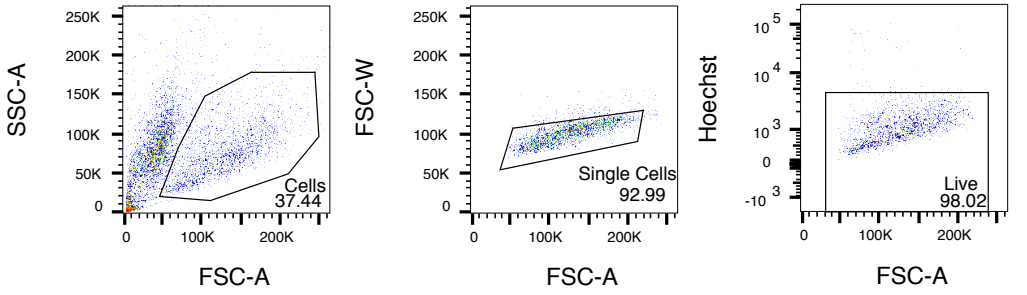
When looking at the average percentage of live cells from total cells, it was seen that at 24 hours post-transfection there was a significant reduction in the percentage of live cells across all sgRNA transfections (Figure 6.5a). The reduction in live cells following pMax transfection was not statistically significant, most likely due to the variability in pMax transfection viabilities. However, at 48 hours post-transfection, only the transfections with sgRNA against *CLP1*, *COPS2* and *WDR26* remained significantly reduced in viability (Figure 6.5b), suggesting that the negative effect on viability was only transient.

Interestingly, when comparing the percentage of live cells to the transfection with pMax, the transfection with sgRNA against *COPS2* showed a significant increase in viability 48 hours post-transfection (Figure 6.5b). This implies that either the targeting of *COPS2* may have conferred an advantage in viability, or that the targeted cells may have died out, leaving behind non-targeted cells that had increased viability.

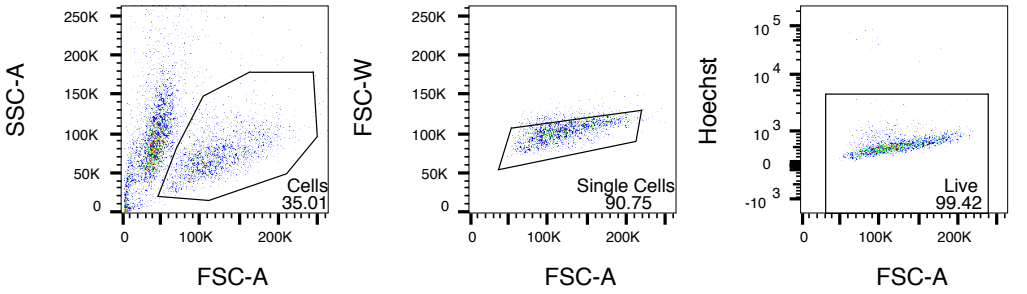
When looking at the transfections using a range of sgRNAs against the 6 genes of interest, typically a moderate increase in viability was seen at 48 hours post-transfection, when compared to 24 hours post-transfection (Figure 6.6). In general, a similar change in viability was seen across all sgRNAs used for a gene (Figure 6.6).



(a) Mock Transfection

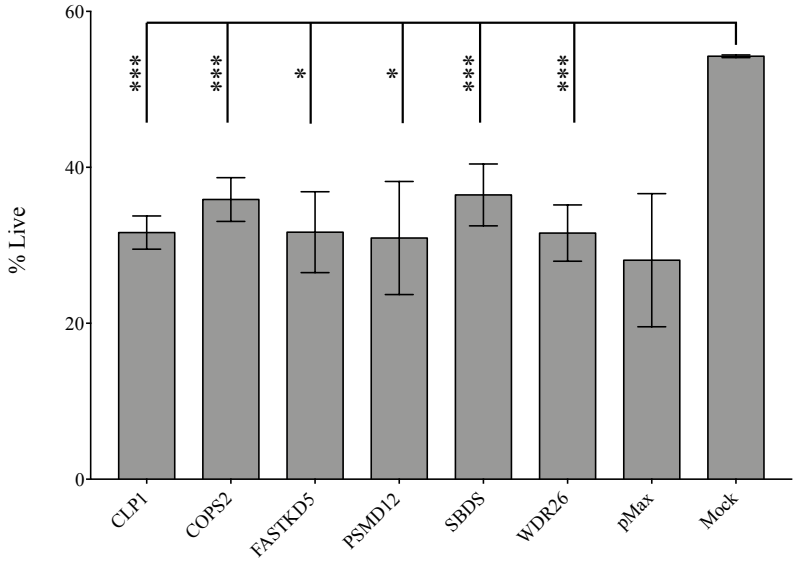


(b) pMax Transfection

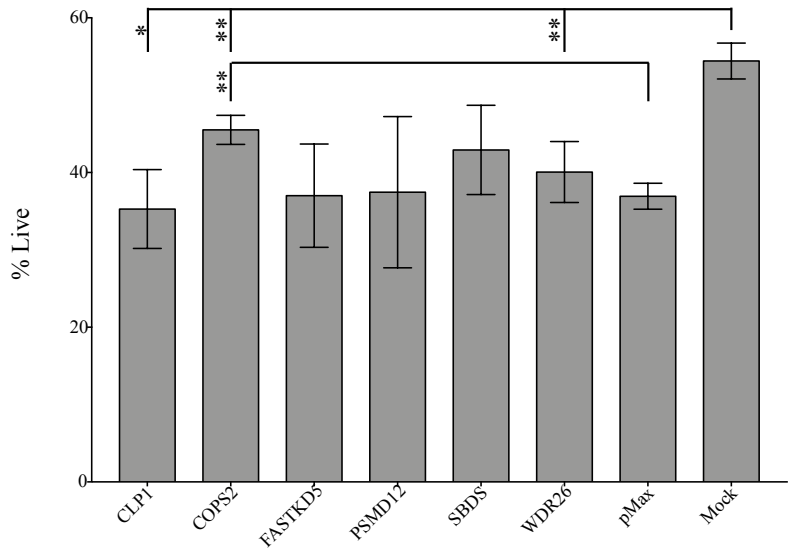


(c) sgRNA *CLP1* Transfection

Figure 6.4: Effect of CRISPR-Cas9 Transfection of K562 Cells on Viability. K562 cells were transfected with PX458 plasmids containing sgRNA against *CLP1*, *CLP1*, *COPS2*, *FASTKD5*, *PSMD12*, *SBDS*, *WDR26*. As controls, K562 cells were mock transfected (i.e. electroporated with no plasmid), and transfected with the pMax GFP plasmid (Lonza). Transfected cells were analysed by flow cytometry at 24 and 48 hours post-transfection. Representative profiles are shown from (a) a mock transfection, (b) a transfection with pMax and (c) transfection with sgRNA against *CLP1*. Frequencies indicate the frequency of the parent gate.

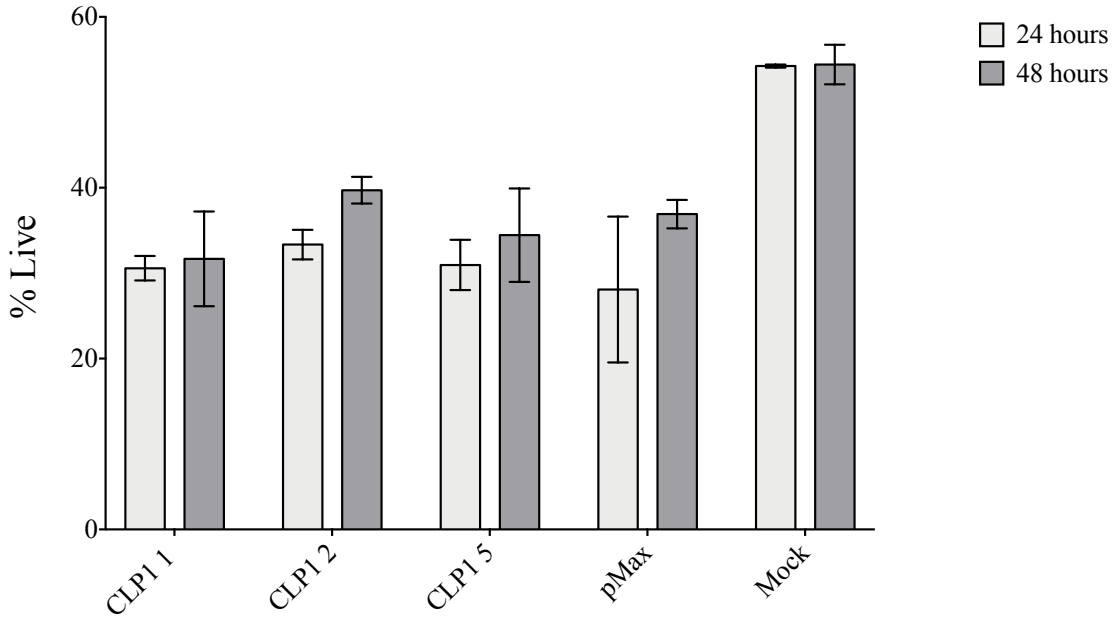


(a) 24 hours

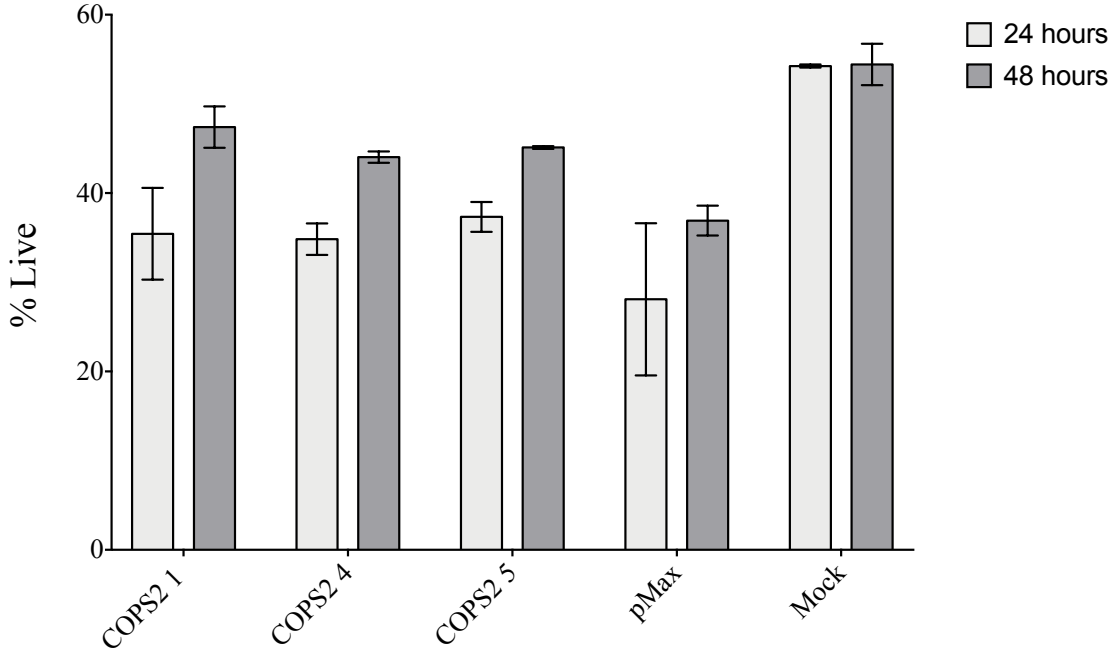


(b) 48 hours

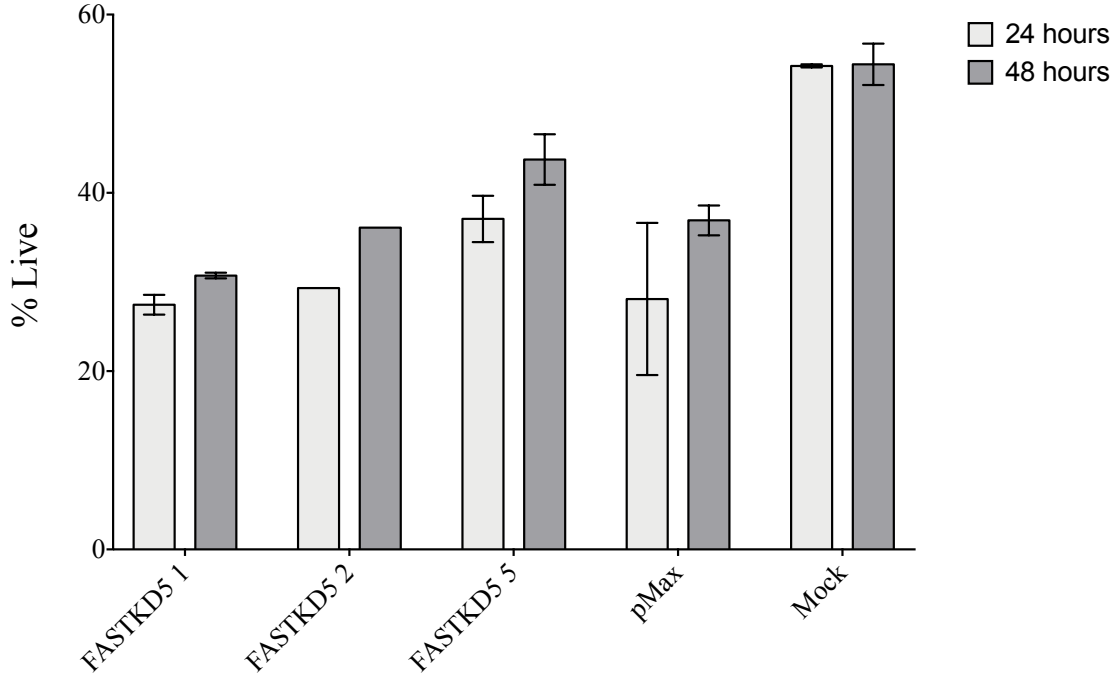
Figure 6.5: Summary of Effect of CRISPR-Cas9 Transfection of K562 Cells on Viability. K562 cells were transfected with PX458 plasmids containing sgRNA against *CLP1*, *COPS2*, *FASTKD5*, *PSMD12*, *SBDS*, *WDR26*. As controls, K562 cells were mock transfected (i.e. electroporated with no plasmid), and transfected with the pMax GFP plasmid (Lonza). 3–5 sgRNAs were tested per gene and 2 technical replicates per sgRNA carried out. Transfected cells were analysed by flow cytometry at 24 and 48 hours post-transfection, and the proportion of live cells out of total events (% Live) calculated. Shown is the average proportion of live cells across the sgRNAs tested and technical replicates. p-adjusted value > 0.05: not significant, p-adjusted value = 0.01–0.05: significant (*), p-adjusted value = 0.001–0.01, very significant (**) and p-adjusted value $<$ 0.001: extremely significant (***). (t-test with Bonferroni correction for multiple testing).



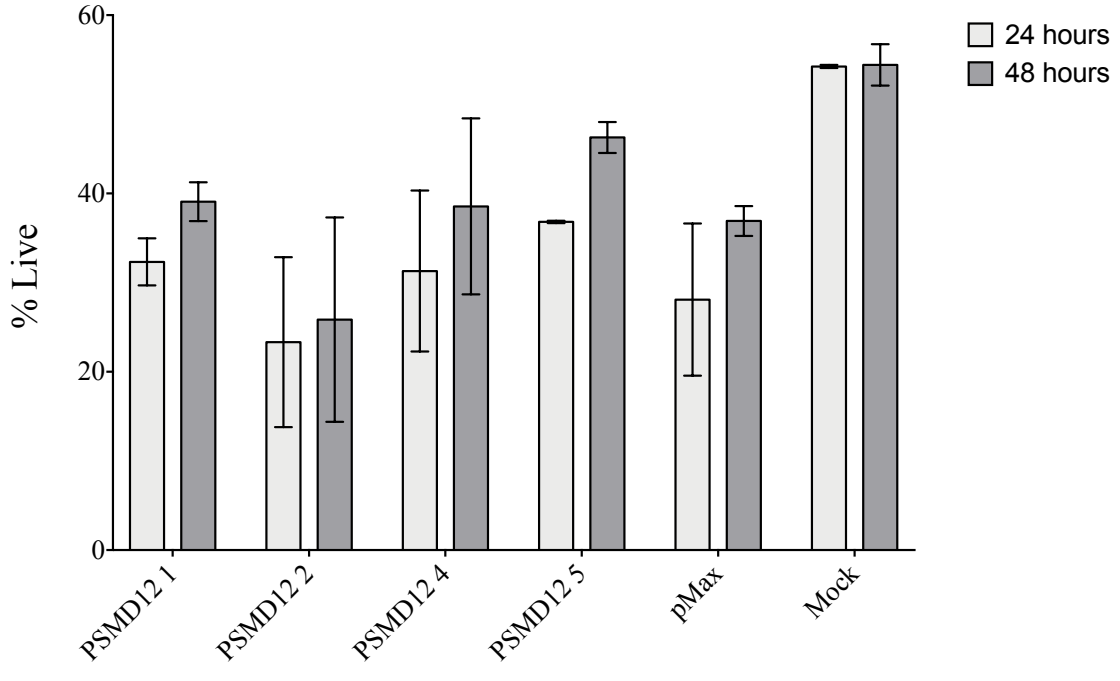
(a) *CLP1*



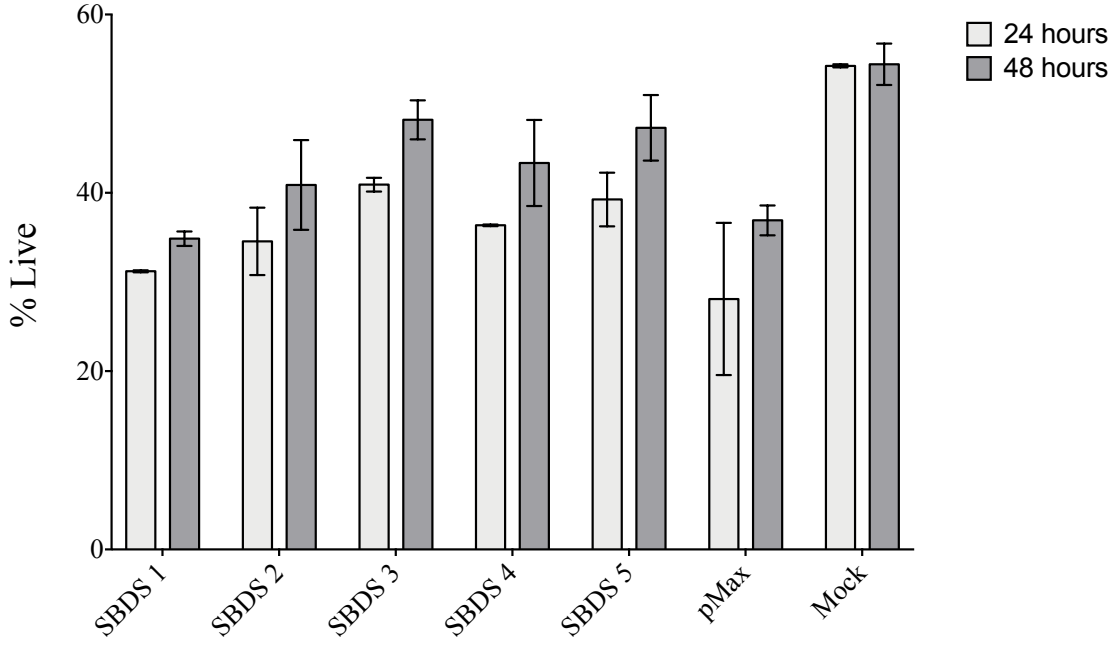
(b) *COPS2*



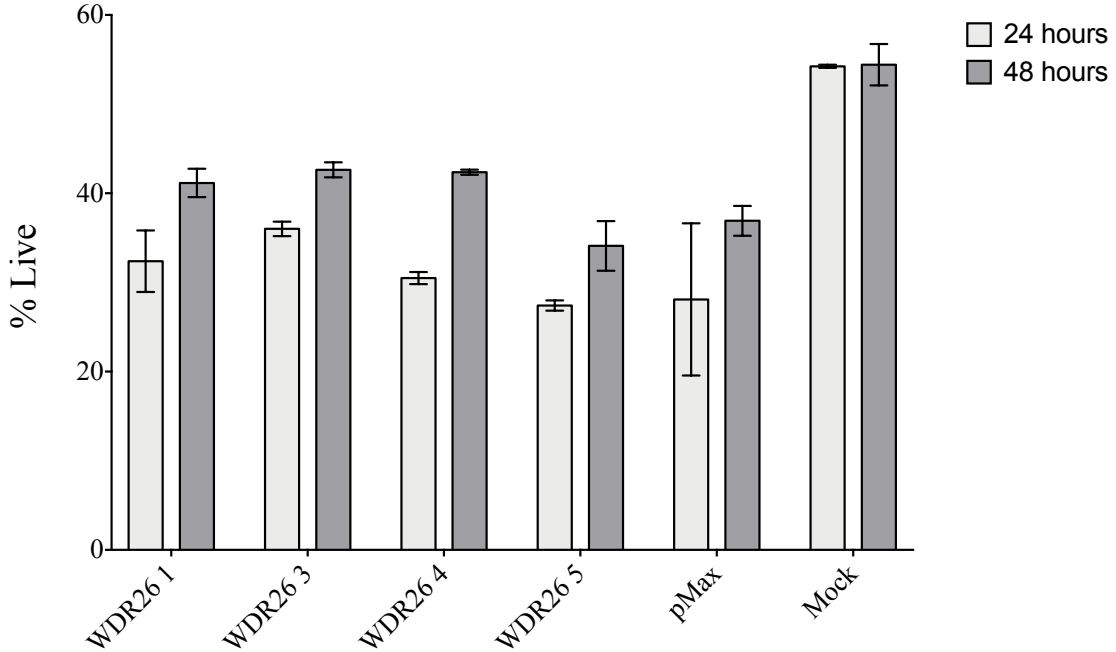
(c) *FASTKD5*



(d) *PSMD12*



(e) *SBDS*



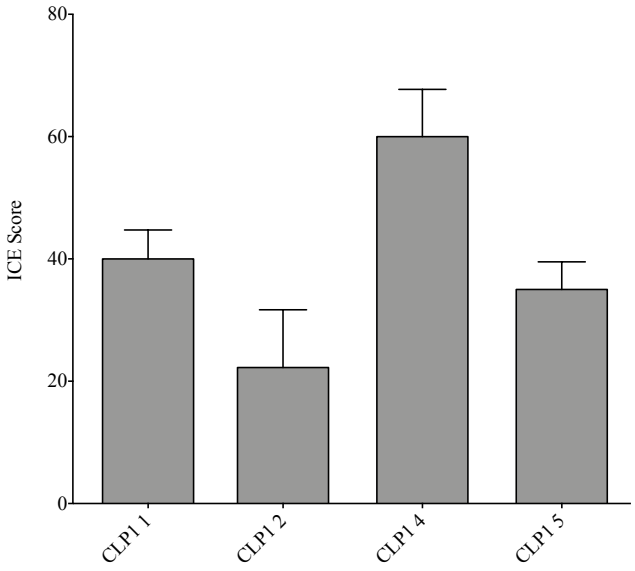
(f) *WDR26*

Figure 6.6: CRISPR-Cas9 Transfection of K562 Cells: Effect of sgRNAs on Viability. K562 cells were transfected with PX458 plasmids containing sgRNA against (a) *CLP1*, (b) *COPS2*, (c) *FASTKD5*, (d) *PSMD12*, (e) *SBDS* and (f) *WDR26*. As controls, K562 cells were mock transfected (i.e. electroporated with no plasmid), and transfected with the pMax GFP plasmid (Lonza). 3–5 sgRNAs were tested per gene and 2 technical replicates per sgRNA carried out. Transfected cells were analysed by flow cytometry at 24 and 48 hours post-transfection, and the proportion of live cells out of total events (% Live) calculated. Shown is the average proportion of live cells across technical replicates. sgRNAs are as indicated in Table A.3.

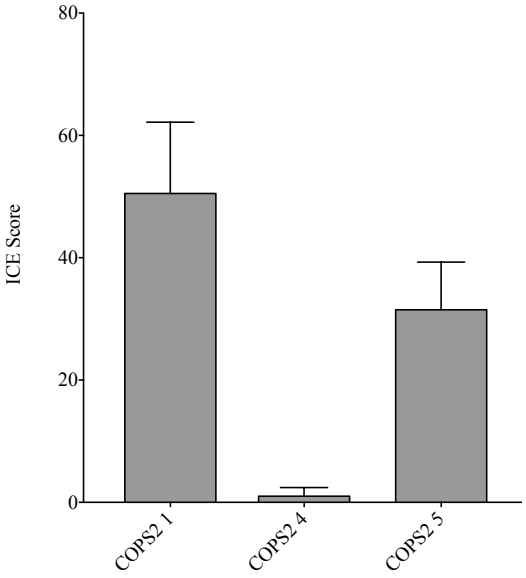
6.3.2 Inference of CRISPR Editing Analysis

Next the extent of gene editing following transfection of K562 cells with sgRNA was assessed. Gene editing was assessed using Inference of CRISPR Editing (ICE) Scores [90], where the higher the score, the higher the extent of editing. All ICE scores are shown in Table A.29, and the average ICE scores (across technical replicates and forward and reverse Sanger sequencing reads) are shown in Figure 6.7. ICE scores are shown for all sgRNAs whose potentially edited gene region could be amplified and successful Sanger sequencing obtained.

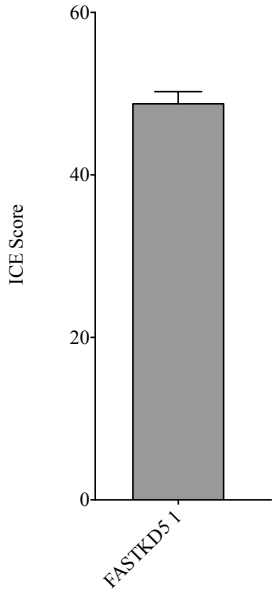
A range of ICE scores was seen across sgRNAs for a gene, with scores ranging from 5 to 77. The sgRNAs with the highest ICE scores for each gene were: *CLP1* 4, *COPS2* 1, *FASTKD5* 1, *PSMD12* 2 and *WDR26* 2. The sgRNAs targeting *SBDS* did not achieve sufficiently high enough ICE scores to be taken forward. ICE scores could only be elucidated for one sgRNA targeting *FASTKD5* due to mutations already being present in the *FASTKD5* gene in the K562 cell line used.



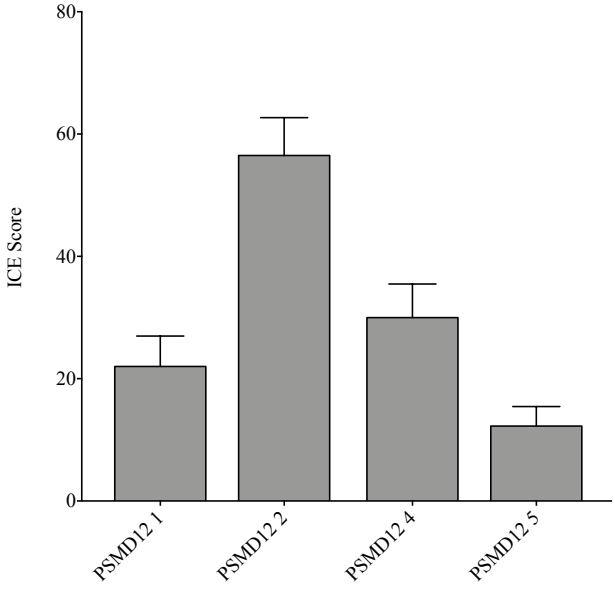
(a) *CLP1*



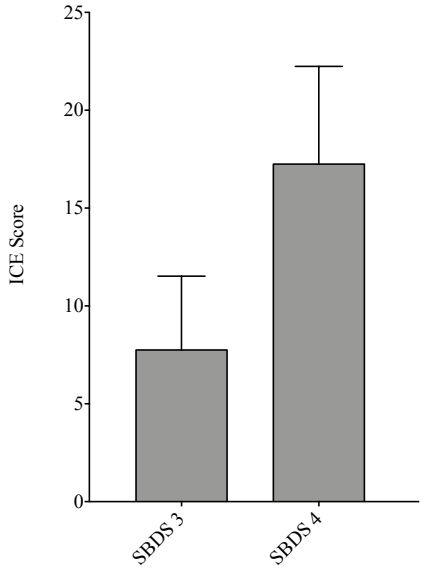
(b) *COPS2*



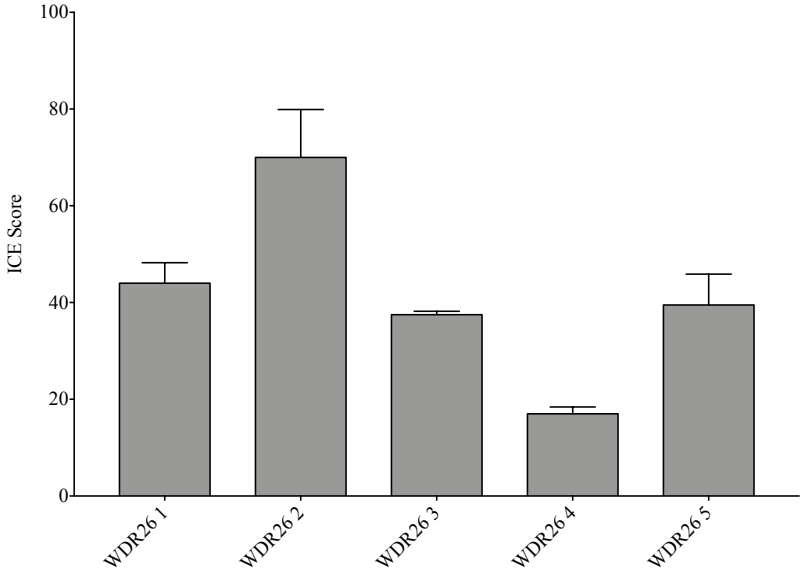
(c) *FASTKD5*



(d) *PSMD12*



(e) *SBDS*

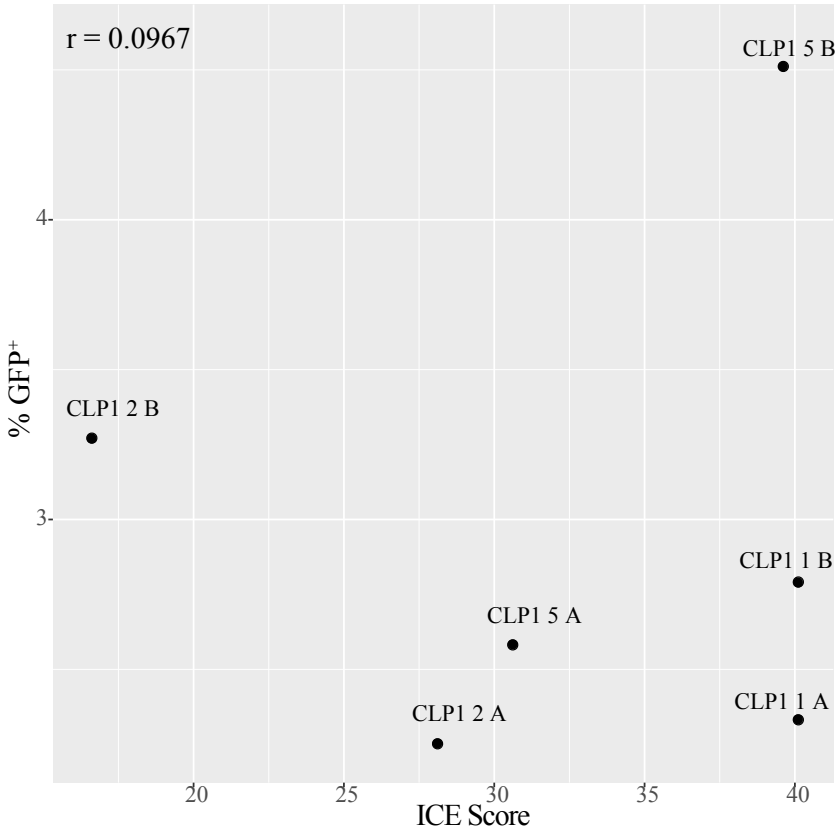


(f) *WDR26*

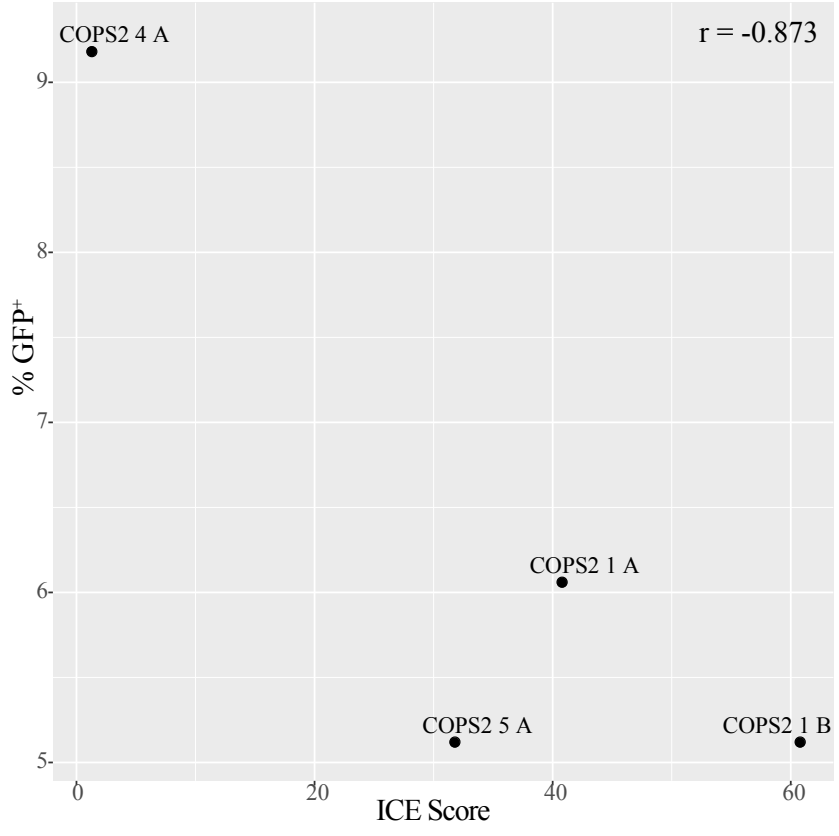
Figure 6.7: CRISPR Editing of K562 Cells: Summary of ICE Scores. K562 cells were transfected with PX458 plasmids containing sgRNA against *CLP1*, *COPS2*, *FASTKD5*, *PSMD12*, *SBDS* and *WDR26*. 48 hours post-transfection DNA was extracted for sequencing to assess gene editing. Gene editing was assessed using ICE Scores [90]. Shown are the average ICE scores across technical replicates and forward and reverse Sanger sequencing reads. See Table A.29 for ICE scores. sgRNAs are as indicated in Table A.3.

6.3.3 Relationship Between ICE Scores and GFP Positivity

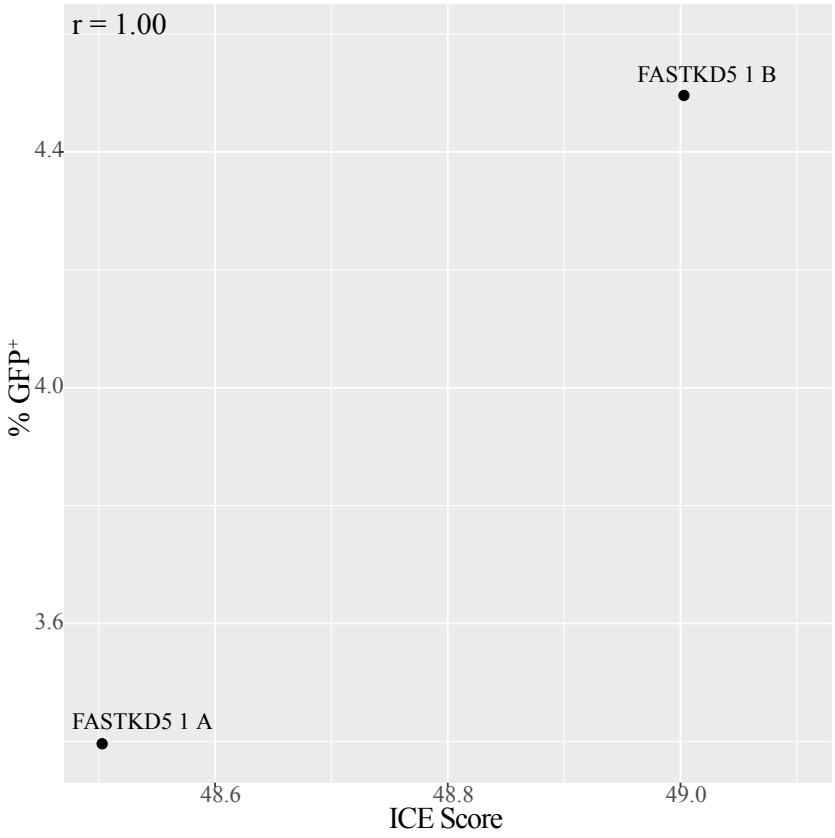
The PX458 plasmid used in transfections also encoded a GFP, and the relationship between the ICE score and GFP positivity was investigated. Although GFP positivity does not indicate the extent of editing, it is associated with the transfection efficiency. The ICE score and GFP positivity of transfected cells are shown in Figure 6.8, and it was seen that the relationship between these factors was variable. For some of the sgRNAs there was either no correlation or even a negative correlation between ICE score and GFP positivity (Figures 6.8a, 6.8b and 6.8f). However for others, there was a strong positive correlation, where the higher the GFP positivity, the higher the ICE score (Figures 6.8c, 6.8d and 6.8e).



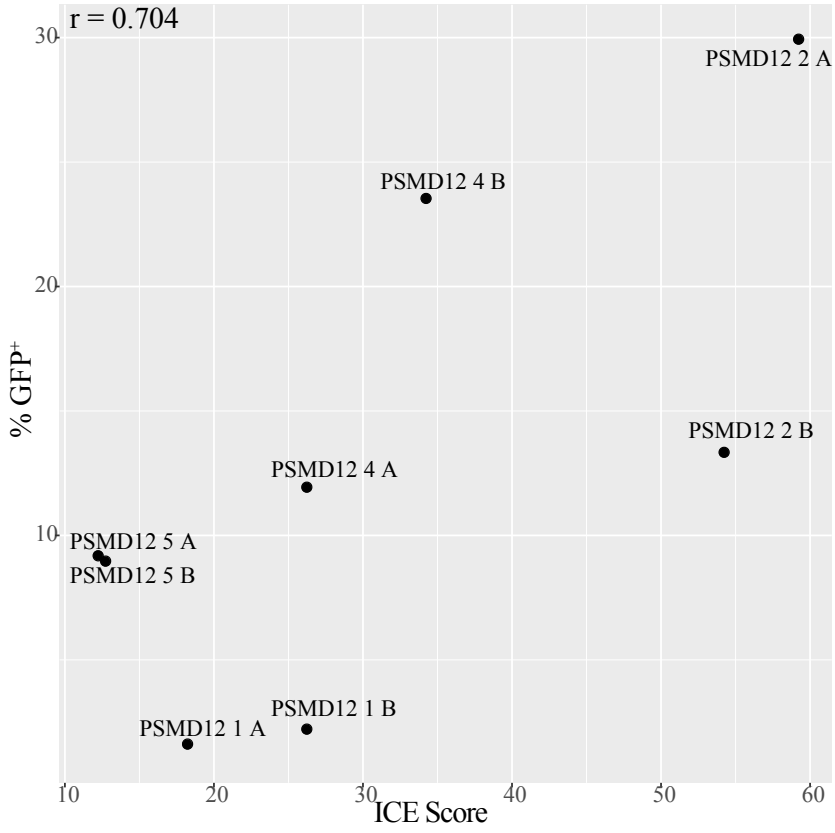
(a) *CLP1*



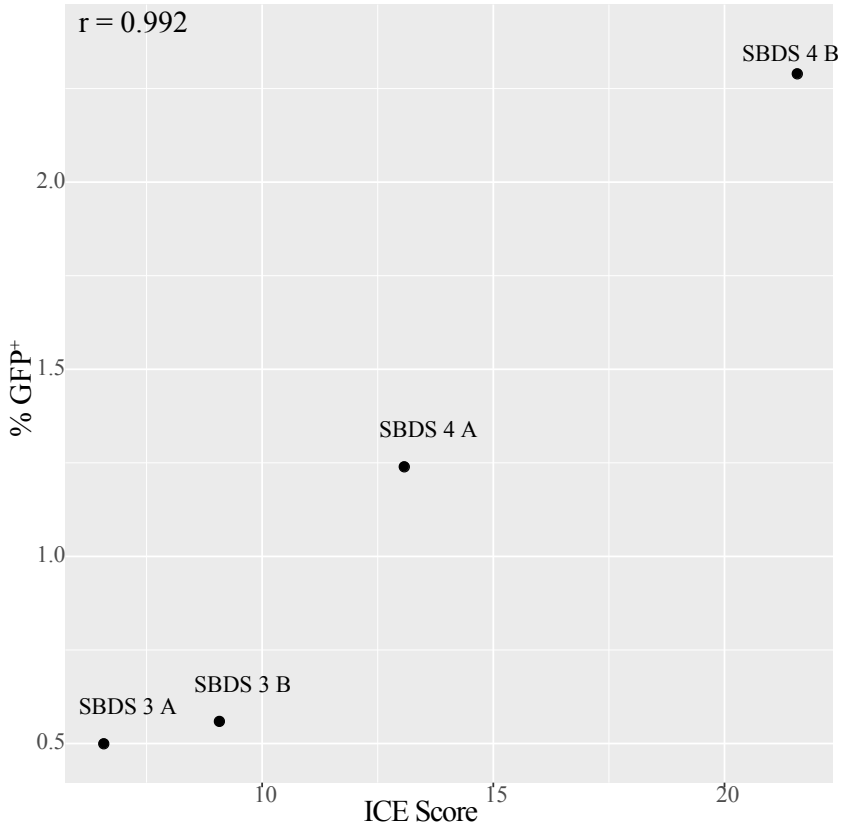
(b) *COPS2*



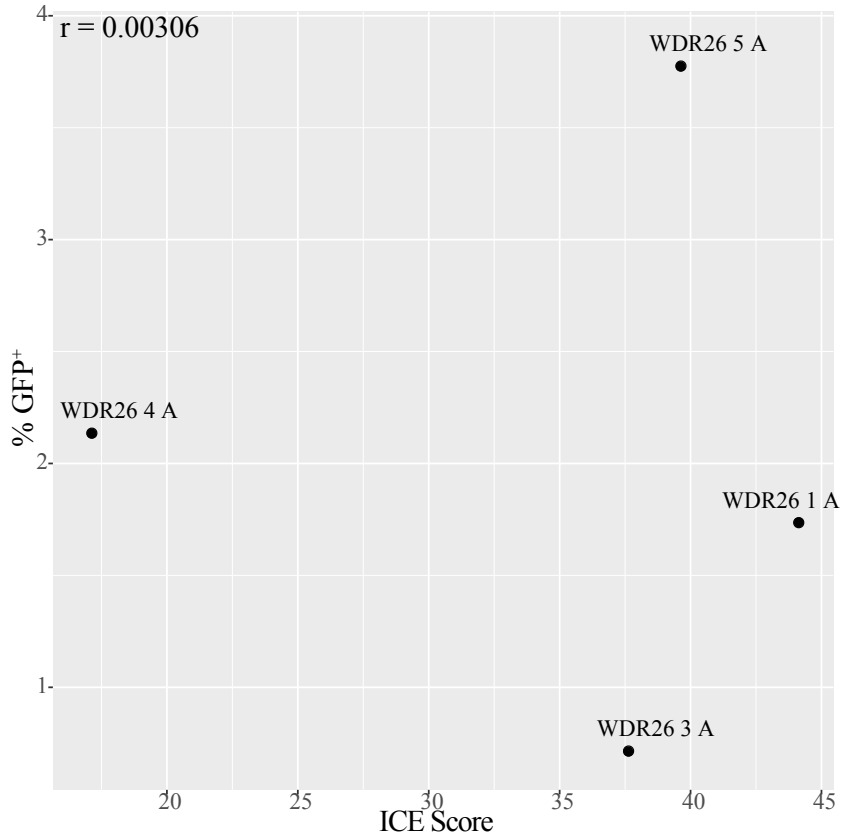
(c) *FASTKD5*



(d) *PSMD12*



(e) *SBDS*



(f) *WDR26*

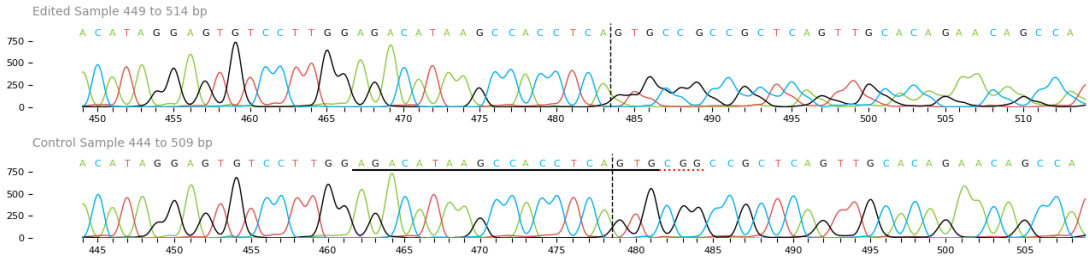
Figure 6.8: Correlation between GFP$^+$ Cells and ICE Scores. K562 cells were transfected with PX458 plasmids containing sgRNA against (a) *CLP1*, (b) *COPS2*, (c) *FASTKD5*, (d) *PSMD12*, (e) *SBDS* and (f) *WDR26*. 48 hours post-transfection the cells were analysed by flow cytometry to assess GFP positivity, and DNA extracted for sequencing to assess gene editing. Gene editing was assessed using ICE Scores [90]. Shown are the average ICE scores across forward and reverse reads from Sanger sequencing. A and B represent technical replicates. The sgRNAs are as indicated in Table A.3. r indicates the Pearson correlation co-efficient.

6.3.4 Indels Introduced in K652 by CRISPR Editing

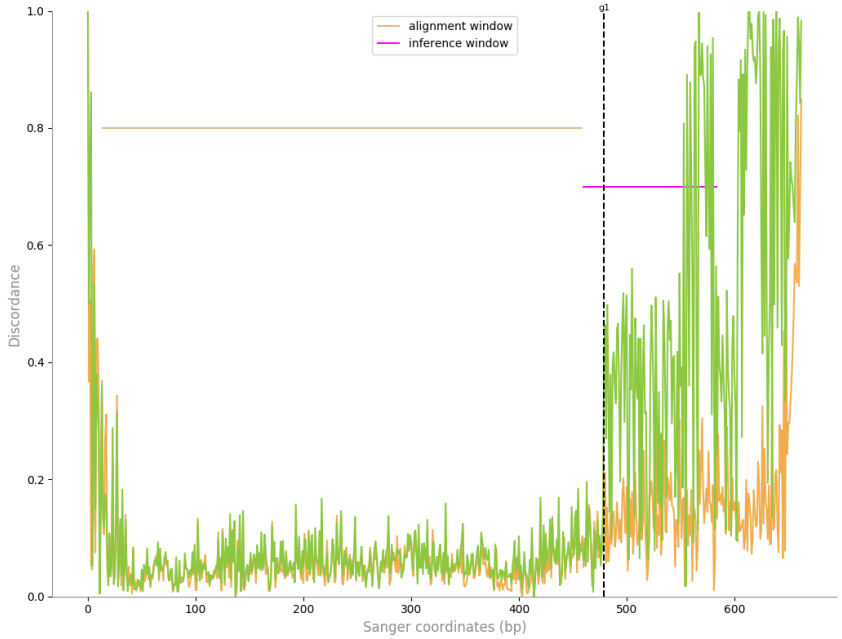
To gain an insight into the consequences of CRISPR editing, the nature of the generated indels in K562 cells was examined. ICE software [90] was used to compare control (unedited) and CRISPR edited sequences around the sgRNA site for each gene of interest, and to identify the discordance of the alignment between the control and CRISPR edited sequences. ICE software was also used to infer the distribution of indels in the CRISPR edited samples (Figures 6.9, 6.10, 6.11, 6.12, 6.13 and 6.14).

The effect of CRISPR editing was clearly seen both by comparing the sequence traces between control and edited sequences and also by looking at the discordance between the sequences. For sgRNAs with highly efficient editing, for example sgRNA *PSMD12* 2 (Figure 6.12a and 6.12b), the deterioration of the sequencing trace and the increase in discordance can be seen immediately following the cut site. However, for sgRNAs with low editing efficiencies, such as *SBDS* 4, this was not seen (Figure 6.13a and 6.13b).

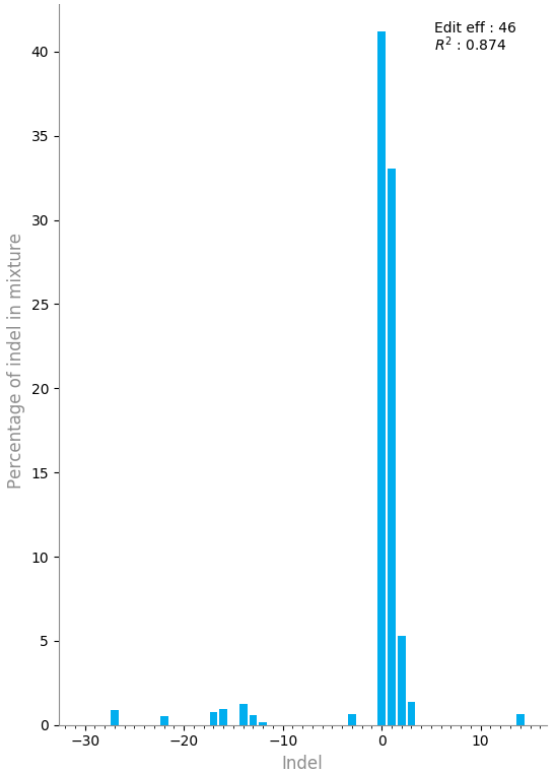
When looking at the indel distribution, it was seen that the majority of indels were small in size. For most sgRNAs the majority of indels were less than 10 nucleotides in size (Figures 6.9c, 6.10c, 6.11c and 6.14c). However, for sgRNA *PSMD12* 2, a minor population of cells showed deletions larger than 10 nucleotides (Figure 6.12c).



(a) *CLP1*: Control and CRISPR Edited Sequences

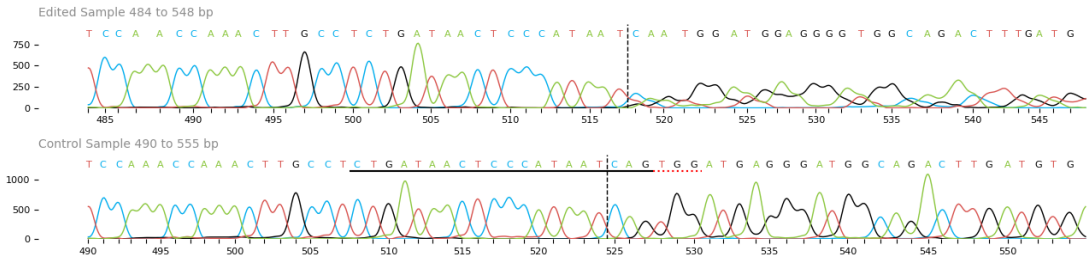


(b) *CLP1*: Discordance Between Control and CRISPR Edited Sequences

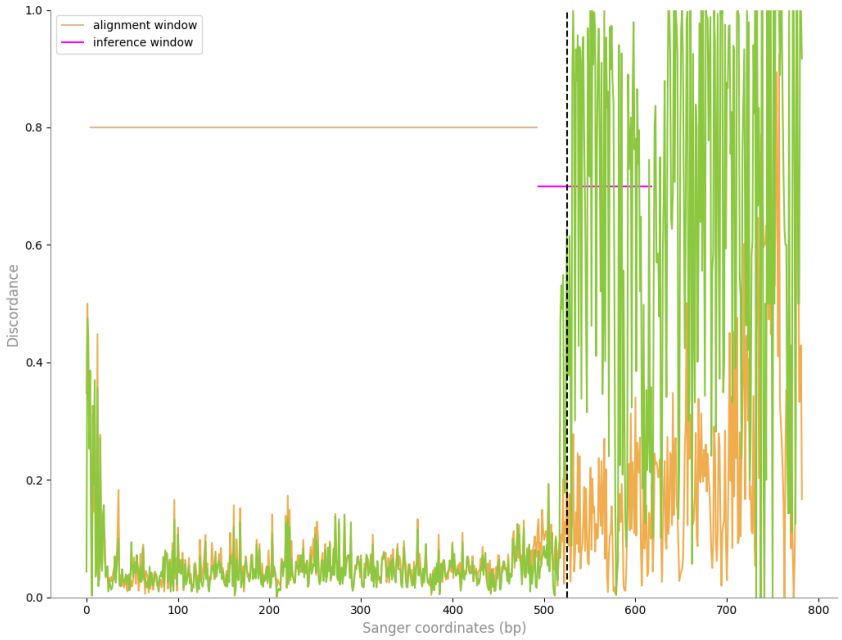


(c) *CLP1*: Indel Distribution

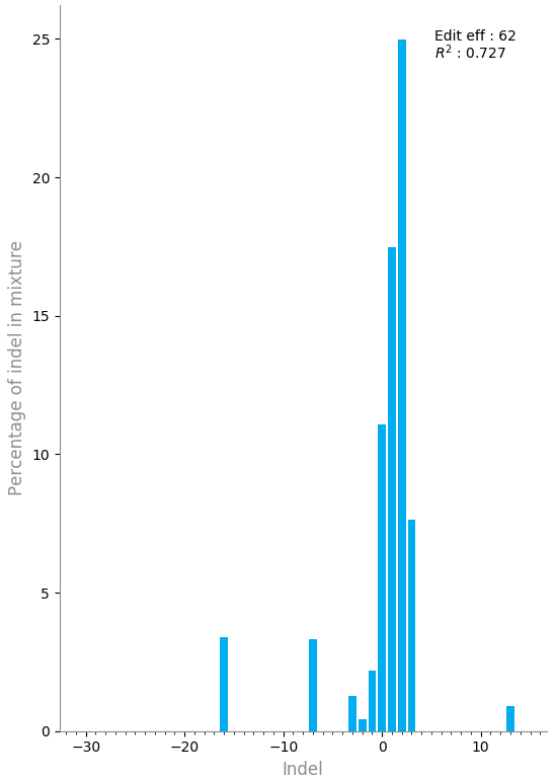
Figure 6.9: CRISPR Editing of *CLP1* in K562 Cells. K562 cells were transfected with PX458 plasmid containing sgRNA against *CLP1*, and 48 hours post-transfection DNA was extracted for sequencing to assess gene editing. Gene editing was assessed using ICE Scores [90]. (a) The Sanger sequence traces for the control and CRISPR edited sequences around the sgRNA site in *CLP1*. The sgRNA site and PAM site are indicated by the black and red lines respectively. The black dotted line represents the actual cut site. (b) The discordance of the alignment per base between the control (orange line) and edited (green line) sequences in the inference window (the region around the cut site). (c) The inferred distribution of indels in the CRISPR edited sample. Shown are representative plots from editing using the sgRNA CLP1 1. The sgRNAs are as indicated in Table A.3. ICE analysis and plots generated using ICE software (Synthego, [90]).



(a) *COPS2*: Control and CRISPR Edited Sequences

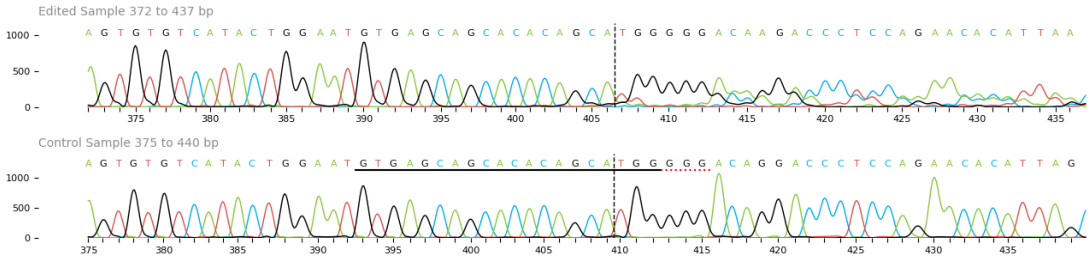


(b) *COPS2*: Discordance Between Control and CRISPR Edited Sequences

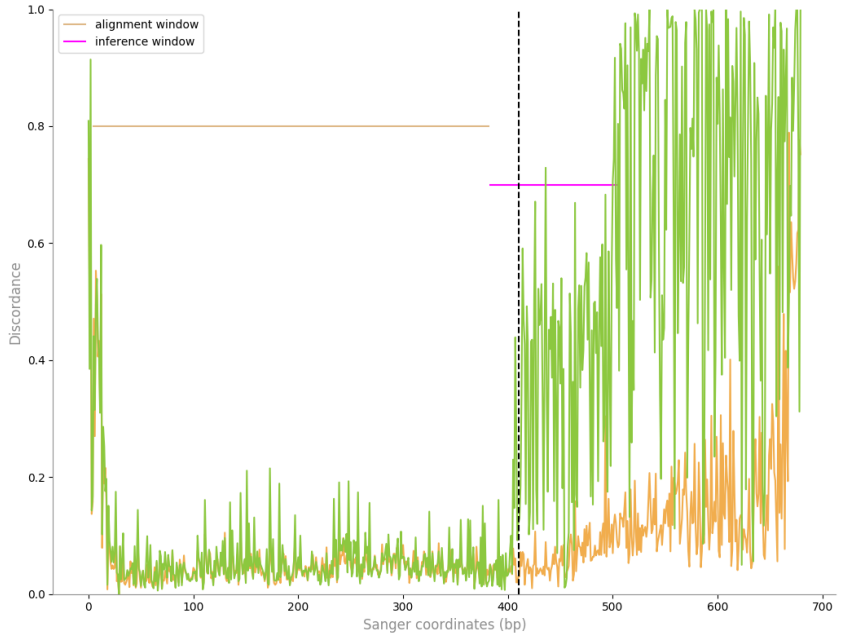


(c) *COPS2*: Indel Distribution

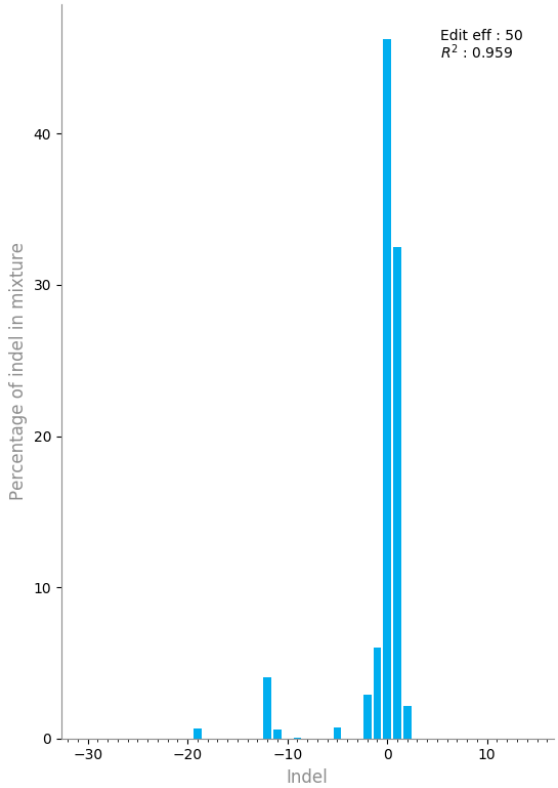
Figure 6.10: CRISPR Editing of *COPS2* in K562 Cells. K562 cells were transfected with PX458 plasmid containing sgRNA against *COPS2*, and 48 hours post-transfection DNA was extracted for sequencing to assess gene editing. Gene editing was assessed using ICE Scores [90]. (a) The Sanger sequence traces for the control and CRISPR edited sequences around the sgRNA site in *COPS2*. The sgRNA site and PAM site are indicated by the black and red lines respectively. The black dotted line represents the actual cut site. (b) The discordance of the alignment per base between the control (orange line) and edited (green line) sequences in the inference window (the region around the cut site). (c) The inferred distribution of indels in the CRISPR edited sample. Shown are representative plots from editing using the sgRNA COPS2 1. The sgRNAs are as indicated in Table A.3. ICE analysis and plots generated using ICE software (Synthego, [90]).



(a) *FASTKD5*: Control and CRISPR Edited Sequences

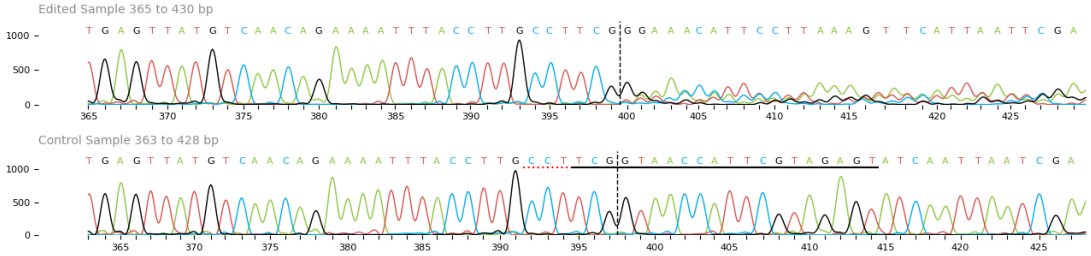


(b) *FASTKD5*: Discordance Between Control and CRISPR Edited Sequences

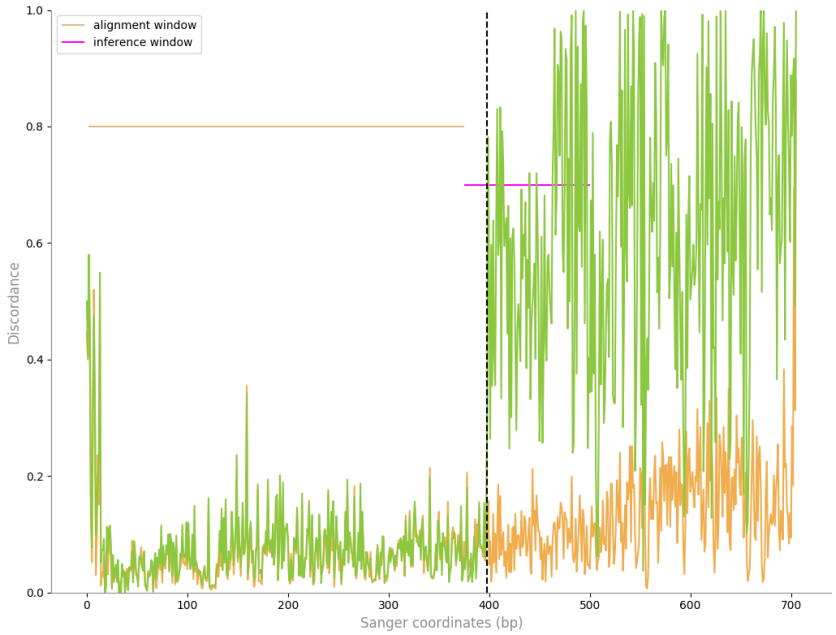


(c) *FASTKD5*: Indel Distribution

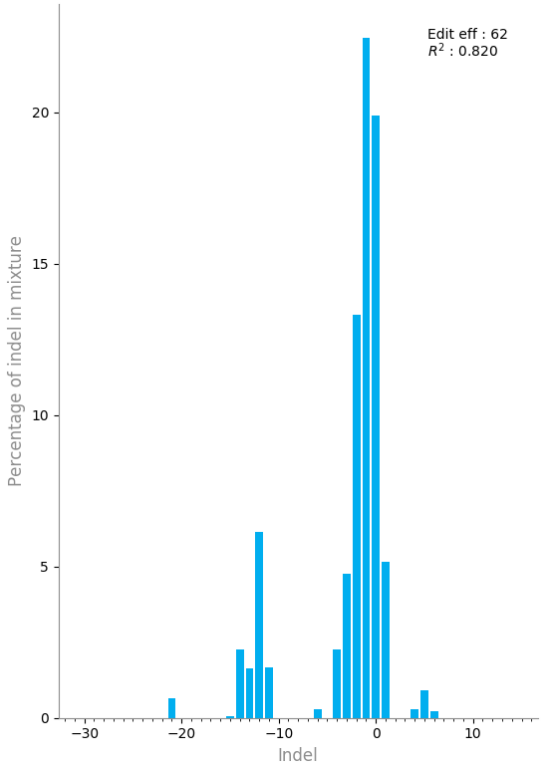
Figure 6.11: CRISPR Editing of *FASTKD5* in K562 Cells. K562 cells were transfected with PX458 plasmid containing sgRNA against *FASTKD5*, and 48 hours post-transfection DNA was extracted for sequencing to assess gene editing. Gene editing was assessed using ICE Scores [90]. (a) The Sanger sequence traces for the control and CRISPR edited sequences around the sgRNA site in *FASTKD5*. The sgRNA site and PAM site are indicated by the black and red lines respectively. The black dotted line represents the actual cut site. (b) The discordance of the alignment per base between the control (orange line) and edited (green line) sequences in the inference window (the region around the cut site). (c) The inferred distribution of indels in the CRISPR edited sample. Shown are representative plots from editing using the sgRNA FASTKD5 1. The sgRNAs are as indicated in Table A.3. ICE analysis and plots generated using ICE software (Synthego, [90]).



(a) *PSMD12*: Control and CRISPR Edited Sequences

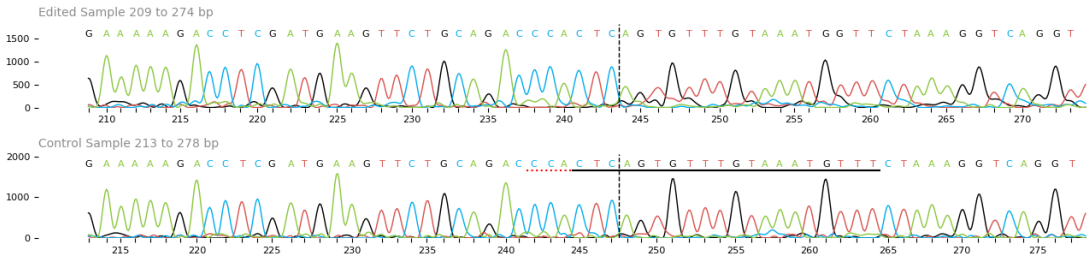


(b) *PSMD12*: Discordance Between Control and CRISPR Edited Sequences

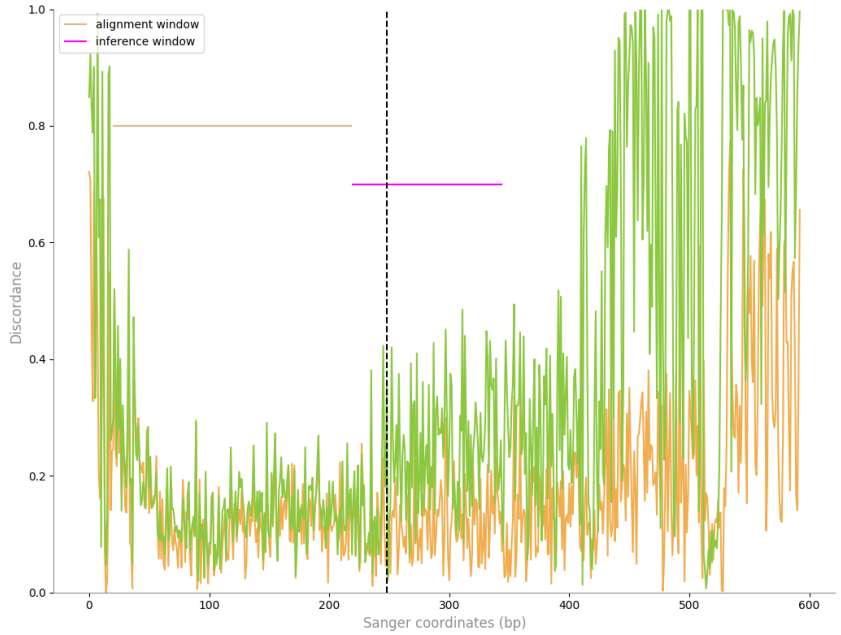


(c) *PSMD12*: Indel Distribution

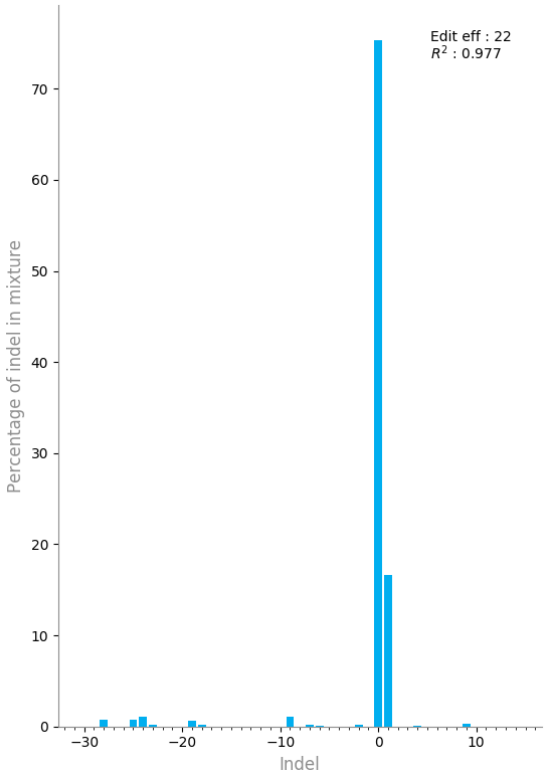
Figure 6.12: CRISPR Editing of *PSMD12* in K562 Cells. K562 cells were transfected with PX458 plasmid containing sgRNA against *PSMD12*, and 48 hours post-transfection DNA was extracted for sequencing to assess gene editing. Gene editing was assessed using ICE Scores [90]. (a) The Sanger sequence traces for the control and CRISPR edited sequences around the sgRNA site in *PSMD12*. The sgRNA site and PAM site are indicated by the black and red lines respectively. The black dotted line represents the actual cut site. (b) The discordance of the alignment per base between the control (orange line) and edited (green line) sequences in the inference window (the region around the cut site). (c) The inferred distribution of indels in the CRISPR edited sample. Shown are representative plots from editing using the sgRNA PSMD12 2. The sgRNAs are as indicated in Table A.3. ICE analysis and plots generated using ICE software (Synthego, [90]).



(a) *SBDS*: Control and CRISPR Edited Sequences

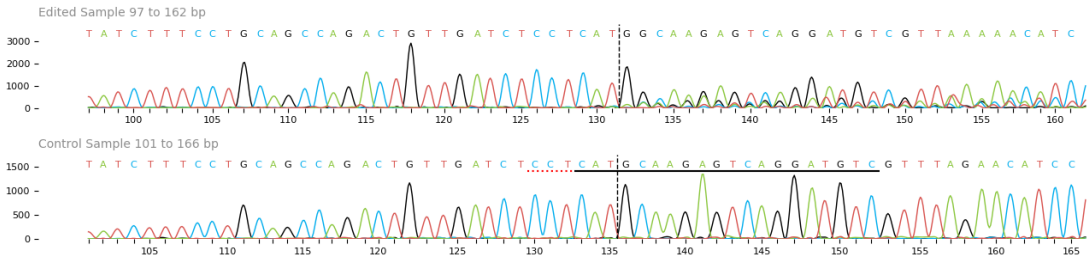


(b) *SBDS*: Discordance Between Control and CRISPR Edited Sequences

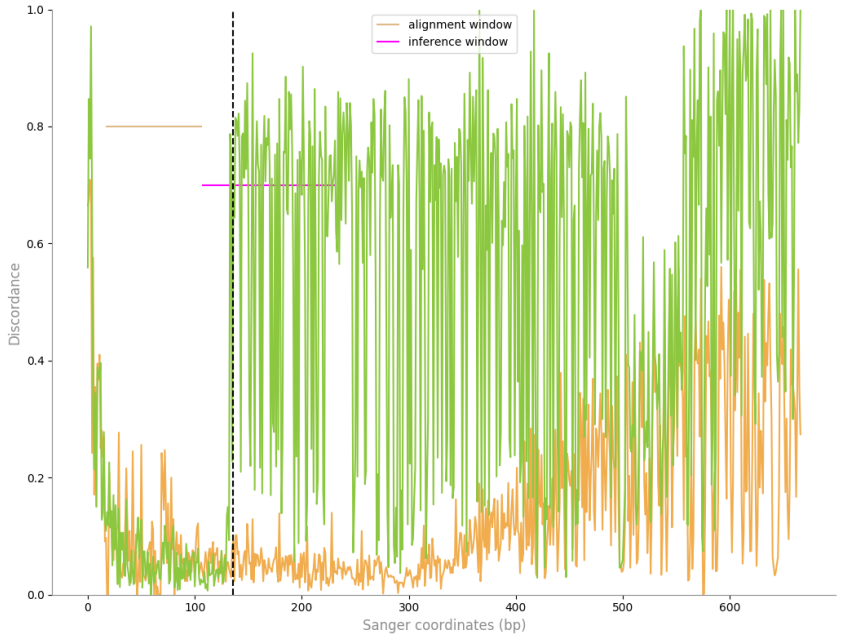


(c) *SBDS*: Indel Distribution

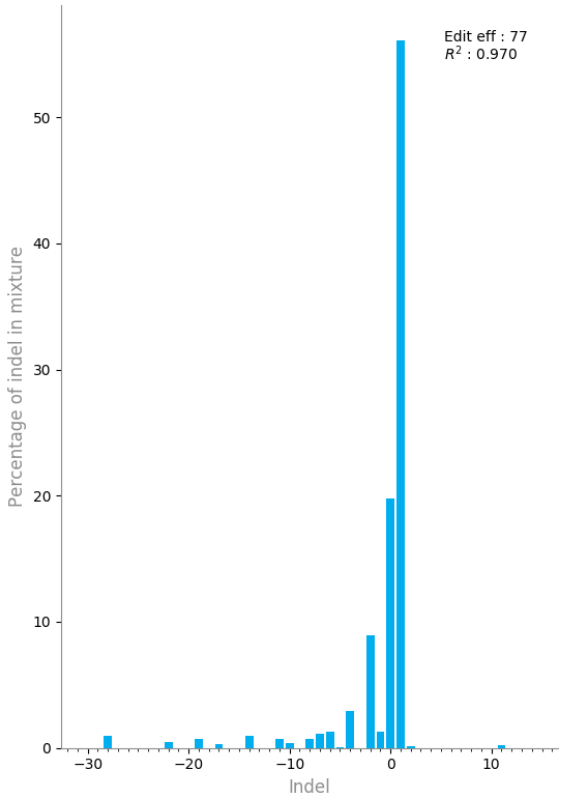
Figure 6.13: CRISPR Editing of *SBDS* in K562 Cells. K562 cells were transfected with PX458 plasmid containing sgRNA against *SBDS*, and 48 hours post-transfection DNA was extracted for sequencing to assess gene editing. Gene editing was assessed using ICE Scores [90]. (a) The Sanger sequence traces for the control and CRISPR edited sequences around the sgRNA site in *SBDS*. The sgRNA site and PAM site are indicated by the black and red lines respectively. The black dotted line represents the actual cut site. (b) The discordance of the alignment per base between the control (orange line) and edited (green line) sequences in the inference window (the region around the cut site). (c) The inferred distribution of indels in the CRISPR edited sample. Shown are representative plots from editing using the sgRNA SBDS 4. The sgRNAs are as indicated in Table A.3. ICE analysis and plots generated using ICE software (Synthego, [90]).



(a) *WDR26*: Control and CRISPR Edited Sequences



(b) *WDR26*: Discordance Between Control and CRISPR Edited Sequences



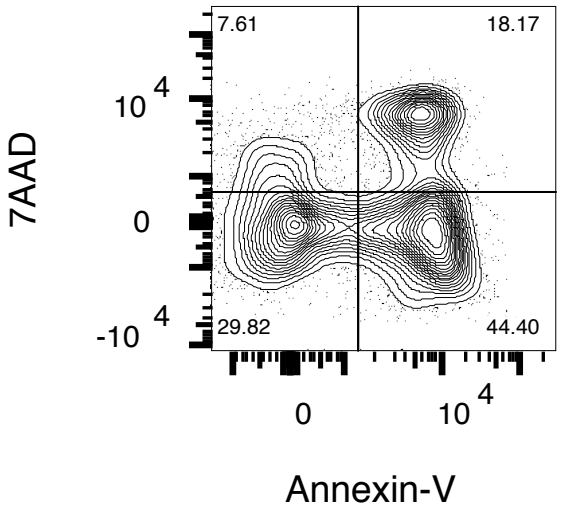
(c) *WDR26*: Indel Distribution

Figure 6.14: CRISPR Editing of *WDR26* in K562 Cells. K562 cells were transfected with PX458 plasmid containing sgRNA against *WDR26*, and 48 hours post-transfection DNA was extracted for sequencing to assess gene editing. Gene editing was assessed using ICE Scores [90]. (a) The Sanger sequence traces for the control and CRISPR edited sequences around the sgRNA site in *WDR26*. The sgRNA site and PAM site are indicated by the black and red lines respectively. The black dotted line represents the actual cut site. (b) The discordance of the alignment per base between the control (orange line) and edited (green line) sequences in the inference window (the region around the cut site). (c) The inferred distribution of indels in the CRISPR edited sample. Shown are representative plots from editing using the sgRNA WDR26 2. The sgRNAs are as indicated in Table A.3. ICE analysis and plots generated using ICE software (Synthego, [90]).

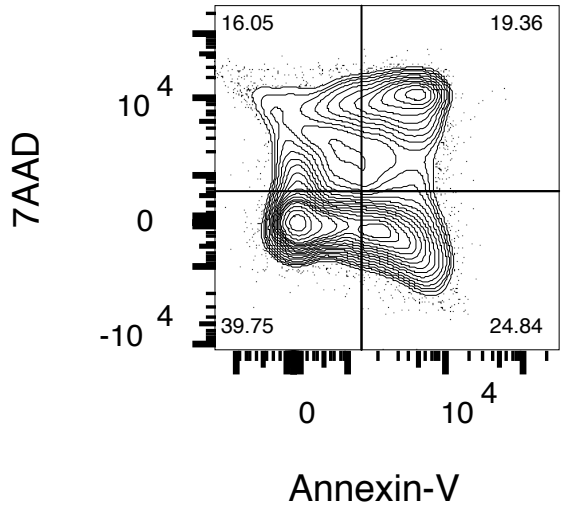
6.4 Optimisation of the *In Vitro* Culture of Primary AML

A challenge faced when aiming to edit primary AML cells is their *in vitro* culture. In order to facilitate the *in vitro* editing of LSCs by CRISPR-Cas9, the culture of primary AML cells was optimised. Firstly, the effects of co-culturing AML cells in the presence of a stromal layer and/or in the presence of compounds previously identified to maintain AML cells *in vitro* were investigated (Figure 6.15 and 6.16). For co-cultures, a human BM derived MSC line, immortalised by the expression of *TERT*, was used [84]. The small compounds Stemregenin 1 (SR1) and UM171 previously identified by Pabst et al. were used in isolation or in combination with MSCs [85] (see Chapter 2.15 for full culture conditions). Either primary AML cells or patient derived xenograft (PDX) cells were used.

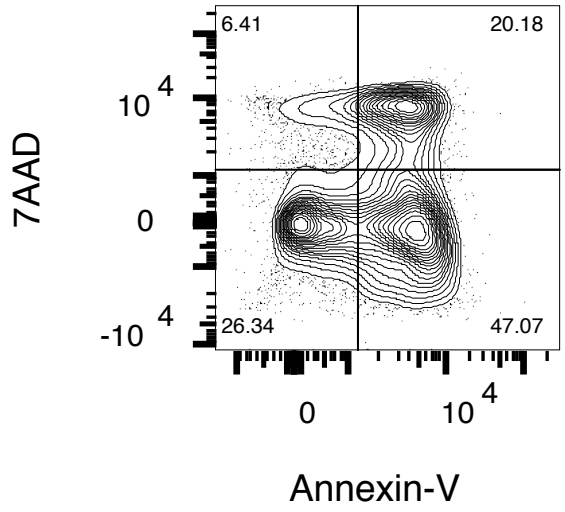
It was first noted that the proportion of cells entering apoptosis (identified as being Annexin-V^+ $7AAD^-$) was potentially reduced when cultured in the presence of MSCs (Figures 6.15 and 6.16). PDX samples BIR344 and BIR517 showed a significant decrease in Annexin-V^+ cells when cultured in the presence of MSCs, either in the presence or absence of SR1 and UM171 (Figures 6.15, 6.16 and Table A.30). This difference was more marked at earlier time-points of the culture and decreased at later time-points. However, for PDX sample STB257 this reduction in Annexin-V^+ cells was not observed.



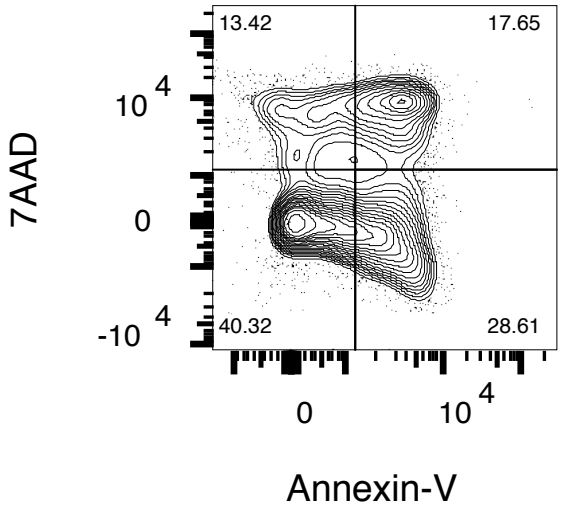
(a) BIR344: No MSC



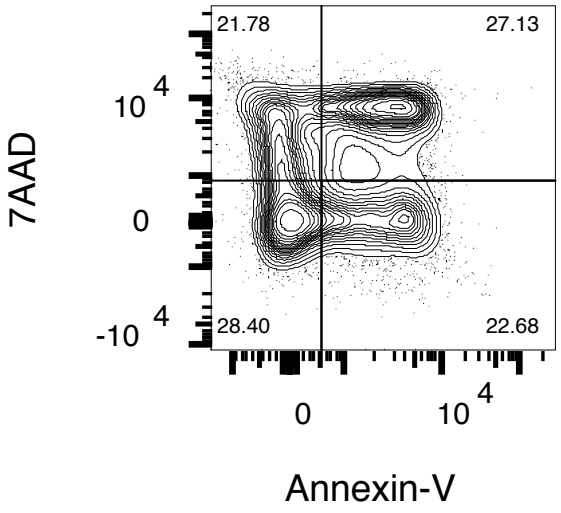
(b) BIR344: MSC



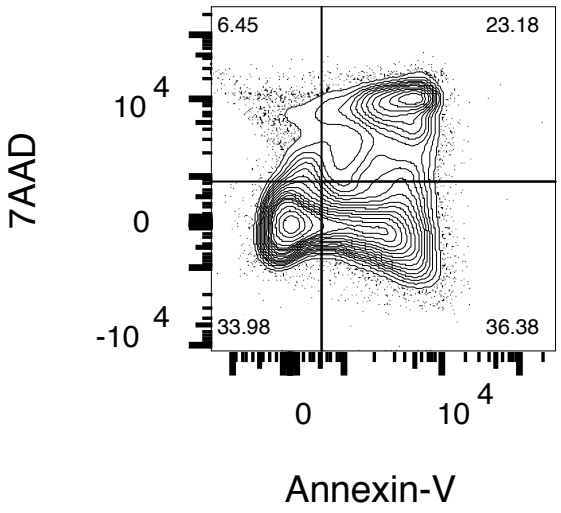
(c) BIR517: No MSC



(d) BIR517: MSC

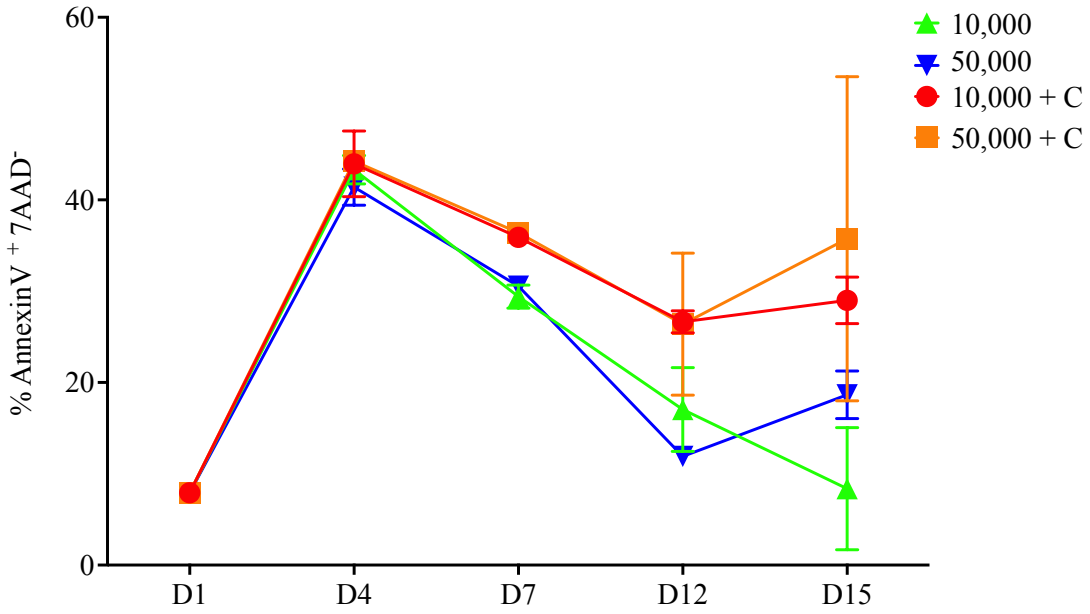


(e) STB257: No MSC

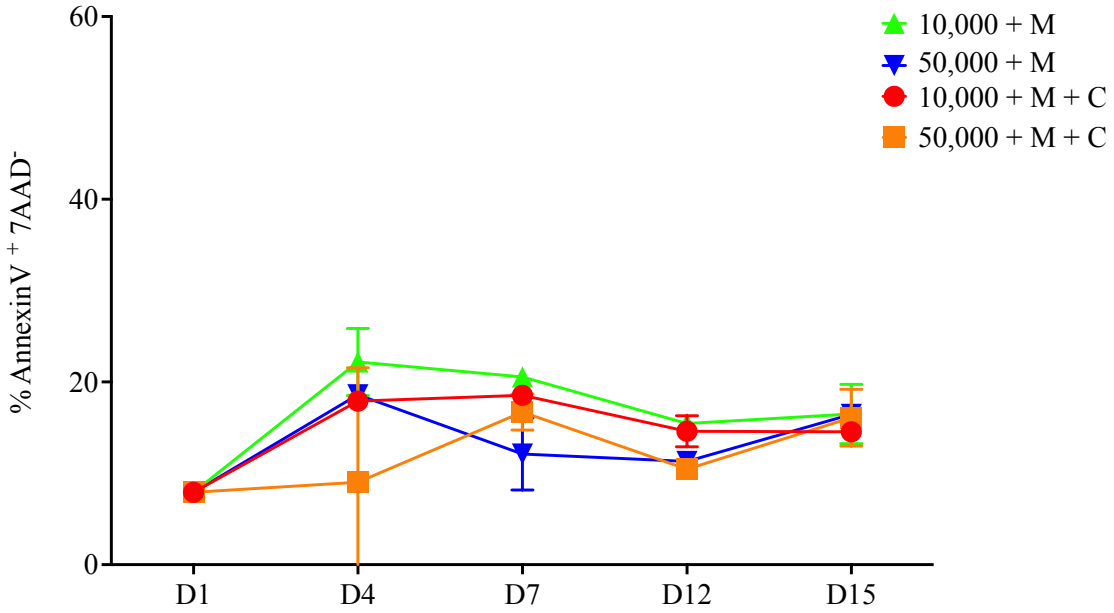


(f) STB257: MSC

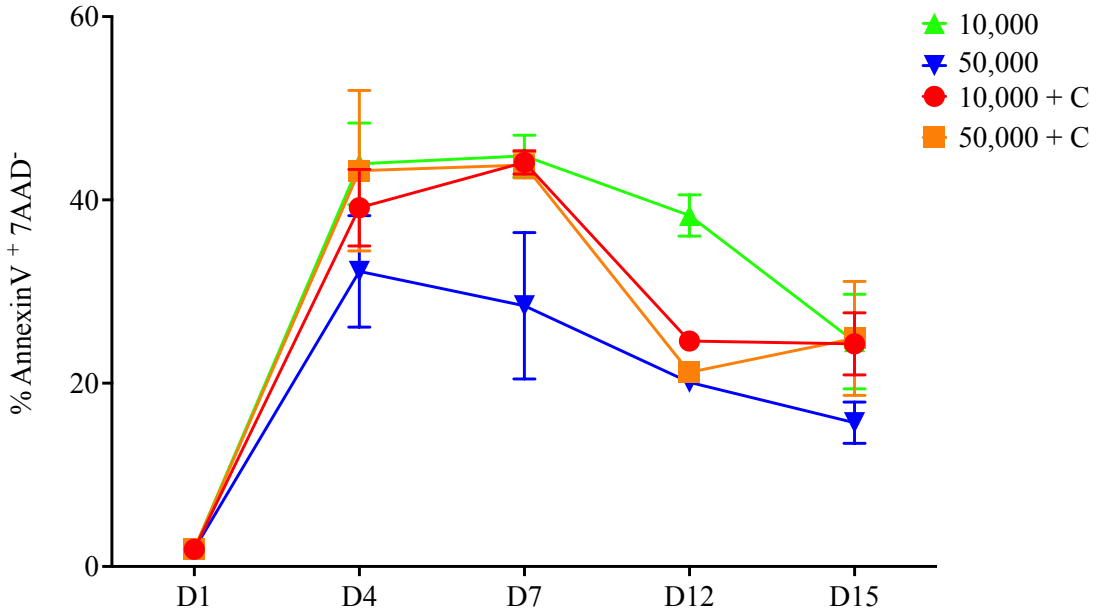
Figure 6.15: Optimising AML Culture: Effect of MSC on Viability. The effect of AML culture in the presence and absence of a human BM derived MSC stromal layer [84] on Annexin-V$^+$ is shown. PDX samples were cultured in the presence and absence of an MSC stromal layer for 15 days. Immunophenotypic profiles showing the expression of Annexin-V and 7AAD are shown from Day 4 of culture. Either 10,000 or 50,000 PDX cells were plated on Day 1 of culture, and representative plots from 10,000 cells are shown. All plots are gated on all events, excluding mouse CD45$^+$ and MSC-GFP$^+$ events. Frequencies indicate the frequency of the parent gate.



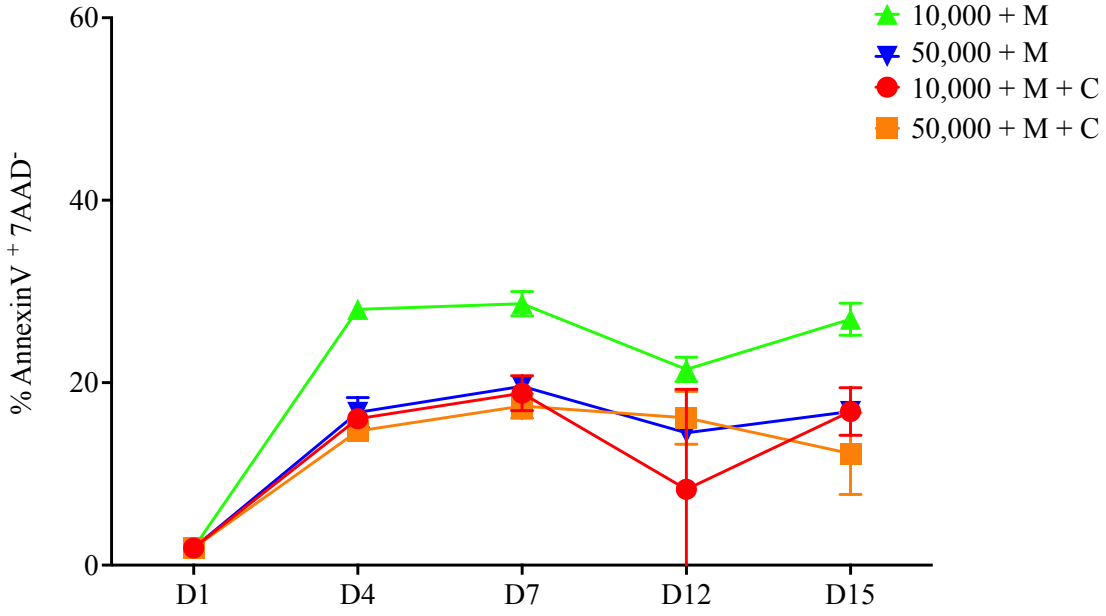
(a) BIR344: No MSC



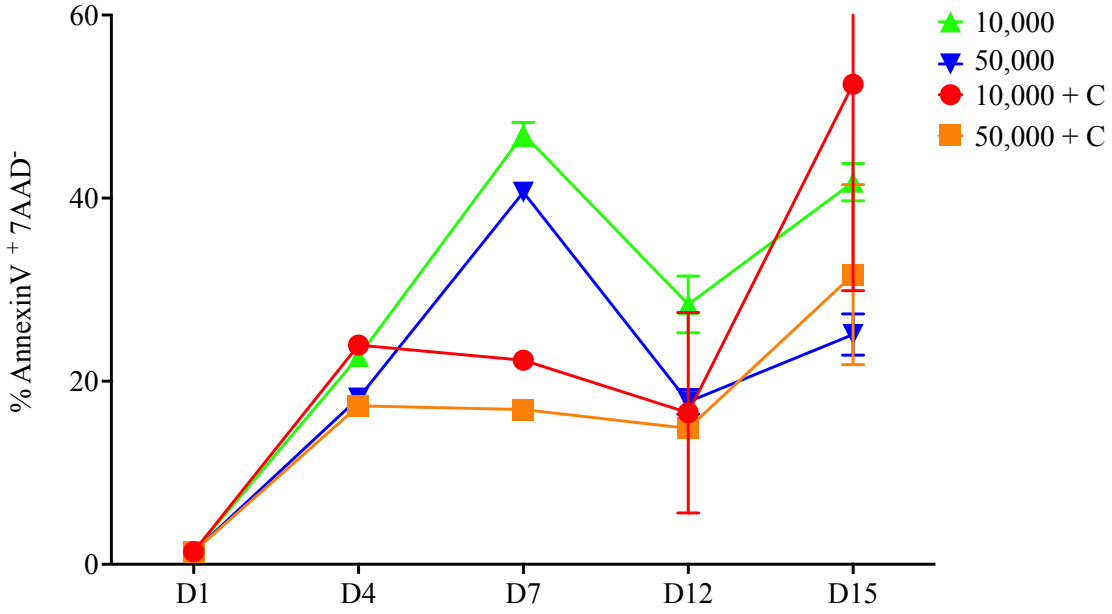
(b) BIR344: MSC



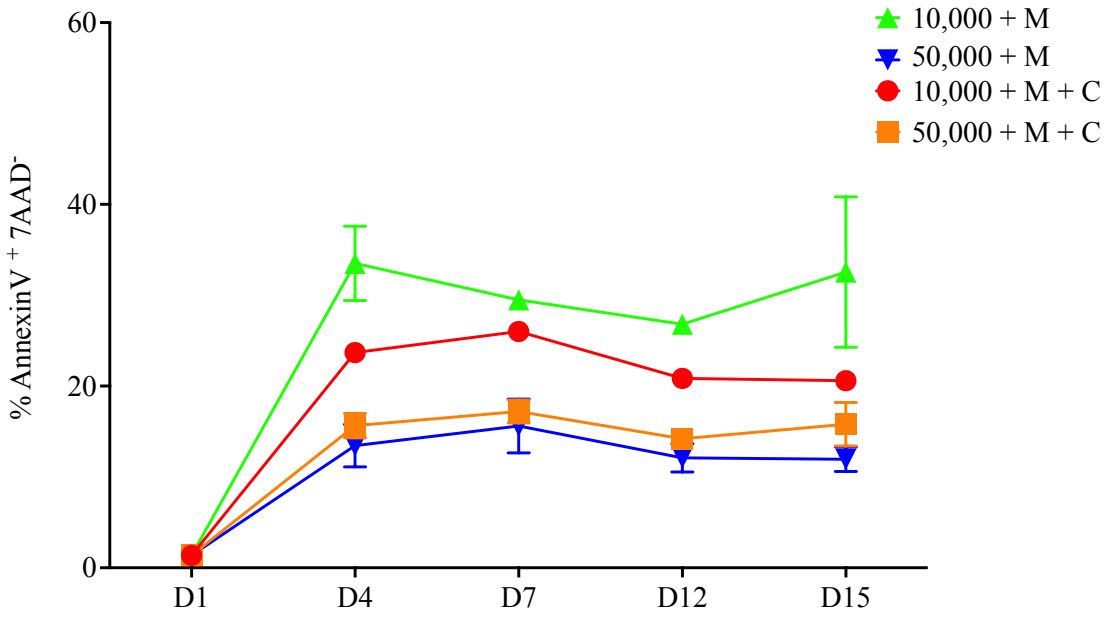
(c) BIR517: No MSC



(d) BIR517: MSC



(e) STB257: No MSC



(f) STB257: MSC

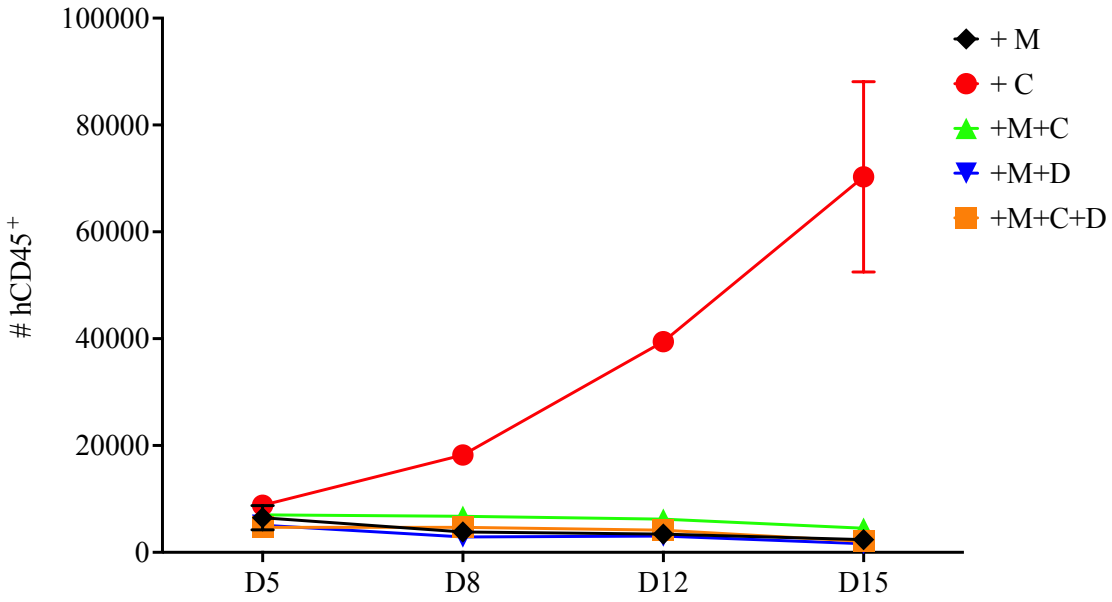
Figure 6.16: Optimising AML Culture: Summary of Effect of MSC on Viability. The effect of AML culture in the presence and absence of a human BM derived MSC stromal layer [84] on Annexin-V$^+$ is shown. PDX samples were cultured in the presence and absence of an MSC stromal layer for 15 days. Either 10,000 or 50,000 PDX cells were plated on D1 of culture (see key). Cultures were in the presence and absence of SR1 and UM171 (see Chapter 2.15 for full culture conditions). Error bars represent the standard deviation from two technical replicates. C: SR1 and UM171 compounds, M: MSC.

Following on from this, the effect of co-culture in the presence of MSCs on differentiation was assessed. Here an additional culture condition in the presence of Dimethyloxalylglycine (DMOG) was included. DMOG partially mimics hypoxic conditions by stabilising hypoxia-inducible factor (HIF), and it has been suggested it improves the culture of HSPCs [86].

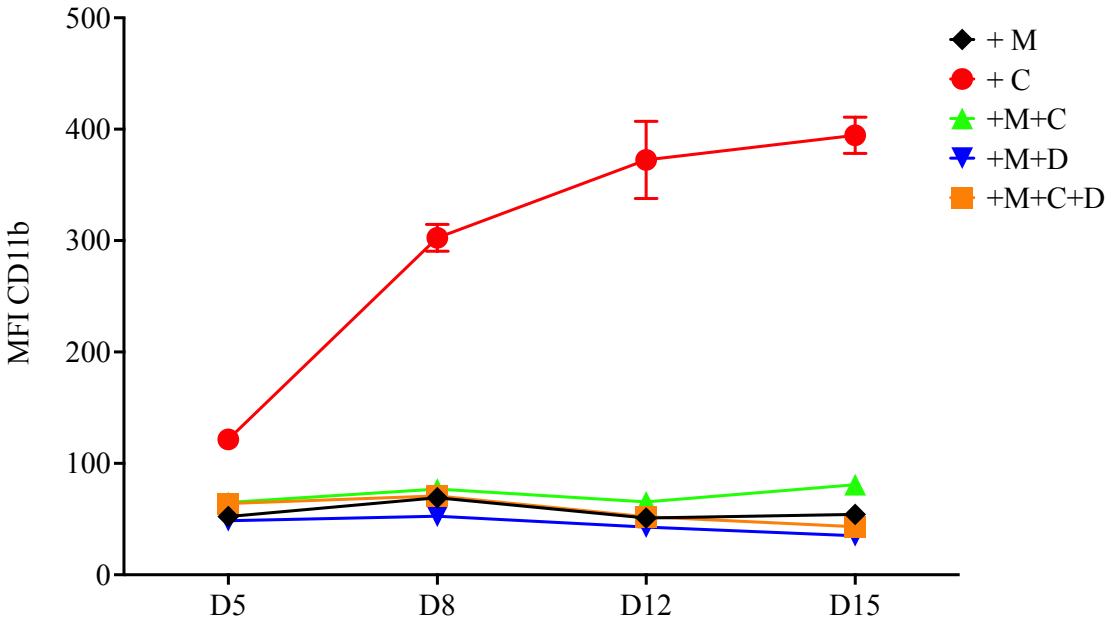
It was seen, for some samples, that the culture condition in the absence of MSCs was more permissive for the growth of differentiated cells (Figure 6.17). PDX sample STB257 was cultured in conditions in the presence/absence of MSCs, SR1 and UM171 and/or DMOG. CD11b was used as a marker of myeloid differentiation. Here an outgrowth of $CD11b^+$ cells was seen in the absence of MSCs (Figure 6.17). In the absence of MSCs, the number of cells (Figure 6.17a), the mean fluorescent intensity (MFI) of CD11b (Figure 6.17b) and the proportion of $CD34^-CD11b^+$ cells (Figure 6.17c) were observed to increase with time, at much higher levels when compared to all other culture conditions. Taken together this suggests cell expansion in this culture condition, but of more differentiated myeloid cells.

The differentiation of primary AML in *in vitro* culture conditions was further investigated. To characterise the differentiated cells the monocytic and granulocytic markers CD14 and CD15, respectively, were used. It was seen that the extent of differentiation was variable across AML samples, with the condition that minimised differentiation varying across samples. Both PDX and primary AML samples were cultured in conditions in the presence/absence of MSCs, SR1 and UM171 and/or DMOG.

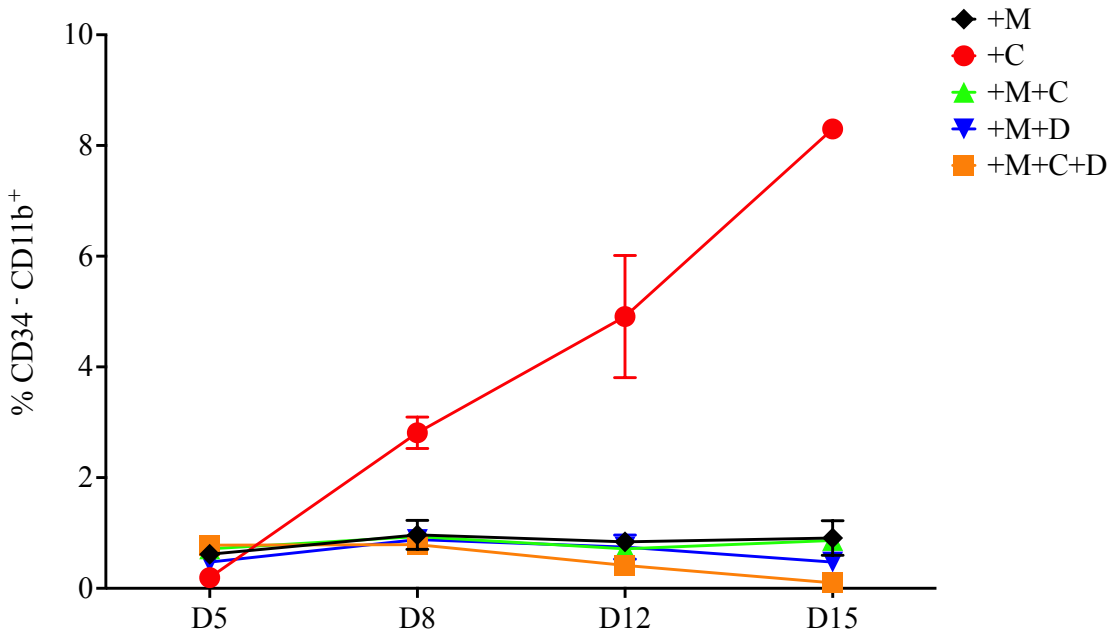
For PDX sample BIR344 and primary AML sample BIR473, it was seen that in any condition in the presence of compounds, differentiation was reduced (Figures 6.18a and 6.18d). For the primary AML sample OX1462 it was seen that in any condition in the presence of DMOG, differentiation was reduced (Figure 6.18f). For the PDX sample STB257 and the primary AML sample BRI2122 it was seen that in any condition in the presence of MSCs, differentiation was reduced (Figure 6.18c and 6.18e). For PDX sample BIR517 it appeared that any condition in the presence of MSCs has the converse effect of increasing differentiation (Figure 6.18b).



