Novel insights into megakaryopoiesis, thrombopoiesis and acute coronary thrombosis: transcriptome profiling of the haematopoietic stem cell, megakaryocyte and platelet

Fizzah Aziz Choudry

This dissertation is submitted to the University of Cambridge for the degree of Doctor of Philosophy

September 2018

Hughes Hall, Cambridge

SUMMARY

Novel insights into megakaryopoiesis, thrombopoiesis and acute coronary thrombosis: transcriptome profiling of the haematopoietic stem cell, megakaryocyte and platelet

Fizzah Aziz Choudry

The aim of this project was to investigate the transcriptome of human haematopoietic stem cells (HSCs), megakaryocytes and platelets to gain insights into steady state and accelerated thrombopoiesis that occurs in states of haemostatic demand and in thrombosis by applying these findings to the pathological setting of acute coronary thrombosis.

To investigate transcriptional heterogeneity within the human HSC population, single cell RNA sequencing was performed in human bone marrow HSCs. Transcriptionally distinct subpopulations were identified including two megakaryocyte biased subsets with potentially differing functional relevance. Both populations expressed megakaryocyte specific transcripts, one of which also co-expressed common myeloid and megakaryocyte-erythroid progenitor transcripts while the other did not.

This study represents the first interrogation of the human bone marrow megakaryocyte transcriptome. Cells were collected from healthy human bone marrow and analysed by low input and single cell RNA sequencing. To identify novel drivers of megakaryocyte maturation, the human bone marrow megakaryocyte transcriptome was compared to that of megakaryocytes cultured from human CD34+ cells, a process known to generate immature megakaryocytes. Transcriptional signatures associated with increasing megakaryocyte ploidy were then investigated. Increasing megakaryocyte ploidy level was found to be associated with an upregulation of transcripts involved in translation and protein processing as well as expression of a number of transmembrane receptors which might have functional relevance.

Finally, the pathological setting of acute coronary thrombosis was used as a model for accelerated thrombopoiesis. Megakaryocyte and platelet transcriptomes were compared between patients with acute myocardial infarction (AMI) as well as severe coronary disease and a control group. The transcriptional signature relating to disease compared to control in megakaryocytes included upregulation of platelet activation related transcripts in megakaryocytes isolated from patients with AMI and severe coronary artery disease.

DECLARATION

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared below and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared below and specified in the text

It does not exceed the prescribed word limit for the relevant Degree Committee.

The following collaborations have contributed to this work:

1. Nbeal2$^{+/-}$ mice used for megakaryocyte viability experiments were generated by the Wellcome Trust Sanger Institute and provided by Dr C Ghevaert's group, Department of Haematology, University of Cambridge, UK.
2. NOD/LtSz-scidIL2Rg^{null} mice (Jackson Laboratory) used for xenotransplant experiments were provided by Dr E Laurenti's group, Department of Haematology, University of Cambridge, UK. All xenotransplantation work was performed in collaboration with Dr E Laurenti's group.
3. Platelet activation and function work was performed in collaboration with Dr K Downes' group, Department of Haematology, University of Cambridge, UK.
4. Statistical analysis of RNA sequencing data was performed by Dr F Bagger, European Bioinformatics Institute, Hinxton, UK, and Dr R Petersen and Dr L Grassi, Department of Haematology, University of Cambridge, UK.

Supervisors:
Willem Ouwehand (University of Cambridge)
Anthony Mathur (Queen Mary, University of London)

ACKNOWLEDGEMENTS

I would like to thank my supervisors Willem Ouwehand, Anthony Mathur and John Martin who not only inspired me to take on this exciting project in the first place but who have provided ongoing guidance and support throughout. Particular thanks must go to Mattia Frontini for the motivation, intellectual stimulation, continued encouragement and technical advice without which this work would not have been possible.

I would also like to thank Iain Macaulay for taking me under his wing at the Sanger Institute, teaching me about the sequencing world and helping me take on the challenge of megakaryocyte single cell RNA-seq. Thanks also goes to Frederik Bagger for his tireless help with RNA-seq analysis. I must also thank all the members of Willem's group and collaborators who have contributed time, materials and ideas to help with this work, in particular, Samantha Farrow, Kate Downes and Elisa Laurenti. I would like to thank Rakesh Uppal and the rest of the Cardiothoracic team at Barts for welcoming me to theatres so often for samples and to the Barts Research team for their support in my endeavours. And of course thanks goes to the patients and volunteers who consented to be in this study.

A special thanks goes to my family who have lived every bit of this long journey with me. The long hours, the uncertainty of experiments, the endless writing and the time away from the world. They have looked after me and supported me in every way from the very beginning and this is as much their thesis as it is mine. Thank you for understanding, being patient and doing this with me.

Finally, I would like to thank the Medical Research Council for their financial support.

ABBREVIATONS AND CONVENTIONS

Abbreviations

ACS	Acute coronary syndrome
ADP	Adenosine diphosphate
AMI	Acute myocardial infarction
APC	Allophycocyanin
B2M	B2 microglobulin
Bp	Base pairs
BV	Brilliant violet
CAD	Coronary artery disease
CD	Cluster of differentiation
cDNA	Copy deoxyribonucleic acid
CFU	Colony-forming unit
CLP	Common lymphoid progenitor
CMP	Common myeloid progenitor
CRP-XL	Crosslinked collagen related peptide
Cy	Cyanine
DAPI	4',6-Diamidino-2-Phenylindole
DNA	Deoxyribonucleic acid
ECG	Electrocardiogram
ERCC	External RNA Controls Consortium
FACS	Fluorescence activated cell sorting
FITC	Fluorescein isothyocyanate
GAPDH	Glyceraldehyde-3-phosphate dehydrogenase
GFP	Green fluorescent protein
GMP	Granulocyte-macrophage progenitor
GO	Gene ontology
GP	Glycoprotein
G&T-seq	Genome and transcriptome sequencing
GWAS	Genome-wide association study
Hb	Haemoglobin
HGNC	HUGO gene nomenclature committee
HSA	Human serum albumin
HSC	Haematopoietic stem cell
IPF	Immature platelet fraction
iPSC	Induced pluripotent stem cells
LMPP	Lymphoid primed multipotent progenitor
LV	Left ventricle

MegE	Megakaryocyte-Erythroid lineage
MEP	Megakaryocyte-erythroid progenitor
MK	Megakaryocyte
MPL	Myeloproliferative leukaemia protein
MPP	Multipotent progenitor
MPV	Mean platelet volume
mRNA	Messenger ribonucleic acid
NSG	NOD scid gamma
NSTEMI	Non ST-elevation myocardial infarction
PBS	Phosphate buffered saline
PCA	Principal component analysis
PCT	Plateletcrit
PDW	Platelet distribution width
PE	Phycoerythrin
PerCP	Peridinin-Chlorophyll-protein
P-LCR	Platelet large cell ratio
Poly(A)	Poly-adenylated
Plt	Platelet
PLT	Platelet count
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
rRNA	Ribosomal ribonucleic acid
RT-qPCR	Real time quantitative polymerase chain reaction
STEMI	ST-elevation myocardial infarction
SVM	Support vector machine
TPO	Thrombopoietin
TRAP-6	Thrombin receptor activating peptide - 6
WC	White