(a) Number of hCD45$^+$ Cells



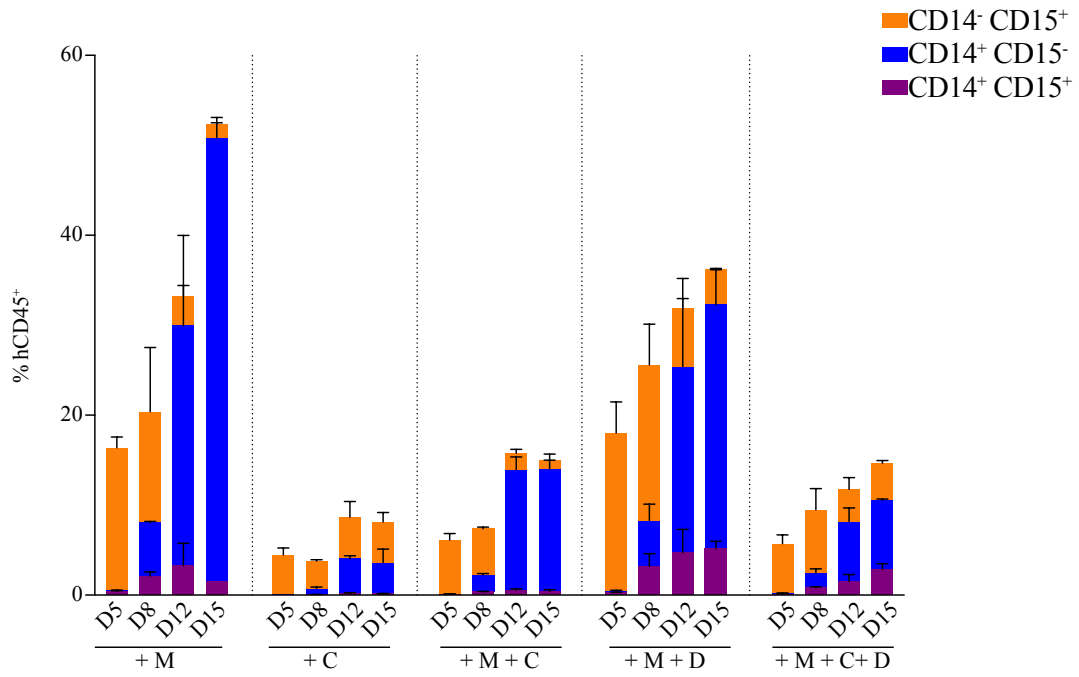
(b) MFI CD11b



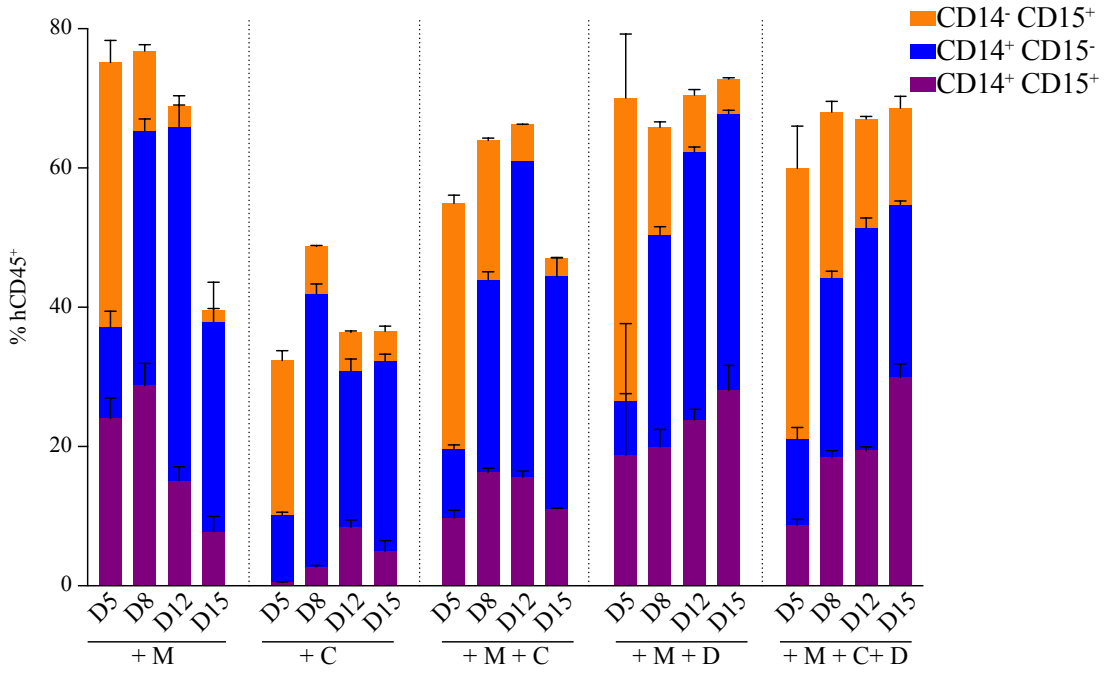
(c) CD34-CD11b+

Figure 6.17: Optimising AML Culture: Effect of SR1 & UM171 in the Absence of MSCs. PDX sample STB257 was cultured in various conditions (see Chapter 2.15 for full culture conditions). The effect of culturing in the absence of MSCs, but presence of SR1 and UM171 (+C) can be seen on (a) the number of human(h)CD45$^+$ cells, (b) the mean fluorescent intensity (MFI) of CD11b and (c) the proportion of hCD45$^+$ cells that are CD34$^-$CD11b$^+$. Error bars represent the standard deviation from two technical replicates. C: SR1 and UM171 compounds, M: MSC, D: DMOG.

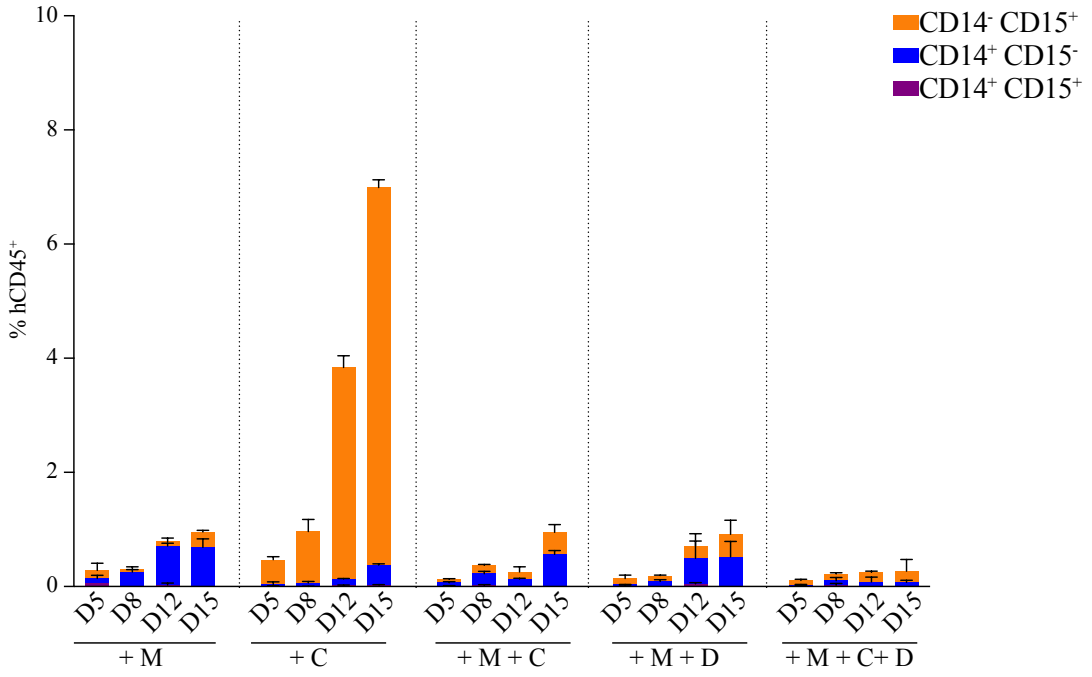
Therefore, the condition which minimised differentiation across the most samples tested was culture in the presence of MSCs, SR1 and UM171 and DMOG.



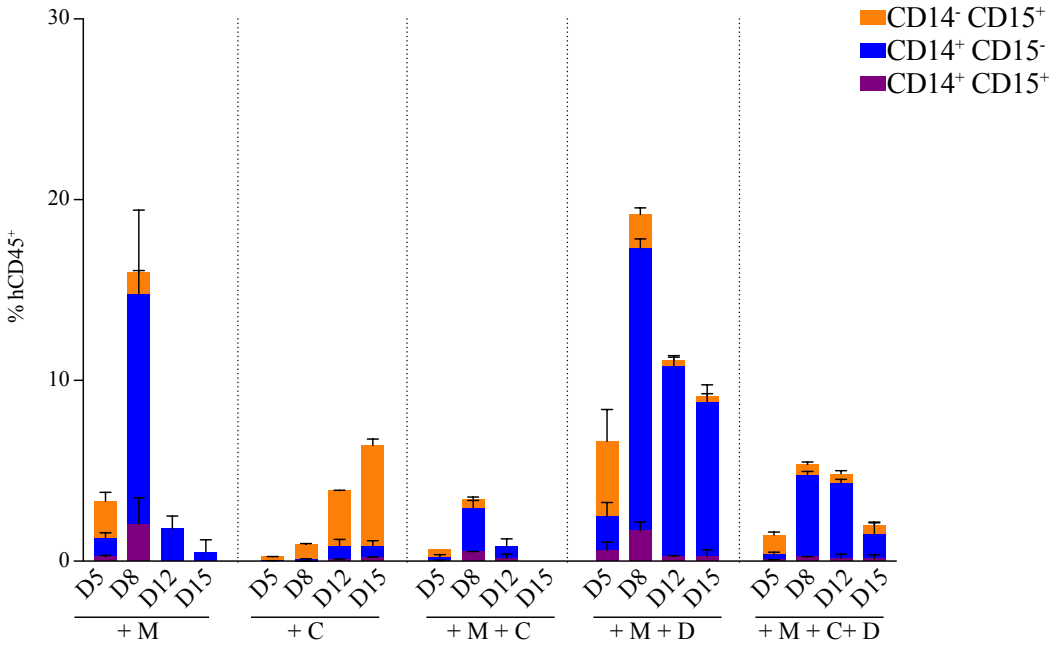
(a) BIR344



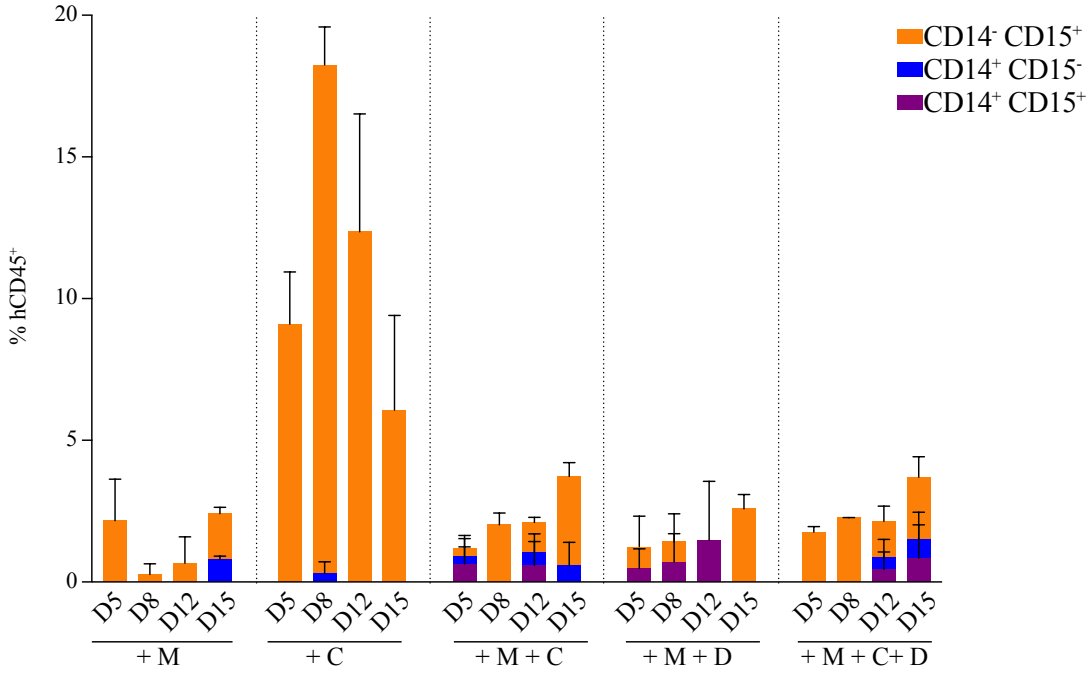
(b) BIR517



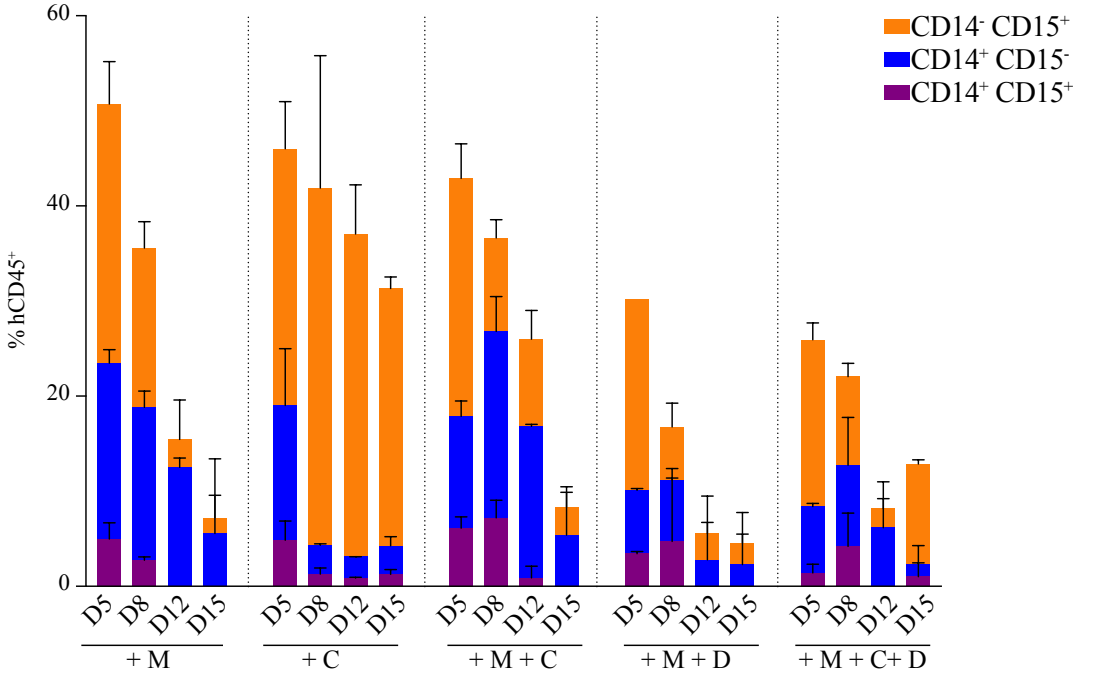
(c) STB257



(d) BIR473



(e) BRI2122



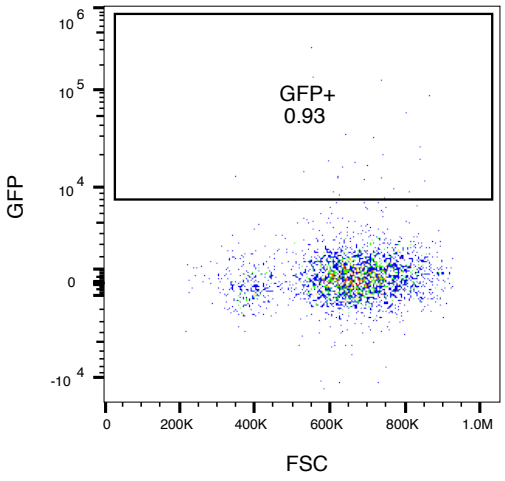
(f) OX1462

Figure 6.18: Optimising AML Culture: CD14 and CD15 Immunophenotype. PDX and primary AML samples were cultured in various conditions (see Chapter 2.15 for full culture conditions). Error bars represent the standard deviation from two technical replicates. CD14/CD15 immunophenotype shown as a percentage of hCD45$^+$ cells. C: SR1 and UM171 compounds, M: MSC, D: DMOG.

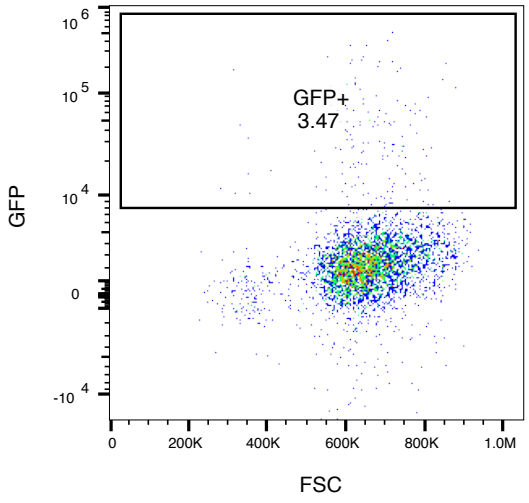
6.5 Lentiviral Transduction of Primary HSPCs

In order to introduce the CRISPR machinery into primary AML cells, the lentiviral transduction of primary HSPCs was tested. The pL-CRISPR.SFFV.GFP lentiviral plasmid was used to test the transduction efficiency of normal CB, PDX and primary AML cells. Cells were infected with lentivirus at a multiplicity of infection (MOI) ranging from 0 to 100 and analysed by flow cytometry 72 hours post-transduction. GFP positivity was used as a measure of transduction efficiency (see Chapter 2.18 for full transduction conditions).

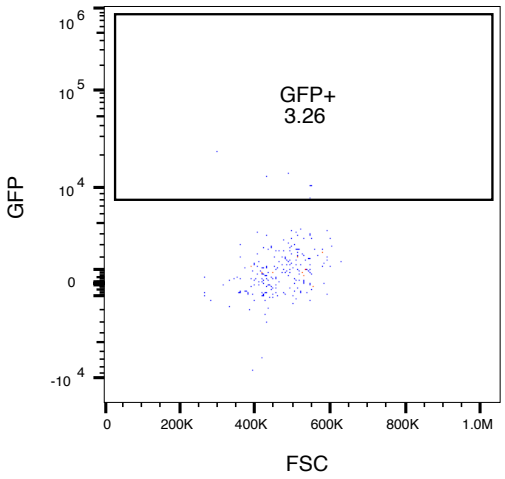
At a maximum MOI of 100, only minimal levels of transduction were seen. Normal $CD34^+$ CB cells showed the lowest transduction efficiency at ~1% (Figure 6.19a). Both the primary AML and PDX samples showed similar slightly higher levels of transduction of ~3%; however, the survival of the PDX sample in this experiment was poor (Figures 6.19b and 6.19c). The difference in surviving GFP^+ cells was seen more clearly when looking at the number of GFP^+ cells detected by flow cytometry (Figure 6.20a). There was a trend observed for both the number, and the proportion of GFP^+ cells, to increase with increasing MOI (Figure 6.20). It can be concluded from these results that, to achieve CRISPR editing of primary AML, transduction conditions require optimisation in order to maximise transduction efficiencies.



(a) Normal CB

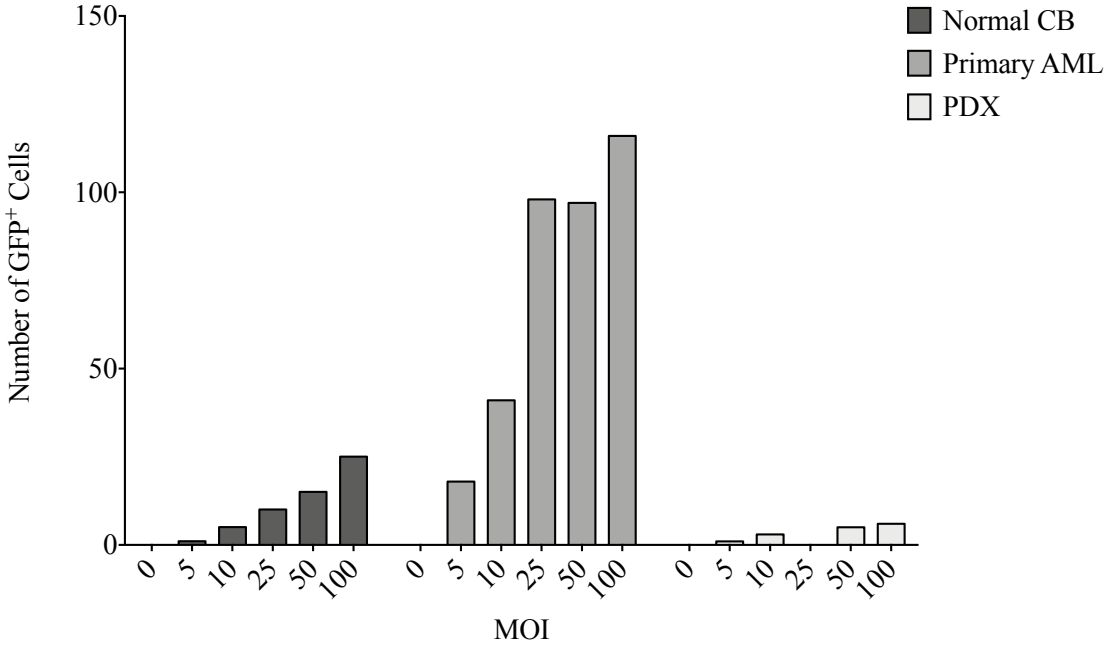


(b) Primary AML

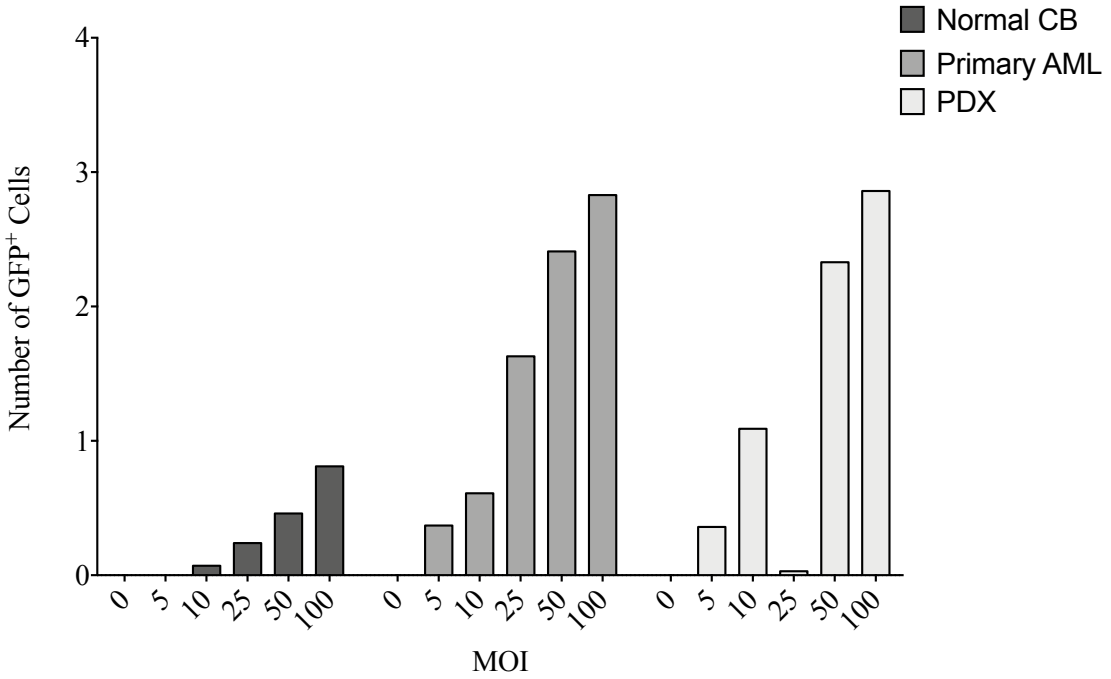


(c) PDX

Figure 6.19: Lentiviral Transduction of Primary HSPC. Primary HSPC were transduced by spinfection with the lentivirus pL-CRISPR.SFFV.GFP [91] backbone at various MOIs. Primary (a) normal CB, (b) AML and (c) PDX cells were used. Normal CB was cultured in conditions described in Chapter 2.18, and AML/PDX samples were cultured in conditions identified above (Chapter 6.4). 72 hours post-transduction cells were analysed by flow cytometry to assess GFP positivity indicating the extent of transduction. Shown is the GFP positivity at an MOI of 100. Frequencies indicate the frequency of the parent gate.



(a) Number of GFP$^+$ Cells



(b) Percentage of GFP$^+$ Cells

Figure 6.20: Summary of Lentiviral Transduction of Primary HSPC. Primary HSPC were transduced by spinfection with the lentivirus pL-CRISPR.SFFV.GFP [91] backbone at various MOIs. Primary normal CB, AML and a PDX cells were used. Normal CB was cultured in conditions described in Chapter 2.18, and AML/PDX samples were cultured in conditions identified above (Chapter 6.4). 72 hours post-transduction cells were analysed by flow cytometry to assess GFP positivity indicating the extent of transduction. Shown are (a) the number and (b) the percentage of GFP^+ cells.

6.6 Conclusion on the Functional Validation of Putative Leukaemia-Associated Targets

DEGs that were upregulated in AML were identified in Chapter 5.5.4. By integrating these genes with data from genome-wide CRISPR screens in AML cell lines, it was demonstrated that six of these genes may potentially be essential to the survival of leukaemic cells. It was shown that these genes could be targeted by CRISPR-Cas9 in the leukaemic K562 cell line. Genomic editing was achieved to varying extents across the six genes of interest and it was seen that, in most cases, editing resulted in the introduction of small indels. Preliminary work was carried out on the optimisation of the culture of primary human AML cells *in vitro*. Across the samples tested here, it was seen that the optimal condition for minimising myeloid differentiation was a co-culture in the presence of MSCs, with the small compounds SR1 and UM171 and the HIF-stabilising compound DMOG. Preliminary work was also done in an attempt to lentivirally transduce primary HSPCs. This technique requires optimisation in order to target primary HSPCs by CRISPR-Cas9. Taken together, this work presents a framework for targeting genes potentially involved in leukaemogenesis.

Chapter 7

Discussion

Contents

7.1 Normal Human LM Differentiation 267
7.1.1 Identifying Rare Multi-Lineage Progenitors 269
7.1.2 Studying Normal *In Vivo* Haematopoiesis 269
7.1.3 Eosinophil, Mast-Cell, Basophil and Dendritic Cell Lineages 270
7.1.4 Hierarchical Relationships Between LM Progenitors . . 271
7.1.5 Ontogeny of HSPC Samples and Enriching for LM Progenitors 272
7.1.6 Normal LM Progenitors and Leukaemic Transformation 272
7.2 Leukaemia-Propagating Potential of AML LSCs 273
7.2.1 Immunophenotype Screening of AML 273
7.2.2 Limitations of Xenotransplantation into NSG Mice and Alternative Mouse Models 273
7.2.3 Aberrant CD19 Expression in AML 275
7.2.4 The Effect of the Immunophenotypic LSC Population on Leukaemic Engraftment 276
7.3 Transcriptional Analysis of Normal and Leukaemic HSPC 277
7.3.1 The Normal Counterpart of the LSC 279
7.3.2 Identification of Upregulated Genes in the LSC 280
7.3.3 Loss of the Normal LMPP-GMP Transcriptional Changes 281
7.3.4 Epigenetic, Proteomic and Single-Cell Characterisation of the LSC 281
7.4 Functional Validation of Putative Leukaemic-Associated Targets 283
7.4.1 Optimising the Culture of Primary AML 283
7.4.2 Lentiviral Infection and Editing of Primary AML 285
7.5 Conclusion 287

7.1 Normal Human LM Differentiation

In Chapter 3, the normal human LM progenitors were studied, focussing on the LMPP, MLP and GMP. A full characterisation was carried out on freshly isolated CB LMPP, MLP and GMP cells both functionally and transcriptionally. At a

population level it was seen that these populations were distinct and separable, with clear immunophenotypic differences.

At a single-cell level, following on from an exhaustive analysis of 4,598 single LM progenitor cells, it was observed that these populations were highly heterogenous. The majority of single-cells were uni-lineage progenitors, that generated only one mature cell type. However, rarer bi-lineage and multi-lineage progenitors were detected, most frequently among the LMPP and GMP. The MLP represented the most functionally pure progenitor, being primarily a lymphoid-progenitor. Most rare was combined lymphoid and myeloid potential; however, this was detected, most commonly from the LMPP.

Single-cell transcriptional analysis showed that LMPPs, MLPs and GMPs formed a transcriptional continuum. MLPs and GMPs represented the two extremes of this spectrum, being the most lymphoid and myeloid progenitors respectively. Intersecting these two populations was the LMPP, which showed the most balanced profile.

By using flow cytometry index data to integrate the functional and transcriptional analyses, it was possible to devise new purification strategies for more functionally pure populations. The $CD38^{hi}$ GMPs were isolated, which had almost exclusively myeloid potential. Additionally, two sub-populations within the $CD10^{-}$ $CD45RA^{+}$ LMPP population were identified: the $CD10^{hi}$ $CD45RA^{hi}$ ($LMPP^{ly}$) cells and the $CD10^{lo}$ $CD45RA^{lo}$ ($LMPP^{mix}$) cells. The $LMPP^{ly}$ retained the majority of the lymphoid output from the traditional LMPP, and the $LMPP^{mix}$ retained the myeloid and LM potential.

Prior to this work, several issues had prevented clear understanding of previously identified LM progenitors. Firstly, HSPC populations have traditionally been isolated based on their immunophenotypic definitions. These are usually due to historical precedent, rather than using cell-surface markers that play a functional role. Functionally relevant cell-surface markers are utilised to isolate populations of mature immune cells, such as CD4 and CD8 T cells. To further improve the definitions of the HSPC populations, novel cell-surface markers should be identified. These should be functionally relevant, with the aim of isolating functionally pure

and uniform populations. To achieve this aim, further single-cell RNA-Seq and proteomic analysis of human HSPCs is needed.

Another limitation in previous work has been that, although several LM progenitors have been reported and studied [22, 24–27, 29], a side by side comparison of these populations had not been done. This had resulted in an inconsistency across the field regarding the definitions and study of the LM progenitors. By carrying out a characterisation of the three most widely reported LM progenitors, this study has clarified their immunophenotypic definitions, functional potential and transcriptional signatures.

7.1.1 Identifying Rare Multi-Lineage Progenitors

Traditionally haematopoietic progenitors downstream of the multi-potent HSC and MPP, such as the CLP, CMP and LMPP, were considered to be multi-lineage [13–16, 20, 22]. However following on from this, it was uncovered that these populations may be heterogenous, containing lineage biased cells [8, 195]. This lineage priming has been observed as early as the HSC, particularly involving megakaryocyte/platelet biased HSCs [9–11]. However more recently it has been proposed that the majority of these supposedly multi-lineage progenitors are a heterogenous collection of uni-lineage cells [12, 18, 28]. The data from Chapter 3 suggests that although the majority of single-progenitor cells are uni-lineage, rare multi-lineage progenitor cells do exist.

7.1.2 Studying Normal *In Vivo* Haematopoiesis

By studying the LM progenitors in a novel humanised ossicle mouse model, ~10–100-fold higher human engraftment was observed than has been seen previously [24, 28, 29]. Interestingly, in this assay the LMPP produced a higher myeloid than lymphoid output. This was in direct contrast to its *in vitro* activity, where it was a strongly lymphoid biased LM progenitor. This is even more striking when considering that traditional immunodeficient NSG models better support lymphopoiesis. It is possible that the ossicle assay may be particularly efficient at promoting myelopoiesis [57]. It is clear that the LMPP must retain a degree of

plasticity, allowing it to adapt its output to its conditions. It may be this plasticity that results in the transformation of the LMPP to the LSC in AML [22].

Another factor to be considered is the potential perturbation of haematopoiesis caused by a transplantation setting. It had originally been thought that a small number of multi-potent HSCs were responsible for driving the majority of haematopoiesis. However, through the use of steady-state tracking of haematopoiesis in mouse models, without irradiation or transplantation, it was seen that this was not the case [196, 197]. In reality it was observed that the majority of haematopoiesis was generated by a large number of long-lived progenitors. This resulted in a poly-clonal haematopoietic output and highlights the importance of studying progenitors.

Although the LM progenitors' *in vivo* potential was studied using the humanised ossicle mouse model, this approach is limited by both the transplantation setting and the limitations of not being in humans. To gain an insight into the unperturbed *in vivo* activity of the LM progenitors, primary human HSPCs and their mature progeny should be isolated and somatic mutations used as barcodes to reconstruct a phylogenetic tree of haematopoiesis (as proposed by Lee-Six et al., unpublished, [198]). This approach would also overcome the limitations of the *in vitro* read-out of lineage potential. Each culture condition is limited by being an artificial set of conditions, designed to push immature cells into the differentiation of particular lineages. This may force progenitors into producing an output they would not typically generate in unperturbed, steady-state *in vivo* conditions. Additionally, the absence of an output *in vitro*, as often seen by the MLP, may indicate the full potential of this population may not have been fully observed. Again, by studying these populations in steady-state *in vivo* conditions, a better indication of their full potential may be revealed.

7.1.3 Eosinophil, Mast-Cell, Basophil and Dendritic Cell Lineages

24 different lineage-associated potentials were detected from single-LM progenitor cells. However, in these experiments the eosinophil, mast-cell, basophil and DC lineages were not assessed. Therefore, it is possible there is an even higher level

of heterogeneity than suggested in Chapter 3. It has been demonstrated in mouse myelopoiesis that there is a segregation of the myeloid lineages, and it is likely this also occurs in humans [60]. This is supported by work done on the cell-surface marker CD133 in human haematopoiesis [199]. Here it was suggested that the neutrophil and macrophage output is generated by $CD133^{+}$ LMPP and GMP progenitors. On the other hand, eosinophil and basophil output was seen to arise from $CD133^{lo/-}$ $CD34^{+}$ erythro-myeloid progenitors (EMPs).

Regarding DC output, the progenitors to DCs have been described in humans [92, 200]. These included progenitors that produced: granulocytes, monocytes, and DCs (GMDP), monocytes and DCs (MDP) and only DCs (CDP). All three progenitors fell within the immunophenotypic GMP and were defined as follows: GMDP: $CD34^{+}$ $CD38^{+}$ $CD45RA^{+}$ $CD10^{-}$ $CD123^{int}$, MDP: $CD34^{+}$ $CD38^{+}$ $CD45RA^{+}$ $CD123^{int}$ $CD115^{+}$ and CDP: $CD34^{+}$ $CD38^{+}$ $CD45RA^{+}$ $CD123^{hi}$ $CD115^{-}$. Although when first described, it was proposed that the MLP was a DC progenitor, only monocyte-derived DCs were studied [24]. More recently the DC potential of the LM progenitors, has been more rigorously studied, and it was shown that the LMPP, MLP and GMP all have $CD141^{+}$ conventional DC, $CD1c^{+}$ conventional DC and plasmacytoid DC potential [200]. When considering the combined LM and DC output at the single-cell level, the progenitors were determined to be highly heterogenous and mostly oligo- and uni-lineage progenitors [200].

7.1.4 Hierarchical Relationships Between LM Progenitors

Determining the hierarchical relationships between the LM progenitors is a complex task that remains to be completed. Preliminary work has been done in the context of the DC progenitors [200]. Using DC culture conditions and xenotransplantation into NSG mice, the hierarchical relationships between the HSPC and DC progenitors were investigated. However, this was based only on immunophenotype, and the MLP in these conditions was defined by CD7 expression rather than CD10. It is challenging to solely use immunophenotype to define cell type in these assays. This is due to CD38 being the only cell-surface marker that differentiates the LMPP

and GMP. It has been shown that upon culture and cell cycle activation, CD38 is upregulated in $CD34^+$ cells [201]. To fully study hierarchical relationships between the LM progenitors and other HSPC populations, a combination of live-cell tracking experiments [202, 203] and sequential transcriptional profiling will be needed.

7.1.5 Ontogeny of HSPC Samples and Enriching for LM Progenitors

All experiments in Chapter 3 were done using fresh CB samples. It has been suggested by Notta et al. that the potential multi-potency of HSPCs is not constant across development [12]. When comparing HSPC from fetal liver, CB and BM, it was seen that as development progressed, the proportion of multi-potent cells decreased and the proportion of uni-potent cells increased. To understand the multi-lineage potential of the LM progenitors in adult haematopoiesis, a comparison of the data from CB should be done with BM. It is possible that the CB data underestimates the extent to which the LM progenitors are composed of uni-lineage single-cells.

Furthermore, both the LMPP and MLP were seen to be extremely rare in CB and even more so in adult BM. Therefore, to achieve robust data on these progenitors, fresh CB and flow sorting to enrich for these populations was used. This enabled the detection of the rare multi-lineage and LM single-cell progenitors. An unbiased study of the $CD34^+$ compartment or of the $CD34^+CD38^{+/-}$ compartments is unlikely to provide adequate representation of these rare populations [18, 28]. For example, a single-cell RNA-Seq studied carried out by Velten et al. studied 467 single $CD34^+CD38^-$ cells from one individual and 261 from a second individual [28]. Only 10 and 11 single MLPs were studied from the first and second individuals respectively, and in this study the MLP included both the $CD10^+$ MLP and $CD10^-$ LMPP. Therefore, the rare multi-lineage LMPP cells were highly unlikely to have been detected using the unbiased approach of this study.

7.1.6 Normal LM Progenitors and Leukaemic Transformation

Finally, it has been demonstrated that in AML there is often a differentiation arrest at the LMPP and GMP stage. In the majority of cases of AML, typically more than 80% of cells express CD34 [22]. Within these $CD34^+$ cases of AML, ~85% of patients

will have an arrest and expansion at the LMPP and GMP stage, and it has been shown that these LMPP- and GMP-like populations contained LSC activity [22]. This is particularly remarkable when considering how rare the LMPP is in adult BM. It is possible that this small healthy pool of LMPPs may be tightly controlled to minimise oncogenic transformation. It is possible that many cells pass through an LMPP stage, but it may be a very transitory population, with cells not remaining as an LMPP for extended periods of time. It is also intriguing that the LMPP-like LSC is found so commonly in myeloid leukaemia when considering its highly lymphoid nature that was observed. However, it was clear that the LMPP retains the ability to generate a substantial myeloid output, as seen in the ossicle assay. It may be this myeloid programme that is hijacked in AML. Further study of how the healthy LMPP differentiates into normal lymphoid and myeloid cells may provide insight into potential differentiating therapies for LMPP-arrested leukaemias.

7.2 Leukaemia-Propagating Potential of AML LSCs

7.2.1 Immunophenotype Screening of AML

The immunophenotypic profile of ~50 $CD34^+$ AML cases was studied. Prior to this, the MLP had not been studied in the context of AML. However, this screen demonstrated that the MLP is rarely expanded in $CD34^+$ AML. Taking into account the highly lymphoid nature of the MLP, in particular towards B cells, this is unsurprising. On the other hand, the LMPP, which has been shown to retain considerable myeloid potential, was often expanded in AML, often times alongside the GMP. The GMP was most commonly expanded in AML, as to be expected by its being predominantly a myeloid progenitor. As previously described, a small proportion of cases showed expansion of the $CD45RA^-$ compartments [22].

7.2.2 Limitations of Xenotransplantation into NSG Mice and Alternative Mouse Models

A major limitation with the xenotransplantation of AML into mouse models is the large proportion of samples that do not engraft mice, with only two thirds of samples engrafting in optimised assays [102–104]. As suggested by the transcriptional

profiles in Chapter 5, this does not appear to be due to global gene expression differences, implying the human AML cells may not have been able to adapt to the mouse microenvironment.

It has been suggested by Paczulla et al. that many AML samples have a long latency period, and engraftment is only detectable at time-points beyond the normal detection period [104]. It was suggested that favourable-risk samples took an average of ~28 weeks, intermediate-risk ~22 weeks and adverse-risk ~18 weeks for leukaemia to be detectable. In Chapter 4, all xenotransplantation assays were terminated at ~24 weeks. It is possible that the AML samples that were not seen to engraft may have engrafted if analysed at time points beyond this. Paczulla et al. monitored mice for up to one year for the development of leukaemia. However, an assay with an extensive readout period of one year may not be suitable for a routine assay for LSC function.

One method in which to improve the engraftment of human AML in mouse models is to use the ossicle assay. Not only has this humanised microenvironment been shown to improve the engraftment of normal HSPCs, as shown in Chapter 3, it has also been shown to improve the engraftment of human AML [57]. It was seen that 13 out of 15 AML samples tested were able to engraft mice with human leukaemia, and in 12 of these samples, the human leukaemic chimerism was seen to be over 95%. Furthermore, it was observed that the leukaemia initiating cell (LIC) frequency was at higher levels when evaluated using the ossicle assays [57]. To better assess the leukaemia-propagating potential of the AML samples studied, xenotransplantations into the ossicle assay should be done. There is preliminary data that some of the non-engrafting AML samples from Chapter 4 may engraft in the ossicle assay (data not shown).

An alternative to the ossicle assay is the use of engineered bioscaffolds designed to resemble human BM niches [204–207]. It has been proposed that there are several advantages of using bioscaffolds over traditional xenotransplantation into immunocompromised mice. A high proportion (~75%) of samples (including favourable-risk samples) were seen to engraft mice with leukaemia. Self-renewal

was seen to be better preserved when using bioscaffolds, as assessed by serial transplantation and stem-cell transcriptional signatures. Finally, bioscaffolds were suggested to better preserve the clonal heterogeneity as seen in original patient samples [207].

There are also some suggested advantages of using bioscaffolds over ossicles. Firstly, since the ossicle assay requires pre-injection of MSCs and formation of the ossicle bone structure, the ossicle experiment may be lengthier than the bioscaffold experiment. Additionally, it has been proposed that the formation of a bone structure may not be necessary for engraftment [206]. Furthermore, the ossicle assay is traditionally done using human BM derived MSCs, whereas it has been suggested that bioscaffolds may be more flexible and could utilise a range of stromal cell types, including malignant stromal cells [206].

Many studies have suggested that the limited engraftment of human samples in mouse models is due to a lack of human cytokines necessary for HSPC proliferation, survival and differentiation [208–210]. Therefore, several efforts have been made to generate mouse models that express human cytokines. These include NSGS mice that have transgenic expression of hSCF, hGM-CSF and IL-3 [208]. Additionally, mice have been generated with hM-CSF, hIL-3, hGM-CSF and hTPO knocked into their respective mouse loci [209, 210] (termed MITRG mice). This strain was further developed to also contain a bacterial artificial chromosome (BAC) transgene encoding human SIRPα, minimising the phagocytosis of CD47-expressing human HSPCs (termed MISTRG mice) [210]. It is possible that the engraftment of the AML samples studied may have been improved in these models, potentially with different kinetics of engraftment.

7.2.3 Aberrant CD19 Expression in AML

In half of the engrafting AML samples in Chapter 4, some of the human engraftment was seen to express the lymphoid B cell marker CD19. It was demonstrated that this was unlikely to be normal B cell engraftment generated by residual normal HSPC cells, due to the presence of leukaemic mutations in the CD19$^+$ cells and

abnormal co-expression of CD34. It should be noted that $CD19^+$ cells would not have been injected in xenotransplantations, because these would have been excluded by the flow sorting strategy (since CD19 was included in the lineage cocktail).

The aberrant expression of CD19 by myeloid blasts in AML has previously been described [105–110]. This is most typically associated with t(8;21) AML [105, 106, 108]; however, it has been suggested that CD19 may also be expressed by AML lacking t(8;21) [110]. It has been proposed that the aberrant expression of CD19 may result from the expression of the TF PAX5, a known regulator of CD19 [108]. However, the mechanism by which PAX5 is upregulated, potentially involving RUNX1/ RUNX1T1 is not known. It is agreed that this CD19 expression is not usually associated with other features of lymphocytic leukaemia or with engagement of a B cell differentiation programme [105, 108]. It is possible that the aberrant CD19 expression may be related to the LM nature of the transformed LMPP- and GMP-like LSC populations, and the abnormal arrest in both lymphoid and myeloid differentiation.

7.2.4 The Effect of the Immunophenotypic LSC Population on Leukaemic Engraftment

When leukaemic engraftment was observed in Chapter 4, broadly similar engraftment was seen across all transplantations from the same AML sample. Firstly, for samples where some, but not all, injected populations gave rise to human engraftment, this appeared to be a stochastic event, most likely due to the limitations of xenotransplantation discussed above. Secondly, when leukaemic engraftment was seen (above the 0.1% threshold for positive human engraftment), the level of human chimerism was often similar across injections for a sample, regardless of the injected population. Similarly, the lymphoid/myeloid nature and kinetics of engraftment was often similar across all injections of a sample, regardless of the injected population. This observation was particularly interesting when seen from sample STB257, where all three of the LM progenitors were transplanted, and all three gave rise to similar mixed LM engraftment. Finally, the genotype of the engrafted cells was seen to be consistent across all injections of a sample, regardless of the injected population.

The HSPC immunophenotype of the engrafted leukaemia was often seen to be consistent with that of the original patient sample. Again, the HSPC immunophenotype appeared to remain similar across all injections from a sample, regardless of the injected population. Interestingly, in sample BIR344, minor immunophenotypic HSC and MPP populations were seen in the original patient sample, yet lost post-engraftment. This HSC/MPP loss was even observed when the HSC- and MPP-like populations were injected, with the resulting engraftment showing an LMPP and GMP dominant immunophenotype. Also of note, following xenotransplantation, sample STB257 displayed a predominantly BNK progenitor like immunophenotype. This $CD10^+$ $CD38^+$ engraftment was generated by the LMPP, MLP and GMP leukaemic populations. The upregulation of CD38 may be related to the association of CD38 with the cell cycle [201].

Taken together, these findings suggest that the immunophenotype of the LSC population may not play a fundamental role in determining the nature of leukaemic engraftment. It is implied that, for each leukaemic sample, only one genetic clone engrafted mice, regardless of which immunophenotypic population was injected. Furthermore, the consistent LM balance of the engraftment across all injected populations for a sample suggests this is irrespective of the original lineage-bias of the injected population. Finally, the consistent HSPC immunophenotype observed post-engraftment suggests that normal immunophenotypic population relationships may be dysregulated. However, it should be noted that only primary transplantations have been studied here, and functional differences between LSC populations may become apparent in serial transplantations.

7.3 Transcriptional Analysis of Normal and Leukaemic HSPC

In Chapters 3 and 5, the transcriptional profiles of normal CB and BM HSPC, respectively, were studied. It was seen that these populations were distinct and separable, even when unbiased analyses were carried out that did not pre-assume population identity. Gene expression programmes for the major lineages were

identified and an immature, stem-cell signature was generated. The observed separation of the HSPC populations is in general agreement with previous work on these (and similar) populations [25, 37, 94]. However, it is now clear that within these populations there are high levels of transcriptional heterogeneity. Single-cell work has started to uncover the continuum of haematopoietic differentiation and sub-populations primed towards mature lineages (see Chapter 3 and [28]).

When looking at the transcriptional profiles of leukaemic samples, a complex variation across samples was seen. Firstly, it was clear that there was a strong patient effect, with HSPC populations from the same AML patient clustering together across all types of analysis. Secondly, when unbiased approaches in analysis were used, clustering of the LMPP-, GMP- and MLP-like leukaemic populations was not seen. When the sources of variation across the leukaemic samples were investigated, it was seen that these were complex. This variation potentially involved the stage at which LM differentiation arrest had occurred, with lymphoid/myeloid-associated genes seen to be highly variable. Furthermore, genes previously associated with both solid and haematological cancers were identified to be highly variable, implying possible variation in the mechanisms of cancer proliferation and survival.

When comparing the transcriptional profiles of AML samples with normal HSPCs, a clear separation between leukaemic and non-leukaemic HSPCs was seen. This is not surprising, given that the cancerous state of AML is likely to be associated with major dysregulation of normal gene expression. Whereas normal HSPC showed distinct, lineage-associated patterns of gene expression, these genes in leukaemia showed a dysregulated pattern. As suggested previously, the leukaemic transcriptional profile was more progenitor-like, but with upregulation of stem-associated genes [22]. It could be proposed that differentiation arrest occurs at a progenitor stage, and leukaemic transformation results in re-activation of abnormal self-renewal, resulting in the LMPP- and GMP-like LSC populations. The clinical relevance of the stemness of the LSC has been demonstrated by Eppert et al. [47], where it was shown that enrichment for an HSC or LSC gene signature correlated with a poor prognosis [47].

7.3.1 The Normal Counterpart of the LSC

Identifying the normal counterpart of the LSC populations in AML was seen to be a complex question. It is clear that leukaemic transformation causes major dysregulation of the leukaemic transcriptional profile; therefore, the LSC's resemblance to normal HSPC populations is minimal. It can be seen that the LSC is an AML-specific population that sits in a novel position in the transcriptional landscape of normal haematopoiesis. It was observed that the LSC populations had progenitor-like profiles that most closely resembled immature myeloid progenitors. However, on a myeloid axis of differentiation it was seen that the LSCs do not fully resemble normal GMPs, due to an arrest in myeloid differentiation. There was a trend for LMPP-like LSCs to lie closer to normal LMPPs, and for GMP-like LSCs to lie closer to GMPs (as suggested previously [22]). However, this was a mild separation and it is clear that the LSCs represent highly dysregulated LM progenitors.

This is in agreement with work done by Rapin et al. that describes the complexity of identifying the normal counterpart of AML cells [132]. Here not only the HSPC populations were considered, but mature myeloid cells and their precursors were also included. However, the GMP was the only LM progenitor included in this analysis, and the more lymphoid-biased LM progenitors were not included. It was seen that the normal counterpart of the AML cells varied with patient and was somewhat dependent on cytogenetic sub-type. However even within a cytogenetic group, several cell types were often identified as the normal counterpart, ranging from the HSC to mature monocytes. As with the analysis in Chapter 5, it could be observed that the AML samples had dysregulated transcriptional profiles and occupied a novel position in the landscape of normal haematopoiesis. It should be noted that this work was done using previously published data on bulk AML cells, where the majority of cells would represent leukaemic blasts. Although informative, this is limited in shedding light on the rarer LSC population.

7.3.2 Identification of Upregulated Genes in the LSC

In order to elucidate the transcriptional changes that occur during leukaemic transformation, DEGs were identified in the LSC populations. It had been observed that across normal HSPCs there was a wide range of transcriptional diversity associated with the various mature lineages. Therefore, a naive comparison of the LSC populations with all HSPC populations is likely to struggle to identify meaningful genes associated with leukaemogenesis. This is due to confounding factors such as the lack of erythrocyte and megakaryocytic potential from the LSCs. Consequently, two more focussed approaches were taken: a comparison of the normal LMPP and GMP with the leukaemic LMPP and GMP, and a comparison of leukaemic HSPC with the normal HSC. A large overlap in the genes identified by the two approaches was observed. Due to the LSCs' displaying a more progenitor-like transcriptional profile, the genes identified from the former comparison were taken forward for downstream work discussed in Chapter 6.

Since the majority of published work comparing AML with normal HSPCs has used bulk AML cells, it is difficult to compare this with the DEGs identified in Chapter 5. However, Majeti et al. have taken a more refined approach to comparing the LSC population with normal HSCs [128]. Although only the $CD38^-$ LSC was considered here, similarities can be seen between this dataset and that shown in Chapter 5. These similarities are seen even though Majeti et al. have compared the LSC with the HSC, and work in Chapter 5 focussed on a comparison with LM progenitors.

For example, enrichment of apoptosis related pathways was identified by Majeti et al., and this could be related to the role of PKR in an apoptosis pathway identified here. Regulation of actin cytoskeleton was another pathway enriched in the Majeti et al. data, and this could be related to the Substance P mediated cytoskeleton remodelling pathway observed here. Interleukin receptor activity was an enriched functional group in the Majeti et al. data, and CNTF and LIF (both members of the interleukin 6 family) were identified here in a pathway enriched in AML. The enriched functional group of protein kinase activity in the Majeti et al. data

could encompass several of the enriched pathways seen here. This includes the PKR activity identified and any of the signalling pathways identified that involve protein kinases. Lastly, the Jak-Stat pathway was identified by Majeti et al. as being enriched, and it is known that LIF and CNTF (identified here) are able to activate the Jak-Stat pathway. The similarities between these two independent studies imply that a common leukaemogenic phenotype may have been identified.

7.3.3 Loss of the Normal LMPP-GMP Transcriptional Changes

It is clear in normal haematopoiesis that the LMPP and GMP represent two functionally and transcriptionally distinct and separate populations. However, when looking at the LMPP- and GMP-like populations in AML, this transcriptional distinction was mostly lost. Not only was there very little transcriptional difference between the two populations, the transcriptional changes seen between the normal LMPP and GMP were not retained. This implies that the normal functionality of the LMPP and GMP populations had been lost upon leukaemic transformation, and there is now minimal difference between the two LSC populations. When looking into the small transcriptional differences between the leukaemic LMPP and GMP, it was seen that these genes were enriched for cell cycle associated pathways. This implied that the GMP-like LSC may represent a more actively cycling LSC population than the potentially more quiescent LMPP-like LSC. It could follow that the GMP-like LSC is potentially downstream of the LMPP-like LSC in the leukaemic hierarchy (as suggested previously [22]). This cell cycle phenotype remains to be functionally validated, however is in agreement with previous work comparing the $CD34^{+}CD38^{+}$ and $CD34^{+}CD38^{-}$ AML populations and the known upregulation of CD38 with cell cycle [120, 201].

7.3.4 Epigenetic, Proteomic and Single-Cell Characterisation of the LSC

It should be noted that transcriptional analysis is able to explain only one aspect of the leukaemic phenotype. For a more complete understanding of leukaemic transformation and survival, analysis both at the epigenetic and protein levels

should be carried out. Epigenetics is known to play a major role in AML, with mutations in epigenetic regulators commonly occurring in AML and dysregulated epigenetic profiles having been observed (reviewed here [168]).

Through the development of mass cytometry, some groups have begun to interrogate the heterogeneity of AML at the protein level [211, 212]. This single-cell analysis has allowed intra-tumour heterogeneity to be uncovered, revealing that cell surface markers were poor indicators of the signalling state of AML cells. When considering the signalling state of cells (based on intracellular signalling molecules) Inferred Functionally Primitive Cells (IFPCs) could be identified that correlated with poor survival [212]. This is in accordance with the work discussed in Chapter 4, where it was suggested that there is a decoupling between cell surface marker expression and leukaemic function.

It has been shown that within normal HSPC populations there is a great deal of heterogeneity at the single-cell level (see Chapter 3 and [28]). Therefore, to better understand the leukaemic transcriptional profile, single-cell transcriptional analysis should be carried out on primary human AML. This should be done not only on leukaemic blasts, but additionally on enriched LSC populations that may not be captured by a completely unbiased analysis. This single-cell data should then be compared with normal single-cell HSPC data in order to identify, for example, where in the continuum of normal differentiation single leukaemic cells are arrested.

Although the single-cell RNA-Seq of normal human HSPC has been widely studied, the analysis of primary human AML by single-cell RNA-Seq is more limited. A small, proof of concept study has carried out single-cell RNA-Seq on 20 single-AML cells [213]. In a slightly larger study, single-cell RNA-Seq was carried out on two AML patients both pre- and post-allogeneic hematopoietic stem cell transplant [214]. Here sub-populations within AML could be identified, with the kinetics of these populations analysed pre- and post-transplantation. However, the implications of these findings are limited, due to only preliminary t-SNE (t-Distributed Stochastic Neighbor Embedding) and clustering analysis being carried out, and due to one of the patient's being diagnosed with erythroleukaemia [214].

7.4 Functional Validation of Putative Leukaemic-Associated Targets

By integrating the DEGs identified in Chapter 5 with genome-wide CRISPR screens in AML cell lines, six potential targets of interest were identified. These six genes were seen to have the most severe drop-out across the majority of the AML cell lines tested, indicating a potential essentiality of these genes. To confirm that these putative targets were not essential in all cells, two non-AML CRISPR screens were also interrogated [80, 81]. Although there was some evidence that *CLP1* may have been essential across all cells, when K562 cells were edited *in vitro*, this was not seen.

It was seen that the electroporation of the plasmid containing the CRISPR machinery or the control pMax plasmid into K562 cells caused a transient decrease in viability. It is well known that transfection of plasmids by electroporation into haematopoietic cells causes a decrease in viability [182, 215]. The decrease in viability following K562 transfection was mostly seen to recover at later time points. Editing of K562 cells was observed at varying levels, with ICE Scores ranging from 5 to 77. When looking into the nature of these edits, they were mostly seen to be smaller indels, but some larger editing was observed.

7.4.1 Optimising the Culture of Primary AML

The *in vitro* culture of primary AML remains a challenge [85, 86, 216–222]. Although some improvements were made regarding the culture of these cells, as discussed in Chapter 6, further optimisation is required. Several issues need to be overcome before extensive *in vitro* work can be done on primary AML cells. Firstly, cell death remains an issue with many cells dying upon culture. Secondly, when cells can be maintained in culture, immature LSC populations are rarely expanded. This is in contrast with the highly proliferative ability of AML cells *in vivo*, although it could be argued that the LSC may exist in a quiescent state, with the leukaemic blasts being more proliferative. Finally, an excess of differentiation is often observed when culturing AML *in vitro*, with the end result of differentiation being cell death.

This, again, is in contrast to the severe differentiation block observed *in vivo*. This implies that the maintenance of AML is highly niche dependent, since once taken out of its microenvironment, it is unable to maintain its leukaemic properties.

One of the first factors to be considered when culturing AML cells is whether a stromal layer of cells should be used. Given the potential niche-dependent nature of AML, it follows that the presence of a feeder cell layer may prove beneficial. There have been several attempts to culture AML with a stromal cell layer [217–220]. A range of cell types have been used to support leukaemic growth, including the mouse MS5 cell line [218, 219], the human HS27 fibroblast cell line [220] and primary human MSCs [217]. Different AML starting populations have been used in conjunction with a stromal layer, including bulk AML cells [218, 220] and $CD34^+$ cells [217, 219]. Various cytokine cocktails and serums were used in these culture conditions, with one of the cultures being carried out in hypoxic conditions [218].

Several groups have proposed culture systems without a stromal layer [85, 86, 221]. Most notably, Pabst et al. have carried out a high-throughput screen and identified small compounds which have been proposed to maintain LSC activity *in vitro* [85]. One of these compounds was SR1, which is known to act through the aryl-hydrocarbon receptor (AhR) pathway. The second of these was UM729, which collaborated with AhR suppressors, but was proposed to be independent of this pathway [85]. Following on from this, UM171, an analogue of UM729, was developed and described to be 10–20 times more potent [223]. UM171 was also shown to maintain the self-renewal of normal HSCs.

Finally, it has been proposed that using small molecules to mimic hypoxia can be beneficial to maintaining HSCs *in vitro* [86]. The compound DMOG partially mimics hypoxic conditions by stabilising HIF and was used in conjunction with SR1 by Jackson et al. to maintain HSCs *in vitro* [86].

In Chapter 6 several of the proposed culture conditions were tested to gain insight into an optimal condition for the culture of LSCs. Firstly, it was seen that, for some AML samples, the presence of an MSC stromal layer improved initial viability. It was observed for these samples that, especially at earlier time points, the

proportion of cells entering apoptosis was lower. This could imply that the presence of an MSC stromal layer forms a more supportive environment, that minimises the initial shock experienced by cells upon cell culture.

Following on from this, it was observed that in culture conditions that lacked an MSC stromal layer, an outgrowth of more differentiated cells occurred. This suggests that conditions lacking MSCs did not sufficiently support the immature leukaemic populations. In contrast, they were more permissive for the proliferation of cells entering differentiation. If expansion is the only goal of culture, then this may be suitable. However, if the aim is to maintain the LSC population, then this is of limited suitability.

Finally, when looking across several AML samples, either PDX or patient samples that had not undergone transplantation, it could be concluded that optimal culture conditions were very sample-dependent. The extent of monocytic and/or granulocytic differentiation was variable across conditions, and the condition with minimal differentiation varied with sample. This is most likely due to different mechanisms of leukaemogenesis in different patient samples, leading to different dependencies for survival in culture.

7.4.2 Lentiviral Infection and Editing of Primary AML

Preliminary experiments were presented in Chapter 6 aiming to lentivirally transduce primary haematopoietic cells. It was clear that this work requires optimisation in order to edit primary LSCs. There has been a great deal of work done on optimising the infection of primary $CD34^+$ normal HSPCs [183–185]. There are many experimental factors that can be varied to optimise transduction efficiency, including: the presence/absence and type of serum, cytokine combinations, pre-stimulation time, MOI, time of transduction, spinfection and RetroNectin (as compared in [183, 185]). Through the optimisation of these conditions, transduction efficiencies of up to ~90% have been reported [183]. However, it should be noted that these efficiencies were calculated from myeloid colonies derived from transduced cells, rather than directly from transduced cells.

In the experiments presented in Chapter 6, primary HSPC were spinfected with lentivirus then returned to culture with an MSC stromal layer. It is possible that the MSCs may have had a negative effect on transfection efficiency, despite having a potentially positive effect on viability. Future work should directly compare lentiviral transduction in culture conditions in both the presence and absence of MSCs. In addition to this, the use of RetroNectin may also improve transduction efficiency. Work in Chapter 6 was similar to published optimised protocols in the use of SR1 and UM171 to maintain stemness in culture [85, 183, 223]. Cytokines used for culture were comparable to what has been used in previous work [183–185], but may still require optimisation.

In addition to the lentiviral infection of normal HSPCs, the infection of primary AML cells has been published [224–226]. Firstly, a proof of principle study has been carried out that demonstrated transduction efficiencies ranging from ~10–40% following two rounds of spinfection [224]. Other studies have transduced primary AML cells with the aim of restoring functional PU.1 or C/EBPα proteins ([225] and [226] respectively). In both studies, transduction efficiencies were below 30%. In the case of C/EBPα, three rounds of transduction were done; remarkably, it was seen that, in *in vitro* serial replating experiments, self-renewal was maintained [226].

An alternative to using a lentiviral vector to introduce the CRISPR machinery into primary AML cells would be to use the RNP system. Optimised protocols have been published for the use of RNP with normal $CD34^+$ cells [89]. However, the editing of primary AML cells by RNP has not been published. This method may potentially achieve higher rates of editing; however, editing is transient, with the sgRNA not being incorporated into the genome. Furthermore, if edited cells were to be purified, this would require the incorporation of a fluorescent reporter through the use of an AAV. This incorporation would rely on the less efficient HR method of DNA repair, rather than the NHEJ method typically used when carrying out CRISPR knock-outs.

7.5 Conclusion

Presented here is a comprehensive characterisation and study of the LM progenitors both in normal and leukaemic haematopoiesis. Firstly, these populations were studied in normal haematopoiesis, where single-cell analyses allowed the heterogeneity and continuum underlying normal LM differentiation to be uncovered. Moving onto the LM progenitors in leukaemic haematopoiesis, it was demonstrated that the majority of AML cases showed an LM progenitor-like profile. The leukaemia-propagating ability of these leukaemic LM progenitors was demonstrated in xenotransplantation assays and the immunophenotype, genotype and kinetics of leukaemic engraftment was characterised. When looking into the transcriptional profiles of the LM progenitors in AML, it was seen that these were dysregulated progenitor-like profiles, with an immature LM profile. By comparing the normal and leukaemic LM progenitors, a set of genes upregulated in AML could be identified. This gene set was integrated with genome-wide CRISPR screens, allowing the identification of six putative targets of interest. Further work is required to confirm these targets are functionally essential to LSCs, and the mechanism by which they may facilitate leukaemogenesis. By targeting these genes in a clinical setting, it is possible that treatments against the LSC may be developed, with the goal of truly eliminating the leukaemic disease.

Appendices

Chapter A

Appendices

Contents

A.1 Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells 291
A.2 Flow Cytometric Sorting and Analysis 325
A.3 *In Vitro* LM Differentiation Assays 329
A.4 CRISPR-Cas9 Editing of Leukaemic Cells 330
A.5 *In Vitro* Assays of LM Progenitors 333
A.6 Refining the LMPP and GMP Populations: *In Vitro* Assays 340
A.7 Bulk RNA-Seq of Normal CB HSPC 347
A.8 Immunophenotyping of Normal and Leukaemic BM . 347
A.9 Genetic and Cytogenetic Profiles of AML Samples . . 363
A.10 Summary of Xenotransplantation of AML Experiments 367
A.11 Immunophenotype of Human Engraftment 371
A.12 RNA-Seq Analysis of BM HSPCs 389
A.13 CRISPR-Cas9 Editing of K562 Cells 419
A.14 Optimising AML Culture 421

A.1 Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells

nature immunology

ARTICLES

https://doi.org/10.1038/s41590-017-0001-2

Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells

Dimitris Karamitros[1,2], Bilyana Stoilova[1,2], Zahra Aboukhalil[1], Fiona Hamey[3], Andreas Reinisch[4], Marina Samitsch[1], Lynn Quek[1,2,5], Georg Otto[1], Emmanouela Repapi[1], Jessica Doondeea[1,2], Batchimeg Usukhbayar[1,2], Julien Calvo[6], Stephen Taylor[1], Nicolas Goardon[1], Emmanuelle Six[7], Francoise Pflumio[6], Catherine Porcher[1], Ravindra Majeti[4], Berthold Göttgens[3] and Paresh Vyas[1,2,5]*

The hierarchy of human hemopoietic progenitor cells that produce lymphoid and granulocytic–monocytic (myeloid) lineages is unclear. Multiple progenitor populations produce lymphoid and myeloid cells, but they remain incompletely characterized. Here we demonstrated that lympho-myeloid progenitor populations in cord blood — lymphoid-primed multi-potential progenitors (LMPPs), granulocyte-macrophage progenitors (GMPs) and multi-lymphoid progenitors (MLPs) — were functionally and transcriptionally distinct and heterogeneous at the clonal level, with progenitors of many different functional potentials present. Although most progenitors had the potential to develop into only one mature cell type ('uni-lineage potential'), bi- and rarer multi-lineage progenitors were present among LMPPs, GMPs and MLPs. Those findings, coupled with single-cell expression analyses, suggest that a continuum of progenitors execute lymphoid and myeloid differentiation, rather than only uni-lineage progenitors' being present downstream of stem cells.

Human hemopoiesis produces 10×10^9 new, terminally mature blood cells daily, production that is also rapidly responsive to external change. Most of this production generates red cells, short-lived myeloid cells and platelets. It also replenishes long-lived acquired immune cells and natural killer (NK) cells of the innate immune system. Dysregulation of this complex process can lead to hemopoietic deficiency, immunodeficiency and blood cancer. Active debate continues about the heterogeneity and plasticity of hemopoietic cell populations at steady state and in response to stimuli. At the hierarchy apex lie multi-potent hemopoietic stem cell (HSC) populations, which are heterogeneous in their differentiation potential, cell cycle, self-renewal capacity, stability over time and contribution to hemopoiesis at steady state versus during transplantation[1–11]. Downstream of long-term HSCs in the mouse are heterogeneous short-term HSCs, multipotent progenitors (MPPs) and early lineage-biased progenitors[5,7,12–14]. The short-term HSC–MPP populations in humans have not been fully defined[15,16]. In terms of lineage-potential restriction, the erythroid and megakaryocyte fates probably diverge early from other myeloid and lymphoid potentials in mice[14,17–20] and humans[21–25] and might arise directly from either HSCs[6] or downstream MPPs[14,16,26].

Focusing on the first human lympho-myeloid progenitors downstream of HSCs and MPPs, two progenitor populations have been identified in the immature Lin$^-$CD34$^+$CD38$^-$CD90$^{neg–lo}$ compartment. These include Lin$^-$CD34$^+$CD38$^-$CD90$^{neg–lo}$CD45RA$^+$CD10$^-$ lymphoid-primed multi-potential progenitors (LMPPs) with granulocytic, monocytic, B cell and T cell potential but that are unable to generate erythrocytes or megakaryocytes[22]. Such data support published studies showing that human CD34$^+$CD10$^-$ cells retain lympho-myeloid potential and progressively lose myeloid potential with expression of the cell-surface marker CD10 (refs. [27,28]). In contrast, the multi-lymphoid progenitors (MLPs), initially reported as Lin$^-$CD34$^+$CD38$^-$CD90$^{neg–lo}$CD45RA$^+$CD10$^+$ cells, have lymphoid (B cell, T cell and NK cell), monocytic and dendritic cell (DC) potential but cannot make granulocytes[21]. However, CD10$^-$ MLP populations[29] have been reported that might overlap the LMPPs. Within the Lin$^-$CD34$^+$CD38$^+$CD45RA$^+$ compartment, there are at least two lympho-myeloid progenitors: a CD62L^{hi}CD10$^-$ lymphoid-primed progenitor with lymphoid, monocytic and DC potential[23], and the granulocyte-monocyte progenitor (GMP) (Lin$^-$CD34$^+$CD38$^+$CD45RA$^+$CD123$^+$). The GMP population includes both a CD62L^{hi} subpopulation and a CD62L^{lo} subpopulation and has mainly myeloid potential but retains residual lymphoid potential[22,30], consistent with the mouse pre-granulocyte-monocyte progenitor[31]. Finally, the human Lin$^-$CD34$^+$CD38$^+$CD45RA$^+$ compartment also includes a CD10$^+$ subpopulation with T cell, B cell, NK cell and DC potential but lacking myeloid potential[32]. Such observations have raised questions about whether these progenitor populations are pure or heterogeneous, how distinct they are and the nature of the functional, transcriptional and hierarchical relationships among them.

Lympho-myeloid progenitors in the Lin$^-$CD34$^+$ CD45RA$^+$ CD90$^{neg–lo}$ compartment that can be either CD38$^+$ or CD38$^-$ and either CD10$^+$ or CD10$^-$ have been described. Here we directly and rigorously compared the in vitro and in vivo functional potential and transcriptional programs of human LMPPs, MLPs and GMPs.

[1]MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK. [2]Oxford Biomedical Research Centre, University of Oxford, Oxford, UK. [3]Department of Haematology and Wellcome Trust–Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK. [4]Division of Hematology, Stanford Institute for Stem Cell Biology and Regenerative Medicine, California, USA. [5]Department of Hematology, OUH NHS Trust, Oxford, UK. [6]UMR967 INSERM/CEA, Université Paris 7/Université Paris 11, Paris, France. [7]UMR1163, Imagine Institute, Paris Descartes -Sorbonne Paris Cité University, Paris, France. Dimitris Karamitros and Bilyana Stoilova contributed equally to this work. *e-mail: paresh.vyas@imm.ox.ac.uk

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

We found that these progenitors were distinct and heterogeneous. Single-cell gene-expression analysis indicated a continuum of progenitors with lymphoid and myeloid potential downstream of HSCs. The bulk of the multi-lineage lympho-myeloid progenitors were in a sub-compartment of Lin$^-$CD34$^+$CD38$^-$CD90^{neg-lo}CD45RA$^+$CD10$^-$ LMPPs (called 'LMPP^{mix} cells' here) with the lowest expression, among LMPPs, of the cell-surface markers CD10 and CD45RA.

Results

In vitro assays of the potential of distinct lympho-myeloid progenitors. We improved published flow-cytometry staining and sorting strategies[21,22] to prospectively purify eight human hemopoietic stem–progenitor cell (HSPC) populations (Supplementary Table 1 and Supplementary Fig. 1a) in human cord blood (CB) and bone marrow (BM). These HSPC populations included the following: HSCs (Lin$^-$CD34$^+$CD38$^-$CD90$^+$CD45RA$^-$CD10$^-$), MPPs (Lin$^-$CD34$^+$CD38$^-$CD90$^-$CD45RA$^-$CD10$^-$), LMPPs (Lin$^-$CD34$^+$CD38$^-$CD90^{neg-lo}CD45RA$^+$CD10$^-$), MLPs (Lin$^-$CD34$^+$CD38$^-$CD90^{neg-lo}CD45RA$^+$CD10$^+$), common myeloid progenitors (CMPs) (Lin$^-$CD34$^+$CD38$^+$CD123$^+$CD45RA$^-$CD10$^-$), GMPs (Lin$^-$CD34$^+$CD38$^+$CD123$^+$CD45RA$^+$CD10$^-$), megakaryocyte erythroid progenitor (MEPs) (Lin$^-$CD34$^+$CD38$^+$CD123$^-$CD45RA$^-$CD10$^-$), and B cell–NK cell (B-NK) progenitors (Lin$^-$CD34$^+$CD38$^+$CD90$^-$CD45RA$^+$CD10$^+$) (Supplementary Table 1 and Supplementary Fig. 1a). Among CB Lin$^-$CD34$^+$ cells, these eight HSPC populations accounted for 82% of cells; the remaining cells did not constitute separate populations (Supplementary Table 1 and Supplementary Fig. 1a). The Lin$^-$CD34$^+$CD38$^-$CD45RA$^+$ compartment contained a mixture of CD10$^-$ LMPPs and CD10$^+$ MLPs (Supplementary Fig. 1a). Furthermore, the more-mature Lin$^-$CD34$^+$CD38$^+$ compartment was separated into the CD10$^+$ B-NK progenitor population[32] and CD10$^-$ heterogeneous myeloid progenitors (Supplementary Fig. 1a). Immunophenotypic LMPPs and MLPs were rare (Supplementary Table 1). As assessed with analysis gates, they constituted 0.2% of the BM Lin$^-$CD34$^+$ compartment and ~2 cells among 1×10^5 BM mononuclear cells (MNCs) (Supplementary Table 1). Although they were more frequent in CB, they still constituted only ~1 cell among 1×10^4 CB MNCs (Supplementary Table 1). GMPs were 20-fold more abundant in CB and 100-fold more abundant in BM than were LMPPs or MLPs (~1.5–2 cells among 1×10^3 MNCs) (Supplementary Table 1).

As the frequency of adult BM LMPPs and MLPs was extremely low, we used fresh CB cells as a source of HSPCs. Cells were double-sorted to high purity (of >99%, except CMPs, with a purity of 97%). In methylcellulose-based colony-forming-unit assays, the LMPPs and MLPs had lower myeloid clonogenic potential (6% (LMPPs) and <1% (MLPs)) than that of GMPs (31%) (Fig. 1a,b). GMPs and LMPPs generated granulocyte colonies, monocyte-macrophage colonies and granulocyte and monocyte-macrophage colonies with no or minimal erythrocyte potential (<0.5%) (Fig. 1a). MLPs only generated very few monocyte colonies (Fig. 1a), consistent with published data[15,21,22,33].

We analyzed the lymphoid and myeloiddifferentiation potential of a population of 150 LMPPs, MLPs and GMPs using an optimized, new in vitro liquid culture on MS-5 mouse stromal cells supplemented with the cytokines SCF, G-CSF, FLT3L, IL-2 and IL-15, and the prostaglandin-synthesis inhibitor DUP-697 (called 'SGF15/2 culture' here). By performing a kinetic analysis of lineage output, we determined that 2 weeks was the optimal time at which to detect human CD45$^+$CD15$^+$ (hCD45$^+$CD15$^+$) neutrophil-granulocytes (Gs), hCD45$^+$CD14$^+$ monocyte-macrophages (Ms), hCD45$^+$CD19$^+$ B cells and hCD45$^+$CD56$^+$ NK cells in the culture (Fig. 1c and Supplementary Fig. 1b). Gs, Ms, B cells or NK cells purified by flow cytometry from the in vitro SGF15/2 culture expressed the appropriate lineage-associated genes (Fig. 1d and Supplementary Table 2). Therefore, we analyzed all subsequent limiting-dilution and single-cell cultures at week 2 to capture all four outputs: myeloid (G and M) and lymphoid (B and NK). We assessed the production of T cells by LMPP, MLP and GMP populations at weeks 5 and 7 using an in vitro liquid culture assay on OP9 stromal cells expressing the human Notch receptor ligand DL1 (OP9-hDL1 cells) in the presence of SCF, FLT3L and IL-7 (called 'SF7a culture' here) (Fig. 1e and Supplementary Fig. 1c). LMPPs, GMPs and MLPs generated hCD7$^+$CD1a$^+$ cells of the T cell lineage, including hCD7$^+$CD1a$^+$ CD4$^+$CD8$^+$ (double-positive) T cells and hCD7$^+$CD1a$^+$ CD4$^-$CD8$^+$ or CD4$^+$CD8$^-$ (single-positive) T cells (Fig. 1e and Supplementary Fig. 1c). T cell subpopulations purified by flow cytometry expressed the appropriate lineage-associated genes (Fig. 1f and Supplementary Table 2). In summary, we established conditions with which to prospectively purify eight HSPC populations and to test their in vitro potential to develop into myeloid and lymphoid lineages.

Functional heterogeneity of lympho-myeloid progenitors. We next used four in vitro liquid culture assays to assess the clonal potential of CB LMPPs, MLPs and GMPs (Supplementary Fig. 2a,b). First, we performed limiting-dilution assays of SGF15/2 cultures and assessed lineage output by flow cytometry (Supplementary Fig. 2c). One in two LMPPs produced B cells and one in three LMPPs produced NK cells, but only one in five LMPPs produced monocytes and one in ten LMPPs produced granulocytes (Table 1 and Supplementary Fig. 2d). GMPs generated myeloid cells with higher frequency than did LMPPs (one in two for Ms and one in four for Gs) and generated lymphoid cells at lower frequency than did LMPPs (one in twenty-two for B cells and one in eight for NK cells) (Table 1 and Supplementary Fig. 2d). 1 in 11 MLP cells produced B cells and 1 in 18 generated NK cells, whereas myeloid output was rare (Table 1), which indicated that MLPs were lymphoid biased. Bi-lineage and multi-lineage cells were detected at a lower frequency (1 in 6 to 1 in 789) (Table 1 and Supplementary Table 3a).

As the limiting-dilution assay does not rigorously define the frequency of multi-lineage functional potential at a clonal level, we assessed lympho-myeloid (B, NK, G and M) potential in the optimized SGF15/2 culture. We assessed the potential of 1,136 LMPPs, 710 MLPs and 1,622 GMPs as single cells, isolated from 22 CB donors (a total of 6.3×10^9 MNCs), to provide robust quantitative data, especially for rare functional potentials (Supplementary Table 3b). At the single-cell level, LMPPs and GMPs had higher cloning efficiency (54% and 71%, respectively) than that of MLPs (11%) (Fig. 2a). LMPPs and MLPs were mainly lymphoid progenitors, whereas GMPs were mainly myeloid progenitors (Fig. 2a). Focusing on productive wells (defined as wells that produced any type of mature output), we found that 69% of LMPPs, 88% of MLPs and 63% of GMPs produced uni-lineage output (Fig. 2b). When there was uni-lineage output, 92% of LMPPs had lymphoid output (B or NK) and 8% had myeloid output (G or M) (Fig. 2b). The MLPs were almost exclusively lymphoid progenitors with very low myeloid output (3%) (Fig. 2b). 79% of GMP cells had myeloid output, and 21% had lymphoid output (Fig. 2b). Bi-lineage output was detected in 24% of LMPPs, 12% of MLPs and 33% of GMPs (Fig. 2c), and output of three or more lineages was rare (6% of LMPPs, 0% of MLPs and 3% of GMPs) (Fig. 2d). Only 8% of all plated LMPPs, 7% of GMPs and hardly any MLPs (0.3%) exhibited combined lympho-myeloid potential (Fig. 2e). The lympho-myeloid output of LMPPs was significantly higher than that of GMPs ($P=0.0125$) or MLPs ($P=0.0019$, Supplementary Table 3c).

We also assessed the lympho-myeloid (B, NK, G and M) potential of 96 LMPPs, 52 MLPs and 110 GMPs as single cells in a third in vitro liquid culture assay with MS-5 stroma in the presence of the cytokines SCF, IL-7, thrombopoietin, IL-2, GM-CSF, G-CSF and M-CSF (called 'S7T2GM/G/M culture' here) that was used to define the MLPs[21] (Supplementary Fig. 2b). We obtained results similar to those of SGF15/2 culture, but S7T2GM/G/M cultures were less

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

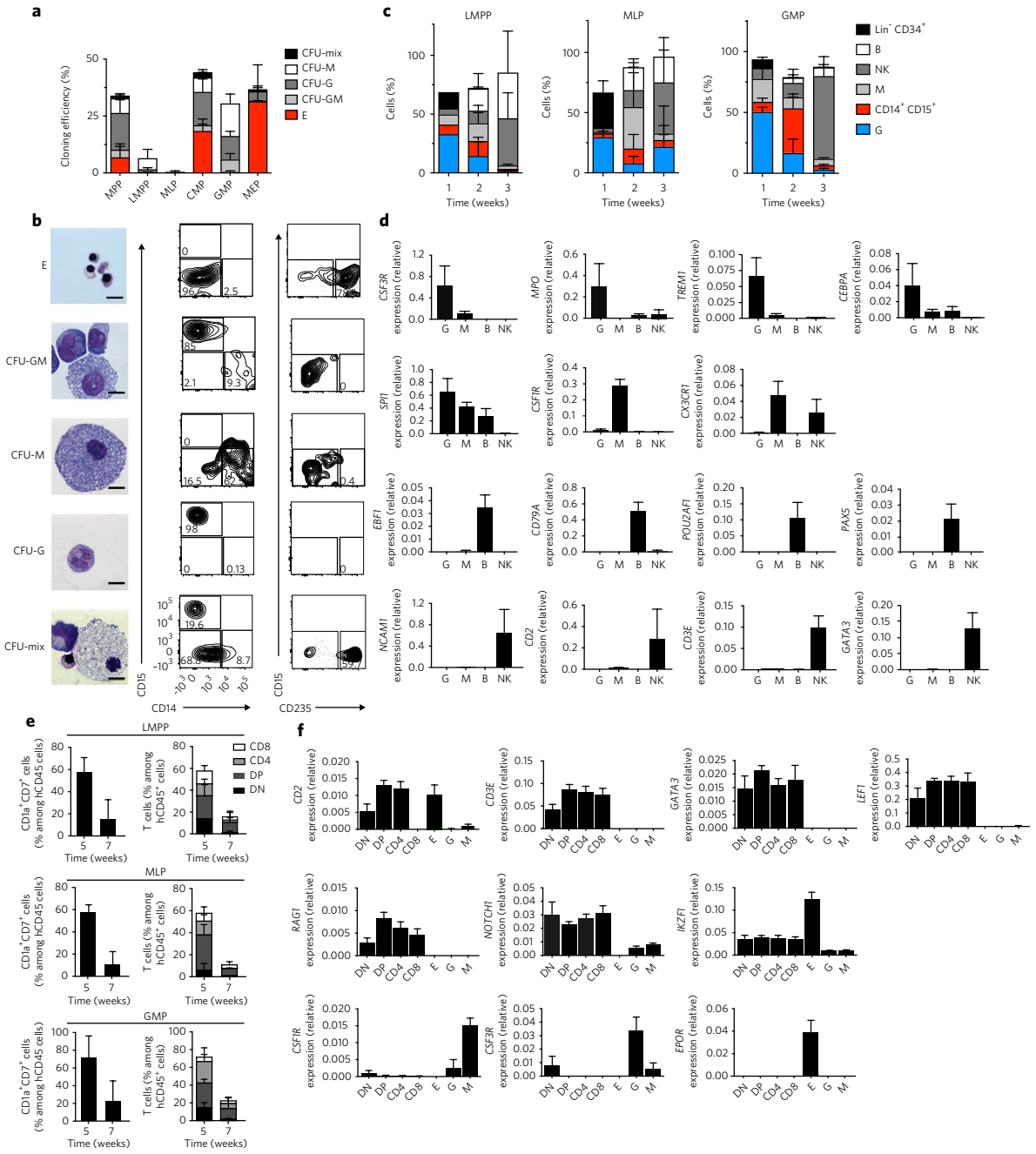


Fig. 1 | Human CB lympho-myeloid populations have distinct functional potential in vitro. a, Colony-forming unit (CFU) assay of the cloning efficiency and myelo-erythroid lineage affiliation of CB HSPCs (150 cells plated; horizontal axis), showing the frequency of mixed erythro-myeloid colonies (CFU-mix), monocyte-macrophage colonies (CFU-M), granulocyte colonies (CFU-G), granulocyte and monocyte-macrophage colonies (CFU-GM) and erythroid colonies (burst-forming units-erythroid and CFU-E (E)) (key). **b**, Microscopy showing the morphology of cells from the CFU assay in **a** (left margin), stained with May–Grunwald–Giemsa (left), and flow cytometry of cells harvested from the colony types in **a** (middle and right). Scale bars (left), 10 μm. Numbers in outlined areas (middle and right) indicate percent $CD15^+CD14^-$ neutrophil granulocytes (top left), $CD15^-CD14^+$ monocyte-macrophages (bottom right) or $CD15^-CD14^-$ cells (bottom left) (middle column), or $CD15^-CD14^-CD235^+$ erythroid cells (bottom right) (right column). **c**, Lineage output (key) of LMPPs, MLPs and GMPs (above plot; 150 each) after 1, 2 or 3 weeks (horizontal axis) in SGF15/2 culture, presented as frequency among $hCD45^+$ cells (flow cytometry of 2-week cultures, Supplementary Fig. 1b). **d**, Gene expression by output cells from **c** (horizontal axis), purified by flow cytometry; results are presented relative to those of the control genes *GAPDH* and *ACTB*. **e**, T cell output by LMPPs, MLPs and GMPs (above plots) in bulk after 5 or 7 weeks (horizontal axis) in SF7a culture, showing the frequency of $CD7^+CD1a^+$ cells of the T cell lineage (left), and of $CD7^+CD1a^+$ $CD4^-CD8^+$ cells (CD8), $CD7^+CD1a^+$ $CD4^+CD8^-$ cells (CD4), $CD7^+CD1a^+$ $CD4^+CD8^+$ cells (DP) and $CD7^+CD1a^+$ $CD4^-CD8^-$ cells (DN) (key) (right) among $hCD45^+$ cells (flow cytometry of 5-week cultures, Supplementary Fig. 1c). **f**, Gene expression by output T cells from **e**, purified by flow cytometry (left half of horizontal axis; as in key in **e**), and by control mature non-T cells obtained from the sorting of cells from E, G and M colonies from **a** (right half of horizontal axis); results are presented relative to those of *GAPDH* and *ACTB*. Data are representative of three experiments (**a**; error bars, mean + s.d. of five donors), two experiments (**b**), two experiments (**c**; mean + s.d of three donors), two experiments (**d**; mean + s.e.m. of three donors), three experiments (**e**; mean + s.d. of five donors) or three experiments (**f**; mean + s.e.m. of three donors).

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

permissive for granulocytic output (Supplementary Fig. 2e–i and Supplementary Table 3d). Most of the output from LMPPs was uni-lineage, with rarer bi-lineage and less frequent multi-lineage output (Supplementary Fig. 2e–i and Supplementary Table 3d). MLPs exhibited only lymphoid uni-lineage output (Supplementary Fig. 2e–i and Supplementary Table 3d).

Finally, we assessed the lympho-myeloid potential of 215 LMPPs, 197 MLPs and 219 GMPs as single cells in a fourth assay, an independent culture condition optimized for the detection of combined lymphoid (B, NK and T) and myeloid (M and G) potential. Single LMPPs, MLPs and GMPs were cultured in the presence of SCF, FLT3L and IL-7 on an MS-5 stromal cell line that can be induced to express hDL1 by doxycycline treatment (MS-5–hDL1^{IND} stroma) (called 'SF7b/Dox culture' here). The B-NK-M-G output was analyzed at 3 weeks and the T cell output was analyzed at 6 weeks (Fig. 2f–j and Supplementary Table 3e). Uni-lineage T cell output was detected in LMPP and MLP populations (3% of productive wells) but was almost completely absent in GMP populations (<0.1%) (Fig. 2g). T cell output combined with other lymphoid output was detected for 1–5% of LMPPs and MLPs and rarely for GMPs (Fig. 2g). Lympho-myeloid output was detected only for LMPPs (14%) (Fig. 2i). Overall, 24 functionally different progenitor types were identified in the three single-cell in vitro clonal assays; all 24 progenitor types were observed for LMPPs, and only subsets of those were noted for MLPs and GMPs (Supplementary Fig. 2j).

Table 1 | Frequency of lineage output

Output	LMPP	MLP	GMP
B	2	11	22
NK	3	18	8
M	5	194	2
G	10	394	4
B_NK	6	38	31
M_G	16	ND	7
B_M	10	392	32
B_G	13	789	38
M_NK	9	392	11
NK_G	15	ND	17
B_NK_M	16	392	44
B_NK_G	18	ND	49
B_M_G	22	ND	43
NK_M_G	23	ND	20
B_NK_M_G	29	ND	56

Limiting-dilution assay of the frequency of various lineage outputs (far left column) of LMPPs, MLPs and GMPs (above columns), presented as the 'X' value in the calculation '1 in X cells can give rise to (the output specified)'. ND, not detected. Data are representative of two experiments with four donors.

Ossicle assay defines the in vivo potential of LMPPs, MLPs and GMPs. Successful single-cell transplantation of human progenitors in xenotransplantation assays is not feasible. Furthermore, direct injection of progenitor populations into immunodeficient mice yields low engraftment (<0.1%)[15,21,22,25]. Therefore, we assessed in vivo progenitor function in a new humanized ossicle model[34]. Human BM-derived mesenchymal stromal cells were subcutaneously injected into immunodeficient mice, where they form a humanized ossicle over 8 weeks. LMPPs, MLPs and GMPs were injected into the ossicle, and lineage output was analyzed 1 and 2 weeks after injection (Supplementary Fig. 3a,b). Engraftment was detected at both time points, with greater engraftment of hCD45$^+$hCD33$^+$hCD14$^+$ cells (M), hCD45$^+$hCD33$^+$hCD15$^+$ cells (G) and hCD45$^+$hCD33$^-$hCD19$^+$ cells (B) at week 2 than at week 1 (data not shown). All subsequent analyses were done at 2 weeks after transplantation. As the number of cells injected varied (~300–60,000 cells, depending on the progenitor subset; Supplementary Fig. 3c), we report the mean cell engraftment per 1,000 transplanted cells here. GMPs had the greatest mean engraftment (2.6%), followed by LMPPs (1.4%) and MLPs (0.2%) (Fig. 3a). LMPPs produced more CD33$^+$ myeloid cells (82%) than CD19$^+$ B cells (17%) (Fig. 3b,c). MLPs generated more B cells (78% ± 5.9%) than myeloid cells (19% ± 6.7%) (mean ± s.e.m.) (Fig. 3b,c). There was no correlation between the number of the transplanted cells and the ratio of lymphoid cells to myeloid cells (Supplementary Fig. 3d). GMPs generated mainly myeloid cells (97%) (Fig. 3b,c). Myeloid cells generated from LMPPs and GMPs expressed markers of monocytes (CD14) and granulocytes (CD15) (Fig. 3c). No CD14$^+$ or CD15$^+$ cells were detected as being produced from MLPs (Fig. 3c). Morphological analysis of the engrafted cells confirmed that CD15$^+$ cells were granulocytic and CD14$^+$ cells were monocytic (Fig. 3d). CD14$^+$CD15$^+$ cells, generated by LMPPs and GMPs (Fig. 3c), were immature myeloid cells by morphology (Fig. 3d). Thus, LMPPs, MLPs and GMPs had different lymphoid potential and myeloid potential in the humanized ossicle assay.

LMPP, MLP and GMP transcriptional programs correlate with function. We performed RNA sequencing of human CB HSPC populations (HSC, MPP, LMPP, MLP, CMP, GMP and MEP). Hierarchical clustering using all expressed genes separated LMPPs and MLPs from the other HSPCs (Fig. 4a and Supplementary Fig. 4a). HSCs and MPPs clustered away from mature progenitors in this analysis (Fig. 4a and Supplementary Fig. 4a). We used analysis of variance (ANOVA) to identify genes with differential expression ('differentially expressed genes' (DEGs)) across the seven HSPC populations (Supplementary Table 4). We performed principal-component analysis (PCA) using all expressed genes or between 300 and 10,000 of the genes with the most significantly differential expression ('most significant DEGs'; ranked by ANOVA *P* value) (Fig. 4b). The best separation of HSPC populations on a PCA plot was achieved with the 300 most significant DEGs (Fig. 4b). Principal component 1 (PC1) separated HSPCs by lineage potential, and PC2 separated HSPCs by maturation (Fig. 4b). By comparing the eigenvalues of the 300 most significant DEGs against those from a randomized data set, we demonstrated that PC1–PC3 captured most of the variation between populations (Supplementary Fig. 4b,c). We also identified the genes with the greatest variance across all populations without assuming population identity (Fig. 4b and Supplementary Fig. 4d). PCA plotting using this gene set produced similar results (Fig. 4b and Supplementary Fig. 4d). The loadings plot for the PCA using the 300 most significant DEGs identified genes associated with stem cells (*HLF*, *MECOM* and *NFIB*), lymphoid cells (*IGJ*, *IRF8* and *MME*) and erythroid-megakaryocytic cells (*HBD* and *HPGDS*) (Fig. 4c). Hierarchical clustering using that same gene set separated HSPC populations (Fig. 4d). GMPs had the highest expression of *ELANE*, *MPO* and *PRTN3*, whereas LMPPs and MLPs shared expression of many lymphoid-cell-associated genes (for example, *IL-7R*, *LCK*, *SYK*, *ADA*, *HLX*, *LST1* and *ITGAL*) (Fig. 4d).

Transcriptional relatedness between HSPC populations, without assumption of any hierarchical relationships, was further analyzed through pairwise comparisons. The most closely related populations were HSCs and MPPs (only 13 DEGs separated these populations), while LMPPs and MLPs were also closely related (85 DEGs) (Fig. 4e, Supplementary Fig. 4e and Supplementary Tables 5–11). GMPs were most closely related to CMPs (40 DEGs) but retained similarity to LMPPs (183 DEGs) (Fig. 4e, Supplementary Fig. 4e and Supplementary Tables 5–11). We derived gene-expression signatures for LMPPs, MLPs and GMPs from DEGs in one-versus-all

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

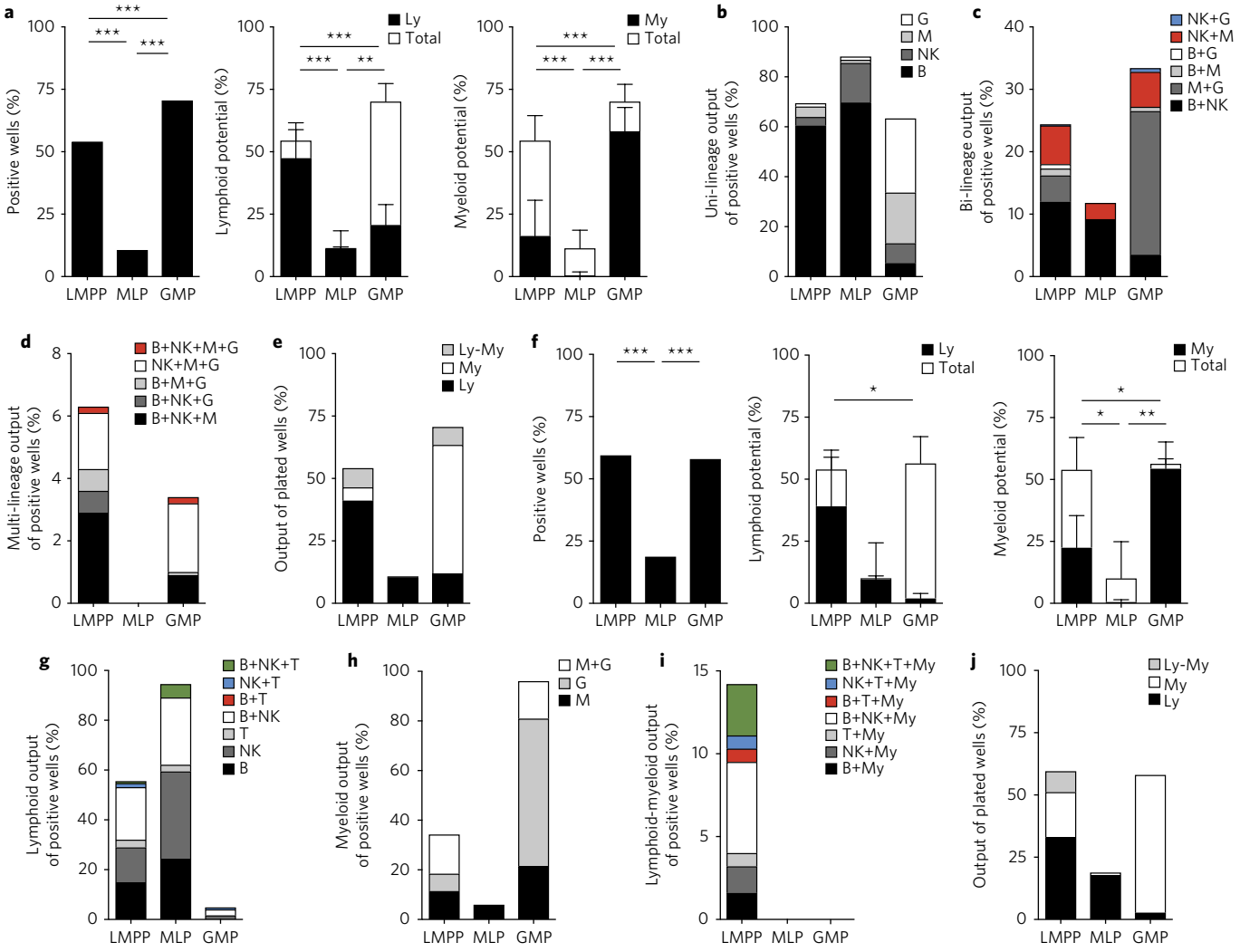


Fig. 2 | CB LMPPs and GMPs are lympho-myeloid progenitors, while MLPs are mainly lymphoid progenitors in clonal in vitro assays. a, Total cloning efficiency of single LMPPs (615 of 1,136 cells), MLPs (76 of 710 cells) and GMPs (1,145 of 1,622 cells) in SGF15/2 culture (left), and cloning efficiency of lymphoid lineages (middle) and myeloid lineages (right), showing total cloning efficiency (Total) and proportion of lymphoid potential (lymphoid plus mixed; Ly) or myeloid potential (myeloid plus mixed clones; My) (key). **b**–**d**, Uni-lineage output (**b**), bi-lineage output (**c**) and multi-lineage output (**d**) of single CB LMPPs, MLPs and GMPs (horizontal axis) in SGF15/2 culture, presented as the proportion of cells of various lineages (key) in wells that produced cells (positive wells). **e**, Lymphoid, myeloid and lympho-myeloid (Ly-My) output (key) of LMPPs, MLPs and GMPs (horizontal axis) in SGF15/2 culture, presented as the frequency among all plated cells. **f**, Total cloning efficiency of single CB LMPPs (128 of 215 cells), MLPs (37 of 197 cells) and GMPs (127 of 219 cells) in SF7b/Dox culture (left), and cloning efficiency of lymphoid lineages (middle) and myeloid lineages (right), showing total cloning efficiency and proportion of lymphoid or myeloid potential (key). **g**–**i**, Uni-lineage output (**g**), bi-lineage output (**h**) and multi-lineage output (**i**) of single CB LMPPs, MLPs and GMPs (horizontal axis) in SF7b/Dox culture, presented as the frequency of cells of various lineages (key) in positive wells. **j**, Lymphoid, myeloid and lympho-myeloid output (key) of LMPPs, MLPs and GMPs (horizontal axis) in SF7b/Dox culture, presented as the frequency among all plated cells. NS, not significant ($P > 0.05$); *$P = 0.01$–0.05, **$P = 0.001$–0.01 and ***$P < 0.001$ (Fisher's exact test (**a**,**f**, left), Students *t*-test (**a**,**f**, middle and right). Data are representative of nine experiments with 22 donors (**a**, left, and **b**–**e**), nine experiments (**a**, middle and right; mean + s.d. of 22 donors), one experiment with three donors (**f**, left, and **g**–**j**) or one experiment (**f**, middle and right; mean + s.d. of three donors).

population comparisons filtered for uniquely expressed genes (Fig. 4f and Supplementary Table 12a–c). The GMP signature included many myeloid-cell-associated genes, and the MLP signature included many lymphoid-cell-associated genes (Fig. 4d). In contrast, the LMPP signature included both lymphoid-cell-associated genes (*ETS1*, *EBF1* and *CYTIP*) and myeloid-cell-associated genes (*TRPM2*, *S100A8*, *PADI4* and *ALOX15B*) (Fig. 4f).

To confirm those findings, we investigated the profiles of LMPPs, GMPs and MLPs using published gene sets[25]. GMPs expressed genes associated with immature myeloid cells, whereas LMPPs and MLPs expressed genes associated with B cells, monocytes and DCs but not those associated with neutrophils (with the exception of *FOSB*) (Supplementary Fig. 4f). Additionally, in a MetaCore enrichment analysis, the GMPs showed enrichment for pathways (expressed a significant number of components of these pathways) associated with myeloid maturation (for example, granulocyte development (false-discovery rate (FDR) = 0.0136)), whereas MLPs showed enrichment for lymphopoiesis pathways (for example, Notch signaling (FDR < 0.001)) (Supplementary Table 13). LMPPs showed more balanced enrichment for both lymphoid pathways and myeloid pathways (for example, M-CSF signaling (FDR < 0.001) and BCR signaling (FDR = 0.049)) (Supplementary Table 13).

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

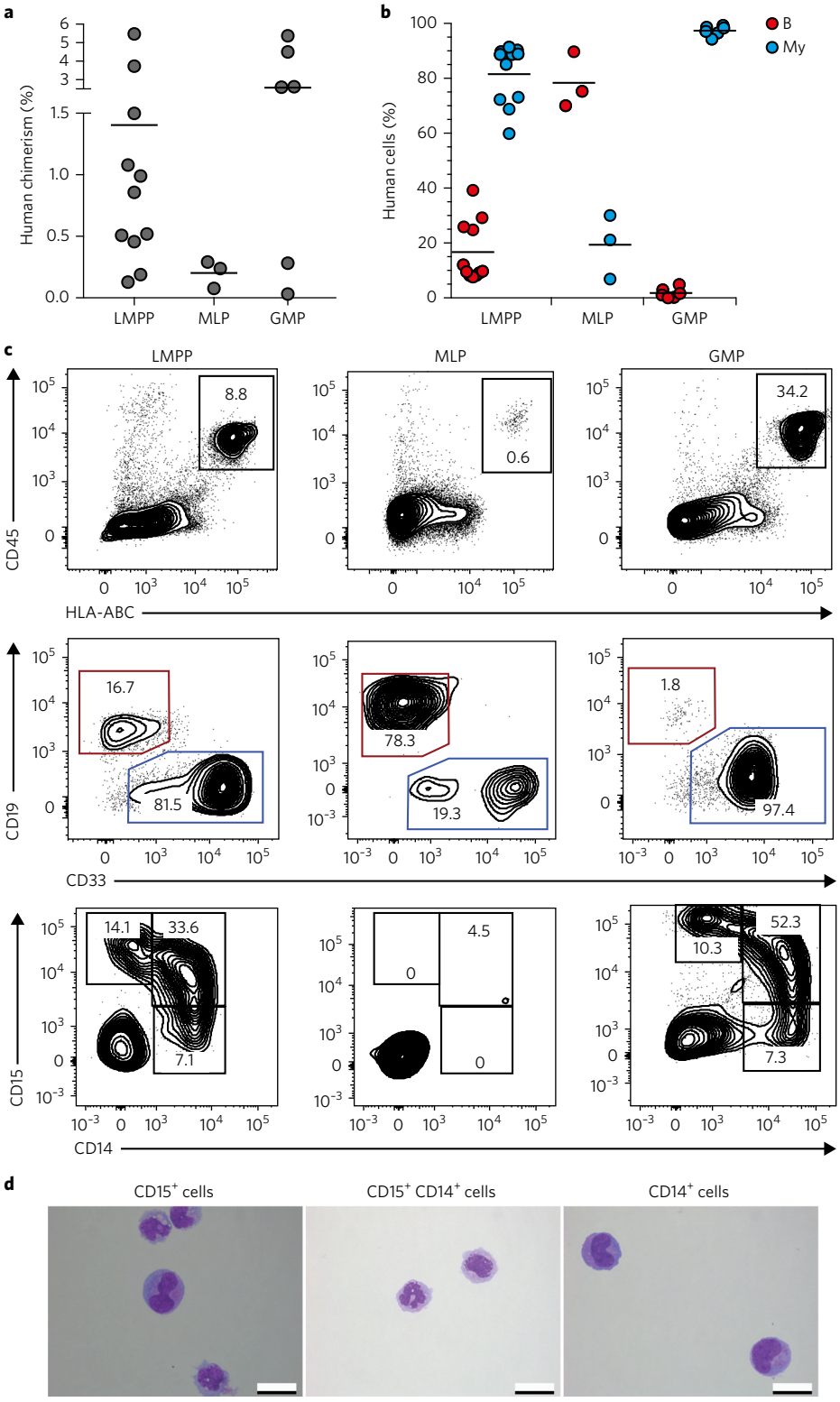


Fig. 3 | Human CB LMPPs, MLPs and GMPs have distinct differentiation potential in vivo. a, Engraftment of human LMPPs, MLPs and GMPs (horizontal axis) into humanized ossicles in immunodeficient mice 2 weeks after transplantation, normalized to 1,000 transplanted cells and presented as percent human chimerism. **b**, Frequency of B cells and myeloid cells (key) among human CD45$^+$HLA-ABC$^+$ cells generated by LMPPs, MLPs and GMPs (horizonal axis) after transplantation as in **a**. **c**, Flow cytometry of cells from host mice at 2 weeks after transplantation as in **a**. Numbers in outlined areas indicate percent human CD45$^+$HLA-ABC$^+$ cells (human engraftment) (top row); CD19$^+$ (B) cells (top left) or CD33$^+$ (myeloid) cells (bottom right) (middle row); or CD15$^+$CD14$^-$ granulocytes (top left), CD15$^+$CD14$^+$ immature myeloid cells (top right) or CD15$^-$CD14$^+$ monocytes (bottom right) (bottom row). **d**, Microscopy of CD15$^+$CD14$^-$, CD15$^+$CD14$^+$ and CD15$^-$CD14$^+$ myeloid cells (above images) generated by LMPPs at 2 weeks after transplantation as in **a**, stained with May–Grunwald–Giemsa. Scale bars, 20 μm. Each symbol (**a**, **b**) represents an individual recipient mouse; small horizontal lines indicate the mean. Data are from ten experiments (**a**–**c**; mean of $n=11$ donors (LMPP), $n=3$ donors (MLP) or $n=6$ donors (GMP)) or two experiments with two donors (**d**).

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

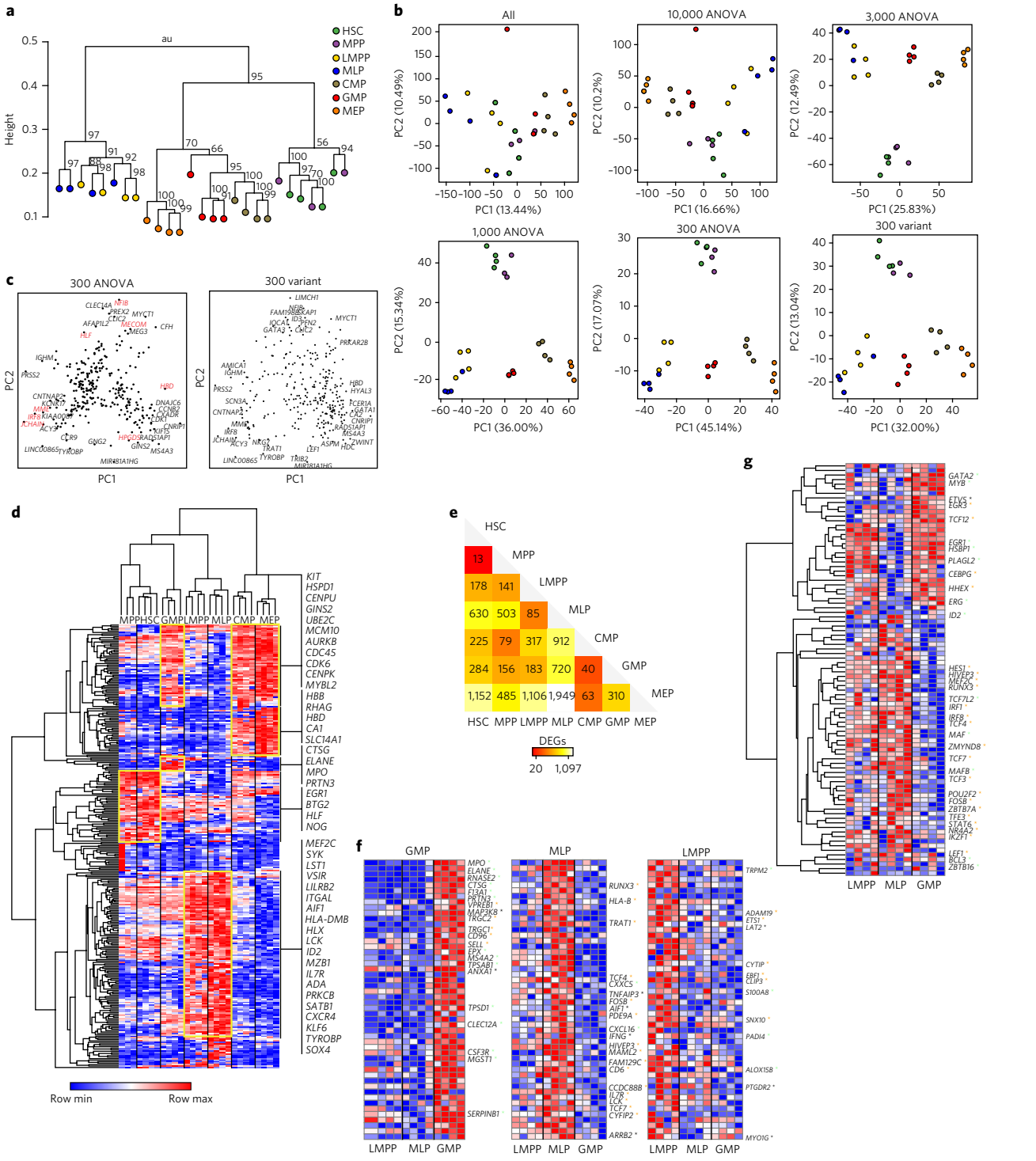


Fig. 4 | Distinct transcriptional patterns of human CB HSPC populations. a, Hierarchical clustering of HSPC populations (key) using all genes and 1,000 bootstrap permutation analyses, with height calculated as (1 – [correlation co-efficient]); numbers above dendrogram bars in plot indicate approximate unbiased *P* values (au). **b**, PCA of CB HSPCs with all genes (top left) or the 300, 3,000 or 10,000 most significant DEGs, ranked by ANOVA *P* value (called, for example, 'top 300 ANOVA genes' subsequently; above plots), or the 300 genes with the greatest variance in expression (300 variant; bottom right); numbers in parentheses along axes indicate the proportion of variance represented by PC1 and PC2. **c**, Loadings plot showing the genes with the most extreme loadings scores for the PCA run with the top 300 ANOVA genes (left) or the 300 genes with the greatest variance in expression (right). **d**, Expression (key; normalized per gene) of the top 300 ANOVA genes (right margin; one per row) by HSPC populations (columns), showing clusters with distinct expression patterns across HSPC populations (yellow outlines); dendrograms (top and left margins) indicate hierarchical clustering (of columns and rows, respectively). **e**, Number of DEGs (key; specific number in plot) in various comparisons of HSPC populations (rows and columns). **f**,**g**, Expression (as in **d**) of the 50 most significant DEGs, ranked by *P* value, in the LMPP, MLP and GMP gene signatures (above plots) (**f**), and of transcription-factor-encoding genes expressed differentially across HSPC populations (**g**); asterisk colors indicate genes associated with the lymphoid lineage (orange) or myeloid lineage (green) or genes encoding products associated with immunological function (black). Data are representative of three experiments with three donors (MPPs) or four donors (all other cell populations).

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

We used two approaches to pinpoint transcription factors driving the programs noted above. First, we identified transcription factors expressed differentially by MLPs relative to their expression by GMPs (Supplementary Fig. 4g). Second, we assessed the expression of hemopoietic transcription factors identified previously[35] (Fig. 4g). In both analyses, GMPs expressed mainly genes encoding myeloid-cell-associated transcription factors (for example, *ERG*, *GATA2*, *MYB* and *EGR1*), while lymphoid-cell-associated transcription factors (for example, *HES1*, *RUNX3*, *POU2F2*, *LEF1*, *IKZF1*, *IRF8*, *TCF4*) had their highest expression in MLPs (Fig. 4g and Supplementary Fig. 4g). LMPPs showed expression of both lymphoid-cell-associated transcription factors and myeloid-cell-associated transcription factors (Fig. 4g and Supplementary Fig. 4g). A similar trend was seen for genes encoding cytokine and chemokine receptors (Supplementary Fig. 4h). Therefore, the transcriptional programs of LMPPs, MLPs and GMPs reflected their functional potential.

Single-cell RNA analysis reveals a continuum of differentiation. To begin to separate distinct progenitors in the heterogeneous GMP, LMPP and MLP populations, we sorted single cells by index flow cytometry for functional analysis, RNA sequencing and quantitative RT-PCR (qRT-PCR). Index data allowed correlation of function and transcriptional state[36] (Fig. 5a). First, we profiled the expression of 96 genes encoding lineage-associated transcriptional regulators and cell-surface and lineage-associated markers (Supplementary Table 14) by a total of 919 single LMPPs, MLPs and GMPs. Genes with low variance and levels of detection were excluded. The expression of 74 genes was taken forward for analysis. Hierarchical clustering assigned GMPs, LMPPs and MLPs to three clusters: cluster 1 (543 cells) was mainly MLPs and LMPPs, cluster 2 (150 cells) was a mixture of GMPs, LMPPs and MLPs, and cluster 3 (226 cells) was mainly GMPs (Supplementary Fig. 5a,b). Cluster 1 showed higher expression of lymphoid-cell-associated genes than that of the other clusters, and cluster 3 showed higher expression of myeloid-cell-associated genes than that of the other clusters (Supplementary Fig. 5b). Cluster 2 had a mixed lympho-myeloid expression profile (Supplementary Fig. 5b). The cellular composition of each gene-expression cluster mirrored the single-cell functional output (Supplementary Fig. 5b).

We performed dimensionality reduction of gene-expression data using a diffusion- map method adapted for single cell data[37,38]. We indicated progenitor identity on the diffusion map (Supplementary Fig. 5c) and found that MLPs, LMPPs and GMPs cells formed a continuum, in agreement with the hierarchical clustering (Supplementary Fig. 5a). Next, we colored the diffusion map by cluster assignment (Supplementary Fig. 5c). Cluster 2 was positioned between clusters 1 and 3, in agreement with its mixed lympho-myeloid transcriptional signature (Supplementary Fig. 5b).

To overcome gene-selection bias in qRT-PCR data, we performed single-cell RNA sequencing and correlated this with function for 91 LMPPs, 110 MLPs and 119 GMPs from two different donors (157 cells and 163 cells from each donor). Clustering using the union (combined set) of variable genes (above the level of technical variance calculated from external control RNA) within each donor identified three clusters (Fig. 5b and Supplementary Fig. 5d). Most cells in cluster 1 were MLPs and most cells in cluster 3 were GMPs, while cluster 2 comprised LMPPs and GMPs (Supplementary Fig. 5e). Cluster 1 showed high expression of lymphoid-cell-associated genes (for example, *MME*, *JCHAIN* and *ABCA1*) (Fig. 5b and Supplementary Fig. 5d). Cluster 3 showed higher expression of myeloid genes (for example, *CPA3, MPO* and *VIM*) than that of the other clusters (Fig. 5b and Supplementary Fig. 5d). Cluster 2 showed a mixed transcriptional signature and higher expression of the hemopoietic-progenitor-cell-associated gene *KIT* that of the other clusters (Fig. 5b and Supplementary Fig. 5d). PCA revealed a transcriptional continuum of LMPP, MLP and GMP populations (Fig. 5c,d). Identical analysis of the second donor produced similar results (Supplementary Fig. 5f). Overall, the single-cell transcriptional profiles of the LMPPs, MLPs and GMPs suggested a continuum of lympho-myeloid differentiation for the currently defined LMPPs, MLPs and GMPs.

Refined sorting strategies purify lymphoid CD10^{hi}CD45RA^{hi} LMPPs and myeloid CD38^{hi} GMPs. As our data showed that current sorting by flow cytometry does not purify functionally homogenous populations, we correlated surface-marker expression with function for the LMPPs and GMPs, as they showed the greatest functional heterogeneity. Flow-cytometry indexing data showed that single LMPPs with lymphoid output had significantly higher expression of CD10 and CD45RA than that of LMPPs with myeloid and lympho-myeloid output (Fig. 6a,b). That result was confirmed by the higher expression of CD10 in single LMPPs in the lymphoid-biased cluster 1 than in single LMPPs in the myeloid-biased cluster 3 (Supplementary Fig. 6a). Therefore, we developed a new strategy for sorting LMPPs by flow cytometry to purify CD10^{hi}CD45RA^{hi} LMPPs (called 'LMPP^{ly} cells' here) and CD10^{lo}CD45RA^{lo} LMPPs (called 'LMPP^{mix} cells' here) (Supplementary Fig. 6b), aiming to maximize the lymphoid-only potential and the mixed myeloid and lympho-myeloid potential, respectively. 26% of total LMPPs were LMPP^{ly} cells and 27% were LMPP^{mix} cells (Fig. 6c). When cultured in SGF15/2 conditions and analyzed after 2 weeks, LMPP^{ly} cells had significantly lower cloning efficiency than that of LMPPs and LMPP^{mix} cells but significantly higher cloning efficiency than that of MLPs (Fig. 6d and Supplementary Table 15; $P<0.0001$, for all comparisons (Fisher's exact test)). LMPP^{ly} cells were lymphoid progenitors with almost no myeloid potential and significantly lower myeloid potential than that of LMPPs or LMPP^{mix} cells (Fig. 6d and Supplementary Table 15; $P=0.0496$ and $P=0.0280$, respectively (Fisher's exact test)). LMPP^{ly} cells had very limited residual lymphoid-myeloid potential (1.6%) (Fig. 6h). LMPP^{mix} cells retained almost all of the myeloid potential and most of the lympho-myeloid potential that the LMPPs had (Fig. 6e–h and Supplementary Table 15). This suggested that functionally, LMPP^{ly} cells were intermediate between LMPPs and MLPs. That was confirmed by a second in vitro culture (SF7b) (Fig. 6i, Supplementary Fig. 6c–f and Supplementary Table 15).

On the basis of flow-cytometry indexing data, GMPs with myeloid-only output had significantly higher expression of the costimulatory molecule CD38 than that of GMPs with lympho-myeloid or lymphoid output (Fig. 7a,b). Concordantly, single GMPs in cluster 3 (with the highest myeloid potential) had significantly higher CD38 expression than that of GMPs in cluster 1 (with the highest lymphoid potential) or cluster 2 (with lymphoid and myeloid potential) (Supplementary Fig. 6g). There was a significant positive correlation between expression of CD38 and expression of a myeloid-cell-associated gene (*MPO*), and a negative correlation between expression of CD38 and expression of lymphoid-cell-associated genes (*MME* and *SELL*), as assessed by single cell qRT-PCR ($P=2.2\times10^{-16}$ and $\rho=0.53$ (*MPO*), $P=7.1\times10^{-5}$ and $\rho=-0.22$ (*MME*), and $P=1.3\times10^{-5}$ and $\rho=-0.24$ (*SELL*) (Spearman's rank correlation coefficient); Supplementary Fig. 6h). To purify a GMP sub-population without lymphoid potential on the basis of CD38 expression, we divided the entire Lin$^-$CD34$^+$ population into CD38^{hi} cells (44% of CD38$^+$ cells), CD38^{lo} cells (15% of CD38$^-$ cells) and CD38^{int} cells (an area between the two new gates) (Fig. 7c and Supplementary Fig. 6i). We further purified CD38^{hi}, CD38^{int} and CD38^{lo} cells to isolate CD38^{hi} GMPs, CD38^{int} (CD38^{int}CD45RA$^+$CD10$^-$) cells and CD38^{lo} LMPPs. CD38^{lo} LMPPs were rare (1 in 1×10^8 MNCs), and no conclusions could be reached about their functional potential (data not shown). We compared the in vitro lineage potential of single CD38^{hi} GMPs (279 cells) and

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

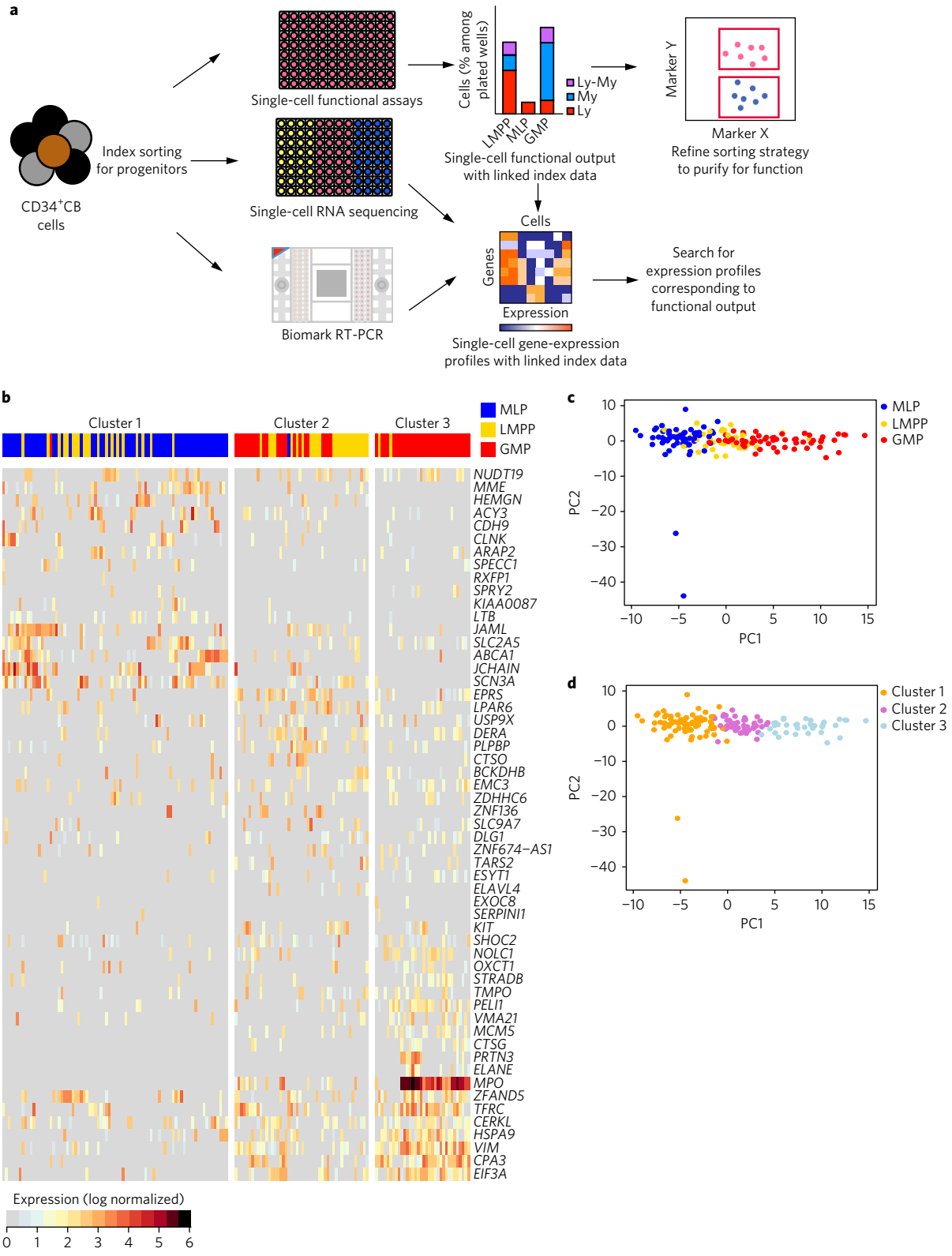


Fig. 5 | Transcriptional heterogeneity of CB lympho-myeloid progenitor cells by single-cell RNA sequencing. a, Strategy for combining single-cell functional analysis, single-cell RNA sequencing and single-cell qRT-PCR based on flow-cytometry index data. **b**, Clustering of single LMPPs, GMPs and MLPs (top right key; one cell per column) by expression (bottom key; log-normalized values) of the 55 genes (right margin; one per row) with the highest and most variable expression between clusters. **c**,**d**, PCA of the 5,388 genes with highly variable expression (above the level of technical variance calculated from external control RNA), assessed by cell type (key) (**c**) or cluster membership (key) (**d**). Data are from one donor of two donors, in one experiment (other donor, Supplementary Fig. 5d–f).

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

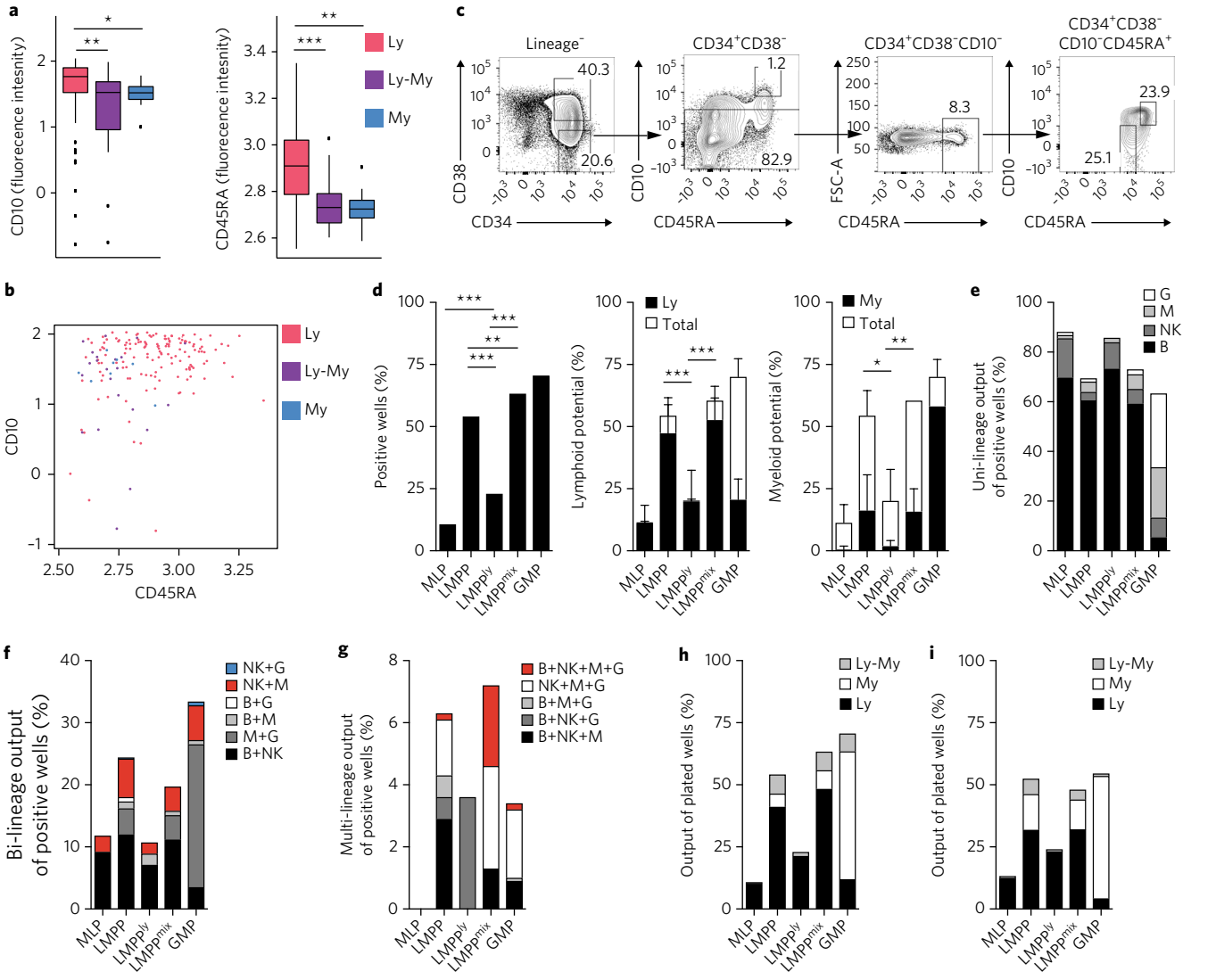


Fig. 6 | New flow-cytometry sorting strategy to purify functional potential in the CB LMPP compartment. a, Expression of the surface markers CD10 and CD45RA on CB LMPPs, grouped by functional output as lymphoid (composed of uni-lymphoid output (B or NK) and bi-lymphoid output (B and NK)), myeloid (composed of uni-myeloid output (M or G) and bi-myeloid output (M and G)) or lympho-myeloid (key); expression values calculated with the 'flowCore::logicleTransform' function in R software. CD10, $P = 0.0052$, lymphoid versus lympho-myeloid, and $P = 0.027$, lymphoid versus myeloid; CD45RA, $P = 4.8 \times 10^{-6}$, lymphoid versus lympho-myeloid, and $P = 0.0027$, lymphoid versus myeloid (Wilcoxon rank sum test). **b**, Expression of CD10 and CD45RA on CB LMPPs, measured by flow cytometry (calculated as in **a**); colors indicate output from functional assays (key). **c**, Revised sorting strategy for CB cells based on expression of CD10 and CD45RA defined by bioinformatics analyses. Numbers adjacent to outlined areas indicate percent cells from the parental gate (above plots). **d**, Total cloning efficiency of single CB MLPs, LMPPs, LMPP^{ly} cells (56 of 244 cells), LMPP^{mix} cells (152 of 240 cells) and GMPs in SGF15/2 culture (left), and cloning efficiency of lymphoid lineages (middle) and myeloid lineages (right) (key; presented as in Fig. 2a). **e**–**g**, Uni-lineage output **e**, bi-lineage output **f** and multi-lineage output **g** of single CB cells as in **d** (horizontal axis) in SGF15/2 culture, presented as the frequency of output cells (key) from positive wells (percent values). **h**,**i**, Lymphoid, myeloid and lympho-myeloid output (key) of CB cells as in **d** (horizontal axis) in SGF15/2 culture **h** or SF7b culture **i**, presented as the frequency among all plated cells. *$P = 0.01$–0.05, **$P = 0.001$–0.01 and ***$P < 0.001$ (Wilcoxon rank-sum test **a**, Fisher's exact test (**d**, left) or Student's *t*-test (**d**, middle and right)). Data are from one experiment with two donors (**a**,**b**; presented as median (middle line), lower and upper quartile (box top and bottom), minimum and maximum without outliers (bars extending above and below), and outliers (symbols above and below)), two experiments with one donor representative of six donors (**c**), two experiments (**d**, middle and right; mean + s.d. of six donors), two experiments with six donors ($n = 22$ donors for LMPPs, MLPs and GMPs; same donors in Fig. 2a–e) (**d**, left, and **e**–**h**) or two experiments with six donors ($n = 9$ donors for LMPPs, MLPs and GMPs, including $n = 3$ donors in Fig. 2f–j) (**i**).

CD38^{int} cells (693 cells) with that of conventionally purified LMPPs (1,136 cells) and GMPs (1,622) using SGF15/2 culture. Whereas the CD38^{hi} GMPs and LMPPs had a similar cloning efficiency of ~55%, the GMPs and CD38^{int} cells had a slightly higher cloning efficiency of ~70% (Fig. 7d). All four populations produced principally uni-lineage output (63–72%) (Fig. 7e). CD38^{hi} GMPs had much lower lymphoid potential ($P < 0.0001$ (Fisher's exact test)) and lympho-myeloid potential ($P = 0.0115$ (Fisher's exact test)) than that of conventionally purified GMPs (Fig. 7e–h and Supplementary Table 15), indicative of a functionally purer population. In summary, among CD10^{-}CD45RA^{+} LMPPs, the phenotype CD10^{hi}CD45RA^{hi} marked progenitors with lymphoid-only potential, and among GMPs, the phenotype CD38^{hi} marked progenitors with myeloid-only potential. Collectively, our single-cell analyses suggested that the progenitor

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

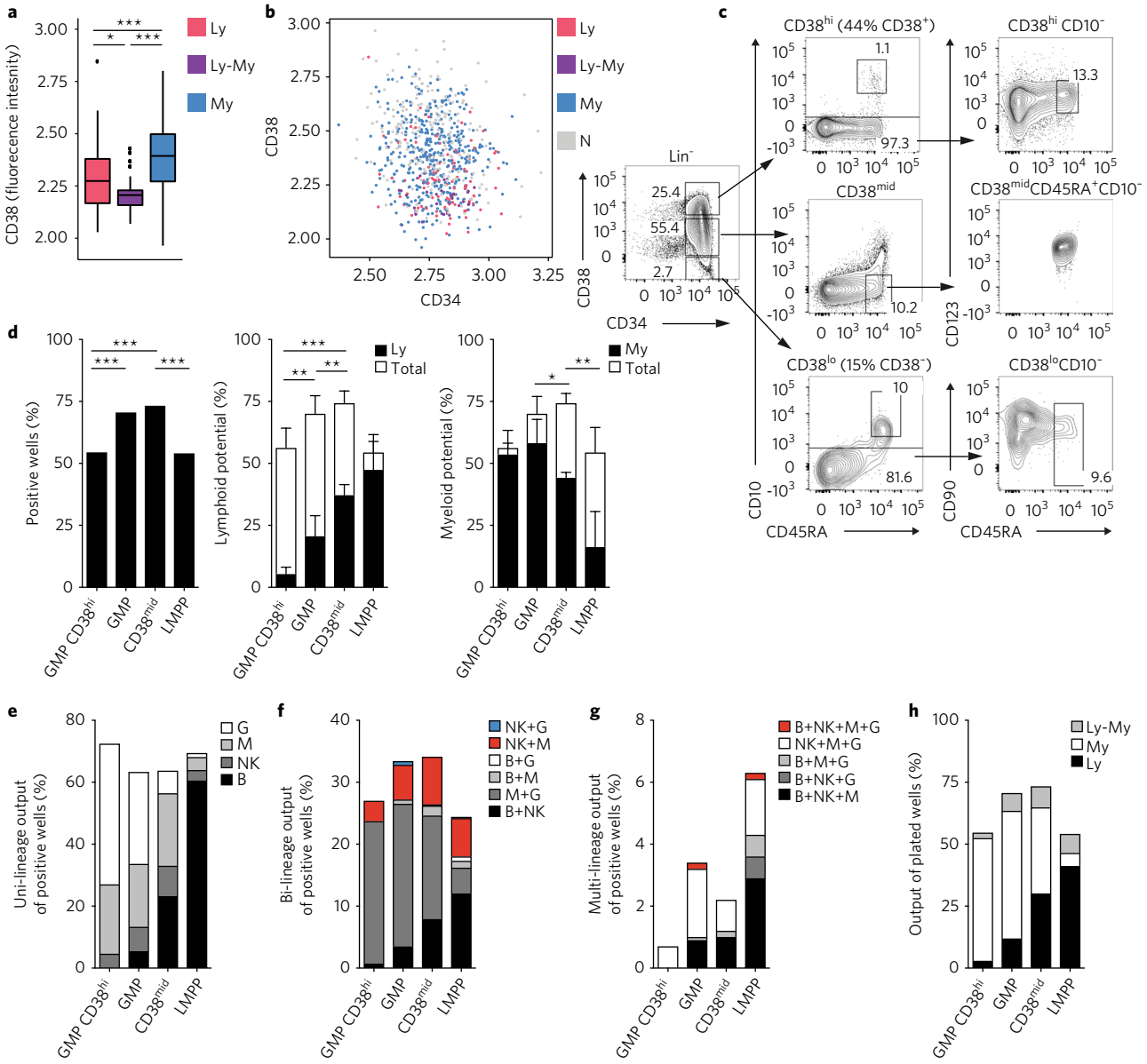


Fig. 7 | New flow-cytometry sorting strategy to purify functional potential in the CB GMP compartment. a, Expression of the surface marker CD38 on CB GMPs (calculated as in Fig. 6a), grouped by functional output (key). $P = 1.57 \times 10^{-11}$ (myeloid versus lymphoid) and $P = 1.6 \times 10^{-8}$ (myeloid versus lympho-myeloid) (Wilcoxon rank-sum test). **b**, Expression of CD38 and CD34 by CB GMPs; colors indicate output from functional assays (key). **c**, Revised sorting strategy for CB cells, based on CD38 expression defined by bioinformatics analysis. Numbers adjacent to outlined areas indicate percent cells in each. **d**, Total cloning efficiency of single CB CD38^{hi} GMPs (152 of 279 cells), CD38^{int} cells (508 of 693 cells) and LMPPs (horizontal axis) in SGF15/2 culture (left), and cloning efficiency of lymphoid lineages (middle) and myeloid lineages (right) of single cells as at left (key; presented as in Fig. 2a). **e**–**g**, Uni-lineage output **e**, bi-lineage output **f** and multi-lineage output **g** of single CB cells as in **d** (horizonal axis), presented as the frequency of output cells (key) in positive wells. **h**, Lymphoid, myeloid and lympho-myeloid output (key) of CB cells as in **d** (horizonal axis), presented as the frequency among all plated cells. $^*P = 0.01$–0.05, $^{**}P = 0.001$–0.01 and $^{***}P < 0.001$ (Wilcoxon rank sum test **a**, Fisher's exact test (**d**, left) or Students *t*-test (**d**, middle and right)). Data are from two experiments with five donors (**a**,**b**; **a** presented as in Fig. 6a), two experiments with one donor representative of four donors (**c**), two experiments (**d**, middle and right; mean + s.d. of four donors) or two experiments with four donors (or $n = 22$ donors for LMPP and GMP controls; same donors in Fig. 2a–e) (**d**, left, and **e**–**h**).

hierarchy downstream of stem cells might be more complex than previous models have suggested (Supplementary Fig. 7).

Discussion

Here we reported on the prospective separation and direct comparison of freshly isolated CB LMPPs, MLPs and GMPs. Our results showed that these lympho-myeloid progenitors were functionally and transcriptionally distinct and heterogeneous at the single-cell level. Although uni-lineage progenitors were most abundant, rarer multi-lineage lympho-myeloid progenitors were detected, most frequently among the LMPPs. Single-cell transcriptional analysis showed that LMPPs, MLPs and GMPs formed a transcriptional continuum. MLPs were positioned at one end of the lympho-myeloid spectrum, with GMPs at the other end, and they intersected with LMPPs, positioned in the middle. By combining functional and transcriptional analyses with flow-cytometry index data, we devised new flow-cytometry purification strategies for the isolation of myeloid CD38^{hi} GMPs and two sub-populations within the CD10$^-$

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

$CD45RA^+$ LMPP population: lymphoid $CD10^{hi}CD45RA^{hi}$ ($LMPP^{ly}$) cells and $CD10^{lo}CD45RA^{lo}$ ($LMPP^{mix}$) cells that retained most of the lympho-myeloid potential of LMPPs.

Several issues have prevented clear understanding of previously identified human lympho-myeloid progenitors. First, these progenitors have been isolated via cell-surface markers, on the basis of historical precedent, rather than via marker purification to functional homogeneity. Second, prospectively isolated lympho-myeloid progenitor populations have never previously been systematically compared, to our knowledge. Third, it is unclear if early progenitor populations downstream of HSCs include only uni-lineage cells[16,25] or also include bi- and multi-lineage progenitors in mice[5,14,17,18,26,39] or humans[21–24]. Fourth, functional assays demonstrate potential rather than actual cell fate in vivo under steady-state conditions. Finally, failure to register functional potential might reflect the inadequacy of an assay rather than the true potential, or indeed fate, of the cell in vivo. Thus, there is uncertainty about how distinct the differently identified progenitors are, and, if they are distinct, what their comparative functional potentials and transcriptional programs are at a clonal level.

Our exhaustive analysis of 4,598 single LMPPs, MLPs and GMPs, as well as populations of these progenitors, showed that they were functionally different in vitro and in vivo when transplanted in mice with humanized ossicles. The novel humanized ossicle model allowed ~10- to 100-fold more human-cell output than reported previously[21,23,25]. The GMP was mainly a progenitor of myeloid cells, with residual B cell and NK cell potential. Residual lymphoid potential was almost completely eliminated in the 44% of GMP cells with the highest CD38 expression (GMP $CD38^{hi}$). The MLP was mainly a progenitor of B cells, NK cells and T cells, with residual monocyte output. The LMPP had lymphoid and myeloid potential. Our new flow-cytometry purification scheme divided the LMPPs into two populations on the basis of expression of CD10 and CD45RA: one was almost entirely lymphoid ($LMPP^{ly}$); the other captured most of the myeloid and lympho-myeloid potential ($LMPP^{mix}$). Interestingly, the LMPPs produced mainly myeloid cells in vivo. Humanized ossicles might be particularly efficient at promoting human myelopoiesis, unlike the immunodeficient NSG mouse strain, which better supports lymphopoiesis.

We detected 24 different lineage-associated potentials in lympho-myeloid progenitors, which probably represents an underestimation, as we did not assess eosinophil, mast-cell, basophil or DC function. Although the majority of progenitors were uni-lineage, bi-lineage and multi-lineage output was seen (in up to 39% and 13%, respectively, of cells in vitro). Lympho-myeloid lineage 'decisions' can occur at multiple levels in HSC populations[1–3] and MPP populations [5,14] and in presumably more-mature LMPP populations[17,18,26], MLP populations and GMP populations[39]. Within the LMPP and GMP populations, true lympho-myeloid progenitors were rare (up to 10–14% of cells) and were concentrated among LMPPs. Notably, no experiments so far, to our knowledge, have directly assessed the hierarchical relationships among lineage-biased HSCs, MPPs and lympho-myeloid progenitors. Quantitative differences between fetal liver and BM have been observed in the multi-lineage output versus uni-lineage output of the broad $CD34^+CD38^+$ and $CD34^+CD38^-$ populations[16]. All our data were obtained with CB, and experiments similar to those reported here will be needed to determine the proportion of uni-lineage, bi- lineage and multi-lineage progenitors in BM.

One separate question is whether diverse lineage-associated progenitors identified in vitro have stably different functions or whether there is plasticity, such that functional output might be stochastically determined or variably 'instructed'. Further single-cell functional analysis of potentially functionally purer populations will be needed, with detailed fate mapping, in mice.

The rarity of LMPPs and MLPs (2 in 1×10^5 BM MNCs and 1 in 1×10^4 CB MNCs, respectively) and the minor proportion of multi- and bi-lineage progenitors among the LMPPs prompted us to study large numbers of single cells to obtain robust information on rare bi- and multi-lineage potentials. The rarity of LMPPs is also noteworthy for two reasons. First, single-cell RNA-sequencing analysis of unfractionated MNCs[39] will require the sequencing of large numbers of cells to provide adequate representation of these rare progenitors. Second, in acute myeloid leukemia, leukemic stem cells are often arrested at an LMPP-like stage, where they can constitute up to 80% of MNCs[22]. Given that, we speculate that the small pool size of healthy LMPPs might be very tightly controlled to minimize oncogenic transformation. Additionally, understanding how healthy LMPPs differentiate might provide insight into novel differentiation therapies for LMPP leukemic stem cells in acute myeloid leukemia.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41593-017-0002-3.

Received: 17 January 2017; Accepted: 3 October 2017;
Published online: 21 November 2017

References

1. Dykstra, B. et al. Long-term propagation of distinct hematopoietic differentiation programs in vivo. *Cell Stem Cell* **1**, 218–229 (2007).
2. Challen, G. A., Boles, N. C., Chambers, S. M. & Goodell, M. A. Distinct hematopoietic stem cell subtypes are differentially regulated by TGF-beta1. *Cell Stem Cell* **6**, 265–278 (2010).
3. Benz, C. et al. Hematopoietic stem cell subtypes expand differentially during development and display distinct lymphopoietic programs. *Cell Stem Cell* **10**, 273–283 (2012).
4. Chen, J. Y. et al. Hoxb5 marks long-term haematopoietic stem cells and reveals a homogenous perivascular niche. *Nature* **530**, 223–227 (2016).
5. Oguro, H., Ding, L. & Morrison, S. J. SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. *Cell Stem Cell* **13**, 102–116 (2013).
6. Sanjuan-Pla, A. et al. Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy. *Nature* **502**, 232–236 (2013).
7. Yamamoto, R. et al. Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. *Cell* **154**, 1112–1126 (2013).
8. Sun, J. et al. Clonal dynamics of native haematopoiesis. *Nature* **514**, 322–327 (2014).
9. Busch, K. et al. Fundamental properties of unperturbed haematopoiesis from stem cells in vivo. *Nature* **518**, 542–546 (2015).
10. Sawai, C. M. et al. Hematopoietic stem cells are the major source of multilineage hematopoiesis in adult animals. *Immunity* **45**, 597–609 (2016).
11. Yu, V. W. et al. Epigenetic memory underlies cell-autonomous heterogeneous behavior of hematopoietic stem cells. *Cell* **167**, 1310–1322 (2016).
12. Wilson, A. et al. Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* **135**, 1118–1129 (2008).
13. Cabezas-Wallscheid, N. et al. Identification of regulatory networks in HSCs and their immediate progeny via integrated proteome, transcriptome, and DNA methylome analysis. *Cell Stem Cell* **15**, 507–522 (2014).
14. Pietras, E. M. et al. Functionally distinct subsets of lineage-biased multipotent progenitors control blood production in normal and regenerative conditions. *Cell Stem Cell* **17**, 35–46 (2015).
15. 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 (2007).
16. Notta, F. et al. Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* **351**, aab2116 (2016).
17. Adolfsson, J. et al. Identification of $Flt3^+$ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. *Cell* **121**, 295–306 (2005).
18. Månsson, R. et al. Molecular evidence for hierarchical transcriptional lineage priming in fetal and adult stem cells and multipotent progenitors. *Immunity* **26**, 407–419 (2007).

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

19. Guo, G. et al. Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire. *Cell Stem Cell* **13**, 492–505 (2013).
20. Naik, S. H. et al. Diverse and heritable lineage imprinting of early haematopoietic progenitors. *Nature* **496**, 229–232 (2013).
21. Doulatov, S. et al. Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development. *Nat. Immunol.* **11**, 585–593 (2010).
22. Goardon, N. et al. Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. *Cancer Cell* **19**, 138–152 (2011).
23. Kohn, L. A. et al. Lymphoid priming in human bone marrow begins before expression of CD10 with upregulation of L-selectin. *Nat. Immunol.* **13**, 963–971 (2012).
24. Görgens, A. et al. Multipotent hematopoietic progenitors divide asymmetrically to create progenitors of the lymphomyeloid and erythromyeloid lineages. *Stem Cell Rep.* **5**, 154–155 (2015).
25. Velten, L. et al. Human haematopoietic stem cell lineage commitment is a continuous process. *Nat. Cell Biol.* **19**, 271–281 (2017).
26. Perié, L., Duffy, K. R., Kok, L., de Boer, R. J. & Schumacher, T. N. The branching point in erythro-myeloid differentiation. *Cell* **163**, 1655–1662 (2015).
27. Berardi, A. C. et al. Individual $CD34^{+}CD38^{low}CD19^{-}CD10^{-}$ progenitor cells from human cord blood generate B lymphocytes and granulocytes. *Blood* **89**, 3554–3564 (1997).
28. Ichii, M. et al. The density of CD10 corresponds to commitment and progression in the human B lymphoid lineage. *PLoS One* **5**, e12954 (2010).
29. Farlik, M. et al. DNA methylation dynamics of human hematopoietic stem cell differentiation. *Cell Stem Cell* **19**, 808–822 (2016).
30. Lee, J. et al. Restricted dendritic cell and monocyte progenitors in human cord blood and bone marrow. *J. Exp. Med.* **212**, 385–399 (2015).
31. Pronk, C. J. et al. Elucidation of the phenotypic, functional, and molecular topography of a myeloerythroid progenitor cell hierarchy. *Cell Stem Cell* **1**, 428–442 (2007).
32. Galy, A., Travis, M., Cen, D., Chen, B. & Human, T. Human T, B, natural killer, and dendritic cells arise from a common bone marrow progenitor cell subset. *Immunity* **3**, 459–473 (1995).
33. Manz, M. G., Miyamoto, T., Akashi, K. & Weissman, I. L. Prospective isolation of human clonogenic common myeloid progenitors. *Proc. Natl. Acad. Sci. USA* **99**, 11872–11877 (2002).
34. Reinisch, A. et al. A humanized bone marrow ossicle xenotransplantation model enables improved engraftment of healthy and leukemic human hematopoietic cells. *Nat. Med.* **22**, 812–821 (2016).
35. Novershtern, N. et al. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* **144**, 296–309 (2011).
36. Wilson, N. K. et al. Combined single-cell functional and gene expression analysis resolves heterogeneity within stem cell populations. *Cell Stem Cell* **16**, 712–724 (2015).
37. Haghverdi, L., Buettner, F. & Theis, F. J. Diffusion maps for high-dimensional single-cell analysis of differentiation data. *Bioinformatics* **31**, 2989–2998 (2015).
38. Moignard, V. et al. Decoding the regulatory network of early blood development from single-cell gene expression measurements. *Nat. Biotechnol.* **33**, 269–276 (2015).
39. Paul, F. et al. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. *Cell* **163**, 1663–1677 (2015).

Acknowledgements

We thank J.C. Zuniga-Pflucker for OP9-DL4 cells used for initial experiments; and the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust (090532/Z/09/Z)) for generation of sequencing data. Supported by the MRC (MHU Award G1000729 and MRC Disease Team Award 4050189188 to P.V.; PhD studentship to Z.A. and F.H.; Single Cell Award MR/M00919X/1), CRUK (Program Grant C7893/A12796 to P.V.; Program Grant C1163/A21762 to B.G.; Development Fund CRUKDF0176-DK to D.K. and P.V.; and CRUK Development Fund C5255/A20758 to B.S. and P.V.), Bloodwise (Specialist Program 13001 and Project Grant 12019), the Oxford Partnership Comprehensive Biomedical Research Centre (NIHR BRC Funding scheme), the US National Institutes of Health (R01CA188055 and U01HL099999 to R.M.), the New York Stem Cell Foundation (Robertson Investigator, R.M.), the Leukemia and Lymphoma Society (Scholar Award to R.M.) and the Austrian Science Fund (FWF) (Erwin-Schroedinger Research fellowship to A.R.).

Author contributions

D.K. and B.S. contributed equally to this work and are listed alphabetically; F.H. and A.R. contributed equally to this work; D.K., B.S., Z.A. and P.V. designed the experiments; D.K., B.S., Z.A., A.R., M.S., L.Q. and N.G. performed experiments and analyzed data; F.H., G.O., Z.A., E.R. and S.T. performed bioinformatics and statistical analysis; J.D. and B.U. prepared samples; J.C., E.S., F.P., C.P., R.M. and B.G. provided reagents and materials; D.K., B.S. and P.V. wrote the paper; and all authors edited the manuscript.

Competing interests

The authors declare no competing financial interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41590-017-0001-2.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to P.V.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

Methods

Collection of samples from healthy donors and patients. BM or CB samples from healthy donors were obtained with informed consent (UK protocol MREC 06/Q1606/ or Administrative Panel on Human Subjects Research Institutional Review Board-approved protocols Stanford IRB no. 18329, no. 6453, and no. 5637). Fresh CB samples were purchased from NHS Cord Blood Bank, UK or from New York Blood Center. They were processed within 16–34 h after collection. Mononuclear cells were isolated and $CD34^+$ fraction was separated as described[40]. Fresh or frozen BM MNCs or $CD34^+$ fractions were used. Human BM stromal cell were obtained from samples according to Medical University of Graz Ethikkommission (Institutional Review Board-approved protocol, MUG Graz IRB no. 19-252). BM mesenchymal stromal cells (MSCs) were isolated and cell populations were expanded as described[34].

Sorting of HSPC populations by flow cytometry. Antibodies used for sorting by flow-cytometry and immunophenotyping are listed in Supplementary Table 16. CB or BM MNCs enriched for $CD34^+$ cells underwent depletion of lineage-marker-positive cells by staining with purified antibodies to human CD2, CD3, CD4, CD7, CD8a, CD11b, CD14, CD19, CD20, CD56 and CD235a, followed by Qdot 605 conjugated goat $F(ab')_2$ anti-mouse IgG (H + L). Cells were also stained for human markers with anti-CD38-FITC, anti-CD45RA-PE or anti-CD45RA-BV650, anti-CD123-PE Cy7, anti-CD90-biotin, anti-CD34-PerCP and anti-CD10-APC. Finally, cells were incubated with streptavidin-conjugated APC-eF780 and Hoechst 33258 (Invitrogen, Loughborough UK; final concentration: 1 µg/ml). For the humanized-ossicle-xenotransplantation assay, $CD34^+$ CB cells were stained with purified antibodies to the human lineage markers CD2, CD3, CD4, CD7, CD8a, CD11b, CD14, CD19, CD20, CD56, CD235a and anti-CD16. All antibodies to lineage markers were PE Cy5 conjugated. Cells were then stained with anti-CD38-PE Cy7, anti-CD90-FITC, anti-CD123-PE, anti-CD34-APC, anti-CD10-APC Cy7, anti-CD45RA-BV605 and propidium-iodide (Thermo Fisher, Waltham MA; final concentration: 1 µg/ml). Unstained, single-stained and fluorescence-minus-one (FMO) controls were used to determine background staining and compensation in each channel. Single-stained controls used the anti-mouse Ig, κ/Negative Control compensation particle set (BD, Oxford UK). CB cells were sorted with an average purity of 99% for in vitro and RNA assays and 96% for humanized ossicle xenotransplantation. Prior to single-cell sorting, single fluorescent beads were deposited directly into a 96-well plate to establish accuracy of single-cell deposition (>99%). Sorting was performed on a BD FACSAria III or BD Fusion, and flow-cytometry analysis was done on LSR Fortessa X20. Data analysis was performed using Diva v8.1 or FlowJo v10.0.06 and v10.0.07r2.

Index sorting for functional and transcriptional analyses. For index sorting, we saved information on the following parameters: FSC, SSC, Hoechst and expression of lineage markers, CD34, CD38, CD45RA, CD10, CD90 and CD123 for each single cell. For 919 index-sorted single cells, we assessed expression of 96 genes by qRT-PCR (Supplementary Fig. 5); 74 passed quality control. Separately, we performed single-cell index sorting and single-cell in vitro functional assays of 3,458 single cells (from Fig. 2, Supplementary Fig. 2, Fig. 6, Fig. 7 and Supplementary Fig. 6). In separate experiments, we index sorted 320 single cells for single-cell RNA-seq (Fig. 5). Using common 'position of the cells' in flow-cytometry plots, we could then map functional potential (i.e., lymphoid, myeloid or lympho-myeloid) to gene expression, cell-surface-marker expression and forward and side scatter. To purify $LMPP^{ly}$ and $LMPP^{mix}$ cells, the thresholds were defined based on maximum expression of CD10 and CD45RA by LMPPs with myeloid output. To purify $CD38^{hi}$ GMPs, thresholds were set using the maximum normalized CD38 expression of GMPs with lymphoid and lympho-myeloid output.

In vitro lympho-myeloid differentiation assay (bulk, single-cell, limiting-dilution assay). For population analysis, MS-5 cells were seeded on a 24-well plate coated with 0.1% gelatin at a density of 2×10^4 cells per well in α-MEM medium (Gibco/Thermo Fisher Scientific Loughborough UK) supplemented with 10% FBS (Hyclone, GE Healthcare, SH30070.03 Amersham Hatfield, UK), 1% penicillin-streptomycin, 1% L-glutamine, 0.1 µM DuP-697 (Cayman Chemical, Ann Arbor, Mi, USA), 20 ng/ml SCF, 10 ng/ml G-CSF, 10 ng/ml FLT3L, 10 ng/ml IL-15 and 10 ng/ml IL-2 (Peprotech London UK, SGF15/2 culture). 24 h after plating of MS-5 cells, 150 highly purified LMPPs, MLPs or GMPs were deposited into each well. Half of the medium was changed every week. Harvested cells were analyzed by flow cytometry at weeks 1, 2 and 3.

The limiting-dilution assay (LDA) was performed by sorting of LMPPs, MLPs or GMPs at different cell doses (1, 2, 5, 10 and 20 cells) from four different CB samples into 96-well plates pre-plated with 2,500 MS-5 cells per well with 100 µl of medium without cytokines. Immediately after sorting, 100 µl of 2 × SGF15/2 culture medium was added to each well. Half of the medium was changed every week. A total of 833 LMPPs, 789 MLPs and 1252 GMPs from four different CB samples were analyzed for the LDA at weeks 2–2.5 (Supplementary Table 2a). Frequency calculations were performed using L-Calc software (Stem Cell Technologies) and were independently verified by ELDA software (http://bioinf.wehi.edu.au/software/elda/). The LDA plots were generated using R, with lines representing the estimates calculated by ELDA software.

For single-cell analysis, single LMPPs, MLPs and GMPs were deposited into 96-well plates pre-plated with 2,500 MS-5 cells per well with 100 µl of medium without cytokines. Medium with 2 × cytokines was added to each well after sorting. Half of the medium was changed every week. After culture for 2–2.5 weeks, flow cytometry was performed, and wells with more than 15 human $CD15^+$, $CD14^+$, $CD56^+$ or $CD19^+$ cells were considered positive (details in Supplementary Table 2b). To compare with published conditions[21], single LMPPs, MLPs and GMPs were cultured for 4 weeks on MS-5 stroma in H5100 medium (StemCell Technologies Cambridge UK) supplemented with 100 ng/ml SCF, 20 ng/ml IL-7, 50 ng/ml thrombopoietin, 10 ng/ml IL-2, 20 ng/ml GM-CSF, 20 ng/ml G-CSF and 10 ng/ml M-CSF (all from Peprotech, S7T2GM/G/M culture) and were analyzed by flow cytometry.

To read lineage readouts for all in vitro lympho-myeloid differentiation assays, harvested cells were stained with anti-human CD15-FITC, anti-human CD14-PE, anti-human CD19-PE Cy7, anti-human CD56-APC or -PE Cy5, anti-human CD45-APC Cy7 and, in some cases, with anti-human CD34-BV605 (Supplementary Table 16).

***In vitro* T cell–differentiation assay.** OP9-hDL1 cells[41] were seeded on a 24-well plate coated with 0.1% gelatin at a density of 2×10^4 cells per well in freshly prepared α-MEM medium (Gibco/Thermo Fisher Scientific, 12000-063) with 20% heat-inactivated FBS (Hyclone, GE Healthcare, SH30070.03 Amersham Hatfield, UK), 1% penicillin-streptomycin, 1% L-glutamine. 24 h after plating of OP9-hDL1 cells, 150 highly purified LMPPs, MLPs or GMPs were deposited into each well and were cultured in the presence of 10 ng/ml SCF, 5 ng/ml FLT3L and 5 ng/ml IL-7 (Peprotech, London, UK, SF7a culture). Cells were dissociated from wells and were transferred to new plates with fresh OP9-hDL1 cells weekly. Harvested cells were analyzed by flow cytometry at weeks 4, 5 and 7. Cells were stained with anti-human CD7-FITC, anti-human CD1a-PE, anti-human CD8-PE Cy7, anti-human CD4-APC and anti-human CD45-APC Cy7 (Supplementary Table 16).

***In vitro* combined T cell–lympho-myeloid differentiation assay.** MS5-$hDL1^{IND}100$ cells[42] (in which hDL1 expression could be induced by the addition of doxycycline) were seeded on 96-well plates coated with 0.1% gelatin at a density of 2,500 cells per well in 100 µl freshly prepared α-MEM medium (Gibco/Thermo Fisher Scientific, Loughborough UK) supplemented with 20% FBS (Hyclone, GE Healthcare, SH30070.03HI, Amersham Hatfield, UK), 1% penicillin-streptomycin and 1% L-glutamine. 24 h after plating of MS5-$hDL1^{IND}$ cells, single LMPPs, MLPs or GMPs were deposited into each well and were cultured in the presence of 20 nM insulin (Sigma-Aldrich, St Louis, MO), 50 ng/ml SCF, 20 ng/ml FLT3L and 10 ng/ml IL-7 (Peprotech London UK, SF7b culture). Fresh medium was added every week.

Cells were harvested at 21 d and were split into two groups: half were used for flow cytometry, and the remaining half were re-seeded onto MS5-$hDL1^{IND}100$ cells and cultured in SF7b/Dox culture with doxycycline (1 µg/ml). Half of the medium was changed twice every week. Fresh doxycycline was added to the cultures three times a week. At 42 d, cells were harvested and flow cytometry was performed. At 21 d, wells with more than 8 human $CD15^+$, $CD14^+$, $CD56^+$ or $CD19^+$ cells were considered positive. At 42 d, flow cytometry with anti-human CD1a, anti-human CD7, anti-human CD4 and anti-human CD8 was performed, and wells with more than 8 $CD7^+$ cells were considered positive for T cell output.

CFU assay. Colony formation was tested as before[22]. Colony identity was confirmed morphologically after cytospin (medium acceleration, 800 r.p.m., 5 min, May–Grunwald–Giemsa stain; Sigma, Poole UK) and by flow cytometry with anti-human CD15-FITC, anti-human CD14-PE and anti-human CD235a-PE Cy5 (Supplementary Table 16).

Humanized-ossicle-xenotransplantation assay. Protocol was performed as previously described[34]. In brief, in vitro–expanded human BM-MSC populations were harvested, resuspended in 60 µl of pooled human platelet lysate (pHPL) and admixed with 240 µl of matrigel-equivalent matrix. The whole matrix-cell mixture was injected subcutaneously to generate humanized ossicle niches. 8–10 weeks after application of BM-MSCs, transplants were evaluated for bone and marrow formation. Mice with established humanized ossicle niches were conditioned with 200 rads of irradiation 12–24 h before transplantation. Variable numbers of LMPPs, MLPs and GMPs from at least three different CB donors (Supplementary Fig. 3c) were transplanted in total volume of 20 µl by direct intraossicle injection. Experiments were performed in accordance with a protocol approved by Stanford's Administrative Panel on Laboratory Animal Care (no. 22264) and in adherence to the US National Institutes of Health's Guide for the Care and Use of Laboratory Animals. Normal multi-lineage engraftment was assessed 1–2 weeks after transplantation and was defined by the presence of myeloid cells ($CD33^+$) and B cells ($CD19^+$) among engrafted human $CD45^+HLA\text{-}ABC^+$ cells. Engrafted mice were stained with anti-human CD14-PE or anti-human CD14-APC Cy7, anti-human CD15-FITC, anti-human HLA-ABC-FITC or anti-human HLA-ABC-PB, anti-human CD19-APC, anti-human CD33-PE and anti-human CD45-V450 (Supplementary Table 16).

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

RNA sequencing of bulk HSPC populations. 100 highly purified HSPCs from CB samples obtained from healthy donors were sorted directly into lysis buffer containing RNAse inhibitor (Clontech St Germain-en-Laye France) and were stored at −80°C before further processing. cDNA synthesis was done with Smarter Ultra low input RNA kit v1 (Clontech) as previously described[43]. Illumina libraries were generated using a Nextera XT DNA sample preparation kit and Index Kit (Illumina Chesterford UK). Library size and quality were checked using Agilent High-Sensitivity DNA chip with Agilent Bioanalyser (Agilent Technologies Stockport UK). The concentration of indexed libraries was determined using a Qubit High-Sensitivity DNA kit (Invitrogen Loughborough, UK). Libraries were pooled to a final concentration of 5–14 nM and were sequenced on an Illumina HiSeq 2000 single-end 50-bp reads.

Bulk and single-cell gene-expression analysis by Dynamic Arrays. Gene-expression analysis was performed as described[40]. TaqMan assays (Applied Biosystems) are listed in Supplementary Tables 2 and 8.

Single-cell RNA sequencing. Single-cell libraries for RNA sequencing were prepared using the Smart-seq2 protocol[44], whereby 23 cycles were used for the cDNA library preamplification. Illumina Nextera XT DNA sample preparation kit and Index Kit (Illumina Chesterford UK) was used for cDNA tagmentation and indexing. ERCC RNA Spike-In Mix (Ambion) was added to the lysis mix at a final dilution of 1:80,000,000. Library size, quality and concentration were checked as done for the bulk RNA sequencing. Libraries were pooled to a final concentration of 7–28 nM, and 78–95 single-cell libraries were combined per pool. Sequencing was done on HiSeq4000 using 75-bp paired-end reads. Each pool contained a library generated from an empty well.

Bioinformatics analysis (bulk RNA-seq, single-cell Biomark and single-cell RNA-seq). For 50-bp single-end bulk RNA sequencing, alignment to the hg38 reference genome was carried out using TopHat v2.0.10 (ref. [45]). Alignments were processed using Picard tools (http://picard.sourceforge.net/). We used R version 3.1.1 http://www.R-project.org. Sequencing reads were filtered for mapq 4 i.e. uniquely mapping reads. This gave a range of 15.1×10^6 to 56.2×10^6 aligned reads. The total number of genes expressed per sample was calculated as an rpkm > 1. The number of expressed genes ranged from 7,707 to 11,350, with an average of 9,800. The count matrix was transformed to $\log_2$(cpm) scale, and principal-component analysis was carried out. An ANOVA-like test was performed, using edgeR package for R, to identify genes expressed differentially by the populations. One CB biological replicate MPP population was excluded because its global gene expression showed more uniquely expressed genes than in the three remaining MPP biological replicates and showed low correlation to the other three replicates. The genes were ranked by their significance (*P* value adjusted for multiple testing), and different numbers of genes were used for PCA and hierarchical clustering of samples. Eigenvalues from PCA were calculated by using the square of the s.d. of the principle components. Differential gene expression for one versus one and one versus all comparisons were calculated using edgeR. For gene-signature generation, a cut-off of a change in expression of over onefold (log values) was used, and genes were ranked by *P* value. Heat maps and associated hierarchical clustering were generated using GENE-E software (Broad Institute) or using the R packages pvclust and heatmap.2 (gplots). MetaCore Pathway Map enrichment analysis (Thomson Reuters, London UK) was carried out on genes expressed differentially by each lympho-myeloid population versus all other populations (one versus all). A *P*-value cut-off of 0.05 was used to identify positively enriched pathway maps.

Analysis of single-cell Biomark data was performed in R version 3.3.1 using data exported from the Fluidigm Data Collection software. For quality control, amplification curves with a Quality Score of <0.65, and any Ct values of >27 were treated as undetected expression. Any cells in which expression of both housekeeping genes (*B2M* and *GAPDH*) was not detected were removed from further analysis ($n=7$ cells). An additional cell was removed as it had a large outlying number of genes detected. The housekeeping gene *ACTB* was also measured in the assay, but unlike *B2M* and *GAPDH*, it did not show robust expression across the majority of cells and therefore was not used in further analysis. Normalized ΔCt values were calculated by subtraction of the mean of Ct values for *B2M* and *GAPDH* in each cell, as previously described[19]. Housekeeping genes were excluded from further analysis. Genes detected in <20 cells, with a variance of <1 across all cells or expressed in none of the MLP, GMP or LMPP ten-cell control samples assayed by qRT-PCR alongside single-cell samples were removed from downstream analysis. Post-quality-control data measured 74 genes in 919 single cells.

Hierarchical clustering was performed on genes and cells by using the hclust function (stats package) with distance measure 1 – Spearman's correlation and agglomeration method Ward.D2. The heat map visualizing the clustering was plotted using the heatmap.2 function (gplots package). Cells were divided into three clusters using the cutree function (stats package) on the hierarchical clustering. A gene was classed as expressed differentially by two clusters if it satisfied two criteria: 1) the magnitude of the change ($\log_2$ fold value) of the mean ΔCt in each cluster was >1, and 2) the adjusted *P* value (Benjamini & Hochberg correction for multiple testing) of two-sided Wilcox test of ΔCt expression values between the two clusters was <0.01. Diffusion maps[46] were used for dimensionality reduction of the single-cell gene-expression data. This method was implemented using the DiffusionMap function from the destiny R package with Euclidean distance[37,47].

Single-cell RNA-sequencing reads were aligned using G-SNAP[48], and mapped reads were assigned to Ensembl genes (release 81; ref. [49]) by using HTSeq[50]. Cells with fewer than 500,000 reads mapping to nuclear genes, greater than 20% of mapped reads mapping to mitochondrial genes or greater than 20% of mapped reads mapping to External RNA Controls Consortium (ERCC) spike-ins or with expression of fewer than 750 different genes with at least ten counts were removed from further analysis. ERCC spike-in controls identified genes exceeding technical variance[51]. From donors 1 and 2, 163 of 166 cells and 157 of 249 cells, respectively, passed quality control. Single-cell profiles were normalized using the scran R package[52], and variable genes were identified as having variation exceeding technical levels[51]. Data showed batch effects between different donors. The Seurat R package (https://github.com/satijalab/seurat) was then used to regress out plate effects from the sequencing data and set more stringent thresholds for variable genes, leading to 1,605 variable genes in donor 1 and 1,273 variable genes in donor 2. Principal-component analysis was performed using Seurat, and clusters were found using the Seurat::FindClusters function on the first ten principal components. Heat maps display the top genes marking these clusters as identified by the Seurat::FindAllMarkers function and were visualized using the gplots::heatmap.2 function.

Statistical analysis. Frequency of populations in flow-cytometry plots gates is the mean of the population across all samples analyzed as indicated. Bar graphs of gene-expression analysis represent mean + s.e.m. or mean + s.d., as indicated in figure legends. Two-tailed Student's unpaired *t*-test and Fisher's exact test (Excel, GraphPad sortware) were used to determine statistical significance in gene-expression-analysis data and single-cell functional assays, respectively. The statistical significance of the *P* value was defined as follows for all *P* value comparisons made: $P>0.05$. not significant; $P=0.01$–0.05, significant (*), $P=0.001$–0.01, very significant (**); and $P<0.001$, extremely significant (***). Wilcoxon rank sum test was done using R. Kruskal-Wallis test, stratified by group, was used to define significant differences between LMPPs, MLPs and GMPs in the single-cell functional assay in SGF15/2 culture and gave the following *P* values: LMPPs, 5×10^{-6}; MLPs, 0.1725; and GMPs, 0.7395. Wilcoxon rank sum test confirmed that there were no outliers among single LMPPs from different CB donors. Prism software was used to plot the gene-expression analysis and single-cell in vitro data. LDA plots were generated using R, and the lines represent the estimates calculated using ELDA software.

Life Sciences Reporting Summary. Further information on experimental design and reagents is available in the Life Sciences Reporting Summary.

Data availability. Bulk RNA sequencing data have been deposited in Arrayexpress (https://www.ebi.ac.uk/arrayexpress/) with accession number E-MTAB-5456. The single-cell RNA-sequencing data accession number is GSE100618. All other source data that support the findings of this study are available from the corresponding author upon request.

References

40. Quek, L. et al. Genetically distinct leukemic stem cells in human CD34⁻ acute myeloid leukemia are arrested at a hemopoietic precursor-like stage. *J. Exp. Med.* **213**, 1513–1535 (2016).
41. Six, E. M. et al. Cytokines and culture medium have a major impact on human in vitro T-cell differentiation. *Blood Cells Mol. Dis.* **47**, 72–78 (2011).
42. Calvo, J., BenYoucef, A., Baijer, J., Rouyez, M. C. & Pflumio, F. Assessment of human multi-potent hematopoietic stem/progenitor cell potential using a single in vitro screening system. *PLoS One* **7**, e50495 (2012).
43. Woll, P. S. et al. Myelodysplastic syndromes are propagated by rare and distinct human cancer stem cells in vivo. *Cancer Cell* **25**, 794–808 (2014).
44. Picelli, S. et al. Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* **9**, 171–181 (2014).
45. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
46. Coifman, R. R. et al. Geometric diffusions as a tool for harmonic analysis and structure definition of data: diffusion maps. *Proc. Natl. Acad. Sci. USA* **102**, 7426–7431 (2005).
47. Angerer, P. et al. destiny: diffusion maps for large-scale single-cell data in R. *Bioinformatics* **32**, 1241–1243 (2016).
48. Wu, T. D. & Nacu, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26**, 873–881 (2010).
49. Yates, A. et al. Ensembl 2016. *Nucleic Acids Res.* **44** D1, D710–D716 (2016).
50. Anders, S., Pyl, P. T. & Huber, W. HTSeq–a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
51. Brennecke, P. et al. Accounting for technical noise in single-cell RNA-seq experiments. *Nat. Methods* **10**, 1093–1095 (2013).
52. Lun, A. T., Bach, K. & Marioni, J. C. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biol.* **17**, 75(2016).

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

nature research

Corresponding author(s): Paresh Vyas

☐ Initial submission ☐ Revised version ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

3 or more biological replicates are routinely used and accepted in the field (including publications in Nature Immunology journal) for analyses of normal primary human hemopoietic stem/progenitor cells.
The sample size selection for each figure was the following:
Figure 1: Sample sizes were: CFU assay, 5 biological replicates; in vitro liquid culture, 150 progenitor cells from 3 populations from 3 biological replicates; bulk T cell culture results, 5 biological replicates
Figure 2: Sample sizes were: LDA studies used 4 biological replicates; single cell functional analyses in theSGF15/2 condition 22 biological replicates; single cell functional studies in the SF7b condition 3 biological replicates
Figure 3: Sample sizes were: In vivo transplantation, LMPP were isolated from 11 biological replicates, MLP from 3 biological replicates, GMP from 6 biological replicates
Figure 4: Sample sizes were: RNA sequencing of normal stem/progenitor populations from 3-4 biological replicates
Figure 5: Sample sizes were: Single cell RNA seq gene expression analysis 2 biological replicates;
Figure 6: Sample sizes were: LMPP revised sorting strategy 6 biological replicates (SGF15/2 condition) and 6 biological replicates (SF7b condition)
Figure 7: Sample sizes were: GMP revised sorting strategy 4 biological replicates.

2. Data exclusions

Describe any data exclusions.

In Figure 1: For CFU assay 1 biological replicate from population CMP was excluded as the FACS-purified population had a low sorting purity (82.9%).
No data were excluded for Figure 2
In Figure 3: Injected FACS-sorted populations with lower than 90% purity were excluded from the engraftment analyses.
In Figure 4: RNA-Seq data from one biological replicate of an MPP population was excluded when compared to the 3 remaining MPP biological replicates as its global gene expression showed higher number of uniquely expressed genes and low correlation to the other three replicates.
In Figure 5: From donors 1 and 2, 163/166 and 157/249 cells passed quality control respectively (p.36)
In Figure 6: LMPP CD38lo (34 single cells) were excluded from further analysis due to the very low abundance of this population (as stated on page 15-16).
In Sup. Figure 5: 8 single cells were excluded according to our QC (as set out on page 36).

3. Replication

Describe whether the experimental findings were reliably reproduced.

All experiments were reliably reproduced and results are represented as mean +/- SEM or +/-SD as appropriate and is indicated. For the single cell functional assay in SGF15/2 condition, where 22 different CB donors were used, Kruskal-Wallis test, stratified by group was applied to define whether there were significant differences between LMPP, MLP and GMP cell outputs. It gave the following p-values: LMPP - 5x10^-6, MLP - 0.1725, GMP - 0.7395, suggesting that there might be an outlier among LMPPs from the different donors. To this end Wilcoxon rank

sum test was used and it confirmed that there was no outlier among single cell LMPPs coming from different CB donors (as stated on page XX in Materials and Methods Section).

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

No randomization method was used because no samples/organisms/participants were allocated into experimental groups.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

No blinding was used because no group allocation was done during data collection and/or analysis.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒ A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ A statement indicating how many times each experiment was replicated
☐	☒ The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒ A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒ The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
☐	☒ A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒ Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

No custom-made software was used for this study.
FACS data analysis was performed using Diva v8.1 or FlowJo v10.0.06 and v10.0.07r2. Frequency calculations for LDA analysis were performed using L-Calc software (Stem Cell Technologies) and independently verified by ELDA software (http://bioinf.wehi.edu.au/software/elda/).
For bulk RAN-seq alignments were processed using Picard tools (http://picard.sourceforge. net/). We used R version 3.1.1 http://www.R-project.org. An ANOVA-like test was performed, using edgeR package for R. Heatmaps and associated hierarchical clustering were generated using GENE-E software (Broad Institute) or using the R packages pvclust and heatmap.2 (gplots). MetaCore Pathway Map (Thomson Reuters, London UK) enrichment analysis was carried out on genes differentially expressed by each lympho-myeloid population versus all other populations.
Analysis of single cell Biomark data was performed in R version 3.3.1 using data exported from the Fluidigm Data Collection software. Hierarchical clustering was performed on genes and cells by using the hclust function (stats package) with distance measure 1 – Spearman's correlation and agglomeration method Ward.D2. The heatmap visualizing the clustering was plotted using the heatmap.2 function (gplots package). Cells were divided into three clusters using the cutree function (stats package) on the hierarchical clustering. Diffusion maps were used for dimensionality reduction of the single cell gene expression data, applying the DiffusionMap function from the destiny R package with Euclidean distance.
Single cell RNA sequencing reads were aligned using G-SNAP and mapped reads were assigned to Ensembl genes (release 81) by using HTSeq. Single cell profiles were normalized using the scran R package. The Seurat R package (https://github.com/satijalab/seurat) was used to regress the plate effects from the

sequencing data. Principal component analysis was performed using Seurat, and clusters found using the Seurat::FindClusters function on the first 10 principal components. Heatmaps display the top genes marking these clusters as identified by the Seurat::FindAllMarkers function and were visualized using the gplots::heatmap.2 function.
Two-tailed students unpaired t-test and Fisher's exact test (Excel, GraphPad sortware) were used to determine statistical significance in gene expression analysis data and single cell functional assays respectively. Prism software was used to plot the gene expression analysis and single cell in vitro data. LDA plots were generated using R and the lines represent the estimates calculated using ELDA software.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.	No unique materials were used.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).	All antibodies used are given in Supplementary Table 10, including clones, fluorochromes, companies and dilutions used. All antibodies were titrated and validated using positive and negative control samples.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.	MS5 cells were imported from DSZM, OP9-hDL1 was a gift from E. Six lab, MS5-DL1ind100 were a gift from F. Pflumio lab.
b. Describe the method of cell line authentication used.	We have validated DL1 expression in the OP9-hDL1 and MS5-DL1ind100 cell lines by FACS analysis.
c. Report whether the cell lines were tested for mycoplasma contamination.	We have not tested for mycoplasma contamination.
d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.	None of the cell lines used in this study are listed by ICLAC.

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.	6-12 weeks old NOD.Cg-Prkdcscid Il2rgtm1Wjl/SzJ (NSG from Jackson Laboratory,Bar Harbor, ME) mice (irrelevant of sex) were used in the study.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.	Human bone marrow was obtained from elderly individuals (age >60 years) after hip replacement surgeries. Covariate-relevant population characteristics not relevant for human cord blood samples.

nature research

Corresponding author(s): Paresh Vyas

☐ Initial submission ☐ Revised version ☒ Final submission

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

☒ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ 3. All plots are contour plots with outliers or pseudocolor plots.

☒ 4. A numerical value for number of cells or percentage (with statistics) is provided.

Methodological details

5. Describe the sample preparation.	Fresh CB samples were processed within 16-34h after collection. Mononuclear cells were isolated by ficoll density gradient. CD34+ cells were purified using a CD34 Microbead kit and magnetic-activated cell-sorting separation columns (Miltenyi Biotec). Fresh or frozen BM MNCs or CD34+ fractions were used. Cells were stained as described in Materials & Methods and then analysed.
6. Identify the instrument used for data collection.	Sorting was performed on BD Aria III or BD Fusion and FACS analysis was done on LSR Fortessa X20
7. Describe the software used to collect and analyze the flow cytometry data.	Data analysis was performed using Diva v8.1 or FlowJo v10.0.06 and v10.0.07r2
8. Describe the abundance of the relevant cell populations within post-sort fractions.	CB cells were sorted with average purity 99% for in vitro and RNA assays and 96% for humanized ossicle xenotransplantation. Prior to single-cell sorts, single fluorescent beads were deposited directly to a 96-well plate to establish accuracy of single cell deposition (>99%).
9. Describe the gating strategy used.	Unstained, single stained and Fluorescence Minus One (FMO) controls were used to determine background staining and compensation in each channel. Supplementary Fig. 1a shows full gating strategy. Gating was on live cells (FSC-A & Hoechst), doublet exclusion (FSC-A & FSC-W) and then on immunophenotypic markers as described.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☒

nature immunology

SUPPLEMENTARY INFORMATION

https://doi.org/10.1038/s41590-017-0001-2

In the format provided by the authors and unedited.

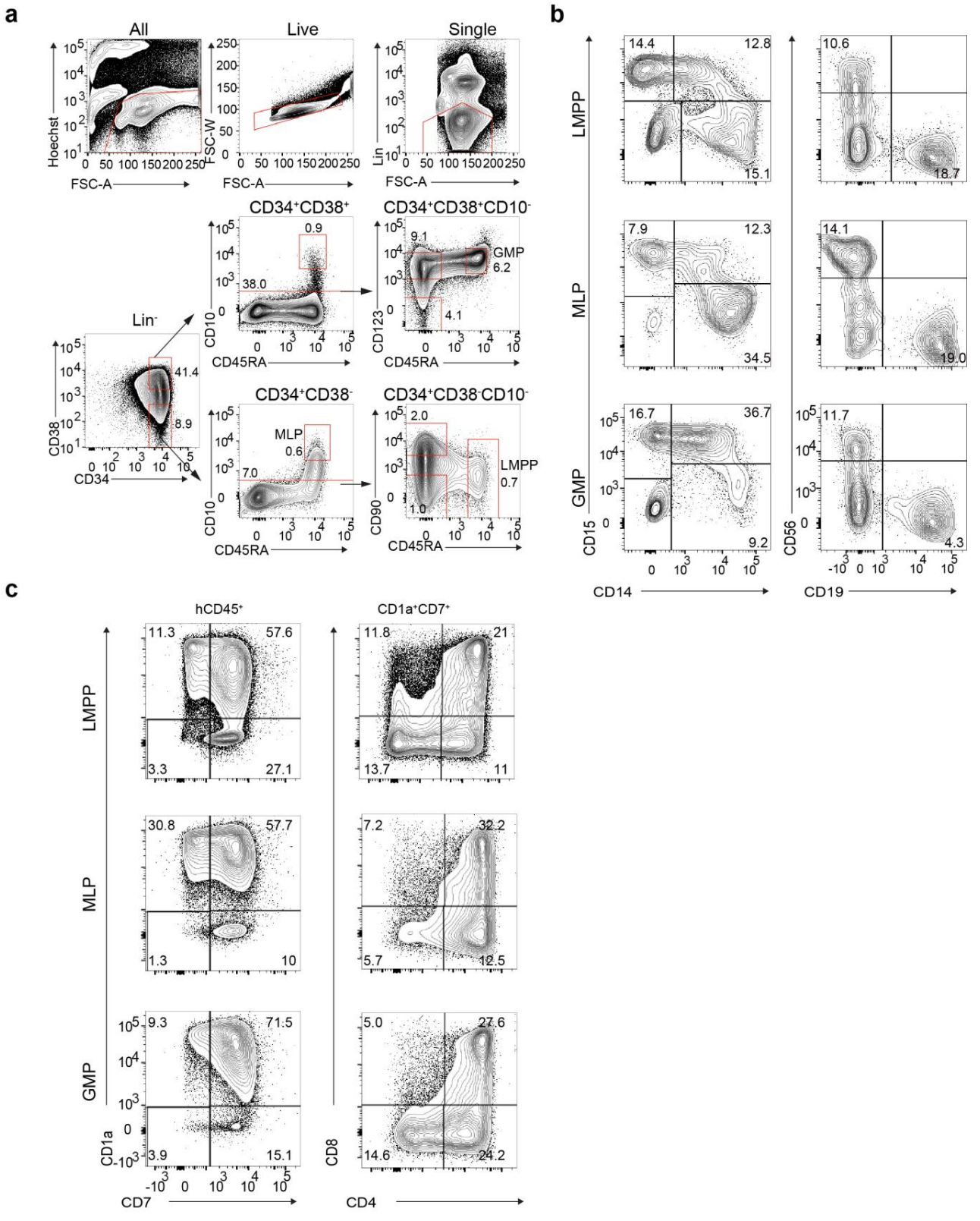
Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells

Dimitris Karamitros[1,2], Bilyana Stoilova[1,2], Zahra Aboukhalil[1], Fiona Hamey[3], Andreas Reinisch[4], Marina Samitsch[1], Lynn Quek[1,2,5], Georg Otto[1], Emmanouela Repapi[1], Jessica Doondeea[1,2], Batchimeg Usukhbayar[1,2], Julien Calvo[6], Stephen Taylor[1], Nicolas Goardon[1], Emmanuelle Six[7], Francoise Pflumio[6], Catherine Porcher[1], Ravindra Majeti[4], Berthold Göttgens[3] and Paresh Vyas[1,2,5]*

[1]MRC Molecular Haematology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, UK. [2]Oxford Biomedical Research Centre, University of Oxford, Oxford, UK. [3]Department of Haematology and Wellcome Trust–Medical Research Council Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK. [4]Division of Hematology, Stanford Institute for Stem Cell Biology and Regenerative Medicine, Stanford, California, USA. [5]Department of Hematology, OUH NHS Trust, Oxford, UK. [6]UMR967 INSERM/CEA/Université Paris 7/Université Paris 11, Paris, France. [7]UMR1163, Paris Descartes–Sorbonne Paris Cité University, Imagine Institute, Paris, France. Dimitris Karamitros and Bilyana Stoilova contributed equally to this work. *e-mail: paresh.vyas@imm.ox.ac.uk

© 2017 Nature America Inc., part of Springer Nature. All rights reserved.

Supplementary Figure 1

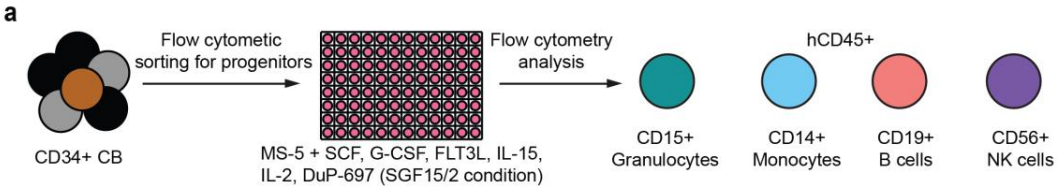


Supplementary Figure 1

Immunophenotypic separation of eight distinct human hemopoietic stem–progenitor cell (HSPC) populations and *in vitro* functional lympho-myeloid potential of LMPPs, MLPs and GMPs.

(a) Representative plot of flow sorting of human CB HSPCs. Population frequencies shown are the mean from 44 CB donors calculated as a percentage of Lin^-CD34^+ compartment. HSC, hemopoietic stem cell; MPP, multipotent progenitor; LMPP, lymphoid-primed multipotent progenitor; MLP, multi-lymphoid progenitor; CMP, common myeloid progenitor; GMP, granulocyte-monocyte progenitor; MEP, megakaryocyte-erythroid progenitor; B/NK, B-NK cell progenitor. **(b)** Representative flow cytometric analysis of the B cell, NK, monocytic and granulocytic output of CB LMPP, MLP and GMP cultured for 2 weeks on SGF15/2. Frequencies shown are mean from 3 CB donors calculated as a percentage of human $CD45^+$ cells. **(c)** Representative flow cytometric analysis of T cell output of CB LMPP, MLP and GMP cultured for 5 weeks on SF7a. Frequencies shown are average from 5 CB donors calculated as a percentage of human $CD45^+$ cells.

Supplementary Figure 2



b Culture condition 1: SGF15/2 LDA (**Supplementary Fig. 2c**) & single cell (**Fig. 2a-e, Fig. 6d-h, Fig. 7d-h**)

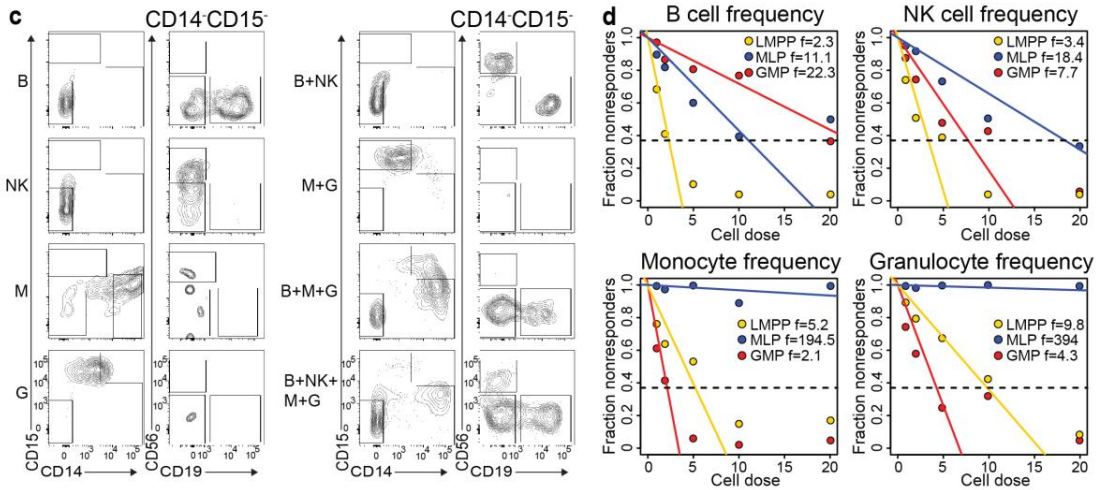
Cytokines	Time of analysis
20 ng/ml SCF 10 ng/ml G-CSF 10 ng/ml FLT3L 10 ng/ml IL-15 10 ng/ml IL-2 10^{-7}M Dup-697	2-2.5 weeks

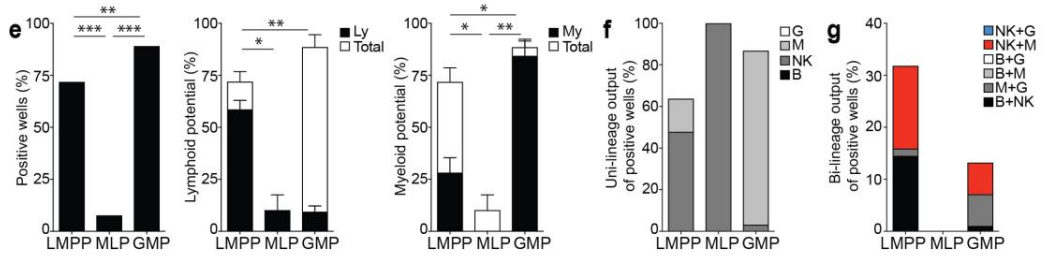
Culture condition 2: S7T2GM/G/M (ref 21) single cell (**Supplementary Fig. 2d-h**)

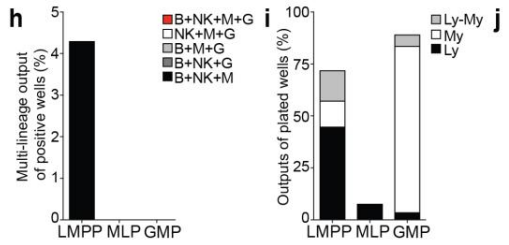
Cytokines	Time of analysis
100 ng/ml SCF 10 ng/ml IL-7 50 ng/ml TPO 10 ng/ml IL-2 20 ng/ml GM-CSF 20 ng/ml G-CSF 10 ng/ml M-CSF	4 weeks

Culture condition 3: SF7b/Dox (ref 42) single cell (**Fig. 2f-j, Fig. 6i, Supplementary Fig. 6c-f**)

Cytokines	Time of analysis
50 ng/ml SCF 20 ng/ml FLT3L 10 ng/ml IL-7 20 ng/ml Insulin 1 μg/ml Doxycyclin (at week 3)	B/NK/M/G: 3 weeks T-cell: 6 weeks







j

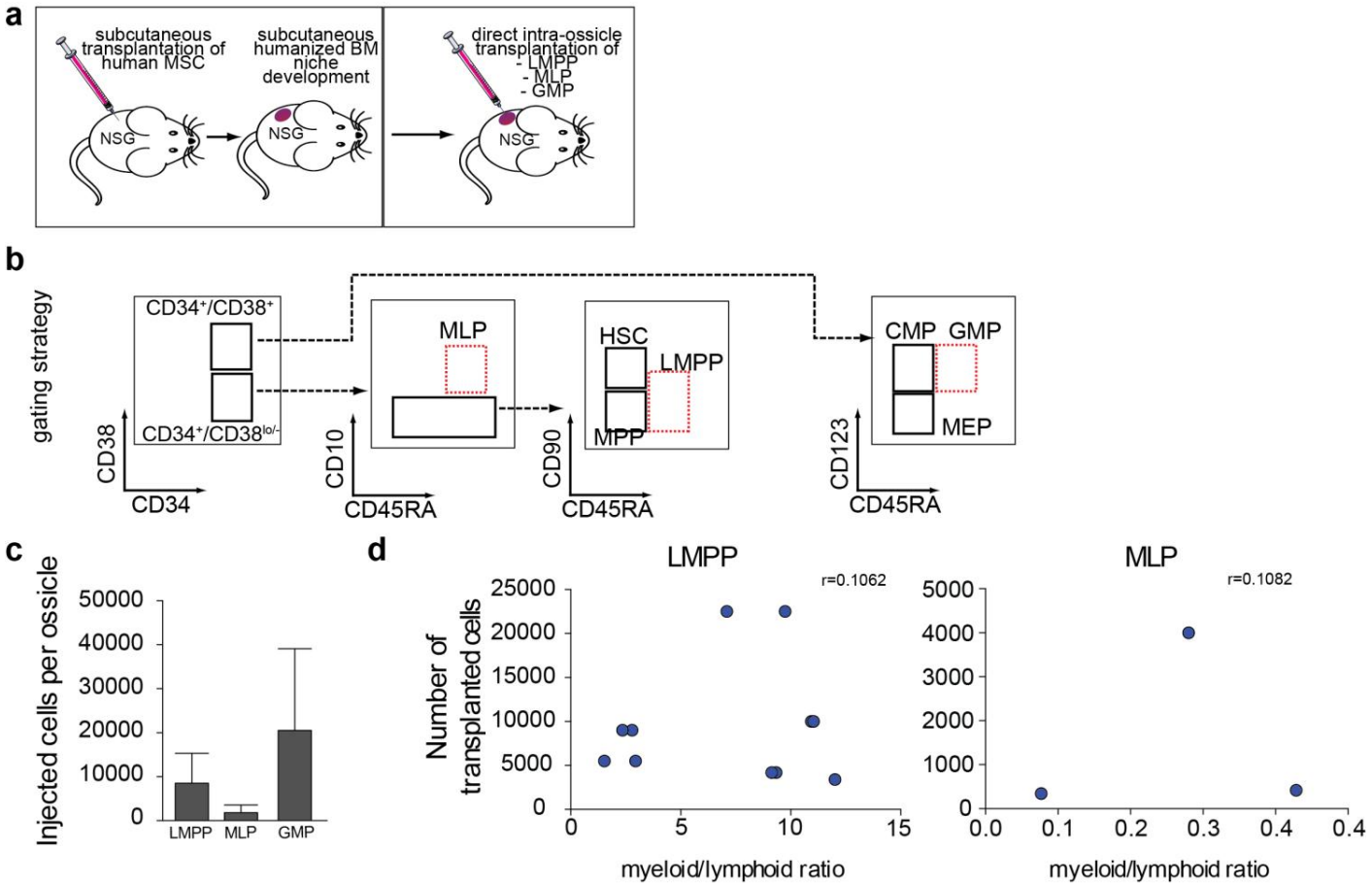
Potentials	LMPP	MLP	GMP
Lymphoid	B, NK, T, B+NK, NK+T, B+NK+T	B, NK, T, B+NK, B+NK+T	B, NK, B+NK, NK+T
Myeloid	M, G, M+G	M, G	M, G, M+G
Ly-My	B+M, B+G, G+T, NK+M, NK+G, B+NK+M, B+NK+G, B+M+G, NK+M+G, B+NK+M+G, B+NK+M+T, NK+M+G+T, B+M+G+T, B+NK+G+T, B+NK+M+G+T	NK+M	B+M, NK+M, NK+G, B+NK+M, B+M+G, NK+M+G, B+NK+M+G
Total no.	24	8	14

Supplementary Figure 2

LMPPs and GMPs are lympho-myeloid progenitors, while MLPs are mainly lymphoid progenitors in quantitative *in vitro* assays.

(a) Experimental strategy for sorting, culture and analysis for quantitative *in vitro* assays in SGF15/2. Similar strategy was used for the other culture conditions. **(b)** LDA and single cell *in vitro* culture conditions used in the study. **(c)** Representative flow cytometric profiles of the outputs from the quantitative *in vitro* assays. Plots are gated on human $CD45^+$ cells. **(d)** LDA plots showing the frequency (f: 1 in X cells can give rise to) of LMPP, MLP and GMP cells with B-cell, NK cell, monocytic and granulocytic potential. Plots are generated in R and the lines represent the estimates calculated using ELDA software. **(e)** Total cloning efficiency (left) of single LMPP, MLP and GMP in S7T2GM/G/M culture (LMPP: 69/96 cells, MLP: 4/52, GMP: 98/110). Significance defined using Fisher's exact test. Cloning efficiency of lymphoid (Ly, middle) and myeloid lineages (My, right). Bars indicate total cloning efficiency; filled portion indicates the proportion of lymphoid (lymphoid plus mixed clones) or myeloid potential (myeloid plus mixed clones). Mean ± SD is shown. Significance is defined using student's t-test. **(f)** Single-, **(g)** bi- and **(h)** multi-lineage single cell outputs in S7T2GM/G/M culture, presented as percentage of the positive wells. **(i)** Lymphoid (Ly), myeloid (My) and lympho-myeloid (Ly-My) outputs presented as percentage of all plated single cell LMPP, MLP and GMP wells in S7T2GM/G/M culture. **(j)** Summary of lineage outputs from single LMPP, MLP and GMP. For the LDA data were obtained from 4 CB donors, for the single cell assay in S7T2GM/G/M culture data were obtained from 2 CB donors.

Supplementary Figure 3

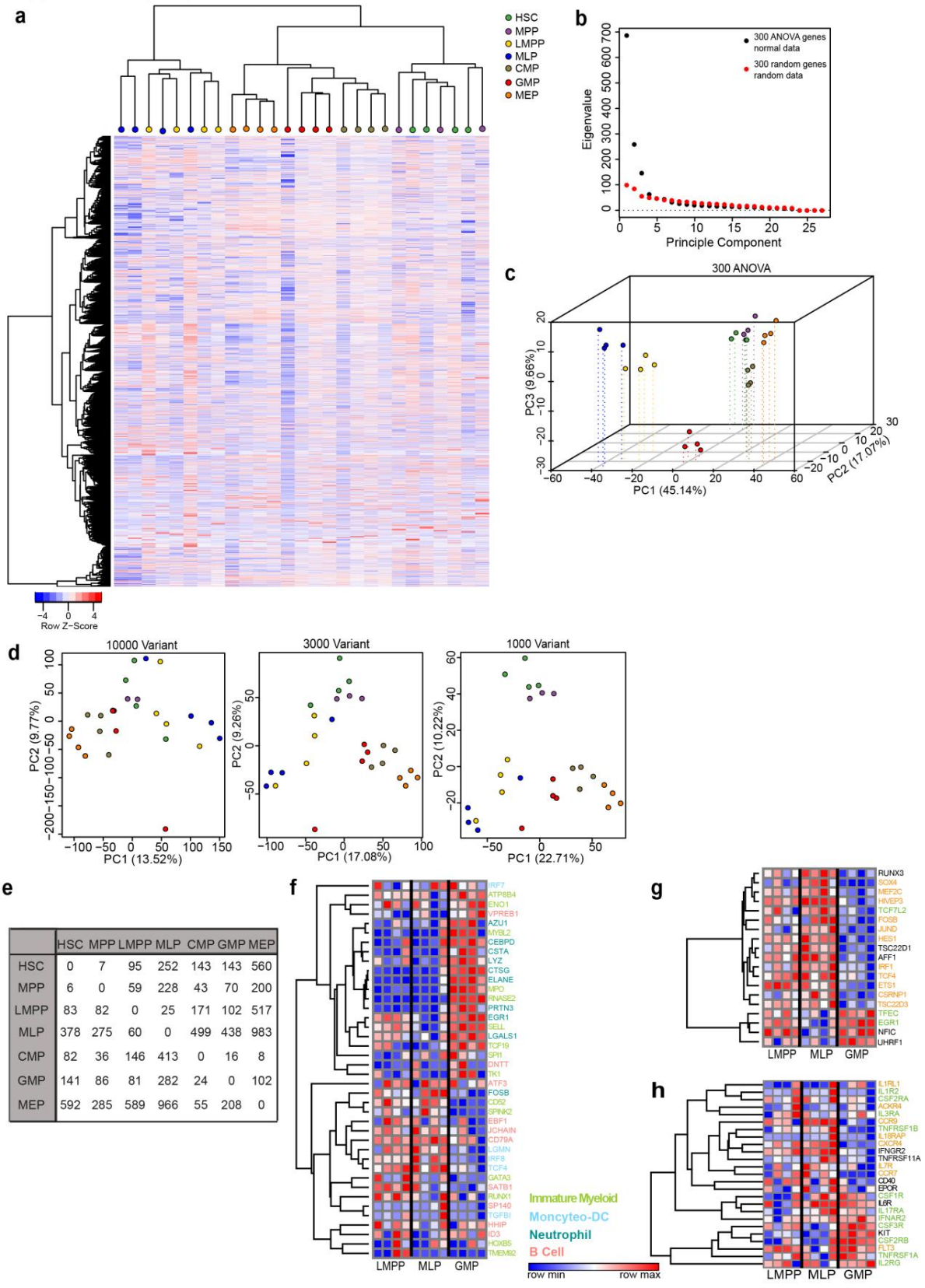


Supplementary Figure 3

Human LMPPs, MLPs and GMPs have distinct differentiation potential *in vivo*.

(a) Experimental strategy for generation of human ossicles in NSG mice and human progenitor transplantation. **(b)** Gating strategy for the sorting of human progenitors for *in vivo* transplantation. **(c)** Number of cells of LMPP, MLP and GMP populations injected per ossicle. **(d)** Correlation between the number of transplanted cells and the myeloid/lymphoid ratio of the output cells.

Supplementary Figure 4



e

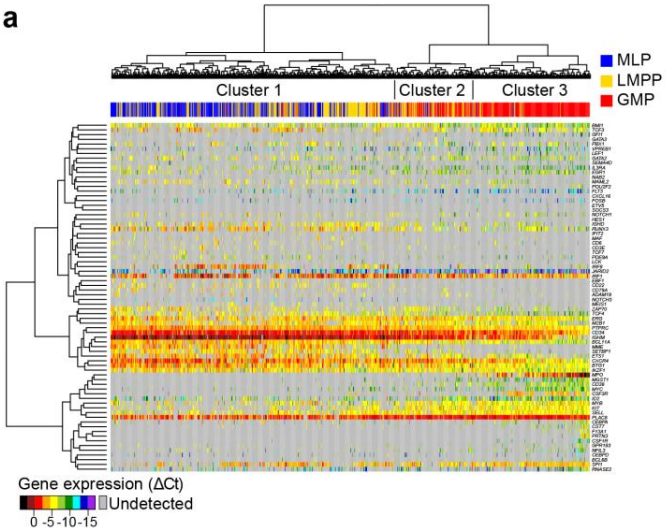
	HSC	MPP	LMPP	MLP	CMP	GMP	MEP
HSC	0	7	95	252	143	143	560
MPP	6	0	59	228	43	70	200
LMPP	83	82	0	25	171	102	517
MLP	378	275	60	0	499	438	983
CMP	82	36	146	413	0	16	8
GMP	141	86	81	282	24	0	102
MEP	592	285	589	966	55	208	0

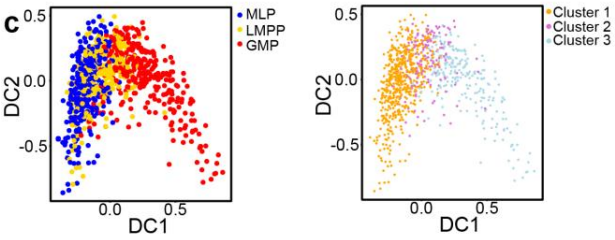
Supplementary Figure 4

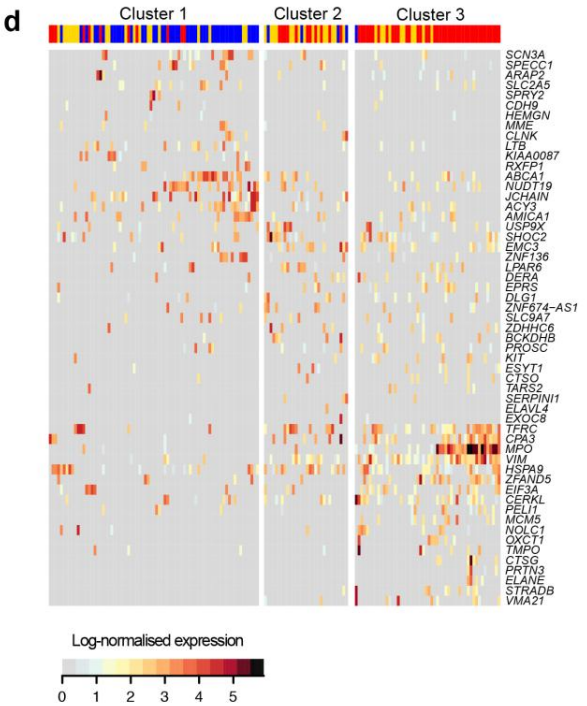
Transcriptional profiling of human HSPCs in bulk shows the distinct transcriptional patterns of human progenitor populations.

(a) Heatmap showing hierarchical clustering and the expression of all genes by HSPC populations. Expression values are normalized per gene. **(b)** Plot comparing the eigenvalues for each principle component (PC) for the PC analysis (PCA) of HSPC using the top 300 ANOVA genes (black) and 300 randomly selected genes from a randomized expression matrix (red). **(c)** 3D PCA plot showing HSPC populations using the top 300 ANOVA genes. Percentage variance for each PC is shown. **(d)** PCA plots showing HSPC populations using 10,000-1,000 genes with the highest variance across all HSPC populations. Percentage variance for each PC is shown. **(e)** Table of differentially expressed genes in one versus one comparisons of HSPCs. Genes are upregulated in a population column versus row. **(f)** Heatmap showing the expression of genes recently identified as being expressed by lineage primed hemopoietic cells[25]. Genes are color-coded according to their classification. Expression values are normalized per gene. **(g-h)** Heatmap showing the expression of **(g)** hemopoietic-related transcription factors that were differentially expressed between the MLP and GMP and **(h)** cytokine and chemokine related genes by the LMPP, GMP and MLP. Genes affiliated with the lymphoid or myeloid lineages are color-coded (lymphoid: orange, myeloid: green) and genes associated with immune function are labeled in black. Expression values are normalized per gene.

Supplementary Figure 5







b

Cell type	Cluster 1	Cluster 2	Cluster 3
GMP	25	80	203
LMPP	230	57	19
MLP	288	13	4

A	B	High A vs B	High B vs A
Cluster 1	Cluster 3	CD22, PDE9A, ADAM19, MAF, CD3E, TCF7, CD6, IRF8, SETBP1, CD79A, MME, NOTCH3, EBF1	MYC, CD38, CSF1R, NFIL3, CSF3R, ID2, POU2F2, BCL6B, CST7, MGST1, PRTN3, SOCS3, CEBPA, ETV5, GFI1, KIT, MPO, F13A1, GPR183, MYB
Cluster 1	Cluster 2	CD22, PDE9A, MAF, CD6, IRF8, MME, EBF1	CD38, BCL6B, MGST1, GATA3, ETV5, KIT, MPO
Cluster 2	Cluster 3	SETBP1, GATA3, MME	MYC, TCF3, CD38, CSF1R, CSF3R, CST7, MGST1, PRTN3, CEBPA, GFI1, MPO, F13A1, GPR183
		Lymphoid Genes	Myeloid Genes

Cell type	Ly	Ly-My	My
GMP	49	27	232
LMPP	229	46	30
MLP	285	10	10

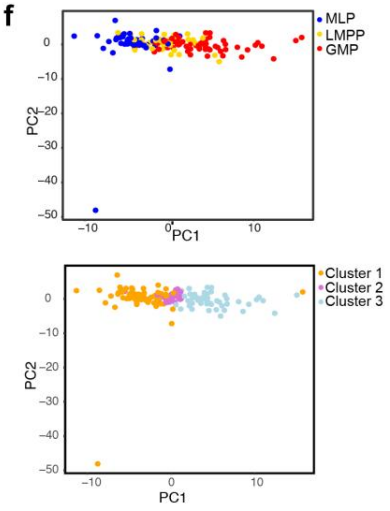
e

Donor 1:

Cell type	Cluster 1	Cluster 2	Cluster 3
GMP	1	23	32
LMPP	20	24	2
MLP	60	1	0

Donor 2:

Cell type	Cluster 1	Cluster 2	Cluster 3
GMP	11	11	41
LMPP	20	15	10
MLP	44	4	1

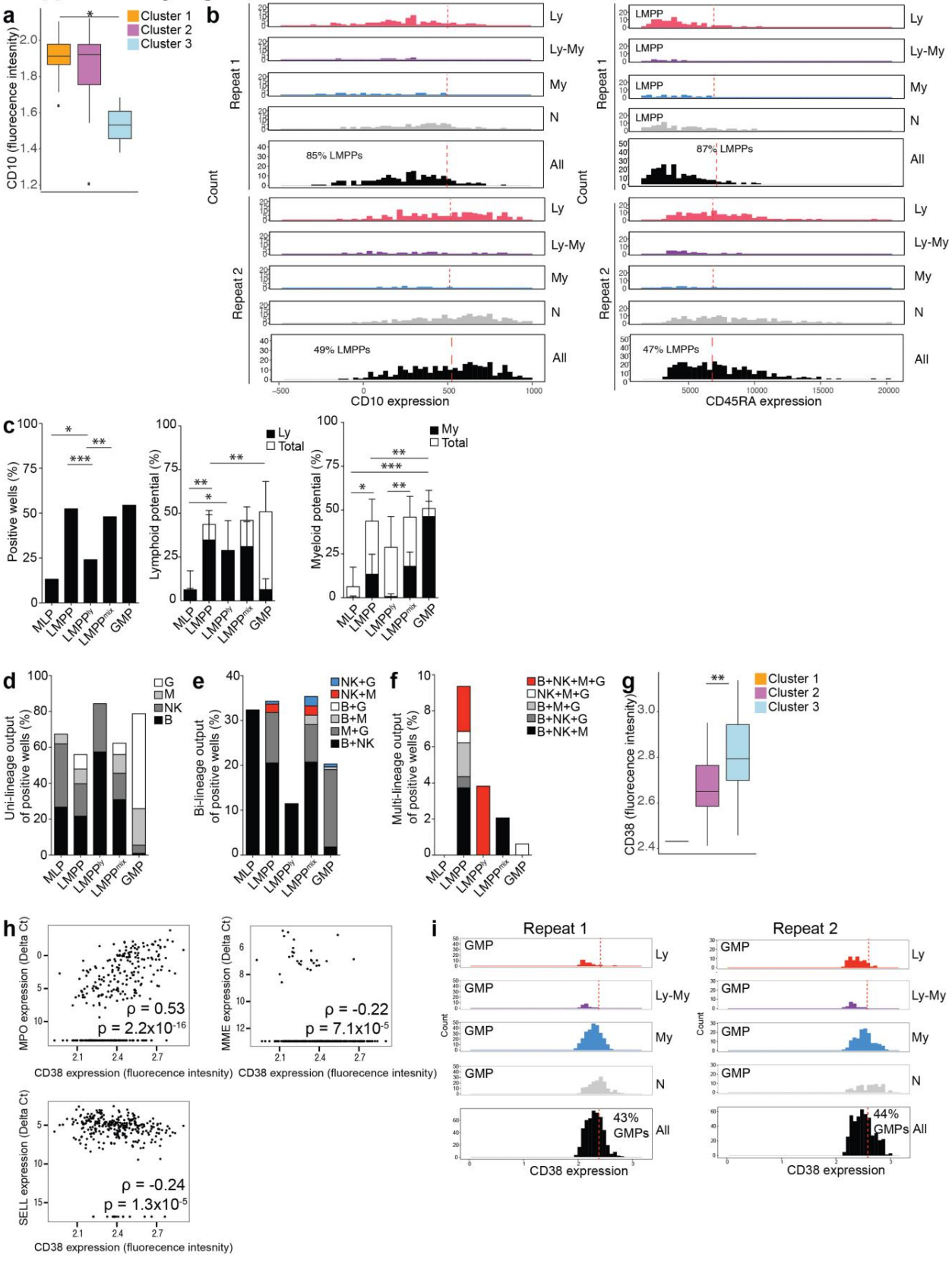


Supplementary Figure 5

Transcriptional heterogeneity of single lympho-myeloid progenitor cells.

(a) Heatmap showing hierarchical clustering on genes (rows) and single cells (columns), based on qRT-PCR gene expression analysis. **(b)** Top, Cell type composition of the 3 clusters identified in the dendrogram in (a). Middle, differentially expressed genes between cluster pairs. Bottom, expected number of cells per cluster, based on the functional assays. **(c)** Diffusion map dimensionality reductions colored in by cell type (left) and cluster membership (right), based on qRT-PCR gene expression analysis. **(d)** Heatmap showing clustering of single LMPP, GMP and MLP using the 55 most highly and variably expressed genes between clusters. Single cell RNA-sequencing data from two donors were used to generate the clusters. The heat map shows clustering from one of two. Data from the other donor is in **Fig. 5b.** Log-normalised gene expression (rows) for each single cell (columns) is shown. **(e)** Cell type composition of the 3 clusters identified in the heatmap in (d) and **Fig. 5b**. **(f)** PCA plot colored by cell type (top) or cluster membership (bottom). For single cell qRT-PCR gene expression analysis: data from 4 CB donors. For single cell RNA sequencing analysis: data from 2 donors (donor 1 in **Fig. 5**, donor 2 in current figure).

Supplementary Figure 6

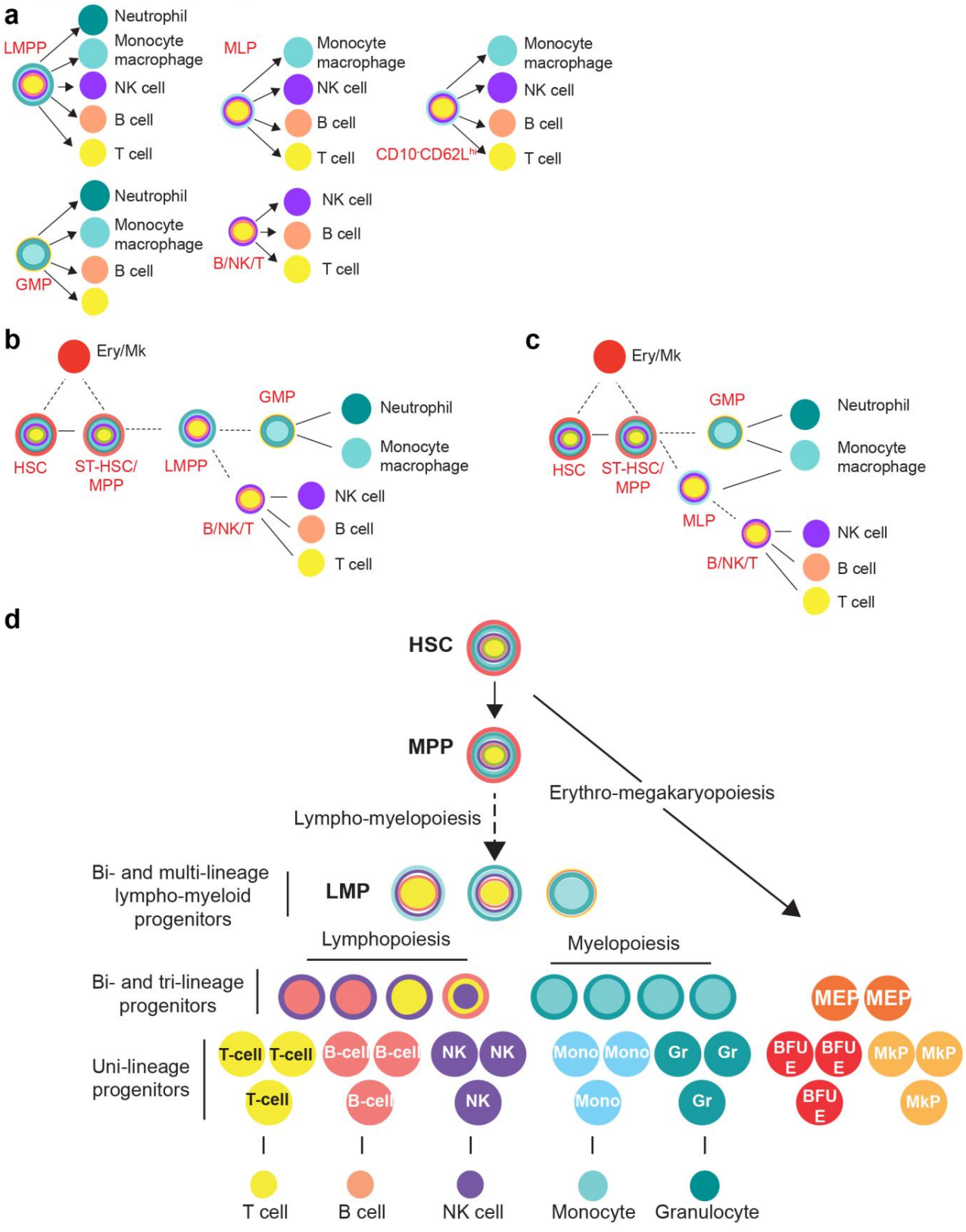


Supplementary Figure 6

Further functional purification of the lympho-myeloid progenitor populations.

(a) CD10 fluorescence intensity in LMPPs in the 3 clusters identified from clustering of single cell RNA Seq data presented **Fig. 5b**. Wilcoxon rank sum test is used to define significance. **(b)** Histograms showing expression of CD10 and CD45RA for LMPP with different functional output. The percentage cut-offs used to define the new sorting strategy for $LMPP^{ly}$ and $LMPP^{mix}$ are shown. Thresholds were defined based on maximum CD10 and CD45RA expression of LMPPs with myeloid output. **(c)** Total cloning efficiency (left) of single MLP, LMPP, $LMPP^{ly}$, $LMPP^{mix}$ and GMP in SF7b culture ($LMPP^{ly}$: 26/108 cells, $LMPP^{mix}$: 48/100). Significance defined using Fisher's exact test. Cloning efficiency of lymphoid (Ly, middle) and myeloid lineages (My, right). Bars indicate total cloning efficiency; filled portion indicates the proportion of lymphoid (lymphoid plus mixed clones) or myeloid potential (myeloid plus mixed clones). Mean ± SD is shown. Significance is defined using students t-test. **(d)** Single-, **(e)** bi- and **(f)** multi-lineage outputs from single cells in SF7b culture, presented as percentage of the positive wells. **(g)** CD38 fluorescence intensity in GMPs in the 3 clusters identified from clustering of single cell RNA Seq data presented **Fig. 5b**. Wilcoxon rank sum test is used to define significance. **(h)** Correlation between expression of selected genes and cell surface marker expression in GMPs. ρ indicates Spearman's rank correlation coefficient values and p is corresponding p-value from cor.test function in R. **(i)** Histograms showing expression of CD38 for GMP with different functional output. The percentage cut-off used to define the new sorting strategy for GMP $CD38^{hi}$ and $CD38^{mid}$ are shown. For functional assay based on the revised LMPP sorting strategy (SF7b culture) data are from 6 CB donors; for LMPP, MLP and GMP controls data from 9 CB donors (including 3 CB donors used for **Fig. 2f-j**).

Supplementary Figure 7



Supplementary Figure 7

Models of human lympho-myeloid differentiation.

(a) Lineage output of lympho-myeloid progenitors **(b-c)** Models of human hemopoietic differentiation and lineage diversification. Dashed arrows - hierarchical relationships not experimentally validated. **(d)** Novel model of human lympho-myeloid differentiation. Multiple, rare, functionally distinct bi- and multi-lineage lympho-myeloid progenitors (LMP) could differentiate either directly from hematopoietic stem cells (HSC) or multi-potent progenitors (MPP). We have used the term LMP, rather than LMPP (lymphoid primed multi-potential progenitor) to describe these lympho-myeloid progenitors, as not all LMP will be lymphoid biased. Multi-lineage LMPs are rare. Lymphoid only and myeloid only progenitors are shown below. Bi-lineage progenitors are more frequent and uni-potent progenitors are most common. The hierarchical relationships between LMP and lymphoid and myeloid only bi- and uni-lineage progenitors remain to be determined. This model also leaves open the question of when commitment to either lymphoid or myeloid fate occurs. Erythro-megakaryocytic differentiation leading to a MEP (megakaryocyte-erythroid progenitor), erythroid (BFU-E) and megakaryocytic progenitor (MkP) could occur directly from either HSC or MPP.

A.2 Flow Cytometric Sorting and Analysis

Table A.1: List of antibodies used for flow cytometric sorting and analysis.

Antibodies	Clone	Fluorochrome	Company	Dilution	Assay
CD2	RPA-2.10	Purified/ PE Cy5	eBioscience, BD	1:50	HSPC sorting
CD3	HIT3a	Purified/ PE Cy5	eBioscience, BD	1:50	HSPC sorting
CD4	RPA-T4	Purified/ PE Cy5	eBioscience, BD	1:50	HSPC sorting
CD7	eBio124-1D1	Purified	eBioscience	1:50	HSPC sorting
CD7	M-T701	PE Cy5	BD	1:50	HSPC sorting
CD8a	RPA-T8	Purified/ PE Cy5	eBioscience, Biolegend, BD	1:50	HSPC sorting
CD11b	ICRF44	Purified/ PE Cy5	eBioscience, BD	1:50	HSPC sorting
CD14	61D3	Purified	eBioscience	1:50	HSPC sorting
CD14	Mϕ P9	PE Cy5	BD	1:50	HSPC sorting
CD16	3G8	PE Cy5	BD	1:50	HSPC sorting
CD19	HIB19	Purified/ PE Cy5	eBioscience, BD	1:50	HSPC sorting
CD20	2H7	Purified/ PE Cy5	eBioscience, BD	1:50	HSPC sorting
CD56	MEM188	Purified	eBioscience	1:50	HSPC sorting
CD56	B159	PE Cy5	BD	1:50	HSPC sorting
CD235a	HIR2	Purified	eBioscience	1:50	HSPC sorting
CD235a	GA-R2	PE Cy5	BD	1:100	HSPC sorting
CD10	HI10a	APC	BD Biosciences	1:40	HSPC sorting
CD10	HI10a	APC Cy7	Biolegend	1:10	HSPC sorting

Continuation of Table A.1					
Antibodies	Clone	Fluorochrome	Company	Dilution	Assay
CD34	581	PerCP	Biolegend	1:20	HSPC sorting
CD34	8G12	APC	BD	1:25	HSPC sorting
CD38	HIT2	FITC	eBioscience	1:20	HSPC sorting
CD38	HB7	PE Cy7	BD	1:50	HSPC sorting
CD90	eBio5E10	Biotin	eBioscience	1:50	HSPC sorting
CD90	5E10	FITC	BD	1:25	HSPC sorting
CD45RA	HI100	PE	eBioscience	1:40	HSPC sorting
CD45RA	HI100	BV650	Biolegend	1:40	HSPC sorting
CD45RA	HI100	BV605	Biolegend	1:20	HSPC sorting
CD123	6H6	PE Cy7	eBioscience	1:10	HSPC sorting
CD123	7G3	PE	BD	1:10	HSPC sorting
CD15	HI98	FITC	eBioscience	1:20	B/NK/M/G assay, CFU assay
CD14	61D3	PE	eBioscience	1:100	B/NK/M/G assay, CFU assay
CD56	MEM-188	APC, PE Cy5	Biolegend	1:20	B/NK/M/G assay
CD19	HIB19	PE Cy7	eBioscience	1:40	B/NK/M/G assay
CD45	HI30	APC eFluor 780	eBioscience	1:40	B/NK/M/G assay, T cell assay, AML engraftment
CD34	581	BV605	Biolegend	1:20	B/NK/M/G assay
CD4	RPA-T4	APC	Biolegend	1:20	T cell assay
CD8	RPA-T8	PE Cy7	eBioscience	1:40	T cell assay
CD7	4H9	FITC	eBioscience	1:20	T cell assay

Continuation of Table A.1					
Antibodies	Clone	Fluorochrome	Company	Dilution	Assay
CD1a	HI149	PE	eBioscience	1:40	T cell assay
CD235	HIR2	PE Cy5	Biolegend	1:100	CFU assay
CD45	HI30	V450	BD	1:25	Ossicle engraftment
HLA-ABC	W6/32	FITC	eBioscience	1:50	Ossicle engraftment
HLA-ABC	W6/23	Pacific Blue	Biolegend	1:40	Ossicle engraftment
CD33	WM53	PE	BD	1:25	Ossicle engraftment
CD19	HIB19	APC	BD	1:25	Ossicle engraftment
CD14	Mϕ P9	PE/APC-Cy7	BD	1:25	Ossicle engraftment
CD15	MMA	FITC	BD	1:50	Ossicle engraftment
hCD45	HI30	APC-eFluor 780	eBioscience	1:40	Leukaemic engraftment
mCD45	30-F11	APC	Biolegend	1:40	Leukaemic engraftment
CD33	WM53	PE	eBioscience	1:40	Leukaemic engraftment
CD19	HIB19	FITC	eBioscience	1:20	Leukaemic engraftment
CD38	HIT2	BV421	Biolegend	1:20	AML Culture
CD33	WM53	APC Cy7	Biolegend	X	AML Culture
Annexin-V	NA	APC	Biolegend	X	AML Culture
CD11b	ICRF44	APC Cy7	Biolegend	X	AML Culture
CD15	HI98	eFluor 450	eBioscience	X	AML Culture
Streptavidin	NA	APC eFluor 780	eBioscience	1:50	HSPC sorting

Continuation of Table A.1					
Antibodies	Clone	Fluorochrome	Company	Dilution	Assay
F(ab')2-Goat anti-Mouse IgG (H+L)	Polyclonal	Qdot 605	Invitrogen	1:50	HSPC sorting
Hoechst 33258 (10 mg/ml)	NA	NA	Invitrogen	1:10000	All assays except Ossicle engraftment
Propidium-Iodide(1 mg/ml)	NA	NA	Thermo Fisher	1:1000	Ossicle engraftment

A.3 *In Vitro* LM Differentiation Assays

Table A.2: Summary of *In Vitro* LM Differentiation Assays.

Time of Analysis	Cytokines
SGF15/2	
2-2.5 weeks	20 ng/ml SCF 10 ng/ml G-CSF 10 ng/ml FLT3L 10 ng/ml IL-15 10 ng/ml IL-2 0.1 µM DuP-697
S7T2GM/G/M	
4 weeks	100 ng/ml SCF 10 ng/ml IL-7 50 ng/ml TPO 10 ng/ml IL-2 20 ng/ml GM-CSF 20 ng/ml G-CSF 10 ng/ml M-CSF
SF7a	
4-7 weeks	10 ng/ml SCF 5 ng/ml FLT3L 5 ng/ml IL-7
SF7b/Dox	
B/NK/M/G: 3 weeks, T cell: 6 weeks	50 ng/ml SCF 20 ng/ml FLT3L 10 ng/ml IL-7 20 ng/ml Insulin 1 µg/ml Doxycyclin (at week 3)

A.4 CRISPR-Cas9 Editing of Leukaemic Cells

Table A.3: sgRNAs for CRISPR-Cas9 Editing of Leukaemic Cells. Benchling (`https://benchling.com/crispr`) software was used to design sgRNAs, which uses on-target (efficiency) and off-target (specificity) scores to rank sgRNAs [87, 88].sgRNAs were designed with the aim of introducing indels into a coding exon of each gene, with the goal of knocking-out the functional protein. sgRNAs were chosen based on their on-target (efficiency) score and 5 sgRNAs were tested for each gene to assess editing.

Gene	#	sgRNA	Exon	Specificity Score	Efficiency Score
CLP1	1	AGACATAAGCCACCTCAGTG	2	65.70	77.39
	2	ACCAAAATGATACACCAGAG	2	65.53	74.18
	3	GAACCAAAATGATACACCAG	2	57.06	73.92
	4	GCGGAAAAGGAAGAAGAGCG	2	46.64	73.02
	5	TAGACTTGCCCACATCAGTG	2	69.34	72.89
COPS2	1	CTGATAACTCCCATAATCAG	7	40.78	67.81
	2	GAAGGTGAAAAAGGAGAATG	3	25.07	67.60
	3	AAAAGAAGATGACCCAAAAG	2	30.91	65.87
	4	AAAATTATATTTAGAACGAG	6	37.41	62.90
	5	TAGTAACTCCGAGCCAAATG	2	44.61	61.77
FASTKD5	1	GTGAGCAGCACACAGCATGG	2	32.28	75.30
	2	TGAAGGAGTAATGAATGCAG	2	35.64	72.62
	3	AAGATGTTGCCAAGATTCTG	2	36.98	71.43
	4	ATTGCCTGGTCAAATCTGTG	2	39.86	71.37
	5	CCCAGGGAAGGAATTCGCTG	2	43.75	71.23
PSMD12	1	AGTAGTGAAGATGTGCTATG	3	28.09	72.27
	2	ACTCTACGAATGGTTACCGA	4	36.57	71.42
	3	TTATGCTTTTGTCCAAAAGG	3	33.74	69.11
	4	ACAGTGCTGTACTTATGTTG	4	41.18	64.32
	5	ACGGGATGTCGATACCATAT	3	47.93	59.68
SBDS	1	GGGTCTGCAGAACTTCATCG	2	37.35	77.67
	2	AGATAGAACGTGCTCACATG	4	29.63	70.30
	3	AACATTTACAAACACTGAGT	2	24.52	62.12
	4	AAACATTTACAAACACTGAG	2	24.23	61.99
	5	CCTTATTGAGAGAGCCATGA	3	28.34	61.12
WDR26	1	GAAGTACCTAGAATACCTGG	4	42.13	81.13
	2	GACATCCTGACTCTTGCATG	2	39.79	70.99
	3	GGAGGATGGCAAGGTCCTGG	4	31.54	67.20
	4	AGAATACCTGGAGGATGGCA	4	33.57	62.83
	5	TACCTAGAATACCTGGAGGA	4	36.64	60.43

Table A.4: Oligonucleotides for the Cloning of sgRNAs into PX458.

Gene	sgRNA #	Oligonucleotide	Direction
CLP1	1	CACCGAGACATAAGCCACCTCAGTG	FWD
	1	AAACCACTGAGGTGGCTTATGTCTC	REV
	2	CACCGACCAAAATGATACACCAGAG	FWD
	2	AAACCTCTGGTGTATCATTTTGGTC	REV
	3	CACCGAACCAAAATGATACACCAG	FWD
	3	AAACCTGGTGTATCATTTTGGTTC	REV
	4	CACCGCGGAAAAGGAAGAAGAGCG	FWD
	4	AAACCGCTCTTCTTCCTTTTCCGC	REV
	5	CACCGTAGACTTGCCCACATCAGTG	FWD
	5	AAACCACTGATGTGGGCAAGTCTAC	REV
COPS2	1	CACCGCTGATAACTCCCATAATCAG	FWD
	1	AAACCTGATTATGGGAGTTATCAGC	REV
	2	CACCGAAGGTGAAAAAGGAGAATG	FWD
	2	AAACCATTCTCCTTTTTCACCTTC	REV
	3	CACCGAAAAGAAGATGACCCAAAAG	FWD
	3	AAACCTTTTGGGTCATCTTCTTTTC	REV
	4	CACCGAAAATTATATTTAGAACGAG	FWD
	4	AAACCTCGTTCTAAATATAATTTTC	REV
	5	CACCGTAGTAACTCCGAGCCAAATG	FWD
	5	AAACCATTTGGCTCGGAGTTACTAC	REV
FASTKD5	1	CACCGTGAGCAGCACACAGCATGG	FWD
	1	AAACCCATGCTGTGTGCTGCTCAC	REV
	2	CACCGTGAAGGAGTAATGAATGCAG	FWD
	2	AAACCTGCATTCATTACTCCTTCAC	REV
	3	CACCGAAGATGTTGCCAAGATTCTG	FWD
	3	AAACCAGAATCTTGGCAACATCTTC	REV
	4	CACCGATTGCCTGGTCAAATCTGTG	FWD
	4	AAACCACAGATTTGACCAGGCAATC	REV
	5	CACCGCCCAGGGAAGGAATTCGCTG	FWD
	5	AAACCAGCGAATTCCTTCCCTGGGC	REV
PSMD12	1	CACCGAGTAGTGAAGATGTGCTATG	FWD
	1	AAACCATAGCACATCTTCACTACTC	REV
	2	CACCGACTCTACGAATGGTTACCGA	FWD
	2	AAACTCGGTAACCATTCGTAGAGTC	REV
	3	CACCGTTATGCTTTTGTCCAAAAGG	FWD
	3	AAACCCTTTTGGACAAAAGCATAAC	REV
	4	CACCGACAGTGCTGTACTTATGTTG	FWD
	4	AAACCAACATAAGTACAGCACTGTC	REV
	5	CACCGACGGGATGTCGATACCATAT	FWD
	5	AAACATATGGTATCGACATCCCGTC	REV
SBDS	1	CACCGGGTCTGCAGAACTTCATCG	FWD
	1	AAACCGATGAAGTTCTGCAGACCC	REV

Continuation of Table A.4			
Gene	sgRNA #	Oligonucleotide	Direction
	2	CACCGAGATAGAACGTGCTCACATG	FWD
	2	AAACCATGTGAGCACGTTCTATCTC	REV
	3	CACCGAACATTTACAAACACTGAGT	FWD
	3	AAACACTCAGTGTTTGTAAATGTTC	REV
	4	CACCGAAACATTTACAAACACTGAG	FWD
	4	AAACCTCAGTGTTTGTAAATGTTTC	REV
	5	CACCGCCTTATTGAGAGAGCCATGA	FWD
	5	AAACTCATGGCTCTCTCAATAAGGC	REV
WDR26	1	CACCGAAGTACCTAGAATACCTGG	FWD
	1	AAACCCAGGTATTCTAGGTACTTC	REV
	2	CACCGACATCCTGACTCTTGCATG	FWD
	2	AAACCATGCAAGAGTCAGGATGTC	REV
	3	CACCGGAGGATGGCAAGGTCCTGG	FWD
	3	AAACCCAGGACCTTGCCATCCTCC	REV
	4	CACCGAGAATACCTGGAGGATGGCA	FWD
	4	AAACTGCCATCCTCCAGGTATTCTC	REV
	5	CACCGTACCTAGAATACCTGGAGGA	FWD
	5	AAACTCCTCCAGGTATTCTAGGTAC	REV

Table A.5: Sequencing Primers for sgRNAs.

Gene	Exon	Primer Sequence	Direction
CLP1	2	AGAGCGAGAAACAGAACTTCG	FWD
	2	TCTGAGACTGACATATTTGGAATTG	REV
COPS2	2	TGACTCCTCCTTCCTCCTCTC	FWD
	2	TTTTCCCCTGAATAACCAACC	REV
	3	TGGCTCTTTATTTTGCTGATGA	FWD
	3	GATCGAGACCATCCTGGCTA	REV
	5	CAAAGATGCACTTTTAGTTTGTTCA	FWD
	5	GCAAAACCCCACCTCTACAA	REV
	7	AAGATTTCCAGGCTAATGCAAA	FWD
	7	GTACGGCTTGGCCTAAACAC	REV
FASTKD5	2	TCATGTGGTTTGGGCATTTA	FWD 1
	2	GCTGTGAACACGGTATTCTGG	REV 1
	2	TGAATTTGTCATGCGGAAAA	FWD 2
	2	CGGGTATGAGGCAAAATCAT	REV 2
PSMD12	2-3	TGTTTTAGGAAGGAAGACTTCAAGA	FWD
	2-3	GCTTACAGTCCCAGCTGCTT	REV
	4	CTGCATCTCTGCATCTGAGG	FWD
	4	TCATGCTTAGGCTTGCTTACC	REV
SBDS	2	TGAGGTTACAGTGACCCGAGA	FWD
	2	CTGGGCAACAGAGCAAGACT	REV
	3	TTTTCTCTCCATCCAGTTACTCA	FWD

Continuation of Table A.5			
Gene	Exon	Primer Sequence	Direction
	3	CTGGCCAACATGGTGAAAC	REV
	4	CCTGGCCTTCACTTTCTTCA	FWD
	4	TCAAGACCGTCCTGGCTAAC	REV
WDR26	2-3	TTTAACCAGGCACATGAAAGC	FWD
	2-3	AGTAGCACTGAAAGCATGAAGAG	REV
	4	GATGTGGAATTGAGCAAAGTCA	FWD
	4	GCCTGGCACATGACAAGTATT	REV

A.5 *In Vitro* Assays of LM Progenitors

Table A.6: Lineage Outputs From LDA of LM Progenitors. An LDA was carried out on the LM progenitors. 1, 2, 5, 10 and 20 LMPP/MLP/GMP cells were plated in SGF15/2 culture conditions and culture output analysed by flow cytometry after 2–2.5 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the lineage output from the LDA. Number of cell plated is shown in columns. Lineage output determined by flow cytometry is shown in rows. See Table 3.2 for LDA frequencies.

Output	1 cell	2 cells	5 cells	10 cells	20 cells
LMPP					
B	33	59	25	14	12
NK	27	49	17	14	12
M	25	36	13	12	10
G	11	20	9	8	11
B+NK	8	32	14	14	12
M+G	5	12	5	7	10
B+M	3	18	12	12	10
B+G	3	14	9	8	11
M+NK	13	23	7	12	10
NK+G	3	16	3	8	11
B+NK+M	0	11	6	12	10
B+NK+G	0	12	3	8	11
B+M+G	1	7	5	7	10
NK+M+G	2	9	1	7	10
B+NK+M+G	0	5	1	7	10
Total plated	103	99	28	14	12
MLP					
B	12	19	12	11	3
NK	6	9	8	9	4
M	0	2	0	2	0
G	0	2	0	0	0
B+NK	2	4	4	7	2
M+G	0	0	0	0	0
B+M	0	1	0	1	0

Continuation of Table A.6					
Output	1 cell	2 cells	5 cells	10 cells	20 cells
B+G	0	1	0	0	0
M+NK	0	1	0	1	0
NK+G	0	0	0	0	0
B+NK+M	0	1	0	1	0
B+NK+G	0	0	0	0	0
B+M+G	0	0	0	0	0
NK+M+G	0	0	0	0	0
B+NK+M+G	0	0	0	0	0
Total plated	124	108	30	18	6
GMP					
B	5	19	9	8	7
NK	18	36	25	20	11
M	56	85	45	34	11
G	36	60	36	24	11
B+NK	3	12	7	6	7
M+G	14	39	33	24	11
B+M	1	11	8	7	7
B+G	0	8	8	6	7
M+NK	9	19	22	20	11
NK+G	5	11	18	13	11
B+NK+M	0	6	6	6	7
B+NK+G	0	5	6	5	7
B+M+G	0	6	7	6	7
NK+M+G	4	6	15	13	11
B+NK+M+G	0	3	5	5	7
Total plated	144	143	48	35	11

Table A.7: Lineage Outputs From Single-Cell SGF15/2 Culture of LM Progenitors. Single-cell SGF15/2 cultures were carried out on the LM progenitors. 1,136 LMPPs, 710 MLPs and 1,622 GMPs as single cells, isolated from 22 CB donors were cultured and culture output analysed by flow cytometry after 2–2.5 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the lineage output from the single-cell cultures. Lineage output determined by flow cytometry is shown in columns. CB replicates are shown in rows.

CB	B	NK	M	G	B + NK	M + G	B + M	B + G	NK + M	NK + G	B + NK + M	B + NK + G	B + M + G	NK + M + G	B + NK + M + G	Positive	Negative	Total plated	Cloning efficiency
LMPP																			
CB969	3	2	2	1.0	1	0	0.0	2.0	1	1.0	0	0.0	0.0	1	0.0	14	5	19	74
CB747	4	0	3	0.0	2	0	0.0	0.0	0	0.0	0	0.0	0.0	0	0.0	9	3	12	75
CB011	6	2	2	2.0	4	0	1.0	0.0	6	0.0	0	0.0	0.0	1	0.0	24	12	36	67
CB131	7	1	0	0.0	1	2	1.0	0.0	4	0.0	0	0.0	1.0	0	0.0	17	19	36	47
CB514	71	2	1	0.0	7	2	2.0	0.0	1	0.0	3	0.0	1.0	0	0.0	90	93	183	49
CB569	3	0	0	1.0	8	0	0.0	0.0	1	0.0	0	4.0	0.0	3	0.0	20	16	36	56
CB173	35	1	3	2.0	3	4	0.0	0.0	2	0.0	0	0.0	0.0	0	0.0	50	26	76	66
CB075	30	2	5	1.0	4	3	0.0	0.0	3	0.0	1	0.0	0.0	1	0.0	50	26	76	66
CB073	15	0	1	0.0	1	0	0.0	0.0	0	0.0	0	0.0	0.0	0	0.0	17	22	39	44
CB385	73	3	3	1.0	18	4	2.0	0.0	11	0.0	7	0.0	1.0	1	0.0	124	55	179	69
CB787	46	0	0	0.0	5	0	0.0	0.0	0	0.0	0	0.0	0.0	0	0.0	51	93	144	35
CB907 + CB911	7	1	2	0.0	2	4	0.0	0.0	5	0.0	3	0.0	1.0	2	1.0	28	8	36	78
CB487	14	3	0	0.0	11	2	0.0	1.0	2	0.0	2	0.0	0.0	0	0.0	35	13	48	73
CB589	7	2	0	0.0	1	1	0.0	0.0	0	0.0	0	0.0	0.0	0	0.0	11	13	24	46
CB985	13	2	0	0.0	2	1	0.0	0.0	1	0.0	0	0.0	0.0	0	0.0	19	29	48	40
CB591	6	0	0	0.0	0	0	0.0	0.0	0	0.0	0	0.0	0.0	0	0.0	6	10	16	38
CB005	12	0	0	0.0	2	0	1.0	0.0	1	0.0	1	0.0	0.0	0	0.0	17	31	48	35
CB073	8	0	0	0.0	0	0	0.0	1.0	0	0.0	0	0.0	0.0	0	0.0	9	23	32	28
CB637	12	0	4	0.0	2	3	0.0	0.0	0	0.0	1	0.0	0.0	2	0.0	24	24	48	50
Absolute Number	372	21	26	8.0	74	26	7.0	4.0	38	1.0	18	4.0	4.0	11	1.0	615	521	1136	54
Percentage From All Wells	33	2	2	0.7	6	2	0.6	0.4	3	0.1	2	0.4	0.4	1	0.1	NA	NA	NA	NA
Percentage From Positive Wells	60	3	4	1.3	12	4	1.1	0.7	6	0.2	3	0.7	0.7	2	0.2	NA	NA	NA	NA
MLP																			
CB969	0	0	0.0	0.0	0	0	0	0	0.0	0	0	0	0	0	0	0	15	15	0
CB747	7	1	0.0	0.0	1	0	0	0	0.0	0	0	0	0	0	0	9	28	37	24
CB011	1	2	0.0	0.0	1	0	0	0	0.0	0	0	0	0	0	0	4	32	36	11
CB131	2	1	0.0	0.0	0	0	0	0	0.0	0	0	0	0	0	0	3	33	36	8
CB514	7	0	0.0	0.0	0	0	0	0	0.0	0	0	0	0	0	0	7	161	168	4
CB569	3	0	0.0	1.0	2	0	0	0	2.0	0	0	0	0	0	0	8	52	60	13
CB173 + CB075 + CB073	11	0	1.0	0.0	0	0	0	0	0.0	0	0	0	0	0	0	12	64	76	16
CB385	7	3	0.0	0.0	2	0	0	0	0.0	0	0	0	0	0	0	12	79	91	13
CB787	4	2	0.0	0.0	1	0	0	0	0.0	0	0	0	0	0	0	7	41	48	15
CB487	2	0	0.0	0.0	0	0	0	0	0.0	0	0	0	0	0	0	2	46	48	4
CB005	0	1	0.0	0.0	0	0	0	0	0.0	0	0	0	0	0	0	1	31	32	3
CB073	6	1	0.0	0.0	0	0	0	0	0.0	0	0	0	0	0	0	7	24	31	23
CB637	3	1	0.0	0.0	0	0	0	0	0.0	0	0	0	0	0	0	4	28	32	12

Continuation of Table A.7

CB	B	NK	M	G	B + NK	M + G	B + M	B + G	NK + M	NK + G	B + NK + M	B + NK + G	B + M + G	NK + M + G	B + NK + M + G	Positive	Negative	Total plated	Cloning efficiency
Absolute Number	53	12	1.0	1.0	7	0	0	0	2.0	0	0	0	0	0	0	76	634	710	11
Percentage From All Wells	8	2	0.1	0.1	1	0	0	0	0.3	0	0	0	0	0	0	NA	NA	NA	NA
Percentage From Positive Wells	70	16	1.3	1.3	9	0	0	0	2.6	0	0	0	0	0	0	NA	NA	NA	NA
GMP																			
CB969	0	2	14	2	0	0	0.0	0	1	1.0	0.0	0	0.0	1	0.0	21	15	36	58
CB747	1	1	13	2	0	3	0.0	0	1	0.0	0.0	0	0.0	0	0.0	21	15	36	58
CB011	0	1	5	7	0	3	0.0	0	3	0.0	0.0	0	0.0	2	0.0	21	15	36	58
CB131	0	1	4	10	3	4	1.0	0	0	0.0	0.0	0	0.0	1	0.0	24	12	36	67
CB514	9	5	21	52	1	22	2.0	0	2	0.0	0.0	0	0.0	5	0.0	119	33	152	78
CB569	1	4	4	6	2	12	0.0	0	2	1.0	0.0	0	0.0	1	0.0	33	17	50	66
CB173	10	3	38	31	0	38	0.0	0	9	0.0	0.0	0	0.0	3	0.0	132	47	179	74
CB075	5	6	28	42	2	13	0.0	0	5	0.0	1.0	0	0.0	0	0.0	102	46	148	69
CB073	1	2	19	36	0	18	0.0	0	1	0.0	0.0	0	0.0	0	0.0	77	69	146	53
CB385	3	16	15	30	4	28	0.0	0	8	0.0	1.0	0	0.0	5	0.0	110	33	143	77
CB787	10	20	10	28	11	26	1.0	0	1	2.0	0.0	0	0.0	0	0.0	109	35	144	76
CB907	4	1	12	7	1	13	0.0	0	6	0.0	0.0	0	0.0	0	0.0	44	16	60	73
CB911	1	5	4	10	1	16	0.0	0	4	0.0	0.0	0	1.0	1	1.0	44	28	72	61
CB909	2	2	6	12	1	9	1.0	0	4	0.0	0.0	0	0.0	0	0.0	37	11	48	77
CB485	2	1	6	15	3	8	0.0	0	0	1.0	2.0	0	0.0	1	1.0	40	8	48	83
CB487	0	5	4	16	2	7	0.0	0	4	0.0	2.0	0	0.0	1	0.0	41	7	48	85
CB589	0	0	9	8	2	5	2.0	0	4	0.0	1.0	0	0.0	0	0.0	31	17	48	65
CB985	8	1	4	7	4	9	0.0	0	1	0.0	3.0	0	0.0	0	0.0	37	11	48	77
CB591	1	7	4	9	2	8	0.0	0	1	0.0	0.0	0	0.0	0	0.0	32	16	48	67
CB005	3	2	2	2	0	7	0.0	0	3	0.0	0.0	0	0.0	0	0.0	19	13	32	59
CB073	0	4	5	3	1	7	1.0	0	1	0.0	0.0	0	0.0	0	0.0	22	10	32	69
CB637	1	2	5	5	0	7	0.0	0	3	2.0	0.0	0	0.0	4	0.0	29	3	32	91
Absolute Number	62	91	232	340	40	263	8.0	0	64	7.0	10.0	0	1.0	25	2.0	1145	477	1622	71
Percentage From All Wells	4	6	14	21	2	16	0.5	0	4	0.4	0.6	0	0.1	2	0.1	NA	NA	NA	NA
Percentage From Positive Wells	5	8	20	30	4	23	0.7	0	6	0.6	0.9	0	0.1	2	0.2	NA	NA	NA	NA

Table A.8: Fisher's Test of Single-Cell SGF15/2 Culture of LM Progenitors. Single-cell SGF15/2 cultures were carried out on the LM progenitors. 1,136 LMPPs, 710 MLPs and 1,622 GMPs as single cells, isolated from 22 CB donors were cultured and culture output analysed by flow cytometry after 2–2.5 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the Fisher's exact test on the lineage output from the single-cell cultures. S: significant, NS: not significant, B: B cell output in any combinations, NK: NK cell output in any combinations, M: M (monocytic) output in any combinations, G: G (granulocytic) output in any combinations, Ly: 1Ly + 2Ly, My: 1My + 2My, Ly-My: lympho-myeloid output in any combinations.

Comparison	B	NK	M	G	Ly	My	Ly-My	Cloning Efficiency
LMPP vs MLP	NS (p=1.0000)	NS (p=1.0000)	S (p<0.0001)	S (p=0.0092)	S (p<0.0001)	NS (p=0.0519)	S (p=0.0019)	S (p<0.0001)
LMPP vs GMP	S (p<0.0001)	S (p=0.0025)	S (p<0.0001)	S (p<0.0001)	S (p<0.0001)	S (p<0.0001)	S (p=0.0125)	S (p<0.0001)
MLP vs GMP	S (p<0.0001)	NS (p=0.1916)	S (p<0.0001)	S (p<0.0001)	S (p<0.0001)	S (p<0.0001)	S (p=0.0268)	S (p<0.0001)

Table A.9: Lineage Outputs From Single-Cell S7T2GM/G/M Culture of LM Progenitors. Single-cell S7T2GM/G/M cultures were carried out on the LM progenitors. 96 LMPPs, 52 MLPs and 110 GMPs as single cells, isolated from 2 CB donors were cultured and culture output analysed by flow cytometry after 2–2.5 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the lineage output from the single-cell cultures. Lineage output determined by flow cytometry is shown in columns. CB replicates are shown in rows.

CB	B	NK	M	G	B + NK	M + G	B + M	B + G	NK + M	NK + G	B + NK + M	B + NK + G	B + M + G	NK + M + G	B + NK + M + G	Positive	Negative	Total plated	Cloning efficiency
LMPP																			
CB485	0	26	6	0	3	0	0	0	6	0	2	0	0	0	0	43	17	60	72
CB487	0	7	5	0	7	1	0	0	5	0	1	0	0	0	0	26	10	36	72
Absolute Number	0	33	11	0	10	1	0	0	11	0	3	0	0	0	0	69	27	96	72
Percentage From All Wells	0	34	12	0	10	1	0	0	12	0	3	0	0	0	0	NA	NA	NA	NA
Percentage From Positive Wells	0	48	16	0	14	1	0	0	16	0	4	0	0	0	0	NA	NA	NA	NA
MLP																			
CB485	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2	37	39	5
CB487	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	2	11	13	15
Absolute Number	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	4	48	52	8
Percentage From All Wells	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	NA	NA	NA	NA
Percentage From Positive Wells	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	NA	NA	NA	NA
GMP																			
CB485	0	2	28	0	1.0	5	0	0	2	0	0	0	0	0	0	38	6	44	86
CB487	0	1	54	0	0.0	1	0	0	4	0	0	0	0	0	0	60	6	66	91
Absolute Number	0	3	82	0	1.0	6	0	0	6	0	0	0	0	0	0	98	12	110	89
Percentage From All Wells	0	3	74	0	0.9	6	0	0	6	0	0	0	0	0	0	NA	NA	NA	NA
Percentage From Positive Wells	0	3	84	0	1.0	6	0	0	6	0	0	0	0	0	0	NA	NA	NA	NA

Table A.10: Lineage Outputs From Single-Cell SF7b/Dox Culture of LM Progenitors. Single-cell SF7b/Dox cultures were carried out on the LM progenitors. 215 LMPPs, 197 MLPs and 219 GMPs as single cells, isolated from 3 CB donors were cultured on an MS-5 stromal cell line that can be induced to express hDL1 by doxycycline treatment. The B-NK-M-G output was analysed at 3 weeks and the T cell output was analysed at 6 weeks. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the lineage output from the single-cell cultures. Lineage output determined by flow cytometry is shown in columns. CB replicates are shown in rows.

CB	B	NK	T	B + NK	B + T	NK + T	B + NK + T	M	G	M + G	B + My	T + My	NK + My	B + NK + My	B + T + My	NK + T + My	B + NK + T + My	Positive	Negative	Total Plated	Cloning Efficiency
LMPP																					
CB991	12	12	2	19	0	1.0	1.0	9	7	14	1.0	1.0	2.0	3	1.0	1.0	4	90	38	128	70
CB989	2	4	1	0	0	1.0	0.0	2	0	2	0.0	0.0	0.0	0	0.0	0.0	0	12	36	48	25
CB723	5	2	1	8	0	0.0	0.0	2	1	2	1.0	0.0	0.0	4	0.0	0.0	0	26	13	39	67
Absolute Number	19	18	4	27	0	2.0	1.0	13	8	18	2.0	1.0	2.0	7	1.0	1.0	4	128	87	215	60
Percentage From All Wells	9	8	2	13	0	0.9	0.5	6	4	8	0.9	0.5	0.9	3	0.5	0.5	2	NA	NA	NA	NA
Percentage From Positive Wells	15	14	3	21	0	1.6	0.8	10	6	14	1.6	0.8	1.6	6	0.8	0.8	3	NA	NA	NA	NA
MLP																					
CB991	9	12	1.0	10	0	0	2	2	0	0	0	0	0	0	0	0	0	36	92	128	28
CB989	0	1	0.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	47	48	2
CB723	0	0	0.0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	21	21	0
Absolute Number	9	13	1.0	10	0	0	2	2	0	0	0	0	0	0	0	0	0	37	160	197	19
Percentage From All Wells	5	7	0.5	5	0	0	1	1	0	0	0	0	0	0	0	0	0	NA	NA	NA	NA
Percentage From Positive Wells	24	35	2.7	27	0	0	5	5	0	0	0	0	0	0	0	0	0	NA	NA	NA	NA
GMP																					
CB991	0	1.0	0	3	0	1.0	0	16	44	13	0	0	0	0	0	0	0	78	50	128	61
CB989	0	0.0	0	0	0	0.0	0	4	20	4	0	0	0	0	0	0	0	28	16	44	64
CB723	0	1.0	0	0	0	0.0	0	7	11	2	0	0	0	0	0	0	0	21	26	47	45
Absolute Number	0	2.0	0	3	0	1.0	0	27	75	19	0	0	0	0	0	0	0	127	92	219	58
Percentage From All Wells	0	0.9	0	1	0	0.5	0	12	34	9	0	0	0	0	0	0	0	NA	NA	NA	NA
Percentage From Positive Wells	0	1.6	0	2	0	0.8	0	21	59	15	0	0	0	0	0	0	0	NA	NA	NA	NA

A.6 Refining the LMPP and GMP Populations: *In Vitro* Assays

Table A.11: Lineage Outputs From Single-Cell SGF15/2 Culture of LMPP^{ly} and LMPP^{mix}. Single-cell SGF15/2 cultures were carried out on the LMPP^{ly} and LMPP^{mix}. Single CB MLPs, LMPPs, LMPP^{ly} cells (56 of 244 cells), LMPP^{mix} cells (152 of 240 cells) and GMPs as single cells were cultured and culture output analysed by flow cytometry after 2–2.5 weeks of culture (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the lineage output from the single-cell cultures. Lineage output determined by flow cytometry is shown in columns. CB replicates are shown in rows.

CB	B	NK	M	G	B + NK	M + G	B + M	B + G	NK + M	NK + G	B + NK + M	B + NK + G	B + M + G	NK + M + G	B + NK + M + G	Positive	Negative	Total plated	Cloning efficiency
LMPP^{ly}																			
CB589	0	2	0	0	1	0	0	0	1	0	0	0	0	0	0	4	20	24	16.7
CB985	6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	7	33	40	17.5
CB591	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	12	12	0.0
CB005	13	0	0	0	1	0	0	0	0	0	0	0	0	0	0	14	58	72	19.4
CB073	10	1	0	0	1	0	0	0	0	0	0	0	0	0	0	12	20	32	37.5
CB637	12	3	1	0	1	0	0	0	0	0	0	2	0	0	0	19	45	64	29.7
Absolute Number	41	6	1	0	4	0	1	0	1	0	0	2	0	0	0	56	188	244	23.0
Percentage From All Wells	16.8	2.5	0.4	0.0	1.6	0.0	0.4	0.0	0.4	0.0	0.0	0.8	0.0	0.0	0.0				
Percentage From Positive Wells	73.2	10.7	1.8	0.0	7.1	0.0	1.8	0.0	1.8	0.0	0.0	3.6	0.0	0.0	0.0				
LMPP^{mix}																			
CB589	0	0	0	0	2	1	0	0	0	0	0	0	0	0	1	4	4	8	50.0
CB985	24	1	1	0	6	2	0	0	0	0	1	0	0	0	0	35	21	56	62.5
CB005	26	5	1	0	6	1	1	0	4	0	0	0	0	1	0	45	27	72	62.5
CB073	13	1	0	1	2	0	0	0	0	0	0	0	0	0	2	19	13	32	59.4
CB637	27	2	7	2	1	2	0	0	2	0	1	0	0	4	1	49	23	72	68.1
Absolute Number	90	9	9	3	17	6	1	0	6	0	2	0	0	5	4	152	88	240	63.3
Percentage From All Wells	37.5	3.8	3.8	1.3	7.1	2.5	0.4	0.0	2.5	0.0	0.8	0.0	0.0	2.1	1.7				
Percentage From Positive Wells	59.2	5.9	5.9	2.0	11.2	3.9	0.7	0.0	3.9	0.0	1.3	0.0	0.0	3.3	2.6				

Table A.12: Fisher's Test of Single-Cell SGF15/2 Culture of LMPP^{ly} and LMPP^{mix}. Single-cell SGF15/2 cultures were carried out on the LMPP^{ly} and LMPP^{mix}. Single CB MLPs, LMPPs, LMPP^{ly} cells (56 of 244 cells), LMPP^{mix} cells (152 of 240 cells) and GMPs as single cells were cultured and culture output analysed by flow cytometry after 2–2.5 weeks of culture (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the Fisher's exact test on the lineage output from the single-cell cultures. S: significant, NS: not significant, B: B cell output in any combinations, NK: NK cell output in any combinations, M: M (monocytic) output in any combinations, G: G (granulocytic) output in any combinations, Ly: 1Ly + 2Ly, My: 1My + 2My, Ly-My: lympho-myeloid output in any combinations.

Comparison	B	NK	M	G	Ly	My	Ly-My	Cloning Efficiency
LMPP vs LMPPmix	NS (p=0.3268)	NS (p=0.8394)	NS (p=0.9123)	NS (p=0.4506)	NS (p=1.0000)	NS (p=0.4545)	NS (p=0.5119)	S (p=0.0100)
LMPP vs LMPPly	NS (p=0.3002)	NS (p=0.4303)	S (p=0.0026)	NS (p=0.2193)	S (p=0.0075)	S (p=0.0496)	NS (p=0.1582)	S (p<0.0001)
LMPPmix vs LMPPly	NS (p=0.1313)	NS (p=0.5973)	S (p=0.0063)	NS (p=0.1089)	S (p=0.0182)	S (p=0.0280)	NS (p=0.4483)	S (p<0.0001)
LMPPly vs MLP	NS (p=0.3675)	NS (p=0.5422)	NS (p=0.6982)	NS (p=0.5740)	NS (p=0.4937)	NS (p=1.0000)	NS (p=0.4006)	S (p<0.0001)

Table A.13: Lineage Outputs From Single-Cell SF7b Culture of LMPP^{ly} and LMPP^{mix}. Single-cell SF7b cultures were carried out on the LMPP^{ly} and LMPP^{mix}. Single CB MLPs, LMPPs, LMPP^{ly} cells (56 of 244 cells), LMPP^{mix} cells (152 of 240 cells) and GMPs as single cells were cultured and culture output analysed by flow cytometry after 2–2.5 weeks of culture (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the lineage output from the single-cell cultures. Lineage output determined by flow cytometry is shown in columns. CB replicates are shown in rows.

CB	B	NK	M	G	B + NK	M + G	B + M	B + G	NK + M	NK + G	B + NK + M	B + NK + G	B + M + G	NK + M + G	B + NK + M + G	Positive	Negative	Total plated	Cloning efficiency
LMPP^{ly}																			
CB317	4	0.0	0	0.0	2.0	0.0	0.0	0	0.0	0.0	0.0	0	0	0	1.0	7	32	25	22
CB65	4	4.0	0	0.0	1.0	0.0	0.0	0	0.0	0.0	0.0	0	0	0	0.0	9	18	9	50
CB097	5	1.0	0	0.0	0.0	0.0	0.0	0	0.0	0.0	0.0	0	0	0	0.0	6	18	12	33
CB511	2	2.0	0	0.0	0.0	0.0	0.0	0	0.0	0.0	0.0	0	0	0	0.0	4	40	36	10
Absolute Number	15	7.0	0	0.0	3.0	0.0	0.0	0	0.0	0.0	0.0	0	0	0	1.0	26	108	82	24
Percentage From All Wells	14	6.5	0	0.0	2.8	0.0	0.0	0	0.0	0.0	0.0	0	0	0	0.9	NA	NA	NA	NA
Percentage From Positive Wells	58	26.9	0	0.0	11.5	0.0	0.0	0	0.0	0.0	0.0	0	0	0	3.8	NA	NA	NA	NA
LMPP^{mix}																			
CB317	0	0.0	0	0.0	1.0	1.0	0.0	0	0.0	0.0	0.0	0	0	0	0.0	2	4	2	50
CB5H	0	3.0	0	2.0	1.0	1.0	1.0	0	0.0	0.0	0.0	0	0	0	0.0	8	20	12	40
CB097	10	3.0	3	1.0	7.0	2.0	0.0	0	1.0	1.0	1.0	0	0	0	0.0	29	44	15	66
CB511	5	1.0	2	0.0	1.0	0.0	0.0	0	0.0	0.0	0.0	0	0	0	0.0	9	32	23	28
Absolute Number	15	7.0	5	3.0	10.0	4.0	1.0	0	1.0	1.0	1.0	0	0	0	0.0	48	100	52	48
Percentage From All Wells	15	7.0	5	3.0	10.0	4.0	1.0	0	1.0	1.0	1.0	0	0	0	0.0	NA	NA	NA	NA
Percentage From Positive Wells	31	14.6	10	6.3	20.8	8.3	2.1	0	2.1	2.1	2.1	0	0	0	0.0	NA	NA	NA	NA

Table A.14: Fisher's Test of Single-Cell SF7b Culture of LMPP^{ly} and LMPP^{mix}. Single-cell SF7b cultures were carried out on the LMPP^{ly} and LMPP^{mix}. Single CB MLPs, LMPPs, LMPP^{ly} cells (56 of 244 cells), LMPP^{mix} cells (152 of 240 cells) and GMPs as single cells were cultured and culture output analysed by flow cytometry after 2–2.5 weeks of culture (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the Fisher's exact test on the lineage output from the single-cell cultures. S: significant, NS: not significant, B: B cell output in any combinations, NK: NK cell output in any combinations, M: M (monocytic) output in any combinations, G: G (granulocytic) output in any combinations, Ly: 1Ly + 2Ly, My: 1My + 2My, Ly-My: lympho-myeloid output in any combinations.

Comparison	B	NK	M	G	Ly	My	Ly-My	Cloning Efficiency
LMPP vs LMPP mix	NS (p=0.6219)	NS (p=0.4140)	NS (p=0.5876)	NS (p=0.2465)	NS (p=0.5005)	NS (p=0.8533)	NS (p=0.6067)	NS (p=0.4895)
LMPP vs LMPP ly	NS (p=0.0548)	NS (p=0.6729)	S (p=0.0033)	S (p=0.0109)	S (p=0.0002)	S (p=0.0008)	NS (p=0.3175)	S (p<0.0001)
LMPP mix vs LMPP ly	NS (p=0.2108)	NS (p=1.0000)	S (p=0.0259)	NS (p=0.1474)	S (p=0.0036)	S (p=0.0059)	NS (p=0.6510)	S (p=0.0005)
LMPP ly vs MLP	NS (p=0.2961)	NS (p=0.0701)	NS (p=1.0000)	NS (p=0.4127)	NS (p=1.0000)	NS (p=0.5074)	NS (p=0.4127)	S (p=0.0136)

Table A.15: Lineage Outputs From Single-Cell SGF15/2 Culture of GMP CD38^{hi}. Single-cell SGF15/2 cultures were carried out on the CB CD38^{hi} GMPs (152 of 279 cells), CD38^{int} cells (508 of 693 cells) and LMPPs and culture output analysed by flow cytometry after 2–2.5 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the lineage output from the single-cell cultures. Lineage output determined by flow cytometry is shown in columns. CB replicates are shown in rows.

CB	B	NK	M	G	B + NK	M + G	B + M	B + G	NK + M	NK + G	B + NK + M	B + NK + G	B + M + G	NK + M + G	B + NK + M + G	Positive	Negative	Total plated	Cloning efficiency
GMP CD38^{hi}																			
CB019	0	3.0	10	14.0	0.0	3	0.0	0.0	1.0	0	0.0	0	0.0	0.0	0	31	35	66	47
CB907	0	3.0	20	29.0	1.0	16	0.0	0.0	2.0	0	0.0	0	0.0	0.0	0	71	60	131	54
CB911	0	1.0	2	11.0	0.0	8	0.0	0.0	1.0	0	0.0	0	0.0	1.0	0	24	11	35	69
CB909	0	0.0	2	15.0	0.0	8	0.0	0.0	1.0	0	0.0	0	0.0	0.0	0	26	21	47	55
Absolute Number	0	7.0	34	69.0	1.0	35	0.0	0.0	5.0	0	0.0	0	0.0	1.0	0	152	127	279	54
Percentage From All Wells	0	2.5	12	24.7	0.4	12	0.0	0.0	1.8	0	0.0	0	0.0	0.4	0	NA	NA	NA	NA
Percentage From Positive Wells	0	4.6	22	45.4	0.7	23	0.0	0.0	3.3	0	0.0	0	0.0	0.7	0	NA	NA	NA	NA
CD38^{mid}																			
CB019	53	26.0	56	19.0	23.0	47	5.0	1.0	24.0	0	3.0	0	1.0	3.0	0	261	120	381	68
CB907	62	22.0	61	17.0	15.0	31	3.0	0.0	14.0	0	2.0	0	0.0	2.0	0	229	59	288	80
CB911	3	2.0	2	1.0	2.0	7	0.0	0.0	1.0	0	0.0	0	0.0	0.0	0	18	6	24	75
Absolute Number	118	50.0	119	37.0	40.0	85	8.0	1.0	39.0	0	5.0	0	1.0	5.0	0	508	185	693	73
Percentage From All Wells	17	7.2	17	5.3	5.8	12	1.2	0.1	5.6	0	0.7	0	0.1	0.7	0	NA	NA	NA	NA
Percentage From Positive Wells	23	9.8	23	7.3	7.9	17	1.6	0.2	7.7	0	1.0	0	0.2	1.0	0	NA	NA	NA	NA

Table A.16: Fisher's Test of Single-Cell SGF15/2 Culture of GMP CD38^{hi}. Single-cell SGF15/2 cultures were carried out on the CB CD38^{hi} GMPs (152 of 279 cells), CD38^{int} cells (508 of 693 cells) and LMPPs and culture output analysed by flow cytometry after 2–2.5 weeks of culture. (see Chapter 2.4 and Table A.2 for culture conditions). Shown is the Fisher's exact test on the lineage output from the single-cell cultures. S: significant, NS: not significant, B: B cell output in any combinations, NK: NK cell output in any combinations, M: M (monocytic) output in any combinations, G: G (granulocytic) output in any combinations, Ly: 1Ly + 2Ly, My: 1My + 2My, Ly-My: lympho-myeloid output in any combinations.

Comparison	B	NK	M	G	Ly	My	Ly-My	Cloning Efficiency
GMP vs GMPhi	S (p<0.0001)	S (p=0.0004)	NS (p=0.4373)	S (p=0.0017)	S (p<0.0001)	S (p<0.0001)	S (p=0.0115)	S (p<0.0001)
GMPhi vs CD38mid	S (p<0.0001)	S (p<0.0001)	NS (p=0.6447)	S (p<0.0001)	S (p<0.0001)	S (p<0.0001)	S (p=0.0047)	S (p<0.0001)
GMP vs CD38mid	S (p<0.0001)	S (p=0.0043)	NS (p=0.6694)	S (p<0.0001)	S (p<0.0001)	S (p<0.0001)	NS (p=0.3890)	NS (p=0.1919)
LMPP vs CD38mid	S (p<0.0001)	NS (p=1.0000)	S (p<0.0001)	S (p<0.0001)	S (p<0.0001)	S (p<0.0001)	NS (p=0.2133)	S (p<0.0001)

A.7 Bulk RNA-Seq of Normal CB HSPC

Table A.17: Number of DEGs Between HSPC Populations. DEGs were identified between HSPC populations and shown is the number of DEGs between all HSPC populations. Genes are upregulated in a population column versus row. See Figure 3.30

	HSC	MPP	LMPP	MLP	CMP	GMP	MEP
HSC	0	7	95	252	143	143	560
MPP	6	0	59	228	43	70	200
LMPP	83	82	0	25	171	102	517
MLP	378	275	60	0	499	438	983
CMP	82	36	146	413	0	16	8
GMP	141	86	81	282	24	0	102
MEP	592	285	589	966	55	208	0

A.8 Immunophenotyping of Normal and Leukaemic BM

Table A.18: Immunophenotype Screening of CD34$^+$ AML. Relative size of HSPC populations in 47 CD34$^+$ AML samples screened for immunophenotype. CD34$^+$ compartment shown as a percentage of the Lin$^-$ compartment. All other HSPC populations shown as a percentage of the CD34$^+$ compartment. Normal size of HSPC populations in normal human BM from Table 3.1 and [54]. Fold change is relative to the average size of the HSPC population in normal human BM.

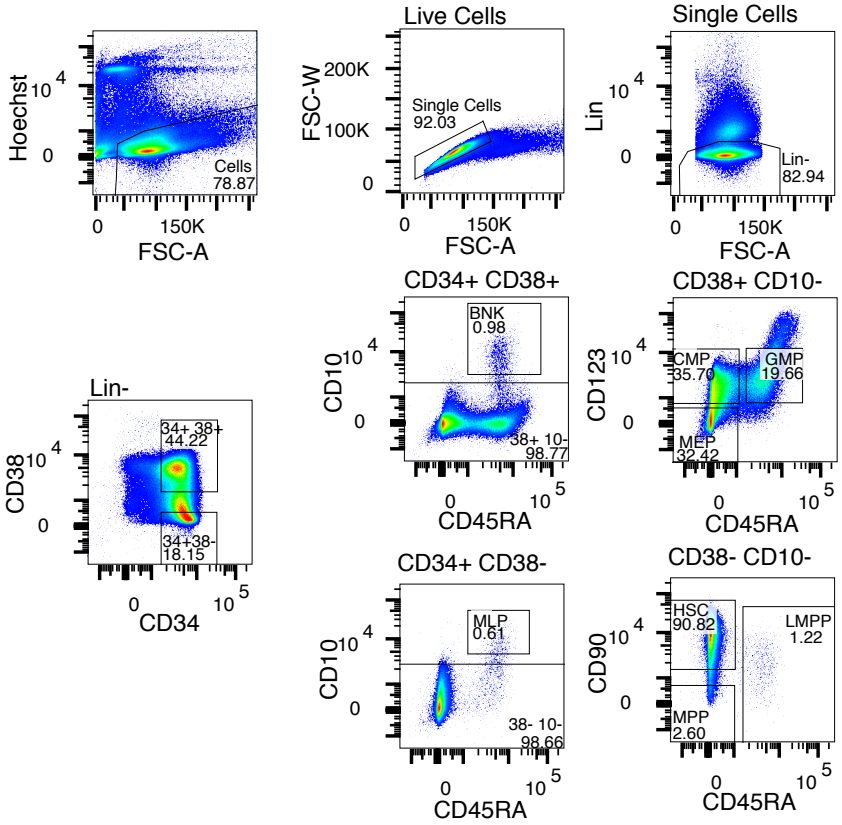
Sample	CD34$^+$	LMPP		GMP		MLP		CD38$^+$ CD45RA$^-$		CD38$^-$ CD45RA$^-$	
		% CD34$^+$	Fold Change	% CD34$^+$	Fold Change	% CD34$^+$	Fold Change	% CD34$^+$	Fold Change	% CD34$^+$	Fold Change
ABE036	86.85	1.90	9.52	91.95	5.17	0.01	0.07	2.65	0.05	0.67	0.04
BEL039	97.58	44.89	224.43	54.55	3.06	0.00	0.02	0.09	0.00	0.20	0.01
BEL048	90.31	43.02	215.09	47.99	2.7	0.00	0.00	0.90	0.02	2.39	0.16
BEL073	57.32	61.84	309.19	36.94	2.08	0.03	0.15	0.05	0.00	0.00	0.00
BEL074	25.38	21.60	107.98	78.33	4.4	0.00	0.00	0.02	0.00	0.05	0.00
BEL085	75.01	0.96	4.82	13.94	0.78	0.02	0.11	67.25	1.36	17.17	1.14
BHR070	86.87	37.68	188.40	61.59	3.46	0.02	0.12	0.07	0.00	0.10	0.01
BIR344	62.72	37.74	188.70	15.82	0.89	7.43	37.17	6.39	0.13	10.54	0.70
BIR352	3.75	24.82	124.08	72.60	4.08	0.63	3.15	0.35	0.01	0.07	0.00
BIR473	98.43	0.13	0.65	96.11	5.4	0.00	0.00	3.76	0.08	0.00	0.00
BIR489	69.47	3.92	19.59	94.81	5.33	0.00	0.00	0.11	0.00	0.06	0.00
BIR492	60.13	0.49	2.43	94.23	5.29	0.00	0.00	0.37	0.01	0.02	0.00
BIR517	85.23	90.97	454.84	6.62	0.37	0.13	0.66	0.08	0.00	0.44	0.03
BIR547	67.44	2.01	10.04	94.63	5.32	0.33	1.65	2.48	0.05	0.24	0.02
BRI1371	97.16	12.06	60.32	71.08	3.99	0.00	0.00	12.17	0.25	0.76	0.05
BRI1481	94.57	16.59	82.94	83.30	4.68	0.00	0.00	0.03	0.00	0.01	0.00
BRI1492	35.72	77.95	389.73	13.67	0.77	0.61	3.03	0.01	0.00	0.32	0.02
BRI1525	84.35	7.63	38.17	78.36	4.40	0.00	0.02	7.64	0.15	1.65	0.11
BRI1952	46.48	1.09	5.46	95.18	5.35	0.01	0.05	3.42	0.07	0.20	0.01
BRI2108	97.83	41.67	208.35	58.25	3.27	0.02	0.08	0.04	0.00	0.02	0.00

Continuation of Table A.18

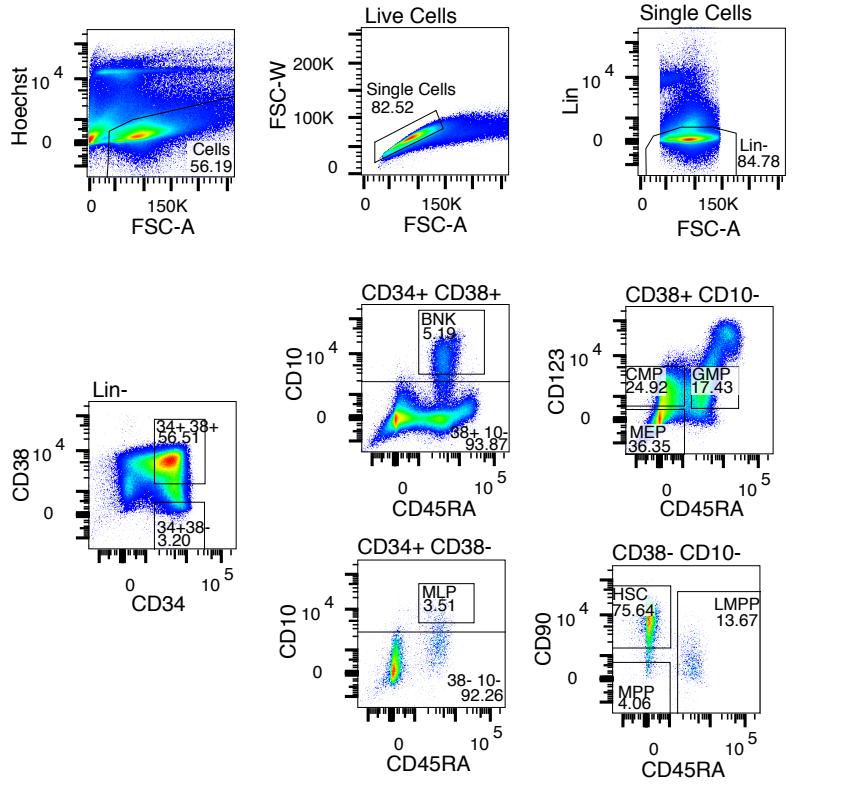
Sample	CD34$^+$	LMPP		GMP		MLP		CD38$^+$ CD45RA$^-$		CD38$^-$ CD45RA$^-$	
		% CD34$^+$	Fold Change	% CD34$^+$	Fold Change	% CD34$^+$	Fold Change	% CD34$^+$	Fold Change	% CD34$^+$	Fold Change
CAR117	88.81	6.28	31.38	27.73	1.56	0.00	0.01	25.01	0.51	32.80	2.17
CAR134	81.11	2.61	13.05	95.75	5.38	0.03	0.15	0.36	0.01	0.13	0.01
CAR143	65.00	20.35	101.77	79.35	4.46	0.01	0.03	0.00	0.00	0.01	0.00
CHR096	70.15	4.28	21.41	94.02	5.28	0.06	0.32	0.16	0.00	0.06	0.00
CUH018	94.10	39.46	197.28	60.36	3.39	0.00	0.01	0.03	0.00	0.02	0.00
GLA032	14.13	65.44	327.19	27.71	1.56	0.05	0.26	0.30	0.01	4.10	0.27
GLA067	70.30	0.15	0.73	95.60	5.37	0.00	0.01	4.02	0.08	0.13	0.01
GLA086	21.19	0.38	1.88	16.76	0.94	0.02	0.11	53.37	1.08	27.81	1.84
GLA102	85.38	2.02	10.10	4.69	0.26	0.01	0.05	58.91	1.19	34.04	2.25
GLA116	97.39	4.28	21.40	72.55	4.08	0.01	0.03	14.71	0.30	0.13	0.01
HAM039	25.11	17.48	87.42	71.98	4.04	2.66	13.30	0.08	0.00	3.48	0.23
HAM061	80.79	38.03	190.14	51.41	2.89	0.19	0.94	2.10	0.04	0.64	0.04
HER103	87.85	3.18	15.92	60.86	3.42	0.00	0.00	29.80	0.60	0.15	0.01
HER127	70.17	0.05	0.23	99.87	5.61	0.00	0.00	0.01	0.00	0.03	0.00
LEE225	48.80	1.31	6.53	54.29	3.05	0.02	0.12	36.28	0.74	0.48	0.03
NST081	93.79	0.12	0.60	78.67	4.42	0.00	0.00	20.59	0.42	0.60	0.04
OX1729	40.50	0.21	1.04	24.16	1.36	0.01	0.06	74.17	1.50	1.24	0.08
OX1908	88.61	58.52	292.58	41.44	2.33	0.01	0.03	0.03	0.00	0.01	0.00
OX2116	91.41	71.23	356.17	10.57	0.59	0.29	1.43	0.16	0.00	10.06	0.67
ROM304	19.00	0.32	1.62	94.86	5.33	0.02	0.09	2.78	0.06	0.35	0.02
RUH038	6.07	0.08	0.38	96.35	5.41	0.01	0.05	0.02	0.00	0.25	0.02
STB128	94.11	2.17	10.84	97.03	5.45	0.00	0.00	0.12	0.00	0.08	0.01
STB207	91.99	0.21	1.04	99.58	5.59	0.00	0.02	0.02	0.00	0.08	0.01
STB257	77.32	25.76	128.79	32.88	1.85	17.52	87.59	1.17	0.02	0.56	0.04

Continuation of Table A.18

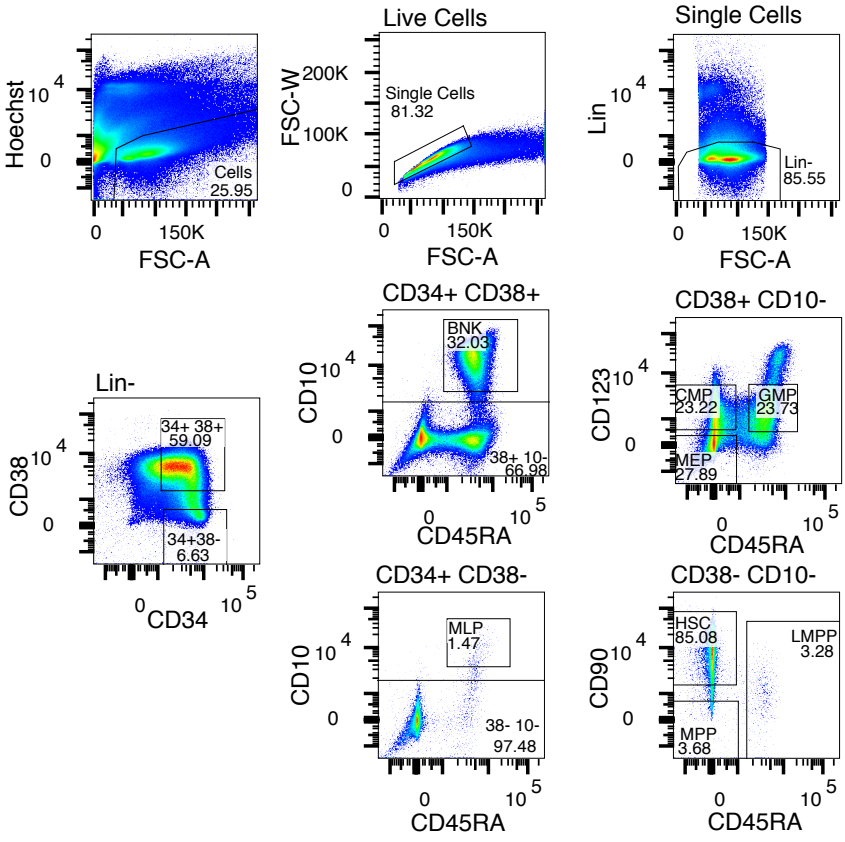
Sample	$CD34^+$	LMPP		GMP		MLP		$CD38^+$ $CD45RA^-$		$CD38^-$ $CD45RA^-$	
		% $CD34^+$	Fold Change	% $CD34^+$	Fold Change	% $CD34^+$	Fold Change	% $CD34^+$	Fold Change	% $CD34^+$	Fold Change
STB264	79.30	36.38	181.91	63.18	3.55	0.04	0.21	0.11	0.00	0.01	0.00
STB270	38.41	42.09	210.46	38.45	2.16	0.00	0.00	17.02	0.35	2.36	0.16
STB291	90.50	18.2	90.99	79.22	4.45	0.00	0.02	0.02	0.00	0.01	0.00



(a) Sample OX671

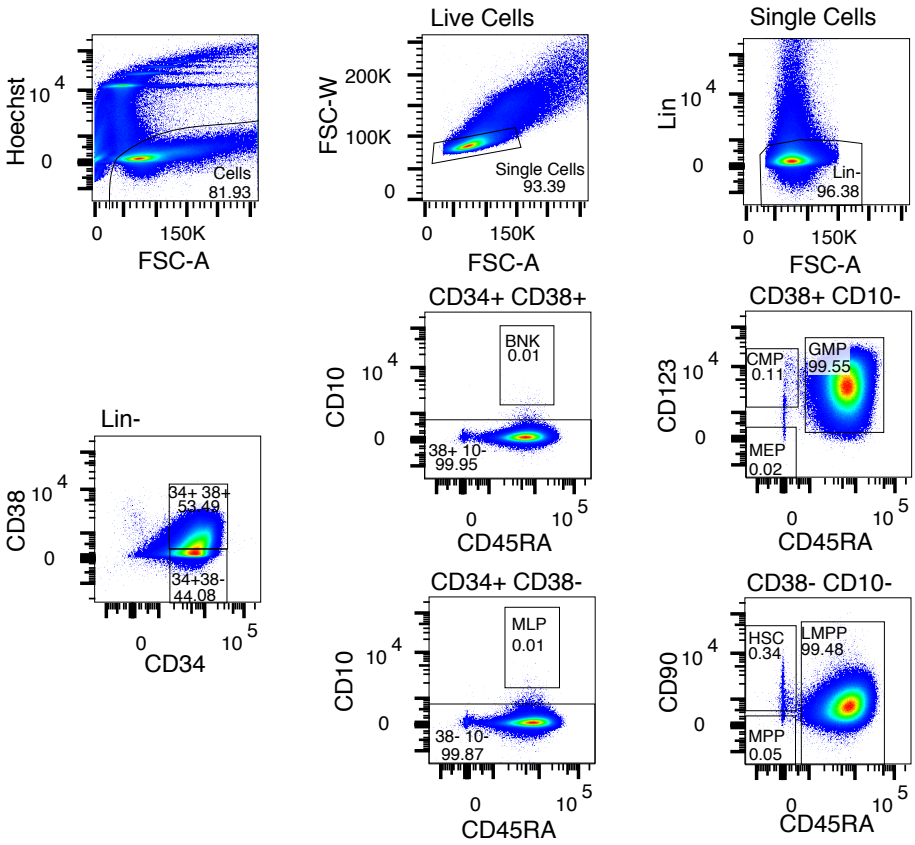


(b) Sample OX700

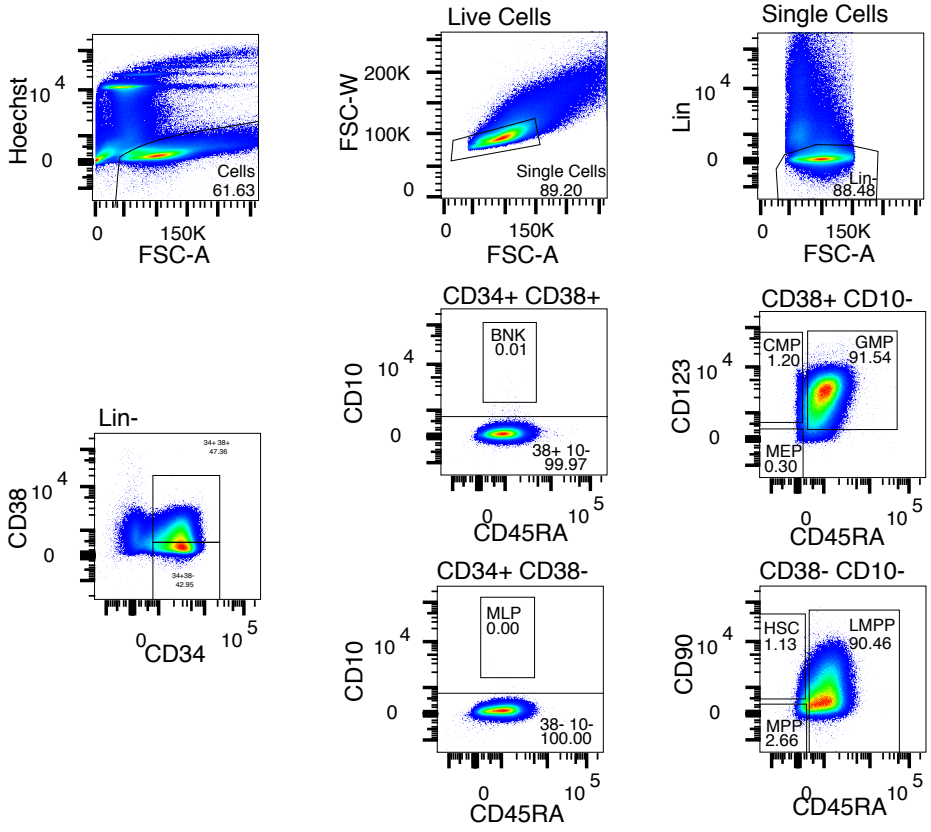


(c) Sample OX701

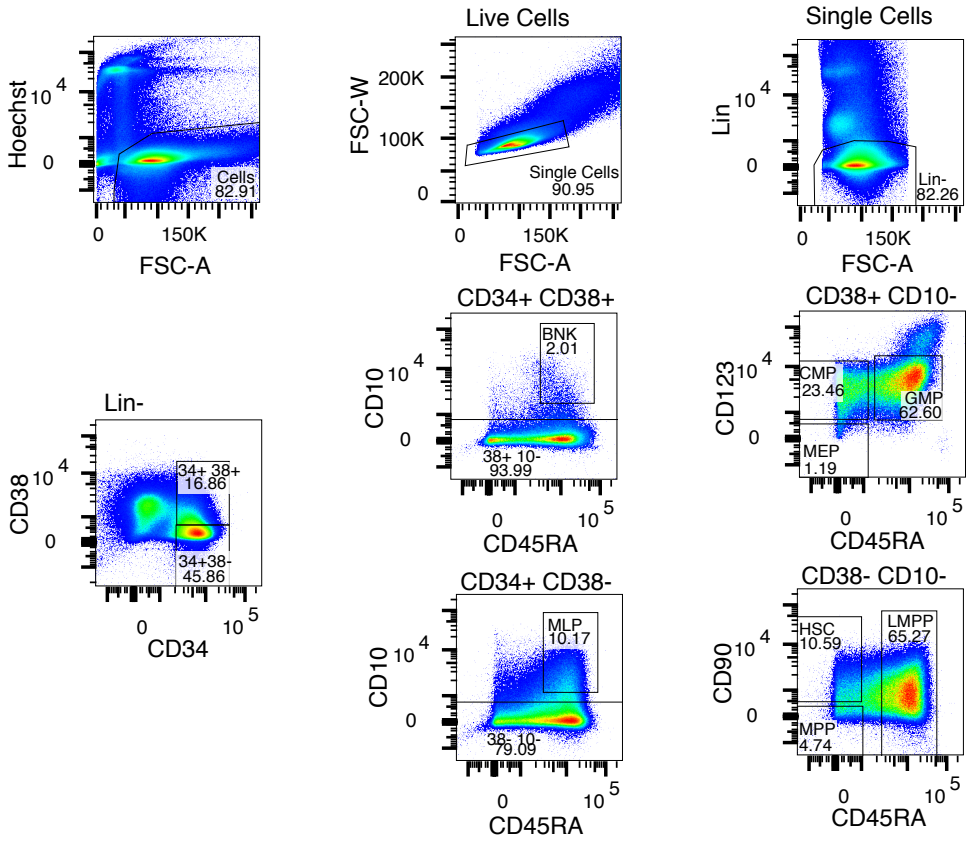
Figure A.1: Immunophenotype of normal BM. HSPC immunophenotype of normal BM samples showing expected HSPC populations. Frequencies indicate the frequency of the parent gate.



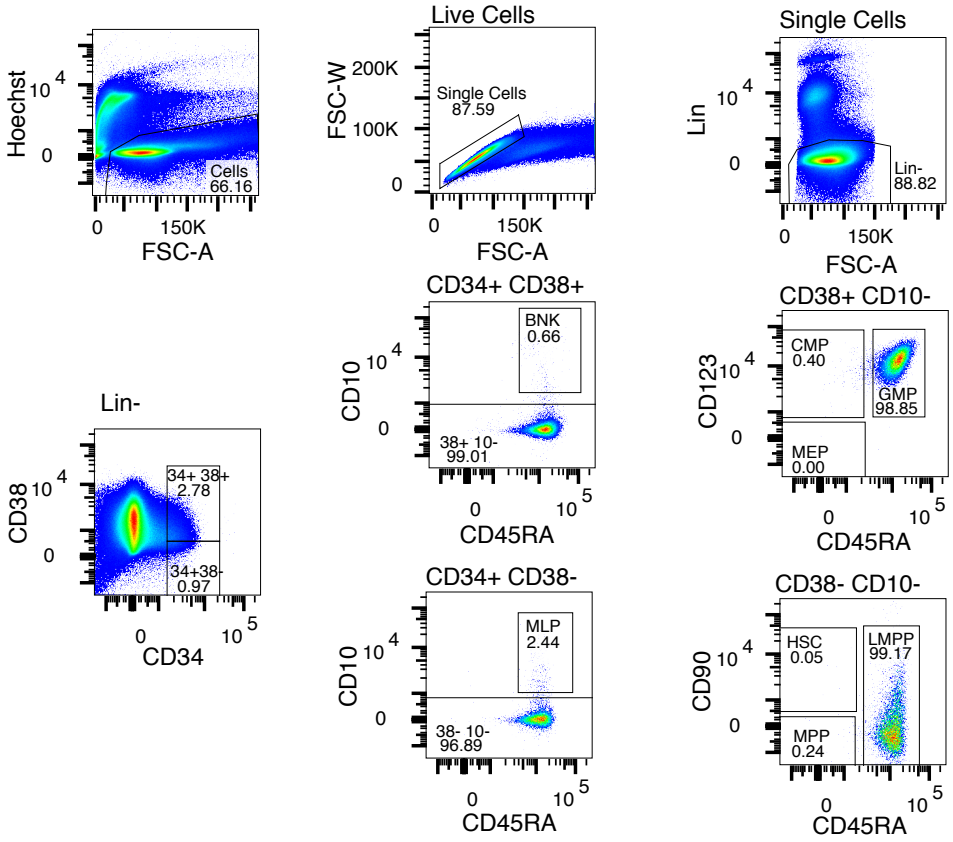
(a) Sample BEL039



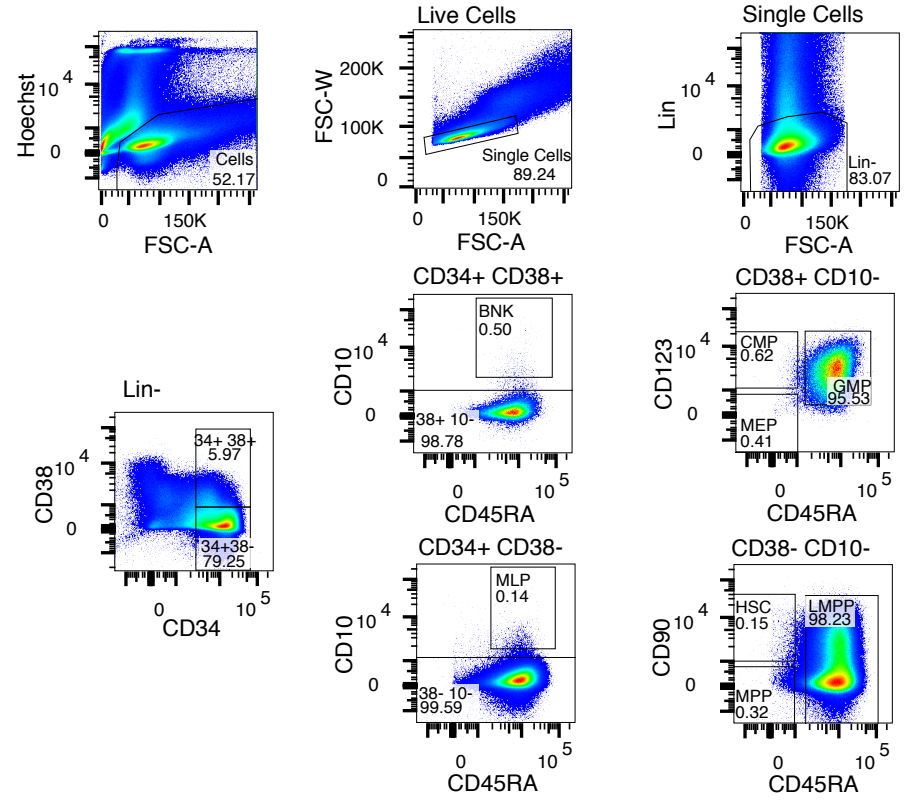
(b) Sample BEL048



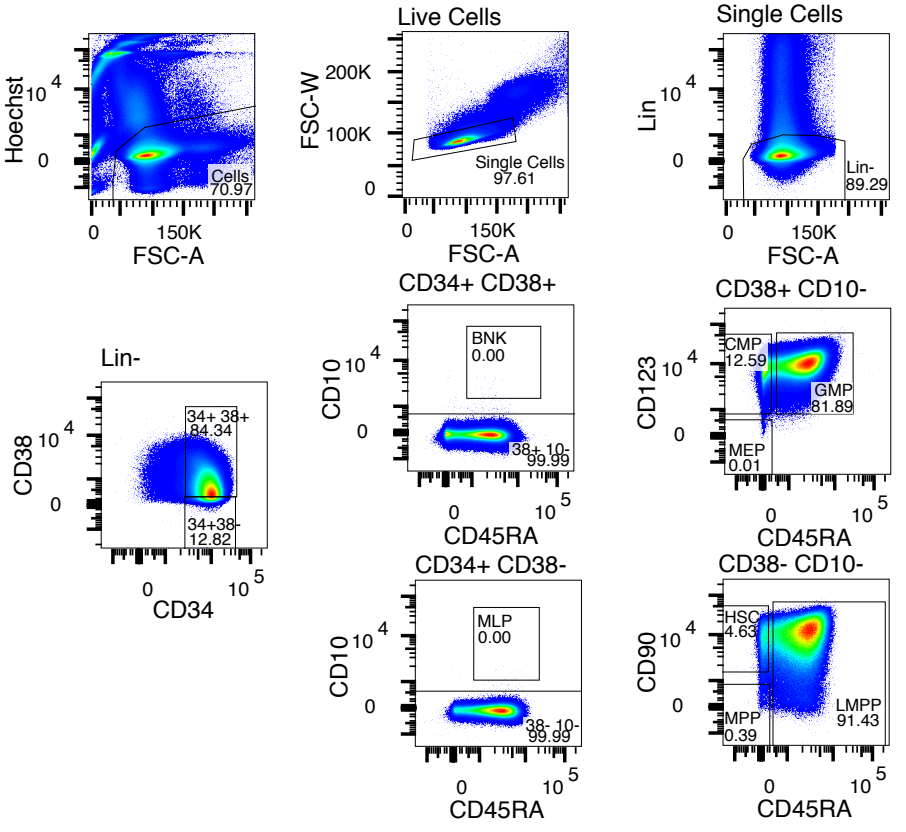
(c) Sample BIR344



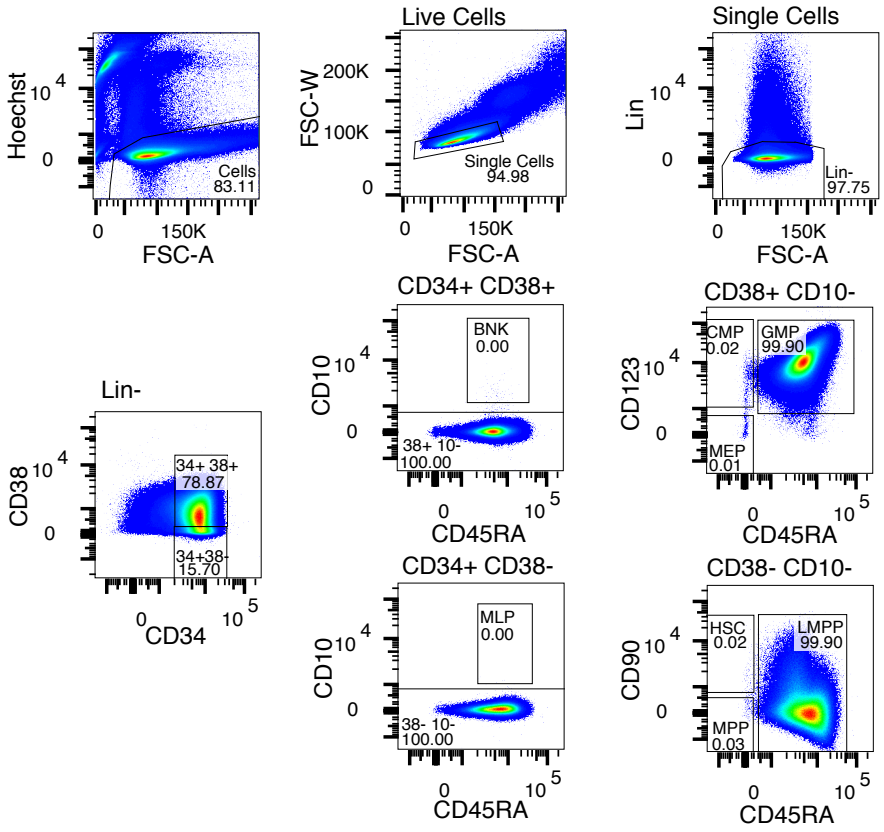
(d) Sample BIR352



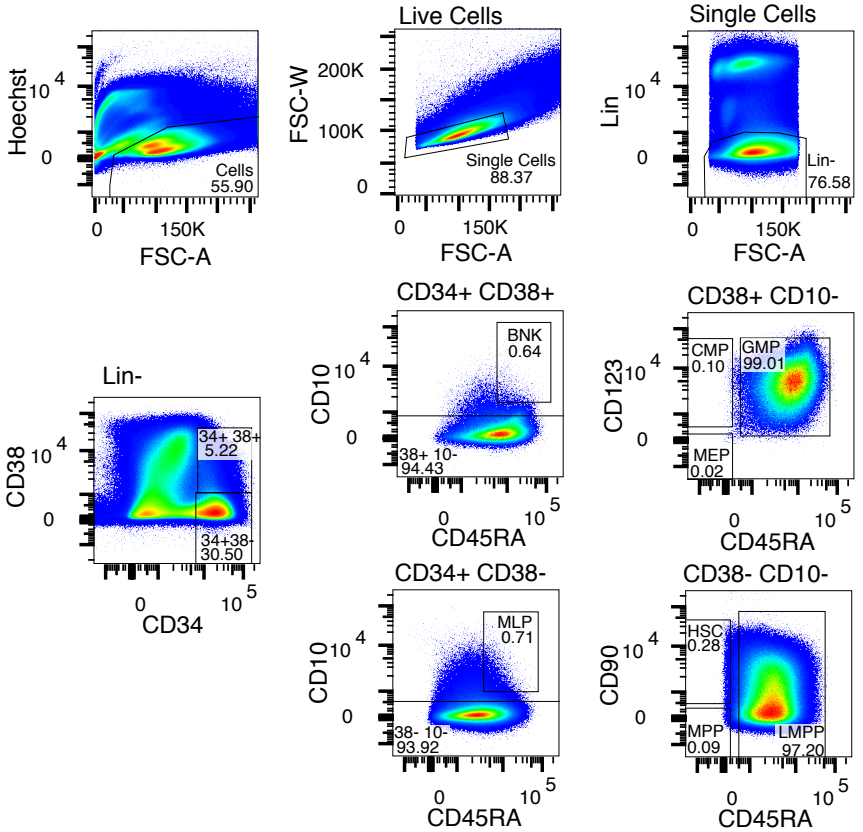
(e) Sample BIR517



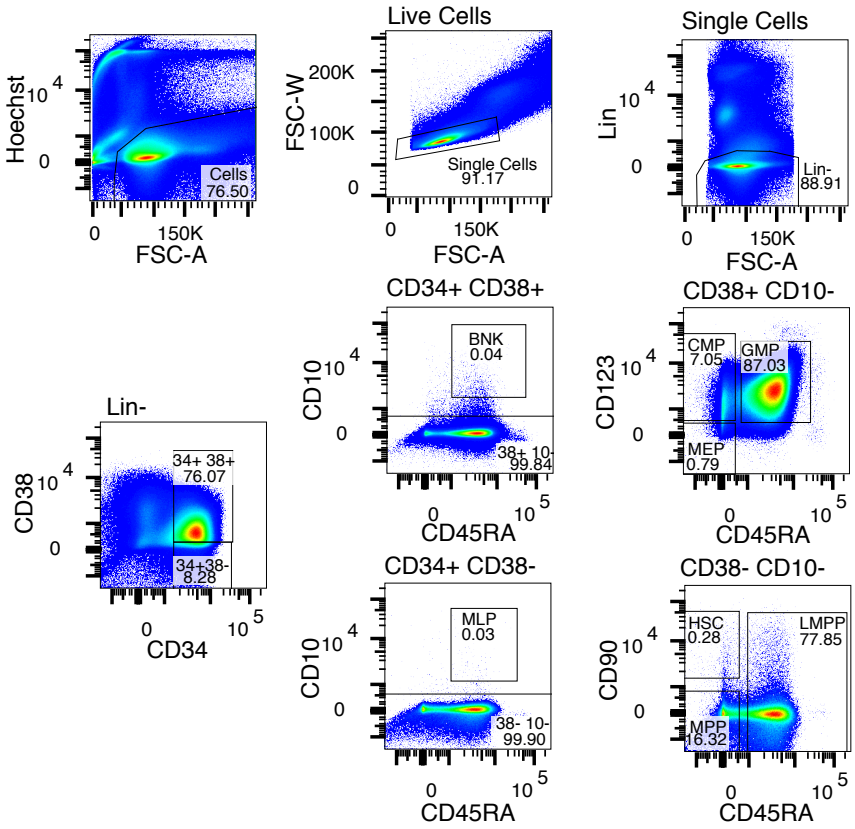
(f) Sample BRI1371



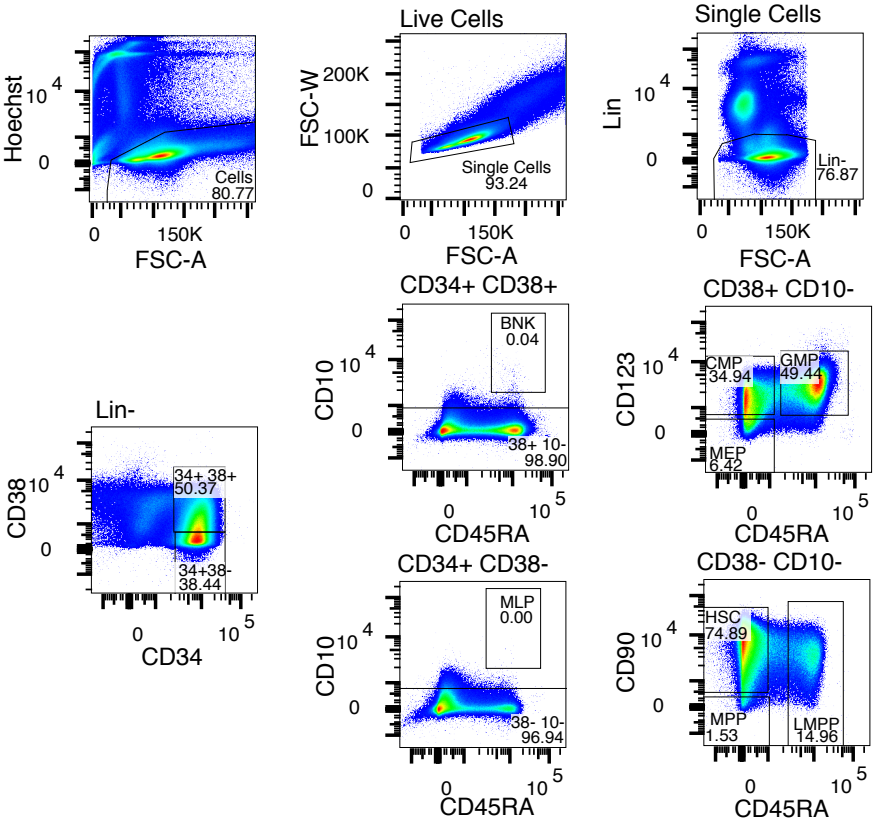
(g) Sample BRI1481



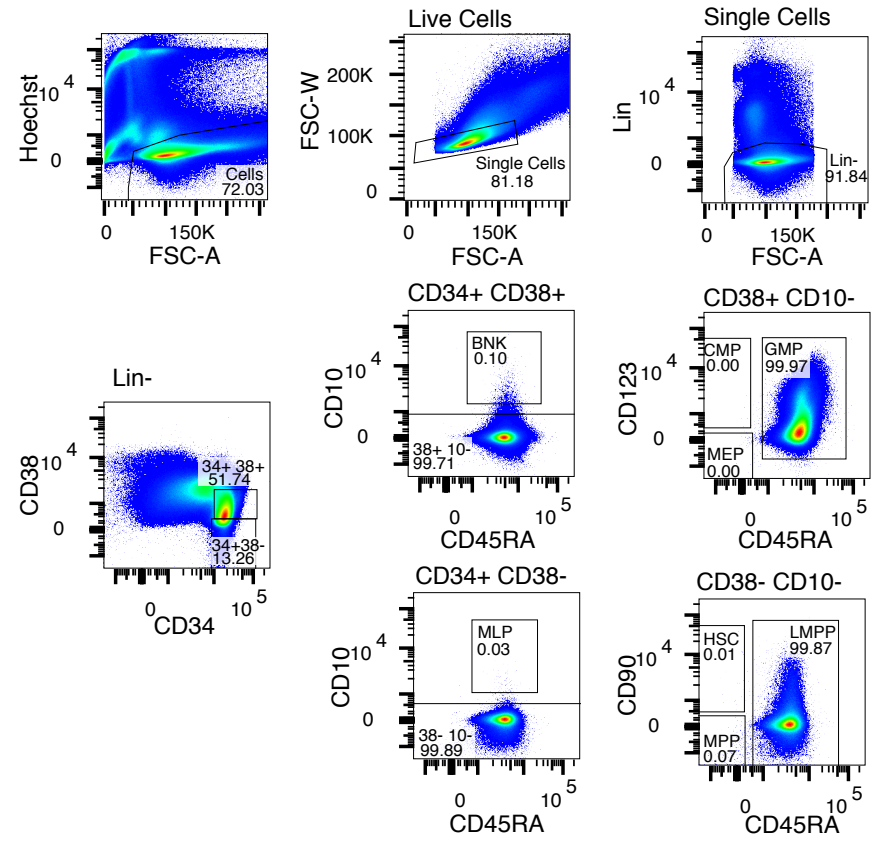
(h) Sample BRI1492



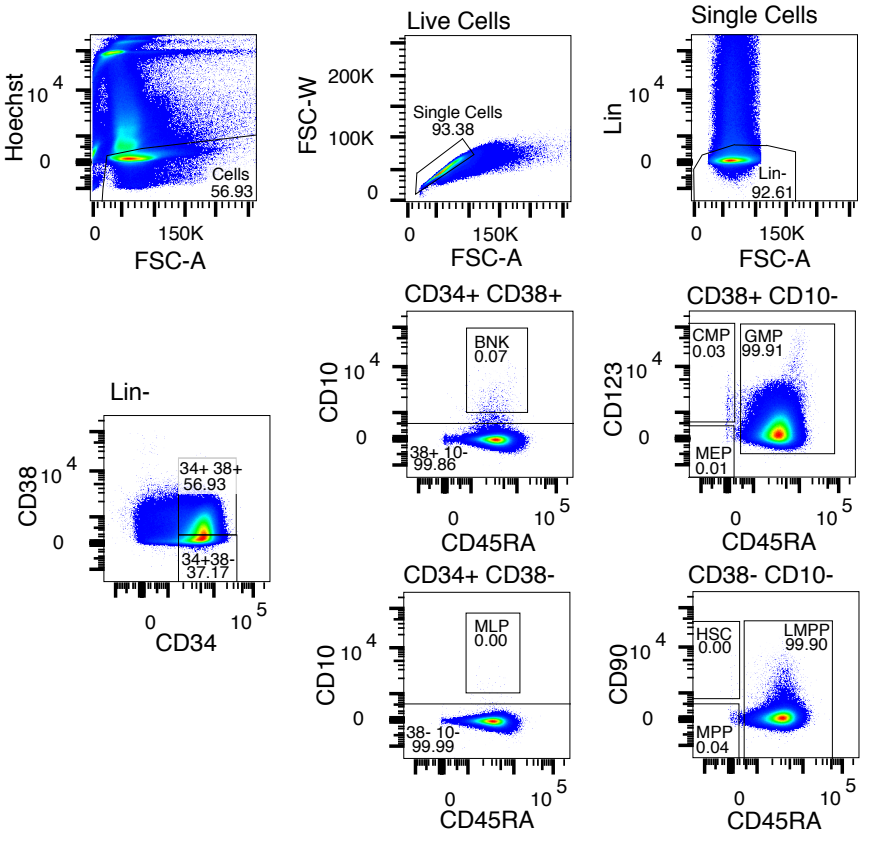
(i) Sample BRI1525



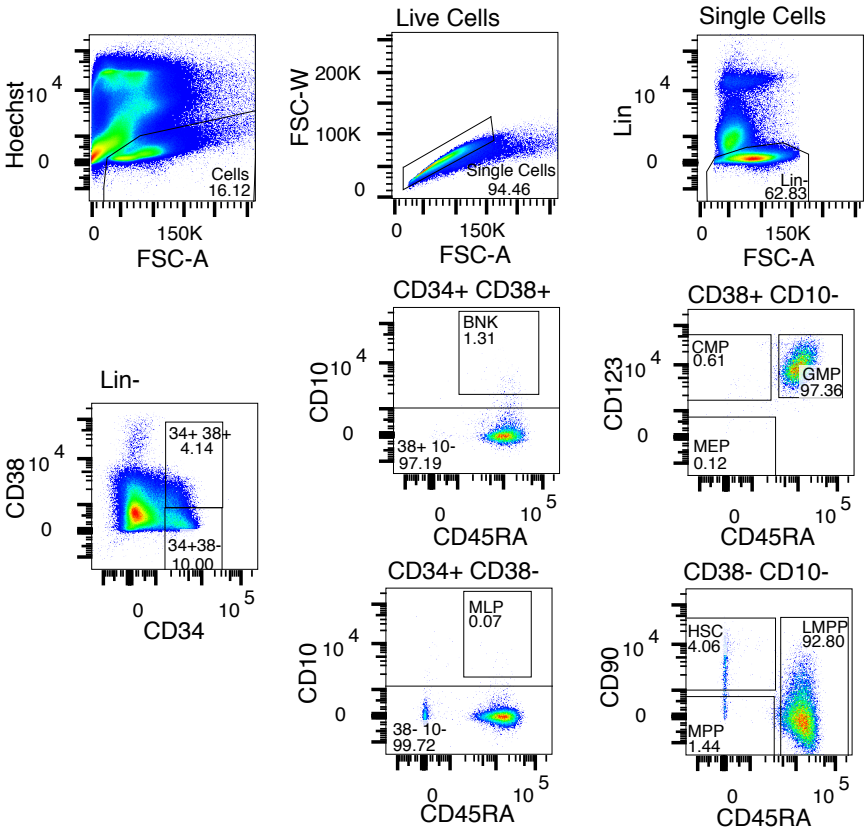
(j) Sample CAR117



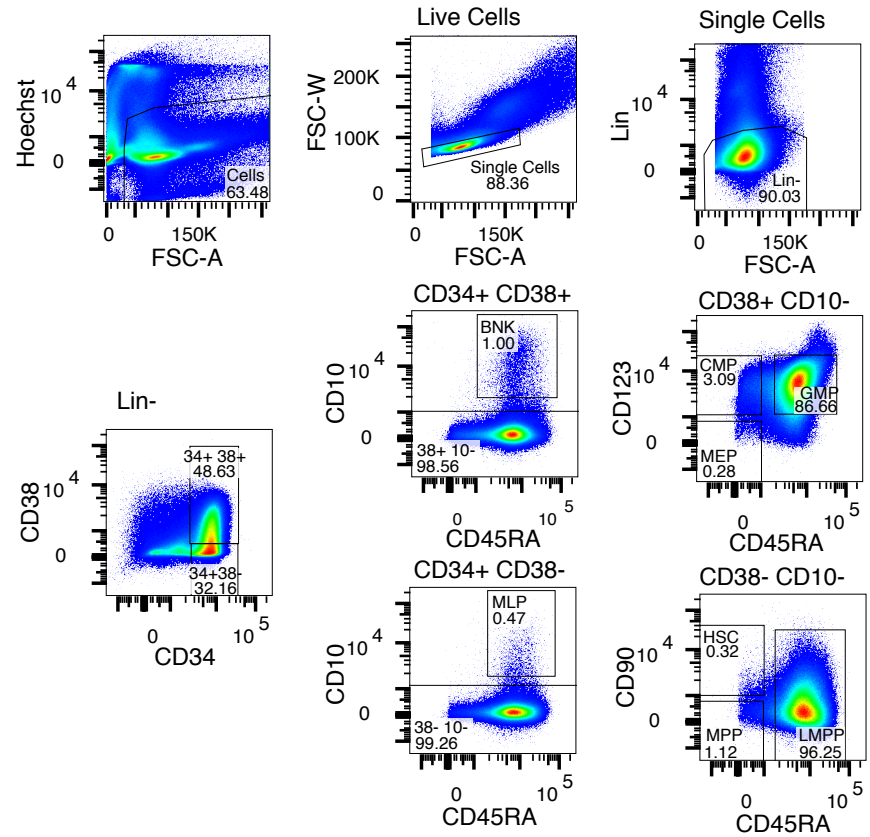
(k) Sample CAR143



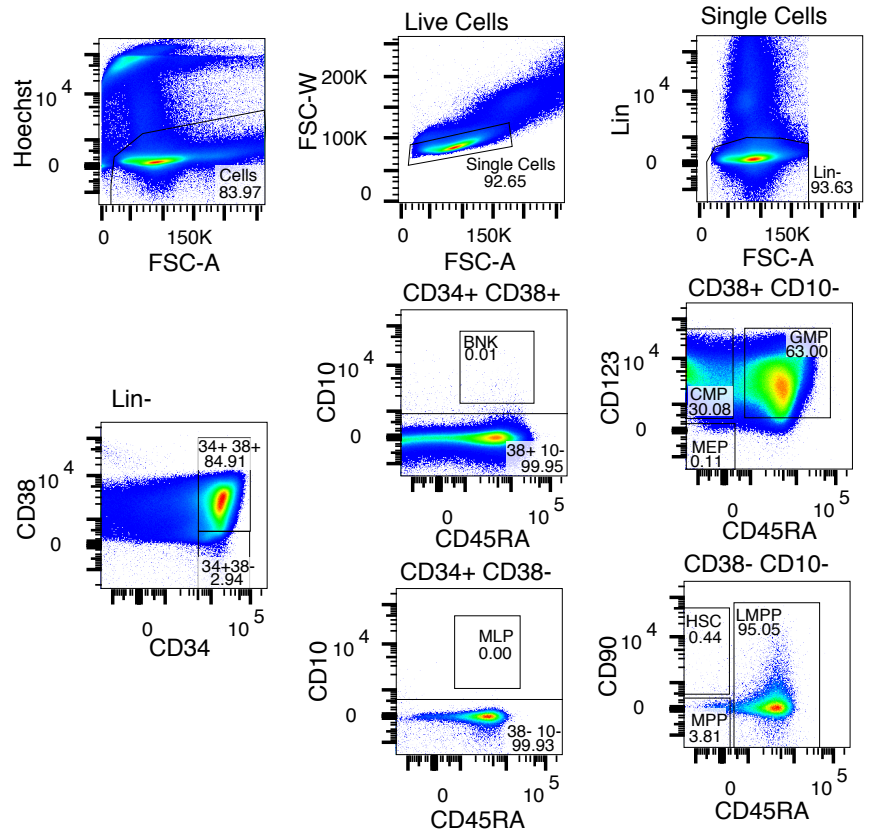
(l) Sample CUH018



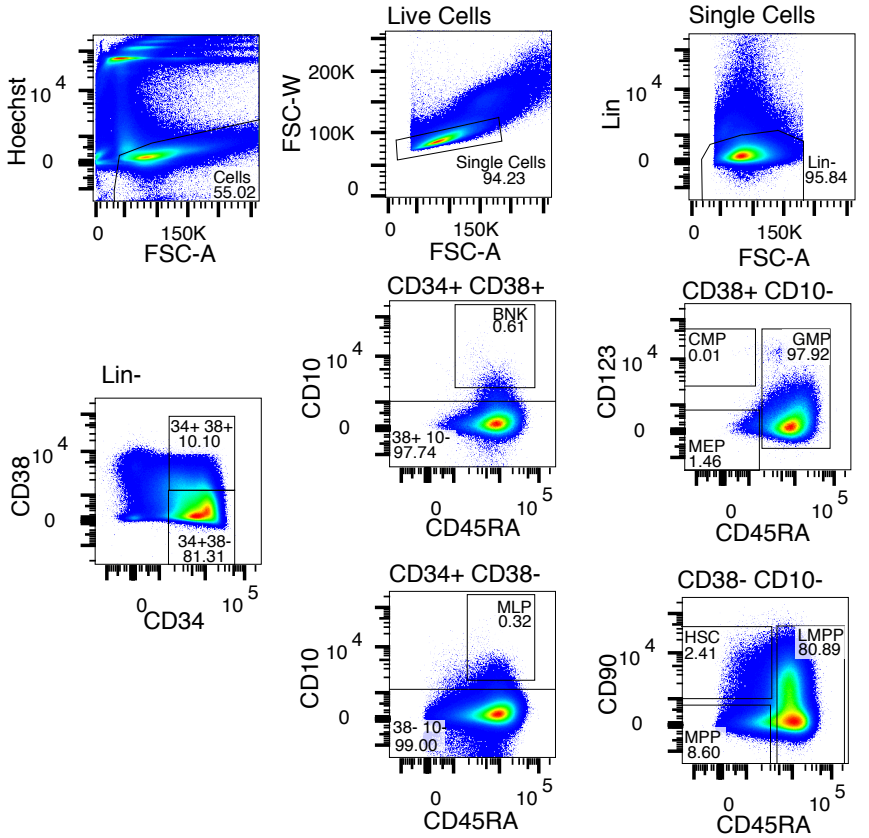
(m) Sample GAL032



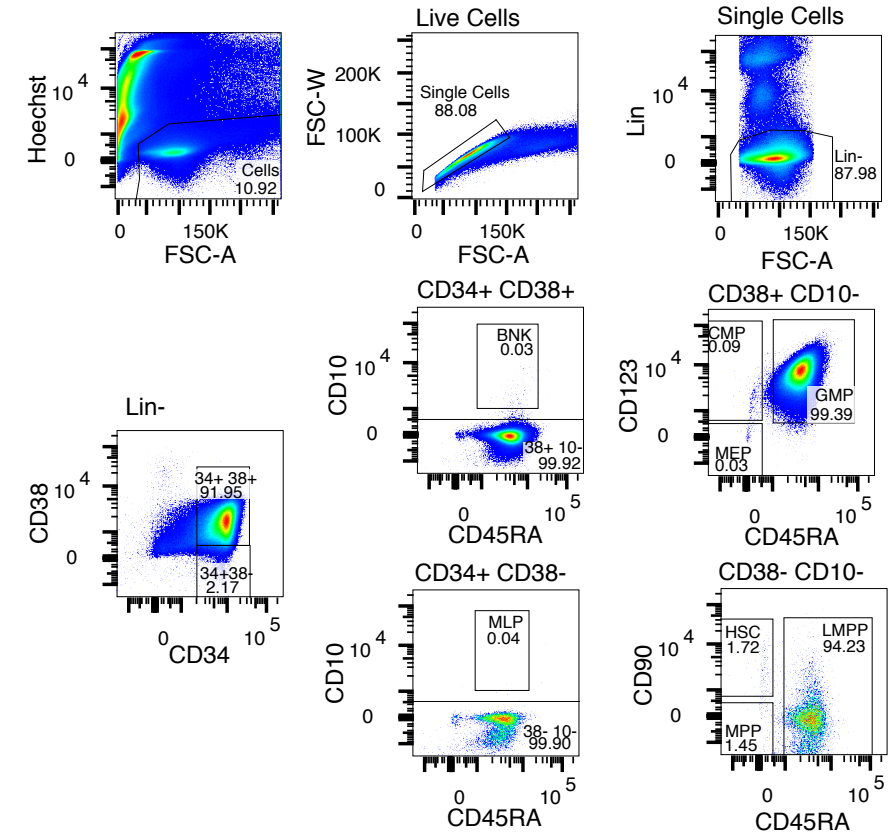
(n) Sample HAM061



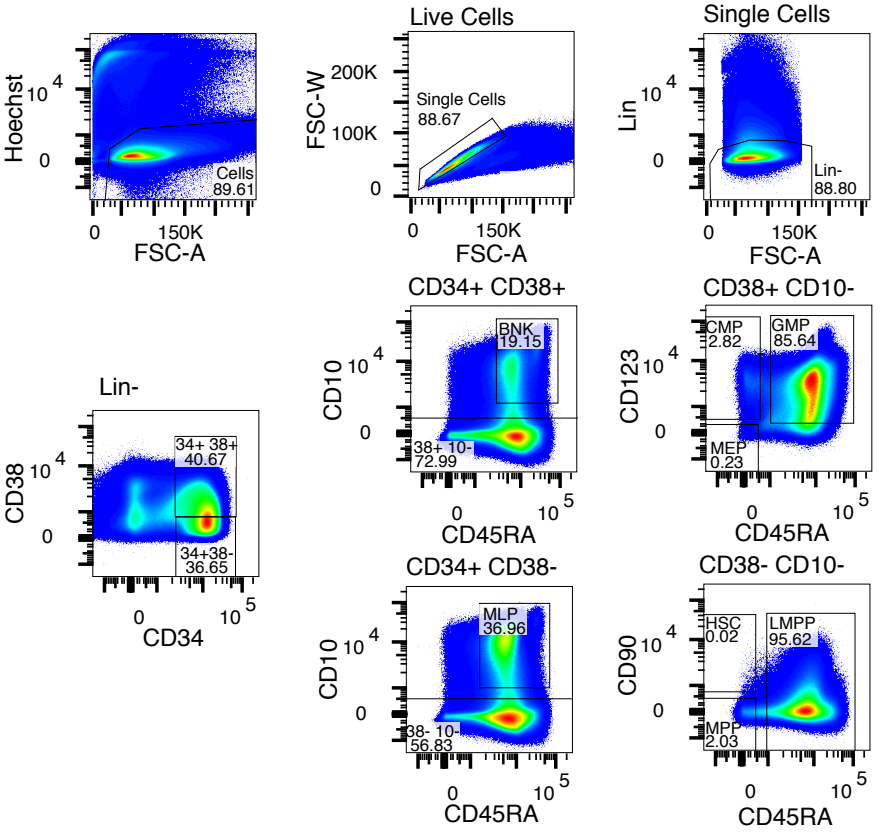
(o) Sample HER103



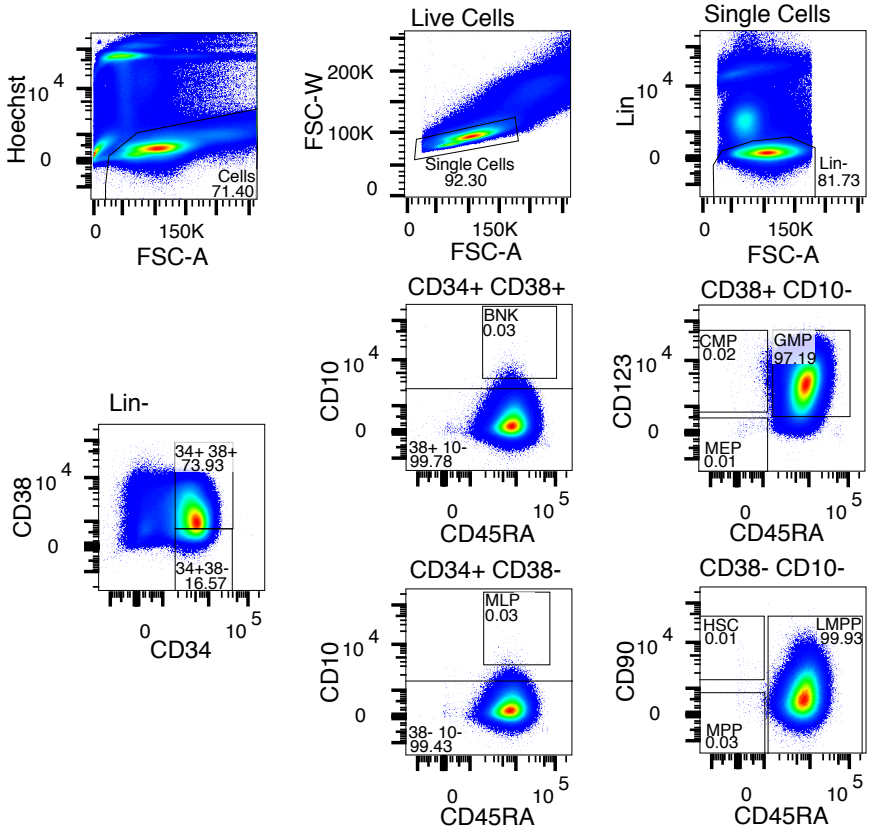
(p) Sample OX2116



(q) Sample STB128



(r) Sample STB257



(s) Sample STB291

Figure A.2: Immunophenotype of AML BM. HSPC immunophenotype of AML BM samples showing expanded HSPC populations. Frequencies indicate the frequency of the parent gate.

A.9 Genetic and Cytogenetic Profiles of AML Samples

Table A.19: Genetic Profile of AML Samples. Genetic profile of AML samples in genes associated with myeloid malignancies, as detected by targeted re-sequencing of genes. wt: wild-type. ITD1: E608delinsDDFREYEYDLKWEFPRE. ITD2: +2CCACGGGGAATATGAATATGATCTCAAATGGGAGTTTCCAAGAGAAAATTTAGAGTTTGGR132HV617F. ITD3: N587delinsNEYFYVDFREDTLPN. ITD4: L576delinsLQMVQVTGSSDNEYFYVDFREYESQL. ITD5: E608delinsYEYDLKWEFPRE.

Sample	**ASXL1**	**BCOR**	**CBL**	**CDKN2A**	**ETV6**	**FLT3**	**IDH1**	**JAK2**	**KIT**	**KRAS**	**MAU2**	**NIPBL**	**NRAS**
BEL039	wt	wt	wt	wt	wt	wt	R132C	wt	wt	wt	wt	wt	wt
BEL048	wt	wt	wt	wt	wt	ITD1	wt	wt	wt	G12D	wt	wt	G12V
BIR344	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	G12D
BIR352	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt
BIR517	wt	p.E1164fs	wt	wt	L398R	wt	wt	V617F	wt	wt	wt	wt	wt
BRI1371	S878fs	wt	wt	wt	wt	wt	wt	wt	wt	L113P	wt	wt	wt
BRI1481	wt	wt	wt	wt	wt	ITD2	R132H	V617F	wt	wt	wt	wt	wt
BRI1492	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt
BRI1525	wt	wt	wt	wt	wt	A680V	wt	V617F	wt	wt	wt	wt	wt
CAR117	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt
CAR143	G680fs	wt	wt	wt	wt	wt	wt	V617F	wt	wt	A610T K247E	G1861S	Y64N
CUH018	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt
GLA032	wt	wt	wt	wt	wt	ITD3	wt	wt	wt	wt	wt	wt	wt
HAM061	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt
HER103	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt
OX2116	wt	wt	Q345E	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt
STB128	wt	wt	wt	wt	wt	V592A	R132H	wt	wt	wt	wt	wt	wt
STB257	wt	wt	wt	wt	wt	ITD4	wt	wt	wt	wt	wt	wt	wt
STB291	wt	wt	wt	R29W	wt	ITD5	wt	wt	wt	wt	wt	wt	Q61K
Sample	**PDS5B**	**PHF6**	**RAD21**	**RUNX1**	**SMC3**	**SRSF2**	**STAG1**	**STAG2**	**TET2**	**TP53**	**U2AF1**	**WT1**	
BEL039	wt	wt	wt	wt	wt	wt	wt	wt	S825X A950fs H949Q	wt	S34F	wt	
BEL048	wt	wt	wt	R139Q	wt	wt	wt	wt	wt	wt	R156H	wt	
BIR344	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	
BIR352	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	
BIR517	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	
BRI1371	wt	wt	wt	D198G	wt	wt	wt	wt	P555S	wt	R83H	wt	
BRI1481	wt	wt	wt	P277fs R376C	wt	wt	wt	wt	wt	wt	wt	wt	
BRI1492	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	
BRI1525	R1269S	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	
CAR117	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	
CAR143	wt	wt	wt	P398L A311fs	wt	K52E	wt	E1201K	wt	wt	wt	wt	
CUH018	wt	wt	wt	S199delins RN R139Q	wt	wt	wt	wt	+1 T S1848fs	wt	wt	wt	
GLA032	wt	G348R	V7A N589S M87T	wt	G947E	wt	I148V	L460F	wt	wt	S34F	wt	
HAM061	wt	wt	wt	R174X	wt	wt	wt	wt	Q1702fs	wt	wt	wt	
HER103	wt	wt	wt	V170fs R178G R180fs	wt	wt	wt	wt	wt	wt	wt	wt	
OX2116	wt	wt	wt	wt	wt	wt	wt	wt	L1151fs	Y31H	wt	S364X	

Continuation of Table A.19

Sample	ASXL1	BCOR	CBL	CDKN2A	ETV6	FLT3	IDH1	JAK2	KIT	KRAS	MAU2	NIPBL	NRAS
STB128	wt	wt	wt	D96fs	wt	wt	wt	wt	wt	wt	wt	wt	
STB257	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	
STB291	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	wt	

Table A.20: Cytogenetic Profile of AML Samples. Cytogenetic profile of AML samples, as detected by a Baseline Cytogenetic Assessment (if performed). ND: not determined.

Sample	Cytogenetic Profile/Risk
BEL039	ND
BEL048	ND
BIR344	ND
BIR352	ND
BIR517	ND
BRI1371	ND
BRI1481	47,XY,+13[2]/46,XY[3]
BRI1492	ND
BRI1525	ND
CAR117	Poor risk
CAR143	Intermediate risk
CUH018	Intermediate risk
GLA032	Intermediate risk
HAM061	ND
HER103	Poor risk
OX2116	47, XXadd(5)(q2?1)add5(q3?5), del(7)(q22),-17, -21, +22, +mar1, +mar2, +mar3[cp2]/49,sl,add(11)(q2?2),-13,-16,+mar3x4[cp2]/48, sdl, -X, -4, +mar4[cp6]
STB128	ND
STB257	Poor risk
STB291	Intermediate risk

Table A.21: AML Baseline Cytogenetic Assessment Key. Risk categories for AML cytogenetics as determined by a Baseline Cytogenetic Assessment.

Risk	Cytogenetic Profile
Favourable Risk	t(8;21)(q22;q22) (Irrespective of additional cytogenetic abnormalities) t(8;21)(q22;q22) (Irrespective of additional cytogenetic abnormalities)
Intermediate Risk	Entities not classified as favourable or adverse
Poor Risk	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) (excluding cases with favourable karyotype, all adverse-risk abnormalities) 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)

A.10 Summary of Xenotransplantation of AML Experiments

Table A.22: Summary of Xenotransplantation of AML Experiments. A summary of all AML xenotransplantation experiments. Engraftment is shown as a percentage of live events, as determined by flow cytometry. To be considered positive for human engraftment, a cut-off of 0.1% engraftment was used. 12 and 16 week time-points represent PB samples and 20, 24 and terminal time-points represent BM samples. Terminal BM samples were only harvested from engrafted mice. NM: no marking, L: left, R: right, LR: left-right. NA: not applicable.

Sample	Status	Ear Clip	Population Injected	Number of Cells Injected	~12wks		~16wks		~20wks		~24wks		Terminal	
					Myeloid Engraftment	Lymphoid Engraftment	Myeloid Engraftment	Lymphoid Engraftment	Myeloid Engraftment	Lymphoid Engraftment	Myeloid Engraftment	Lymphoid Engraftment	Myeloid Engraftment	Lymphoid Engraftment
STB291	No Human Engraftment	LR	LMPP	10,720	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	NA	NA
		1R	GMP	95,000	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	NA	NA
		1L	GMP	95,000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		NM	CD34-	6,320	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	NA	NA
HAM061	No Human Engraftment	1R	LMPP	128,864	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		1L	LMPP	128,864	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		LR	LMPP	128,864	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		2R	GMP	165,640	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		2L	GMP	165,640	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		NM	CD34-	71,757	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
BIR517	Myeloid Engraftment	NM	LMPP	198,195	3.79	0.00	29.78	0.00	NA	NA	NA	NA	89.41	0.02
		1L	LMPP	198,195	4.18	0.02	41.66	0.00	NA	NA	NA	NA	93.33	0.01
		LR	LMPP	198,195	4.15	0.00	28.03	0.00	NA	NA	NA	NA	92.34	0.02
		2R	LMPP	198,195	5.38	0.02	56.53	0.01	NA	NA	NA	NA	89.95	0.02
		1R	CD34-	54,814	0.00	0.00	0.00	0.00	NA	NA	NA	NA	0.00	0.05
GLA032	No Human Engraftment	NM	LMPP	3,845	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		R	CD34-	3,800	0.00	0.00	0.00	0.00	0.02	0.00	0.00	0.02	NA	NA
		L	CD34-	64,000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
OX2116	LM Engraftment	NM	GMP	5,746	0.00	0.00	0.00	0.00	0.00	0.13	0.00	0.00	0.02	0.17
		1R	LMPP	5,746	0.00	0.00	0.00	0.31	0.10	44.05	1.48	49.64	4.15	46.19
		1L	LMPP	447,101	0.00	0.09	0.00	1.26	0.03	65.12	0.12	57.35	NA	NA
		LR	LMPP	447,101	0.00	7.79	0.09	35.47	0.04	64.33	0.08	87.42	0.53	73.60
STB291	Myeloid Engraftment	LR	LMPP	125,330	0.89	0.00	12.69	0.00	96.69	0.00	NA	NA	96.08	0.01
		2R	LMPP	125,330	0.29	0.00	3.30	0.00	86.75	0.00	NA	NA	61.37	0.05
		1L	LMPP	125,330	0.21	0.00	5.45	0.00	18.05	0.02	NA	NA	64.20	0.07
		1R	GMP	125,330	0.04	0.00	0.54	0.00	90.63	0.01	NA	NA	60.02	0.10
		NM	GMP	222,493	0.75	0.00	12.28	0.00	96.59	0.00	NA	NA	91.19	0.07
BEL048	No Human Engraftment	NM	LMPP	62,750	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	NA	NA
		LR	LMPP	62,750	0.00	0.00	0.06	0.14	0.20	0.02	0.00	0.00	NA	NA
		1L	GMP	71,927	0.00	0.00	0.00	0.00	0.01	0.01	0.16	0.00	NA	NA
		R	CD34-	21,633	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	NA	NA
BEL039	LM Engraftment	NM	GMP	120,486	0.09	0.04	0.00	0.06	0.00	0.00	0.00	0.00	0.00	0.16
		L	GMP	120,486	0.03	0.03	0.00	0.00	0.00	0.00	0.00	0.06	0.00	0.22
		R	LMPP	145,844	0.00	0.00	0.02	0.05	0.06	4.61	2.46	17.86	0.47	5.24
		LR	LMPP	145,844	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.05	0.01	0.19
		LL	LMPP	145,844	0.02	0.02	0.00	0.06	0.01	1.84	0.04	5.05	0.15	3.24

Continuation of Table A.22

Sample	Status	Ear Clip	Population Injected	Number of Cells Injected	~12wks		~16wks		~20wks		~24wks		Terminal	
					Myeloid Engraftment	Lymphoid Engraftment	Myeloid Engraftment	Lymphoid Engraftment	Myeloid Engraftment	Lymphoid Engraftment	Myeloid Engraftment	Lymphoid Engraftment	Myeloid Engraftment	Lymphoid Engraftment
CUH018	No Human Engraftment	R	LMPP	200,000	0.00	0.05	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		L	LMPP	148,630	0.06	0.00	0.02	0.00	0.00	0.00	0.00	0.44	NA	NA
		NM	GMP	206,178	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	NA	NA
STB257	LM Engraftment	NM	GMP	717,449	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
		L	GMP	717,449	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
		R	LMPP	700,000	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
		LR	LMPP	1,067,255	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
		NM	LMPP	1,067,255	2.87	3.01	10.95	4.08	23.72	18.34	44.14	7.67	55.95	4.11
		L	MLP	700,000	0.35	0.01	23.50	1.84	20.70	21.65	23.72	16.94	51.93	7.16
		R	MLP	572,210	12.94	1.17	4.28	1.21	15.51	28.12	39.04	19.74	52.17	6.60
		LR	MLP	572,210	2.35	0.29	6.09	1.08	16.53	27.02	22.41	27.85	58.00	6.97
BIR352	Myeloid Engraftment	LR	LMPP	5,714	0.00	0.00	0.00	0.00	1.24	0.00	52.71	0.01	3.69	0.04
		2R	LMPP	134,863	0.09	0.09	0.00	0.00	0.00	0.00	0.04	0.00	NA	NA
		2L	MLP	5,714	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		NM	GMP	11,105	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	NA	NA
		L	CD34-	5,714	0.05	0.02	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		R	CD34-	6,415,206	0.05	0.00	0.00	0.00	0.00	0.00	0.02	0.00	NA	NA
STB257	LM Engraftment	NM	LMPP	289,842	5.59	7.69	8.07	4.33	18.94	27.24	NA	NA	38.49	4.92
		2L	LMPP	515,000	6.29	8.68	9.86	5.69	38.05	12.80	NA	NA	34.84	10.34
		L	MLP	515,103	11.56	1.04	7.52	1.07	25.20	22.64	NA	NA	26.88	9.00
		R	GMP	207,216	7.23	1.67	9.74	0.76	31.69	7.72	NA	NA	36.56	8.65
		LR	GMP	515,000	3.63	10.85	2.74	4.03	10.54	31.00	NA	NA	20.35	7.85
BRI1481	No Human Engraftment	NM	LMPP	353,986	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	NA	NA
		L	LMPP	353,986	0.00	0.02	0.00	0.00	0.00	0.01	0.02	0.03	NA	NA
		R	GMP	350,000	0.07	0.03	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		LR	GMP	721,347	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		2L	GMP	721,347	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
BRI1525	No Human Engraftment	L	LMPP	173,395	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	NA	NA
		NM	GMP	170,000	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	NA	NA
		R	GMP	405,973	0.00	0.00	0.00	0.23	0.00	0.03	0.00	0.00	NA	NA
		LR	GMP	405,973	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	NA	NA
		2L	GMP	405,973	0.00	0.00	0.00	0.00	0.00	0.00	0.04	0.00	NA	NA
BRI1371	No Human Engraftment	LR	LMPP	247,312	0.00	0.00	0.00	0.00	0.01	0.01	0.00	0.00	NA	NA
		2R	LMPP	247,312	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		2L1R	GMP	240,000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		NM	GMP	321,384	0.00	0.00	0.00	0.00	0.01	0.03	0.03	0.00	NA	NA
		L	GMP	321,384	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		R	GMP	321,384	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
HER103	No Human Engraftment	LR	LMPP	243,445	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.01	NA	NA
		2R	LMPP	243,445	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		L	LMPP	243,445	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		2L	GMP	240,000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		NM	GMP	1,902,206	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		1L2R	GMP	1,902,206	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	NA	NA

Continuation of Table A.22

Sample	Status	Ear Clip	Population Injected	Number of Cells Injected	~12wks		~16wks		~20wks		~24wks		Terminal	
					Myeloid Engraftment	Lymphoid Engraftment	Myeloid Engraftment	Lymphoid Engraftment	Myeloid Engraftment	Lymphoid Engraftment	Myeloid Engraftment	Lymphoid Engraftment	Myeloid Engraftment	Lymphoid Engraftment
		R	GMP	1,902,206	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
CAR143	LM Engraftment	NM	LMPP	180,448	0.04	0.00	0.04	0.00	2.94	0.00	3.58	0.03	1.72	0.14
		L	LMPP	180,448	0.31	0.00	0.06	0.00	7.10	0.08	10.14	0.03	1.79	0.18
		R	GMP	173,564	0.04	0.00	0.00	0.00	0.35	0.00	15.38	0.05	1.06	0.11
		LR	GMP	173,564	0.18	0.06	0.01	0.00	2.28	0.00	16.27	0.02	3.10	0.17
BRI1492	No Human Engraftment	NM	LMPP	170,000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		L2R	LMPP	303,539	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		2L	GMP	174,112	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
CAR117	No Human Engraftment	NM	LMPP	73,812	0.01	0.04	0.00	0.27	0.00	0.00	0.00	0.00	NA	NA
		L	GMP	70,000	0.00	0.02	0.00	0.00	0.00	0.00	0.01	0.00	NA	NA
		R	GMP	130,588	0.00	0.10	0.00	0.00	0.00	0.00	0.00	0.13	NA	NA
BIR344	LM Engraftment	NM	LMPP	288,340	0.01	0.06	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		1L	MLP	43,000	22.48	0.18	92.80	1.83	NA	NA	NA	NA	79.12	3.41
		1R	LMPP	43,000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	NA	NA
		LR	GMP	43,000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.02	NA	NA
		2L	HSC	43,000	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	NA	NA
		NM	HSC	42,065	38.32	0.14	94.90	0.73	NA	NA	NA	NA	90.18	1.41
		1L	MLP	246,023	3.24	0.12	78.22	3.70	NA	NA	NA	NA	69.36	7.59
		1R	GMP	32,224	0.00	0.04	0.01	0.00	0.00	0.00	0.00	0.00	NA	NA
		LR	LMPP	288,340	49.13	0.16	89.71	0.67	NA	NA	NA	NA	88.69	0.73
		2L	MPP	43,048	18.29	0.00	95.21	0.09	NA	NA	NA	NA	92.55	0.15

A.11 Immunophenotype of Human Engraftment

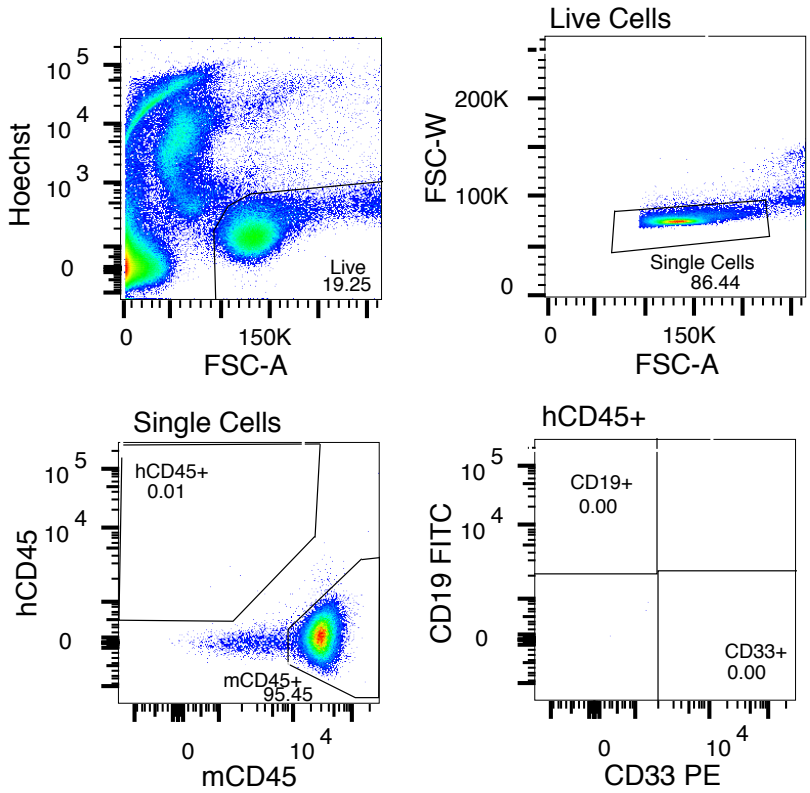
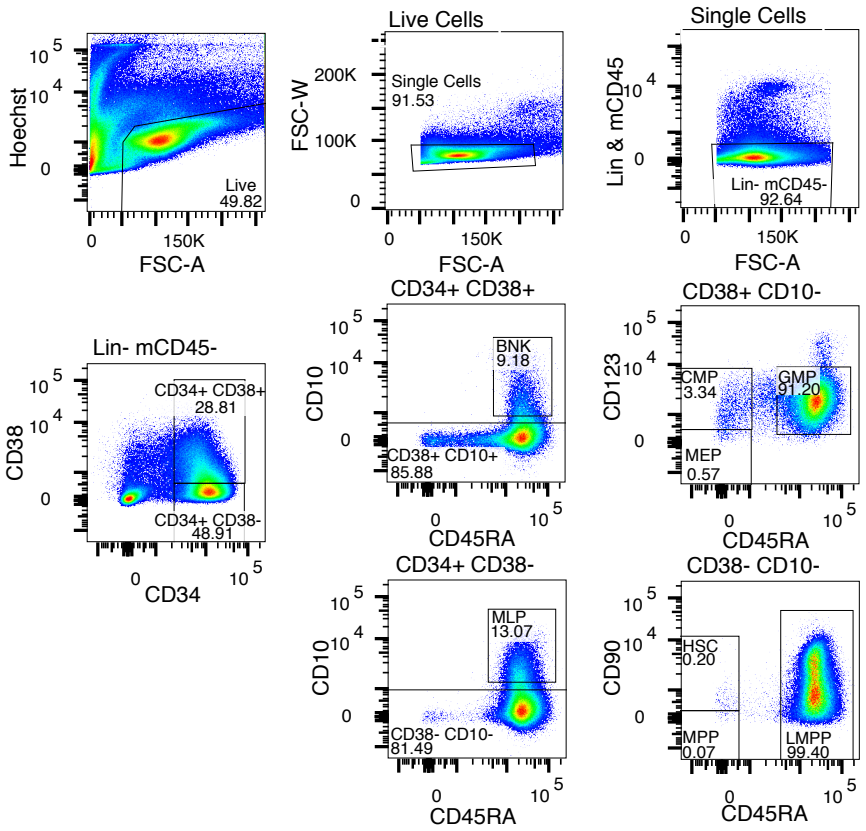
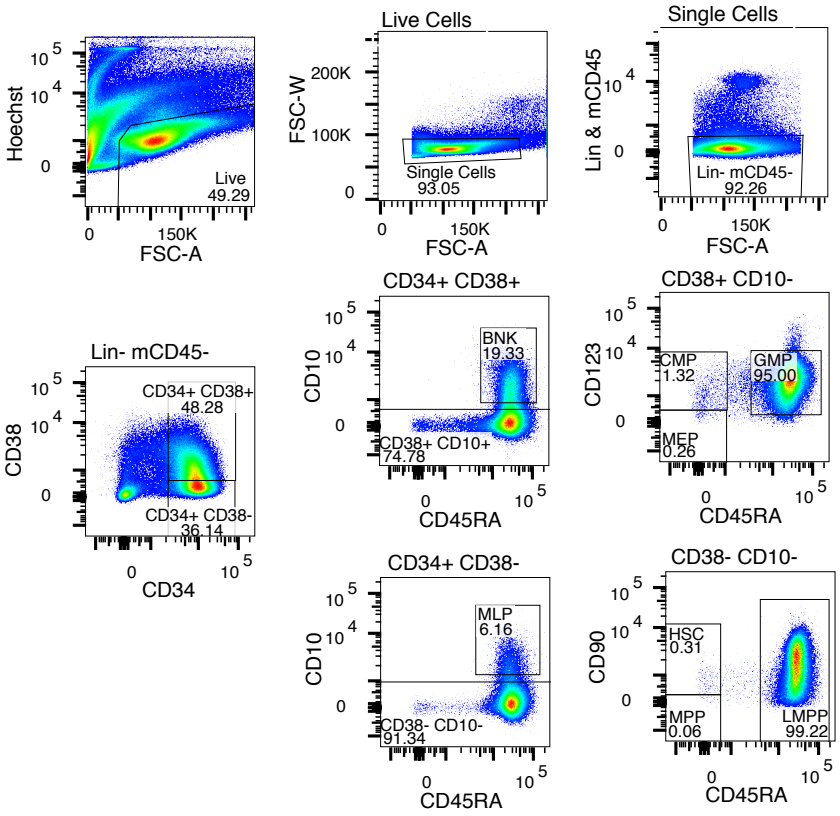


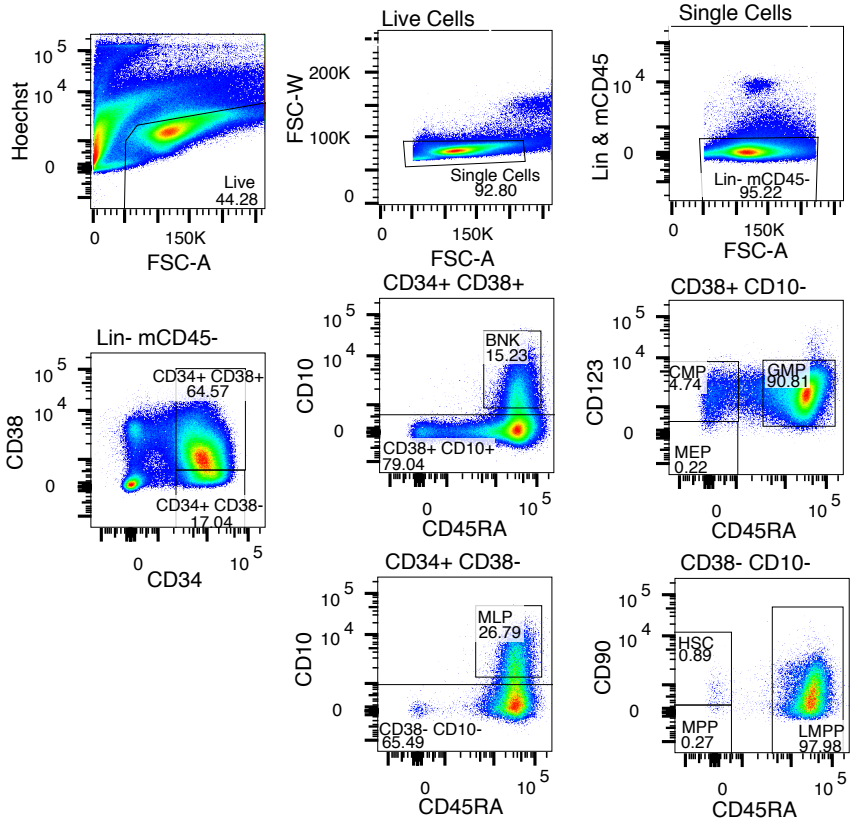
Figure A.3: Absence of Human Engraftment in Xenotransplantation Assays. Immunophenotype of the absence of human engraftment in xenotransplantation assays. Representative flow cytometry analysis of the BM from a mouse transplanted with AML sample STB128, ~24 weeks post-transplantation. Frequencies indicate the frequency of the parent gate.



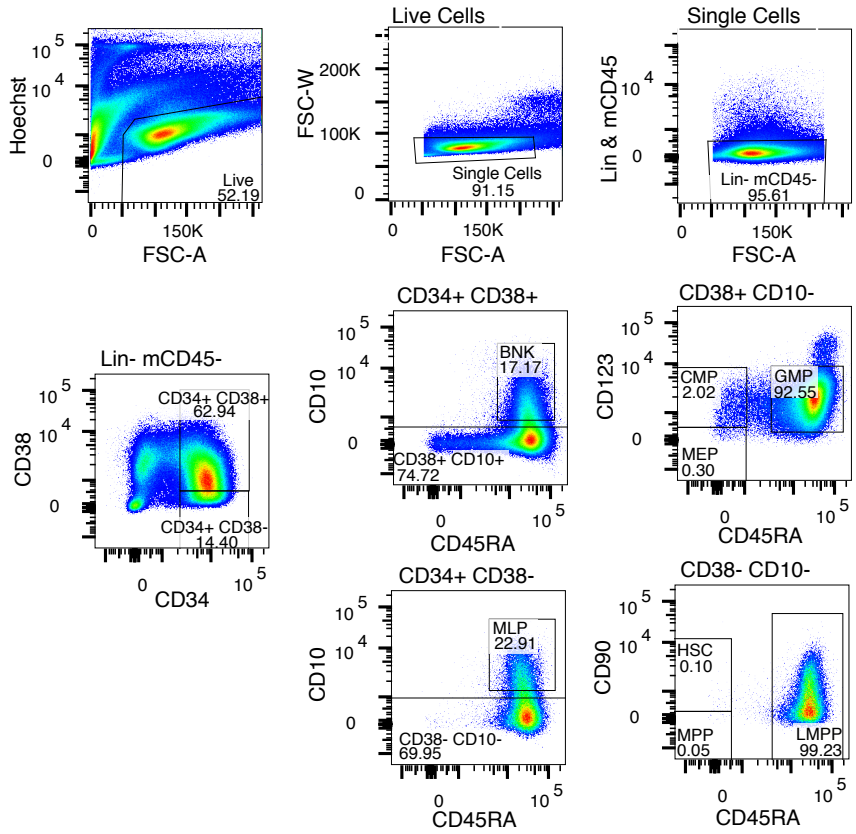
(a) Sample BIR344, MLP Injected



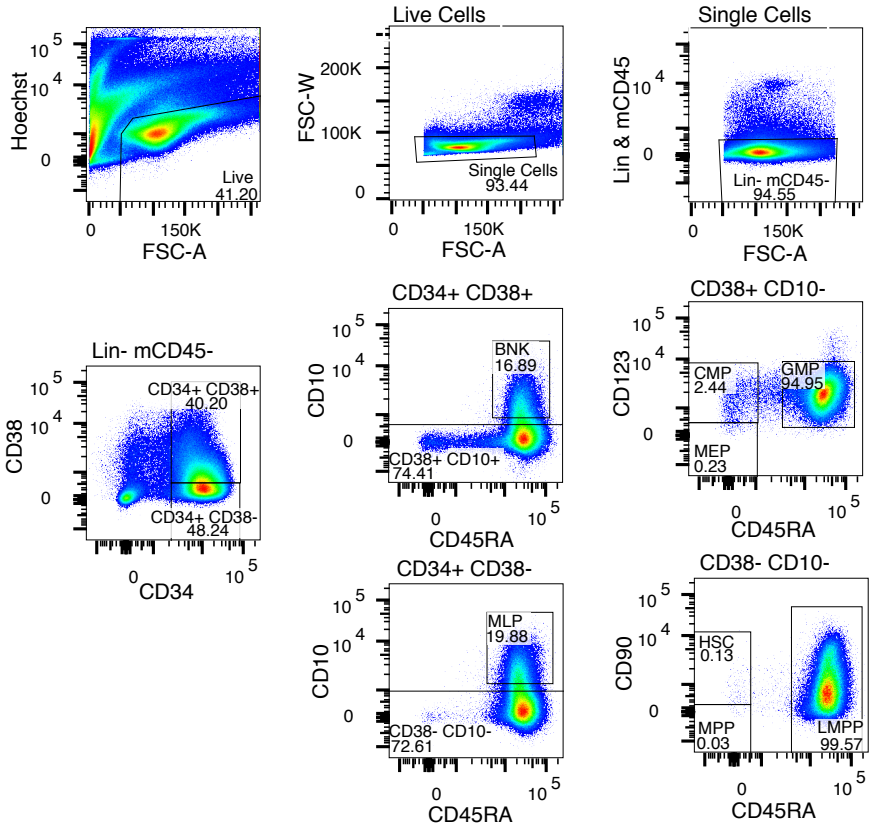
(b) Sample BIR344, MLP Injected



(c) Sample BIR344, MPP Injected

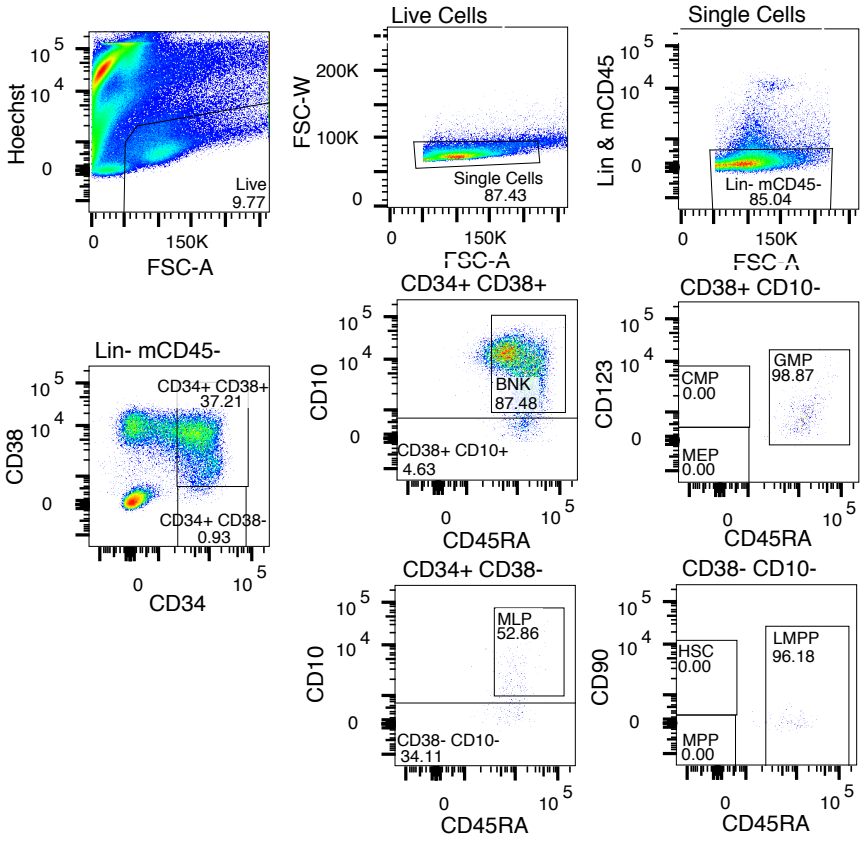


(d) Sample BIR344, LMPP Injected

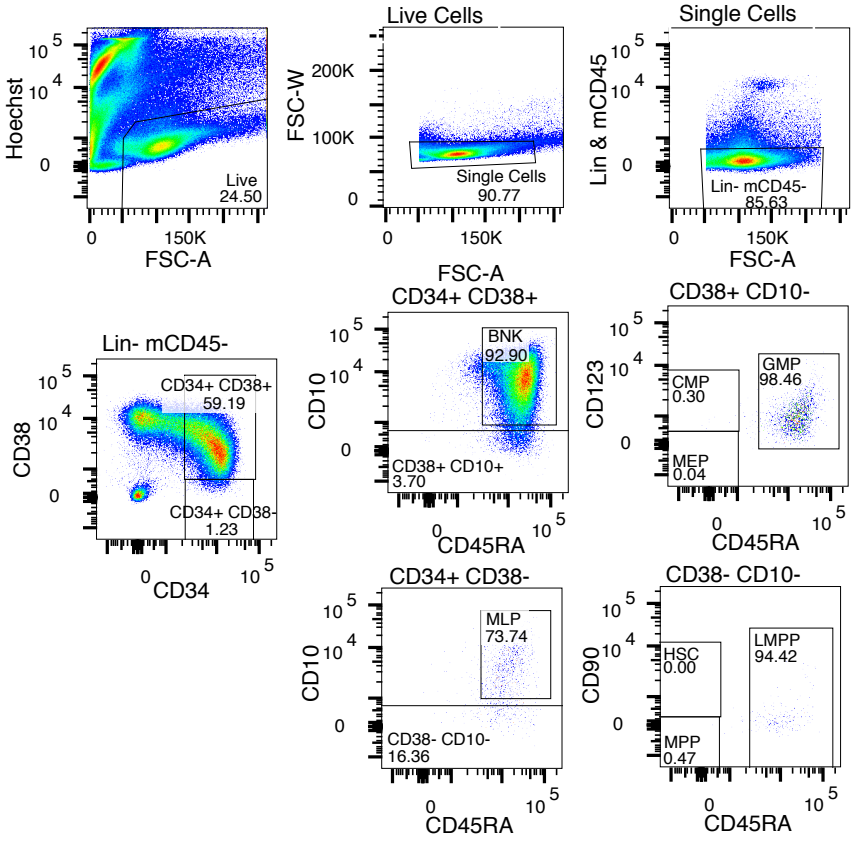


(e) Sample BIR344, HSC Injected

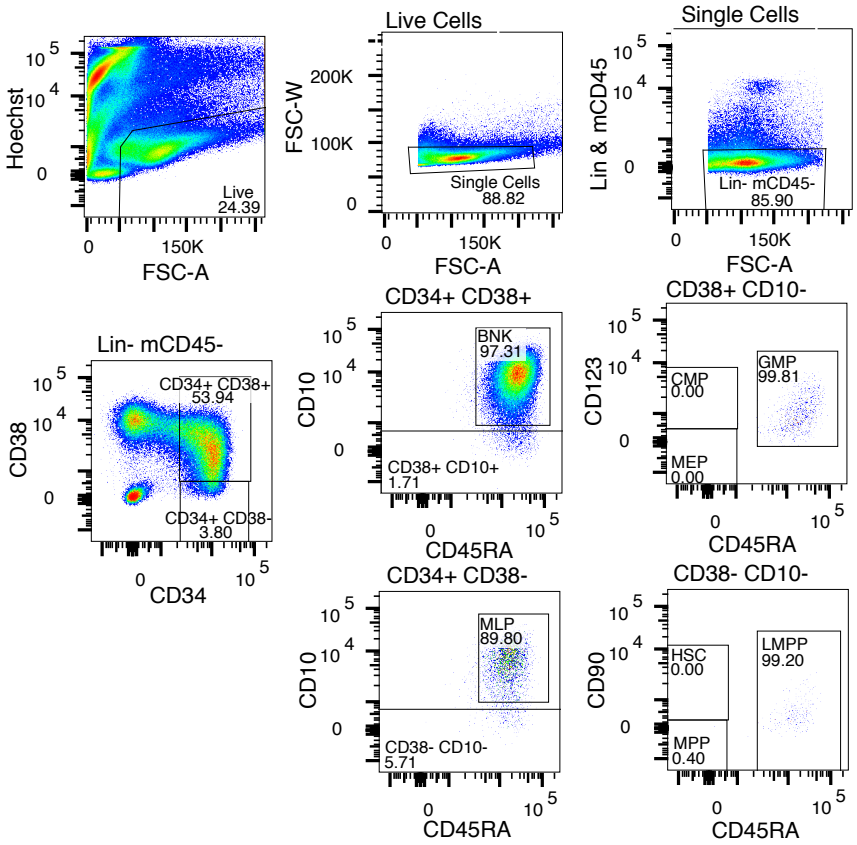
Figure A.4: Post-Engraftment HSPC Immunophenotype of BIR344. Post-engraftment HSPC immunophenotype of sample BIR344 in xenotransplantation assays. The injected population is indicated below plots. HSPC immunophenotyping was carried out on terminal BM samples. Frequencies indicate the frequency of the parent gate.



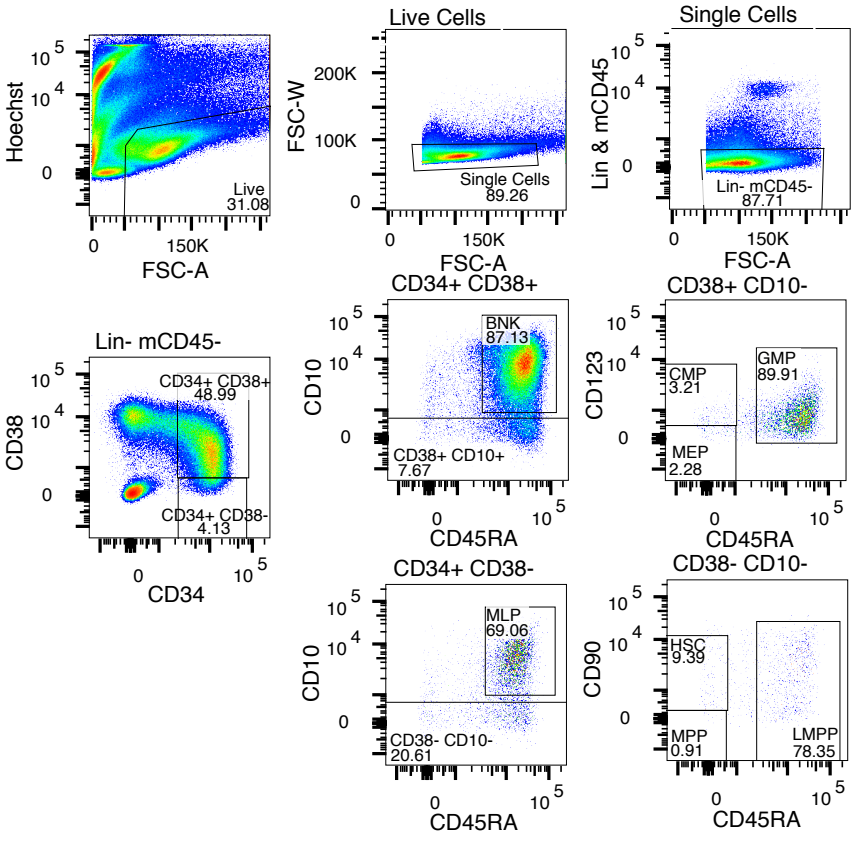
(a) Sample STB257, MLP Injected



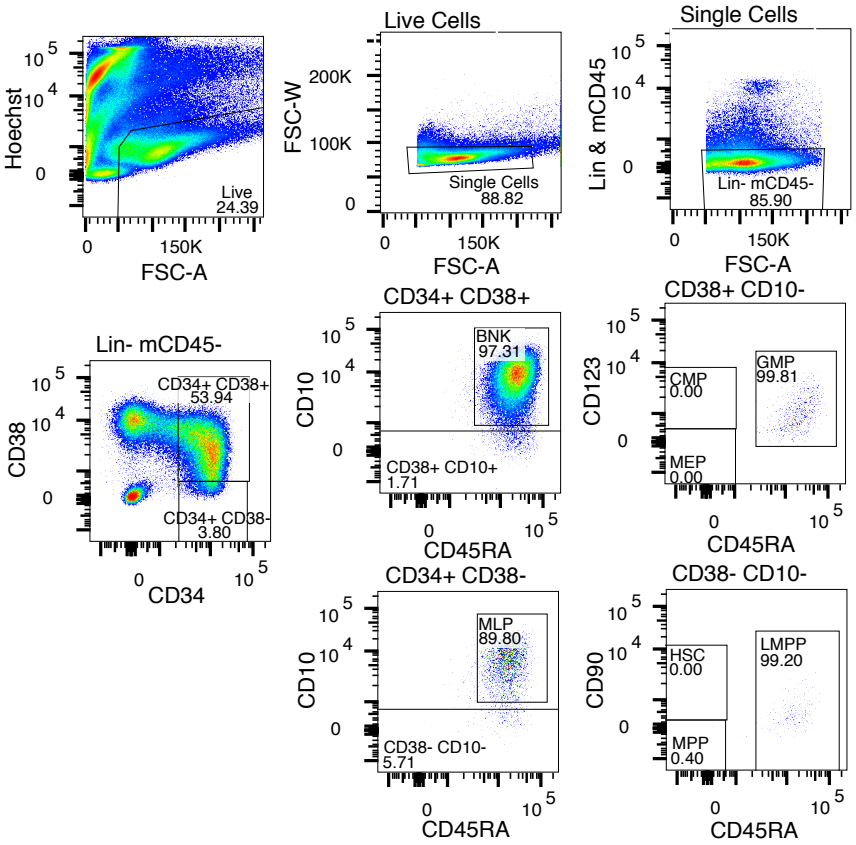
(b) Sample STB257, MLP Injected



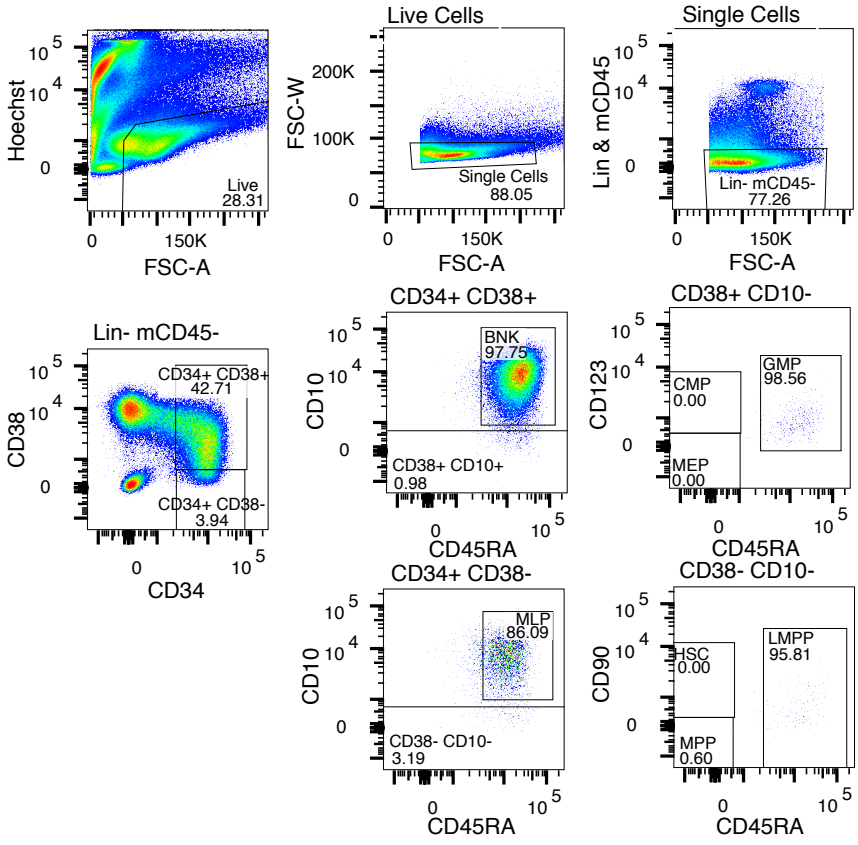
(c) Sample STB257, LMPP Injected



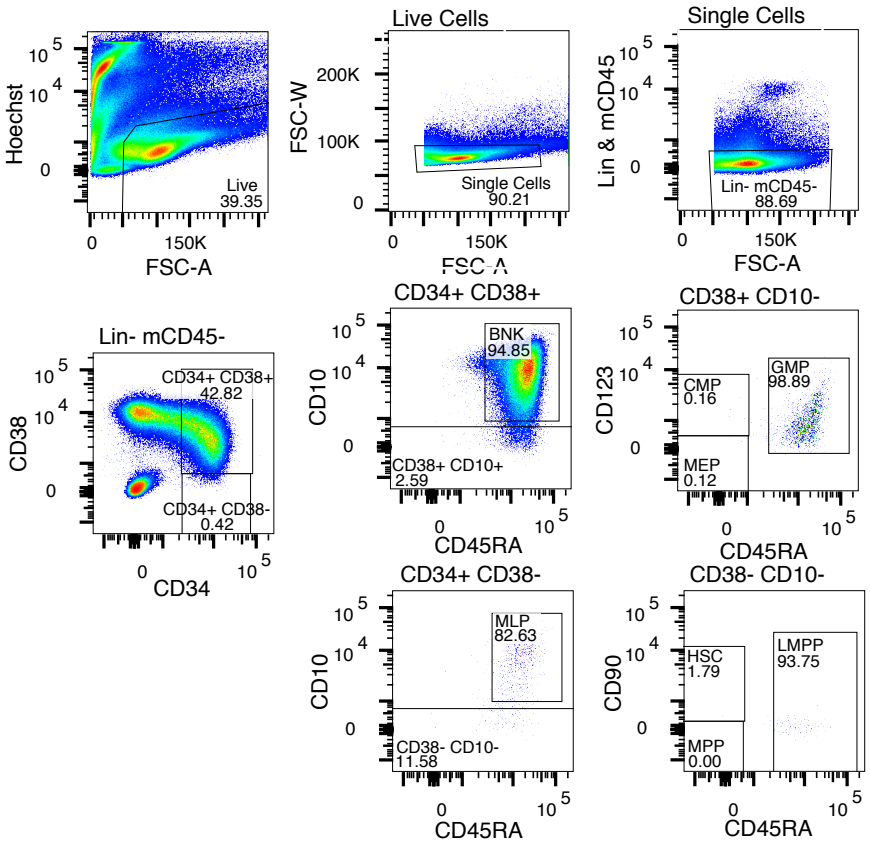
(d) Sample STB257, MLP Injected



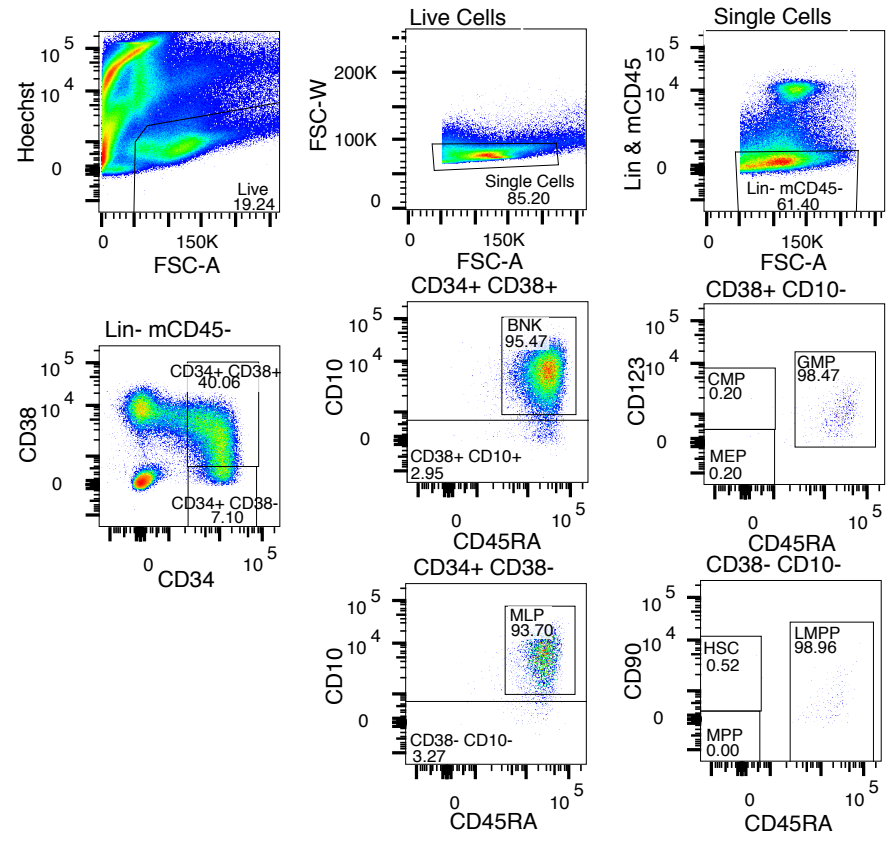
(e) Sample STB257, LMPP Injected



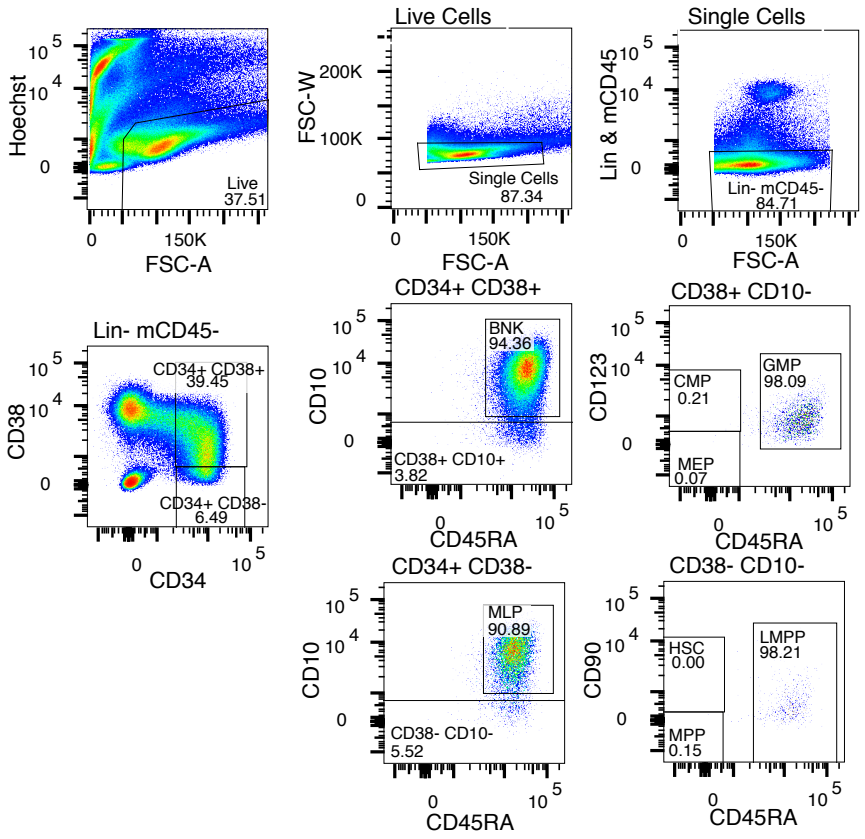
(f) Sample STB257, LMPP Injected



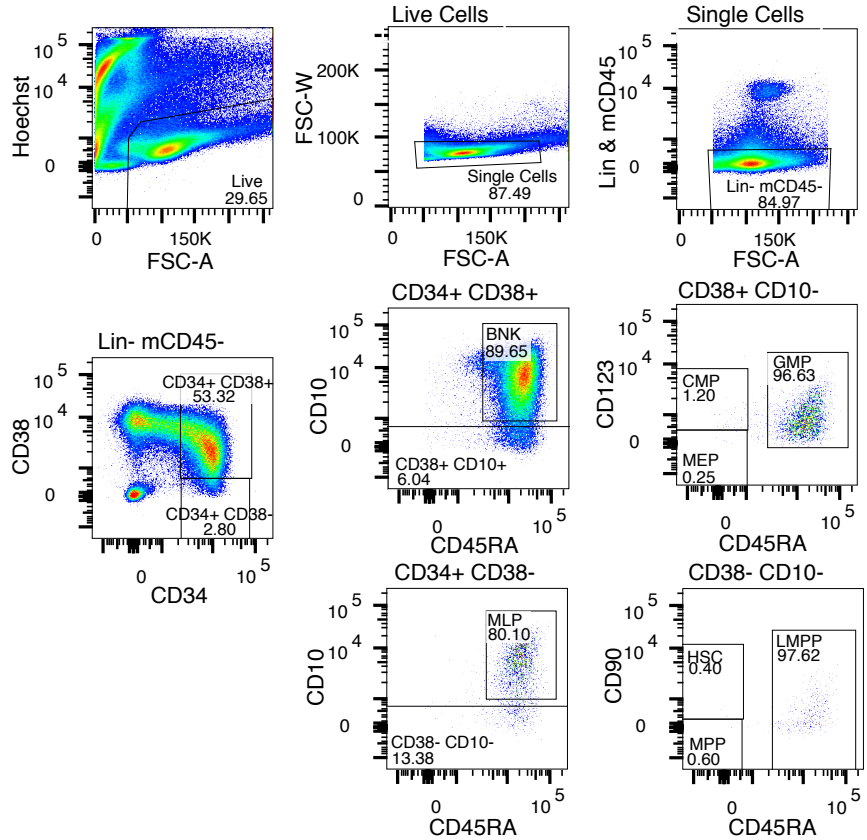
(g) Sample STB257, MLP Injected



(h) Sample STB257, GMP Injected



(i) Sample STB257, LMPP Injected



(j) Sample STB257, GMP Injected

Figure A.5: Post-Engraftment HSPC Immunophenotype of STB257. Post-engraftment HSPC immunophenotype of sample STB257 in xenotransplantation assays. The injected population is indicated below plots. HSPC immunophenotyping was carried out on terminal BM samples. Frequencies indicate the frequency of the parent gate.

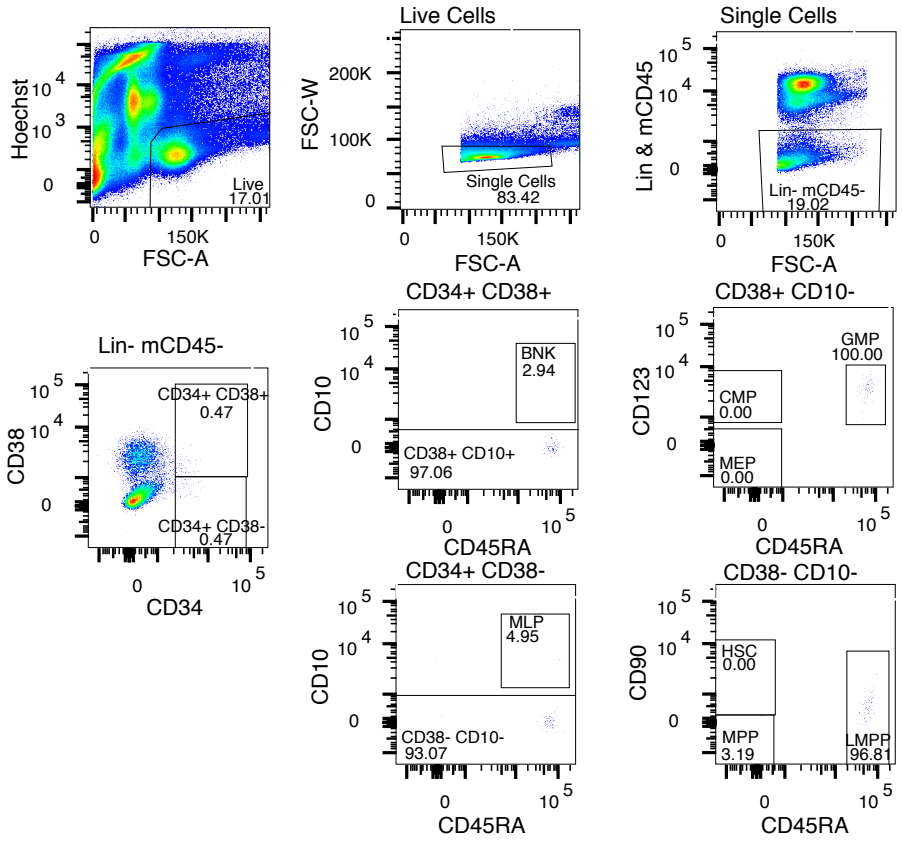
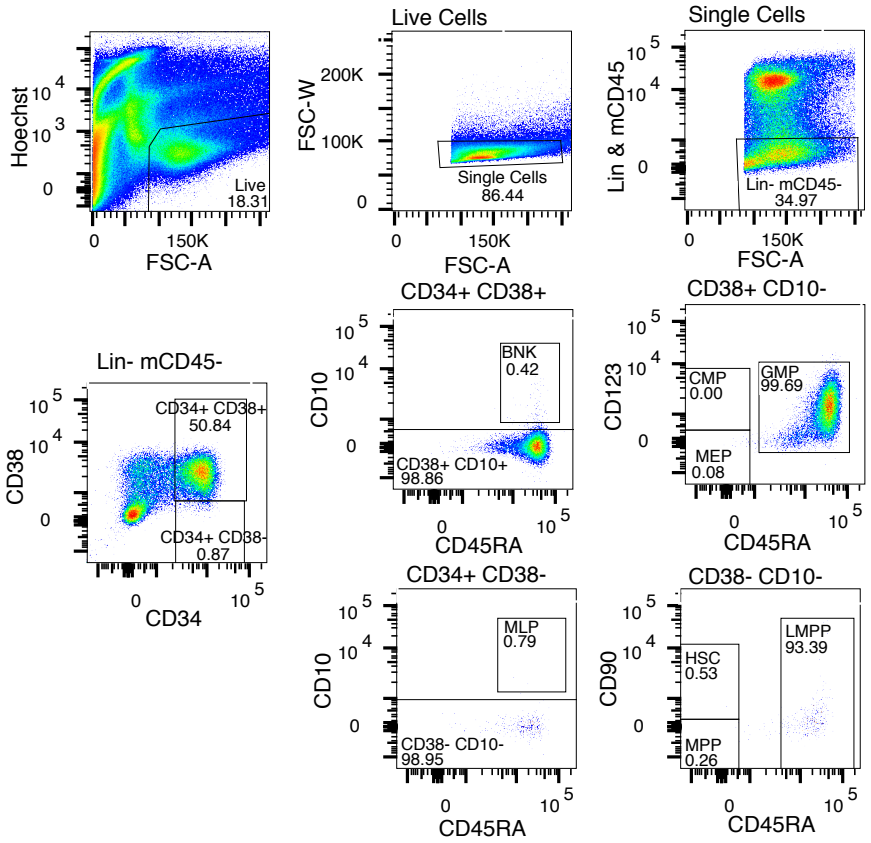
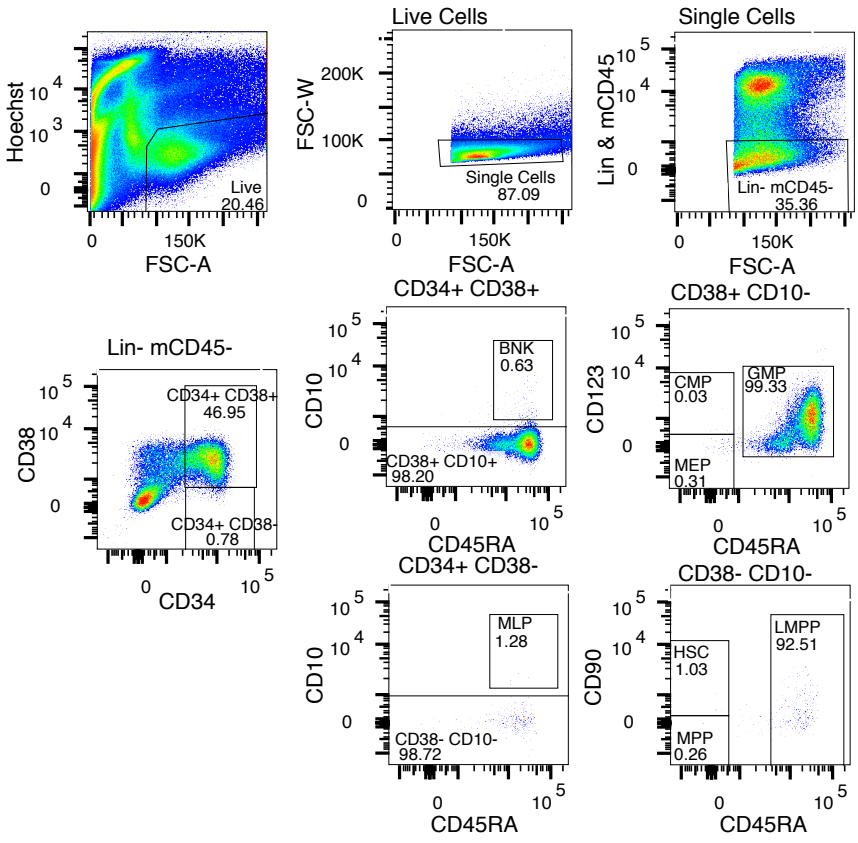


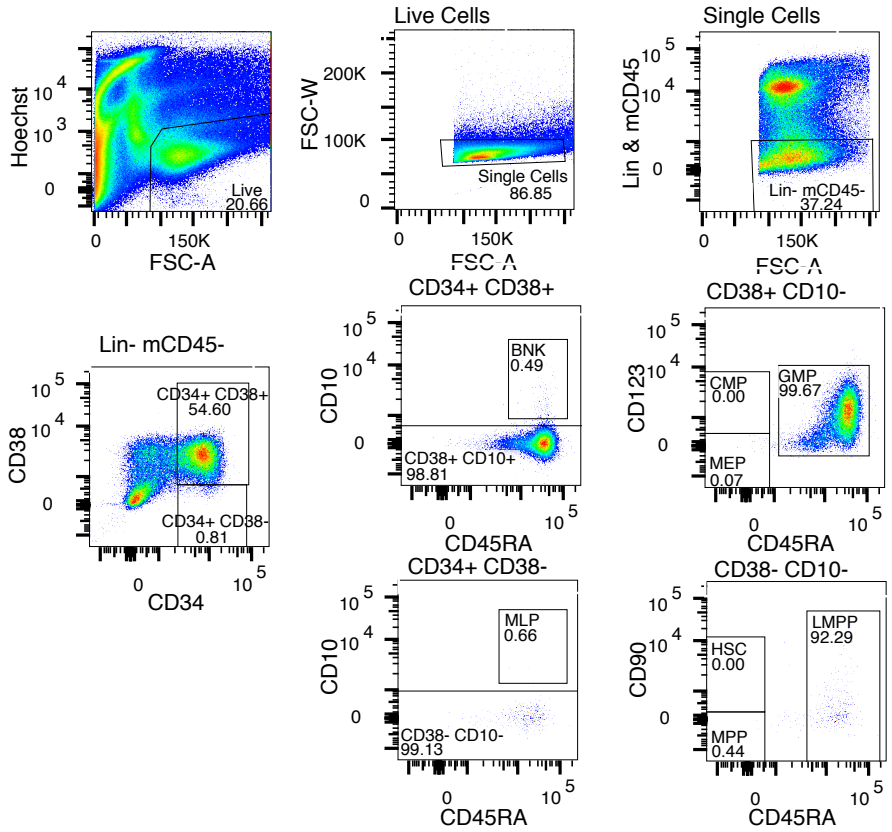
Figure A.6: Post-Engraftment HSPC Immunophenotype of BIR352. Post-engraftment HSPC immunophenotype of sample BIR352 in xenotransplantation assays, following LMPP injection. HSPC immunophenotyping was carried out on terminal BM samples. Frequencies indicate the frequency of the parent gate.



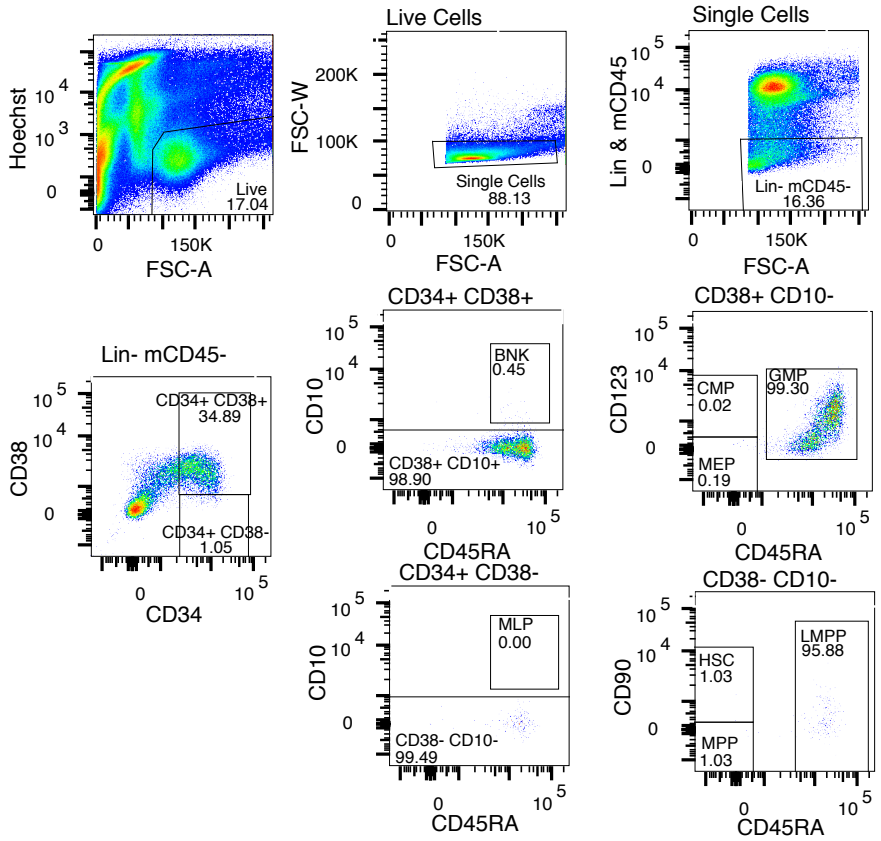
(a) Sample CAR143, LMPP Injected



(b) Sample CAR143, LMPP Injected

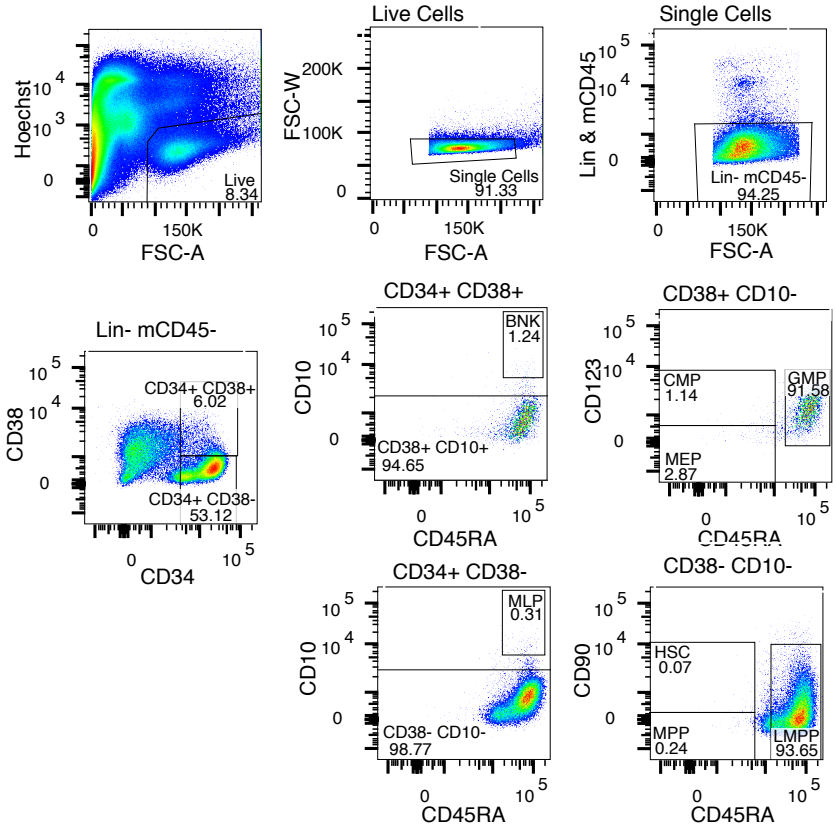


(c) Sample CAR143, GMP Injected

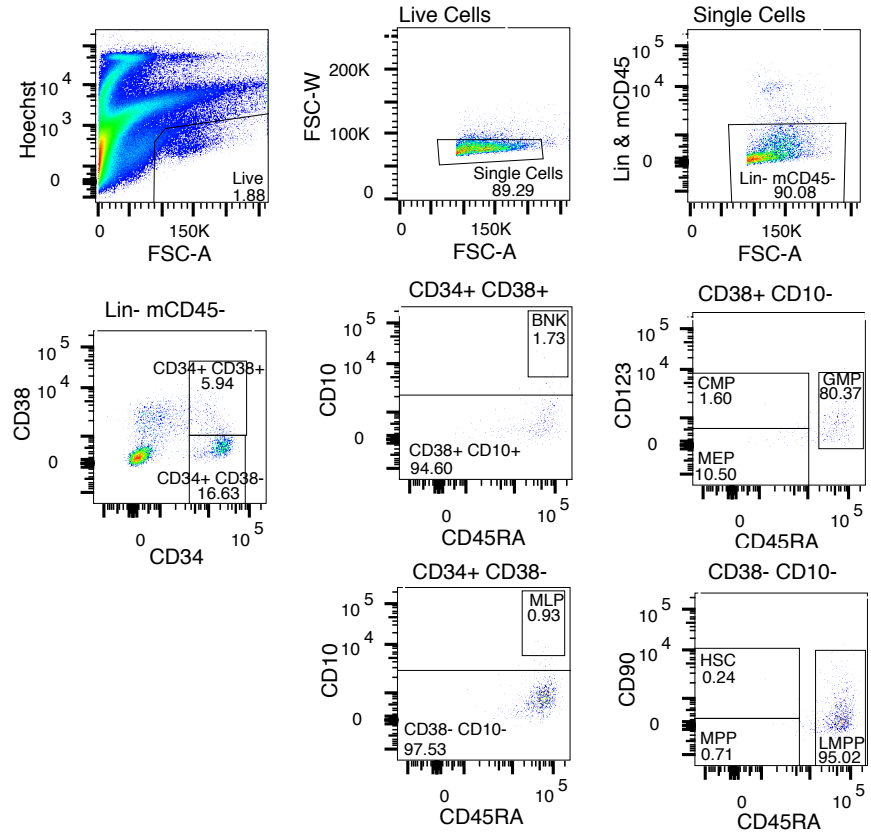


(d) Sample CAR143, GMP Injected

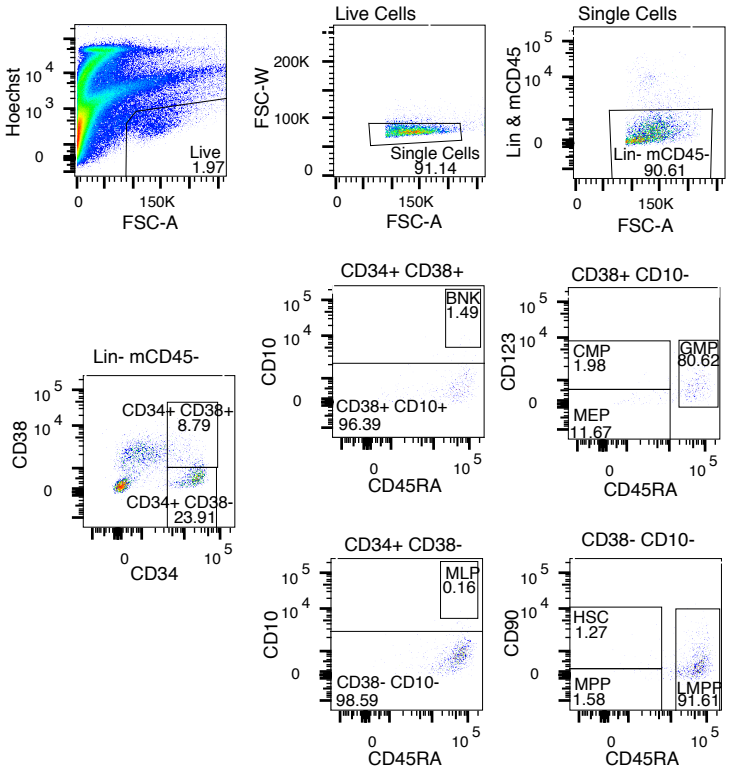
Figure A.7: Post-Engraftment HSPC Immunophenotype of CAR143. Post-engraftment HSPC immunophenotype of sample CAR143 in xenotransplantation assays. The injected population is indicated below plots. HSPC immunophenotyping was carried out on terminal BM samples. Frequencies indicate the frequency of the parent gate.



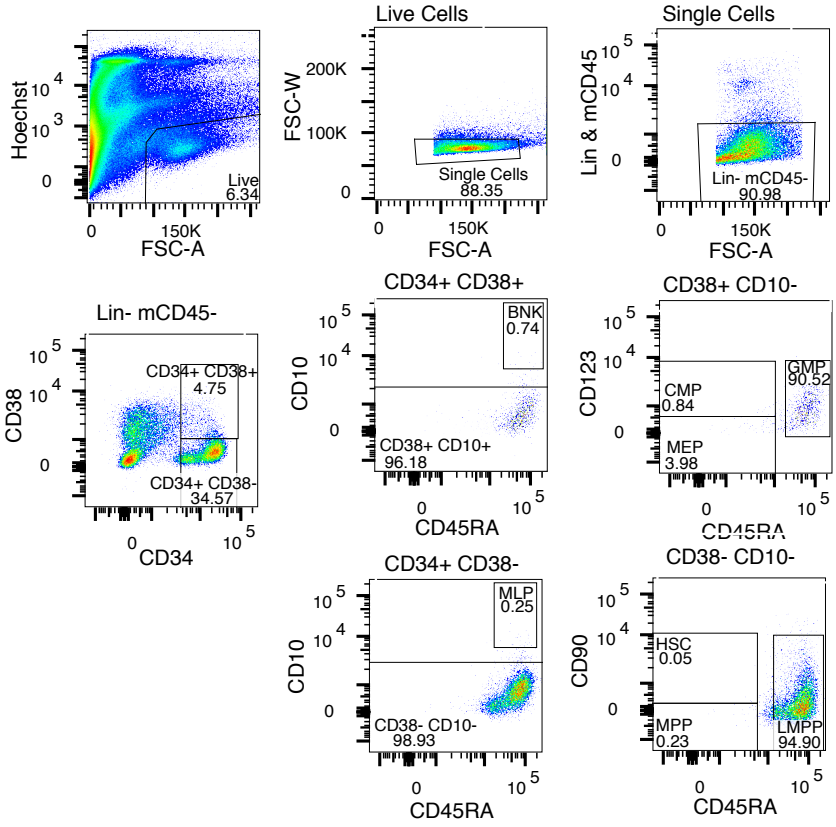
(a) Sample BIR517, LMPP Injected



(b) Sample BIR517, LMPP Injected

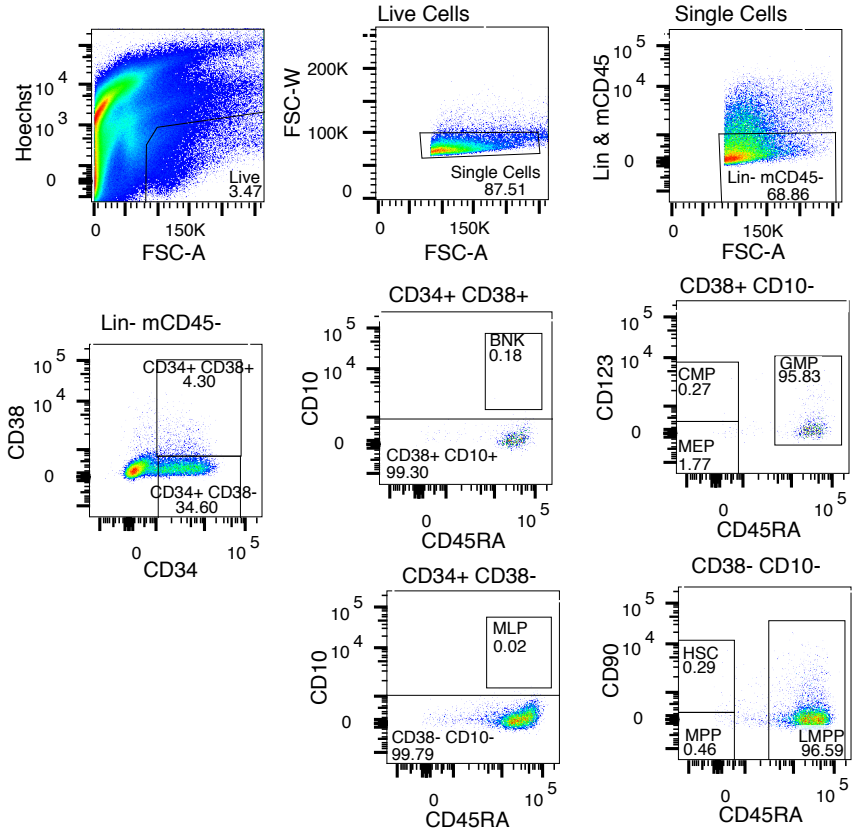


(c) Sample BIR517, LMPP Injected

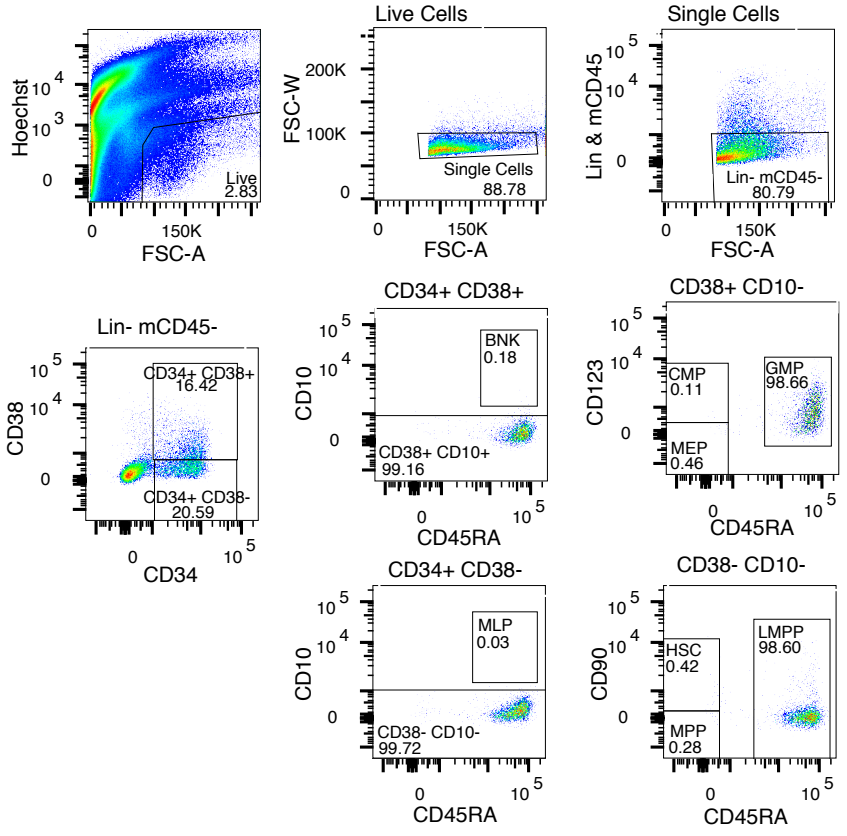


(d) Sample BIR517, LMPP Injected

Figure A.8: Post-Engraftment HSPC Immunophenotype of BIR517. Post-engraftment HSPC immunophenotype of sample BIR517 in xenotransplantation assays. The injected population is indicated below plots. HSPC immunophenotyping was carried out on terminal BM samples. Frequencies indicate the frequency of the parent gate.



(a) Sample OX2116, LMPP Injected



(b) Sample OX2116, LMPP Injected

Figure A.9: Post-Engraftment HSPC Immunophenotype of OX2116. Post-engraftment HSPC immunophenotype of sample OX2116 in xenotransplantation assays. The injected population is indicated below plots. HSPC immunophenotyping was carried out on terminal BM samples. Frequencies indicate the frequency of the parent gate.

A.12 RNA-Seq Analysis of BM HSPCs

Table A.23: Normal BM HSPC Populations: 500 LRT Genes. The 500 most DEGs across normal BM HSPC populations as determined by an LRT in DESEq2. Clusters: 1: early lympho-myeloid, 2: late lymphoid, 3: stem, 4: myeloid, 5: megakaryocytic/erythroid (as indicated in Figure 5.6).

Ensembl	Gene Symbol	Cluster	Ensembl	Gene Symbol	Cluster
ENSG00000171476	HOPX	1	ENSG00000166710	B2M	1
ENSG00000169442	CD52	1	ENSG00000164116	GUCY1A3	1
ENSG00000134324	LPIN1	1	ENSG00000119866	BCL11A	1
ENSG00000204472	AIF1	1	ENSG00000173681	CXorf23	1
ENSG00000019582	CD74	1	ENSG00000125089	SH3TC1	1
ENSG00000204287	HLA-DRA	1	ENSG00000137642	SORL1	1
ENSG00000139278	GLIPR1	1	ENSG00000173821	RNF213	1
ENSG00000231389	HLA-DPA1	1	ENSG00000059588	TARBP1	1
ENSG00000164035	EMCN	1	ENSG00000185189	NRBP2	1
ENSG00000048052	HDAC9	1	ENSG00000075426	FOSL2	1
ENSG00000150347	ARID5B	1	ENSG00000143515	ATP8B2	1
ENSG00000160593	JAML	1	ENSG00000133169	BEX1	1
ENSG00000185215	TNFAIP2	1	ENSG00000189067	LITAF	1
ENSG00000128040	SPINK2	1	ENSG00000070501	POLB	1
ENSG00000042980	ADAM28	1	ENSG00000206503	HLA-A	1
ENSG00000105997	HOXA3	1	ENSG00000153253	SCN3A	2
ENSG00000164484	TMEM200A	1	ENSG00000196549	MME	2
ENSG00000054793	ATP9A	1	ENSG00000126353	CCR7	2
ENSG00000141232	TOB1	1	ENSG00000196628	TCF4	2
ENSG00000183688	RFLNB	1	ENSG00000142669	SH3BGRL3	2
ENSG00000126860	EVI2A	1	ENSG00000262655	SPON1	2
ENSG00000188404	SELL	1	ENSG00000164929	BAALC	2
ENSG00000223865	HLA-DPB1	1	ENSG00000132465	JCHAIN	2
ENSG00000172349	IL16	1	ENSG00000117115	PADI2	2
ENSG00000157110	RBPMS	1	ENSG00000182568	SATB1	2
ENSG00000174059	CD34	1	ENSG00000163219	ARHGAP25	2
ENSG00000042493	CAPG	1	ENSG00000205542	TMSB4X	2
ENSG00000119801	YPEL5	1	ENSG00000275896	PRSS2	2
ENSG00000081320	STK17B	1	ENSG00000012124	CD22	2
ENSG00000091972	CD200	1	ENSG00000109684	CLNK	2
ENSG00000196126	HLA-DRB1	1	ENSG00000140968	IRF8	2
ENSG00000102786	INTS6	1	ENSG00000162894	FCMR	2
ENSG00000145687	SSBP2	1	ENSG00000174469	CNTNAP2	2
ENSG00000134531	EMP1	1	ENSG00000205268	PDE7A	2
ENSG00000153944	MSI2	1	ENSG00000123836	PFKFB2	2
ENSG00000152217	SETBP1	1	ENSG00000148180	GSN	2
ENSG00000204482	LST1	1	ENSG00000100600	LGMN	2
ENSG00000204257	HLA-DMA	1	ENSG00000071575	TRIB2	2
ENSG00000143119	CD53	1	ENSG00000069188	SDK2	2
ENSG00000113369	ARRDC3	1	ENSG00000008513	ST3GAL1	2
ENSG00000104043	ATP8B4	1	ENSG00000100055	CYTH4	2
ENSG00000105639	JAK3	1	ENSG00000162892	IL24	2
ENSG00000101445	PPP1R16B	1	ENSG00000010810	FYN	2
ENSG00000026508	CD44	1	ENSG00000134109	EDEM1	2
ENSG00000198771	RCSD1	1	ENSG00000159314	ARHGAP27	2
ENSG00000104447	TRPS1	1	ENSG00000141480	ARRB2	2
ENSG00000167613	LAIR1	1	ENSG00000271605	MILR1	2
ENSG00000026025	VIM	1	ENSG00000172493	AFF1	2
ENSG00000160712	IL6R	1	ENSG00000135916	ITM2C	2
ENSG00000095932	SMIM24	1	ENSG00000196405	EVL	2
ENSG00000132274	TRIM22	1	ENSG00000168461	RAB31	2
ENSG00000115738	ID2	1	ENSG00000145247	OCIAD2	2
ENSG00000197321	SVIL	1	ENSG00000102007	PLP2	2
ENSG00000129116	PALLD	1	ENSG00000122025	FLT3	2
ENSG00000159840	ZYX	1	ENSG00000234745	HLA-B	2
ENSG00000146285	SCML4	1	ENSG00000107447	DNTT	2
ENSG00000150995	ITPR1	1	ENSG00000112297	AIM1	2
ENSG00000087245	MMP2	1	ENSG00000166963	MAP1A	2
ENSG00000124831	LRRFIP1	1	ENSG00000144642	RBMS3	2
ENSG00000091592	NLRP1	1	ENSG00000141748	ARL5C	2
ENSG00000163346	PBXIP1	1	ENSG00000160191	PDE9A	2
ENSG00000165801	ARHGEF40	1	ENSG00000116741	RGS2	2
ENSG00000154721	JAM2	1	ENSG00000198833	UBE2J1	2
ENSG00000139318	DUSP6	1	ENSG00000102265	TIMP1	2
ENSG00000007062	PROM1	1	ENSG00000055163	CYFIP2	2
ENSG00000138670	RASGEF1B	1	ENSG00000082438	COBLL1	2
ENSG00000119535	CSF3R	1	ENSG00000196839	ADA	2
ENSG00000069424	KCNAB2	1	ENSG00000130592	LSP1	2
ENSG00000170776	AKAP13	1	ENSG00000170476	MZB1	2
ENSG00000171522	PTGER4	1	ENSG00000127507	ADGRE2	2
ENSG00000110237	ARHGEF17	1	ENSG00000158195	WASF2	2
ENSG00000186088	GSAP	1	ENSG00000164300	SERINC5	2
ENSG00000168899	VAMP5	1	ENSG00000152894	PTPRK	2
ENSG00000128335	APOL2	1	ENSG00000124813	RUNX2	2

Continuation of Table A.23

Ensembl	Gene Symbol	Cluster	Ensembl	Gene Symbol	Cluster
ENSG00000179869	ABCA13	1	ENSG00000177663	IL17RA	2
ENSG00000147488	ST18	2	ENSG00000100380	ST13	3
ENSG00000106236	NPTX2	2	ENSG00000174175	SELP	3
ENSG00000128815	NA	2	ENSG00000143995	MEIS1	3
ENSG00000140575	IQGAP1	2	ENSG00000198586	TLK1	3
ENSG00000002586	CD99	2	ENSG00000165275	TRMT10B	3
ENSG00000100167	Sep-03	2	ENSG00000112137	PHACTR1	3
ENSG00000166501	PRKCB	2	ENSG00000135048	TMEM2	3
ENSG00000240891	PLCXD2	2	ENSG00000169398	PTK2	3
ENSG00000204131	NHSL2	2	ENSG00000162366	PDZK1IP1	3
ENSG00000005844	ITGAL	2	ENSG00000115604	IL18R1	3
ENSG00000065413	ANKRD44	2	ENSG00000087086	FTL	3
ENSG00000123146	ADGRE5	2	ENSG00000108924	HLF	3
ENSG00000168675	LDLRAD4	2	ENSG00000008311	AASS	3
ENSG00000197183	NOL4L	2	ENSG00000133321	RARRES3	3
ENSG00000165025	SYK	2	ENSG00000186105	LRRC70	3
ENSG00000135899	SP110	2	ENSG00000180596	HIST1H2BC	3
ENSG00000104081	BMF	2	ENSG00000171490	RSL1D1	3
ENSG00000172260	NEGR1	2	ENSG00000120885	CLU	3
ENSG00000132334	PTPRE	2	ENSG00000232125	DYTN	3
ENSG00000093072	CECR1	2	ENSG00000120594	PLXDC2	3
ENSG00000133246	PRAM1	2	ENSG00000120279	MYCT1	3
ENSG00000166224	SGPL1	2	ENSG00000138722	MMRN1	3
ENSG00000168234	TTC39C	2	ENSG00000169122	FAM110B	3
ENSG00000204983	PRSS1	2	ENSG00000136546	SCN7A	3
ENSG00000035403	VCL	2	ENSG00000116747	TROVE2	3
ENSG00000112081	SRSF3	2	ENSG00000111897	SERINC1	3
ENSG00000142583	SLC2A5	2	ENSG00000151090	THRB	3
ENSG00000136630	HLX	2	ENSG00000204252	HLA-DOA	3
ENSG00000078399	HOXA9	2	ENSG00000105641	SLC5A5	3
ENSG00000213402	PTPRCAP	2	ENSG00000122644	ARL4A	3
ENSG00000211772	NA	2	ENSG00000100650	SRSF5	3
ENSG00000081189	MEF2C	2	ENSG00000162896	PIGR	3
ENSG00000169575	VPREB1	2	ENSG00000155849	ELMO1	3
ENSG00000119328	FAM206A	2	ENSG00000121858	TNFSF10	3
ENSG00000082512	TRAF5	2	ENSG00000147604	RPL7	3
ENSG00000141380	SS18	2	ENSG00000169224	GCSAML	3
ENSG00000204869	LOC400706	2	ENSG00000215301	DDX3X	3
ENSG00000171608	PIK3CD	2	ENSG00000182107	TMEM30B	3
ENSG00000196233	C10orf12	2	ENSG00000008282	SYPL1	3
ENSG00000124780	KCNK17	2	ENSG00000142611	PRDM16	3
ENSG00000170482	SLC23A1	2	ENSG00000172869	DMXL1	3
ENSG00000171643	S100Z	2	ENSG00000123349	PFDN5	3
ENSG00000118946	PCDH17	3	ENSG00000113851	CRBN	3
ENSG00000167552	TUBA1A	3	ENSG00000085563	ABCB1	3
ENSG00000147862	NFIB	3	ENSG00000157554	ERG	3
ENSG00000171843	MLLT3	3	ENSG00000162409	PRKAA2	3
ENSG00000136929	HEMGN	3	ENSG00000111790	FGFR1OP2	3
ENSG00000145708	CRHBP	3	ENSG00000102804	TSC22D1	3
ENSG00000104894	CD37	3	ENSG00000204628	RACK1	3
ENSG00000166482	MFAP4	3	ENSG00000156976	EIF4A2	3
ENSG00000154175	ABI3BP	3	ENSG00000167105	TMEM92	3
ENSG00000064042	LIMCH1	3	ENSG00000157214	STEAP2	3
ENSG00000174748	RPL15	3	ENSG00000018236	CNTN1	3
ENSG00000081138	CDH7	3	ENSG00000185630	PBX1	3
ENSG00000163815	CLEC3B	3	ENSG00000123562	MORF4L2	3
ENSG00000167601	AXL	3	ENSG00000111348	ARHGDIB	3
ENSG00000010404	IDS	3	ENSG00000196139	AKR1C3	3
ENSG00000135535	CD164	3	ENSG00000127955	GNAI1	3
ENSG00000198832	SELENOM	3	ENSG00000095303	PTGS1	3
ENSG00000197959	DNM3	3	ENSG00000159592	GPBP1L1	3
ENSG00000144824	PHLDB2	3	ENSG00000142046	TMEM91	3
ENSG00000137941	TTLL7	3	ENSG00000169403	PTAFR	3
ENSG00000204592	HLA-E	3	ENSG00000178053	MLF1	3
ENSG00000046889	PREX2	3	ENSG00000112308	C6orf62	3
ENSG00000130303	BST2	3	ENSG00000138669	PRKG2	3
ENSG00000144893	MED12L	3	ENSG00000068697	LAPTM4A	3
ENSG00000113712	CSNK1A1	3	ENSG00000169251	NMD3	3
ENSG00000173889	PHC3	3	ENSG00000081237	PTPRC	3
ENSG00000180530	NRIP1	3	ENSG00000188243	COMMD6	3
ENSG00000135426	TESPA1	3	ENSG00000179218	CALR	4
ENSG00000181722	ZBTB20	3	ENSG00000159199	ATP5G1	4
ENSG00000134202	GSTM3	3	ENSG00000120802	TMPO	4
ENSG00000196935	SRGAP1	3	ENSG00000197956	S100A6	4
ENSG00000104408	EIF3E	3	ENSG00000084234	APLP2	4
ENSG00000184226	PCDH9	3	ENSG00000172232	AZU1	4
ENSG00000105855	ITGB8	3	ENSG00000243989	ACY1	4
ENSG00000213366	GSTM2	3	ENSG00000164687	FABP5	4
ENSG00000166046	TCP11L2	3	ENSG00000196209	SIRPB2	4
ENSG00000183283	DAZAP2	3	ENSG00000241468	ATP5J2	4
ENSG00000115457	IGFBP2	4	ENSG00000197561	ELANE	4
ENSG00000245848	CEBPA	4	ENSG00000163737	PF4	4
ENSG00000138829	FBN2	4	ENSG00000100028	SNRPD3	4
ENSG00000072274	TFRC	4	ENSG00000134815	DHX34	4
ENSG00000276043	UHRF1	4	ENSG00000106211	HSPB1	4
ENSG00000149534	MS4A2	4	ENSG00000178921	PFAS	4
ENSG00000049323	LTBP1	4	ENSG00000115956	PLEK	4
ENSG00000130816	DNMT1	4	ENSG00000243147	MRPL33	4
ENSG00000221869	CEBPD	4	ENSG00000100479	POLE2	4
ENSG00000092470	WDR76	4	ENSG00000239672	NME1	4
ENSG00000034510	TMSB10	4	ENSG00000119705	SLIRP	4
ENSG00000169385	RNASE2	4	ENSG00000138442	WDR12	4

Continuation of Table A.23

Ensembl	Gene Symbol	Cluster	Ensembl	Gene Symbol	Cluster
ENSG00000019991	HGF	4	ENSG00000263961	C1orf186	4
ENSG00000158417	EIF5B	4	ENSG00000121552	CSTA	4
ENSG00000105193	RPS16	4	ENSG00000105205	CLC	4
ENSG00000110077	MS4A6A	4	ENSG00000197766	CFD	4
ENSG00000187514	PTMA	4	ENSG00000104738	MCM4	4
ENSG00000113569	NUP155	4	ENSG00000166123	GPT2	4
ENSG00000143977	SNRPG	4	ENSG00000167419	LPO	4
ENSG00000198176	TFDP1	4	ENSG00000066336	SPI1	4
ENSG00000189403	HMGB1	4	ENSG00000135678	CPM	4
ENSG00000124491	F13A1	4	ENSG00000131174	COX7B	4
ENSG00000169397	RNASE3	4	ENSG00000163041	H3F3A	4
ENSG00000173436	MINOS1	4	ENSG00000065328	MCM10	4
ENSG00000179958	DCTPP1	4	ENSG00000198258	UBL5	4
ENSG00000184992	BRI3BP	4	ENSG00000112312	GMNN	4
ENSG00000005379	TSPOAP1	4	ENSG00000198830	HMGN2	4
ENSG00000266074	BAHCC1	4	ENSG00000167088	SNRPD1	4
ENSG00000099901	RANBP1	4	ENSG00000173660	UQCRH	4
ENSG00000090382	LYZ	4	ENSG00000175538	KCNE3	4
ENSG00000156411	C14orf2	4	ENSG00000039068	CDH1	5
ENSG00000139343	SNRPF	4	ENSG00000181467	RAP2B	5
ENSG00000185198	PRSS57	4	ENSG00000182118	FAM89A	5
ENSG00000169020	ATP5I	4	ENSG00000100225	FBXO7	5
ENSG00000108561	C1QBP	4	ENSG00000055044	NOP58	5
ENSG00000155368	DBI	4	ENSG00000181061	HIGD1A	5
ENSG00000163736	PPBP	4	ENSG00000130826	DKC1	5
ENSG00000183011	NAA38	4	ENSG00000149218	ENDOD1	5
ENSG00000166803	KIAA0101	4	ENSG00000111843	TMEM14C	5
ENSG00000168502	MTCL1	4	ENSG00000150593	PDCD4	5
ENSG00000182774	RPS17	4	ENSG00000090013	BLVRB	5
ENSG00000115541	HSPE1	4	ENSG00000135218	CD36	5
ENSG00000228716	DHFR	4	ENSG00000138796	HADH	5
ENSG00000183684	ALYREF	4	ENSG00000105879	CBLL1	5
ENSG00000155966	AFF2	4	ENSG00000087842	PIR	5
ENSG00000113749	HRH2	4	ENSG00000126561	STAT5A	5
ENSG00000051341	POLQ	4	ENSG00000109265	KIAA1211	5
ENSG00000169554	ZEB2	4	ENSG00000129562	DAD1	5
ENSG00000106305	AIMP2	4	ENSG00000132792	CTNNBL1	5
ENSG00000141510	TP53	4	ENSG00000213965	NUDT19	5
ENSG00000149516	MS4A3	4	ENSG00000109971	HSPA8	5
ENSG00000111775	COX6A1	4	ENSG00000150867	PIP4K2A	5
ENSG00000165264	NDUFB6	4	ENSG00000133816	MICAL2	5
ENSG00000164405	UQCRQ	4	ENSG00000140416	TPM1	5
ENSG00000011028	MRC2	4	ENSG00000244509	APOBEC3C	5
ENSG00000105339	DENND3	4	ENSG00000128683	GAD1	5
ENSG00000143321	HDGF	4	ENSG00000029534	ANK1	5
ENSG00000166794	PPIB	4	ENSG00000187715	KBTBD12	5
ENSG00000186652	PRG2	4	ENSG00000124782	RREB1	5
ENSG00000068724	TTC7A	4	ENSG00000177469	PTRF	5
ENSG00000065978	YBX1	4	ENSG00000116161	CACYBP	5
ENSG00000176890	TYMS	4	ENSG00000196914	ARHGEF12	5
ENSG00000122862	SRGN	4	ENSG00000115947	ORC4	5
ENSG00000124172	ATP5E	4	ENSG00000108439	PNPO	5
ENSG00000085840	ORC1	4	ENSG00000141469	SLC14A1	5
ENSG00000110917	MLEC	4	ENSG00000168754	FAM178B	5
ENSG00000166896	ATP23	4	ENSG00000198898	CAPZA2	5
ENSG00000164032	H2AFZ	4	ENSG00000167553	TUBA1C	5
ENSG00000123975	CKS2	4	ENSG00000147872	PLIN2	5
ENSG00000136450	SRSF1	4	ENSG00000162367	TAL1	5
ENSG00000106605	BLVRA	5	ENSG00000131171	SH3BGRL	5
ENSG00000173262	SLC2A14	5	ENSG00000000971	CFH	5
ENSG00000126088	UROD	5	ENSG00000148346	LCN2	5
ENSG00000162599	NFIA	5	ENSG00000165650	PDZD8	5
ENSG00000119865	CNRIP1	5	ENSG00000005961	ITGA2B	5
ENSG00000117322	CR2	5	ENSG00000181163	NPM1	5
ENSG00000181350	LRRC75A	5	ENSG00000164741	DLC1	5
ENSG00000168078	PBK	5	ENSG00000135838	NPL	5
ENSG00000166831	RBPMS2	5	ENSG00000096384	HSP90AB1	5
ENSG00000165672	PRDX3	5	ENSG00000182117	NOP10	5
ENSG00000106327	TFR2	5	ENSG00000137198	GMPR	5
ENSG00000197993	KEL	5	ENSG00000166743	ACSM1	5
ENSG00000120278	PLEKHG1	5	ENSG00000168952	STXBP6	5
ENSG00000114054	PCCB	5	ENSG00000143416	SELENBP1	5
ENSG00000005187	ACSM3	5	ENSG00000273841	TAF9	5
ENSG00000111716	LDHB	5	ENSG00000120915	EPHX2	5
ENSG00000172752	COL6A5	5	ENSG00000165092	ALDH1A1	5
ENSG00000113328	CCNG1	5	ENSG00000012660	ELOVL5	5
ENSG00000076043	REXO2	5	ENSG00000167815	PRDX2	5
ENSG00000196154	S100A4	5			
ENSG00000179639	FCER1A	5			
ENSG00000143158	MPC2	5			
ENSG00000122574	WIPF3	5			
ENSG00000101166	PRELID3B	5			
ENSG00000180537	RNF182	5			
ENSG00000144677	CTDSPL	5			
ENSG00000140307	GTF2A2	5			
ENSG00000162496	DHRS3	5			
ENSG00000164305	CASP3	5			
ENSG00000138449	SLC40A1	5			
ENSG00000066294	CD84	5			
ENSG00000187134	AKR1C1	5			
ENSG00000165702	GFI1B	5			

Table A.24: Normal BM HSPC Populations: Upregulated Stem Genes. DEGs were identified between early stem (HSC and MPP) and progenitor (LMPP, MLP, CMP, GMP, MEP) populations using a Wald test in DESeq2. Upregulated DEGs with a p-adjusted value < 0.05 and a positive logFC are shown.

Ensembl	$\log_2$ Fold Change	padj	Gene Symbol	Ensembl	$\log_2$ Fold Change	padj	Gene Symbol
ENSG00000118946	8.15	1.6e-77	PCDH17	ENSG00000168067	1.80	4.7e-05	MAP4K2
ENSG00000163815	7.32	1.9e-39	CLEC3B	ENSG00000185559	3.50	4.7e-05	DLK1
ENSG00000198832	5.07	1.6e-26	SELENOM	ENSG00000054654	1.23	5.2e-05	SYNE2
ENSG00000081138	3.57	2.1e-21	CDH7	ENSG00000164111	1.29	5.2e-05	ANXA5
ENSG00000154175	3.63	5.7e-19	ABI3BP	ENSG00000182263	2.96	5.2e-05	FIGN
ENSG00000136546	4.22	2.2e-18	SCN7A	ENSG00000187837	1.20	5.2e-05	HIST1H1C
ENSG00000144824	4.33	2.0e-17	PHLDB2	ENSG00000204682	2.58	5.5e-05	CASC10
ENSG00000170961	5.70	1.8e-16	HAS2	ENSG00000272196	1.45	5.5e-05	HIST2H2AA4
ENSG00000164935	5.67	6.2e-15	DCSTAMP	ENSG00000189060	1.71	6.2e-05	H1F0
ENSG00000137941	2.81	1.5e-13	TTLL7	ENSG00000204387	0.65	6.9e-05	C6orf48
ENSG00000171843	2.73	7.7e-13	MLLT3	ENSG00000138669	2.65	8.0e-05	PRKG2
ENSG00000115604	4.59	1.2e-12	IL18R1	ENSG00000071082	0.58	9.7e-05	RPL31
ENSG00000162409	3.16	1.6e-12	PRKAA2	ENSG00000108590	1.64	9.7e-05	MED31
ENSG00000152601	0.96	2.1e-11	MBNL1	ENSG00000138085	0.86	1.0e-04	ATRAID
ENSG00000134202	2.52	4.2e-11	GSTM3	ENSG00000159648	2.26	1.1e-04	TEPP
ENSG00000046889	3.98	8.6e-11	PREX2	ENSG00000198589	1.20	1.2e-04	LRBA
ENSG00000087086	0.77	1.1e-10	FTL	ENSG00000198774	2.82	1.3e-04	RASSF9
ENSG00000105835	1.90	1.9e-10	NAMPT	ENSG00000174099	1.48	1.3e-04	MSRB3
ENSG00000169435	3.08	5.1e-10	RASSF6	ENSG00000112137	1.29	1.4e-04	PHACTR1
ENSG00000169122	4.45	5.8e-10	FAM110B	ENSG00000080166	3.26	1.4e-04	DCT
ENSG00000181722	1.51	5.8e-10	ZBTB20	ENSG00000023445	1.80	1.6e-04	BIRC3
ENSG00000109475	0.71	9.3e-10	RPL34	ENSG00000135426	1.46	1.6e-04	TESPA1
ENSG00000155849	1.70	1.9e-09	ELMO1	ENSG00000065621	1.60	1.6e-04	GSTO2
ENSG00000133687	2.85	2.3e-09	TMTC1	ENSG00000151650	1.63	1.9e-04	VENTX
ENSG00000064042	4.37	2.6e-09	LIMCH1	ENSG00000141293	2.58	2.1e-04	SKAP1
ENSG00000150556	4.69	5.1e-09	LYPD6B	ENSG00000105143	1.12	2.1e-04	SLC1A6
ENSG00000008282	1.45	5.8e-09	SYPL1	ENSG00000139083	0.74	2.4e-04	ETV6
ENSG00000154330	2.32	6.2e-09	PGM5	ENSG00000273802	1.72	2.5e-04	HIST1H2BG
ENSG00000157214	3.19	8.6e-09	STEAP2	ENSG00000170500	2.18	2.6e-04	LONRF2
ENSG00000244694	3.98	9.7e-09	PTCHD4	ENSG00000184226	2.37	2.6e-04	PCDH9
ENSG00000180596	1.60	2.7e-08	HIST1H2BC	ENSG00000162641	2.83	2.7e-04	AKNAD1
ENSG00000141431	4.48	3.6e-08	ASXL3	ENSG00000147862	2.96	2.7e-04	NFIB
ENSG00000085276	1.99	3.7e-08	MECOM	ENSG00000167552	1.60	2.8e-04	TUBA1A
ENSG00000169239	1.30	6.0e-08	CA5B	ENSG00000143545	1.15	3.0e-04	RAB13
ENSG00000188243	1.06	7.0e-08	COMMD6	ENSG00000120279	2.50	3.0e-04	MYCT1
ENSG00000213366	1.77	7.5e-08	GSTM2	ENSG00000108924	2.97	3.1e-04	HLF
ENSG00000120885	2.96	7.9e-08	CLU	ENSG00000102543	1.31	3.2e-04	CDADC1
ENSG00000186105	2.74	9.7e-08	LRRC70	ENSG00000118596	1.38	3.2e-04	SLC16A7
ENSG00000130303	2.07	1.1e-07	BST2	ENSG00000149806	0.63	3.4e-04	FAU
ENSG00000158122	1.02	1.2e-07	AAED1	ENSG00000133321	1.48	3.4e-04	RARRES3
ENSG00000174748	0.68	1.2e-07	RPL15	ENSG00000185019	1.38	3.4e-04	UBOX5
ENSG00000232125	2.30	1.2e-07	DYTN	ENSG00000166387	2.50	3.5e-04	PPFIBP2
ENSG00000178053	3.05	1.6e-07	MLF1	ENSG00000169398	2.03	3.8e-04	PTK2
ENSG00000142611	2.51	2.7e-07	PRDM16	ENSG00000115896	2.95	4.0e-04	PLCL1
ENSG00000154678	3.67	3.6e-07	PDE1C	ENSG00000123094	2.15	4.0e-04	RASSF8
ENSG00000182899	0.58	3.6e-07	RPL35A	ENSG00000184515	2.43	4.1e-04	BEX5
ENSG00000157554	1.34	4.2e-07	ERG	ENSG00000154783	2.55	4.5e-04	FGD5
ENSG00000109756	1.82	8.7e-07	RAPGEF2	ENSG00000161011	0.91	4.6e-04	SQSTM1
ENSG00000105855	2.48	8.9e-07	ITGB8	ENSG00000178567	0.92	5.0e-04	EPM2AIP1
ENSG00000142676	0.52	1.3e-06	RPL11	ENSG00000163283	1.97	5.2e-04	ALPP
ENSG00000018236	3.22	1.6e-06	CNTN1	ENSG00000188313	0.91	5.3e-04	PLSCR1
ENSG00000071967	1.71	2.0e-06	CYBRD1	ENSG00000127074	1.32	5.3e-04	RGS13
ENSG00000182916	3.94	2.4e-06	TCEAL7	ENSG00000206560	1.15	5.4e-04	ANKRD28
ENSG00000182107	3.60	3.0e-06	TMEM30B	ENSG00000142875	1.25	5.8e-04	PRKACB
ENSG00000154065	3.02	3.2e-06	ANKRD29	ENSG00000088899	3.05	5.9e-04	LZTS3
ENSG00000167858	3.89	3.3e-06	TEKT1	ENSG00000112096	0.66	5.9e-04	SOD2
ENSG00000154269	3.50	4.2e-06	ENPP3	ENSG00000180626	2.05	6.1e-04	ZNF594
ENSG00000180530	1.89	4.2e-06	NRIP1	ENSG00000169740	1.29	6.1e-04	ZNF32
ENSG00000184905	3.60	4.2e-06	TCEAL2	ENSG00000204592	1.54	6.1e-04	HLA-E
ENSG00000197992	3.59	4.5e-06	CLEC9A	ENSG00000159640	1.77	6.4e-04	ACE
ENSG00000122644	1.61	4.7e-06	ARL4A	ENSG00000165169	1.66	6.5e-04	DYNLT3
ENSG00000135048	2.19	4.7e-06	TMEM2	ENSG00000172794	1.06	6.8e-04	RAB37
ENSG00000159871	1.89	5.3e-06	LYPD5	ENSG00000154277	2.54	6.9e-04	UCHL1
ENSG00000166405	2.03	5.4e-06	RIC3	ENSG00000174080	1.84	7.0e-04	CTSF
ENSG00000143153	1.53	5.7e-06	ATP1B1	ENSG00000115602	1.97	7.1e-04	IL1RL1
ENSG00000009724	3.06	5.8e-06	MASP2	ENSG00000102409	0.85	7.1e-04	BEX4
ENSG00000120833	2.36	6.7e-06	SOCS2	ENSG00000073756	1.72	7.2e-04	PTGS2
ENSG00000156482	0.56	9.3e-06	RPL30	ENSG00000164116	1.65	7.3e-04	GUCY1A3
ENSG00000162366	3.52	9.9e-06	PDZK1IP1	ENSG00000121207	1.69	7.5e-04	LRAT
ENSG00000153234	1.89	1.0e-05	NR4A2	ENSG00000138696	2.50	7.6e-04	BMPR1B
ENSG00000142046	1.94	1.2e-05	TMEM91	ENSG00000089327	0.77	8.0e-04	FXYD5
ENSG00000145708	3.07	1.2e-05	CRHBP	ENSG00000123562	0.91	8.0e-04	MORF4L2
ENSG00000153714	3.70	1.3e-05	LURAP1L	ENSG00000139433	0.85	8.3e-04	GLTP
ENSG00000147604	0.71	1.3e-05	RPL7	ENSG00000138722	1.93	8.4e-04	MMRN1
ENSG00000132321	3.68	1.4e-05	IQCA1	ENSG00000168497	1.85	8.5e-04	SDPR
ENSG00000169403	2.54	1.9e-05	PTAFR	ENSG00000068697	1.03	8.6e-04	LAPTM4A
ENSG00000018408	1.41	2.3e-05	WWTR1	ENSG00000115594	2.26	8.9e-04	IL1R1
ENSG00000163682	0.56	2.6e-05	RPL9	ENSG00000136895	1.82	8.9e-04	GARNL3
ENSG00000123349	0.63	3.6e-05	PFDN5	ENSG00000167601	2.06	9.9e-04	AXL
ENSG00000124570	1.29	4.1e-05	SERPINB6	ENSG00000168993	2.87	9.9e-04	CPLX1
ENSG00000133134	1.34	4.2e-05	BEX2	ENSG00000138131	2.91	1.0e-03	LOXL4
ENSG00000134243	2.30	4.4e-05	SORT1	ENSG00000183741	1.05	1.0e-03	CBX6
ENSG00000222009	1.91	4.6e-05	BTBD19	ENSG00000011454	1.17	1.0e-03	GPR21

Continuation of Table A.24

Ensembl	$\log_2$ Fold Change	padj	Gene Symbol	Ensembl	$\log_2$ Fold Change	padj	Gene Symbol
ENSG00000079337	2.88	1.0e-03	RAPGEF3	ENSG00000170962	2.53	4.9e-03	PDGFD
ENSG00000151623	2.83	1.1e-03	NR3C2	ENSG00000113811	0.66	5.1e-03	SELENOK
ENSG00000119844	1.02	1.1e-03	AFTPH	ENSG00000189241	0.85	5.2e-03	TSPYL1
ENSG00000156976	0.85	1.2e-03	EIF4A2	ENSG00000147535	0.80	5.3e-03	PLPP5
ENSG00000259417	2.18	1.2e-03	LINC01314	ENSG00000197448	0.73	5.4e-03	GSTK1
ENSG00000122223	1.69	1.2e-03	CD244	ENSG00000186409	0.80	5.4e-03	CCDC30
ENSG00000182103	2.31	1.2e-03	FAM181B	ENSG00000101000	1.37	5.5e-03	PROCR
ENSG00000161958	2.49	1.2e-03	FGF11	ENSG00000181847	2.22	5.5e-03	TIGIT
ENSG00000155980	2.77	1.2e-03	KIF5A	ENSG00000165816	2.37	5.5e-03	VWA2
ENSG00000061918	1.89	1.2e-03	GUCY1B3	ENSG00000079308	1.69	5.5e-03	TNS1
ENSG00000150627	1.65	1.2e-03	WDR17	ENSG00000027075	1.26	5.6e-03	PRKCH
ENSG00000123240	1.59	1.2e-03	OPTN	ENSG00000165406	1.18	5.6e-03	Mar-08
ENSG00000142089	1.09	1.2e-03	IFITM3	ENSG00000144713	0.46	5.6e-03	RPL32
ENSG00000150783	2.78	1.5e-03	TEX12	ENSG00000111728	1.30	5.6e-03	ST8SIA1
ENSG00000166816	2.84	1.5e-03	LDHD	ENSG00000136451	0.94	5.7e-03	VEZF1
ENSG00000115268	0.54	1.5e-03	RPS15	ENSG00000167105	2.07	5.7e-03	TMEM92
ENSG00000172164	1.41	1.5e-03	SNTB1	ENSG00000149273	0.47	5.7e-03	RPS3
ENSG00000100316	0.68	1.5e-03	RPL3	ENSG00000166046	2.03	5.8e-03	TCP11L2
ENSG00000104714	1.05	1.6e-03	ERICH1	ENSG00000173114	2.21	5.8e-03	LRRN3
ENSG00000183401	1.10	1.7e-03	CCDC159	ENSG00000131459	2.48	5.9e-03	GFPT2
ENSG00000022267	1.29	1.8e-03	FHL1	ENSG00000166902	1.21	5.9e-03	MRPL16
ENSG00000174599	2.72	1.8e-03	TRAM1L1	ENSG00000069431	0.75	6.4e-03	ABCC9
ENSG00000165475	1.34	1.8e-03	CRYL1	ENSG00000133142	0.70	6.4e-03	TCEAL4
ENSG00000112149	1.65	1.9e-03	CD83	ENSG00000178460	0.55	6.4e-03	MCMDC2
ENSG00000171160	2.03	1.9e-03	MORN4	ENSG00000165502	0.48	6.4e-03	RPL36AL
ENSG00000232388	0.63	1.9e-03	LINC00493	ENSG00000164609	0.73	6.4e-03	SLU7
ENSG00000134480	1.17	1.9e-03	CCNH	ENSG00000114124	1.41	6.5e-03	GRK7
ENSG00000010626	1.79	1.9e-03	LRRC23	ENSG00000162191	0.83	6.5e-03	UBXN1
ENSG00000162896	1.32	2.0e-03	PIGR	ENSG00000110492	1.96	6.5e-03	MDK
ENSG00000100650	1.00	2.0e-03	SRSF5	ENSG00000154229	1.43	6.7e-03	PRKCA
ENSG00000281039	1.81	2.0e-03	NA	ENSG00000172243	1.37	6.8e-03	CLEC7A
ENSG00000102362	1.89	2.0e-03	SYTL4	ENSG00000143995	1.53	6.8e-03	MEIS1
ENSG00000178913	0.91	2.0e-03	TAF7	ENSG00000085982	1.15	7.0e-03	USP40
ENSG00000143947	0.45	2.1e-03	RPS27A	ENSG00000091947	1.02	7.0e-03	TMEM101
ENSG00000167658	0.76	2.1e-03	EEF2	ENSG00000198515	2.02	7.6e-03	CNGA1
ENSG00000164125	2.73	2.1e-03	FAM198B	ENSG00000155962	1.28	7.6e-03	CLIC2
ENSG00000091073	1.96	2.2e-03	DTX2	ENSG00000163521	1.56	7.8e-03	GLB1L
ENSG00000003436	0.92	2.2e-03	TFPI	ENSG00000250722	1.69	7.9e-03	SELENOP
ENSG00000174032	1.34	2.2e-03	SLC25A30	ENSG00000270629	0.74	8.0e-03	NBPF14
ENSG00000160325	2.58	2.2e-03	CACFD1	ENSG00000177954	0.42	8.2e-03	RPS27
ENSG00000105641	1.89	2.3e-03	SLC5A5	ENSG00000120690	0.73	8.2e-03	ELF1
ENSG00000006042	2.05	2.3e-03	TMEM98	ENSG00000114391	0.39	8.5e-03	RPL24
ENSG00000140285	2.43	2.5e-03	FGF7	ENSG00000093167	0.78	8.5e-03	LRRFIP2
ENSG00000180730	2.32	2.5e-03	SHISA2	ENSG00000023909	1.13	8.6e-03	GCLM
ENSG00000120725	1.68	2.5e-03	SIL1	ENSG00000166780	1.70	8.8e-03	C16orf45
ENSG00000255112	1.05	2.6e-03	CHMP1B	ENSG00000112306	0.50	8.9e-03	RPS12
ENSG00000004139	1.70	2.7e-03	SARM1	ENSG00000114784	0.65	9.0e-03	EIF1B
ENSG00000187037	2.31	2.7e-03	GPR141	ENSG00000109452	1.47	9.0e-03	INPP4B
ENSG00000277075	1.48	2.8e-03	HIST1H2AE	ENSG00000187021	1.09	9.0e-03	PNLIPRP1
ENSG00000198763	0.49	3.0e-03	ND2	ENSG00000152219	0.72	9.1e-03	ARL14EP
ENSG00000187391	1.08	3.0e-03	MAGI2	ENSG00000100938	0.94	9.1e-03	GMPR2
ENSG00000164197	2.04	3.1e-03	RNF180	ENSG00000221944	0.65	9.4e-03	TIGD1
ENSG00000185155	1.71	3.1e-03	MIXL1	ENSG00000198839	0.68	9.4e-03	ZNF277
ENSG00000168769	0.96	3.2e-03	TET2	ENSG00000100380	0.74	9.5e-03	ST13
ENSG00000063046	0.44	3.2e-03	EIF4B	ENSG00000178562	0.85	9.6e-03	CD28
ENSG00000171490	0.87	3.2e-03	RSL1D1	ENSG00000099338	1.08	9.8e-03	CATSPERG
ENSG00000196632	2.02	3.2e-03	WNK3	ENSG00000237515	0.58	9.8e-03	SHISA9
ENSG00000111261	0.88	3.3e-03	MANSC1	ENSG00000174448	1.98	1.0e-02	STARD6
ENSG00000069667	1.74	3.4e-03	RORA	ENSG00000003393	1.35	1.0e-02	ALS2
ENSG00000121858	1.68	3.4e-03	TNFSF10	ENSG00000117036	1.51	1.0e-02	ETV3
ENSG00000214575	1.66	3.4e-03	CPEB1	ENSG00000174607	1.12	1.0e-02	UGT8
ENSG00000183166	1.94	3.5e-03	CALN1	ENSG00000134755	1.53	1.1e-02	DSC2
ENSG00000140600	2.57	3.6e-03	SH3GL3	ENSG00000111077	1.86	1.1e-02	TNS2
ENSG00000059377	1.00	3.6e-03	TBXAS1	ENSG00000185442	1.40	1.1e-02	FAM174B
ENSG00000120594	1.74	3.6e-03	PLXDC2	ENSG00000136052	2.27	1.1e-02	SLC41A2
ENSG00000106025	2.59	3.6e-03	TSPAN12	ENSG00000127952	0.91	1.1e-02	STYXL1
ENSG00000044115	1.30	3.6e-03	CTNNA1	ENSG00000120093	1.36	1.1e-02	HOXB3
ENSG00000139946	1.15	3.7e-03	PELI2	ENSG00000148057	1.17	1.1e-02	IDNK
ENSG00000117450	0.84	3.7e-03	PRDX1	ENSG00000060138	0.67	1.1e-02	YBX3
ENSG00000177875	2.67	3.7e-03	CCDC184	ENSG00000198765	1.28	1.1e-02	SYCP1
ENSG00000152128	2.38	3.8e-03	TMEM163	ENSG00000186472	2.30	1.1e-02	PCLO
ENSG00000102359	2.45	3.8e-03	SRPX2	ENSG00000254553	1.39	1.1e-02	NA
ENSG00000106034	1.92	3.8e-03	CPED1	ENSG00000100234	1.79	1.2e-02	TIMP3
ENSG00000175395	1.99	3.9e-03	ZNF25	ENSG00000122406	0.56	1.2e-02	RPL5
ENSG00000171747	1.76	3.9e-03	LGALS4	ENSG00000151150	1.51	1.2e-02	ANK3
ENSG00000075391	1.68	4.0e-03	RASAL2	ENSG00000151090	2.29	1.2e-02	THRB
ENSG00000055208	0.89	4.1e-03	TAB2	ENSG00000163565	0.85	1.2e-02	IFI16
ENSG00000182230	0.55	4.1e-03	FAM153B	ENSG00000104408	0.60	1.2e-02	EIF3E
ENSG00000174206	1.45	4.2e-03	C12orf66	ENSG00000281106	1.91	1.2e-02	LINC00282
ENSG00000196935	1.02	4.5e-03	SRGAP1	ENSG00000170889	0.40	1.2e-02	RPS9
ENSG00000072682	1.89	4.5e-03	P4HA2	ENSG00000144802	1.17	1.3e-02	NFKBIZ
ENSG00000117400	2.47	4.6e-03	MPL	ENSG00000176542	0.96	1.3e-02	USF3
ENSG00000104375	0.74	4.7e-03	STK3	ENSG00000144935	2.23	1.3e-02	TRPC1
ENSG00000226430	2.54	4.9e-03	USP17L7	ENSG00000132669	1.78	1.3e-02	RIN2
ENSG00000155158	1.59	1.3e-02	TTC39B	ENSG00000108018	1.36	2.6e-02	SORCS1
ENSG00000087076	2.37	1.3e-02	HSD17B14	ENSG00000095303	2.04	2.6e-02	PTGS1
ENSG00000184678	1.49	1.3e-02	HIST2H2BE	ENSG00000106546	1.25	2.6e-02	AHR
ENSG00000173812	0.79	1.4e-02	EIF1	ENSG00000092096	2.08	2.6e-02	SLC22A17
ENSG00000105137	2.33	1.4e-02	SYDE1	ENSG00000129226	1.55	2.6e-02	CD68
ENSG00000251322	1.48	1.4e-02	SHANK3	ENSG00000170421	1.12	2.6e-02	KRT8
ENSG00000107485	2.36	1.4e-02	GATA3	ENSG00000186393	1.11	2.7e-02	KRT26

Continuation of Table A.24

Ensembl	log_2 Fold Change	padj	Gene Symbol	Ensembl	log_2 Fold Change	padj	Gene Symbol
ENSG00000113389	1.09	1.4e-02	NPR3	ENSG00000120686	0.72	2.7e-02	UFM1
ENSG00000237651	1.13	1.4e-02	C2orf74	ENSG00000005187	1.23	4.4e-02	ACSM3
ENSG00000223443	2.31	1.4e-02	USP17L2	ENSG00000107771	1.11	4.4e-02	CCSER2
ENSG00000166681	0.59	1.4e-02	BEX3	ENSG00000134419	0.33	4.4e-02	RPS15A
ENSG00000113648	0.86	1.4e-02	H2AFY	ENSG00000181690	1.75	4.4e-02	PLAG1
ENSG00000164691	1.06	1.4e-02	TAGAP	ENSG00000080298	0.96	4.4e-02	RFX3
ENSG00000188859	1.58	1.4e-02	FAM78B	ENSG00000134853	1.48	4.4e-02	PDGFRA
ENSG00000132010	1.85	1.5e-02	ZNF20	ENSG00000081237	0.95	4.4e-02	PTPRC
ENSG00000086205	2.24	1.5e-02	FOLH1	ENSG00000105738	0.68	4.4e-02	SIPA1L3
ENSG00000106299	1.46	1.5e-02	WASL	ENSG00000177200	0.51	4.5e-02	CHD9
ENSG00000000457	1.21	1.5e-02	SCYL3	ENSG00000111674	1.17	4.5e-02	ENO2
ENSG00000137486	0.97	1.5e-02	ARRB1	ENSG00000177565	0.74	4.5e-02	TBL1XR1
ENSG00000178852	1.47	1.5e-02	EFCAB13	ENSG00000180432	1.32	4.5e-02	CYP8B1
ENSG00000174175	2.27	1.5e-02	SELP	ENSG00000167778	1.52	4.6e-02	SPRYD3
ENSG00000179144	1.86	1.5e-02	GIMAP7	ENSG00000078967	0.85	4.6e-02	UBE2D4
ENSG00000204065	2.32	1.5e-02	TCEAL5	ENSG00000139323	0.86	4.6e-02	POC1B
ENSG00000197746	0.87	1.5e-02	PSAP	ENSG00000101096	0.97	4.6e-02	NFATC2
ENSG00000254772	0.59	1.6e-02	EEF1G	ENSG00000130844	1.04	4.6e-02	ZNF331
ENSG00000283563	1.12	1.6e-02	ZCWPW2	ENSG00000143515	1.02	2.7e-02	ATP8B2
ENSG00000198189	0.80	1.6e-02	HSD17B11	ENSG00000143622	1.00	2.7e-02	RIT1
ENSG00000155366	1.87	1.7e-02	RHOC	ENSG00000169495	0.71	2.7e-02	HTRA4
ENSG00000104164	0.48	1.7e-02	BLOC1S6	ENSG00000136110	2.14	2.7e-02	LECT1
ENSG00000109046	0.56	1.7e-02	WSB1	ENSG00000151208	1.61	2.7e-02	DLG5
ENSG00000010404	1.47	1.7e-02	IDS	ENSG00000165626	1.58	2.7e-02	BEND7
ENSG00000008311	1.56	1.7e-02	AASS	ENSG00000170209	1.65	2.8e-02	ANKK1
ENSG00000099622	0.67	1.7e-02	CIRBP	ENSG00000101335	2.16	2.8e-02	MYL9
ENSG00000123427	1.42	1.7e-02	METTL21B	ENSG00000111679	0.97	2.8e-02	PTPN6
ENSG00000067066	0.65	1.7e-02	SP100	ENSG00000010278	1.81	2.8e-02	CD9
ENSG00000151914	1.32	1.7e-02	DST	ENSG00000134352	1.26	2.8e-02	IL6ST
ENSG00000164035	2.05	1.8e-02	EMCN	ENSG00000115561	0.78	2.9e-02	CHMP3
ENSG00000152207	1.54	1.8e-02	CYSLTR2	ENSG00000172578	0.96	2.9e-02	KLHL6
ENSG00000153162	2.29	1.8e-02	BMP6	ENSG00000101276	1.69	2.9e-02	SLC52A3
ENSG00000122026	0.44	1.8e-02	RPL21	ENSG00000152778	1.75	2.9e-02	IFIT5
ENSG00000131781	1.16	1.8e-02	FMO5	ENSG00000156508	0.54	2.9e-02	EEF1A1
ENSG00000236287	0.94	1.8e-02	ZBED5	ENSG00000162105	1.39	2.9e-02	SHANK2
ENSG00000268163	2.29	1.8e-02	NA	ENSG00000196531	0.33	2.9e-02	NACA
ENSG00000152527	1.38	1.8e-02	PLEKHH2	ENSG00000198467	1.81	2.9e-02	TPM2
ENSG00000107186	2.04	1.8e-02	MPDZ	ENSG00000168283	0.92	3.0e-02	BMI1
ENSG00000174749	0.52	1.9e-02	C4orf32	ENSG00000152952	1.18	3.0e-02	PLOD2
ENSG00000106246	0.90	1.9e-02	PTCD1	ENSG00000259120	1.67	3.0e-02	SMIM6
ENSG00000149131	1.01	1.9e-02	SERPING1	ENSG00000148143	1.38	3.0e-02	ZNF462
ENSG00000089916	0.56	1.9e-02	GPATCH2L	ENSG00000000971	2.00	3.0e-02	CFH
ENSG00000283632	0.98	1.9e-02	NA	ENSG00000173809	1.64	3.0e-02	TDRD12
ENSG00000046651	0.64	1.9e-02	OFD1	ENSG00000175197	1.29	3.1e-02	DDIT3
ENSG00000106617	0.83	1.9e-02	PRKAG2	ENSG00000185022	1.24	3.1e-02	MAFF
ENSG00000258130	1.49	1.9e-02	LOC613038	ENSG00000198918	0.40	3.1e-02	RPL39
ENSG00000065717	1.89	2.0e-02	TLE2	ENSG00000171657	1.31	3.1e-02	GPR82
ENSG00000182195	1.99	2.0e-02	LDOC1	ENSG00000204252	1.45	3.1e-02	HLA-DOA
ENSG00000145949	2.00	2.0e-02	MYLK4	ENSG00000154133	1.93	3.2e-02	ROBO4
ENSG00000123358	1.39	2.0e-02	NR4A1	ENSG00000106723	0.81	3.2e-02	SPIN1
ENSG00000154265	1.17	2.0e-02	ABCA5	ENSG00000146350	0.83	3.2e-02	TBC1D32
ENSG00000148175	0.94	2.0e-02	STOM	ENSG00000181007	0.97	3.2e-02	ZFP82
ENSG00000138109	1.26	2.0e-02	CYP2C9	ENSG00000017260	0.67	3.3e-02	ATP2C1
ENSG00000141756	2.20	2.1e-02	FKBP10	ENSG00000185250	1.46	3.3e-02	PPIL6
ENSG00000100307	1.64	2.1e-02	CBX7	ENSG00000197948	1.15	3.3e-02	FCHSD1
ENSG00000177575	0.61	2.1e-02	CD163	ENSG00000169756	0.87	3.4e-02	LIMS1
ENSG00000254122	2.21	2.1e-02	PCDHGB7	ENSG00000276966	0.92	3.4e-02	HIST1H4E
ENSG00000109654	1.94	2.1e-02	TRIM2	ENSG00000100767	1.21	3.4e-02	PAPLN
ENSG00000119684	0.98	2.1e-02	MLH3	ENSG00000118473	1.80	3.4e-02	SGIP1
ENSG00000168658	1.97	2.1e-02	VWA3B	ENSG00000183696	1.23	3.4e-02	UPP1
ENSG00000072952	1.97	2.2e-02	MRVI1	ENSG00000197958	0.53	3.4e-02	RPL12
ENSG00000111348	0.63	2.2e-02	ARHGDIB	ENSG00000075884	0.56	3.4e-02	ARHGAP15
ENSG00000138678	2.10	2.2e-02	GPAT3	ENSG00000128000	1.29	3.5e-02	ZNF780B
ENSG00000231500	0.49	2.2e-02	RPS18	ENSG00000205707	0.64	3.5e-02	ETFRF1
ENSG00000137642	1.34	2.2e-02	SORL1	ENSG00000152404	0.93	3.5e-02	CWF19L2
ENSG00000272325	0.60	2.2e-02	NUDT3	ENSG00000155816	1.10	3.5e-02	FMN2
ENSG00000087884	0.53	2.3e-02	AAMDC	ENSG00000008283	1.30	3.5e-02	CYB561
ENSG00000161835	1.85	2.4e-02	GRASP	ENSG00000256537	0.95	3.5e-02	SMIM10L1
ENSG00000196159	2.19	2.4e-02	FAT4	ENSG00000133112	0.50	3.6e-02	TPT1
ENSG00000132356	1.04	2.4e-02	PRKAA1	ENSG00000136854	1.66	3.6e-02	STXBP1
ENSG00000141504	1.17	2.4e-02	SAT2	ENSG00000166343	1.62	3.6e-02	MSS51
ENSG00000164989	1.02	2.4e-02	CCDC171	ENSG00000133401	1.43	3.6e-02	PDZD2
ENSG00000115875	0.51	2.5e-02	SRSF7	ENSG00000144567	0.57	3.6e-02	FAM134A
ENSG00000258818	1.75	2.5e-02	RNASE4	ENSG00000050426	0.70	3.7e-02	LETMD1
ENSG00000122642	1.55	2.5e-02	FKBP9	ENSG00000152785	1.64	3.7e-02	BMP3
ENSG00000142512	1.21	2.5e-02	SIGLEC10	ENSG00000158163	2.08	3.7e-02	DZIP1L
ENSG00000104231	0.95	2.5e-02	ZFAND1	ENSG00000274290	1.07	3.7e-02	HIST1H2BE
ENSG00000107185	1.09	4.1e-02	RGP1	ENSG00000156521	1.15	3.8e-02	TYSND1
ENSG00000149564	1.67	4.1e-02	ESAM	ENSG00000152582	1.03	3.8e-02	SPEF2
ENSG00000179083	2.05	4.1e-02	FAM133A	ENSG00000164587	0.38	3.8e-02	RPS14
ENSG00000203950	0.86	4.1e-02	FAM127B	ENSG00000102316	1.03	3.8e-02	MAGED2
ENSG00000253767	1.02	4.1e-02	PCDHGA8	ENSG00000198822	2.08	3.8e-02	GRM3
ENSG00000157873	1.91	4.2e-02	TNFRSF14	ENSG00000101901	0.69	3.8e-02	ALG13
ENSG00000197046	1.28	4.2e-02	SIGLEC15	ENSG00000067606	1.88	3.8e-02	PRKCZ
ENSG00000232629	1.59	4.2e-02	HLA-DQB2	ENSG00000116117	1.94	3.8e-02	PARD3B
ENSG00000104894	0.97	4.2e-02	CD37	ENSG00000154822	1.45	3.8e-02	PLCL2
ENSG00000124839	1.12	4.2e-02	RAB17	ENSG00000177697	1.01	3.9e-02	CD151
ENSG00000242950	0.65	4.2e-02	ERVW-1	ENSG00000217128	0.68	3.9e-02	FNIP1
ENSG00000106560	1.13	4.2e-02	GIMAP2	ENSG00000164070	1.02	3.9e-02	HSPA4L
ENSG00000104765	0.81	4.3e-02	BNIP3L	ENSG00000143952	1.00	3.9e-02	VPS54
ENSG00000137154	0.53	4.3e-02	RPS6	ENSG00000089692	2.00	3.9e-02	LAG3

Continuation of Table A.24

Ensembl	log_2 Fold Change	padj	Gene Symbol	Ensembl	log_2 Fold Change	padj	Gene Symbol
ENSG00000129187	0.67	4.3e-02	DCTD	ENSG00000120915	1.64	4.0e-02	EPHX2
ENSG00000170271	1.49	4.3e-02	FAXDC2	ENSG00000063177	0.35	4.0e-02	RPL18
ENSG00000137509	0.86	4.3e-02	PRCP	ENSG00000130921	0.65	4.0e-02	C12orf65
ENSG00000132718	1.01	4.6e-02	SYT11				
ENSG00000154188	1.18	4.7e-02	ANGPT1				
ENSG00000090376	0.85	4.7e-02	IRAK3				
ENSG00000163638	1.50	4.7e-02	ADAMTS9				
ENSG00000216937	1.28	4.7e-02	CCDC7				
ENSG00000051382	0.77	4.7e-02	PIK3CB				
ENSG00000171928	0.89	4.7e-02	TVP23B				
ENSG00000282988	1.35	4.7e-02	NA				
ENSG00000021645	1.25	4.8e-02	NRXN3				
ENSG00000130475	1.24	4.8e-02	FCHO1				
ENSG00000147601	0.53	4.8e-02	TERF1				
ENSG00000159592	0.73	4.8e-02	GPBP1L1				
ENSG00000151702	1.02	4.8e-02	FLI1				
ENSG00000119508	1.62	4.9e-02	NR4A3				
ENSG00000109771	1.22	4.9e-02	LRP2BP				
ENSG00000145425	0.52	4.9e-02	RPS3A				
ENSG00000243710	1.74	4.9e-02	CFAP57				
ENSG00000132879	1.24	4.9e-02	FBXO44				
ENSG00000061337	1.48	4.9e-02	LZTS1				
ENSG00000128590	0.92	4.9e-02	DNAJB9				
ENSG00000138079	0.95	4.9e-02	SLC3A1				
ENSG00000173258	0.85	4.9e-02	ZNF483				
ENSG00000186446	1.82	5.0e-02	ZNF501				
ENSG00000114107	1.10	5.0e-02	CEP70				
ENSG00000135124	0.92	5.0e-02	P2RX4				
ENSG00000102908	0.87	5.0e-02	NFAT5				
ENSG00000142541	0.27	5.0e-02	RPL13A				
ENSG00000102024	1.51	5.0e-02	PLS3				
ENSG00000186889	1.44	5.0e-02	TMEM17				
ENSG00000196104	1.89	5.0e-02	SPOCK3				
ENSG00000113532	0.84	5.0e-02	ST8SIA4				
ENSG00000138433	0.62	5.0e-02	CIR1				
ENSG00000086288	1.97	5.0e-02	NME8				

Table A.25: DEGs Between Normal LMPP-GMP and Leukaemic LMPP-GMP. DEGs were identified between normal LMPP and GMP and leukaemic LMPP and GMP, using a Wald test in DESeq2. A design formula of ~population + condition was used (where condition indicates normal/leukaemic status). Upregulated and downregulated DEGs with a p-adjusted value < 0.05 are shown.

Ensembl	$\log_2$ Fold Change	padj	Gene Symbol	Ensembl	$\log_2$ Fold Change	padj	Gene Symbol
ENSG00000123728	2.84	1.3e-13	RAP2C	ENSG00000243710	-3.26	2.4e-06	CFAP57
ENSG00000168092	2.15	4.0e-13	PAFAH1B2	ENSG00000113742	2.64	2.7e-06	CPEB4
ENSG00000197170	2.10	4.4e-13	PSMD12	ENSG00000135018	2.04	2.7e-06	UBQLN1
ENSG00000165410	2.87	5.4e-12	CFL2	ENSG00000116489	1.35	2.8e-06	CAPZA1
ENSG00000196549	-3.26	6.5e-12	MME	ENSG00000276043	-2.20	2.9e-06	UHRF1
ENSG00000169641	2.28	1.3e-11	LUZP1	ENSG00000152256	2.58	3.7e-06	PDK1
ENSG00000205189	2.61	7.1e-11	ZBTB10	ENSG00000135679	1.24	4.0e-06	MDM2
ENSG00000076053	1.71	1.4e-10	RBM7	ENSG00000133703	1.94	4.0e-06	KRAS
ENSG00000110367	1.69	3.7e-10	DDX6	ENSG00000162923	2.30	4.6e-06	WDR26
ENSG00000172239	2.28	3.7e-10	PAIP1	ENSG00000084463	1.65	5.3e-06	WBP11
ENSG00000112200	2.43	3.7e-10	ZNF451	ENSG00000196352	1.89	5.7e-06	CD55
ENSG00000164463	2.53	3.7e-10	CREBRF	ENSG00000132475	2.25	5.7e-06	H3F3B
ENSG00000151332	2.00	1.5e-09	MBIP	ENSG00000115084	1.65	5.9e-06	SLC35F5
ENSG00000173812	1.82	1.6e-09	EIF1	ENSG00000129484	1.83	5.9e-06	PARP2
ENSG00000162971	1.99	2.0e-09	TYW5	ENSG00000157106	1.34	5.9e-06	SMG1
ENSG00000114784	2.41	6.8e-09	EIF1B	ENSG00000182831	1.86	5.9e-06	C16orf72
ENSG00000126524	2.59	6.8e-09	SBDS	ENSG00000138750	1.42	6.0e-06	NUP54
ENSG00000109332	1.71	7.8e-09	UBE2D3	ENSG00000013441	2.04	6.2e-06	CLK1
ENSG00000147650	2.50	1.0e-08	LRP12	ENSG00000068745	1.20	6.5e-06	IP6K2
ENSG00000175104	2.88	1.0e-08	TRAF6	ENSG00000109686	2.29	6.5e-06	SH3D19
ENSG00000180667	2.78	1.0e-08	YOD1	ENSG00000163466	1.16	6.5e-06	ARPC2
ENSG00000179454	2.41	1.4e-08	KLHL28	ENSG00000100519	1.43	6.6e-06	PSMC6
ENSG00000139620	1.99	2.2e-08	KANSL2	ENSG00000010244	1.42	7.1e-06	ZNF207
ENSG00000152242	2.77	3.6e-08	C18orf25	ENSG00000022840	1.74	7.1e-06	RNF10
ENSG00000029363	1.53	4.7e-08	BCLAF1	ENSG00000009307	1.36	7.3e-06	CSDE1
ENSG00000163444	1.52	5.1e-08	TMEM183A	ENSG00000255529	2.02	7.4e-06	POLR2M
ENSG00000119396	1.70	6.1e-08	RAB14	ENSG00000165525	0.92	7.6e-06	NEMF
ENSG00000165637	1.32	6.7e-08	VDAC2	ENSG00000196504	1.34	8.1e-06	PRPF40A
ENSG00000100644	2.81	6.9e-08	HIF1A	ENSG00000140941	2.32	8.5e-06	MAP1LC3B
ENSG00000082153	1.89	7.3e-08	BZW1	ENSG00000278311	1.38	9.0e-06	GGNBP2
ENSG00000178150	-2.31	7.3e-08	ZNF114	ENSG00000198961	1.64	9.1e-06	PJA2
ENSG00000162616	2.61	8.1e-08	DNAJB4	ENSG00000205838	-3.78	9.1e-06	TTC23L
ENSG00000100614	2.14	8.8e-08	PPM1A	ENSG00000090061	1.69	9.2e-06	CCNK
ENSG00000101782	2.42	8.8e-08	RIOK3	ENSG00000113734	1.93	9.3e-06	BNIP1
ENSG00000114978	1.59	8.8e-08	MOB1A	ENSG00000165312	2.87	9.3e-06	OTUD1
ENSG00000162664	1.59	8.8e-08	ZNF326	ENSG00000174469	-1.79	9.3e-06	CNTNAP2
ENSG00000196470	2.16	8.8e-08	SIAH1	ENSG00000107819	2.32	9.6e-06	SFXN3
ENSG00000166200	2.25	1.0e-07	COPS2	ENSG00000215301	1.35	9.6e-06	DDX3X
ENSG00000034677	1.79	1.2e-07	RNF19A	ENSG00000116747	1.53	9.6e-06	TROVE2
ENSG00000176853	2.24	1.4e-07	FAM91A1	ENSG00000137876	1.50	9.6e-06	RSL24D1
ENSG00000114405	-4.34	1.7e-07	C3orf14	ENSG00000139112	2.51	9.7e-06	GABARAPL1
ENSG00000215251	2.09	2.2e-07	FASTKD5	ENSG00000166046	2.39	1.0e-05	TCP11L2
ENSG00000169251	1.87	2.4e-07	NMD3	ENSG00000187735	1.03	1.0e-05	TCEA1
ENSG00000138757	1.65	2.4e-07	G3BP2	ENSG00000189241	1.28	1.0e-05	TSPYL1
ENSG00000140367	2.15	2.7e-07	UBE2Q2	ENSG00000120616	1.32	1.1e-05	EPC1
ENSG00000069956	1.94	2.9e-07	MAPK6	ENSG00000137962	1.96	1.2e-05	ARHGAP29
ENSG00000132680	1.79	3.1e-07	KIAA0907	ENSG00000135114	3.29	1.3e-05	OASL
ENSG00000197747	3.81	3.2e-07	S100A10	ENSG00000141682	2.63	1.4e-05	PMAIP1
ENSG00000279765	2.18	3.5e-07	NA	ENSG00000101290	1.73	1.6e-05	CDS2
ENSG00000110852	2.59	3.6e-07	CLEC2B	ENSG00000184838	-3.13	1.6e-05	PRR16
ENSG00000023734	1.73	3.8e-07	STRAP	ENSG00000100647	1.52	1.7e-05	SUSD6
ENSG00000156735	1.67	3.9e-07	BAG4	ENSG00000096401	1.65	1.8e-05	CDC5L
ENSG00000116030	1.44	4.0e-07	SUMO1	ENSG00000139433	1.29	1.8e-05	GLTP
ENSG00000123091	2.38	4.0e-07	RNF11	ENSG00000197561	-3.69	1.8e-05	ELANE
ENSG00000137947	1.95	4.0e-07	GTF2B	ENSG00000125772	1.84	1.9e-05	GPCPD1
ENSG00000108654	1.35	4.3e-07	DDX5	ENSG00000184007	1.08	1.9e-05	PTP4A2
ENSG00000070831	1.89	4.6e-07	CDC42	ENSG00000162511	1.81	1.9e-05	LAPTM5
ENSG00000116209	1.07	4.6e-07	TMEM59	ENSG00000153201	1.70	1.9e-05	RANBP2
ENSG00000067248	1.47	4.7e-07	DHX29	ENSG00000120438	1.21	2.3e-05	TCP1
ENSG00000168175	1.50	4.7e-07	MAPK1IP1L	ENSG00000121879	2.00	2.3e-05	PIK3CA
ENSG00000179021	1.67	4.7e-07	C3orf38	ENSG00000132356	1.71	2.3e-05	PRKAA1
ENSG00000116731	1.83	6.2e-07	PRDM2	ENSG00000097033	1.25	2.3e-05	SH3GLB1
ENSG00000078747	2.18	6.2e-07	ITCH	ENSG00000138669	-2.57	2.3e-05	PRKG2
ENSG00000079332	2.06	6.5e-07	SAR1A	ENSG00000138593	1.84	2.4e-05	SECISBP2L
ENSG00000182481	2.02	8.0e-07	KPNA2	ENSG00000181467	1.75	2.5e-05	RAP2B
ENSG00000137055	1.76	8.3e-07	PLAA	ENSG00000134684	1.02	2.5e-05	YARS
ENSG00000154447	2.37	8.3e-07	SH3RF1	ENSG00000111252	2.11	2.6e-05	SH2B3
ENSG00000185885	2.79	8.3e-07	IFITM1	ENSG00000114796	2.23	2.6e-05	KLHL24
ENSG00000147872	2.85	8.6e-07	PLIN2	ENSG00000165282	-2.13	2.6e-05	PIGO
ENSG00000159346	1.80	1.0e-06	ADIPOR1	ENSG00000242550	-2.75	2.7e-05	SERPINB10
ENSG00000138069	1.62	1.1e-06	RAB1A	ENSG00000164690	-3.19	2.8e-05	SHH
ENSG00000273841	1.59	1.1e-06	TAF9	ENSG00000117614	1.56	2.8e-05	SYF2
ENSG00000083799	2.01	1.4e-06	CYLD	ENSG00000136381	1.21	2.9e-05	IREB2
ENSG00000132912	1.53	1.5e-06	DCTN4	ENSG00000150991	1.65	2.9e-05	UBC
ENSG00000137965	2.24	1.6e-06	IFI44	ENSG00000111615	1.49	3.0e-05	KRR1
ENSG00000005812	2.14	1.6e-06	FBXL3	ENSG00000144566	1.55	3.1e-05	RAB5A
ENSG00000169155	2.52	1.6e-06	ZBTB43	ENSG00000107951	1.45	3.1e-05	MTPAP
ENSG00000278053	1.52	1.6e-06	DDX52	ENSG00000065548	1.37	3.3e-05	ZC3H15
ENSG00000114742	1.80	1.7e-06	WDR48	ENSG00000163636	1.15	3.3e-05	PSMD6
ENSG00000135655	1.48	1.7e-06	USP15	ENSG00000177311	-1.70	3.3e-05	ZBTB38
ENSG00000184203	1.77	1.9e-06	PPP1R2	ENSG00000198492	1.14	3.5e-05	YTHDF2
ENSG00000091317	1.68	1.9e-06	CMTM6	ENSG00000109270	1.58	3.5e-05	LAMTOR3

Continuation of Table A.24

Ensembl	log_2 Fold Change	padj	Gene Symbol	Ensembl	log_2 Fold Change	padj	Gene Symbol
ENSG00000152332	1.58	2.2e-06	UHMK1	ENSG00000137040	2.10	3.6e-05	RANBP6
ENSG00000143761	1.34	2.4e-06	ARF1	ENSG00000143622	2.30	3.7e-05	RIT1
ENSG00000122692	0.76	3.7e-05	SMU1	ENSG00000135900	1.29	1.8e-04	MRPL44
ENSG00000155508	1.54	3.7e-05	CNOT8	ENSG00000147100	-1.79	1.8e-04	SLC16A2
ENSG00000122482	1.11	3.7e-05	ZNF644	ENSG00000156671	1.87	1.8e-04	SAMD8
ENSG00000143549	1.14	4.0e-05	TPM3	ENSG00000168610	1.19	1.8e-04	STAT3
ENSG00000162702	1.43	4.0e-05	ZNF281	ENSG00000104856	2.71	1.9e-04	RELB
ENSG00000185721	1.15	4.1e-05	DRG1	ENSG00000067334	1.04	1.9e-04	DNTTIP2
ENSG00000163660	1.42	4.2e-05	CCNL1	ENSG00000092108	0.93	1.9e-04	SCFD1
ENSG00000178301	3.06	4.5e-05	AQP11	ENSG00000121749	1.40	1.9e-04	TBC1D15
ENSG00000067064	1.78	4.7e-05	IDI1	ENSG00000105849	1.26	1.9e-04	TWISTNB
ENSG00000136451	1.39	4.8e-05	VEZF1	ENSG00000179119	1.63	1.9e-04	SPTY2D1
ENSG00000155545	2.02	4.8e-05	MIER3	ENSG00000179869	-2.17	1.9e-04	ABCA13
ENSG00000115091	1.24	4.8e-05	ACTR3	ENSG00000213281	1.44	1.9e-04	NRAS
ENSG00000152556	-2.27	5.1e-05	PFKM	ENSG00000151883	1.90	2.0e-04	PARP8
ENSG00000189079	1.50	5.1e-05	ARID2	ENSG00000128322	-2.74	2.0e-04	IGLL1
ENSG00000186184	1.05	5.5e-05	POLR1D	ENSG00000116754	1.04	2.1e-04	SRSF11
ENSG00000113851	1.54	5.6e-05	CRBN	ENSG00000175197	2.50	2.1e-04	DDIT3
ENSG00000215717	1.52	5.6e-05	TMEM167B	ENSG00000054267	1.48	2.3e-04	ARID4B
ENSG00000184922	1.59	5.7e-05	FMNL1	ENSG00000169629	1.84	2.3e-04	RGPD8
ENSG00000086666	1.32	5.8e-05	ZFAND6	ENSG00000077721	1.43	2.4e-04	UBE2A
ENSG00000115540	1.73	5.8e-05	MOB4	ENSG00000114416	0.90	2.4e-04	FXR1
ENSG00000117139	1.60	5.9e-05	KDM5B	ENSG00000198265	1.10	2.4e-04	HELZ
ENSG00000174485	1.92	6.2e-05	DENND4A	ENSG00000277726	1.52	2.4e-04	NA
ENSG00000188529	1.07	6.2e-05	SRSF10	ENSG00000120063	1.71	2.4e-04	GNA13
ENSG00000068697	1.32	6.2e-05	LAPTM4A	ENSG00000059728	2.04	2.4e-04	MXD1
ENSG00000154451	2.49	6.4e-05	GBP5	ENSG00000131171	1.04	2.4e-04	SH3BGRL
ENSG00000136758	1.08	6.5e-05	YME1L1	ENSG00000090060	1.08	2.4e-04	PAPOLA
ENSG00000101166	1.66	6.7e-05	PRELID3B	ENSG00000155957	-1.08	2.4e-04	TMBIM4
ENSG00000145780	2.07	6.7e-05	FEM1C	ENSG00000085832	1.18	2.5e-04	EPS15
ENSG00000239305	2.04	6.7e-05	RNF103	ENSG00000184900	1.24	2.6e-04	SUMO3
ENSG00000081154	1.38	6.7e-05	PCNP	ENSG00000119446	1.62	2.6e-04	RBM18
ENSG00000261857	2.97	6.8e-05	MIA	ENSG00000158615	1.52	2.6e-04	PPP1R15B
ENSG00000115524	0.86	6.8e-05	SF3B1	ENSG00000137135	-1.64	2.7e-04	ARHGEF39
ENSG00000198160	1.96	6.9e-05	MIER1	ENSG00000150316	1.21	2.7e-04	CWC15
ENSG00000283467	2.22	7.3e-05	FAM65B	ENSG00000108381	1.25	2.8e-04	ASPA
ENSG00000066654	1.33	7.3e-05	THUMPD1	ENSG00000162819	1.72	2.8e-04	BROX
ENSG00000102753	1.80	7.6e-05	KPNA3	ENSG00000136436	1.51	2.9e-04	CALCOCO2
ENSG00000167461	0.99	7.7e-05	RAB8A	ENSG00000095951	1.65	2.9e-04	HIVEP1
ENSG00000143751	1.80	7.8e-05	SDE2	ENSG00000127903	-1.56	2.9e-04	ZNF835
ENSG00000177839	1.28	7.9e-05	PCDHB9	ENSG00000183283	1.31	2.9e-04	DAZAP2
ENSG00000180979	1.27	7.9e-05	LRRC57	ENSG00000213145	2.62	2.9e-04	CRIP1
ENSG00000117523	0.90	8.0e-05	PRRC2C	ENSG00000054654	1.04	3.0e-04	SYNE2
ENSG00000101417	-1.25	8.1e-05	PXMP4	ENSG00000089818	1.54	3.1e-04	NECAP1
ENSG00000062194	1.28	9.5e-05	GPBP1	ENSG00000130830	1.27	3.1e-04	MPP1
ENSG00000165219	1.36	9.5e-05	GAPVD1	ENSG00000198898	1.35	3.1e-04	CAPZA2
ENSG00000272886	1.94	9.7e-05	DCP1A	ENSG00000101624	1.45	3.1e-04	CEP76
ENSG00000258890	1.31	9.8e-05	CEP95	ENSG00000129083	1.37	3.1e-04	COPB1
ENSG00000188487	-2.46	1.1e-04	INSC	ENSG00000067066	0.90	3.1e-04	SP100
ENSG00000099246	1.76	1.1e-04	RAB18	ENSG00000141401	-1.79	3.2e-04	IMPA2
ENSG00000067560	1.09	1.1e-04	RHOA	ENSG00000183255	1.17	3.2e-04	PTTG1IP
ENSG00000172007	2.24	1.1e-04	RAB33B	ENSG00000156304	1.32	3.2e-04	SCAF4
ENSG00000119725	1.56	1.1e-04	ZNF410	ENSG00000164743	-2.66	3.3e-04	C8orf48
ENSG00000156232	1.81	1.1e-04	WHAMM	ENSG00000095574	1.74	3.4e-04	IKZF5
ENSG00000205084	-1.80	1.2e-04	TMEM231	ENSG00000169762	1.66	3.4e-04	TAPT1
ENSG00000082258	1.63	1.2e-04	CCNT2	ENSG00000111596	0.93	3.6e-04	CNOT2
ENSG00000085117	1.33	1.2e-04	CD82	ENSG00000048405	1.84	3.6e-04	ZNF800
ENSG00000137502	3.18	1.3e-04	RAB30	ENSG00000114120	1.42	3.6e-04	SLC25A36
ENSG00000197579	1.65	1.3e-04	TOPORS	ENSG00000169826	1.56	3.6e-04	CSGALNACT2
ENSG00000169239	1.68	1.3e-04	CA5B	ENSG00000066557	1.61	3.6e-04	LRRC40
ENSG00000256229	-1.41	1.3e-04	ZNF486	ENSG00000168438	1.38	3.6e-04	CDC40
ENSG00000113387	1.12	1.3e-04	SUB1	ENSG00000137814	1.13	3.6e-04	HAUS2
ENSG00000123983	1.95	1.3e-04	ACSL3	ENSG00000101654	1.37	3.7e-04	RNMT
ENSG00000119231	1.38	1.4e-04	SENP5	ENSG00000004961	1.94	3.8e-04	HCCS
ENSG00000111711	2.19	1.4e-04	GOLT1B	ENSG00000178127	1.33	3.8e-04	NDUFV2
ENSG00000111647	1.87	1.4e-04	UHRF1BP1L	ENSG00000136504	1.27	3.9e-04	KAT7
ENSG00000112339	1.37	1.4e-04	HBS1L	ENSG00000177694	-2.31	3.9e-04	NAALADL2
ENSG00000095485	1.43	1.4e-04	CWF19L1	ENSG00000117226	1.82	4.0e-04	GBP3
ENSG00000129226	2.27	1.4e-04	CD68	ENSG00000261740	1.10	4.1e-04	BOLA2-SMG1P6
ENSG00000139233	1.03	1.4e-04	LLPH	ENSG00000112343	1.31	4.2e-04	TRIM38
ENSG00000275895	2.71	1.4e-04	U2AF1L5	ENSG00000122257	1.31	4.2e-04	RBBP6
ENSG00000120686	1.48	1.5e-04	UFM1	ENSG00000158604	1.12	4.2e-04	TMED4
ENSG00000196422	-3.08	1.5e-04	PPP1R26	ENSG00000164924	1.31	4.4e-04	YWHAZ
ENSG00000112308	1.41	1.5e-04	C6orf62	ENSG00000133460	-1.01	4.4e-04	SLC2A11
ENSG00000062716	1.94	1.5e-04	VMP1	ENSG00000153922	1.73	4.5e-04	CHD1
ENSG00000130066	1.96	1.5e-04	SAT1	ENSG00000101558	1.05	4.5e-04	VAPA
ENSG00000093167	1.48	1.5e-04	LRRFIP2	ENSG00000140332	1.33	4.5e-04	TLE3
ENSG00000067900	0.97	1.6e-04	ROCK1	ENSG00000184860	-1.81	4.6e-04	SDR42E1
ENSG00000146457	1.52	1.7e-04	WTAP	ENSG00000186205	-1.08	4.6e-04	Mar-01
ENSG00000183696	1.71	1.7e-04	UPP1	ENSG00000083828	2.13	4.6e-04	ZNF586
ENSG00000112282	1.32	1.7e-04	MED23	ENSG00000152944	1.40	4.8e-04	MED21
ENSG00000144597	2.11	1.7e-04	EAF1	ENSG00000197872	2.77	4.8e-04	FAM49A
ENSG00000033122	2.16	1.7e-04	LRRC7	ENSG00000170185	1.46	4.8e-04	USP38
ENSG00000134265	1.33	1.7e-04	NAPG	ENSG00000159377	0.96	4.8e-04	PSMB4
ENSG00000132950	1.83	1.7e-04	ZMYM5	ENSG00000169905	1.34	4.8e-04	TOR1AIP2
ENSG00000101367	1.36	1.8e-04	MAPRE1	ENSG00000186075	2.04	4.9e-04	ZPBP2
ENSG00000101557	1.33	1.8e-04	USP14	ENSG00000182197	-2.63	5.0e-04	EXT1
ENSG00000100941	0.91	5.0e-04	PNN	ENSG00000162613	1.07	1.1e-03	FUBP1
ENSG00000151893	1.43	5.0e-04	CACUL1	ENSG00000005381	-2.90	1.1e-03	MPO
ENSG00000115520	1.36	5.1e-04	COQ10B	ENSG00000009724	2.54	1.1e-03	MASP2

Continuation of Table A.24

Ensembl	log_2 Fold Change	padj	Gene Symbol	Ensembl	log_2 Fold Change	padj	Gene Symbol
ENSG00000163531	-0.94	5.3e-04	NFASC	ENSG00000139679	2.68	1.1e-03	LPAR6
ENSG00000123505	1.49	5.3e-04	AMD1	ENSG00000213516	1.70	1.1e-03	RBMXL1
ENSG00000213625	1.58	5.3e-04	LEPROT	ENSG00000147251	1.14	1.1e-03	DOCK11
ENSG00000112531	1.22	5.5e-04	QKI	ENSG00000179954	2.40	1.1e-03	SSC5D
ENSG00000104312	1.90	5.5e-04	RIPK2	ENSG00000196233	0.89	1.1e-03	C10orf12
ENSG00000131876	1.14	5.5e-04	SNRPA1	ENSG00000164035	-2.02	1.1e-03	EMCN
ENSG00000256683	1.57	5.5e-04	ZNF350	ENSG00000133773	1.00	1.1e-03	CCDC59
ENSG00000129518	1.13	5.5e-04	EAPP	ENSG00000167996	1.90	1.1e-03	FTH1
ENSG00000136754	1.17	5.6e-04	ABI1	ENSG00000205571	1.29	1.1e-03	SMN1
ENSG00000153898	-1.57	5.6e-04	MCOLN2	ENSG00000153914	1.01	1.2e-03	SREK1
ENSG00000163521	-1.58	5.6e-04	GLB1L	ENSG00000180228	1.15	1.2e-03	PRKRA
ENSG00000114933	1.04	5.6e-04	INO80D	ENSG00000205302	1.29	1.2e-03	SNX2
ENSG00000134352	2.31	5.7e-04	IL6ST	ENSG00000138674	0.70	1.2e-03	SEC31A
ENSG00000120705	1.61	5.7e-04	ETF1	ENSG00000129515	1.21	1.2e-03	SNX6
ENSG00000133657	1.67	5.7e-04	ATP13A3	ENSG00000175931	-1.97	1.2e-03	UBE2O
ENSG00000152380	1.39	5.7e-04	FAM151B	ENSG00000138107	1.05	1.2e-03	ACTR1A
ENSG00000155330	1.39	5.7e-04	C16orf87	ENSG00000137959	2.57	1.2e-03	IFI44L
ENSG00000164898	-1.57	5.7e-04	FMC1	ENSG00000119689	1.24	1.3e-03	DLST
ENSG00000170852	1.55	5.7e-04	KBTBD2	ENSG00000213639	1.31	1.3e-03	PPP1CB
ENSG00000176542	1.35	5.7e-04	USF3	ENSG00000168961	1.11	1.3e-03	LGALS9
ENSG00000106052	1.34	5.8e-04	TAX1BP1	ENSG00000114127	1.33	1.3e-03	XRN1
ENSG00000111011	1.08	5.8e-04	RSRC2	ENSG00000178074	1.46	1.3e-03	C2orf69
ENSG00000189266	1.44	5.8e-04	PNRC2	ENSG00000107104	-1.68	1.3e-03	KANK1
ENSG00000111832	0.74	5.8e-04	RWDD1	ENSG00000147231	-2.55	1.3e-03	CXorf57
ENSG00000131263	1.30	6.0e-04	RLIM	ENSG00000168301	2.12	1.3e-03	KCTD6
ENSG00000143862	1.63	6.2e-04	ARL8A	ENSG00000174010	1.77	1.3e-03	KLHL15
ENSG00000177084	-1.39	6.2e-04	POLE	ENSG00000104765	1.32	1.3e-03	BNIP3L
ENSG00000244405	2.63	6.2e-04	ETV5	ENSG00000113328	1.65	1.4e-03	CCNG1
ENSG00000137776	0.79	6.4e-04	SLTM	ENSG00000138279	0.79	1.4e-03	ANXA7
ENSG00000055332	1.18	6.4e-04	EIF2AK2	ENSG00000154864	2.46	1.4e-03	PIEZO2
ENSG00000276966	1.35	6.5e-04	HIST1H4E	ENSG00000134109	1.43	1.4e-03	EDEM1
ENSG00000148943	1.21	6.7e-04	LIN7C	ENSG00000136143	1.61	1.4e-03	SUCLA2
ENSG00000185728	1.65	6.8e-04	YTHDF3	ENSG00000171103	1.08	1.4e-03	TRMT61B
ENSG00000113163	1.42	6.9e-04	COL4A3BP	ENSG00000197045	1.39	1.4e-03	GMFB
ENSG00000151092	1.06	7.1e-04	NGLY1	ENSG00000068912	1.54	1.4e-03	ERLEC1
ENSG00000169018	1.28	7.3e-04	FEM1B	ENSG00000136986	0.96	1.4e-03	DERL1
ENSG00000159140	1.07	7.4e-04	SON	ENSG00000169045	1.28	1.4e-03	HNRNPH1
ENSG00000077254	0.86	7.4e-04	USP33	ENSG00000198408	0.98	1.4e-03	MGEA5
ENSG00000019995	2.13	7.6e-04	ZRANB1	ENSG00000055732	1.34	1.4e-03	MCOLN3
ENSG00000023287	1.44	7.7e-04	RB1CC1	ENSG00000168876	1.67	1.5e-03	ANKRD49
ENSG00000071243	1.48	7.8e-04	ING3	ENSG00000134900	1.31	1.5e-03	TPP2
ENSG00000139973	1.49	8.0e-04	SYT16	ENSG00000141446	1.23	1.5e-03	ESCO1
ENSG00000137575	1.46	8.2e-04	SDCBP	ENSG00000136527	1.09	1.5e-03	TRA2B
ENSG00000129315	1.29	8.3e-04	CCNT1	ENSG00000198612	1.25	1.5e-03	COPS8
ENSG00000124688	1.59	8.4e-04	MAD2L1BP	ENSG00000129460	1.01	1.5e-03	NGDN
ENSG00000134107	2.05	8.4e-04	BHLHE40	ENSG00000109684	-2.14	1.5e-03	CLNK
ENSG00000165626	-1.91	8.4e-04	BEND7	ENSG00000152601	0.96	1.5e-03	MBNL1
ENSG00000091164	1.37	8.4e-04	TXNL1	ENSG00000166266	1.24	1.5e-03	CUL5
ENSG00000182013	-1.46	8.4e-04	PNMAL1	ENSG00000175356	-2.07	1.5e-03	SCUBE2
ENSG00000150593	1.33	8.6e-04	PDCD4	ENSG00000146729	1.16	1.5e-03	GBAS
ENSG00000175073	1.43	8.6e-04	VCPIP1	ENSG00000113712	1.21	1.6e-03	CSNK1A1
ENSG00000170915	-2.32	8.6e-04	PAQR8	ENSG00000117133	1.24	1.6e-03	RPF1
ENSG00000181982	-2.00	8.6e-04	CCDC149	ENSG00000172318	2.07	1.6e-03	B3GALT1
ENSG00000183090	-2.86	8.8e-04	FREM3	ENSG00000269897	2.12	1.6e-03	COMMD3-BMI1
ENSG00000182944	0.86	8.8e-04	EWSR1	ENSG00000111371	1.38	1.6e-03	SLC38A1
ENSG00000124486	1.05	8.8e-04	USP9X	ENSG00000118482	0.96	1.6e-03	PHF3
ENSG00000134758	1.59	8.8e-04	RNF138	ENSG00000051523	2.05	1.7e-03	CYBA
ENSG00000217128	1.53	8.8e-04	FNIP1	ENSG00000127993	1.13	1.7e-03	RBM48
ENSG00000198380	1.51	8.9e-04	GFPT1	ENSG00000144366	1.30	1.7e-03	GULP1
ENSG00000116752	1.37	9.3e-04	BCAS2	ENSG00000181555	1.21	1.7e-03	SETD2
ENSG00000185630	-1.70	9.3e-04	PBX1	ENSG00000196083	1.98	1.7e-03	IL1RAP
ENSG00000187837	1.71	9.4e-04	HIST1H1C	ENSG00000173575	1.27	1.7e-03	CHD2
ENSG00000188647	1.46	9.4e-04	PTAR1	ENSG00000198189	1.03	1.7e-03	HSD17B11
ENSG00000111897	1.38	9.5e-04	SERINC1	ENSG00000116132	-2.72	1.8e-03	PRRX1
ENSG00000006634	1.31	9.9e-04	DBF4	ENSG00000107438	2.44	1.8e-03	PDLIM1
ENSG00000081479	-2.29	9.9e-04	LRP2	ENSG00000126814	-1.92	1.8e-03	TRMT5
ENSG00000102805	0.93	9.9e-04	CLN5	ENSG00000109819	-2.13	1.8e-03	PPARGC1A
ENSG00000142227	1.66	9.9e-04	EMP3	ENSG00000120690	1.64	1.8e-03	ELF1
ENSG00000112245	1.49	9.9e-04	PTP4A1	ENSG00000147145	2.60	1.8e-03	LPAR4
ENSG00000196850	1.51	9.9e-04	PPTC7	ENSG00000143443	1.82	1.9e-03	C1orf56
ENSG00000135334	1.55	1.0e-03	AKIRIN2	ENSG00000186288	2.48	1.9e-03	PABPC1L2A
ENSG00000156030	1.50	1.0e-03	ELMSAN1	ENSG00000198887	1.28	1.9e-03	SMC5
ENSG00000259511	1.49	1.0e-03	UBE2Q2L	ENSG00000133997	1.07	1.9e-03	MED6
ENSG00000105669	1.29	1.0e-03	COPE	ENSG00000151461	0.77	1.9e-03	UPF2
ENSG00000136819	1.02	1.0e-03	C9orf78	ENSG00000162695	1.31	1.9e-03	SLC30A7
ENSG00000145414	1.09	1.0e-03	NAF1	ENSG00000168778	-1.27	1.9e-03	TCTN2
ENSG00000165997	1.76	1.0e-03	ARL5B	ENSG00000172243	2.46	1.9e-03	CLEC7A
ENSG00000174776	2.35	1.0e-03	WDR49	ENSG00000119922	2.54	1.9e-03	IFIT2
ENSG00000163960	1.44	1.0e-03	UBXN7	ENSG00000151239	1.53	1.9e-03	TWF1
ENSG00000185787	0.81	1.0e-03	MORF4L1	ENSG00000138434	1.79	1.9e-03	SSFA2
ENSG00000100225	1.06	1.0e-03	FBXO7	ENSG00000138802	1.31	1.9e-03	SEC24B
ENSG00000112149	2.13	1.1e-03	CD83	ENSG00000135919	-2.05	1.9e-03	SERPINE2
ENSG00000138433	1.07	1.9e-03	CIR1	ENSG00000114316	0.77	3.2e-03	USP4
ENSG00000168288	1.04	1.9e-03	MMADHC	ENSG00000164031	1.28	3.3e-03	DNAJB14
ENSG00000116560	1.02	1.9e-03	SFPQ	ENSG00000164327	1.37	3.3e-03	RICTOR
ENSG00000113384	-1.54	2.0e-03	GOLPH3	ENSG00000279906	0.83	3.3e-03	NA
ENSG00000167470	1.44	2.0e-03	MIDN	ENSG00000139946	1.55	3.4e-03	PELI2
ENSG00000138385	0.66	2.0e-03	SSB	ENSG00000124193	1.00	3.4e-03	SRSF6
ENSG00000137267	2.73	2.0e-03	TUBB2A	ENSG00000147894	1.51	3.4e-03	C9orf72
ENSG00000025800	1.12	2.0e-03	KPNA6	ENSG00000135164	1.23	3.4e-03	DMTF1

Continuation of Table A.24

Ensembl	log_2 Fold Change	padj	Gene Symbol	Ensembl	log_2 Fold Change	padj	Gene Symbol
ENSG00000234545	0.89	2.0e-03	FAM133B	ENSG00000120833	2.33	3.5e-03	SOCS2
ENSG00000188375	2.21	2.0e-03	H3F3C	ENSG00000179361	1.66	3.5e-03	ARID3B
ENSG00000134759	1.17	2.0e-03	ELP2	ENSG00000158796	1.16	3.5e-03	DEDD
ENSG00000172113	-1.38	2.0e-03	NME6	ENSG00000275591	-0.92	3.5e-03	XKR5
ENSG00000265354	0.89	2.0e-03	TIMM23	ENSG00000081237	1.28	3.5e-03	PTPRC
ENSG00000106686	0.81	2.0e-03	SPATA6L	ENSG00000150281	-1.66	3.6e-03	CTF1
ENSG00000167419	-2.02	2.1e-03	LPO	ENSG00000117115	-2.08	3.6e-03	PADI2
ENSG00000124104	-1.85	2.1e-03	SNX21	ENSG00000119917	2.34	3.6e-03	IFIT3
ENSG00000036257	1.21	2.1e-03	CUL3	ENSG00000179598	-1.62	3.6e-03	PLD6
ENSG00000011566	1.85	2.1e-03	MAP4K3	ENSG00000164758	1.23	3.6e-03	MED30
ENSG00000164287	1.00	2.1e-03	CDC20B	ENSG00000116237	-1.66	3.7e-03	ICMT
ENSG00000133872	0.91	2.1e-03	SARAF	ENSG00000114115	-1.83	3.7e-03	RBP1
ENSG00000138032	1.17	2.1e-03	PPM1B	ENSG00000143727	0.82	3.7e-03	ACP1
ENSG00000111605	0.94	2.1e-03	CPSF6	ENSG00000161243	-0.67	3.7e-03	FBXO27
ENSG00000083520	1.43	2.2e-03	DIS3	ENSG00000101109	1.11	3.7e-03	STK4
ENSG00000221963	1.30	2.2e-03	APOL6	ENSG00000155816	1.65	3.8e-03	FMN2
ENSG00000187595	-1.40	2.2e-03	ZNF385C	ENSG00000170647	-2.18	3.8e-03	TMEM133
ENSG00000065243	1.06	2.2e-03	PKN2	ENSG00000198865	-2.55	3.8e-03	CCDC152
ENSG00000104998	-1.51	2.2e-03	IL27RA	ENSG00000152778	2.11	3.8e-03	IFIT5
ENSG00000010017	1.55	2.2e-03	RANBP9	ENSG00000214265	2.04	3.9e-03	SNURF
ENSG00000132823	1.73	2.2e-03	OSER1	ENSG00000141452	1.94	3.9e-03	C18orf8
ENSG00000183735	1.54	2.2e-03	TBK1	ENSG00000128989	1.13	3.9e-03	ARPP19
ENSG00000089234	1.33	2.3e-03	BRAP	ENSG00000104320	1.12	4.0e-03	NBN
ENSG00000139826	1.30	2.3e-03	ABHD13	ENSG00000154582	0.74	4.0e-03	TCEB1
ENSG00000085274	1.39	2.3e-03	MYNN	ENSG00000183747	0.87	4.0e-03	ACSM2A
ENSG00000196646	1.69	2.3e-03	ZNF136	ENSG00000203867	-0.83	4.0e-03	RBM20
ENSG00000105556	-1.62	2.3e-03	MIER2	ENSG00000058272	0.78	4.0e-03	PPP1R12A
ENSG00000220205	0.98	2.3e-03	VAMP2	ENSG00000008018	0.77	4.0e-03	PSMB1
ENSG00000135605	1.56	2.4e-03	TEC	ENSG00000091844	1.67	4.0e-03	RGS17
ENSG00000151247	0.95	2.4e-03	EIF4E	ENSG00000100664	0.87	4.0e-03	EIF5
ENSG00000171161	-1.20	2.4e-03	ZNF672	ENSG00000104497	1.71	4.0e-03	SNX16
ENSG00000244115	1.70	2.4e-03	DNAJC25-GNG10	ENSG00000108510	1.20	4.0e-03	MED13
ENSG00000117505	1.28	2.4e-03	DR1	ENSG00000105708	1.83	4.1e-03	ZNF14
ENSG00000059378	1.90	2.4e-03	PARP12	ENSG00000185651	0.85	4.2e-03	UBE2L3
ENSG00000156875	1.34	2.4e-03	MFSD14A	ENSG00000106785	-1.49	4.2e-03	TRIM14
ENSG00000005483	1.21	2.4e-03	KMT2E	ENSG00000164754	0.95	4.3e-03	RAD21
ENSG00000092148	1.01	2.4e-03	HECTD1	ENSG00000170558	-2.54	4.3e-03	CDH2
ENSG00000102908	1.49	2.5e-03	NFAT5	ENSG00000127947	1.29	4.3e-03	PTPN12
ENSG00000065060	1.36	2.5e-03	UHRF1BP1	ENSG00000117036	1.63	4.3e-03	ETV3
ENSG00000152520	1.60	2.5e-03	PAN3	ENSG00000139218	0.87	4.3e-03	SCAF11
ENSG00000032219	1.53	2.5e-03	ARID4A	ENSG00000173039	1.24	4.3e-03	RELA
ENSG00000206560	1.31	2.5e-03	ANKRD28	ENSG00000178397	-2.45	4.3e-03	FAM220A
ENSG00000003436	1.56	2.5e-03	TFPI	ENSG00000144736	-1.43	4.3e-03	SHQ1
ENSG00000066422	1.58	2.5e-03	ZBTB11	ENSG00000112739	0.86	4.4e-03	PRPF4B
ENSG00000108559	1.40	2.5e-03	NUP88	ENSG00000092203	0.95	4.5e-03	TOX4
ENSG00000070495	1.37	2.6e-03	JMJD6	ENSG00000130726	-0.88	4.5e-03	TRIM28
ENSG00000095787	0.94	2.6e-03	WAC	ENSG00000151012	2.15	4.5e-03	SLC7A11
ENSG00000117500	1.42	2.6e-03	TMED5	ENSG00000171988	1.62	4.5e-03	JMJD1C
ENSG00000170540	1.02	2.6e-03	ARL6IP1	ENSG00000273820	-2.13	4.6e-03	USP27X
ENSG00000197647	1.59	2.6e-03	ZNF433	ENSG00000137955	1.10	4.6e-03	RABGGTB
ENSG00000025156	1.78	2.6e-03	HSF2	ENSG00000118418	1.02	4.6e-03	HMGN3
ENSG00000203952	1.61	2.6e-03	CCDC160	ENSG00000125970	1.21	4.6e-03	RALY
ENSG00000152291	0.73	2.6e-03	TGOLN2	ENSG00000163898	-1.25	4.6e-03	LIPH
ENSG00000130208	-2.57	2.7e-03	APOC1	ENSG00000126804	1.66	4.7e-03	ZBTB1
ENSG00000173269	1.35	2.7e-03	MMRN2	ENSG00000072274	1.52	4.7e-03	TFRC
ENSG00000136286	1.91	2.7e-03	MYO1G	ENSG00000169955	-1.23	4.7e-03	ZNF747
ENSG00000147443	2.64	2.7e-03	DOK2	ENSG00000266086	1.25	4.7e-03	NA
ENSG00000150403	1.13	2.7e-03	TMCO3	ENSG00000105856	1.72	4.8e-03	HBP1
ENSG00000173889	1.26	2.7e-03	PHC3	ENSG00000141367	1.04	4.8e-03	CLTC
ENSG00000140319	0.79	2.9e-03	SRP14	ENSG00000132953	1.06	4.8e-03	XPO4
ENSG00000091137	-1.53	2.9e-03	SLC26A4	ENSG00000160953	-1.83	4.8e-03	MUM1
ENSG00000168214	1.04	2.9e-03	RBPJ	ENSG00000120457	0.84	4.8e-03	KCNJ5
ENSG00000106546	1.81	2.9e-03	AHR	ENSG00000168036	1.42	4.8e-03	CTNNB1
ENSG00000136881	2.01	2.9e-03	BAAT	ENSG00000084093	0.89	4.8e-03	REST
ENSG00000138663	0.83	2.9e-03	COPS4	ENSG00000107854	1.27	4.9e-03	TNKS2
ENSG00000124226	1.09	3.0e-03	RNF114	ENSG00000134375	1.01	4.9e-03	TIMM17A
ENSG00000124333	1.38	3.0e-03	VAMP7	ENSG00000125952	1.30	4.9e-03	MAX
ENSG00000148296	-0.97	3.0e-03	SURF6	ENSG00000136997	-2.13	4.9e-03	MYC
ENSG00000096746	0.85	3.0e-03	HNRNPH3	ENSG00000136003	1.14	4.9e-03	ISCU
ENSG00000105610	-2.58	3.1e-03	KLF1	ENSG00000243279	1.81	4.9e-03	PRAF2
ENSG00000170759	1.12	3.1e-03	KIF5B	ENSG00000117616	1.17	5.0e-03	RSRP1
ENSG00000129084	0.85	3.1e-03	PSMA1	ENSG00000107443	1.54	5.0e-03	CCNJ
ENSG00000165029	2.29	3.1e-03	ABCA1	ENSG00000261408	-1.44	5.1e-03	CDK3
ENSG00000182566	-2.35	3.1e-03	CLEC4G	ENSG00000108312	-1.15	5.2e-03	UBTF
ENSG00000068878	1.45	3.1e-03	PSME4	ENSG00000059758	1.45	5.2e-03	CDK17
ENSG00000185883	1.32	3.2e-03	ATP6V0C	ENSG00000119801	1.41	5.2e-03	YPEL5
ENSG00000196924	1.94	3.2e-03	FLNA	ENSG00000145103	-1.65	5.2e-03	ILDR1
ENSG00000151491	2.06	5.2e-03	EPS8	ENSG00000165507	2.36	8.1e-03	C10orf10
ENSG00000176422	-1.14	5.2e-03	SPRYD4	ENSG00000133816	1.78	8.2e-03	MICAL2
ENSG00000203910	-1.97	5.2e-03	C1orf146	ENSG00000151553	1.15	8.2e-03	FAM160B1
ENSG00000167925	-1.42	5.3e-03	GHDC	ENSG00000181652	-2.20	8.3e-03	ATG9B
ENSG00000008294	1.32	5.3e-03	SPAG9	ENSG00000164037	1.86	8.3e-03	SLC9B1
ENSG00000267697	1.76	5.3e-03	LUZP6	ENSG00000076248	-2.00	8.3e-03	UNG
ENSG00000263020	1.59	5.4e-03	LY6G5B	ENSG00000132294	1.65	8.5e-03	EFR3A
ENSG00000133858	1.35	5.4e-03	ZFC3H1	ENSG00000177105	0.91	8.5e-03	RHOG
ENSG00000145623	1.10	5.4e-03	OSMR	ENSG00000002586	1.25	8.6e-03	CD99
ENSG00000164548	1.31	5.4e-03	TRA2A	ENSG00000111142	0.84	8.7e-03	METAP2
ENSG00000169375	1.42	5.4e-03	SIN3A	ENSG00000162592	1.16	8.7e-03	CCDC27
ENSG00000186185	-1.47	5.5e-03	KIF18B	ENSG00000144655	1.15	8.9e-03	CSRNP1
ENSG00000172409	2.12	5.5e-03	CLP1	ENSG00000168883	1.19	8.9e-03	USP39

Continuation of Table A.24

Ensembl	log_2 Fold Change	padj	Gene Symbol	Ensembl	log_2 Fold Change	padj	Gene Symbol
ENSG00000259529	1.78	5.5e-03	NA	ENSG00000160563	1.46	8.9e-03	MED27
ENSG00000028203	1.25	5.5e-03	VEZT	ENSG00000197536	1.50	9.0e-03	C5orf56
ENSG00000161547	0.82	5.7e-03	SRSF2	ENSG00000154556	-1.27	9.0e-03	SORBS2
ENSG00000177565	1.09	5.7e-03	TBL1XR1	ENSG00000076108	0.83	9.1e-03	BAZ2A
ENSG00000139725	1.13	5.8e-03	RHOF	ENSG00000122882	1.06	9.1e-03	ECD
ENSG00000160058	1.45	5.8e-03	BSDC1	ENSG00000176597	1.18	9.2e-03	B3GNT5
ENSG00000183054	1.50	5.8e-03	RGPD5	ENSG00000131051	0.77	9.2e-03	RBM39
ENSG00000160932	1.68	5.9e-03	LY6E	ENSG00000144224	1.18	9.2e-03	UBXN4
ENSG00000147586	-1.20	6.0e-03	MRPS28	ENSG00000108774	1.32	9.2e-03	RAB5C
ENSG00000100567	0.70	6.0e-03	PSMA3	ENSG00000160211	1.92	9.2e-03	G6PD
ENSG00000011600	1.53	6.0e-03	TYROBP	ENSG00000064692	1.75	9.4e-03	SNCAIP
ENSG00000143436	1.24	6.0e-03	MRPL9	ENSG00000136950	1.29	9.5e-03	ARPC5L
ENSG00000166471	0.98	6.0e-03	TMEM41B	ENSG00000153933	1.45	9.6e-03	DGKE
ENSG00000143621	0.82	6.0e-03	ILF2	ENSG00000123384	-1.27	9.6e-03	LRP1
ENSG00000162642	1.18	6.0e-03	C1orf52	ENSG00000180398	0.76	9.6e-03	MCFD2
ENSG00000227268	-1.22	6.1e-03	KLLN	ENSG00000178381	1.27	9.7e-03	ZFAND2A
ENSG00000242852	2.27	6.2e-03	ZNF709	ENSG00000043093	1.01	9.7e-03	DCUN1D1
ENSG00000135241	1.53	6.3e-03	PNPLA8	ENSG00000127481	0.60	9.8e-03	UBR4
ENSG00000058673	0.80	6.4e-03	ZC3H11A	ENSG00000146859	1.61	9.8e-03	TMEM140
ENSG00000167004	1.25	6.4e-03	PDIA3	ENSG00000100109	1.32	1.0e-02	TFIP11
ENSG00000198060	1.48	6.4e-03	MARCH5	ENSG00000170584	1.07	1.0e-02	NUDCD2
ENSG00000142669	1.15	6.4e-03	SH3BGRL3	ENSG00000283228	1.63	1.0e-02	CACNB4
ENSG00000120727	0.79	6.5e-03	PAIP2	ENSG00000108256	0.91	1.0e-02	NUFIP2
ENSG00000078699	-1.09	6.5e-03	CBFA2T2	ENSG00000122958	1.03	1.0e-02	VPS26A
ENSG00000156256	1.16	6.6e-03	USP16	ENSG00000170144	0.67	1.0e-02	HNRNPA3
ENSG00000109606	0.93	6.7e-03	DHX15	ENSG00000132323	0.99	1.0e-02	ILKAP
ENSG00000162231	1.02	6.7e-03	NXF1	ENSG00000137747	1.34	1.0e-02	TMPRSS13
ENSG00000110321	0.92	6.8e-03	EIF4G2	ENSG00000167552	1.13	1.0e-02	TUBA1A
ENSG00000083896	1.02	6.9e-03	YTHDC1	ENSG00000163519	2.40	1.1e-02	TRAT1
ENSG00000146223	-0.74	6.9e-03	RPL7L1	ENSG00000125037	0.86	1.1e-02	EMC3
ENSG00000136738	1.53	6.9e-03	STAM	ENSG00000173875	1.27	1.1e-02	ZNF791
ENSG00000162852	1.26	6.9e-03	CNST	ENSG00000141664	1.59	1.1e-02	ZCCHC2
ENSG00000102081	1.53	7.0e-03	FMR1	ENSG00000198625	0.92	1.1e-02	MDM4
ENSG00000109133	0.85	7.0e-03	TMEM33	ENSG00000088356	1.12	1.1e-02	PDRG1
ENSG00000130669	-1.08	7.0e-03	PAK4	ENSG00000056097	0.88	1.1e-02	ZFR
ENSG00000132388	1.25	7.0e-03	UBE2G1	ENSG00000075223	-2.24	1.1e-02	SEMA3C
ENSG00000164125	-2.47	7.0e-03	FAM198B	ENSG00000170873	1.94	1.1e-02	MTSS1
ENSG00000197223	1.42	7.0e-03	C1D	ENSG00000057663	0.98	1.1e-02	ATG5
ENSG00000198467	1.63	7.0e-03	TPM2	ENSG00000278828	2.06	1.1e-02	HIST1H3H
ENSG00000164022	0.79	7.0e-03	AIMP1	ENSG00000131791	1.56	1.1e-02	PRKAB2
ENSG00000166128	1.84	7.0e-03	RAB8B	ENSG00000118564	0.96	1.1e-02	FBXL5
ENSG00000091527	0.98	7.2e-03	CDV3	ENSG00000173848	1.38	1.1e-02	NET1
ENSG00000112218	-1.44	7.2e-03	GPR63	ENSG00000182220	0.71	1.1e-02	ATP6AP2
ENSG00000167378	1.38	7.2e-03	IRGQ	ENSG00000197372	1.33	1.1e-02	ZNF675
ENSG00000119335	0.79	7.3e-03	SET	ENSG00000168283	1.73	1.2e-02	BMI1
ENSG00000119048	1.35	7.3e-03	UBE2B	ENSG00000267060	-1.05	1.2e-02	PTGES3L
ENSG00000175895	1.49	7.3e-03	PLEKHF2	ENSG00000163564	2.09	1.2e-02	PYHIN1
ENSG00000126353	-2.40	7.4e-03	CCR7	ENSG00000065518	0.97	1.2e-02	NDUFB4
ENSG00000137414	1.72	7.4e-03	FAM8A1	ENSG00000015479	0.90	1.2e-02	MATR3
ENSG00000111237	0.87	7.4e-03	VPS29	ENSG00000102580	1.17	1.2e-02	DNAJC3
ENSG00000103199	-1.30	7.5e-03	ZNF500	ENSG00000081803	-2.38	1.2e-02	CADPS2
ENSG00000277758	2.33	7.5e-03	SYT15	ENSG00000136891	1.01	1.2e-02	TEX10
ENSG00000167595	-0.83	7.6e-03	PROSER3	ENSG00000167460	1.09	1.2e-02	TPM4
ENSG00000140299	1.49	7.6e-03	BNIP2	ENSG00000143507	1.40	1.2e-02	DUSP10
ENSG00000154229	1.89	7.6e-03	PRKCA	ENSG00000130803	1.03	1.2e-02	ZNF317
ENSG00000114107	1.68	7.6e-03	CEP70	ENSG00000168958	0.92	1.2e-02	MFF
ENSG00000185658	1.02	7.6e-03	BRWD1	ENSG00000129534	0.99	1.2e-02	MIS18BP1
ENSG00000100350	-1.51	7.6e-03	FOXRED2	ENSG00000186523	-1.71	1.2e-02	FAM86B1
ENSG00000157429	-1.21	7.6e-03	ZNF19	ENSG00000110958	0.65	1.2e-02	PTGES3
ENSG00000155926	-2.19	7.7e-03	SLA	ENSG00000122862	-1.31	1.2e-02	SRGN
ENSG00000136108	1.53	7.7e-03	CKAP2	ENSG00000090857	-1.01	1.2e-02	PDPR
ENSG00000158874	1.60	7.7e-03	APOA2	ENSG00000168374	0.96	1.2e-02	ARF4
ENSG00000163738	1.22	7.8e-03	MTHFD2L	ENSG00000196655	1.13	1.2e-02	TRAPPC4
ENSG00000182162	-1.24	8.0e-03	P2RY8	ENSG00000166912	1.09	1.2e-02	MTMR10
ENSG00000124659	0.97	8.0e-03	TBCC	ENSG00000156273	1.28	1.2e-02	BACH1
ENSG00000033800	0.67	8.0e-03	PIAS1	ENSG00000110324	2.19	1.3e-02	IL10RA
ENSG00000135549	1.86	8.0e-03	PKIB	ENSG00000114354	1.07	1.3e-02	TFG
ENSG00000142961	-2.49	8.0e-03	MOB3C	ENSG00000121807	-1.90	1.3e-02	CCR2
ENSG00000185811	-0.88	8.0e-03	IKZF1	ENSG00000155719	1.17	1.3e-02	OTOA
ENSG00000113558	0.86	8.0e-03	SKP1	ENSG00000158483	-2.12	1.3e-02	FAM86C1
ENSG00000173276	1.20	8.1e-03	ZBTB21	ENSG00000185246	1.00	1.3e-02	PRPF39
ENSG00000213719	0.83	8.1e-03	CLIC1	ENSG00000132952	1.44	1.3e-02	USPL1
ENSG00000148634	0.88	1.3e-02	HERC4	ENSG00000174574	0.87	1.8e-02	AKIRIN1
ENSG00000096717	1.16	1.3e-02	SIRT1	ENSG00000175352	1.68	1.8e-02	NRIP3
ENSG00000106803	0.62	1.3e-02	SEC61B	ENSG00000108840	1.23	1.8e-02	HDAC5
ENSG00000187522	1.22	1.3e-02	HSPA14	ENSG00000141510	-1.15	1.8e-02	TP53
ENSG00000001630	1.03	1.3e-02	CYP51A1	ENSG00000069424	-0.88	1.8e-02	KCNAB2
ENSG00000110318	-1.24	1.3e-02	CEP126	ENSG00000163961	1.07	1.8e-02	RNF168
ENSG00000161010	0.77	1.3e-02	C5orf45	ENSG00000203710	1.19	1.8e-02	CR1
ENSG00000164609	1.20	1.3e-02	SLU7	ENSG00000047315	0.97	1.8e-02	POLR2B
ENSG00000213064	0.67	1.3e-02	SFT2D2	ENSG00000160321	-1.17	1.8e-02	ZNF208
ENSG00000070729	1.14	1.3e-02	CNGB1	ENSG00000113595	1.45	1.8e-02	TRIM23
ENSG00000139154	0.98	1.3e-02	AEBP2	ENSG00000082781	2.18	1.9e-02	ITGB5
ENSG00000172183	1.88	1.3e-02	ISG20	ENSG00000107789	-2.15	1.9e-02	MINPP1
ENSG00000173480	1.08	1.3e-02	ZNF417	ENSG00000147677	0.71	1.9e-02	EIF3H
ENSG00000197622	1.49	1.4e-02	CDC42SE1	ENSG00000148483	1.18	1.9e-02	TMEM236
ENSG00000122641	2.24	1.4e-02	INHBA	ENSG00000164933	1.14	1.9e-02	SLC25A32
ENSG00000125454	-1.79	1.4e-02	SLC25A19	ENSG00000165516	1.01	1.9e-02	KLHDC2
ENSG00000138182	1.00	1.4e-02	KIF20B	ENSG00000115762	1.08	1.9e-02	PLEKHB2
ENSG00000162645	1.64	1.4e-02	GBP2	ENSG00000121236	-1.65	1.9e-02	TRIM6
ENSG00000157322	-1.59	1.4e-02	CLEC18A	ENSG00000139496	1.09	1.9e-02	NUP58

Continuation of Table A.24

Ensembl	$\log_2$ Fold Change	padj	Gene Symbol	Ensembl	$\log_2$ Fold Change	padj	Gene Symbol
ENSG00000136940	1.09	1.4e-02	PDCL	ENSG00000188739	0.72	1.9e-02	RBM34
ENSG00000130589	1.69	1.4e-02	HELZ2	ENSG00000183648	-0.82	1.9e-02	NDUFB1
ENSG00000180596	1.64	1.4e-02	HIST1H2BC	ENSG00000142166	0.95	1.9e-02	IFNAR1
ENSG00000134215	1.06	1.4e-02	VAV3	ENSG00000204152	0.78	1.9e-02	TIMM23B
ENSG00000112996	0.99	1.4e-02	MRPS30	ENSG00000263961	1.39	1.9e-02	C1orf186
ENSG00000119787	1.43	1.4e-02	ATL2	ENSG00000086589	0.80	1.9e-02	RBM22
ENSG00000162851	1.03	1.4e-02	TFB2M	ENSG00000072694	2.07	1.9e-02	FCGR2B
ENSG00000125347	1.49	1.4e-02	IRF1	ENSG00000184905	-1.98	2.0e-02	TCEAL2
ENSG00000197471	-0.93	1.4e-02	SPN	ENSG00000159239	0.97	2.0e-02	C2orf81
ENSG00000080709	1.83	1.4e-02	KCNN2	ENSG00000106105	0.97	2.0e-02	GARS
ENSG00000134755	2.03	1.4e-02	DSC2	ENSG00000151276	1.74	2.0e-02	MAGI1
ENSG00000147036	2.23	1.4e-02	LANCL3	ENSG00000153207	1.30	2.0e-02	AHCTF1
ENSG00000114125	0.88	1.4e-02	RNF7	ENSG00000128789	0.78	2.0e-02	PSMG2
ENSG00000102218	1.63	1.4e-02	RP2	ENSG00000214756	1.64	2.0e-02	METTL12
ENSG00000004487	0.87	1.5e-02	KDM1A	ENSG00000108786	-1.22	2.0e-02	HSD17B1
ENSG00000126705	-1.20	1.5e-02	AHDC1	ENSG00000100600	-1.54	2.0e-02	LGMN
ENSG00000180423	1.52	1.5e-02	HARBI1	ENSG00000197324	1.29	2.0e-02	LRP10
ENSG00000083937	1.03	1.5e-02	CHMP2B	ENSG00000119953	1.04	2.0e-02	SMNDC1
ENSG00000113407	0.81	1.5e-02	TARS	ENSG00000155090	1.33	2.0e-02	KLF10
ENSG00000121716	-1.41	1.5e-02	PILRB	ENSG00000081913	1.32	2.0e-02	PHLPP1
ENSG00000276085	2.23	1.5e-02	CCL3L1	ENSG00000113575	0.98	2.0e-02	PPP2CA
ENSG00000080371	1.21	1.5e-02	RAB21	ENSG00000146757	1.17	2.0e-02	ZNF92
ENSG00000154721	-1.70	1.5e-02	JAM2	ENSG00000171843	1.71	2.0e-02	MLLT3
ENSG00000168264	0.94	1.5e-02	IRF2BP2	ENSG00000187627	2.16	2.0e-02	RGPD1
ENSG00000197461	1.76	1.5e-02	PDGFA	ENSG00000197465	-2.26	2.0e-02	GYPE
ENSG00000136813	0.77	1.5e-02	KIAA0368	ENSG00000179933	-1.37	2.1e-02	C14orf119
ENSG00000161057	0.95	1.5e-02	PSMC2	ENSG00000170802	1.07	2.1e-02	FOXN2
ENSG00000124019	1.71	1.5e-02	FAM124B	ENSG00000145332	-1.35	2.1e-02	KLHL8
ENSG00000135476	-1.65	1.5e-02	ESPL1	ENSG00000121741	0.99	2.1e-02	ZMYM2
ENSG00000049245	1.41	1.5e-02	VAMP3	ENSG00000117569	1.22	2.1e-02	PTBP2
ENSG00000143702	0.75	1.5e-02	CEP170	ENSG00000132471	1.06	2.1e-02	WBP2
ENSG00000196459	0.71	1.5e-02	TRAPPC2	ENSG00000135870	1.14	2.1e-02	RC3H1
ENSG00000162654	1.17	1.5e-02	GBP4	ENSG00000100393	0.78	2.1e-02	EP300
ENSG00000145241	0.79	1.6e-02	CENPC	ENSG00000100353	0.69	2.1e-02	EIF3D
ENSG00000160710	0.64	1.6e-02	ADAR	ENSG00000204856	-1.33	2.1e-02	FAM216A
ENSG00000184678	1.36	1.6e-02	HIST2H2BE	ENSG00000154319	-1.93	2.1e-02	FAM167A
ENSG00000161996	-1.01	1.6e-02	WDR90	ENSG00000223547	1.34	2.1e-02	ZNF844
ENSG00000250361	-2.29	1.6e-02	GYPB	ENSG00000087074	1.47	2.1e-02	PPP1R15A
ENSG00000117335	0.76	1.6e-02	CD46	ENSG00000145391	1.29	2.1e-02	SETD7
ENSG00000106245	1.07	1.6e-02	BUD31	ENSG00000108379	-1.85	2.1e-02	WNT3
ENSG00000115966	1.12	1.6e-02	ATF2	ENSG00000103121	0.86	2.1e-02	CMC2
ENSG00000126945	1.11	1.6e-02	HNRNPH2	ENSG00000165669	0.79	2.1e-02	FAM204A
ENSG00000162711	-1.66	1.6e-02	NLRP3	ENSG00000083223	0.99	2.1e-02	ZCCHC6
ENSG00000171222	1.44	1.6e-02	SCAND1	ENSG00000187608	1.33	2.1e-02	ISG15
ENSG00000061794	0.93	1.6e-02	MRPS35	ENSG00000101638	0.91	2.1e-02	ST8SIA5
ENSG00000241128	-1.85	1.6e-02	NA	ENSG00000106992	1.44	2.1e-02	AK1
ENSG00000160877	-2.03	1.7e-02	NACC1	ENSG00000163380	1.73	2.2e-02	LMOD3
ENSG00000178567	1.11	1.7e-02	EPM2AIP1	ENSG00000172831	-1.45	2.2e-02	CES2
ENSG00000134744	0.86	1.7e-02	ZCCHC11	ENSG00000173726	0.80	2.2e-02	TOMM20
ENSG00000166747	1.13	1.7e-02	AP1G1	ENSG00000022976	-0.92	2.2e-02	ZNF839
ENSG00000102738	-1.33	1.7e-02	MRPS31	ENSG00000058600	0.90	2.2e-02	POLR3E
ENSG00000072364	1.07	1.7e-02	AFF4	ENSG00000090382	-1.94	2.2e-02	LYZ
ENSG00000087086	0.87	1.7e-02	FTL	ENSG00000136161	2.02	2.2e-02	RCBTB2
ENSG00000089916	1.08	1.7e-02	GPATCH2L	ENSG00000152382	1.64	2.2e-02	TADA1
ENSG00000116254	-2.19	1.7e-02	CHD5	ENSG00000170370	1.52	2.2e-02	EMX2
ENSG00000123562	1.00	1.7e-02	MORF4L2	ENSG00000152147	-1.03	2.2e-02	GEMIN6
ENSG00000170959	0.97	1.7e-02	DCDC1	ENSG00000172262	0.73	2.2e-02	ZNF131
ENSG00000150093	1.16	1.7e-02	ITGB1	ENSG00000140623	-1.80	2.2e-02	SEPT12
ENSG00000137497	-0.73	1.7e-02	NUMA1	ENSG00000221955	1.26	2.2e-02	SLC12A8
ENSG00000111679	1.19	1.7e-02	PTPN6	ENSG00000280987	1.13	2.2e-02	MATR3
ENSG00000154478	-1.17	1.7e-02	GPR26	ENSG00000215475	-1.56	2.2e-02	SIAH3
ENSG00000105993	1.15	1.7e-02	DNAJB6	ENSG00000122068	0.95	2.2e-02	FYTTD1
ENSG00000084623	0.78	1.8e-02	EIF3I	ENSG00000133874	1.75	2.2e-02	RNF122
ENSG00000131013	1.07	1.8e-02	PPIL4	ENSG00000213782	0.91	2.2e-02	DDX47
ENSG00000147316	0.93	1.8e-02	MCPH1	ENSG00000168298	1.16	2.3e-02	HIST1H1E
ENSG00000164687	-1.25	1.8e-02	FABP5	ENSG00000164291	-1.52	2.3e-02	ARSK
ENSG00000278129	1.23	2.3e-02	ZNF8	ENSG00000024048	0.85	2.9e-02	UBR2
ENSG00000174780	0.70	2.3e-02	SRP72	ENSG00000180008	1.31	2.9e-02	SOCS4
ENSG00000177200	0.66	2.3e-02	CHD9	ENSG00000186862	-0.75	2.9e-02	PDZD7
ENSG00000158092	0.90	2.3e-02	NCK1	ENSG00000189060	1.73	2.9e-02	H1F0
ENSG00000070501	1.00	2.3e-02	POLB	ENSG00000099326	-1.45	3.0e-02	MZF1
ENSG00000147162	0.73	2.3e-02	OGT	ENSG00000138375	-1.55	3.0e-02	SMARCAL1
ENSG00000036054	1.16	2.3e-02	TBC1D23	ENSG00000141232	1.39	3.0e-02	TOB1
ENSG00000150459	0.80	2.3e-02	SAP18	ENSG00000091986	0.80	3.0e-02	CCDC80
ENSG00000118058	0.77	2.3e-02	KMT2A	ENSG00000112773	1.64	3.0e-02	FAM46A
ENSG00000102230	-1.60	2.3e-02	PCYT1B	ENSG00000136156	0.88	3.0e-02	ITM2B
ENSG00000185046	-1.06	2.3e-02	ANKS1B	ENSG00000258839	-2.13	3.0e-02	MC1R
ENSG00000097096	-1.27	2.3e-02	SYDE2	ENSG00000117000	1.50	3.0e-02	RLF
ENSG00000196092	1.69	2.3e-02	PAX5	ENSG00000141696	-0.85	3.0e-02	P3H4
ENSG00000134330	0.96	2.3e-02	IAH1	ENSG00000163877	1.47	3.0e-02	SNIP1
ENSG00000151465	0.83	2.3e-02	CDC123	ENSG00000110429	1.21	3.0e-02	FBXO3
ENSG00000131732	-1.18	2.4e-02	ZCCHC9	ENSG00000123213	-1.04	3.0e-02	NLN
ENSG00000136937	0.92	2.4e-02	NCBP1	ENSG00000163874	1.79	3.0e-02	ZC3H12A
ENSG00000126457	0.85	2.4e-02	PRMT1	ENSG00000101695	0.97	3.1e-02	RNF125
ENSG00000173262	1.49	2.4e-02	SLC2A14	ENSG00000133048	1.65	3.1e-02	CHI3L1
ENSG00000116514	1.70	2.4e-02	RNF19B	ENSG00000204118	-1.41	3.1e-02	NAP1L6
ENSG00000279374	-1.60	2.4e-02	NA	ENSG00000047365	1.18	3.1e-02	ARAP2
ENSG00000162004	-2.13	2.4e-02	CCDC78	ENSG00000182378	-0.86	3.1e-02	PLCXD1
ENSG00000204869	0.95	2.4e-02	LOC400706	ENSG00000134058	1.00	3.1e-02	CDK7
ENSG00000169914	1.18	2.4e-02	OTUD3	ENSG00000266074	-1.46	3.1e-02	BAHCC1
ENSG00000172209	1.80	2.5e-02	GPR22	ENSG00000161980	-1.26	3.1e-02	POLR3K

Continuation of Table A.24

Ensembl	log_2 Fold Change	padj	Gene Symbol	Ensembl	log_2 Fold Change	padj	Gene Symbol
ENSG00000215788	1.55	2.5e-02	TNFRSF25	ENSG00000147065	0.70	3.1e-02	MSN
ENSG00000173250	-1.94	2.5e-02	GPR151	ENSG00000152782	-1.45	3.1e-02	PANK1
ENSG00000065534	-1.54	2.5e-02	MYLK	ENSG00000066427	0.51	3.2e-02	ATXN3
ENSG00000112419	1.32	2.5e-02	PHACTR2	ENSG00000090863	-1.20	3.2e-02	GLG1
ENSG00000137075	1.18	2.5e-02	RNF38	ENSG00000163714	0.66	3.2e-02	U2SURP
ENSG00000164305	1.51	2.5e-02	CASP3	ENSG00000055070	0.96	3.2e-02	SZRD1
ENSG00000109381	0.75	2.5e-02	ELF2	ENSG00000164024	0.83	3.2e-02	METAP1
ENSG00000102103	0.72	2.5e-02	PQBP1	ENSG00000077147	0.92	3.2e-02	TM9SF3
ENSG00000177685	-2.14	2.5e-02	CRACR2B	ENSG00000113369	1.47	3.2e-02	ARRDC3
ENSG00000101084	0.92	2.5e-02	C20orf24	ENSG00000095932	-0.96	3.2e-02	SMIM24
ENSG00000149231	1.22	2.5e-02	CCDC82	ENSG00000136573	2.05	3.2e-02	BLK
ENSG00000176018	1.61	2.5e-02	LYSMD3	ENSG00000173674	0.71	3.3e-02	EIF1AX
ENSG00000108055	0.87	2.6e-02	SMC3	ENSG00000145287	-0.88	3.3e-02	PLAC8
ENSG00000137710	1.19	2.6e-02	RDX	ENSG00000149257	-1.27	3.3e-02	SERPINH1
ENSG00000132357	-1.43	2.6e-02	CARD6	ENSG00000168216	0.99	3.3e-02	LMBRD1
ENSG00000140400	-1.05	2.6e-02	MAN2C1	ENSG00000101144	1.11	3.3e-02	BMP7
ENSG00000221914	1.02	2.6e-02	PPP2R2A	ENSG00000115306	1.15	3.3e-02	SPTBN1
ENSG00000167674	-1.13	2.6e-02	HDGFRP2	ENSG00000176974	-1.34	3.3e-02	SHMT1
ENSG00000101745	0.79	2.6e-02	ANKRD12	ENSG00000144895	0.82	3.3e-02	EIF2A
ENSG00000137460	1.02	2.6e-02	FHDC1	ENSG00000105879	1.38	3.3e-02	CBLL1
ENSG00000139505	1.48	2.6e-02	MTMR6	ENSG00000074054	0.91	3.3e-02	CLASP1
ENSG00000148444	1.21	2.6e-02	COMMD3	ENSG00000125846	-1.64	3.3e-02	ZNF133
ENSG00000136146	0.99	2.6e-02	MED4	ENSG00000170074	-1.92	3.3e-02	FAM153A
ENSG00000141279	0.74	2.6e-02	NPEPPS	ENSG00000115825	1.00	3.3e-02	PRKD3
ENSG00000164938	1.56	2.6e-02	TP53INP1	ENSG00000183726	0.72	3.4e-02	TMEM50A
ENSG00000166105	0.75	2.6e-02	GLB1L3	ENSG00000271503	1.60	3.4e-02	CCL5
ENSG00000114902	0.72	2.6e-02	SPCS1	ENSG00000135968	0.66	3.4e-02	GCC2
ENSG00000105202	-0.99	2.7e-02	FBL	ENSG00000080822	1.03	3.4e-02	CLDND1
ENSG00000165195	1.66	2.7e-02	PIGA	ENSG00000089737	0.73	3.4e-02	DDX24
ENSG00000064989	1.78	2.7e-02	CALCRL	ENSG00000164167	-1.53	3.5e-02	LSM6
ENSG00000101336	1.93	2.7e-02	HCK	ENSG00000110917	-1.24	3.5e-02	MLEC
ENSG00000153879	1.10	2.7e-02	CEBPG	ENSG00000103150	-1.46	3.5e-02	MLYCD
ENSG00000058729	1.57	2.7e-02	RIOK2	ENSG00000177225	-0.68	3.5e-02	PDDC1
ENSG00000088448	1.08	2.7e-02	ANKRD10	ENSG00000148843	-1.29	3.5e-02	PDCD11
ENSG00000105821	0.92	2.7e-02	DNAJC2	ENSG00000165813	0.98	3.5e-02	CCDC186
ENSG00000196072	0.89	2.7e-02	BLOC1S2	ENSG00000166780	-1.81	3.5e-02	C16orf45
ENSG00000153561	0.99	2.7e-02	RMND5A	ENSG00000112486	1.65	3.5e-02	CCR6
ENSG00000141542	-1.35	2.7e-02	RAB40B	ENSG00000010278	1.99	3.5e-02	CD9
ENSG00000140575	1.09	2.7e-02	IQGAP1	ENSG00000112078	0.91	3.5e-02	KCTD20
ENSG00000157510	1.56	2.7e-02	AFAP1L1	ENSG00000133226	0.71	3.5e-02	SRRM1
ENSG00000165732	0.82	2.7e-02	DDX21	ENSG00000187994	-1.17	3.5e-02	RINL
ENSG00000213699	-1.27	2.8e-02	SLC35F6	ENSG00000168769	0.88	3.5e-02	TET2
ENSG00000075651	1.69	2.8e-02	PLD1	ENSG00000169398	1.82	3.5e-02	PTK2
ENSG00000135218	1.79	2.8e-02	CD36	ENSG00000197496	-1.37	3.5e-02	SLC2A10
ENSG00000104327	-1.79	2.8e-02	CALB1	ENSG00000053372	-1.31	3.5e-02	MRTO4
ENSG00000006576	1.29	2.8e-02	PHTF2	ENSG00000044574	1.40	3.5e-02	HSPA5
ENSG00000141380	0.77	2.8e-02	SS18	ENSG00000188522	-1.31	3.5e-02	FAM83G
ENSG00000156508	0.68	2.8e-02	EEF1A1	ENSG00000111602	-1.20	3.5e-02	TIMELESS
ENSG00000172339	-0.75	2.8e-02	ALG14	ENSG00000144061	1.15	3.6e-02	NPHP1
ENSG00000131781	1.38	2.8e-02	FMO5	ENSG00000187699	-1.33	3.6e-02	C2orf88
ENSG00000163376	2.10	2.8e-02	KBTBD8	ENSG00000274386	1.55	3.6e-02	LOC100129924
ENSG00000146247	0.94	2.8e-02	PHIP	ENSG00000152454	1.96	3.6e-02	ZNF256
ENSG00000145723	-1.45	2.8e-02	GIN1	ENSG00000157601	1.56	3.6e-02	MX1
ENSG00000120992	1.11	2.8e-02	LYPLA1	ENSG00000140678	1.34	3.6e-02	ITGAX
ENSG00000122644	1.39	2.9e-02	ARL4A	ENSG00000182149	0.96	3.6e-02	IST1
ENSG00000100796	0.71	2.9e-02	PPP4R3A	ENSG00000109118	1.09	3.6e-02	PHF12
ENSG00000143158	0.84	2.9e-02	MPC2	ENSG00000155886	-0.77	3.6e-02	SLC24A2
ENSG00000104671	1.17	2.9e-02	DCTN6	ENSG00000064012	0.81	3.7e-02	CASP8
ENSG00000120948	0.80	2.9e-02	TARDBP	ENSG00000131944	-0.99	3.7e-02	FAAP24
ENSG00000157045	1.41	2.9e-02	NTAN1	ENSG00000164944	0.94	3.7e-02	KIAA1429
ENSG00000162517	0.93	3.7e-02	PEF1	ENSG00000119414	0.90	4.5e-02	PPP6C
ENSG00000116977	0.83	3.7e-02	LGALS8	ENSG00000154727	0.93	4.5e-02	GABPA
ENSG00000100603	0.90	3.7e-02	SNW1	ENSG00000181625	-1.10	4.5e-02	SLX1B
ENSG00000134153	1.17	3.7e-02	EMC7	ENSG00000198743	1.23	4.5e-02	SLC5A3
ENSG00000164329	0.96	3.7e-02	PAPD4	ENSG00000262180	1.91	4.5e-02	OCLM
ENSG00000186638	-2.08	3.8e-02	KIF24	ENSG00000006459	1.72	4.5e-02	KDM7A
ENSG00000081019	0.95	3.8e-02	RSBN1	ENSG00000170855	0.83	4.5e-02	TRIAP1
ENSG00000028116	0.80	3.8e-02	VRK2	ENSG00000139323	0.85	4.6e-02	POC1B
ENSG00000078304	0.94	3.8e-02	PPP2R5C	ENSG00000136802	1.06	4.6e-02	LRRC8A
ENSG00000059804	1.57	3.8e-02	SLC2A3	ENSG00000103018	-1.17	4.6e-02	CYB5B
ENSG00000105835	1.26	3.8e-02	NAMPT	ENSG00000108018	1.90	4.6e-02	SORCS1
ENSG00000104518	-1.18	3.8e-02	GSDMD	ENSG00000172795	1.24	4.6e-02	DCP2
ENSG00000197863	-1.39	3.8e-02	ZNF790	ENSG00000082074	1.72	4.6e-02	FYB
ENSG00000213024	-1.10	3.8e-02	NUP62	ENSG00000171940	1.24	4.6e-02	ZNF217
ENSG00000028137	-1.23	3.8e-02	TNFRSF1B	ENSG00000134748	0.75	4.6e-02	PRPF38A
ENSG00000117899	-1.25	3.9e-02	MESDC2	ENSG00000137033	1.66	4.6e-02	IL33
ENSG00000196715	-1.35	3.9e-02	VKORC1L1	ENSG00000179335	0.87	4.6e-02	CLK3
ENSG00000090565	-1.54	3.9e-02	RAB11FIP3	ENSG00000187010	1.34	4.6e-02	RHD
ENSG00000100281	0.91	3.9e-02	HMGXB4	ENSG00000128534	0.80	4.7e-02	LSM8
ENSG00000125304	1.11	3.9e-02	TM9SF2	ENSG00000145216	0.77	4.7e-02	FIP1L1
ENSG00000135637	-0.99	4.0e-02	CCDC142	ENSG00000156802	1.06	4.7e-02	ATAD2
ENSG00000136485	-1.02	4.0e-02	DCAF7	ENSG00000173020	1.01	4.7e-02	GRK2
ENSG00000185947	1.44	4.0e-02	ZNF267	ENSG00000136367	-2.03	4.7e-02	ZFHX2
ENSG00000126261	0.68	4.0e-02	UBA2	ENSG00000169504	1.36	4.7e-02	CLIC4
ENSG00000163659	1.15	4.1e-02	TIPARP	ENSG00000139915	-2.01	4.7e-02	MDGA2
ENSG00000134825	-0.71	4.1e-02	TMEM258	ENSG00000153037	0.76	4.7e-02	SRP19
ENSG00000006534	-1.10	4.1e-02	ALDH3B1	ENSG00000188227	-0.87	4.7e-02	ZNF793
ENSG00000110713	0.97	4.1e-02	NUP98	ENSG00000157181	0.88	4.7e-02	C1orf27
ENSG00000122786	1.66	4.1e-02	CALD1	ENSG00000104884	-1.20	4.7e-02	ERCC2
ENSG00000123836	-1.64	4.1e-02	PFKFB2	ENSG00000115806	1.05	4.7e-02	GORASP2
ENSG00000087266	0.89	4.1e-02	SH3BP2	ENSG00000107077	0.93	4.8e-02	KDM4C

Continuation of Table A.24

Ensembl	log_2 Fold Change	padj	Gene Symbol	Ensembl	log_2 Fold Change	padj	Gene Symbol
ENSG00000155506	-0.82	4.1e-02	LARP1	ENSG00000241484	0.97	4.8e-02	ARHGAP8
ENSG00000116350	0.80	4.1e-02	SRSF4	ENSG00000177485	1.11	4.8e-02	ZBTB33
ENSG00000146909	-1.07	4.1e-02	NOM1	ENSG00000197555	1.61	4.8e-02	SIPA1L1
ENSG00000139910	1.48	4.1e-02	NOVA1	ENSG00000143368	0.76	4.8e-02	SF3B4
ENSG00000144848	0.80	4.1e-02	ATG3	ENSG00000015568	0.92	4.9e-02	RGPD5
ENSG00000095059	0.83	4.2e-02	DHPS	ENSG00000159307	1.32	4.9e-02	SCUBE1
ENSG00000115128	0.68	4.2e-02	SF3B6	ENSG00000072415	1.00	4.9e-02	MPP5
ENSG00000115604	1.84	4.2e-02	IL18R1	ENSG00000176358	1.15	4.9e-02	TAC4
ENSG00000131408	1.32	4.2e-02	NR1H2	ENSG00000048162	-1.13	5.0e-02	NOP16
ENSG00000145782	0.85	4.2e-02	ATG12	ENSG00000071054	1.11	5.0e-02	MAP4K4
ENSG00000146833	-0.84	4.2e-02	TRIM4	ENSG00000173825	-1.85	5.0e-02	TIGD3
ENSG00000188994	0.75	4.2e-02	ZNF292	ENSG00000162062	-1.97	5.0e-02	C16orf59
ENSG00000138468	1.03	4.2e-02	SENP7				
ENSG00000156171	0.87	4.2e-02	DRAM2				
ENSG00000159873	1.11	4.2e-02	CCDC117				
ENSG00000174945	1.58	4.2e-02	AMZ1				
ENSG00000119314	0.68	4.2e-02	PTBP3				
ENSG00000141424	1.09	4.2e-02	SLC39A6				
ENSG00000132781	-1.32	4.3e-02	MUTYH				
ENSG00000025796	0.79	4.3e-02	SEC63				
ENSG00000027697	1.10	4.3e-02	IFNGR1				
ENSG00000069431	0.64	4.3e-02	ABCC9				
ENSG00000141569	-1.01	4.3e-02	TRIM65				
ENSG00000147050	1.14	4.3e-02	KDM6A				
ENSG00000170035	1.41	4.3e-02	UBE2E3				
ENSG00000152795	0.48	4.3e-02	HNRNPDL				
ENSG00000141522	1.06	4.3e-02	ARHGDIA				
ENSG00000168259	0.65	4.3e-02	DNAJC7				
ENSG00000007944	1.23	4.3e-02	MYLIP				
ENSG00000077458	1.10	4.3e-02	FAM76B				
ENSG00000110422	1.17	4.3e-02	HIPK3				
ENSG00000162772	1.57	4.3e-02	ATF3				
ENSG00000182389	1.57	4.3e-02	CACNB4				
ENSG00000182704	0.63	4.3e-02	TSKU				
ENSG00000197245	1.45	4.3e-02	FAM110D				
ENSG00000100364	-1.15	4.4e-02	KIAA0930				
ENSG00000136542	1.49	4.4e-02	GALNT5				
ENSG00000165025	-0.96	4.4e-02	SYK				
ENSG00000141030	0.91	4.4e-02	COPS3				
ENSG00000161011	1.07	4.4e-02	SQSTM1				
ENSG00000134294	0.91	4.4e-02	SLC38A2				
ENSG00000108588	1.11	4.4e-02	CCDC47				
ENSG00000169429	1.69	4.4e-02	CXCL8				
ENSG00000149823	-0.87	4.4e-02	VPS51				
ENSG00000160145	1.40	4.4e-02	KALRN				
ENSG00000172766	1.32	4.4e-02	NAA16				
ENSG00000174804	0.89	4.4e-02	FZD4				
ENSG00000131370	1.62	4.4e-02	SH3BP5				
ENSG00000047346	1.19	4.4e-02	FAM214A				
ENSG00000132182	-1.17	4.5e-02	NUP210				
ENSG00000237515	0.56	4.5e-02	SHISA9				
ENSG00000122299	0.95	4.5e-02	ZC3H7A				
ENSG00000145020	-1.32	4.5e-02	AMT				
ENSG00000124766	1.21	4.5e-02	SOX4				

Table A.26: DEGs Between Normal HSC and Leukaemic HSPC. DEGs were identified between normal HSC and leukaemic HSPC, using a Wald test in DESeq2. A design formula of ~condition was used (where condition indicates normal/leukaemic status). Upregulated and downregulated DEGs with a p-adjusted value < 0.05 are shown.

Ensembl	$\log_2$ Fold Change	padj	Gene Symbol	Ensembl	$\log_2$ Fold Change	padj	Gene Symbol
ENSG00000118946	-8.26	2.5e-47	PCDH17	ENSG00000114405	-4.15	0.00010	C3orf14
ENSG00000081138	-4.43	3.5e-37	CDH7	ENSG00000162889	2.57	0.00013	MAPKAPK2
ENSG00000144057	-5.26	2.9e-34	ST6GAL2	ENSG00000155330	2.09	0.00016	C16orf87
ENSG00000011600	5.00	1.0e-12	TYROBP	ENSG00000167513	2.17	0.00016	CDT1
ENSG00000164035	-5.00	1.9e-11	EMCN	ENSG00000196850	2.47	0.00017	PPTC7
ENSG00000144824	-4.40	7.2e-11	PHLDB2	ENSG00000076604	3.69	0.00018	TRAF4
ENSG00000140575	3.50	6.2e-10	IQGAP1	ENSG00000101782	2.54	0.00018	RIOK3
ENSG00000152256	4.64	8.4e-10	PDK1	ENSG00000119396	1.80	0.00018	RAB14
ENSG00000132912	2.49	1.5e-09	DCTN4	ENSG00000139505	3.15	0.00018	MTMR6
ENSG00000101084	2.99	2.8e-09	C20orf24	ENSG00000206190	4.03	0.00018	ATP10A
ENSG00000154175	-3.47	3.7e-09	ABI3BP	ENSG00000101417	-1.63	0.00019	PXMP4
ENSG00000136546	-3.91	7.5e-09	SCN7A	ENSG00000108924	-4.48	0.00022	HLF
ENSG00000163521	-3.30	7.7e-09	GLB1L	ENSG00000186469	3.38	0.00024	GNG2
ENSG00000133116	-6.29	1.3e-08	KL	ENSG00000119689	1.98	0.00025	DLST
ENSG00000134202	-3.53	1.8e-08	GSTM3	ENSG00000197045	2.16	0.00025	GMFB
ENSG00000065809	5.15	1.9e-08	FAM107B	ENSG00000122034	2.33	0.00026	GTF3A
ENSG00000185630	-3.51	2.6e-08	PBX1	ENSG00000198832	-3.99	0.00029	SELENOM
ENSG00000162409	-3.42	2.9e-08	PRKAA2	ENSG00000067900	1.32	0.00032	ROCK1
ENSG00000196632	-3.52	3.6e-08	WNK3	ENSG00000100614	2.12	0.00032	PPM1A
ENSG00000184905	-5.54	8.9e-08	TCEAL2	ENSG00000148841	2.70	0.00032	ITPRIP
ENSG00000162664	2.31	1.3e-07	ZNF326	ENSG00000165152	-3.89	0.00036	TMEM246
ENSG00000126264	3.33	1.3e-07	HCST	ENSG00000019991	3.90	0.00037	HGF
ENSG00000198901	3.82	1.4e-07	PRC1	ENSG00000108179	2.91	0.00037	PPIF
ENSG00000243710	-4.94	1.5e-07	CFAP57	ENSG00000023734	1.82	0.00038	STRAP
ENSG00000082074	4.94	1.9e-07	FYB	ENSG00000081913	2.65	0.00040	PHLPP1
ENSG00000165626	-3.77	1.9e-07	BEND7	ENSG00000120616	1.51	0.00040	EPC1
ENSG00000106803	1.58	2.7e-07	SEC61B	ENSG00000143476	3.68	0.00040	DTL
ENSG00000111276	4.12	3.3e-07	CDKN1B	ENSG00000168259	1.44	0.00040	DNAJC7
ENSG00000113734	3.23	4.1e-07	BNIP1	ENSG00000162851	2.57	0.00041	TFB2M
ENSG00000107362	3.63	8.9e-07	ABHD17B	ENSG00000138772	-3.56	0.00041	ANXA3
ENSG00000250486	-5.42	8.9e-07	FAM218A	ENSG00000068305	2.26	0.00041	MEF2A
ENSG00000147862	-5.08	9.0e-07	NFIB	ENSG00000110367	1.51	0.00041	DDX6
ENSG00000132321	-4.92	1.3e-06	IQCA1	ENSG00000132819	3.52	0.00041	RBM38
ENSG00000105143	-1.54	1.5e-06	SLC1A6	ENSG00000139970	-3.76	0.00042	RTN1
ENSG00000002586	2.79	1.6e-06	CD99	ENSG00000162896	-1.68	0.00042	PIGR
ENSG00000107104	-3.16	1.8e-06	KANK1	ENSG00000147138	3.89	0.00042	GPR174
ENSG00000169435	-3.03	1.9e-06	RASSF6	ENSG00000070831	2.00	0.00044	CDC42
ENSG00000121207	-1.76	1.9e-06	LRAT	ENSG00000115216	2.24	0.00044	NRBP1
ENSG00000157214	-3.42	2.1e-06	STEAP2	ENSG00000120885	-3.61	0.00044	CLU
ENSG00000163444	1.78	2.8e-06	TMEM183A	ENSG00000122861	3.90	0.00044	PLAU
ENSG00000164125	-5.12	2.8e-06	FAM198B	ENSG00000163466	1.38	0.00044	ARPC2
ENSG00000105339	2.89	2.8e-06	DENND3	ENSG00000164070	-1.51	0.00044	HSPA4L
ENSG00000179021	2.06	4.1e-06	C3orf38	ENSG00000174776	3.47	0.00044	WDR49
ENSG00000141522	2.63	5.2e-06	ARHGDIA	ENSG00000136504	1.86	0.00046	KAT7
ENSG00000163781	2.78	5.2e-06	TOPBP1	ENSG00000067248	1.52	0.00047	DHX29
ENSG00000144771	3.90	5.2e-06	LRTM1	ENSG00000182481	2.23	0.00047	KPNA2
ENSG00000161249	-2.64	5.2e-06	DMKN	ENSG00000278311	1.65	0.00047	GGNBP2
ENSG00000153898	-2.45	6.0e-06	MCOLN2	ENSG00000066336	3.34	0.00050	SPI1
ENSG00000105374	4.50	6.6e-06	NKG7	ENSG00000154127	3.41	0.00051	UBASH3B
ENSG00000111371	2.70	6.6e-06	SLC38A1	ENSG00000182718	3.37	0.00051	ANXA2
ENSG00000143507	3.18	6.6e-06	DUSP10	ENSG00000273841	1.69	0.00051	TAF9
ENSG00000170271	-2.72	9.0e-06	FAXDC2	ENSG00000241128	-3.59	0.00052	NA
ENSG00000196839	2.76	1.2e-05	ADA	ENSG00000154447	2.49	0.00053	SH3RF1
ENSG00000136950	2.59	1.4e-05	ARPC5L	ENSG00000162702	1.80	0.00053	ZNF281
ENSG00000184007	1.52	1.6e-05	PTP4A2	ENSG00000006634	1.91	0.00053	DBF4
ENSG00000100711	-3.10	1.6e-05	ZFYVE21	ENSG00000100281	1.94	0.00053	HMGXB4
ENSG00000135900	2.05	1.6e-05	MRPL44	ENSG00000131876	1.59	0.00054	SNRPA1
ENSG00000068724	2.86	1.6e-05	TTC7A	ENSG00000112200	2.04	0.00057	ZNF451
ENSG00000106648	2.66	1.9e-05	GALNTL5	ENSG00000115091	1.53	0.00057	ACTR3
ENSG00000147872	3.72	2.0e-05	PLIN2	ENSG00000081803	-4.03	0.00063	CADPS2
ENSG00000182107	-4.11	2.2e-05	TMEM30B	ENSG00000108064	1.38	0.00063	TFAM
ENSG00000184838	-3.90	2.2e-05	PRR16	ENSG00000140600	-4.12	0.00064	SH3GL3
ENSG00000183032	-1.64	2.3e-05	SLC25A21	ENSG00000105669	1.83	0.00064	COPE
ENSG00000169641	1.96	2.6e-05	LUZP1	ENSG00000138069	1.72	0.00067	RAB1A
ENSG00000113387	1.74	2.6e-05	SUB1	ENSG00000109332	1.57	0.00070	UBE2D3
ENSG00000196083	3.60	2.6e-05	IL1RAP	ENSG00000176915	0.90	0.00070	ANKLE2
ENSG00000164823	2.59	2.7e-05	OSGIN2	ENSG00000198774	-3.27	0.00072	RASSF9
ENSG00000279765	2.59	3.0e-05	NA	ENSG00000164287	1.59	0.00072	CDC20B
ENSG00000137135	-2.21	3.2e-05	ARHGEF39	ENSG00000129355	3.39	0.00074	CDKN2D
ENSG00000140367	2.47	3.4e-05	UBE2Q2	ENSG00000131174	1.17	0.00074	COX7B
ENSG00000213145	4.21	3.6e-05	CRIP1	ENSG00000196470	2.08	0.00074	SIAH1
ENSG00000109381	1.62	4.0e-05	ELF2	ENSG00000111647	2.45	0.00074	UHRF1BP1L
ENSG00000151725	3.45	5.2e-05	CENPU	ENSG00000138182	1.98	0.00074	KIF20B
ENSG00000133657	2.64	5.4e-05	ATP13A3	ENSG00000155508	1.88	0.00074	CNOT8
ENSG00000111670	2.09	6.0e-05	GNPTAB	ENSG00000197245	3.01	0.00074	FAM110D
ENSG00000259305	4.02	6.0e-05	ZHX1-C8orf76	ENSG00000118193	3.27	0.00078	KIF14
ENSG00000204152	1.69	6.0e-05	TIMM23B	ENSG00000139496	2.07	0.00078	NUP58
ENSG00000124104	-3.33	6.5e-05	SNX21	ENSG00000137449	3.69	0.00079	CPEB2
ENSG00000179869	-3.08	6.6e-05	ABCA13	ENSG00000204869	1.87	0.00080	LOC400706
ENSG00000184203	2.26	6.6e-05	PPP1R2	ENSG00000170759	1.77	0.00082	KIF5B
ENSG00000135842	2.87	7.6e-05	FAM129A	ENSG00000163950	2.41	0.00087	SLBP

Continuation of Table A.24

Ensembl	$\log_2$ Fold Change	padj	Gene Symbol	Ensembl	$\log_2$ Fold Change	padj	Gene Symbol
ENSG00000111640	1.41	8.6e-05	GAPDH	ENSG00000140941	2.56	0.00088	MAP1LC3B
ENSG00000072274	2.84	8.8e-05	TFRC	ENSG00000077420	1.46	0.00089	APBB1IP
ENSG00000180730	-3.17	8.8e-05	SHISA2	ENSG00000116489	1.44	0.00089	CAPZA1
ENSG00000159618	3.67	8.9e-04	ADGRG5	ENSG00000067560	1.26	0.00310	RHOA
ENSG00000197170	1.59	8.9e-04	PSMD12	ENSG00000166780	-3.31	0.00310	C16orf45
ENSG00000156802	2.39	9.7e-04	ATAD2	ENSG00000143549	1.25	0.00324	TPM3
ENSG00000168092	1.55	9.7e-04	PAFAH1B2	ENSG00000064989	3.15	0.00327	CALCRL
ENSG00000186184	1.25	9.7e-04	POLR1D	ENSG00000111011	1.29	0.00327	RSRC2
ENSG00000162923	2.45	1.0e-03	WDR26	ENSG00000138757	1.43	0.00327	G3BP2
ENSG00000184922	1.96	1.0e-03	FMNL1	ENSG00000147408	3.36	0.00327	CSGALNACT1
ENSG00000176723	-1.76	1.0e-03	ZNF843	ENSG00000148735	2.23	0.00327	PLEKHS1
ENSG00000151553	1.85	1.1e-03	FAM160B1	ENSG00000261740	1.37	0.00327	BOLA2-SMG1P6
ENSG00000211772	3.53	1.1e-03	NA	ENSG00000143751	1.98	0.00333	SDE2
ENSG00000144566	1.75	1.1e-03	RAB5A	ENSG00000124795	1.27	0.00338	DEK
ENSG00000187021	-1.82	1.1e-03	PNLIPRP1	ENSG00000058262	1.77	0.00338	SEC61A1
ENSG00000102172	1.65	1.2e-03	SMS	ENSG00000100647	1.52	0.00338	SUSD6
ENSG00000123358	-2.93	1.2e-03	NR4A1	ENSG00000102265	2.14	0.00338	TIMP1
ENSG00000034677	1.67	1.2e-03	RNF19A	ENSG00000108256	1.47	0.00338	NUFIP2
ENSG00000153246	-1.85	1.3e-03	PLA2R1	ENSG00000111711	2.68	0.00338	GOLT1B
ENSG00000130592	2.11	1.3e-03	LSP1	ENSG00000177694	-2.88	0.00338	NAALADL2
ENSG00000100097	3.65	1.3e-03	LGALS1	ENSG00000178691	1.37	0.00343	SUZ12
ENSG00000166387	-3.07	1.3e-03	PPFIBP2	ENSG00000169762	2.00	0.00344	TAPT1
ENSG00000149328	-2.14	1.3e-03	GLB1L2	ENSG00000172578	-1.46	0.00344	KLHL6
ENSG00000189376	2.29	1.3e-03	C8orf76	ENSG00000124688	1.73	0.00355	MAD2L1BP
ENSG00000155980	-3.74	1.4e-03	KIF5A	ENSG00000163735	-3.68	0.00357	CXCL5
ENSG00000034510	1.26	1.4e-03	TMSB10	ENSG00000177105	1.38	0.00358	RHOG
ENSG00000134375	1.62	1.4e-03	TIMM17A	ENSG00000139233	1.21	0.00369	LLPH
ENSG00000065548	1.57	1.4e-03	ZC3H15	ENSG00000182103	-2.79	0.00379	FAM181B
ENSG00000069956	1.78	1.4e-03	MAPK6	ENSG00000153201	1.81	0.00380	RANBP2
ENSG00000134109	1.93	1.5e-03	EDEM1	ENSG00000143702	1.24	0.00386	CEP170
ENSG00000162496	3.57	1.5e-03	DHRS3	ENSG00000155957	-1.22	0.00386	TMBIM4
ENSG00000197579	2.02	1.5e-03	TOPORS	ENSG00000180979	1.44	0.00386	LRRC57
ENSG00000174010	2.46	1.5e-03	KLHL15	ENSG00000135535	-1.78	0.00387	CD164
ENSG00000137497	-1.35	1.5e-03	NUMA1	ENSG00000165480	3.41	0.00387	SKA3
ENSG00000198380	2.13	1.5e-03	GFPT1	ENSG00000152380	1.68	0.00393	FAM151B
ENSG00000123384	-2.27	1.6e-03	LRP1	ENSG00000167460	1.75	0.00393	TPM4
ENSG00000135018	2.05	1.6e-03	UBQLN1	ENSG00000186660	1.84	0.00393	ZFP91
ENSG00000055332	1.61	1.6e-03	EIF2AK2	ENSG00000198178	2.08	0.00394	CLEC4C
ENSG00000205542	1.50	1.7e-03	TMSB4X	ENSG00000076108	1.27	0.00405	BAZ2A
ENSG00000059769	2.78	1.7e-03	DNAJC25	ENSG00000140829	2.24	0.00413	DHX38
ENSG00000106462	2.49	1.7e-03	EZH2	ENSG00000268738	-2.40	0.00421	HSFX2
ENSG00000123091	2.20	1.7e-03	RNF11	ENSG00000137462	3.12	0.00421	TLR2
ENSG00000119682	1.96	1.7e-03	AREL1	ENSG00000227268	-1.76	0.00425	KLLN
ENSG00000068745	1.19	1.8e-03	IP6K2	ENSG00000174004	2.70	0.00426	NRROS
ENSG00000198520	2.47	1.9e-03	C1orf228	ENSG00000103546	-3.44	0.00439	SLC6A2
ENSG00000115339	3.60	1.9e-03	GALNT3	ENSG00000165732	1.42	0.00441	DDX21
ENSG00000162616	2.28	1.9e-03	DNAJB4	ENSG00000054967	2.55	0.00462	RELT
ENSG00000115548	2.12	1.9e-03	KDM3A	ENSG00000130640	1.65	0.00465	TUBGCP2
ENSG00000277632	3.59	1.9e-03	CCL3	ENSG00000132334	2.46	0.00467	PTPRE
ENSG00000151332	1.49	1.9e-03	MBIP	ENSG00000151247	1.28	0.00476	EIF4E
ENSG00000183090	-3.47	1.9e-03	FREM3	ENSG00000197496	-2.35	0.00476	SLC2A10
ENSG00000164463	1.99	2.1e-03	CREBRF	ENSG00000119326	3.21	0.00477	CTNNAL1
ENSG00000172239	1.60	2.1e-03	PAIP1	ENSG00000205208	1.96	0.00486	C4orf46
ENSG00000180530	-1.93	2.1e-03	NRIP1	ENSG00000150316	1.41	0.00487	CWC15
ENSG00000136161	3.44	2.2e-03	RCBTB2	ENSG00000185480	2.52	0.00487	PARPBP
ENSG00000005022	1.43	2.2e-03	SLC25A5	ENSG00000113407	1.33	0.00488	TARS
ENSG00000035403	1.86	2.3e-03	VCL	ENSG00000198826	2.85	0.00488	ARHGAP11A
ENSG00000100644	2.37	2.3e-03	HIF1A	ENSG00000087842	-2.43	0.00493	PIR
ENSG00000151239	2.17	2.3e-03	TWF1	ENSG00000076344	-3.27	0.00500	RGS11
ENSG00000136758	1.30	2.3e-03	YME1L1	ENSG00000188529	1.14	0.00511	SRSF10
ENSG00000137947	1.80	2.3e-03	GTF2B	ENSG00000235162	3.34	0.00514	C12orf75
ENSG00000179833	1.26	2.3e-03	SERTAD2	ENSG00000096063	1.46	0.00531	SRPK1
ENSG00000075651	3.07	2.4e-03	PLD1	ENSG00000100280	1.92	0.00538	AP1B1
ENSG00000170340	2.01	2.4e-03	B3GNT2	ENSG00000142347	1.91	0.00538	MYO1F
ENSG00000163507	2.99	2.5e-03	KIAA1524	ENSG00000091317	1.59	0.00543	CMTM6
ENSG00000173020	2.02	2.5e-03	GRK2	ENSG00000160953	-2.61	0.00559	MUM1
ENSG00000196233	1.16	2.5e-03	C10orf12	ENSG00000152242	2.33	0.00562	C18orf25
ENSG00000078237	3.16	2.5e-03	TIGAR	ENSG00000136436	1.73	0.00571	CALCOCO2
ENSG00000187522	1.95	2.6e-03	HSPA14	ENSG00000127423	2.46	0.00572	AUNIP
ENSG00000198265	1.32	2.6e-03	HELZ	ENSG00000110318	-1.85	0.00579	CEP126
ENSG00000166596	-2.19	2.6e-03	CFAP52	ENSG00000108559	1.81	0.00589	NUP88
ENSG00000198492	1.29	2.7e-03	YTHDF2	ENSG00000169826	1.91	0.00589	CSGALNACT2
ENSG00000239998	2.89	2.7e-03	LILRA2	ENSG00000122122	2.45	0.00592	SASH3
ENSG00000005379	3.34	2.7e-03	TSPOAP1	ENSG00000082153	1.56	0.00594	BZW1
ENSG00000162378	1.51	2.7e-03	ZYG11B	ENSG00000157106	1.30	0.00596	SMG1
ENSG00000114904	3.10	2.7e-03	NEK4	ENSG00000165816	-3.09	0.00622	VWA2
ENSG00000117410	1.60	2.7e-03	ATP6V0B	ENSG00000138593	1.79	0.00624	SECISBP2L
ENSG00000090932	-3.23	2.8e-03	DLL3	ENSG00000166402	-3.27	0.00624	TUB
ENSG00000159346	1.71	2.8e-03	ADIPOR1	ENSG00000150627	-2.04	0.00624	WDR17
ENSG00000110324	3.50	2.9e-03	IL10RA	ENSG00000085840	3.10	0.00625	ORC1
ENSG00000062485	0.83	2.9e-03	CS	ENSG00000132507	1.72	0.00629	EIF5A
ENSG00000067082	1.93	2.9e-03	KLF6	ENSG00000107938	2.05	0.00631	EDRF1
ENSG00000144677	-2.85	2.9e-03	CTDSPL	ENSG00000177427	-2.00	0.00631	MIEF2
ENSG00000172530	1.80	2.9e-03	BANP	ENSG00000101654	1.51	0.00639	RNMT
ENSG00000011454	-1.71	3.0e-03	GPR21	ENSG00000118418	1.43	0.00639	HMGN3
ENSG00000141682	2.69	3.0e-03	PMAIP1	ENSG00000111696	2.42	0.00639	NT5DC3
ENSG00000163660	1.57	3.0e-03	CCNL1	ENSG00000273820	-2.94	0.00639	USP27X
ENSG00000185345	-1.72	3.0e-03	PARK2	ENSG00000141367	1.47	0.00641	CLTC
ENSG00000125970	1.76	6.6e-03	RALY	ENSG00000275052	0.79	0.01091	PPP4R3B
ENSG00000174599	-2.93	6.6e-03	TRAM1L1	ENSG00000271605	2.76	0.01102	MILR1

Continuation of Table A.24

Ensembl	log_2 Fold Change	padj	Gene Symbol	Ensembl	log_2 Fold Change	padj	Gene Symbol
ENSG00000132849	-2.19	6.6e-03	PATJ	ENSG00000090061	1.50	0.01102	CCNK
ENSG00000163029	1.61	6.6e-03	SMC6	ENSG00000177125	2.18	0.01132	ZBTB34
ENSG00000059804	2.70	6.7e-03	SLC2A3	ENSG00000116731	1.53	0.01138	PRDM2
ENSG00000116030	1.29	6.7e-03	SUMO1	ENSG00000100650	-0.74	0.01145	SRSF5
ENSG00000204851	-3.20	6.7e-03	PNMAL2	ENSG00000138663	1.08	0.01159	COPS4
ENSG00000112245	1.81	6.7e-03	PTP4A1	ENSG00000092853	2.72	0.01174	CLSPN
ENSG00000168488	1.58	6.7e-03	ATXN2L	ENSG00000100519	1.28	0.01179	PSMC6
ENSG00000186197	-3.30	6.9e-03	EDARADD	ENSG00000189159	1.87	0.01179	HN1
ENSG00000007237	2.12	6.9e-03	GAS7	ENSG00000182578	3.02	0.01184	CSF1R
ENSG00000160321	-1.82	6.9e-03	ZNF208	ENSG00000069399	2.70	0.01185	BCL3
ENSG00000116560	1.31	6.9e-03	SFPQ	ENSG00000135679	1.08	0.01185	MDM2
ENSG00000112799	3.28	7.0e-03	LY86	ENSG00000144848	1.35	0.01216	ATG3
ENSG00000125037	1.37	7.0e-03	EMC3	ENSG00000154734	-2.57	0.01218	ADAMTS1
ENSG00000142669	1.64	7.0e-03	SH3BGRL3	ENSG00000124562	1.06	0.01224	SNRPC
ENSG00000153914	1.27	7.0e-03	SREK1	ENSG00000010244	1.27	0.01224	ZNF207
ENSG00000151465	1.34	7.1e-03	CDC123	ENSG00000099783	1.24	0.01242	HNRNPM
ENSG00000112739	1.18	7.1e-03	PRPF4B	ENSG00000103534	2.19	0.01242	TMC5
ENSG00000127903	-1.68	7.3e-03	ZNF835	ENSG00000196655	1.66	0.01243	TRAPPC4
ENSG00000153037	1.49	7.3e-03	SRP19	ENSG00000164935	-2.72	0.01259	DCSTAMP
ENSG00000123983	2.05	7.5e-03	ACSL3	ENSG00000185728	1.74	0.01259	YTHDF3
ENSG00000127334	2.53	7.6e-03	DYRK2	ENSG00000180667	2.27	0.01260	YOD1
ENSG00000197343	1.67	7.6e-03	ZNF655	ENSG00000077721	1.48	0.01270	UBE2A
ENSG00000123473	2.47	7.7e-03	STIL	ENSG00000151883	1.95	0.01270	PARP8
ENSG00000143862	2.06	7.7e-03	ARL8A	ENSG00000121895	2.17	0.01271	TMEM156
ENSG00000051523	2.61	7.8e-03	CYBA	ENSG00000184602	3.02	0.01271	SNN
ENSG00000136381	1.25	8.0e-03	IREB2	ENSG00000185885	2.17	0.01279	IFITM1
ENSG00000187391	-1.36	8.3e-03	MAGI2	ENSG00000197959	-3.29	0.01309	DNM3
ENSG00000084463	1.52	8.3e-03	WBP11	ENSG00000139190	2.79	0.01318	VAMP1
ENSG00000166405	-2.53	8.3e-03	RIC3	ENSG00000179119	1.71	0.01325	SPTY2D1
ENSG00000111605	1.19	8.3e-03	CPSF6	ENSG00000078177	1.45	0.01331	N4BP2
ENSG00000080166	-3.02	8.3e-03	DCT	ENSG00000103035	1.36	0.01331	PSMD7
ENSG00000168209	2.63	8.4e-03	DDIT4	ENSG00000010017	1.90	0.01341	RANBP9
ENSG00000172322	3.21	8.4e-03	CLEC12A	ENSG00000147251	1.28	0.01341	DOCK11
ENSG00000175895	1.97	8.5e-03	PLEKHF2	ENSG00000068878	1.86	0.01364	PSME4
ENSG00000156475	-2.33	8.5e-03	PPP2R2B	ENSG00000138434	2.15	0.01365	SSFA2
ENSG00000140455	1.46	8.7e-03	USP3	ENSG00000142634	1.51	0.01366	EFHD2
ENSG00000141480	1.70	8.7e-03	ARRB2	ENSG00000082258	1.66	0.01379	CCNT2
ENSG00000156232	1.90	8.7e-03	WHAMM	ENSG00000144354	2.73	0.01381	CDCA7
ENSG00000179862	3.17	8.7e-03	CITED4	ENSG00000084093	1.17	0.01401	REST
ENSG00000134453	1.54	8.9e-03	RBM17	ENSG00000095015	1.99	0.01404	MAP3K1
ENSG00000181555	1.50	8.9e-03	SETD2	ENSG00000105137	-2.81	0.01410	SYDE1
ENSG00000123728	1.62	8.9e-03	RAP2C	ENSG00000134758	1.75	0.01416	RNF138
ENSG00000213462	2.07	9.0e-03	ERV3-1	ENSG00000174574	1.30	0.01433	AKIRIN1
ENSG00000137812	2.83	9.1e-03	KNL1	ENSG00000114784	1.68	0.01436	EIF1B
ENSG00000112339	1.43	9.1e-03	HBS1L	ENSG00000013441	1.69	0.01440	CLK1
ENSG00000128309	1.26	9.1e-03	MPST	ENSG00000186523	-2.39	0.01456	FAM86B1
ENSG00000101557	1.49	9.2e-03	USP14	ENSG00000095303	-2.81	0.01491	PTGS1
ENSG00000141298	1.58	9.2e-03	SSH2	ENSG00000091651	1.26	0.01503	ORC6
ENSG00000176014	2.93	9.2e-03	TUBB6	ENSG00000276043	2.14	0.01507	UHRF1
ENSG00000182013	-1.68	9.2e-03	PNMAL1	ENSG00000107951	1.28	0.01508	MTPAP
ENSG00000120279	-3.31	9.3e-03	MYCT1	ENSG00000258818	-2.34	0.01508	RNASE4
ENSG00000099246	1.79	9.3e-03	RAB18	ENSG00000265190	2.61	0.01508	ANXA8
ENSG00000114450	1.50	9.3e-03	GNB4	ENSG00000068308	1.70	0.01532	OTUD5
ENSG00000149823	-1.57	9.3e-03	VPS51	ENSG00000182985	1.64	0.01532	CADM1
ENSG00000022840	1.56	9.3e-03	RNF10	ENSG00000184983	0.95	0.01532	NDUFA6
ENSG00000140057	-1.36	9.3e-03	AK7	ENSG00000153404	-2.61	0.01541	PLEKHG4B
ENSG00000123136	1.29	9.6e-03	DDX39A	ENSG00000167378	1.87	0.01543	IRGQ
ENSG00000134602	1.45	9.6e-03	STK26	ENSG00000135077	2.39	0.01549	HAVCR2
ENSG00000166979	-2.27	9.7e-03	EVA1C	ENSG00000142185	3.02	0.01549	TRPM2
ENSG00000137040	2.01	9.7e-03	RANBP6	ENSG00000170144	0.93	0.01558	HNRNPA3
ENSG00000133703	1.73	9.7e-03	KRAS	ENSG00000176532	-2.89	0.01565	PRR15
ENSG00000169155	2.16	9.7e-03	ZBTB43	ENSG00000083799	1.65	0.01577	CYLD
ENSG00000067167	1.39	9.8e-03	TRAM1	ENSG00000112530	-1.99	0.01583	PACRG
ENSG00000198833	1.76	9.8e-03	UBE2J1	ENSG00000178074	1.70	0.01583	C2orf69
ENSG00000006042	-3.00	9.9e-03	TMEM98	ENSG00000051341	2.57	0.01583	POLQ
ENSG00000024422	-2.35	9.9e-03	EHD2	ENSG00000091513	1.78	0.01583	TF
ENSG00000159259	1.57	1.0e-02	CHAF1B	ENSG00000089818	1.61	0.01591	NECAP1
ENSG00000152433	-2.85	1.0e-02	ZNF547	ENSG00000138668	1.00	0.01596	HNRNPD
ENSG00000008083	1.39	1.0e-02	JARID2	ENSG00000221963	1.54	0.01598	APOL6
ENSG00000185559	-2.94	1.0e-02	DLK1	ENSG00000100902	1.17	0.01604	PSMA6
ENSG00000100941	1.04	1.0e-02	PNN	ENSG00000167105	-2.46	0.01604	TMEM92
ENSG00000204118	-2.20	1.0e-02	NAP1L6	ENSG00000182944	0.92	0.01606	EWSR1
ENSG00000187735	0.97	1.0e-02	TCEA1	ENSG00000059728	2.12	0.01622	MXD1
ENSG00000176853	1.77	1.0e-02	FAM91A1	ENSG00000076555	-1.37	0.01622	ACACB
ENSG00000198838	-1.91	1.1e-02	RYR3	ENSG00000178741	1.73	0.01633	COX5A
ENSG00000152253	1.30	1.1e-02	SPC25	ENSG00000151092	1.20	0.01665	NGLY1
ENSG00000102858	1.25	1.1e-02	MGRN1	ENSG00000129518	1.17	0.01667	EAPP
ENSG00000152270	2.43	1.1e-02	PDE3B	ENSG00000164032	1.47	0.01667	H2AFZ
ENSG00000115808	1.97	1.1e-02	STRN	ENSG00000062716	1.91	0.01687	VMP1
ENSG00000198298	-2.93	1.1e-02	ZNF485	ENSG00000119707	1.04	0.01687	RBM25
ENSG00000107968	2.20	1.1e-02	MAP3K8	ENSG00000083457	1.33	0.01702	ITGAE
ENSG00000148362	2.39	1.1e-02	C9orf142	ENSG00000102158	1.57	0.01707	MAGT1
ENSG00000159871	-1.48	1.1e-02	LYPD5	ENSG00000139620	1.49	0.01707	KANSL2
ENSG00000018236	-2.18	1.7e-02	CNTN1	ENSG00000137218	2.06	0.02784	FRS3
ENSG00000095574	1.82	1.7e-02	IKZF5	ENSG00000148143	-2.58	0.02789	ZNF462
ENSG00000100600	2.06	1.8e-02	LGMN	ENSG00000144908	2.24	0.02798	ALDH1L1
ENSG00000102024	-2.43	1.8e-02	PLS3	ENSG00000146833	-1.26	0.02810	TRIM4
ENSG00000166803	2.52	1.8e-02	KIAA0101	ENSG00000151014	1.96	0.02810	NOCT
ENSG00000129128	1.30	1.8e-02	SPCS3	ENSG00000107443	1.85	0.02819	CCNJ
ENSG00000141505	-1.39	1.8e-02	ASGR1	ENSG00000171160	-2.40	0.02819	MORN4
ENSG00000161547	1.07	1.8e-02	SRSF2	ENSG00000171791	2.12	0.02819	BCL2

Continuation of Table A.24

Ensembl	log_2 Fold Change	padj	Gene Symbol	Ensembl	log_2 Fold Change	padj	Gene Symbol
ENSG00000087470	1.23	1.8e-02	DNM1L	ENSG00000172687	1.59	0.02819	NA
ENSG00000126524	1.80	1.8e-02	SBDS	ENSG00000188647	1.53	0.02819	PTAR1
ENSG00000163376	3.01	1.9e-02	KBTBD8	ENSG00000185721	1.01	0.03707	DRG1
ENSG00000165702	-2.15	1.9e-02	GFI1B	ENSG00000140623	-2.07	0.03711	SEPT12
ENSG00000068400	1.31	1.9e-02	GRIPAP1	ENSG00000108055	1.15	0.03716	SMC3
ENSG00000162599	-2.82	1.9e-02	NFIA	ENSG00000118939	1.48	0.03727	UCHL3
ENSG00000181856	-1.82	1.9e-02	SLC2A4	ENSG00000180257	2.33	0.03727	ZNF816
ENSG00000143621	1.06	1.9e-02	ILF2	ENSG00000171877	2.77	0.03731	FRMD5
ENSG00000135114	2.94	1.9e-02	OASL	ENSG00000204252	-2.04	0.03743	HLA-DOA
ENSG00000197299	2.00	1.9e-02	BLM	ENSG00000197312	2.46	0.02898	DDI2
ENSG00000170242	1.26	1.9e-02	USP47	ENSG00000164944	1.42	0.02918	KIAA1429
ENSG00000008018	0.94	1.9e-02	PSMB1	ENSG00000138600	1.34	0.02919	SPPL2A
ENSG00000153922	1.80	1.9e-02	CHD1	ENSG00000173011	2.45	0.02920	TADA2B
ENSG00000111261	-1.26	1.9e-02	MANSC1	ENSG00000151461	0.88	0.02930	UPF2
ENSG00000125812	2.69	1.9e-02	GZF1	ENSG00000167778	-2.88	0.02954	SPRYD3
ENSG00000154721	-2.30	1.9e-02	JAM2	ENSG00000101624	1.65	0.02975	CEP76
ENSG00000259417	-2.23	2.0e-02	LINC01314	ENSG00000115128	1.03	0.02975	SF3B6
ENSG00000102606	2.16	2.0e-02	ARHGEF7	ENSG00000130024	1.44	0.02975	PHF10
ENSG00000137804	2.51	2.0e-02	NUSAP1	ENSG00000134709	-1.97	0.02975	HOOK1
ENSG00000165525	0.73	2.0e-02	NEMF	ENSG00000134744	1.15	0.02975	ZCCHC11
ENSG00000173542	1.78	2.0e-02	MOB1B	ENSG00000183684	1.25	0.03011	ALYREF
ENSG00000154582	0.90	2.1e-02	TCEB1	ENSG00000089820	1.30	0.03012	ARHGAP4
ENSG00000116604	2.06	2.1e-02	MEF2D	ENSG00000168411	1.81	0.03021	RFWD3
ENSG00000151491	2.58	2.1e-02	EPS8	ENSG00000156875	1.55	0.03022	MFSD14A
ENSG00000184445	1.47	2.1e-02	KNTC1	ENSG00000213366	-1.96	0.03022	GSTM2
ENSG00000120705	1.70	2.2e-02	ETF1	ENSG00000124659	1.17	0.03034	TBCC
ENSG00000170852	1.58	2.2e-02	KBTBD2	ENSG00000245848	2.66	0.03034	CEBPA
ENSG00000125772	1.44	2.2e-02	GPCPD1	ENSG00000178381	1.60	0.03037	ZFAND2A
ENSG00000101290	1.54	2.2e-02	CDS2	ENSG00000129204	-1.75	0.03044	USP6
ENSG00000036054	1.57	2.2e-02	TBC1D23	ENSG00000173744	1.58	0.03044	AGFG1
ENSG00000278053	1.01	2.2e-02	DDX52	ENSG00000150093	1.52	0.03051	ITGB1
ENSG00000077458	1.68	2.2e-02	FAM76B	ENSG00000124469	-2.16	0.03082	CEACAM8
ENSG00000188585	-3.10	2.2e-02	LINC00083	ENSG00000122482	1.01	0.03108	ZNF644
ENSG00000196504	1.13	2.2e-02	PRPF40A	ENSG00000138674	0.75	0.03108	SEC31A
ENSG00000076641	2.59	2.3e-02	PAG1	ENSG00000275895	2.50	0.03108	U2AF1L5
ENSG00000114796	1.99	2.3e-02	KLHL24	ENSG00000071967	-1.52	0.03110	CYBRD1
ENSG00000116754	1.04	2.3e-02	SRSF11	ENSG00000147650	1.56	0.03118	LRP12
ENSG00000137815	1.27	2.3e-02	RTF1	ENSG00000258102	2.31	0.03136	MAP1LC3B2
ENSG00000180198	1.11	2.3e-02	RCC1	ENSG00000132823	1.90	0.03147	OSER1
ENSG00000075415	0.92	2.3e-02	SLC25A3	ENSG00000143147	-2.13	0.03147	GPR161
ENSG00000066382	-2.30	2.3e-02	MPPED2	ENSG00000157554	-1.62	0.03149	ERG
ENSG00000136527	1.24	2.3e-02	TRA2B	ENSG00000172795	1.86	0.03176	DCP2
ENSG00000032219	1.76	2.3e-02	ARID4A	ENSG00000107036	1.63	0.03178	RIC1
ENSG00000186472	-3.06	2.3e-02	PCLO	ENSG00000145700	-2.67	0.03180	ANKRD31
ENSG00000104888	1.67	2.3e-02	SLC17A7	ENSG00000173327	1.46	0.03180	MAP3K11
ENSG00000134245	-2.30	2.4e-02	WNT2B	ENSG00000177311	-1.50	0.03180	ZBTB38
ENSG00000167601	-2.20	2.4e-02	AXL	ENSG00000113916	2.34	0.03190	BCL6
ENSG00000176407	1.46	2.4e-02	KCMF1	ENSG00000182557	2.49	0.03192	SPNS3
ENSG00000143093	1.46	2.4e-02	STRIP1	ENSG00000163932	2.62	0.03203	PRKCD
ENSG00000183255	1.16	2.4e-02	PTTG1IP	ENSG00000170509	2.14	0.03203	HSD17B13
ENSG00000156273	1.74	2.4e-02	BACH1	ENSG00000186399	-1.65	0.03203	GOLGA8R
ENSG00000164754	1.16	2.4e-02	RAD21	ENSG00000196345	-2.10	0.03231	ZKSCAN7
ENSG00000131370	2.48	2.4e-02	SH3BP5	ENSG00000122642	-2.48	0.03241	FKBP9
ENSG00000143761	1.09	2.4e-02	ARF1	ENSG00000070010	1.17	0.03250	UFD1L
ENSG00000152556	-2.10	2.4e-02	PFKM	ENSG00000093072	2.42	0.03250	CECR1
ENSG00000204711	-2.01	2.4e-02	C9orf135	ENSG00000105708	2.09	0.03250	ZNF14
ENSG00000166200	1.62	2.5e-02	COPS2	ENSG00000113575	1.36	0.03250	PPP2CA
ENSG00000065978	1.10	2.5e-02	YBX1	ENSG00000158710	1.32	0.03253	TAGLN2
ENSG00000241973	1.06	2.5e-02	PI4KA	ENSG00000132970	-2.94	0.03271	WASF3
ENSG00000169021	1.29	2.6e-02	UQCRFS1	ENSG00000129467	-2.30	0.03272	ADCY4
ENSG00000239305	1.81	2.6e-02	RNF103	ENSG00000096401	1.35	0.03277	CDC5L
ENSG00000057757	1.23	2.6e-02	PITHD1	ENSG00000172757	1.13	0.03290	CFL1
ENSG00000167470	1.61	2.6e-02	MIDN	ENSG00000197128	-2.17	0.03303	ZNF772
ENSG00000133773	1.03	2.6e-02	CCDC59	ENSG00000100239	1.59	0.03312	PPP6R2
ENSG00000177839	1.16	2.6e-02	PCDHB9	ENSG00000175575	-1.22	0.03320	PAAF1
ENSG00000065621	-1.34	2.6e-02	GSTO2	ENSG00000183856	2.83	0.03336	IQGAP3
ENSG00000104738	1.94	2.7e-02	MCM4	ENSG00000280230	2.02	0.03336	NA
ENSG00000177842	-2.19	2.7e-02	ZNF620	ENSG00000164938	2.13	0.03343	TP53INP1
ENSG00000100567	0.90	2.7e-02	PSMA3	ENSG00000103353	1.77	0.03353	UBFD1
ENSG00000117724	2.29	2.7e-02	CENPF	ENSG00000150403	1.27	0.03353	TMCO3
ENSG00000158796	1.30	2.7e-02	DEDD	ENSG00000213928	1.81	0.03353	IRF9
ENSG00000111615	1.18	2.7e-02	KRR1	ENSG00000173575	1.34	0.03421	CHD2
ENSG00000144031	-2.49	2.7e-02	ANKRD53	ENSG00000155893	1.56	0.03459	PXYLP1
ENSG00000182983	-1.92	2.7e-02	ZNF662	ENSG00000149269	1.64	0.03469	PAK1
ENSG00000072694	2.80	2.7e-02	FCGR2B	ENSG00000150687	1.04	0.03472	PRSS23
ENSG00000153207	1.70	2.7e-02	AHCTF1	ENSG00000123066	1.41	0.03474	MED13L
ENSG00000165609	1.17	2.8e-02	NUDT5	ENSG00000120438	1.00	0.03495	TCP1
ENSG00000185019	-1.68	3.6e-02	UBOX5	ENSG00000184857	-2.38	0.03501	TMEM186
ENSG00000102760	2.30	3.6e-02	RGCC	ENSG00000152291	0.86	0.03508	TGOLN2
ENSG00000172296	-1.43	3.6e-02	SPTLC3	ENSG00000111832	0.78	0.03532	RWDD1
ENSG00000160410	1.34	3.6e-02	SHKBP1	ENSG00000145675	1.61	0.03532	PIK3R1
ENSG00000154330	-2.21	3.7e-02	PGM5	ENSG00000179915	-1.65	0.03532	NRXN1
ENSG00000130816	1.34	3.7e-02	DNMT1	ENSG00000196935	-1.10	0.03552	SRGAP1
ENSG00000163462	-1.91	3.7e-02	TRIM46	ENSG00000121390	1.28	0.03565	PSPC1
ENSG00000163636	0.91	3.7e-02	PSMD6	ENSG00000132432	1.04	0.04873	SEC61G
ENSG00000169598	-1.40	3.7e-02	DFFB	ENSG00000004139	-2.06	0.04917	SARM1
ENSG00000162511	1.43	3.8e-02	LAPTM5	ENSG00000094804	2.54	0.04961	CDC6
ENSG00000119318	1.37	3.8e-02	RAD23B	ENSG00000132522	1.17	0.04962	GPS2
ENSG00000129534	1.24	3.8e-02	MIS18BP1	ENSG00000152082	1.70	0.04962	MZT2B
ENSG00000197646	-2.36	3.8e-02	PDCD1LG2	ENSG00000068489	1.29	0.04962	PRR11
ENSG00000111775	1.07	3.8e-02	COX6A1	ENSG00000152332	1.13	0.04991	UHMK1

Continuation of Table A.24

Ensembl	log_2 Fold Change	padj	Gene Symbol	Ensembl	log_2 Fold Change	padj	Gene Symbol
ENSG00000182247	2.79	3.8e-02	UBE2E2	ENSG00000170873	2.31	0.04993	MTSS1
ENSG00000048405	1.78	3.9e-02	ZNF800	ENSG00000197747	2.44	0.04999	S100A10
ENSG00000186862	-1.05	3.9e-02	PDZD7				
ENSG00000004487	1.08	3.9e-02	KDM1A				
ENSG00000152689	1.94	3.9e-02	RASGRP3				
ENSG00000163655	1.04	3.9e-02	GMPS				
ENSG00000104312	1.79	3.9e-02	RIPK2				
ENSG00000165355	1.54	3.9e-02	FBXO33				
ENSG00000102572	1.57	3.9e-02	STK24				
ENSG00000146858	1.30	3.9e-02	ZC3HAV1L				
ENSG00000240053	1.99	3.9e-02	LY6G5B				
ENSG00000159873	1.60	4.0e-02	CCDC117				
ENSG00000103187	2.09	4.0e-02	COTL1				
ENSG00000129195	2.33	4.0e-02	FAM64A				
ENSG00000091527	1.16	4.1e-02	CDV3				
ENSG00000101017	-2.64	4.1e-02	CD40				
ENSG00000104131	1.31	4.1e-02	EIF3J				
ENSG00000132963	1.46	4.1e-02	POMP				
ENSG00000170915	-2.31	4.1e-02	PAQR8				
ENSG00000107447	2.77	4.1e-02	DNTT				
ENSG00000136891	1.30	4.1e-02	TEX10				
ENSG00000181626	-1.61	4.1e-02	ANKRD62				
ENSG00000105355	1.26	4.1e-02	PLIN3				
ENSG00000123146	1.92	4.1e-02	ADGRE5				
ENSG00000159377	0.89	4.1e-02	PSMB4				
ENSG00000107679	-1.50	4.2e-02	PLEKHA1				
ENSG00000102753	1.57	4.2e-02	KPNA3				
ENSG00000159788	1.28	4.2e-02	RGS12				
ENSG00000105364	1.68	4.2e-02	MRPL4				
ENSG00000113742	2.07	4.2e-02	CPEB4				
ENSG00000166902	-1.41	4.2e-02	MRPL16				
ENSG00000104714	-1.21	4.2e-02	ERICH1				
ENSG00000099999	-1.38	4.2e-02	RNF215				
ENSG00000106588	0.93	4.2e-02	PSMA2				
ENSG00000267221	-1.25	4.3e-02	ZNF385C				
ENSG00000073169	2.39	4.3e-02	SELENOO				
ENSG00000123595	1.76	4.3e-02	RAB9A				
ENSG00000108654	0.86	4.3e-02	DDX5				
ENSG00000109756	-1.49	4.3e-02	RAPGEF2				
ENSG00000185823	1.40	4.4e-02	NPAP1				
ENSG00000135828	-1.83	4.4e-02	RNASEL				
ENSG00000131747	2.74	4.4e-02	TOP2A				
ENSG00000065243	1.12	4.4e-02	PKN2				
ENSG00000115020	1.26	4.5e-02	PIKFYVE				
ENSG00000159958	1.54	4.5e-02	TNFRSF13C				
ENSG00000005961	-2.15	4.5e-02	ITGA2B				
ENSG00000198939	-2.73	4.5e-02	ZFP2				
ENSG00000198003	-2.65	4.5e-02	CCDC151				
ENSG00000180776	1.64	4.5e-02	ZDHHC20				
ENSG00000029363	0.97	4.6e-02	BCLAF1				
ENSG00000136810	0.83	4.6e-02	TXN				
ENSG00000145041	1.69	4.6e-02	DCAF1				
ENSG00000265354	1.05	4.6e-02	TIMM23				
ENSG00000099998	1.94	4.6e-02	GGT5				
ENSG00000185651	0.93	4.6e-02	UBE2L3				
ENSG00000162630	2.67	4.6e-02	B3GALT2				
ENSG00000107262	1.34	4.6e-02	BAG1				
ENSG00000120802	1.24	4.6e-02	TMPO				
ENSG00000121073	1.18	4.7e-02	SLC35B1				
ENSG00000081377	-2.43	4.7e-02	CDC14B				
ENSG00000197647	1.71	4.7e-02	ZNF433				
ENSG00000064692	2.07	4.8e-02	SNCAIP				
ENSG00000039068	-1.05	4.8e-02	CDH1				
ENSG00000144567	-1.19	4.8e-02	FAM134A				

Table A.27: DEGs Between Normal LMPP and GMP. DEGs were identified between normal LMPP and GMP, using a Wald test in DESeq2. A design formula of ~replicate + population was used. Upregulated and downregulated DEGs with a p-adjusted value < 0.05 are shown.

Ensembl	$\log_2$ Fold Change	padj	Gene Symbol	Ensembl	$\log_2$ Fold Change	padj	Gene Symbol
ENSG00000275896	8.33	5.7e-98	PRSS2	ENSG00000131042	4.53	5.1e-09	LILRB2
ENSG00000197561	-8.72	1.5e-87	ELANE	ENSG00000111790	2.41	5.8e-09	FGFR1OP2
ENSG00000169385	-9.45	1.0e-80	RNASE2	ENSG00000174500	-4.11	6.2e-09	GCSAM
ENSG00000136929	6.28	4.5e-73	HEMGN	ENSG00000117586	2.68	1.0e-08	TNFSF4
ENSG00000124491	-9.04	1.1e-61	F13A1	ENSG00000168497	2.62	1.1e-08	SDPR
ENSG00000171476	4.42	5.5e-54	HOPX	ENSG00000122862	-2.64	1.1e-08	SRGN
ENSG00000153253	7.30	1.1e-49	SCN3A	ENSG00000100983	-2.48	1.3e-08	GSS
ENSG00000071575	7.40	5.0e-48	TRIB2	ENSG00000167851	2.30	1.4e-08	CD300A
ENSG00000090382	-6.48	7.3e-47	LYZ	ENSG00000133106	4.63	1.5e-08	EPSTI1
ENSG00000196415	-8.68	1.5e-45	PRTN3	ENSG00000010165	-2.83	1.8e-08	METTL13
ENSG00000169397	-8.59	2.2e-45	RNASE3	ENSG00000119535	-2.18	2.3e-08	CSF3R
ENSG00000172232	-7.49	2.5e-45	AZU1	ENSG00000185630	2.94	2.3e-08	PBX1
ENSG00000084110	-8.15	1.2e-39	HAL	ENSG00000111012	-4.89	3.5e-08	CYP27B1
ENSG00000121552	-7.68	1.1e-34	CSTA	ENSG00000104447	2.21	5.0e-08	TRPS1
ENSG00000169442	3.28	1.2e-31	CD52	ENSG00000105854	2.14	5.7e-08	PON2
ENSG00000221869	-7.44	7.7e-28	CEBPD	ENSG00000116260	-3.30	5.8e-08	QSOX1
ENSG00000171509	6.96	1.5e-27	RXFP1	ENSG00000112299	4.27	6.2e-08	VNN1
ENSG00000163563	-7.37	1.9e-27	MNDA	ENSG00000204482	2.15	6.5e-08	LST1
ENSG00000005381	-6.37	3.3e-26	MPO	ENSG00000150867	2.25	6.9e-08	PIP4K2A
ENSG00000123836	3.18	7.0e-25	PFKFB2	ENSG00000153904	-4.72	6.9e-08	DDAH1
ENSG00000149516	-6.90	7.2e-25	MS4A3	ENSG00000136603	2.89	7.5e-08	SKIL
ENSG00000262655	6.74	3.9e-24	SPON1	ENSG00000100644	1.79	7.7e-08	HIF1A
ENSG00000113749	-6.98	3.6e-23	HRH2	ENSG00000078399	2.47	8.0e-08	HOXA9
ENSG00000135535	3.26	6.6e-23	CD164	ENSG00000111052	-4.58	8.3e-08	LIN7A
ENSG00000100678	6.52	9.8e-23	SLC8A3	ENSG00000138829	-4.78	8.6e-08	FBN2
ENSG00000145708	4.02	7.3e-21	CRHBP	ENSG00000179869	3.06	8.6e-08	ABCA13
ENSG00000136869	-6.74	1.8e-19	TLR4	ENSG00000154188	2.31	9.6e-08	ANGPT1
ENSG00000126353	6.06	1.6e-18	CCR7	ENSG00000174059	1.75	9.6e-08	CD34
ENSG00000012124	4.65	1.3e-17	CD22	ENSG00000084234	-1.75	1.3e-07	APLP2
ENSG00000150347	2.60	1.6e-17	ARID5B	ENSG00000137628	3.04	1.3e-07	DDX60
ENSG00000124780	4.24	2.7e-17	KCNK17	ENSG00000154920	-4.71	1.4e-07	EME1
ENSG00000183690	4.94	3.8e-17	EFHC2	ENSG00000128040	1.62	1.4e-07	SPINK2
ENSG00000152217	5.93	6.9e-17	SETBP1	ENSG00000124275	2.55	1.5e-07	MTRR
ENSG00000242550	-4.32	2.8e-16	SERPINB10	ENSG00000133101	-4.69	1.6e-07	CCNA1
ENSG00000144893	3.01	7.1e-16	MED12L	ENSG00000121236	-3.57	1.6e-07	TRIM6
ENSG00000179218	-2.21	8.9e-16	CALR	ENSG00000172738	-4.36	1.6e-07	TMEM217
ENSG00000176076	-6.14	6.6e-15	KCNE5	ENSG00000065809	-2.59	1.7e-07	FAM107B
ENSG00000127955	3.28	8.5e-15	GNAI1	ENSG00000155097	2.57	1.8e-07	ATP6V1C1
ENSG00000134324	2.43	5.7e-14	LPIN1	ENSG00000139915	4.34	2.1e-07	MDGA2
ENSG00000008394	-2.75	6.0e-14	MGST1	ENSG00000203880	1.90	2.1e-07	PCMTD2
ENSG00000042980	2.51	6.7e-14	ADAM28	ENSG00000008311	3.21	2.2e-07	AASS
ENSG00000139163	2.39	9.4e-14	ETNK1	ENSG00000204524	4.46	2.2e-07	ZNF805
ENSG00000169224	3.44	2.7e-13	GCSAML	ENSG00000150760	2.85	2.3e-07	DOCK1
ENSG00000143669	-2.97	3.6e-13	LYST	ENSG00000124882	-1.94	2.6e-07	EREG
ENSG00000196139	2.80	4.2e-13	AKR1C3	ENSG00000102265	2.14	2.6e-07	TIMP1
ENSG00000211772	2.58	4.4e-13	NA	ENSG00000119801	2.02	2.7e-07	YPEL5
ENSG00000163751	-3.72	5.4e-13	CPA3	ENSG00000163536	4.42	2.9e-07	SERPINI1
ENSG00000115738	2.05	7.1e-13	ID2	ENSG00000170260	-4.01	3.0e-07	ZNF212
ENSG00000164938	3.72	8.9e-13	TP53INP1	ENSG00000196549	3.18	3.1e-07	MME
ENSG00000117399	-5.24	5.1e-12	CDC20	ENSG00000082074	-4.21	3.5e-07	FYB
ENSG00000164035	3.96	8.8e-12	EMCN	ENSG00000197008	2.34	3.5e-07	ZNF138
ENSG00000197208	-5.72	1.1e-11	SLC22A4	ENSG00000184226	2.10	3.6e-07	PCDH9
ENSG00000196628	2.10	1.2e-11	TCF4	ENSG00000124193	1.77	4.2e-07	SRSF6
ENSG00000082438	2.92	1.9e-11	COBLL1	ENSG00000162894	4.05	4.4e-07	FCMR
ENSG00000143995	2.26	2.2e-11	MEIS1	ENSG00000082512	1.82	4.8e-07	TRAF5
ENSG00000162892	5.57	2.2e-11	IL24	ENSG00000197766	-3.88	5.1e-07	CFD
ENSG00000197183	2.60	3.2e-11	NOL4L	ENSG00000110848	1.61	6.4e-07	CD69
ENSG00000162374	5.39	3.4e-11	ELAVL4	ENSG00000012048	-2.25	7.1e-07	BRCA1
ENSG00000105997	3.27	3.6e-11	HOXA3	ENSG00000169252	-3.86	9.7e-07	ADRB2
ENSG00000109501	5.35	7.2e-11	WFS1	ENSG00000008513	1.96	1.1e-06	ST3GAL1
ENSG00000100116	-4.16	1.2e-10	GCAT	ENSG00000070018	3.27	1.3e-06	LRP6
ENSG00000165140	-5.43	2.4e-10	FBP1	ENSG00000172869	2.03	1.4e-06	DMXL1
ENSG00000113100	3.85	2.8e-10	CDH9	ENSG00000103202	-2.60	1.6e-06	NME4
ENSG00000147465	-4.39	2.9e-10	STAR	ENSG00000002586	1.50	1.6e-06	CD99
ENSG00000172322	-4.21	4.8e-10	CLEC12A	ENSG00000179862	-4.44	1.6e-06	CITED4
ENSG00000119866	2.08	6.5e-10	BCL11A	ENSG00000134575	4.23	1.9e-06	ACP2
ENSG00000099219	2.13	8.0e-10	ERMP1	ENSG00000138722	2.34	1.9e-06	MMRN1
ENSG00000066294	-2.49	1.2e-09	CD84	ENSG00000005379	-3.78	2.0e-06	TSPOAP1
ENSG00000102804	2.05	1.2e-09	TSC22D1	ENSG00000141232	1.92	2.2e-06	TOB1
ENSG00000109684	3.32	1.2e-09	CLNK	ENSG00000090104	2.24	2.2e-06	RGS1
ENSG00000196371	-3.95	1.7e-09	FUT4	ENSG00000156928	-3.48	2.2e-06	MALSU1
ENSG00000168306	5.16	1.9e-09	ACOX2	ENSG00000105472	-2.75	2.3e-06	CLEC11A
ENSG00000112414	3.18	1.9e-09	ADGRG6	ENSG00000123643	-4.31	2.5e-06	SLC36A1
ENSG00000121807	-5.19	1.9e-09	CCR2	ENSG00000170476	2.04	2.6e-06	MZB1
ENSG00000143321	-2.52	2.0e-09	HDGF	ENSG00000170893	-3.85	3.4e-06	TRH
ENSG00000007038	-4.41	2.0e-09	PRSS21	ENSG00000196730	2.23	3.4e-06	DAPK1
ENSG00000140968	-2.97	2.2e-09	IRF8	ENSG00000166501	1.69	3.5e-06	PRKCB
ENSG00000149346	2.74	2.3e-09	SLX4IP	ENSG00000185624	-2.46	4.0e-06	P4HB
ENSG00000159685	4.79	2.5e-09	CHCHD6	ENSG00000165168	-4.36	4.1e-06	CYBB
ENSG00000109674	-4.55	2.5e-09	NEIL3	ENSG00000101445	4.03	4.5e-06	PPP1R16B
ENSG00000072849	2.34	3.2e-09	DERL2	ENSG00000156030	1.63	4.5e-06	ELMSAN1
ENSG00000196209	-4.41	3.6e-09	SIRPB2	ENSG00000162511	1.57	4.5e-06	LAPTM5
ENSG00000102786	1.98	4.3e-09	INTS6	ENSG00000250361	3.97	5.1e-06	GYPB

Continuation of Table A.24

Ensembl	log_2 Fold Change	padj	Gene Symbol	Ensembl	log_2 Fold Change	padj	Gene Symbol
ENSG00000204420	2.41	4.9e-09	C6orf25	ENSG00000081320	2.01	5.3e-06	STK17B
ENSG00000167600	-4.23	5.7e-06	CYP2S1	ENSG00000136960	2.19	7.5e-05	ENPP2
ENSG00000173889	1.91	7.2e-06	PHC3	ENSG00000101082	2.56	7.6e-05	SLA2
ENSG00000110077	-2.82	7.7e-06	MS4A6A	ENSG00000101199	-3.05	8.0e-05	ARFGAP1
ENSG00000055163	1.68	8.0e-06	CYFIP2	ENSG00000163932	-3.22	8.6e-05	PRKCD
ENSG00000205336	2.45	8.1e-06	ADGRG1	ENSG00000005889	1.43	9.1e-05	ZFX
ENSG00000019991	-2.89	8.1e-06	HGF	ENSG00000150337	-3.87	9.1e-05	FCGR1A
ENSG00000163818	2.03	8.2e-06	LZTFL1	ENSG00000223609	3.70	9.3e-05	HBD
ENSG00000075223	3.72	8.6e-06	SEMA3C	ENSG00000101188	-3.85	9.4e-05	NTSR1
ENSG00000165568	-4.12	9.0e-06	AKR1E2	ENSG00000001561	2.24	9.5e-05	ENPP4
ENSG00000167419	-3.13	9.2e-06	LPO	ENSG00000130520	-2.31	9.5e-05	LSM4
ENSG00000072110	-1.93	9.8e-06	ACTN1	ENSG00000172183	2.88	9.7e-05	ISG20
ENSG00000145349	3.70	9.8e-06	CAMK2D	ENSG00000137942	2.72	1.0e-04	FNBP1L
ENSG00000163219	1.68	9.8e-06	ARHGAP25	ENSG00000114805	3.69	1.0e-04	PLCH1
ENSG00000241978	-3.08	9.8e-06	AKAP2	ENSG00000134285	-1.70	1.0e-04	FKBP11
ENSG00000185215	2.23	1.0e-05	TNFAIP2	ENSG00000112308	1.55	1.0e-04	C6orf62
ENSG00000143870	-1.55	1.0e-05	PDIA6	ENSG00000155090	1.89	1.1e-04	KLF10
ENSG00000165092	2.26	1.1e-05	ALDH1A1	ENSG00000112473	-3.68	1.1e-04	SLC39A7
ENSG00000050767	-4.18	1.1e-05	COL23A1	ENSG00000181690	3.29	1.2e-04	PLAG1
ENSG00000145495	1.52	1.1e-05	MARCH6	ENSG00000050130	1.53	1.2e-04	JKAMP
ENSG00000169554	-1.43	1.1e-05	ZEB2	ENSG00000107447	2.23	1.5e-04	DNTT
ENSG00000137462	-3.32	1.2e-05	TLR2	ENSG00000166224	2.13	1.5e-04	SGPL1
ENSG00000156535	1.92	1.2e-05	CD109	ENSG00000113712	1.22	1.5e-04	CSNK1A1
ENSG00000059804	2.32	1.2e-05	SLC2A3	ENSG00000176160	3.44	1.5e-04	HSF5
ENSG00000150961	-2.28	1.3e-05	SEC24D	ENSG00000100100	2.62	1.5e-04	PIK3IP1
ENSG00000167220	1.69	1.3e-05	HDHD2	ENSG00000198846	2.94	1.5e-04	TOX
ENSG00000165072	-3.93	1.3e-05	MAMDC2	ENSG00000196867	2.96	1.6e-04	ZFP28
ENSG00000121690	-4.02	1.3e-05	DEPDC7	ENSG00000155008	1.48	1.6e-04	APOOL
ENSG00000115457	-3.15	1.3e-05	IGFBP2	ENSG00000237440	2.58	1.6e-04	ZNF737
ENSG00000146918	-2.37	1.4e-05	NCAPG2	ENSG00000164484	3.16	1.7e-04	TMEM200A
ENSG00000011304	-2.10	1.5e-05	PTBP1	ENSG00000259529	-3.16	1.7e-04	NA
ENSG00000145687	1.44	1.5e-05	SSBP2	ENSG00000104388	-1.71	1.7e-04	RAB2A
ENSG00000219545	2.74	1.5e-05	UMAD1	ENSG00000104231	1.38	1.8e-04	ZFAND1
ENSG00000198586	1.40	1.5e-05	TLK1	ENSG00000106524	1.95	1.8e-04	ANKMY2
ENSG00000116741	1.71	1.6e-05	RGS2	ENSG00000122545	1.21	1.8e-04	SEPT7
ENSG00000168502	-3.99	1.6e-05	MTCL1	ENSG00000117533	2.14	1.8e-04	VAMP4
ENSG00000108924	3.54	1.6e-05	HLF	ENSG00000152465	2.48	1.8e-04	NMT2
ENSG00000147576	2.74	1.6e-05	ADHFE1	ENSG00000165476	2.50	1.8e-04	REEP3
ENSG00000185010	-4.13	1.7e-05	F8	ENSG00000089902	-2.16	1.8e-04	RCOR1
ENSG00000266074	-2.22	1.7e-05	BAHCC1	ENSG00000184992	-1.86	1.8e-04	BRI3BP
ENSG00000054793	2.67	1.7e-05	ATP9A	ENSG00000101294	-1.89	1.9e-04	HM13
ENSG00000140022	2.66	1.9e-05	STON2	ENSG00000152760	2.11	1.9e-04	TCTEX1D1
ENSG00000118263	2.35	2.5e-05	KLF7	ENSG00000145703	1.46	2.0e-04	IQGAP2
ENSG00000137959	2.72	2.7e-05	IFI44L	ENSG00000135074	3.65	2.0e-04	ADAM19
ENSG00000135476	-3.06	2.8e-05	ESPL1	ENSG00000110075	1.28	2.2e-04	PPP6R3
ENSG00000110344	2.69	3.2e-05	UBE4A	ENSG00000176014	-2.90	2.3e-04	TUBB6
ENSG00000153815	-3.15	3.3e-05	CMIP	ENSG00000175538	-2.89	2.3e-04	KCNE3
ENSG00000131370	2.59	3.3e-05	SH3BP5	ENSG00000134531	2.17	2.3e-04	EMP1
ENSG00000186088	1.90	3.3e-05	GSAP	ENSG00000185745	2.91	2.3e-04	IFIT1
ENSG00000089127	2.37	3.4e-05	OAS1	ENSG00000185760	-2.94	2.4e-04	KCNQ5
ENSG00000145247	1.70	3.4e-05	OCIAD2	ENSG00000110108	-2.87	2.5e-04	TMEM109
ENSG00000106560	1.55	3.4e-05	GIMAP2	ENSG00000151967	1.51	2.5e-04	SCHIP1
ENSG00000116747	1.44	3.5e-05	TROVE2	ENSG00000181804	3.23	2.5e-04	SLC9A9
ENSG00000128322	-2.08	3.6e-05	IGLL1	ENSG00000099256	3.14	2.5e-04	PRTFDC1
ENSG00000172493	1.47	3.7e-05	AFF1	ENSG00000103187	-2.27	2.5e-04	COTL1
ENSG00000243147	-1.49	3.7e-05	MRPL33	ENSG00000184384	2.82	2.5e-04	MAML2
ENSG00000185198	-2.59	4.2e-05	PRSS57	ENSG00000170801	-3.66	2.6e-04	HTRA3
ENSG00000169896	2.96	4.4e-05	ITGAM	ENSG00000162695	1.69	2.8e-04	SLC30A7
ENSG00000160593	3.29	4.5e-05	JAML	ENSG00000090863	1.55	2.9e-04	GLG1
ENSG00000081154	1.40	4.5e-05	PCNP	ENSG00000136997	-1.86	2.9e-04	MYC
ENSG00000166886	-3.81	4.6e-05	NAB2	ENSG00000213639	1.15	2.9e-04	PPP1CB
ENSG00000163453	-2.41	4.8e-05	IGFBP7	ENSG00000130741	1.25	2.9e-04	EIF2S3
ENSG00000101347	-3.16	5.0e-05	SAMHD1	ENSG00000170542	1.78	2.9e-04	SERPINB9
ENSG00000139675	2.71	5.0e-05	HNRNPA1L2	ENSG00000107738	2.00	3.0e-04	C10orf54
ENSG00000064042	3.63	5.2e-05	LIMCH1	ENSG00000166598	-1.24	3.0e-04	HSP90B1
ENSG00000110917	-1.58	5.3e-05	MLEC	ENSG00000127081	3.53	3.0e-04	ZNF484
ENSG00000134352	2.03	5.5e-05	IL6ST	ENSG00000146285	2.10	3.0e-04	SCML4
ENSG00000163535	-2.33	5.5e-05	SGO2	ENSG00000186105	1.71	3.1e-04	LRRC70
ENSG00000123728	1.97	5.7e-05	RAP2C	ENSG00000113389	1.80	3.1e-04	NPR3
ENSG00000134109	1.58	5.9e-05	EDEM1	ENSG00000276600	-3.45	3.1e-04	RAB7B
ENSG00000167552	1.51	6.0e-05	TUBA1A	ENSG00000152102	1.80	3.3e-04	FAM168B
ENSG00000158796	1.76	6.0e-05	DEDD	ENSG00000214357	-2.71	3.4e-04	NEURL1B
ENSG00000152642	-3.47	6.1e-05	GPD1L	ENSG00000088448	1.29	3.4e-04	ANKRD10
ENSG00000163393	-3.05	6.1e-05	SLC22A15	ENSG00000173611	1.43	3.4e-04	SCAI
ENSG00000168300	1.34	6.1e-05	PCMTD1	ENSG00000185650	2.65	3.5e-04	ZFP36L1
ENSG00000197933	3.43	6.2e-05	ZNF823	ENSG00000272602	2.08	3.5e-04	ZNF595
ENSG00000154310	3.41	6.4e-05	TNIK	ENSG00000149573	3.00	3.5e-04	MPZL2
ENSG00000254087	1.55	6.7e-05	LYN	ENSG00000091651	-2.03	3.6e-04	ORC6
ENSG00000004468	-2.63	6.8e-05	CD38	ENSG00000118508	-2.83	3.7e-04	RAB32
ENSG00000169752	-3.43	6.8e-05	NRG4	ENSG00000147689	-1.71	3.7e-04	FAM83A
ENSG00000171862	1.30	6.8e-05	PTEN	ENSG00000116649	-2.51	3.7e-04	SRM
ENSG00000152256	1.76	7.1e-05	PDK1	ENSG00000168209	2.27	3.7e-04	DDIT4
ENSG00000158470	-2.16	7.1e-05	B4GALT5	ENSG00000164684	-3.05	3.9e-04	ZNF704
ENSG00000179630	-3.79	7.1e-05	LACC1	ENSG00000104419	-2.33	3.9e-04	NDRG1
ENSG00000067445	1.97	7.4e-05	TRO	ENSG00000265808	1.22	4.1e-04	SEC22B
ENSG00000172578	1.76	7.4e-05	KLHL6	ENSG00000185499	-3.31	4.1e-04	MUC1
ENSG00000090060	1.43	4.2e-04	PAPOLA	ENSG00000052795	1.89	9.8e-04	FNIP2
ENSG00000165757	3.13	4.2e-04	KIAA1462	ENSG00000091972	1.64	9.8e-04	CD200
ENSG00000175556	-3.59	4.2e-04	LONRF3	ENSG00000165684	-3.24	9.8e-04	SNAPC4
ENSG00000175471	1.52	4.2e-04	MCTP1	ENSG00000173542	1.61	9.8e-04	MOB1B
ENSG00000117519	2.44	4.3e-04	CNN3	ENSG00000180573	1.67	9.9e-04	HIST1H2AC

Continuation of Table A.24

Ensembl	log_2 Fold Change	padj	Gene Symbol	Ensembl	log_2 Fold Change	padj	Gene Symbol
ENSG00000197142	1.60	4.3e-04	ACSL5	ENSG00000185477	2.32	1.0e-03	GPRIN3
ENSG00000178202	-2.01	4.4e-04	KDELC2	ENSG00000160336	2.12	1.0e-03	ZNF761
ENSG00000255833	-3.57	4.4e-04	TIFAB	ENSG00000043093	1.22	1.0e-03	DCUN1D1
ENSG00000197321	1.58	4.4e-04	SVIL	ENSG00000141380	1.40	1.0e-03	SS18
ENSG00000085563	2.67	4.5e-04	ABCB1	ENSG00000104765	1.27	1.1e-03	BNIP3L
ENSG00000181381	1.91	4.6e-04	DDX60L	ENSG00000124613	2.73	1.1e-03	ZNF391
ENSG00000080822	1.70	4.6e-04	CLDND1	ENSG00000164329	1.29	1.1e-03	PAPD4
ENSG00000120738	-1.49	4.6e-04	EGR1	ENSG00000115325	-2.75	1.1e-03	DOK1
ENSG00000087245	2.80	4.6e-04	MMP2	ENSG00000108654	1.04	1.1e-03	DDX5
ENSG00000184588	1.77	4.6e-04	PDE4B	ENSG00000110031	2.09	1.1e-03	LPXN
ENSG00000129534	-1.25	4.7e-04	MIS18BP1	ENSG00000104894	1.34	1.2e-03	CD37
ENSG00000115520	1.56	4.8e-04	COQ10B	ENSG00000119471	2.16	1.2e-03	HSDL2
ENSG00000244687	1.20	4.8e-04	UBE2V1	ENSG00000169242	3.20	1.2e-03	EFNA1
ENSG00000128346	3.50	5.0e-04	C22orf23	ENSG00000162607	-1.40	1.2e-03	USP1
ENSG00000140263	-2.07	5.1e-04	SORD	ENSG00000143119	1.39	1.2e-03	CD53
ENSG00000116675	-3.40	5.3e-04	DNAJC6	ENSG00000165702	2.01	1.2e-03	GFI1B
ENSG00000144802	-1.29	5.4e-04	NFKBIZ	ENSG00000136877	-2.02	1.2e-03	FPGS
ENSG00000275004	2.39	5.4e-04	ZNF280B	ENSG00000014914	-3.36	1.2e-03	MTMR11
ENSG00000111897	1.46	5.5e-04	SERINC1	ENSG00000166974	1.27	1.2e-03	MAPRE2
ENSG00000164331	2.36	5.5e-04	ANKRA2	ENSG00000111678	-1.37	1.2e-03	C12orf57
ENSG00000244754	1.24	5.5e-04	N4BP2L2	ENSG00000178761	1.48	1.2e-03	FAM219B
ENSG00000245848	-2.32	5.6e-04	CEBPA	ENSG00000120832	2.58	1.3e-03	MTERF2
ENSG00000124788	1.99	5.7e-04	ATXN1	ENSG00000110768	1.64	1.3e-03	GTF2H1
ENSG00000197279	2.00	5.7e-04	ZNF165	ENSG00000116752	1.21	1.3e-03	BCAS2
ENSG00000170325	-2.84	5.7e-04	PRDM10	ENSG00000129116	2.24	1.3e-03	PALLD
ENSG00000099204	1.41	5.9e-04	ABLIM1	ENSG00000102007	1.23	1.3e-03	PLP2
ENSG00000119401	3.48	6.0e-04	TRIM32	ENSG00000153044	-2.05	1.3e-03	CENPH
ENSG00000024048	1.30	6.0e-04	UBR2	ENSG00000120156	3.29	1.4e-03	TEK
ENSG00000066654	1.36	6.0e-04	THUMPD1	ENSG00000173193	1.20	1.4e-03	PARP14
ENSG00000166669	1.28	6.0e-04	ATF7IP2	ENSG00000132589	1.65	1.4e-03	FLOT2
ENSG00000185885	2.12	6.0e-04	IFITM1	ENSG00000162366	3.23	1.4e-03	PDZK1IP1
ENSG00000106952	3.05	6.0e-04	TNFSF8	ENSG00000165995	1.55	1.4e-03	CACNB2
ENSG00000149970	-3.47	6.0e-04	CNKSR2	ENSG00000106080	-2.18	1.5e-03	FKBP14
ENSG00000196116	-3.06	6.0e-04	TDRD7	ENSG00000213672	-2.89	1.5e-03	NCKIPSD
ENSG00000196071	3.50	6.1e-04	OR2L13	ENSG00000037749	2.83	1.5e-03	MFAP3
ENSG00000170776	1.35	6.1e-04	AKAP13	ENSG00000159658	1.70	1.5e-03	EFCAB14
ENSG00000148835	-3.25	6.2e-04	TAF5	ENSG00000177359	-2.24	1.5e-03	OVOS2
ENSG00000065534	1.73	6.3e-04	MYLK	ENSG00000062194	1.24	1.5e-03	GPBP1
ENSG00000170482	2.84	6.3e-04	SLC23A1	ENSG00000204381	-3.02	1.5e-03	LAYN
ENSG00000140905	-1.99	6.4e-04	GCSH	ENSG00000158457	-2.51	1.5e-03	TSPAN33
ENSG00000104043	-1.24	6.4e-04	ATP8B4	ENSG00000168421	1.65	1.6e-03	RHOH
ENSG00000170365	2.32	6.4e-04	SMAD1	ENSG00000136111	2.12	1.6e-03	TBC1D4
ENSG00000106066	-1.57	6.4e-04	CPVL	ENSG00000186591	1.58	1.6e-03	UBE2H
ENSG00000187164	2.37	6.4e-04	SHTN1	ENSG00000065978	-1.39	1.7e-03	YBX1
ENSG00000118564	1.39	6.7e-04	FBXL5	ENSG00000110934	1.88	1.7e-03	BIN2
ENSG00000139055	2.10	6.7e-04	ERP27	ENSG00000006555	2.50	1.7e-03	TTC22
ENSG00000168675	1.32	6.8e-04	LDLRAD4	ENSG00000179029	-1.28	1.7e-03	TMEM107
ENSG00000179051	-1.45	6.8e-04	RCC2	ENSG00000183918	2.86	1.7e-03	SH2D1A
ENSG00000067369	1.29	6.8e-04	TP53BP1	ENSG00000151725	-1.72	1.7e-03	CENPU
ENSG00000100055	2.29	6.8e-04	CYTH4	ENSG00000106804	2.07	1.7e-03	C5
ENSG00000135378	-2.20	7.1e-04	PRRG4	ENSG00000092853	-1.80	1.8e-03	CLSPN
ENSG00000111834	3.00	7.2e-04	RSPH4A	ENSG00000198189	1.16	1.8e-03	HSD17B11
ENSG00000189077	3.33	7.2e-04	TMEM120A	ENSG00000141198	2.53	1.9e-03	TOM1L1
ENSG00000130816	-1.42	7.5e-04	DNMT1	ENSG00000184678	2.13	1.9e-03	HIST2H2BE
ENSG00000100097	-2.61	7.5e-04	LGALS1	ENSG00000179958	-1.62	2.0e-03	DCTPP1
ENSG00000132849	1.61	7.8e-04	PATJ	ENSG00000158710	1.03	2.0e-03	TAGLN2
ENSG00000167004	-1.58	7.9e-04	PDIA3	ENSG00000146410	-2.30	2.0e-03	MTFR2
ENSG00000185189	3.36	8.0e-04	NRBP2	ENSG00000170271	-2.31	2.0e-03	FAXDC2
ENSG00000198795	1.82	8.0e-04	ZNF521	ENSG00000156171	1.79	2.0e-03	DRAM2
ENSG00000081177	2.47	8.0e-04	EXD2	ENSG00000184368	-3.17	2.0e-03	MAP7D2
ENSG00000163577	2.52	8.0e-04	EIF5A2	ENSG00000102524	-1.79	2.0e-03	TNFSF13B
ENSG00000119979	-1.71	8.1e-04	FAM45A	ENSG00000145287	-1.40	2.0e-03	PLAC8
ENSG00000071539	-2.71	8.2e-04	TRIP13	ENSG00000091592	1.89	2.1e-03	NLRP1
ENSG00000111335	2.09	8.2e-04	OAS2	ENSG00000005844	1.29	2.1e-03	ITGAL
ENSG00000081237	1.30	8.3e-04	PTPRC	ENSG00000105639	1.27	2.1e-03	JAK3
ENSG00000016864	-2.09	8.3e-04	GLT8D1	ENSG00000146476	2.35	2.1e-03	ARMT1
ENSG00000041353	2.91	8.4e-04	RAB27B	ENSG00000175826	-1.74	2.1e-03	CTDNEP1
ENSG00000108518	-1.52	8.4e-04	PFN1	ENSG00000113851	1.29	2.1e-03	CRBN
ENSG00000212127	3.18	8.4e-04	TAS2R14	ENSG00000142961	2.86	2.1e-03	MOB3C
ENSG00000205269	-2.56	8.5e-04	TMEM170B	ENSG00000182718	-2.40	2.1e-03	ANXA2
ENSG00000063587	1.91	8.5e-04	ZNF275	ENSG00000144369	2.02	2.1e-03	FAM171B
ENSG00000138670	1.49	8.6e-04	RASGEF1B	ENSG00000162676	-2.37	2.1e-03	GFI1
ENSG00000198851	3.37	8.8e-04	CD3E	ENSG00000124299	2.36	2.1e-03	PEPD
ENSG00000108561	-1.62	9.0e-04	C1QBP	ENSG00000111348	1.07	2.1e-03	ARHGDIB
ENSG00000100554	1.47	9.2e-04	ATP6V1D	ENSG00000175857	1.31	2.2e-03	GAPT
ENSG00000166130	-2.10	9.3e-04	IKBIP	ENSG00000169251	1.49	2.2e-03	NMD3
ENSG00000196233	1.14	9.3e-04	C10orf12	ENSG00000119862	-3.06	2.2e-03	LGALSL
ENSG00000134594	3.32	9.3e-04	RAB33A	ENSG00000182831	1.13	2.3e-03	C16orf72
ENSG00000134262	-1.63	9.6e-04	AP4B1	ENSG00000125520	-2.69	2.3e-03	SLC2A4RG
ENSG00000081189	1.50	9.8e-04	MEF2C	ENSG00000147162	1.35	2.3e-03	OGT
ENSG00000198298	-2.66	2.3e-03	ZNF485	ENSG00000169314	2.83	4.8e-03	C22orf15
ENSG00000089123	1.88	2.4e-03	TASP1	ENSG00000170222	1.74	4.8e-03	ADPRM
ENSG00000135048	1.78	2.4e-03	TMEM2	ENSG00000092964	1.37	4.8e-03	DPYSL2
ENSG00000163933	-2.30	2.4e-03	RFT1	ENSG00000129204	2.18	4.8e-03	USP6
ENSG00000175591	-3.21	2.4e-03	P2RY2	ENSG00000198786	0.98	4.8e-03	ND5
ENSG00000198160	1.13	2.4e-03	MIER1	ENSG00000242574	1.26	4.8e-03	HLA-DMB
ENSG00000138468	1.40	2.4e-03	SENP7	ENSG00000075420	-1.90	4.8e-03	FNDC3B
ENSG00000173276	1.62	2.4e-03	ZBTB21	ENSG00000142675	2.89	4.8e-03	CNKSR1
ENSG00000225697	-2.67	2.4e-03	SLC26A6	ENSG00000165699	1.50	4.8e-03	TSC1
ENSG00000112773	-1.40	2.5e-03	FAM46A	ENSG00000215021	-1.26	4.8e-03	PHB2
ENSG00000213965	1.24	2.5e-03	NUDT19	ENSG00000132952	1.42	4.8e-03	USPL1

Continuation of Table A.24

Ensembl	log_2 Fold Change	padj	Gene Symbol	Ensembl	log_2 Fold Change	padj	Gene Symbol
ENSG00000163644	1.41	2.7e-03	PPM1K	ENSG00000123096	-2.73	4.9e-03	SSPN
ENSG00000155903	1.67	2.7e-03	RASA2	ENSG00000139725	1.70	4.9e-03	RHOF
ENSG00000131269	1.49	2.7e-03	ABCB7	ENSG00000147905	1.22	4.9e-03	ZCCHC7
ENSG00000183763	-3.04	2.8e-03	TRAIP	ENSG00000140199	1.36	4.9e-03	SLC12A6
ENSG00000108582	-1.63	2.8e-03	CPD	ENSG00000119523	1.73	5.0e-03	ALG2
ENSG00000104660	1.36	2.8e-03	LEPROTL1	ENSG00000166780	2.30	5.0e-03	C16orf45
ENSG00000145016	1.80	2.9e-03	RUBCN	ENSG00000132824	1.25	5.1e-03	SERINC3
ENSG00000172244	-2.91	2.9e-03	C5orf34	ENSG00000156970	-1.94	5.2e-03	BUB1B
ENSG00000260916	1.52	2.9e-03	CCPG1	ENSG00000102910	1.09	5.3e-03	LONP2
ENSG00000117569	1.11	3.0e-03	PTBP2	ENSG00000131171	0.97	5.3e-03	SH3BGRL
ENSG00000113369	1.38	3.0e-03	ARRDC3	ENSG00000119328	1.71	5.4e-03	FAM206A
ENSG00000196428	1.20	3.0e-03	TSC22D2	ENSG00000166794	-1.24	5.5e-03	PPIB
ENSG00000070882	2.41	3.0e-03	OSBPL3	ENSG00000257365	-1.86	5.5e-03	FNTB
ENSG00000168556	-1.91	3.1e-03	ING2	ENSG00000162819	1.59	5.6e-03	BROX
ENSG00000119969	-1.21	3.1e-03	HELLS	ENSG00000000457	1.40	5.6e-03	SCYL3
ENSG00000131018	1.30	3.1e-03	SYNE1	ENSG00000100528	1.11	5.6e-03	CNIH1
ENSG00000179115	-1.53	3.1e-03	FARSA	ENSG00000132640	1.46	5.6e-03	BTBD3
ENSG00000187240	2.22	3.1e-03	DYNC2H1	ENSG00000153283	-2.32	5.6e-03	CD96
ENSG00000162852	1.32	3.1e-03	CNST	ENSG00000183283	1.04	5.6e-03	DAZAP2
ENSG00000171097	-2.47	3.1e-03	KYAT1	ENSG00000147180	1.46	5.7e-03	ZNF711
ENSG00000188580	2.22	3.1e-03	NKAIN2	ENSG00000051341	-1.69	5.7e-03	POLQ
ENSG00000048162	-1.26	3.2e-03	NOP16	ENSG00000198898	1.07	5.9e-03	CAPZA2
ENSG00000176714	2.64	3.2e-03	CCDC121	ENSG00000164405	-1.06	5.9e-03	UQCRQ
ENSG00000196411	-2.27	3.2e-03	EPHB4	ENSG00000116852	-2.09	5.9e-03	KIF21B
ENSG00000197653	2.76	3.2e-03	DNAH10	ENSG00000173681	1.64	5.9e-03	CXorf23
ENSG00000186665	-3.08	3.2e-03	C17orf58	ENSG00000128739	1.48	6.0e-03	SNRPN
ENSG00000127993	1.59	3.2e-03	RBM48	ENSG00000164116	1.20	6.0e-03	GUCY1A3
ENSG00000128908	-1.79	3.2e-03	INO80	ENSG00000136888	1.05	6.0e-03	ATP6V1G1
ENSG00000048540	-3.08	3.3e-03	LMO3	ENSG00000157110	2.04	6.0e-03	RBPMS
ENSG00000198286	2.98	3.3e-03	CARD11	ENSG00000100982	1.36	6.0e-03	PCIF1
ENSG00000196405	1.55	3.3e-03	EVL	ENSG00000181467	1.09	6.0e-03	RAP2B
ENSG00000170779	-2.15	3.3e-03	CDCA4	ENSG00000132475	0.99	6.1e-03	H3F3B
ENSG00000114541	1.34	3.3e-03	FRMD4B	ENSG00000138180	-2.65	6.1e-03	CEP55
ENSG00000145428	-3.02	3.3e-03	RNF175	ENSG00000099308	2.01	6.1e-03	MAST3
ENSG00000143443	1.19	3.4e-03	C1orf56	ENSG00000122824	-2.95	6.1e-03	NUDT10
ENSG00000159461	-1.14	3.4e-03	AMFR	ENSG00000242802	-2.50	6.2e-03	AP5Z1
ENSG00000125089	2.52	3.4e-03	SH3TC1	ENSG00000179943	-2.86	6.2e-03	FIZ1
ENSG00000137726	3.08	3.5e-03	FXYD6	ENSG00000110888	1.57	6.3e-03	CAPRIN2
ENSG00000133116	-3.10	3.5e-03	KL	ENSG00000138031	-1.81	6.4e-03	ADCY3
ENSG00000104756	1.36	3.6e-03	KCTD9	ENSG00000163737	2.95	6.4e-03	PF4
ENSG00000166482	2.36	3.6e-03	MFAP4	ENSG00000153574	-2.04	6.4e-03	RPIA
ENSG00000112742	-2.78	3.7e-03	TTK	ENSG00000187325	1.41	6.4e-03	TAF9B
ENSG00000106211	-1.37	3.7e-03	HSPB1	ENSG00000168899	-1.63	6.5e-03	VAMP5
ENSG00000010327	-2.88	3.8e-03	STAB1	ENSG00000197381	-2.65	6.5e-03	ADARB1
ENSG00000100600	2.52	3.8e-03	LGMN	ENSG00000265972	1.09	6.5e-03	TXNIP
ENSG00000163602	1.64	3.8e-03	RYBP	ENSG00000164946	2.70	6.5e-03	FREM1
ENSG00000168077	-3.08	3.8e-03	SCARA3	ENSG00000205758	1.27	6.5e-03	CRYZL1
ENSG00000178950	-2.00	3.8e-03	GAK	ENSG00000171848	-1.83	6.6e-03	RRM2
ENSG00000198624	1.80	3.8e-03	CCDC69	ENSG00000204315	-2.95	6.6e-03	FKBPL
ENSG00000010404	1.08	4.0e-03	IDS	ENSG00000182578	-2.50	6.6e-03	CSF1R
ENSG00000013441	1.14	4.0e-03	CLK1	ENSG00000090376	-1.18	6.6e-03	IRAK3
ENSG00000170035	1.32	4.0e-03	UBE2E3	ENSG00000166432	-1.39	6.7e-03	ZMAT1
ENSG00000188042	1.66	4.1e-03	ARL4C	ENSG00000171320	-1.89	6.7e-03	ESCO2
ENSG00000197062	1.99	4.1e-03	ZSCAN26	ENSG00000109184	1.21	6.7e-03	DCUN1D4
ENSG00000188229	-1.16	4.1e-03	TUBB4B	ENSG00000100395	-1.61	6.7e-03	L3MBTL2
ENSG00000250565	-2.77	4.2e-03	ATP6V1E2	ENSG00000163069	2.66	6.7e-03	SGCB
ENSG00000139324	2.06	4.4e-03	TMTC3	ENSG00000196776	1.23	6.7e-03	CD47
ENSG00000112851	1.13	4.4e-03	ERBIN	ENSG00000163104	1.19	6.8e-03	SMARCAD1
ENSG00000170891	1.60	4.4e-03	CYTL1	ENSG00000167778	2.15	6.8e-03	SPRYD3
ENSG00000179304	-2.23	4.5e-03	FAM156B	ENSG00000110002	2.23	6.9e-03	VWA5A
ENSG00000115165	2.13	4.5e-03	CYTIP	ENSG00000185437	2.81	6.9e-03	SH3BGR
ENSG00000223658	3.06	4.5e-03	C1GALT1C1L	ENSG00000170027	-1.28	7.0e-03	YWHAG
ENSG00000178038	2.68	4.5e-03	ALS2CL	ENSG00000136999	2.39	7.1e-03	NOV
ENSG00000173801	1.37	4.5e-03	JUP	ENSG00000103351	1.46	7.1e-03	CLUAP1
ENSG00000179912	1.19	4.5e-03	R3HDM2	ENSG00000052749	-1.54	7.1e-03	RRP12
ENSG00000155636	2.43	4.5e-03	RBM45	ENSG00000132792	1.49	7.2e-03	CTNNBL1
ENSG00000115758	-1.53	4.6e-03	ODC1	ENSG00000176595	-2.38	7.2e-03	KBTBD11
ENSG00000125812	-2.20	4.6e-03	GZF1	ENSG00000034713	1.18	7.3e-03	GABARAPL2
ENSG00000112715	-3.03	4.7e-03	VEGFA	ENSG00000066084	1.17	7.3e-03	DIP2B
ENSG00000161791	2.37	4.7e-03	FMNL3	ENSG00000100612	1.07	7.3e-03	DHRS7
ENSG00000242259	-1.58	4.7e-03	C22orf39	ENSG00000231389	1.01	7.3e-03	HLA-DPA1
ENSG00000171843	1.30	4.8e-03	MLLT3	ENSG00000133026	1.88	7.4e-03	MYH10
ENSG00000119403	-1.68	4.8e-03	PHF19	ENSG00000176986	-2.01	7.4e-03	SEC24C
ENSG00000119711	1.29	4.8e-03	ALDH6A1	ENSG00000128951	-1.23	7.4e-03	DUT
ENSG00000183011	-1.32	7.4e-03	NAA38	ENSG00000197555	-2.07	1.2e-02	SIPA1L1
ENSG00000185946	1.56	7.4e-03	RNPC3	ENSG00000167528	1.73	1.2e-02	ZNF641
ENSG00000153707	2.54	7.5e-03	PTPRD	ENSG00000133561	2.24	1.2e-02	GIMAP6
ENSG00000105339	-1.71	7.5e-03	DENND3	ENSG00000079257	1.72	1.2e-02	LXN
ENSG00000129422	2.46	7.6e-03	MTUS1	ENSG00000115977	-1.31	1.2e-02	AAK1
ENSG00000198932	2.72	7.7e-03	GPRASP1	ENSG00000089775	1.22	1.2e-02	ZBTB25
ENSG00000154545	2.79	7.8e-03	MAGED4	ENSG00000116539	1.08	1.2e-02	ASH1L
ENSG00000188486	-2.37	7.8e-03	H2AFX	ENSG00000137875	-2.79	1.2e-02	BCL2L10
ENSG00000166128	1.65	7.8e-03	RAB8B	ENSG00000138814	1.05	1.2e-02	PPP3CA
ENSG00000128335	1.33	7.8e-03	APOL2	ENSG00000165527	1.23	1.2e-02	ARF6
ENSG00000166016	-2.90	7.8e-03	ABTB2	ENSG00000084444	1.71	1.2e-02	FAM234B
ENSG00000139154	1.11	7.9e-03	AEBP2	ENSG00000206535	-2.79	1.2e-02	LNP1
ENSG00000198399	1.08	7.9e-03	ITSN2	ENSG00000185015	1.86	1.2e-02	CA13
ENSG00000211898	2.65	8.0e-03	NA	ENSG00000166352	1.43	1.2e-02	C11orf74
ENSG00000127824	2.11	8.0e-03	TUBA4A	ENSG00000023287	1.30	1.2e-02	RB1CC1
ENSG00000121281	1.94	8.1e-03	ADCY7	ENSG00000060982	1.01	1.2e-02	BCAT1
ENSG00000163510	1.25	8.2e-03	CWC22	ENSG00000167112	-1.36	1.2e-02	TRUB2

Continuation of Table A.24

Ensembl	log_2 Fold Change	padj	Gene Symbol	Ensembl	log_2 Fold Change	padj	Gene Symbol
ENSG00000132950	1.76	8.4e-03	ZMYM5	ENSG00000170396	-2.68	1.2e-02	ZNF804A
ENSG00000133742	2.75	8.4e-03	CA1	ENSG00000185658	0.99	1.2e-02	BRWD1
ENSG00000158089	-2.14	8.4e-03	GALNT14	ENSG00000122042	1.94	1.2e-02	UBL3
ENSG00000187607	1.32	8.4e-03	ZNF286A	ENSG00000115145	1.54	1.2e-02	STAM2
ENSG00000120280	-1.23	8.4e-03	CXorf21	ENSG00000164715	1.70	1.2e-02	LMTK2
ENSG00000108828	-1.25	8.4e-03	VAT1	ENSG00000237765	1.25	1.2e-02	FAM200B
ENSG00000124486	1.08	8.4e-03	USP9X	ENSG00000114654	2.32	1.3e-02	EFCC1
ENSG00000133731	1.42	8.7e-03	IMPA1	ENSG00000138496	1.47	1.3e-02	PARP9
ENSG00000068831	-1.70	8.7e-03	RASGRP2	ENSG00000102547	2.76	1.3e-02	CAB39L
ENSG00000131051	0.97	8.7e-03	RBM39	ENSG00000178741	-1.50	1.3e-02	COX5A
ENSG00000167645	-1.96	8.7e-03	YIF1B	ENSG00000110911	1.38	1.3e-02	SLC11A2
ENSG00000168066	-1.17	8.7e-03	SF1	ENSG00000104738	-1.60	1.3e-02	MCM4
ENSG00000100321	-1.14	8.9e-03	SYNGR1	ENSG00000111875	1.46	1.3e-02	ASF1A
ENSG00000273802	1.21	8.9e-03	HIST1H2BG	ENSG00000135845	1.31	1.3e-02	PIGC
ENSG00000103671	1.52	8.9e-03	TRIP4	ENSG00000157514	1.11	1.3e-02	TSC22D3
ENSG00000149948	2.06	9.0e-03	HMGA2	ENSG00000127418	-2.71	1.3e-02	FGFRL1
ENSG00000165029	1.69	9.0e-03	ABCA1	ENSG00000163808	-2.41	1.3e-02	KIF15
ENSG00000128000	1.94	9.2e-03	ZNF780B	ENSG00000113971	1.43	1.3e-02	NPHP3
ENSG00000156671	1.27	9.2e-03	SAMD8	ENSG00000101017	2.45	1.3e-02	CD40
ENSG00000156976	0.87	9.2e-03	EIF4A2	ENSG00000117425	-1.25	1.3e-02	PTCH2
ENSG00000168234	1.28	9.2e-03	TTC39C	ENSG00000143882	-2.76	1.3e-02	ATP6V1C2
ENSG00000221923	1.66	9.2e-03	ZNF880	ENSG00000127920	2.50	1.3e-02	GNG11
ENSG00000164430	-1.79	9.2e-03	MB21D1	ENSG00000101057	-2.44	1.3e-02	MYBL2
ENSG00000173334	-2.35	9.2e-03	TRIB1	ENSG00000085224	1.10	1.3e-02	ATRX
ENSG00000168310	1.86	9.3e-03	IRF2	ENSG00000174456	1.30	1.3e-02	C12orf76
ENSG00000075651	-1.93	9.3e-03	PLD1	ENSG00000141748	2.54	1.3e-02	ARL5C
ENSG00000137563	-1.27	9.3e-03	GGH	ENSG00000228716	-1.39	1.3e-02	DHFR
ENSG00000196950	1.07	9.3e-03	SLC39A10	ENSG00000250312	1.94	1.3e-02	ZNF718
ENSG00000213096	1.60	9.3e-03	ZNF254	ENSG00000174944	2.34	1.3e-02	P2RY14
ENSG00000125354	0.96	9.5e-03	Sep-06	ENSG00000185905	0.98	1.3e-02	C16orf54
ENSG00000079215	-2.63	9.5e-03	SLC1A3	ENSG00000105856	1.44	1.4e-02	HBP1
ENSG00000153944	0.94	9.5e-03	MSI2	ENSG00000065357	1.40	1.4e-02	DGKA
ENSG00000164687	-0.94	9.5e-03	FABP5	ENSG00000116525	2.68	1.4e-02	TRIM62
ENSG00000123684	1.62	9.5e-03	LPGAT1	ENSG00000159374	-2.58	1.4e-02	M1AP
ENSG00000105726	-2.47	9.6e-03	ATP13A1	ENSG00000197451	-1.23	1.4e-02	HNRNPAB
ENSG00000108641	-2.84	9.7e-03	B9D1	ENSG00000095209	1.36	1.4e-02	TMEM38B
ENSG00000110871	1.68	9.8e-03	COQ5	ENSG00000104327	2.32	1.4e-02	CALB1
ENSG00000110330	1.28	9.9e-03	BIRC2	ENSG00000138613	1.65	1.4e-02	APH1B
ENSG00000135002	-1.61	9.9e-03	RFK	ENSG00000142864	-0.90	1.4e-02	SERBP1
ENSG00000076356	2.83	1.0e-02	PLXNA2	ENSG00000151612	1.79	1.4e-02	ZNF827
ENSG00000103353	-1.81	1.0e-02	UBFD1	ENSG00000163947	1.92	1.4e-02	ARHGEF3
ENSG00000110107	-2.02	1.0e-02	PRPF19	ENSG00000198961	1.26	1.4e-02	PJA2
ENSG00000158966	2.03	1.0e-02	CACHD1	ENSG00000215301	0.89	1.4e-02	DDX3X
ENSG00000087206	1.23	1.0e-02	UIMC1	ENSG00000169429	-1.07	1.4e-02	CXCL8
ENSG00000169575	1.46	1.0e-02	VPREB1	ENSG00000130270	-2.09	1.4e-02	ATP8B3
ENSG00000108349	1.49	1.1e-02	CASC3	ENSG00000268043	1.56	1.4e-02	NBPF12
ENSG00000132205	-1.63	1.1e-02	EMILIN2	ENSG00000118513	-1.35	1.4e-02	MYB
ENSG00000167900	-1.92	1.1e-02	TK1	ENSG00000101333	2.12	1.4e-02	PLCB4
ENSG00000001460	-1.62	1.1e-02	STPG1	ENSG00000172349	1.49	1.4e-02	IL16
ENSG00000114742	1.38	1.1e-02	WDR48	ENSG00000140836	1.43	1.4e-02	ZFHX3
ENSG00000186918	1.43	1.1e-02	ZNF395	ENSG00000148110	-1.51	1.4e-02	MFSD14B
ENSG00000133619	-2.81	1.1e-02	KRBA1	ENSG00000165025	1.10	1.4e-02	SYK
ENSG00000108179	-2.01	1.1e-02	PPIF	ENSG00000105819	0.94	1.4e-02	PMPCB
ENSG00000167208	1.30	1.1e-02	SNX20	ENSG00000227057	-1.70	1.4e-02	WDR46
ENSG00000150054	2.41	1.1e-02	MPP7	ENSG00000114904	-1.74	1.4e-02	NEK4
ENSG00000198453	1.56	1.1e-02	ZNF568	ENSG00000120963	1.02	1.4e-02	ZNF706
ENSG00000167377	2.10	1.1e-02	ZNF23	ENSG00000150093	1.24	1.4e-02	ITGB1
ENSG00000182179	1.15	1.1e-02	UBA7	ENSG00000177119	-2.02	1.4e-02	ANO6
ENSG00000146021	2.13	1.1e-02	KLHL3	ENSG00000004455	-1.16	1.4e-02	AK2
ENSG00000167447	-2.71	1.1e-02	SMG8	ENSG00000114374	1.83	1.4e-02	USP9Y
ENSG00000179021	1.47	1.1e-02	C3orf38	ENSG00000156136	1.09	1.4e-02	DCK
ENSG00000156858	-2.02	1.1e-02	PRR14	ENSG00000196449	-2.11	1.5e-02	YRDC
ENSG00000101166	1.23	1.1e-02	PRELID3B	ENSG00000142227	1.69	1.5e-02	EMP3
ENSG00000125148	-1.53	1.1e-02	MT2A	ENSG00000197808	-2.42	1.5e-02	ZNF461
ENSG00000161267	-1.95	1.1e-02	BDH1	ENSG00000061987	1.48	1.5e-02	MON2
ENSG00000079691	1.82	1.1e-02	CARMIL1	ENSG00000025800	1.04	1.5e-02	KPNA6
ENSG00000133872	1.00	1.1e-02	SARAF	ENSG00000151838	2.66	1.5e-02	CCDC175
ENSG00000160193	-2.76	1.2e-02	WDR4	ENSG00000159593	1.07	1.5e-02	NAE1
ENSG00000227507	2.65	1.5e-02	LTB	ENSG00000163527	0.89	2.2e-02	STT3B
ENSG00000196459	1.09	1.5e-02	TRAPPC2	ENSG00000135912	-2.12	2.2e-02	TTLL4
ENSG00000091527	0.90	1.5e-02	CDV3	ENSG00000170509	2.04	2.2e-02	HSD17B13
ENSG00000157107	2.11	1.5e-02	FCHO2	ENSG00000149054	-1.90	2.2e-02	ZNF215
ENSG00000114988	1.71	1.5e-02	LMAN2L	ENSG00000174109	-2.04	2.2e-02	C16orf91
ENSG00000140379	1.76	1.5e-02	BCL2A1	ENSG00000240891	2.34	2.2e-02	PLCXD2
ENSG00000174776	-2.39	1.5e-02	WDR49	ENSG00000068724	-1.43	2.3e-02	TTC7A
ENSG00000100206	-1.95	1.6e-02	DMC1	ENSG00000114861	0.85	2.3e-02	FOXP1
ENSG00000123352	1.30	1.6e-02	SPATS2	ENSG00000159388	1.39	2.3e-02	BTG2
ENSG00000019582	0.95	1.6e-02	CD74	ENSG00000164169	1.83	2.3e-02	PRMT9
ENSG00000164032	-1.16	1.6e-02	H2AFZ	ENSG00000173933	0.95	2.3e-02	RBM4
ENSG00000134802	1.32	1.6e-02	SLC43A3	ENSG00000107864	-1.57	2.3e-02	CPEB3
ENSG00000159314	1.26	1.6e-02	ARHGAP27	ENSG00000168710	-1.19	2.3e-02	AHCYL1
ENSG00000162714	1.29	1.6e-02	ZNF496	ENSG00000127507	1.04	2.3e-02	ADGRE2
ENSG00000077044	-2.10	1.6e-02	DGKD	ENSG00000150527	1.15	2.3e-02	MIA2
ENSG00000166262	2.20	1.6e-02	FAM227B	ENSG00000136051	1.04	2.3e-02	KIAA1033
ENSG00000225698	2.68	1.6e-02	NA	ENSG00000137806	1.51	2.3e-02	NDUFAF1
ENSG00000141349	-1.38	1.6e-02	G6PC3	ENSG00000198939	2.36	2.3e-02	ZFP2
ENSG00000183337	1.36	1.6e-02	BCOR	ENSG00000277161	-2.05	2.3e-02	PIGW
ENSG00000069493	1.25	1.6e-02	CLEC2D	ENSG00000188690	-1.55	2.3e-02	UROS
ENSG00000112972	1.20	1.6e-02	HMGCS1	ENSG00000115241	-1.18	2.3e-02	PPM1G
ENSG00000196460	-1.75	1.6e-02	RFX8	ENSG00000213930	1.16	2.3e-02	GALT
ENSG00000064651	1.09	1.7e-02	SLC12A2	ENSG00000137575	1.11	2.4e-02	SDCBP

Continuation of Table A.24

Ensembl	log_2 Fold Change	padj	Gene Symbol	Ensembl	log_2 Fold Change	padj	Gene Symbol
ENSG00000091157	-1.58	1.7e-02	WDR7	ENSG00000241399	-1.33	2.4e-02	CD302
ENSG00000198791	0.99	1.7e-02	CNOT7	ENSG00000145725	0.84	2.4e-02	PPIP5K2
ENSG00000105492	2.65	1.7e-02	SIGLEC6	ENSG00000071626	-1.58	2.4e-02	DAZAP1
ENSG00000168297	-1.21	1.7e-02	PXK	ENSG00000124731	-2.57	2.4e-02	TREM1
ENSG00000103591	1.50	1.7e-02	AAGAB	ENSG00000148655	-1.76	2.4e-02	C10orf11
ENSG00000165275	1.35	1.7e-02	TRMT10B	ENSG00000169184	2.24	2.4e-02	MN1
ENSG00000179299	2.54	1.7e-02	NSUN7	ENSG00000196357	2.52	2.4e-02	ZNF565
ENSG00000100365	1.32	1.7e-02	NCF4	ENSG00000150995	1.24	2.4e-02	ITPR1
ENSG00000111671	-2.22	1.7e-02	SPSB2	ENSG00000196187	1.52	2.4e-02	TMEM63A
ENSG00000156500	-1.57	1.7e-02	FAM122C	ENSG00000067533	0.97	2.4e-02	RRP15
ENSG00000131153	-1.24	1.7e-02	GINS2	ENSG00000148840	-2.24	2.4e-02	PPRC1
ENSG00000198677	1.10	1.7e-02	TTC37	ENSG00000134313	1.17	2.4e-02	KIDINS220
ENSG00000180530	1.02	1.8e-02	NRIP1	ENSG00000105643	1.56	2.5e-02	ARRDC2
ENSG00000183691	-1.55	1.8e-02	NOG	ENSG00000232653	-1.58	2.5e-02	GOLGA8N
ENSG00000119318	-1.29	1.8e-02	RAD23B	ENSG00000108679	1.54	2.5e-02	LGALS3BP
ENSG00000136040	1.18	1.8e-02	PLXNC1	ENSG00000113504	1.74	2.5e-02	SLC12A7
ENSG00000139117	1.11	1.8e-02	CPNE8	ENSG00000124459	1.80	2.5e-02	ZNF45
ENSG00000213654	1.32	1.8e-02	GPSM3	ENSG00000163110	1.43	2.5e-02	PDLIM5
ENSG00000198169	1.47	1.9e-02	ZNF251	ENSG00000135046	-0.79	2.5e-02	ANXA1
ENSG00000109534	-1.87	1.9e-02	GAR1	ENSG00000166532	1.13	2.5e-02	RIMKLB
ENSG00000134825	-0.91	1.9e-02	TMEM258	ENSG00000082996	1.13	2.5e-02	RNF13
ENSG00000166963	1.18	1.9e-02	MAP1A	ENSG00000102796	1.66	2.5e-02	DHRS12
ENSG00000173486	-1.62	1.9e-02	FKBP2	ENSG00000134815	-2.55	2.5e-02	DHX34
ENSG00000204576	1.37	1.9e-02	PRR3	ENSG00000083799	0.98	2.5e-02	CYLD
ENSG00000012963	-1.34	1.9e-02	UBR7	ENSG00000114098	1.14	2.6e-02	ARMC8
ENSG00000114796	1.10	1.9e-02	KLHL24	ENSG00000107262	-1.23	2.6e-02	BAG1
ENSG00000113240	1.18	1.9e-02	CLK4	ENSG00000117143	-1.94	2.6e-02	UAP1
ENSG00000145780	-1.67	1.9e-02	FEM1C	ENSG00000255587	-2.42	2.6e-02	RAB44
ENSG00000171262	-1.13	1.9e-02	FAM98B	ENSG00000125845	2.55	2.6e-02	BMP2
ENSG00000182903	1.46	1.9e-02	ZNF721	ENSG00000204287	0.89	2.6e-02	HLA-DRA
ENSG00000185909	2.45	1.9e-02	KLHDC8B	ENSG00000197982	-2.45	2.6e-02	C1orf122
ENSG00000165304	-2.35	2.0e-02	MELK	ENSG00000182054	-1.22	2.6e-02	IDH2
ENSG00000187922	2.63	2.0e-02	LCN10	ENSG00000253251	2.29	2.6e-02	TRAPPC13
ENSG00000136270	-1.40	2.0e-02	TBRG4	ENSG00000087589	1.81	2.6e-02	CASS4
ENSG00000167797	-1.71	2.0e-02	CDK2AP2	ENSG00000151332	1.49	2.6e-02	MBIP
ENSG00000112667	-1.60	2.0e-02	DNPH1	ENSG00000185404	1.15	2.6e-02	SP140L
ENSG00000116001	1.36	2.0e-02	TIA1	ENSG00000106086	-2.16	2.6e-02	PLEKHA8
ENSG00000133627	-1.92	2.0e-02	ACTR3B	ENSG00000089820	-1.62	2.6e-02	ARHGAP4
ENSG00000168461	1.88	2.0e-02	RAB31	ENSG00000149679	-2.27	2.7e-02	CABLES2
ENSG00000117597	1.61	2.0e-02	DIEXF	ENSG00000197119	-1.58	2.7e-02	SLC25A29
ENSG00000182983	1.02	2.0e-02	ZNF662	ENSG00000140545	-2.17	2.7e-02	MFGE8
ENSG00000102760	-1.44	2.0e-02	RGCC	ENSG00000182568	0.81	2.7e-02	SATB1
ENSG00000164136	2.09	2.0e-02	IL15	ENSG00000241128	-2.53	2.7e-02	NA
ENSG00000120693	1.87	2.0e-02	SMAD9	ENSG00000276043	-1.10	2.7e-02	UHRF1
ENSG00000154102	-2.35	2.0e-02	C16orf74	ENSG00000105205	-2.53	2.7e-02	CLC
ENSG00000145050	-2.10	2.0e-02	MANF	ENSG00000141968	1.34	2.7e-02	VAV1
ENSG00000154721	1.82	2.0e-02	JAM2	ENSG00000173585	2.53	2.7e-02	CCR9
ENSG00000105483	1.05	2.0e-02	CARD8	ENSG00000046604	1.18	2.7e-02	DSG2
ENSG00000163950	-1.09	2.1e-02	SLBP	ENSG00000182162	-0.98	2.7e-02	P2RY8
ENSG00000122884	1.44	2.1e-02	P4HA1	ENSG00000091409	-1.30	2.7e-02	ITGA6
ENSG00000085832	0.98	2.1e-02	EPS15	ENSG00000142669	0.82	2.7e-02	SH3BGRL3
ENSG00000147894	1.15	2.1e-02	C9orf72	ENSG00000115306	0.94	2.8e-02	SPTBN1
ENSG00000185842	-1.73	2.1e-02	DNAH14	ENSG00000169900	-2.52	2.8e-02	PYDC1
ENSG00000189144	1.72	2.1e-02	ZNF573	ENSG00000204161	2.47	2.8e-02	C10orf128
ENSG00000204131	1.73	2.1e-02	NHSL2	ENSG00000170949	1.35	2.8e-02	ZNF160
ENSG00000162595	2.61	2.1e-02	DIRAS3	ENSG00000137876	0.88	2.8e-02	RSL24D1
ENSG00000184216	-1.38	2.1e-02	IRAK1	ENSG00000067334	0.93	2.8e-02	DNTTIP2
ENSG00000138081	1.05	2.1e-02	FBXO11	ENSG00000129351	-0.92	2.8e-02	ILF3
ENSG00000143376	-1.34	2.1e-02	SNX27	ENSG00000100162	-1.52	2.8e-02	CENPM
ENSG00000141646	1.06	2.1e-02	SMAD4	ENSG00000116138	1.40	2.8e-02	DNAJC16
ENSG00000010810	1.44	2.2e-02	FYN	ENSG00000165525	0.96	2.8e-02	NEMF
ENSG00000136144	1.17	2.2e-02	RCBTB1	ENSG00000171121	-2.04	2.8e-02	KCNMB3
ENSG00000171365	-2.16	2.8e-02	CLCN5	ENSG00000146842	1.19	3.6e-02	TMEM209
ENSG00000205560	-2.28	2.8e-02	CPT1B	ENSG00000177685	-2.44	3.6e-02	CRACR2B
ENSG00000204086	-2.52	2.8e-02	RPA4	ENSG00000196329	-2.43	3.6e-02	GIMAP5
ENSG00000176182	-2.49	2.8e-02	MYPOP	ENSG00000124541	-1.82	3.6e-02	RRP36
ENSG00000155957	0.83	2.8e-02	TMBIM4	ENSG00000135842	-1.38	3.6e-02	FAM129A
ENSG00000071242	2.15	2.9e-02	RPS6KA2	ENSG00000148488	1.94	3.6e-02	ST8SIA6
ENSG00000166197	-0.91	2.9e-02	NOLC1	ENSG00000204257	0.84	3.6e-02	HLA-DMA
ENSG00000064115	-1.02	2.9e-02	TM7SF3	ENSG00000140688	-1.70	3.6e-02	C16orf58
ENSG00000102575	-2.50	2.9e-02	ACP5	ENSG00000123144	-1.25	3.6e-02	C19orf43
ENSG00000104689	1.96	2.9e-02	TNFRSF10A	ENSG00000037042	2.09	3.6e-02	TUBG2
ENSG00000124198	-1.13	2.9e-02	ARFGEF2	ENSG00000155016	1.87	3.6e-02	CYP2U1
ENSG00000145569	-2.04	2.9e-02	FAM105A	ENSG00000170522	1.77	3.6e-02	ELOVL6
ENSG00000180071	-2.41	2.9e-02	ANKRD18A	ENSG00000182944	-0.88	3.6e-02	EWSR1
ENSG00000242028	-1.36	2.9e-02	HYPK	ENSG00000186501	1.73	3.6e-02	TMEM222
ENSG00000109762	2.50	2.9e-02	SNX25	ENSG00000106392	1.28	3.6e-02	C1GALT1
ENSG00000137767	1.31	2.9e-02	SQRDL	ENSG00000196876	2.43	3.6e-02	SCN8A
ENSG00000148516	0.99	2.9e-02	ZEB1	ENSG00000115464	0.89	3.7e-02	USP34
ENSG00000163902	-0.93	2.9e-02	RPN1	ENSG00000157954	-1.38	3.7e-02	WIPI2
ENSG00000147121	1.44	2.9e-02	KRBOX4	ENSG00000084652	1.04	3.7e-02	TXLNA
ENSG00000113648	-0.79	2.9e-02	H2AFY	ENSG00000108091	1.37	3.7e-02	CCDC6
ENSG00000180957	0.90	2.9e-02	PITPNB	ENSG00000122694	1.58	3.7e-02	GLIPR2
ENSG00000211584	1.62	2.9e-02	SLC48A1	ENSG00000137409	-1.59	3.7e-02	MTCH1
ENSG00000005175	1.10	2.9e-02	RPAP3	ENSG00000198015	1.08	3.7e-02	MRPL42
ENSG00000100603	1.20	2.9e-02	SNW1	ENSG00000169398	1.82	3.7e-02	PTK2
ENSG00000079385	1.84	3.0e-02	CEACAM1	ENSG00000070718	1.46	3.7e-02	AP3M2
ENSG00000154277	-2.36	3.0e-02	UCHL1	ENSG00000065328	-1.59	3.7e-02	MCM10
ENSG00000198901	-1.12	3.0e-02	PRC1	ENSG00000077616	2.08	3.7e-02	NAALAD2
ENSG00000140443	1.11	3.0e-02	IGF1R	ENSG00000165732	-0.81	3.8e-02	DDX21
ENSG00000136868	-1.51	3.0e-02	SLC31A1	ENSG00000258227	-2.37	4.3e-02	CLEC5A

Continuation of Table A.24

Ensembl	$\log_2$ Fold Change	padj	Gene Symbol	Ensembl	$\log_2$ Fold Change	padj	Gene Symbol
ENSG00000130635	2.38	3.0e-02	COL5A1	ENSG00000106070	-1.30	4.3e-02	GRB10
ENSG00000167384	1.26	3.0e-02	ZNF180	ENSG00000275052	1.03	4.4e-02	PPP4R3B
ENSG00000092531	1.24	3.0e-02	SNAP23	ENSG00000154309	2.19	4.4e-02	DISP1
ENSG00000113719	1.16	3.0e-02	ERGIC1	ENSG00000180879	-0.92	4.4e-02	SSR4
ENSG00000100647	1.15	3.0e-02	SUSD6	ENSG00000089091	-1.58	4.4e-02	DZANK1
ENSG00000166199	1.43	3.0e-02	ALKBH3	ENSG00000039319	1.11	4.4e-02	ZFYVE16
ENSG00000050344	-1.38	3.0e-02	NFE2L3	ENSG00000178234	-1.60	4.4e-02	GALNT11
ENSG00000100650	1.03	3.0e-02	SRSF5	ENSG00000131368	-1.07	4.4e-02	MRPS25
ENSG00000178229	-1.72	3.0e-02	ZNF543	ENSG00000136205	1.22	4.4e-02	TNS3
ENSG00000135363	-1.10	3.1e-02	LMO2	ENSG00000115421	1.64	4.4e-02	PAPOLG
ENSG00000023839	-2.44	3.1e-02	ABCC2	ENSG00000149313	1.20	4.4e-02	AASDHPPT
ENSG00000173436	-0.96	3.1e-02	MINOS1	ENSG00000164300	1.14	4.4e-02	SERINC5
ENSG00000175756	-1.67	3.1e-02	AURKAIP1	ENSG00000116691	-2.36	4.5e-02	MIIP
ENSG00000116815	-1.61	3.1e-02	CD58	ENSG00000164327	0.94	4.5e-02	RICTOR
ENSG00000237441	1.31	3.1e-02	RGL2	ENSG00000100380	0.79	4.5e-02	ST13
ENSG00000132326	0.94	3.1e-02	PER2	ENSG00000166402	2.35	4.5e-02	TUB
ENSG00000129515	0.96	3.1e-02	SNX6	ENSG00000103249	-2.04	4.5e-02	CLCN7
ENSG00000203760	-1.69	3.1e-02	CENPW	ENSG00000142875	0.92	4.5e-02	PRKACB
ENSG00000160219	1.20	3.1e-02	GAB3	ENSG00000151893	1.19	4.5e-02	CACUL1
ENSG00000144118	1.61	3.2e-02	RALB	ENSG00000079616	-1.46	4.5e-02	KIF22
ENSG00000135049	-1.44	3.2e-02	AGTPBP1	ENSG00000239672	-0.95	4.5e-02	NME1
ENSG00000134146	1.37	3.2e-02	DPH6	ENSG00000111215	1.44	4.5e-02	PRR4
ENSG00000125266	2.47	3.2e-02	EFNB2	ENSG00000023041	1.16	4.5e-02	ZDHHC6
ENSG00000172301	-1.56	3.2e-02	COPRS	ENSG00000135899	1.13	4.6e-02	SP110
ENSG00000164144	1.77	3.2e-02	ARFIP1	ENSG00000175274	2.22	4.6e-02	TP53I11
ENSG00000283486	-1.77	3.2e-02	NA	ENSG00000081803	2.35	4.6e-02	CADPS2
ENSG00000138785	1.25	3.3e-02	INTS12	ENSG00000102359	2.41	3.9e-02	SRPX2
ENSG00000114346	-1.53	3.3e-02	ECT2	ENSG00000214706	-1.97	3.9e-02	IFRD2
ENSG00000119965	-1.96	3.4e-02	C10orf88	ENSG00000136770	1.42	3.9e-02	DNAJC1
ENSG00000164828	1.76	3.4e-02	SUN1	ENSG00000172667	-1.39	3.9e-02	ZMAT3
ENSG00000178163	1.95	3.4e-02	ZNF518B	ENSG00000103404	-1.88	3.9e-02	USP31
ENSG00000171314	-0.92	3.4e-02	PGAM1	ENSG00000088986	-0.87	3.9e-02	DYNLL1
ENSG00000175166	-1.15	3.4e-02	PSMD2	ENSG00000198804	0.77	3.9e-02	COX1
ENSG00000112303	2.30	3.4e-02	VNN2	ENSG00000104774	-1.75	4.0e-02	MAN2B1
ENSG00000149218	1.78	3.4e-02	ENDOD1	ENSG00000100632	-1.00	4.0e-02	ERH
ENSG00000256269	1.53	3.4e-02	HMBS	ENSG00000109743	-1.94	4.0e-02	BST1
ENSG00000124766	0.92	3.4e-02	SOX4	ENSG00000148331	-1.61	4.0e-02	ASB6
ENSG00000166164	0.92	3.4e-02	BRD7	ENSG00000155660	-1.06	4.0e-02	PDIA4
ENSG00000113456	-1.32	3.5e-02	RAD1	ENSG00000247746	-2.22	4.0e-02	USP51
ENSG00000119411	-2.15	3.5e-02	BSPRY	ENSG00000141219	1.76	4.0e-02	C17orf80
ENSG00000130254	-1.40	3.5e-02	SAFB2	ENSG00000185339	2.40	4.0e-02	TCN2
ENSG00000161558	-2.08	3.5e-02	TMEM143	ENSG00000066027	1.15	4.0e-02	PPP2R5A
ENSG00000163913	-1.97	3.5e-02	IFT122	ENSG00000053900	0.92	4.0e-02	ANAPC4
ENSG00000164611	-1.36	3.5e-02	PTTG1	ENSG00000184937	-2.40	4.0e-02	WT1
ENSG00000118503	0.87	3.5e-02	TNFAIP3	ENSG00000146425	1.04	4.0e-02	DYNLT1
ENSG00000124762	2.08	3.5e-02	CDKN1A	ENSG00000136811	1.23	4.0e-02	ODF2
ENSG00000096872	1.50	3.5e-02	IFT74	ENSG00000103150	-1.93	4.0e-02	MLYCD
ENSG00000107897	-1.83	3.5e-02	ACBD5	ENSG00000104728	1.83	4.0e-02	ARHGEF10
ENSG00000166147	1.70	3.5e-02	FBN1	ENSG00000119929	1.02	4.0e-02	CUTC
ENSG00000141522	-1.96	3.5e-02	ARHGDIA	ENSG00000136717	2.13	4.0e-02	BIN1
ENSG00000088451	1.20	3.6e-02	TGDS	ENSG00000140386	1.08	4.0e-02	SCAPER
ENSG00000100129	0.82	3.6e-02	EIF3L	ENSG00000168081	2.39	4.1e-02	PNOC
ENSG00000126860	1.19	3.6e-02	EVI2A	ENSG00000163636	0.99	4.1e-02	PSMD6
ENSG00000162654	0.94	3.6e-02	GBP4	ENSG00000119688	1.09	4.1e-02	ABCD4
ENSG00000168016	1.56	3.6e-02	TRANK1	ENSG00000117054	1.00	4.1e-02	ACADM
ENSG00000119684	1.16	3.6e-02	MLH3	ENSG00000144840	-1.60	4.1e-02	RABL3
ENSG00000133316	-1.03	4.3e-02	WDR74	ENSG00000206053	-1.18	4.1e-02	HN1L
ENSG00000187800	2.34	4.3e-02	PEAR1	ENSG00000165632	1.48	4.1e-02	TAF3
ENSG00000123064	-1.46	4.3e-02	DDX54	ENSG00000166446	1.99	4.1e-02	CDYL2
ENSG00000162889	-1.37	4.3e-02	MAPKAPK2	ENSG00000167992	-2.12	4.1e-02	VWCE
ENSG00000116489	0.73	4.3e-02	CAPZA1	ENSG00000089335	-1.63	4.1e-02	ZNF302
ENSG00000064225	-1.26	4.3e-02	ST3GAL6	ENSG00000168268	-1.59	4.1e-02	NT5DC2
ENSG00000115947	1.02	4.3e-02	ORC4	ENSG00000180354	1.25	4.1e-02	MTURN
ENSG00000119541	1.08	4.6e-02	VPS4B	ENSG00000136732	-1.10	4.2e-02	GYPC
ENSG00000025293	0.96	4.6e-02	PHF20	ENSG00000175356	2.12	4.2e-02	SCUBE2
ENSG00000087903	-1.28	4.6e-02	RFX2	ENSG00000003756	0.80	4.2e-02	RBM5
ENSG00000110436	-1.95	4.6e-02	SLC1A2	ENSG00000101266	0.93	4.2e-02	CSNK2A1
ENSG00000131238	0.93	4.6e-02	PPT1	ENSG00000111536	2.38	4.2e-02	IL26
ENSG00000184428	-1.33	4.6e-02	TOP1MT	ENSG00000081985	1.31	4.2e-02	IL12RB2
ENSG00000137309	-1.16	4.6e-02	HMGA1	ENSG00000167967	-2.15	4.2e-02	E4F1
ENSG00000140749	-1.60	4.6e-02	IGSF6	ENSG00000130363	1.32	4.2e-02	RSPH3
ENSG00000136937	-1.17	4.6e-02	NCBP1	ENSG00000139278	0.93	4.2e-02	GLIPR1
ENSG00000030582	-0.95	4.7e-02	GRN	ENSG00000161203	0.97	4.2e-02	AP2M1
ENSG00000161542	1.15	4.7e-02	PRPSAP1	ENSG00000204217	1.33	4.2e-02	BMPR2
ENSG00000160888	-1.26	4.7e-02	IER2	ENSG00000158715	-2.34	4.3e-02	SLC45A3
ENSG00000074966	1.78	4.7e-02	TXK	ENSG00000172765	1.28	4.3e-02	TMCC1
ENSG00000104442	1.02	4.7e-02	ARMC1	ENSG00000109805	-2.13	4.3e-02	NCAPG
ENSG00000196954	1.16	4.7e-02	CASP4	ENSG00000135452	1.11	4.3e-02	TSPAN31
ENSG00000144451	1.38	4.7e-02	SPAG16	ENSG00000166123	-1.74	4.3e-02	GPT2
ENSG00000170037	-1.74	4.7e-02	CNTROB	ENSG00000037637	-1.35	4.3e-02	FBXO42
ENSG00000174796	1.27	4.7e-02	THAP6	ENSG00000038210	1.54	4.3e-02	PI4K2B
ENSG00000164975	1.26	4.7e-02	SNAPC3	ENSG00000052126	1.06	4.3e-02	PLEKHA5
ENSG00000204983	2.33	4.7e-02	PRSS1	ENSG00000052802	1.48	4.3e-02	MSMO1
ENSG00000155926	0.88	4.7e-02	SLA	ENSG00000163811	-0.87	5.0e-02	WDR43
ENSG00000164050	-2.11	4.7e-02	PLXNB1	ENSG00000074842	-1.46	5.0e-02	MYDGF
ENSG00000177565	0.86	4.7e-02	TBL1XR1	ENSG00000084676	1.07	5.0e-02	NCOA1
ENSG00000075945	1.37	4.7e-02	KIFAP3	ENSG00000085274	1.33	5.0e-02	MYNN
ENSG00000109814	-1.58	4.7e-02	UGDH	ENSG00000157349	1.06	5.0e-02	DDX19B
ENSG00000173890	-2.08	4.7e-02	GPR160	ENSG00000156345	-2.30	5.0e-02	CDK20
ENSG00000007129	1.52	4.7e-02	CEACAM21	ENSG00000181666	1.23	5.0e-02	HKR1
ENSG00000124171	1.58	4.8e-02	PARD6B				

Continuation of Table A.24

Ensembl	log_2 Fold Change	padj	Gene Symbol	Ensembl	log_2 Fold Change	padj	Gene Symbol
ENSG00000145012	1.03	4.8e-02	LPP				
ENSG00000161888	-1.71	4.8e-02	SPC24				
ENSG00000126524	1.13	4.8e-02	SBDS				
ENSG00000128604	-1.14	4.8e-02	IRF5				
ENSG00000174469	1.62	4.8e-02	CNTNAP2				
ENSG00000117676	-1.32	4.8e-02	RPS6KA1				
ENSG00000106636	-1.87	4.8e-02	YKT6				
ENSG00000106261	0.86	4.8e-02	ZKSCAN1				
ENSG00000177409	1.11	4.8e-02	SAMD9L				
ENSG00000166401	-1.99	4.8e-02	SERPINB8				
ENSG00000072778	-1.01	4.9e-02	ACADVL				
ENSG00000073584	0.93	4.9e-02	SMARCE1				
ENSG00000187243	2.20	4.9e-02	MAGED4B				
ENSG00000239306	-1.30	4.9e-02	RBM14				
ENSG00000253729	-0.81	4.9e-02	PRKDC				
ENSG00000015676	-1.10	4.9e-02	NUDCD3				
ENSG00000011638	1.16	4.9e-02	TMEM159				
ENSG00000128585	0.99	4.9e-02	MKLN1				
ENSG00000134460	1.66	4.9e-02	IL2RA				
ENSG00000146005	2.31	4.9e-02	PSD2				
ENSG00000189079	0.82	4.9e-02	ARID2				
ENSG00000224877	-2.19	4.9e-02	NDUFAF8				
ENSG00000134480	0.85	5.0e-02	CCNH				

Table A.28: DEGs Between Leukaemic LMPP and GMP. DEGs were identified between leukaemic LMPP and GMP, using a Wald test in DESeq2. A design formula of ~replicate + population was used. Upregulated and downregulated DEGs with a p-adjusted value < 0.05 are shown.

Ensembl	$\log_2$ Fold Change	padj	Gene Symbol	Ensembl	$\log_2$ Fold Change	padj	Gene Symbol
ENSG00000004468	-2.71	8.7e-22	CD38	ENSG00000132646	-1.21	0.011	PCNA
ENSG00000139734	-2.20	1.5e-10	DIAPH3	ENSG00000011426	-1.29	0.011	ANLN
ENSG00000166803	-1.98	4.9e-09	KIAA0101	ENSG00000103490	-0.86	0.011	PYCARD
ENSG00000089685	-2.02	2.1e-08	BIRC5	ENSG00000121152	-1.38	0.011	NCAPH
ENSG00000104738	-1.27	2.8e-07	MCM4	ENSG00000132780	-0.64	0.011	NASP
ENSG00000165304	-2.04	7.4e-07	MELK	ENSG00000133401	1.12	0.011	PDZD2
ENSG00000004534	0.42	1.7e-06	RBM6	ENSG00000005022	-0.47	0.012	SLC25A5
ENSG00000276043	-1.38	1.0e-05	UHRF1	ENSG00000130921	0.58	0.012	C12orf65
ENSG00000168411	-0.80	1.7e-05	RFWD3	ENSG00000198856	-0.60	0.012	OSTC
ENSG00000102804	0.77	1.8e-05	TSC22D1	ENSG00000121067	0.67	0.012	SPOP
ENSG00000092853	-1.64	3.0e-05	CLSPN	ENSG00000169083	1.07	0.012	AR
ENSG00000070081	-0.66	4.2e-05	NUCB2	ENSG00000115762	-0.59	0.012	PLEKHB2
ENSG00000005381	-1.80	1.9e-04	MPO	ENSG00000137501	0.83	0.012	SYTL2
ENSG00000051341	-1.57	1.9e-04	POLQ	ENSG00000197345	-0.57	0.012	MRPL21
ENSG00000100354	0.53	1.9e-04	TNRC6B	ENSG00000080839	-1.02	0.012	RBL1
ENSG00000198830	-0.74	2.4e-04	HMGN2	ENSG00000110367	0.38	0.013	DDX6
ENSG00000163297	0.88	2.7e-04	ANTXR2	ENSG00000147996	0.47	0.013	CBWD5
ENSG00000003436	0.70	3.6e-04	TFPI	ENSG00000196776	0.43	0.013	CD47
ENSG00000145287	-0.97	4.2e-04	PLAC8	ENSG00000123416	-0.61	0.013	TUBA1B
ENSG00000164104	-0.88	4.4e-04	HMGB2	ENSG00000118640	-0.48	0.013	VAMP8
ENSG00000107447	-1.37	5.5e-04	DNTT	ENSG00000136709	0.68	0.013	WDR33
ENSG00000117632	-0.91	7.1e-04	STMN1	ENSG00000163918	-1.18	0.014	RFC4
ENSG00000090382	-1.61	1.0e-03	LYZ	ENSG00000163808	-1.34	0.014	KIF15
ENSG00000119535	-0.93	1.0e-03	CSF3R	ENSG00000169057	0.89	0.014	MECP2
ENSG00000018189	1.09	1.4e-03	RUFY3	ENSG00000116688	-1.12	0.015	MFN2
ENSG00000123975	-0.62	1.4e-03	CKS2	ENSG00000160213	-0.62	0.015	CSTB
ENSG00000157514	0.75	1.5e-03	TSC22D3	ENSG00000169871	0.77	0.015	TRIM56
ENSG00000075239	-1.00	1.5e-03	ACAT1	ENSG00000116604	-0.57	0.015	MEF2D
ENSG00000101096	1.11	1.6e-03	NFATC2	ENSG00000099622	0.31	0.015	CIRBP
ENSG00000126267	-0.53	1.6e-03	COX6B1	ENSG00000180530	0.57	0.015	NRIP1
ENSG00000143995	0.81	1.6e-03	MEIS1	ENSG00000213203	1.02	0.016	GIMAP1
ENSG00000155368	-0.58	1.6e-03	DBI	ENSG00000179387	0.97	0.017	ELMOD2
ENSG00000104889	-1.10	1.6e-03	RNASEH2A	ENSG00000151893	0.43	0.017	CACUL1
ENSG00000183054	0.74	1.7e-03	RGPD5	ENSG00000148773	-1.31	0.017	MKI67
ENSG00000112312	-1.28	1.8e-03	GMNN	ENSG00000164300	0.73	0.017	SERINC5
ENSG00000137804	-1.42	1.8e-03	NUSAP1	ENSG00000171603	-0.74	0.017	CLSTN1
ENSG00000187239	0.38	1.8e-03	FNBP1	ENSG00000005379	-1.14	0.018	TSPOAP1
ENSG00000152601	0.49	1.9e-03	MBNL1	ENSG00000014641	-0.71	0.018	MDH1
ENSG00000100162	-1.14	2.5e-03	CENPM	ENSG00000084110	-1.31	0.018	HAL
ENSG00000151773	0.34	2.5e-03	CCDC122	ENSG00000113851	0.48	0.018	CRBN
ENSG00000085840	-1.45	2.6e-03	ORC1	ENSG00000144354	-1.16	0.018	CDCA7
ENSG00000113648	-0.56	2.6e-03	H2AFY	ENSG00000152520	0.26	0.018	PAN3
ENSG00000142657	-0.48	2.6e-03	PGD	ENSG00000184009	-0.50	0.018	ACTG1
ENSG00000178999	-1.52	2.7e-03	AURKB	ENSG00000272325	0.45	0.018	NUDT3
ENSG00000183856	-1.52	2.9e-03	IQGAP3	ENSG00000139218	0.19	0.019	SCAF11
ENSG00000161888	-0.83	2.9e-03	SPC24	ENSG00000139343	-0.50	0.019	SNRPF
ENSG00000228716	-1.12	3.2e-03	DHFR	ENSG00000163792	0.54	0.020	TCF23
ENSG00000180354	0.96	3.5e-03	MTURN	ENSG00000164032	-0.61	0.020	H2AFZ
ENSG00000227471	0.53	3.5e-03	AKR1B15	ENSG00000100368	-1.27	0.020	CSF2RB
ENSG00000100448	-1.47	4.0e-03	CTSG	ENSG00000134333	-0.38	0.021	LDHA
ENSG00000117650	-1.29	4.2e-03	NEK2	ENSG00000154188	0.88	0.021	ANGPT1
ENSG00000165480	-1.43	4.2e-03	SKA3	ENSG00000160014	-0.51	0.021	CALM2
ENSG00000108510	0.35	4.3e-03	MED13	ENSG00000033170	0.39	0.022	FUT8
ENSG00000138092	-1.07	4.3e-03	CENPO	ENSG00000047621	-1.01	0.022	C12orf4
ENSG00000166451	-1.28	4.7e-03	CENPN	ENSG00000269313	0.66	0.022	MAGIX
ENSG00000198795	1.18	5.4e-03	ZNF521	ENSG00000164220	1.21	0.022	F2RL2
ENSG00000124486	0.39	5.4e-03	USP9X	ENSG00000147364	0.71	0.022	FBXO25
ENSG00000126524	0.36	6.0e-03	SBDS	ENSG00000166949	0.83	0.022	SMAD3
ENSG00000174197	0.78	6.1e-03	MGA	ENSG00000166508	-0.40	0.022	MCM7
ENSG00000107829	0.67	6.2e-03	FBXW4	ENSG00000112308	0.23	0.023	C6orf62
ENSG00000122862	-0.80	6.2e-03	SRGN	ENSG00000165494	0.47	0.023	PCF11
ENSG00000116747	0.42	6.2e-03	TROVE2	ENSG00000169020	-0.39	0.023	ATP5I
ENSG00000162664	0.39	6.7e-03	ZNF326	ENSG00000204131	0.89	0.023	NHSL2
ENSG00000120963	0.56	6.8e-03	ZNF706	ENSG00000067082	0.51	0.024	KLF6
ENSG00000128951	-0.98	7.1e-03	DUT	ENSG00000167554	1.18	0.024	ZNF610
ENSG00000139163	0.73	7.4e-03	ETNK1	ENSG00000180817	-0.58	0.024	PPA1
ENSG00000034152	-1.03	7.4e-03	MAP2K3	ENSG00000035720	-0.90	0.024	STAP1
ENSG00000094804	-1.36	7.4e-03	CDC6	ENSG00000114933	0.54	0.025	INO80D
ENSG00000118193	-1.00	7.4e-03	KIF14	ENSG00000079482	0.34	0.025	OPHN1
ENSG00000136247	-0.69	7.4e-03	ZDHHC4	ENSG00000213865	0.60	0.025	C8orf44
ENSG00000138160	-1.28	7.4e-03	KIF11	ENSG00000113140	-0.87	0.025	SPARC
ENSG00000182831	0.35	7.7e-03	C16orf72	ENSG00000114656	0.49	0.025	KIAA1257
ENSG00000162852	0.64	8.4e-03	CNST	ENSG00000101624	-0.86	0.026	CEP76
ENSG00000114861	0.34	8.5e-03	FOXP1	ENSG00000178104	0.55	0.026	PDE4DIP
ENSG00000115464	0.45	8.5e-03	USP34	ENSG00000161204	-0.71	0.026	ABCF3
ENSG00000113580	0.56	8.8e-03	NR3C1	ENSG00000168894	-0.39	0.027	RNF181
ENSG00000099901	-0.74	9.5e-03	RANBP1	ENSG00000173812	0.30	0.027	EIF1
ENSG00000151366	-0.53	9.5e-03	NDUFC2	ENSG00000174371	-1.23	0.027	EXO1
ENSG00000100028	-0.63	9.8e-03	SNRPD3	ENSG00000105339	-0.63	0.028	DENND3
ENSG00000164611	-0.96	9.8e-03	PTTG1	ENSG00000137309	-0.62	0.028	HMGA1
ENSG00000176890	-1.17	9.8e-03	TYMS	ENSG00000159140	0.29	0.028	SON
ENSG00000095932	-0.66	9.9e-03	SMIM24	ENSG00000111669	-0.40	0.028	TPI1
ENSG00000156970	-1.29	1.0e-02	BUB1B	ENSG00000173207	-0.99	0.028	CKS1B

Continuation of Table A.24

Ensembl	log_2 Fold Change	padj	Gene Symbol	Ensembl	log_2 Fold Change	padj	Gene Symbol
ENSG00000173517	0.78	1.1e-02	PEAK1	ENSG00000090989	0.90	0.028	EXOC1
ENSG00000119599	0.93	2.8e-02	DCAF4	ENSG00000171530	-0.55	0.042	TBCA
ENSG00000138777	-0.61	2.8e-02	PPA2	ENSG00000084234	-0.52	0.042	APLP2
ENSG00000099956	-0.80	2.9e-02	SMARCB1	ENSG00000106268	-0.68	0.042	NUDT1
ENSG00000130816	-0.70	2.9e-02	DNMT1	ENSG00000108561	-0.78	0.042	C1QBP
ENSG00000106554	-0.80	3.0e-02	CHCHD3	ENSG00000137807	-1.00	0.042	KIF23
ENSG00000121691	-0.59	3.0e-02	CAT	ENSG00000172888	0.65	0.042	ZNF621
ENSG00000134291	-0.94	3.0e-02	TMEM106C	ENSG00000196743	-0.88	0.043	GM2A
ENSG00000169180	-0.87	3.0e-02	XPO6	ENSG00000254004	-0.86	0.043	ZNF260
ENSG00000122482	0.37	3.1e-02	ZNF644	ENSG00000128923	0.72	0.044	FAM63B
ENSG00000164109	-1.00	3.1e-02	MAD2L1	ENSG00000115524	0.33	0.045	SF3B1
ENSG00000077984	-1.20	3.1e-02	CST7	ENSG00000173889	0.39	0.045	PHC3
ENSG00000104765	0.38	3.1e-02	BNIP3L	ENSG00000117399	-1.16	0.045	CDC20
ENSG00000118263	0.58	3.1e-02	KLF7	ENSG00000158604	0.33	0.045	TMED4
ENSG00000121966	-0.88	3.1e-02	CXCR4	ENSG00000271503	-0.77	0.045	CCL5
ENSG00000168575	1.15	3.1e-02	SLC20A2	ENSG00000186026	0.90	0.046	ZNF284
ENSG00000196230	-0.52	3.1e-02	TUBB	ENSG00000141644	0.90	0.046	MBD1
ENSG00000082074	-0.94	3.1e-02	FYB	ENSG00000182885	-1.08	0.046	ADGRG3
ENSG00000132669	0.98	3.1e-02	RIN2	ENSG00000107581	-0.43	0.046	EIF3A
ENSG00000006747	0.35	3.2e-02	SCIN	ENSG00000120992	-0.50	0.047	LYPLA1
ENSG00000070761	0.70	3.2e-02	CFAP20	ENSG00000128050	-0.65	0.047	PAICS
ENSG00000138036	-0.91	3.2e-02	DYNC2LI1	ENSG00000141867	0.66	0.047	BRD4
ENSG00000183166	0.99	3.2e-02	CALN1	ENSG00000106992	0.77	0.049	AK1
ENSG00000182557	-1.00	3.2e-02	SPNS3	ENSG00000166226	-0.37	0.049	CCT2
ENSG00000116539	0.48	3.2e-02	ASH1L	ENSG00000244754	0.41	0.049	N4BP2L2
ENSG00000026652	0.99	3.3e-02	AGPAT4	ENSG00000135686	0.61	0.049	KLHL36
ENSG00000140022	0.78	3.3e-02	STON2	ENSG00000144591	-1.07	0.049	GMPPA
ENSG00000157456	-1.20	3.3e-02	CCNB2	ENSG00000129315	0.55	0.050	CCNT1
ENSG00000176715	0.60	3.6e-02	ACSF3	ENSG00000181924	-0.72	0.050	COA4
ENSG00000141665	0.33	3.7e-02	FBXO15	ENSG00000196428	0.38	0.050	TSC22D2
ENSG00000124193	0.32	3.7e-02	SRSF6				
ENSG00000108848	0.31	3.7e-02	LUC7L3				
ENSG00000118939	-0.64	3.7e-02	UCHL3				
ENSG00000116161	-0.52	3.7e-02	CACYBP				
ENSG00000111665	-0.95	3.8e-02	CDCA3				
ENSG00000172264	0.48	3.8e-02	MACROD2				
ENSG00000186815	-0.92	3.8e-02	TPCN1				
ENSG00000164405	-0.59	3.8e-02	UQCRQ				
ENSG00000181722	0.65	3.8e-02	ZBTB20				
ENSG00000196214	0.64	3.8e-02	ZNF766				
ENSG00000101745	0.52	3.8e-02	ANKRD12				
ENSG00000196262	-0.43	3.8e-02	PPIA				
ENSG00000237515	0.25	3.9e-02	SHISA9				
ENSG00000163453	-0.78	4.0e-02	IGFBP7				
ENSG00000184677	1.00	4.0e-02	ZBTB40				
ENSG00000069020	0.63	4.1e-02	MAST4				

A.13 CRISPR-Cas9 Editing of K562 Cells

Table A.29: Summary of ICE Scores in K562 Cells. K562 cells were transfected with PX458 plasmids containing sgRNA against *CLP1*, *COPS2*, *FASTKD5*, *PSMD12*, *SBDS* and *WDR26*. 48 hours post-transfection DNA was extracted for sequencing to assess gene editing. Gene editing was assessed using ICE Scores [90]. Shown are the ICE scores across technical replicates (A/B) and the average ICE score per sgRNA. sgRNAs are as indicated in Table A.3. Technical replicates A/B.1/2 refer to a technical replicate (A/B) compared to 2 wild-type control sequences (.1/2). F: Sanger sequencing forward read, R: Sanger sequencing reverse read.

Gene	Guide	Replicate	Read	ICE Score
CLP1	1	A	F	34
		A	R	46
		B	F	37
		B	R	43
		Average		40
	2	A	F	38
		A	R	18
		B	F	13
		B	R	20
		Average		22.25
	4	A	F	67
		A	R	68
		B	F	55
		B	R	50
		Average		60
	5	A	F	30
		A	R	31
		B	F	39
		B	R	40
		Average		35
COPS2	1	A	F	42
		A	R	39
		B	F	59
		B	R	62
		Average		50.5
	4	A	F	2
		B	F	0
		Average		1
	5	A	R	37
		B	R	26
		Average		31.5

Continuation of Table A.29				
Gene	Guide	Replicate	Read	ICE Score
FASTKD5	1	A	F	50
		A	R	47
		B	F	50
		B	R	48
		Average		48.75
PSMD12	1	A	F	20
		A	R	16
		B	F	25
		B	R	27
		Average		22
	2	A	F	56
		A	R	62
		B	F	60
		B	R	48
		Average		56.5
	4	A	F	29
		A	R	23
		B	F	36
		B	R	32
		Average		30
	5	A	F	15
		A	R	9
		B	F	15
		B	R	10
		Average		12.25
SBDS	3	A.1	F	8
		A.2	F	5
		B.1	F	13
		B.2	F	5
		Average		7.75
	4	A.1	F	12
		A.2	F	14
		B.1	F	22
		B.2	F	21
		Average		17.25
WDR26	1	A	F	41
		B	F	47
		Average		44
	2	A	F	77
		B	F	63
		Average		70
	3	A	F	38
		B	F	37

Continuation of Table A.29				
Gene	Guide	Replicate	Read	ICE Score
		Average		37.5
	4	A	F	16
		B	F	18
		Average		17
	5	A	F	35
		B	F	44
		Average		39.5

A.14 Optimising AML Culture

Table A.30: Effect of MSC on Annexin-V$^+$ t-test Summary. The effect of AML culture in the presence and absence of a human BM derived MSC stromal layer [84] on Annexin-V$^+$ was investigated (Figure 6.16). PDX samples were cultured in the presence and absence of a MSC stromal layer for 15 days. Either 10,000 or 50,000 PDX cells were plated on D1 of culture. Cultures were in the presence of absence of SR1 and UM171 (see Chapter 2.15 for full culture conditions). Shown are the p-values from a t-test comparing the Annexin-V$^+$ population in each condition in the presence and absence of MSCs. C: SR1 and UM171 compounds, M: MSC.

Comparison	D4	D7	D12	D15
BIR344				
10000 + C vs 10000 + M + C	0.010	0.002	0.015	0.016
50000 + C vs 50000 + M + C	0.058	0.006	0.101	0.263
10000 vs 10000 + M	0.017	0.010	0.672	0.262
50000 vs 50000 + M	0.005	0.022	0.115	0.358
BIR517				
10000 + C vs 10000 + M + C	0.016	0.004	0.172	0.133
50000 + C vs 50000 + M + C	0.044	0.003	0.149	0.145
10000 vs 10000 + M	0.038	0.013	0.012	0.597
50000 vs 50000 + M	0.074	0.261	0.016	0.551
STB257				
10000 + C vs 10000 + M + C	0.760	0.024	0.635	0.270
50000 + C vs 50000 + M + C	0.226	0.817	0.474	0.157
10000 vs 10000 + M	0.065	0.003	0.560	0.266
50000 vs 50000 + M	0.104	0.007	0.060	0.019

References

[1] C. M. Baum et al. “Isolation of a candidate human hematopoietic stem-cell population.” In: *Proceedings of the National Academy of Sciences* 89.7 (1992), pp. 2804–2808. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.89.7.2804.

[2] M Bhatia et al. “Purification of primitive human hematopoietic cells capable of repopulating immune-deficient mice.” In: *Proceedings of the National Academy of Sciences of the United States of America* 94.10 (1997), pp. 5320–5325.

[3] Ravindra Majeti, Christopher Y Park, and Irving L Weissman. “Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood.” In: *Cell stem cell* 1.6 (Dec. 2007), pp. 635–45. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2292126%7B%5C&%7Dtool=pmcentrez%7B%5C&%7Drendertype=abstract.

[4] Faiyaz Notta et al. “Isolation of single human hematopoietic stem cells capable of long-term multilineage engraftment.” In: *Science (New York, N.Y.)* 333.6039 (July 2011), pp. 218–21. URL: http://www.ncbi.nlm.nih.gov/pubmed/21737740.

[5] Brad Dykstra et al. “Long-Term Propagation of Distinct Hematopoietic Differentiation Programs In Vivo”. In: *Cell Stem Cell* 1.2 (2007), pp. 218–229.

[6] Grant A. Challen et al. “Distinct Hematopoietic Stem Cell Subtypes Are Differentially Regulated by TGF-β1”. In: *Cell Stem Cell* 6.3 (2010), pp. 265–278. URL: http://dx.doi.org/10.1016/j.stem.2010.02.002.

[7] Eric M. Pietras et al. “Functionally Distinct Subsets of Lineage-Biased Multipotent Progenitors Control Blood Production in Normal and Regenerative Conditions”. In: *Cell Stem Cell* 17.1 (2015), pp. 35–46.

[8] Robert Månsson et al. “Molecular Evidence for Hierarchical Transcriptional Lineage Priming in Fetal and Adult Stem Cells and Multipotent Progenitors”. In: *Immunity* 26.April (2007), pp. 407–419.

[9] Ryo Yamamoto et al. “Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells.” In: *Cell* 154.5 (Aug. 2013), pp. 1112–26. URL: http://www.ncbi.nlm.nih.gov/pubmed/23993099.

[10] Alejandra Sanjuan-Pla et al. “Platelet-biased stem cells reside at the apex of the haematopoietic stem-cell hierarchy.” In: *Nature* 502.7470 (Oct. 2013), pp. 232–6. URL: http://www.ncbi.nlm.nih.gov/pubmed/23934107.

[11] Joana Carrelha et al. “Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells”. In: *Nature* 554.7690 (2018), pp. 106–111. URL: http://dx.doi.org/10.1038/nature25455.

[12] Faiyaz Notta et al. “Distinct routes of lineage development reshape the human blood hierarchy across ontogeny”. In: *Science* November (2015), pp. 1–16.

[13] Koichi Akashi et al. “A clonogenic common myeloid progenitor that gives rise to all myeloid lineages”. In: *Nature* 404.6774 (2000), pp. 193–197.

[14] Markus G Manz et al. “Prospective isolation of human clonogenic common myeloid progenitors.” In: *Proceedings of the National Academy of Sciences of the United States of America* 99.18 (Sept. 2002), pp. 11872–7. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=129361%7B%5C&%7Dtool=pmcentrez%7B%5C&%7Drendertype=abstract.

[15] M Kondo et al. “Identification of clonogenic common lymphoid progenitors in mouse bone marrow.” In: *Cell* 91.5 (1997), pp. 661–72. URL: http://www.ncbi.nlm.nih.gov/pubmed/9393859.

[16] I Hoebeke et al. “T-, B- and NK-lymphoid, but not myeloid cells arise from human CD34(+)CD38(-)CD7(+) common lymphoid progenitors expressing lymphoid-specific genes.” In: *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, U.K* 21 (2007), pp. 311–319.

[17] Leïla Perié et al. “The Branching Point in Erythro-Myeloid Differentiation”. In: *Cell* 163.7 (2015), pp. 1655–1662.

[18] Franziska Paul et al. “Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors”. In: *Cell* 163.7 (2015), pp. 1663–1677. URL: http://dx.doi.org/10.1016/j.cell.2015.11.013.

[19] M Lu et al. “The common myelolymphoid progenitor: a key intermediate stage in hemopoiesis generating T and B cells”. In: *J Immunol* 169.7 (2002), pp. 3519–3525. URL: http://www.ncbi.nlm.nih.gov/pubmed/12244141.

[20] Jörgen Adolfsson et al. “Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment.” In: *Cell* 121.2 (Apr. 2005), pp. 295–306. URL: http://www.ncbi.nlm.nih.gov/pubmed/15851035.

[21] Anne Y. Lai and Motonari Kondo. “Asymmetrical lymphoid and myeloid lineage commitment in multipotent hematopoietic progenitors”. In: *Journal of Experimental Medicine* 203.8 (2006), pp. 1867–1873. arXiv: NIHMS150003.

[22] Nicolas Goardon et al. “Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia.” In: *Cancer cell* 19.1 (Jan. 2011), pp. 138–52. URL: http://www.ncbi.nlm.nih.gov/pubmed/21251617.

[23] Shalin H Naik et al. “Diverse and heritable lineage imprinting of early haematopoietic progenitors.” In: *Nature* 496 (2013), pp. 229–32. URL: http://www.ncbi.nlm.nih.gov/pubmed/23552896.

[24] Sergei Doulatov et al. “Revised map of the human progenitor hierarchy shows the origin of macrophages and dendritic cells in early lymphoid development.” In: *Nature immunology* 11.7 (July 2010), pp. 585–93. URL: http://www.ncbi.nlm.nih.gov/pubmed/20543838.

[25] Elisa Laurenti et al. “The transcriptional architecture of early human hematopoiesis identifies multilevel control of lymphoid commitment.” In: *Nature immunology* 14.7 (July 2013), pp. 756–63. URL: http://www.ncbi.nlm.nih.gov/pubmed/23708252.

[26] Peter van Galen et al. “Reduced lymphoid lineage priming promotes human hematopoietic stem cell expansion.” In: *Cell stem cell* 14.1 (Jan. 2014), pp. 94–106. URL: http://www.ncbi.nlm.nih.gov/pubmed/24388174.

[27] Matthias Farlik et al. “DNA Methylation Dynamics of Human Hematopoietic Stem Cell Differentiation”. In: *Cell Stem Cell* 0.0 (2016), pp. 224–226. URL: http://linkinghub.elsevier.com/retrieve/pii/S1934590916303605.

[28] Lars Velten et al. “Human haematopoietic stem cell lineage commitment is a continuous process without discrete progenitor populations”. In: *Nature Cell Biology* 19.4 (2017).

[29] Lisa a Kohn et al. “Lymphoid priming in human bone marrow begins before expression of CD10 with upregulation of L-selectin.” In: *Nature immunology* 13.10 (Oct. 2012), pp. 963–71. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3448017%7B%5C&%7Dtool=pmcentrez%7B%5C&%7Drendertype=abstract.

[30] H. Döhner et al. “Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet”. In: *Blood* 115.3 (2010), pp. 453–474. URL: http://bloodjournal.hematologylibrary.org/content/115/3/453.short.

[31] Clémentine Sarkozy et al. “Outcome of older patients with acute myeloid leukemia in first relapse”. In: *American Journal of Hematology* 88.9 (2013), pp. 758–764.

[32] John E Dick. “ASH 50th anniversary review Stem cell concepts renew cancer research”. In: *Stem Cells* 112.13 (2008), pp. 4793–4807.

[33] B. Griffin, J. D. & Lowenberg. “Clonogenic cells in acute myeloblastic leukemia”. In: *Blood* 68.6 (1986), pp. 1185–95.

[34] T Lapidot et al. “A cell initiating human acute myeloid leukaemia after transplantation into SCID mice.” In: *Nature* 367.6464 (1994), pp. 645–648. URL: http://www.nature.com.ez.srv.meduniwien.ac.at/nature/journal/v367/n6464/abs/367645a0.html.

[35] D Bonnet and J E Dick. “Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell.” In: *Nature medicine* 3.7 (1997), pp. 730–737.

[36] David C Taussig et al. “Anti-CD38 antibody mediated clearance of human repopulating cells masks the heterogeneity of leukemia initiating cells Anti-CD38 antibody mediated clearance of human repopulating cells masks the heterogeneity of leukemia initiating cells Corresponding aut”. In: *Blood* 112.3 (2008), pp. 568–576.

[37] Lynn Quek et al. “Genetically distinct leukemic stem cells in human CD34 acute myeloid leukemia are arrested at a hemopoietic precursor-like stage”. In: *The Journal of Experimental Medicine* (2016), jem.20151775. URL: http://www.jem.org/lookup/doi/10.1084/jem.20151775.

[38] a Blair et al. “Lack of expression of Thy-1 (CD90) on acute myeloid leukemia cells with long-term proliferative ability in vitro and in vivo.” In: *Blood* 89.9 (1997), pp. 3104–3112.

[39] Kristin J Hope, Liqing Jin, and John E Dick. “Acute myeloid leukemia originates from a hierarchy of leukemic stem cell classes that differ in self-renewal capacity.” In: *Nature immunology* 5.7 (2004), pp. 738–743.

[40] Max Jan et al. “Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia.” In: *Science translational medicine* 4.149 (2012), 149ra118. URL: http://dx.doi.org/10.1016/j.beha.2014.10.003%7B%5C%7D5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/22932223.

[41] M Ryan Corces-Zimmerman et al. “Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission.” In: *Proceedings of the National Academy of Sciences of the United States of America* 111.7 (2014), pp. 2548–53. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3932921%7B%5C&%7Dtool=pmcentrez%7B%5C&%7Drendertype=abstract.

[42] Liran I. Shlush et al. “Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia”. In: *Nature* 506.7488 (2014), pp. 328–333. arXiv: arXiv:1011.1669v3. URL: http://dx.doi.org/10.1038/nature13038.

[43] Alejandro Chavez et al. “Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes”. In: *New England Journal of Medicine* 371.26 (2014), pp. 2488–2498.

[44] Andrew L Young et al. “Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults”. In: *Nature Communications* 7 (2016), pp. 1–7. URL: http://dx.doi.org/10.1038/ncomms12484.

[45] D. H. Wiseman, B. F. Greystoke, and T. C P Somervaille. “The variety of leukemic stem cells in myeloid malignancy”. In: *Oncogene* 33.24 (2014), pp. 3091–3098. URL: http://dx.doi.org/10.1038/onc.2013.269.

[46] Daniel Thomas and Ravindra Majeti. “Biology and relevance of human acute myeloid leukemia stem cells”. In: 129.12 (2017), pp. 1577–1586.

[47] Kolja Eppert et al. “Stem cell gene expression programs influence clinical outcome in human leukemia.” In: *Nature medicine* 17.9 (2011), pp. 1086–1093. URL: http://dx.doi.org/10.1038/nm.2415.

[48] C T Jordan et al. “The interleukin-3 receptor alpha chain is a unique marker for human acute myelogenous leukemia stem cells”. In: *Leukemia* 14.10 (2000), pp. 1777–1784.

[49] Ugo Testa et al. “Elevated expression of IL-3Ra in acute myelogenous leukemia is associated with enhanced blast proliferation, increased cellularity, and poor prognosis”. In: *Blood* 100.8 (2002), pp. 2980–2988.

[50] C. Craddock et al. “Azacitidine fails to eradicate leukemic stem/progenitor cell populations in patients with acute myeloid leukemia and myelodysplasia”. In: *Leukemia* 27.5 (2013), pp. 1028–1036. URL: http://dx.doi.org/10.1038/leu.2012.312.

[51] Charles Craddock et al. “ Outcome of Azacitidine Therapy in Acute Myeloid Leukemia is not Improved by Concurrent Vorinostat Therapy but is Predicted by a Diagnostic Molecular Signature ”. In: *Clinical Cancer Research* (2017), clincanres.1423.2017. URL: http://clincancerres.aacrjournals.org/lookup/doi/10.1158/1078-0432.CCR-17-1423.

[52] Tzu-Chieh Ho et al. "Evolution of acute myelogenous leukemia stem cell properties following treatment and progression." In: *Blood* 128.13 (2016), blood–2016–02–695312. URL: http://www.ncbi.nlm.nih.gov/pubmed/27421961.

[53] Liran I Shlush et al. "Tracing the origins of relapse in acute myeloid leukaemia to stem cells". In: *Nature Publishing Group* 547.7661 (2017), pp. 104–108. URL: http://dx.doi.org/10.1038/nature22993%20https://www.nature.com/nature/journal/vaop/ncurrent/pdf/nature22993.pdf.

[54] Dimitris Karamitros et al. "Single-cell analysis reveals the continuum of human lympho-myeloid progenitor cells article". In: *Nature Immunology* 19.1 (2018), pp. 85–97.

[55] Emmanuelle M. Six et al. "Cytokines and culture medium have a major impact on human in vitro T-cell differentiation". In: *Blood Cells, Molecules, and Diseases* 47.1 (2011), pp. 72–78. URL: http://dx.doi.org/10.1016/j.bcmd.2011.04.001.

[56] Julien Calvo et al. "Assessment of Human Multi-Potent Hematopoietic Stem/Progenitor Cell Potential Using a Single In Vitro Screening System". In: *PLoS ONE* 7.11 (2012), e50495. URL: http://dx.plos.org/10.1371/journal.pone.0050495.

[57] Andreas Reinisch et al. "A humanized bone marrow ossicle xenotransplantation model enables improved engraftment of healthy and leukemic human hematopoietic cells". In: *Nature Medicine* 22.7 (2016), pp. 812–821. arXiv: arXiv:1011.1669v3. URL: http://www.nature.com/doifinder/10.1038/nm.4103.

[58] Petter S. Woll et al. "Myelodysplastic syndromes are propagated by rare and distinct human cancer stem cells in vivo". In: *Cancer Cell* 25.6 (2014), pp. 794–808.

[59] Simone Picelli et al. "Full-length RNA-seq from single cells using Smart-seq2." In: *Nature protocols* 9.1 (Jan. 2014), pp. 171–81. URL: http://www.ncbi.nlm.nih.gov/pubmed/24385147.

[60] Roy Drissen et al. "Distinct myeloid progenitor–differentiation pathways identified through single-cell RNA sequencing". In: *Nature Immunology* 17.April (2016), pp. 1–13. URL: http://www.nature.com/doifinder/10.1038/ni.3412.

[61] Daehwan Kim et al. "TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions." In: *Genome biology* 14.4 (2013), R36. URL: http://genomebiology.com/2013/14/4/R36.

[62] Guoji Guo et al. "Mapping cellular hierarchy by single-cell analysis of the cell surface repertoire." In: *Cell stem cell* 13.4 (Oct. 2013), pp. 492–505. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3845089%7B%5C&%7Dtool=pmcentrez%7B%5C&%7Drendertype=abstract.

[63] R. R. Coifman et al. "Geometric diffusions as a tool for harmonic analysis and structure definition of data: Multiscale methods". In: *Proceedings of the National Academy of Sciences* 102.21 (2005), pp. 7432–7437. arXiv: NIHMS150003. URL: http://www.pnas.org/cgi/doi/10.1073/pnas.0500896102.

[64] Laleh Haghverdi, Florian Buettner, and Fabian J. Theis. "Diffusion maps for high-dimensional single-cell analysis of differentiation data". In: *Bioinformatics* 31.18 (2015), pp. 2989–2998.

[65] Philipp Angerer et al. "Destiny: Diffusion maps for large-scale single-cell data in R". In: *Bioinformatics* 32.8 (2016), pp. 1241–1243.

[66] Thomas D. Wu and Serban Nacu. "Fast and SNP-tolerant detection of complex variants and splicing in short reads". In: *Bioinformatics* 26.7 (2010), pp. 873–881.

[67] Andrew Yates et al. "Ensembl 2016". In: *Nucleic Acids Research* 44.D1 (2016), pp. D710–D716. arXiv: arXiv:1011.1669v3.

[68] Simon Anders, Paul Theodor Pyl, and Wolfgang Huber. "HTSeq-A Python framework to work with high-throughput sequencing data". In: *Bioinformatics* 31.2 (2015), pp. 166–169. arXiv: /biorxiv.org/content/early/2014/08/19/002824 [http:].

[69] Philip Brennecke et al. "Accounting for technical noise in single-cell RNA-seq experiments." In: *Nature methods* 10.11 (2013), pp. 1093–5. URL: http://www.ncbi.nlm.nih.gov/pubmed/24056876.

[70] Aaron T.L. Lun, Karsten Bach, and John C. Marioni. "Pooling across cells to normalize single-cell RNA sequencing data with many zero counts". In: *Genome Biology* 17.1 (2016), pp. 1–14. URL: http://dx.doi.org/10.1186/s13059-016-0947-7.

[71] Anthony M. Bolger, Marc Lohse, and Bjoern Usadel. "Trimmomatic: A flexible trimmer for Illumina sequence data". In: *Bioinformatics* 30.15 (2014), pp. 2114–2120.

[72] Daehwan Kim, Ben Langmead, and Steven L. Salzberg. "HISAT: A fast spliced aligner with low memory requirements". In: *Nature Methods* 12.4 (2015), pp. 357–360. arXiv: 15334406.

[73] Yang Liao, Gordon K. Smyth, and Wei Shi. "FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features". In: *Bioinformatics* 30.7 (2014), pp. 923–930. arXiv: arXiv:1305.3347v2.

[74] Nicolas L Bray et al. "Near-optimal probabilistic RNA-seq quantification". In: *Nature Biotechnology* 34.5 (2016), pp. 525–527. arXiv: 1505.02710. URL: http://www.nature.com/doifinder/10.1038/nbt.3519.

[75] Michael I Love, Wolfgang Huber, and Simon Anders. "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." In: *Genome biology* 15.12 (2014), p. 550. URL: http://genomebiology.biomedcentral.com/articles/10.1186/s13059-014-0550-8.

[76] Hadley Wickham. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2009. URL: http://ggplot2.org.

[77] Uwe Ligges and M Martin. "Scatterplot3d – an R Package for Visualizing Multivariate Data". In: *Journal of Statistical Software* 8.11 (2003), pp. 1–20. URL: http://ggplot2.org.

[78] Tim Wang et al. "Gene Essentiality Profiling Reveals Gene Networks and Synthetic Lethal Interactions with Oncogenic Ras". In: *Cell* 168.5 (2017), 890–903.e15. URL: http://dx.doi.org/10.1016/j.cell.2017.01.013.

[79] Konstantinos Tzelepis et al. "A CRISPR Dropout Screen Identifies Genetic Vulnerabilities and Therapeutic Targets in Acute Myeloid Leukemia". In: *Cell Reports* 17.4 (2016), pp. 1193–1205. URL: http://dx.doi.org/10.1016/j.celrep.2016.09.079.

[80] Ryan J Park et al. "A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors". In: *Nature Genetics* 49.2 (2017), pp. 193–203. URL: http://www.nature.com/doifinder/10.1038/ng.3741.

[81] Zachary Steinhart et al. "Genome-wide CRISPR screens reveal a Wnt–FZD5 signaling circuit as a druggable vulnerability of RNF43-mutant pancreatic tumors". In: *Nature Medicine* 23.1 (2016), pp. 60–68. URL: http://www.nature.com/doifinder/10.1038/nm.4219.

[82] Traver Hart and Jason Moffat. "BAGEL: a computational framework for identifying essential genes from pooled library screens". In: *BMC Bioinformatics* 17.1 (2016), p. 164. arXiv: arXiv:0907.2398v1. URL: http://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-016-1015-8.

[83] Traver Hart et al. "High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities". In: *Cell* 163.6 (2015), pp. 1515–1526. URL: http://dx.doi.org/10.1016/j.cell.2015.11.015.

[84] K Mihara et al. "Development and functional characterization of human bone marrow mesenchymal cells immortalized by enforced expression of telomerase". In: *Br J Haematol* 120.5 (2003), pp. 846–849.

[85] Caroline Pabst et al. "Identification of small molecules that support human leukemia stem cell activity ex vivo". In: *Nature Methods* 11.4 (2014), pp. 436–442. URL: http://www.nature.com/doifinder/10.1038/nmeth.2847.

[86] Jacob T. Jackson et al. "Hhex Regulates Hematopoietic Stem Cell Self-Renewal and Stress Hematopoiesis via Repression of Cdkn2a". In: *Stem Cells* 35 (2017), pp. 1948–1957.

[87] Patrick D. Hsu et al. "DNA targeting specificity of RNA-guided Cas9 nucleases". In: *Nature Biotechnology* 31.9 (2013), pp. 827–832. arXiv: NIHMS150003.

[88] Ophir Shalem et al. "Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells". In: *Science* 343.6166 (2014), pp. 84–87. arXiv: NIHMS150003. URL: http://www.sciencemag.org/lookup/doi/10.1126/science.1247005.

[89] Rasmus O. Bak, Daniel P. Dever, and Matthew H. Porteus. "CRISPR/Cas9 genome editing in human hematopoietic stem cells". In: *Nature Protocols* 13.2 (2018), pp. 358–376.

[90] Timothy Hsiau et al. "Inference of CRISPR Edits from Sanger Trace Data". In: *bioRxiv* (2018), p. 251082. arXiv: /dx.doi.org/10.1101/251082 [http:]. URL: https://www.biorxiv.org/content/early/2018/01/20/251082.full.pdf+html.

[91] Dirk Heckl et al. “Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing”. In: *Nature Biotechnology* 32.9 (2014), pp. 941–946. arXiv: NIHMS150003.

[92] Jaeyop Lee et al. “Restricted dendritic cell and monocyte progenitors in human cord blood and bone marrow”. In: *Journal of Experimental Medicine* (2015).

[93] Anne Galy et al. “Human T, B, natural killer, and dendritic cells arise from a common bone marrow progenitor cell subset”. In: *Immunity* 3 (1995), pp. 459–473.

[94] Noa Novershtern et al. “Densely interconnected transcriptional circuits control cell states in human hematopoiesis.” In: *Cell* 144.2 (Jan. 2011), pp. 296–309. URL: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3049864%7B%5C&%7Dtool=pmcentrez%7B%5C&%7Drendertype=abstract.

[95] Nicola K. Wilson et al. “Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations”. In: *Cell Stem Cell* (2015), pp. 712–724. URL: http://linkinghub.elsevier.com/retrieve/pii/S1934590915001629.

[96] Victoria Moignard et al. “Decoding the regulatory network of early blood development from single-cell gene expression measurements”. In: *Nature Biotechnology* advance on.3 (2015). URL: http://dx.doi.org/10.1038/nbt.3154.

[97] Charles L. Sawyers, Christopher T. Denny, and Owen N. Witte. “Leukemia and the disruption of normal hematopoiesis”. In: *Cell* 64.2 (1991), pp. 337–350.

[98] James D. Griffin and Bob Lowenberg. “Clonogenic Cells in Acute Myeloblastic Leukemia”. In: *Blood* 68.6 (1986), pp. 1185–1195.

[99] Mauri Keinanen et al. “Clonal Chromosomal Abnormalities Showing Multiple-Cell-Lineage Involvement in Actute Myeloid Leukemia”. In: *New England Journal of Medicine* 318 (1988), p. 1153.

[100] David C Taussig et al. “Leukemia initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34 - fraction Leukemia initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34 fraction”. In: 115.10 (2010), pp. 1976–1985.

[101] E Papaemmanuil et al. “Genomic classification and prognosis in acute myeloid leukemia”. In: *N Engl J Med* in press (2016).

[102] P. V. Sanchez et al. “A robust xenotransplantation model for acute myeloid leukemia”. In: *Leukemia* 23.11 (2009), pp. 2109–2117. arXiv: NIHMS150003. URL: http://dx.doi.org/10.1038/leu.2009.143.

[103] Dominique Bonnet. “Acute myeloid leukemia including favorable-risk group samples engraft in NSG mice: just be patient”. In: *Haematologica* 102.5 (2017), pp. 805–806. URL: http://www.haematologica.org/lookup/doi/10.3324/haematol.2017.165159.

[104] Anna M. Paczulla et al. “Long-term observation reveals high-frequency engraftment of human acute myeloid leukemia in immunodeficient mice”. In: *Haematologica* 102.5 (2017), pp. 854–864.

[105] E.D. Ball et al. “Prognostic value of lymphocyte surface markers in acute myeloid leukemia”. In: *Blood* 77.10 (1991), pp. 2242–2250.

[106] John T. Brandt et al. “Aberrant expression of CD19 as a marker of monocytic lineage in acute myelogenous leukemia”. In: *American Journal of Clinical Pathology* 107.3 (1997), pp. 283–291.

[107] Yen Chuan Hsieh et al. “Aberrant Expression of CD19 and CD43 in a Patient With Therapy-Related Acute Myeloid Leukemia and a History of Mantle Cell Lymphoma”. In: *Kaohsiung Journal of Medical Sciences* 25.7 (2009), pp. 389–394. URL: http://dx.doi.org/10.1016/S1607-551X(09)70532-7.

[108] K Walter et al. “Aberrant expression of CD19 in AML with t(8;21) involves a poised chromatin structure and PAX5.” In: *Oncogene* 29.20 (2010), pp. 2927–2937. URL: http://dx.doi.org/10.1038/onc.2010.56.

[109] Gert J. Ossenkoppele, Arjan A. van de Loosdrecht, and Gerrit Jan Schuurhuis. “Review of the relevance of aberrant antigen expression by flow cytometry in myeloid neoplasms”. In: *British Journal of Haematology* 153.4 (2011), pp. 421–436.

[110] Jawad Francis et al. “CD19 expression in acute leukemia is not restricted to the cytogenetically aberrant populations.” In: *Leukemia & lymphoma* 54.7 (2013), pp. 1517–20. URL: http://www.ncbi.nlm.nih.gov/pubmed/23193950.

[111] Klaus H. Metzeler et al. “An 86-probe-set gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia”. In: *Blood* 112.10 (2008), pp. 4193–4201.

[112] Lars Bullinger et al. “Use of Gene-Expression Profiling to Identify Prognostic Subclasses in Adult Acute Myeloid Leukemia”. In: *The New England journal of medicine* 350.16 (2004), pp. 1605–1616.

[113] Peter J M Valk et al. “Prognostically useful gene-expression profiles in acute myeloid leukemia.” In: *The New England journal of medicine* 350.16 (2004), pp. 1617–1628.

[114] Andrew J Gentles, Sylvia K Plevritis, and Patient Page. “Association of a Leukemic Stem Cell Gene in Acute Myeloid Leukemia”. In: *Journal of the American Medical Association* 304.24 (2012).

[115] Luiza Handschuh et al. “Gene expression profiling of acute myeloid leukemia samples from adult patients with AML-M1 and -M2 through boutique microarrays, real-time PCR and droplet digital PCR”. In: *International Journal of Oncology* 52.3 (2018), pp. 656–678.

[116] Mireia Camós et al. “Gene expression profiling of acute myeloid leukemia with translocation t(8;16)(p11;p13) and MYST3-CREBBP rearrangement reveals a distinctive signature with a specific pattern of HOX gene expression”. In: *Cancer Research* 66.14 (2006), pp. 6947–6954.

[117] Torsten Haferlach et al. “Global approach to the diagnosis of leukemia using gene expression profiling”. In: *Gene Expression* 106.4 (2005), pp. 1189–1198.

[118] Lars Bullinger and Peter J.M. Valk. “Gene expression profiling in acute myeloid leukemia”. In: *Journal of Clinical Oncology* 23.26 (2005), pp. 6296–6305.

[119] Derek L. Stirewalt et al. “Identification of Genes with Abnormal Expression Changes in Acute Myeloid Leukemia”. In: *Genes, chromosomes & cancer* 47.April (2008). arXiv: NIHMS150003.

[120] H. Gal et al. “Gene expression profiles of AML derived stem cells; similarity to hematopoietic stem cells”. In: *Leukemia* 20.12 (2006), pp. 2147–2154.

[121] Torsten Haferlach et al. “Clinical utility of microarray-based gene expression profiling in the diagnosis and subclassification of leukemia: Report from the international microarray innovations in leukemia study group”. In: *Journal of Clinical Oncology* 28.15 (2010), pp. 2529–2537.

[122] Bas J. Wouters et al. “Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome”. In: *Blood* 113.13 (2009), pp. 3088–3091.

[123] A. Kohlmann et al. “Gene expression profiling in AML with normal karyotype can predict mutations for molecular markers and allows novel insights into perturbed biological pathways”. In: *Leukemia* 24.6 (2010), pp. 1216–1220.

[124] The Cancer Genome Atlas Research Network. “Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia”. In: *New England Journal of Medicine* 368.22 (2013), pp. 2059–2074. URL: http://www.nejm.org/doi/10.1056/NEJMoa1301689.

[125] Hendrik J M De Jonge et al. “High VEGFC expression is associated with unique gene expression profiles and predicts adverse prognosis in pediatric and adult acute myeloid leukemia”. In: *Blood* 116.10 (2010), pp. 1747–1754.

[126] Tobias Herold et al. “A 29-gene and cytogenetic score for the prediction of resistance to induction treatment in acute myeloid leukemia”. In: *Haematologica* 103.3 (2017), haematol.2017.178442. URL: http://www.haematologica.org/lookup/doi/10.3324/haematol.2017.178442.

[127] Mei Wang et al. “Development and Validation of a Novel RNA Sequencing–Based Prognostic Score for Acute Myeloid Leukemia”. In: *JNCI: Journal of the National Cancer Institute* 110.May (2018), pp. 1–8. URL: https://academic.oup.com/jnci/advance-article/doi/10.1093/jnci/djy021/4916692.

[128] R Majeti et al. “Dysregulated gene expression networks in human acute myelogenous leukemia stem cells”. In: *Proc Natl Acad Sci U S A* 106.9 (2009), pp. 3396–3401. URL: http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed%7B%5C&%7Did=19218430%7B%5C&%7Dretmode=ref%7B%5C&%7Dcmd=prlinks%7B%5C%%7D5Cnpapers3://publication/doi/10.1073/pnas.0900089106.

[129] Caroline Pabst et al. “GPR56 identifies primary human acute myeloid leukemia cells with high repopulating potential in vivo”. In: *Blood* 127.16 (2018), pp. 2018–2028.

[130] Hans Ulrich Klein et al. “Quantitative comparison of microarray experiments with published leukemia related gene expression signatures”. In: *BMC Bioinformatics* 10 (2009), pp. 1–11.

[131] Brady G. Miller and John A. Stamatoyannopoulos. “Integrative meta-analysis of differential gene expression in acute myeloid leukemia”. In: *PLoS ONE* 5.3 (2010), pp. 1–13.

[132] Nicolas Rapin et al. "Comparing cancer vs normal gene expression profiles identifies new disease entities and common transcriptional programs in AML patients Comparing cancer vs normal gene expression pro fi les identi fi es new disease entities and common transcriptional prog". In: 123.6 (2014), pp. 894–904.

[133] Charlotte Soneson, Michael I. Love, and Mark D. Robinson. "Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences." In: *F1000Research* 4.0 (2015), p. 1521. URL: http://f1000research.com/articles/4-1521/v1%7B%5C%%7D5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/26925227%7B%5C%%7D5Cnhttp://f1000research.com/articles/4-1521/v2.

[134] Thomas Smith. *Why you should use alignment-independent quantification for RNA-Seq*. 2016. URL: https://cgatoxford.wordpress.com/2016/08/17/why-you-should-stop-using-featurecounts-htseq-or-cufflinks2-and-start-using-kallisto-salmon-or-sailfish/.

[135] Philip Ewels et al. "MultiQC: Summarize analysis results for multiple tools and samples in a single report". In: *Bioinformatics* 32.19 (2016), pp. 3047–3048.

[136] Simon Anders and Wolfgang Huber. "Differential expression analysis for sequence count data". In: *Genome Biology* 11.10 (2010), R106. arXiv: 1310.0424.

[137] Elisa Laurenti et al. "CDK6 Levels Regulate Quiescence Exit in Human Hematopoietic Stem Cells". In: *Cell Stem Cell* 16.3 (2015), pp. 302–313. URL: http://linkinghub.elsevier.com/retrieve/pii/S1934590915000181.

[138] Maxim V. Kuleshov et al. "Enrichr: a comprehensive gene set enrichment analysis web server 2016 update". In: *Nucleic acids research* 44.W1 (2016), W90–W97.

[139] Karolina Komorowska et al. "Hepatic Leukemia Factor Maintains Quiescence of Hematopoietic Stem Cells and Protects the Stem Cell Pool during Regeneration". In: *Cell Reports* 21.12 (2017), pp. 3514–3523. URL: https://doi.org/10.1016/j.celrep.2017.11.084.

[140] Ashley P. Ng et al. "Erg is required for self-renewal of hematopoietic stem cells during stress hematopoiesis in mice". In: *Blood* 118.9 (2011), pp. 2454–2461.

[141] Zeenath Unnisa et al. "Meis1 preserves hematopoietic stem cells in mice by limiting oxidative stress". In: *Blood* 120.25 (2012), pp. 4973–4981.

[142] Hanno Hock et al. "Tel/Etv6 is an essential and selective regulator of adult hematopoietic stem cell survival". In: *Genes and Development* 18.19 (2004), pp. 2336–2341.

[143] Jon Mar Björnsson et al. "Reduced proliferative capacity of hematopoietic stem cells deficient in reduced proliferative capacity of hematopoietic stem cells deficient in Hoxb3 and Hoxb4". In: *Molecular and Cellular Biology* 23.11 (2003), pp. 3872–3883.

[144] Julie Lessard and Guy Sauvageau. "Bmi-1 determines the proliferative capacity of normal and leukaemic stem cells". In: *Nature* 423.6937 (2003), pp. 255–260.

[145] Alexander Lachmann et al. "Massive mining of publicly available RNA-seq data from human and mouse". In: *Nature Communications* 9.1 (2018). URL: http://dx.doi.org/10.1038/s41467-018-03751-6.

[146] Elahe Mahdipour, Jayne C Charnock, and Kimberly a Mace. “Hoxa3 promotes the differentiation of hematopoietic progenitor cells into proangiogenic Gr-1+ CD11b+ myeloid cells”. In: *Hematology* 117.3 (2011), pp. 815–826.

[147] B. Argiropoulos and R. K. Humphries. “Hox genes in hematopoiesis and leukemogenesis”. In: *Oncogene* 26.47 (2007), pp. 6766–6776.

[148] Christina M. Ferrell et al. “Activation of Stem-Cell Specific Genes by HOXA9 and HOXA10 Homeodomain Proteins in CD34 ⁺ Human Cord Blood Cells”. In: *Stem Cells* 23.5 (2005), pp. 644–655. URL: http://doi.wiley.com/10.1634/stemcells.2004-0198.

[149] Yoshinori Ohno et al. “Hoxa9 Transduction Induces Hematopoietic Stem and Progenitor Cell Activity through Direct Down-Regulation of Geminin Protein”. In: *PLoS ONE* 8.1 (2013).

[150] Mark L. Sandberg et al. “c-Myb and p300 regulate hematopoietic stem cell proliferation and differentiation”. In: *Developmental Cell* 8.2 (2005), pp. 153–166.

[151] Jianhua Yu et al. “TSC-22 contributes to hematopoietic precursor cell proliferation and repopulation and is epigenetically silenced in large granular lymphocyte leukemia”. In: *Blood* 113.22 (2009), pp. 5558–5567.

[152] Bryan L Krock et al. “The aryl hydrocarbon receptor nuclear translocator is an essential regulator of murine hematopoietic stem cell viability”. In: *Hematopoiesis and Stem Cells* 125.21 (2015), pp. 3263–3273.

[153] Parham Solaimani Kartalaei et al. “Whole-transcriptome analysis of endothelial to hematopoietic stem cell transition reveals a requirement for Gpr56 in HSC generation”. In: *The Journal of Experimental Medicine* 212.1 (2015), pp. 93–106. URL: http://www.jem.org/lookup/doi/10.1084/jem.20140767.

[154] Céline Auvray et al. “HOXC4 homeoprotein efficiently expands human hematopoietic stem cells and triggers similar molecular alterations as HOXB4”. In: *Haematologica* 97.2 (2012), pp. 168–178.

[155] Jennifer N Lilla et al. “Reduced mast cell and basophil numbers and function in Cpa3-Cre ; Mcl-1 fl / fl mice Reduced mast cell and basophil numbers and function in Cpa3-Cre ; Mcl-1 fl / fl mice”. In: 118.26 (2011), pp. 6930–6938.

[156] Ellen V Rothenberg. “Transcriptional Control of Early T and B Cell Developmental Choices”. In: *Annual Review of Immunology* 21.32 (2014), pp. 283–321.

[157] Marc Beyer et al. “High-Resolution Transcriptome of Human Macrophages”. In: *PLoS ONE* 7.9 (2012).

[158] Y. Yang et al. “Targeting phospho-MARCKS overcomes drug-resistance and induces antitumor activity in preclinical models of multiple myeloma”. In: *Leukemia* 29.3 (2015), pp. 715–726. URL: http://dx.doi.org/10.1038/leu.2014.255.

[159] Mizuko Tanaka et al. “Co-expression of S100A14 and S100A16 correlates with a poor prognosis in human breast cancer and promotes cancer cell invasion”. In: *BMC Cancer* 15.1 (2015), pp. 1–14.

[160] Xu Sun et al. "S100A16 is a prognostic marker for colorectal cancer". In: *Journal of Surgical Oncology* August 2017 (2017), pp. 275–283. URL: http://doi.wiley.com/10.1002/jso.24822.

[161] Sócrates Avilés-Vázquez et al. "Global gene expression profiles of hematopoietic stem and progenitor cells from patients with chronic myeloid leukemia: the effect of in vitro culture with or without imatinib". In: *Cancer Medicine* 6.12 (2017), pp. 2942–2956.

[162] Guillaume Bossis et al. "The ROS/SUMO Axis Contributes to the Response of Acute Myeloid Leukemia Cells to Chemotherapeutic Drugs". In: *Cell Reports* 7.6 (2014), pp. 1815–1823. URL: http://www.ncbi.nlm.nih.gov/pubmed/24910433%7B%5C%%7D5Cnhttp://linkinghub.elsevier.com/retrieve/pii/S221112471400391X.

[163] Xiaodong Cheng et al. "PKR inhibits the DNA damage response, and is associated with poor survival in AML and accelerated leukemia in NHD13 mice". In: *Blood* 126.13 (2015), pp. 1585–1594.

[164] Francisco M. Vega and Anne J. Ridley. "Rho GTPases in cancer cell biology". In: *FEBS Letters* 582.14 (2008), pp. 2093–2101.

[165] James C Mulloy et al. "Rho GTPases in hematopoiesis and hemopathies Review article Rho GTPases in hematopoiesis and hemopathies". In: *Blood* 115.5 (2013), pp. 936–947.

[166] F. Esteban et al. "A role for substance P in cancer promotion and progression: A mechanism to counteract intracellular death signals following oncogene activation or DNA damage". In: *Cancer and Metastasis Reviews* 25.1 (2006), pp. 137–145.

[167] Cristina Mayordomo et al. "Targeting of substance P induces cancer cell death and decreases the steady state of EGFR and Her2". In: *Journal of Cellular Physiology* 227.4 (2012), pp. 1358–1366.

[168] Bas J Wouters and Ruud Delwel. "Epigenetics and approaches to targeted epigenetic therapy in acute myeloid leukemia". In: *Blood* 127.1 (2015), pp. 42–53.

[169] Nicos A. Nicola and Jeffrey J. Babon. "Leukemia inhibitory factor (LIF)". In: *Cytokine & Growth Factor Reviews* 26.5 (2015), pp. 533–544. arXiv: 15334406. URL: http://linkinghub.elsevier.com/retrieve/pii/S1359610115000490.

[170] a M Douglas et al. "Expression and function of members of the cytokine receptor superfamily on breast cancer cells." In: *Oncogene* 14.6 (1997), pp. 661–9. URL: http://www.ncbi.nlm.nih.gov/pubmed/9038373.

[171] J Lu et al. "CNTF receptor subunit alpha as a marker for glioma tumor-initiating cells and tumor grade: laboratory investigation". In: *Journal of neurosurgery* 117.6 (2012), pp. 1022–1031. URL: http://www.ncbi.nlm.nih.gov/pubmed/23061382.

[172] L. S. Steelman et al. "Contributions of the Raf/MEK/ERK, PI3K/PTEN/Akt/mTOR and Jak/STAT pathways to leukemia". In: *Leukemia* 22.4 (2008), pp. 686–707.

[173] Xiaomin Wang et al. "mTORC signaling in hematopoiesis". In: *International Journal of Hematology* 103.5 (2016), pp. 510–518.

[174] Yanan Gao et al. “Rheb1 promotes tumor progression through mTORC1 in MLL-AF9- initiated murine acute myeloid leukemia”. In: *Journal of Hematology and Oncology* 9.1 (2016), pp. 1–14. URL: http://dx.doi.org/10.1186/s13045-016-0264-3.

[175] Francesca Zanconato, Michelangelo Cordenonsi, and Stefano Piccolo. “YAP/TAZ at the Roots of Cancer”. In: *Cancer Cell* 29.6 (2016), pp. 783–803. URL: http://dx.doi.org/10.1016/j.ccell.2016.05.005.

[176] Houda Alachkar et al. “Preclinical efficacy of maternal embryonic leucine-zipper kinase (MELK) inhibition in acute myeloid leukemia”. In: *Oncotarget* 5.23 (2014). URL: http://www.oncotarget.com/fulltext/2642.

[177] Mathew C. Easterday et al. “Neural progenitor genes: Germinal zone expression and analysis of genetic overlap in stem cell populations”. In: *Developmental Biology* 264.2 (2003), pp. 309–322.

[178] Martin Jinek et al. “A Programmable Dual-RNA – Guided”. In: *Science* 337.August (2012), pp. 816–822. arXiv: 38.

[179] Seung Woo Cho et al. “Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease”. In: *Nature Biotechnology* 31.3 (2013), pp. 230–232. URL: http://dx.doi.org/10.1038/nbt.2507.

[180] Le Cong et al. “Multiplex Genome Engineering Using CRISPR/Cas Systems”. In: *Science* 339 (2013), p. 819. arXiv: arXiv:1011.1669v3. URL: http://dx.doi.org/10.1016/j.jeem.2011.07.001%7B%5C%%7D0Ahttp://www.ncbi.nlm.nih.gov/pubmed/18801999%7B%5C%%7D0Ahttp://le.uwpress.org/cgi/content/abstract/78/4/522.

[181] Prashant Mali et al. “RNA-Guided Human Genome Engineering via Cas9”. In: *Science* 339.February (2013), pp. 823–827. arXiv: arXiv:1011.1669v3.

[182] Michael C. Gundry et al. “Highly Efficient Genome Editing of Murine and Human Hematopoietic Progenitor Cells by CRISPR/Cas9”. In: *Cell Reports* 17.5 (2016), pp. 1453–1461. arXiv: 15334406. URL: http://dx.doi.org/10.1016/j.celrep.2016.09.092.

[183] Erika Zonari et al. “Efficient Ex Vivo Engineering and Expansion of Highly Purified Human Hematopoietic Stem and Progenitor Cell Populations for Gene Therapy”. In: *Stem Cell Reports* 8.4 (2017), pp. 977–990.

[184] Michelle Millington et al. “Towards a clinically relevant lentiviral transduction protocol for primary human CD34+ hematopoietic stem/progenitor cells”. In: *PLoS ONE* 4.7 (2009).

[185] N Uchida et al. “Optimal conditions for lentiviral transduction of engrafting human CD34 + cells”. In: *Gene Therapy* 18.11 (2011), pp. 1078–1086. URL: http://dx.doi.org/10.1038/gt.2011.63.

[186] Ayal Hendel et al. “Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells”. In: *Nature Biotechnology* 33.9 (2015), pp. 985–989. arXiv: 15334406.

[187] Pankaj K. Mandal et al. "Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9". In: *Cell Stem Cell* 15.5 (2014), pp. 643–652. arXiv: NIHMS150003. URL: http://dx.doi.org/10.1016/j.stem.2014.10.004.

[188] Zuzana Tothova et al. "Multiplex CRISPR/Cas9-Based Genome Editing in Human Hematopoietic Stem Cells Models Clonal Hematopoiesis and Myeloid Neoplasia". In: *Cell Stem Cell* 21.4 (2017), 547–555.e8. URL: https://doi.org/10.1016/j.stem.2017.07.015.

[189] Lin Ye et al. "Genome editing using CRISPR-Cas9 to create the HPFH genotype in HSPCs: An approach for treating sickle cell disease and β-thalassemia". In: *Proceedings of the National Academy of Sciences* 113.38 (2016), pp. 10661–10665. URL: http://www.pnas.org/lookup/doi/10.1073/pnas.1612075113.

[190] Caroline Y. Kuo et al. "Site-Specific Gene Editing of Human Hematopoietic Stem Cells for X-Linked Hyper-IgM Syndrome". In: *Cell Reports* 23.9 (2018), pp. 2606–2616. URL: https://doi.org/10.1016/j.celrep.2018.04.103.

[191] M. A. DeWitt et al. "Selection-free genome editing of the sickle mutation in human adult hematopoietic stem/progenitor cells". In: *Science Translational Medicine* 8.360 (2016), 360ra134–360ra134. arXiv: 15334406. URL: http://stm.sciencemag.org/cgi/doi/10.1126/scitranslmed.aaf9336.

[192] Oliver Brabetz et al. "RNA-Guided CRISPR-Cas9 System-Mediated Engineering of Acute Myeloid Leukemia Mutations". In: *Molecular Therapy - Nucleic Acids* 6.March (2017), pp. 243–248. URL: http://dx.doi.org/10.1016/j.omtn.2016.12.012.

[193] Panpan Hou et al. "A genome-wide CRISPR screen identifies genes critical for resistance to FLT3 inhibitor AC220". In: *Cancer Research* 77.16 (2017), pp. 4402–4413.

[194] Wei Li et al. "MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens". In: *Genome biology* 15.12 (2014), p. 554. arXiv: arXiv:1011.1669v3.

[195] Leïla Perié et al. "The Branching Point in Erythro-Myeloid Differentiation". In: *Cell* 163.7 (2015), pp. 1655–1662.

[196] Katrin Busch et al. "Fundamental properties of unperturbed haematopoiesis from stem cells in vivo". In: *Nature* (2015). URL: http://www.nature.com/doifinder/10.1038/nature14242.

[197] Jianlong Sun et al. "Clonal dynamics of native haematopoiesis". In: *Nature* 514.7522 (2014), pp. 322–327. URL: http://dx.doi.org/10.1038/nature13824.

[198] Henry Lee-Six et al. "Barcoding of human haematopoietic stem and progenitor cells by whole genome sequencing reveals clonal dynamics of haematopoiesis". In: *Experimental Hematology*. Vol. 53. Elsevier, 2017, S42. URL: http://dx.doi.org/10.1016/j.exphem.2017.06.044.

[199] André Görgens et al. "Revision of the human hematopoietic tree: granulocyte subtypes derive from distinct hematopoietic lineages." In: *Cell reports* 3.5 (May 2013), pp. 1539–52. URL: http://www.ncbi.nlm.nih.gov/pubmed/23707063.

[200] Jaeyop Lee et al. “Lineage specification of human dendritic cells is marked by IRF8 expression in hematopoietic stem cells and multipotent progenitors”. In: *Nat Immunol* advance on.8 (2017). URL: http://dx.doi.org/10.1038/ni.3789%7B%5C%%7D5Cnhttp://10.0.4.14/ni.3789%7B%5C%%7D5Cnhttp://www.nature.com/ni/journal/vaop/ncurrent/abs/ni.3789.html%7B%5C#%7Dsupplementary-information.

[201] Joby L. McKenzie et al. “Reversible cell surface expression of CD38 on CD34-positive human hematopoietic repopulating cells”. In: *Experimental Hematology* 35.9 (2007), pp. 1429–1436.

[202] MA Rieger, PS Hoppe, and BM Smejkal. “Hematopoietic cytokines can instruct lineage choice”. In: *Science* 1092.2007 (2009), pp. 2007–2009. URL: http://www.sciencemag.org/content/325/5937/217.short.

[203] Timm Schroeder. “Long-term single-cell imaging of mammalian stem cells.” In: *Nature methods* 8.4 Suppl (Apr. 2011), S30–5. URL: http://www.ncbi.nlm.nih.gov/pubmed/21451514.

[204] Richard W J Groen et al. “Reconstructing the human hematopoietic niche in immunodeficient mice : opportunities for studying primary multiple myeloma Reconstructing the human hematopoietic niche in immunodeficient mice : opportunities for studying primary multiple myeloma”. In: 120.3 (2012), pp. 9–17.

[205] Diana Passaro et al. “Bioengineering of Humanized Bone Marrow Microenvironments in Mouse and Their Visualization by Live Imaging”. In: *Journal of Visualized Experiments* 126 (2017), pp. 1–13. URL: https://www.jove.com/video/55914/bioengineering-humanized-bone-marrow-microenvironments-mouse-their.

[206] Ander Abarrategi et al. “Versatile humanized niche model enables study of normal and malignant human hematopoiesis”. In: *Journal of Clinical Investigation* 127.2 (2017), pp. 543–548.

[207] Antonella Antonelli et al. “Establishing human leukemia xenograft mouse models by implanting human bone marrow – like scaffold-based niches”. In: *Blood* 128.25 (2017), pp. 2949–2960.

[208] M. Wunderlich et al. “AML xenograft efficiency is significantly improved in NOD/SCID-IL2RG mice constitutively expressing human SCF, GM-CSF and IL-3”. In: *Leukemia* 24.10 (2010), pp. 1785–1788. arXiv: 15334406.

[209] Jana M. Ellegast et al. “inv(16) and NPM1mut AMLs engraft human cytokine knock-in mice”. In: *Blood* 128.17 (2016), pp. 2130–2135.

[210] Anthony Rongvaux et al. “Development and function of human innate immune cells in a humanized mouse model”. In: *Nature Biotechnology* 32.4 (2014), pp. 364–372. arXiv: NIHMS150003.

[211] Sean C Bendall et al. “Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum.” In: *Science (New York, N.Y.)* 332.6030 (2011), pp. 687–696.

[212] Jacob H. Levine et al. “Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis”. In: *Cell* (2015), pp. 1–14. URL: http://linkinghub.elsevier.com/retrieve/pii/S0092867415006376.

[213] Benedict Yan et al. “Single-cell genomic profiling of acute myeloid leukemia for clinical use: A pilot study”. In: *Oncology Letters* 13.3 (2017), pp. 1625–1630.

[214] Grace X.Y. Zheng et al. “Massively parallel digital transcriptional profiling of single cells”. In: *Nature Communications* 8 (2017), pp. 1–12. arXiv: 065912. URL: http://dx.doi.org/10.1038/ncomms14049.

[215] Pietro Genovese et al. “Targeted genome editing in human repopulating haematopoietic stem cells”. In: *Nature* 510.7504 (2014), pp. 235–240. URL: http://dx.doi.org/10.1038/nature13420.

[216] L E Ailles, B Gerhard, and D E Hogge. “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.” In: *Blood* 90.7 (1997), pp. 2555–2564.

[217] Sawa Ito et al. “Long term maintenance of myeloid leukemic stem cells cultured with unrelated human mesenchymal stromal cells”. In: *Stem Cell Research* 14.1 (2015), pp. 95–104. URL: http://dx.doi.org/10.1016/j.scr.2014.11.007.

[218] Emmanuel Griessinger et al. “A Niche-Like Culture System Allowing the Maintenance of Primary Human Acute Myeloid Leukemia-Initiating Cells: A New Tool to Decipher Their Chemoresistance and Self-Renewal Mechanisms”. In: *STEM CELLS Translational Medicine* 3.4 (2014), pp. 520–529. URL: http://doi.wiley.com/10.5966/sctm.2013-0166.

[219] Djoke van Gosliga et al. “Establishing long-term cultures with self-renewing acute myeloid leukemia stem/progenitor cells”. In: *Experimental Hematology* 35.10 (2007), pp. 1538–1549.

[220] Jeffery M. Klco et al. “Genomic impact of transient low-dose decitabine treatment on primary AMLcells”. In: *Blood* 121.9 (2013), pp. 1633–1643.

[221] Ujunwa C. Okoye-Okafor et al. “New IDH1 mutant inhibitors for treatment of acute myeloid leukemia”. In: *Nature Chemical Biology* 11.11 (2015), pp. 878–886. arXiv: 15334406.

[222] Hector Mayani, Eugenia Flores-Figueroa, and Antonieta Chávez-González. “In vitro biology of human myeloid leukemia”. In: *Leukemia Research* 33.5 (2009), pp. 624–637.

[223] I. Fares et al. “Pyrimidoindole derivatives are agonists of human hematopoietic stem cell self-renewal”. In: *Science* 345.6203 (2014), pp. 1509–1512. URL: http://www.sciencemag.org/cgi/doi/10.1126/science.1256337.

[224] E. Biagi et al. “Efficient lentiviral transduction of primary human acute myelogenous and lymphoblastic leukemia cells”. In: *Haematologica* 86.1 (2001), pp. 13–16.

[225] S. Durual et al. “Lentiviral PU.1 overexpression restores differentiation in myeloid leukemic blasts”. In: *Leukemia* 21.5 (2007), pp. 1050–1059.

[226] Hein Schepers et al. “Reintroduction of C/EBPa in leukemic CD34+ stem/progenitor cells impairs self-renewal and partially restores myelopoiesis”. In: *Blood* 110 (2007), pp. 1317–1325.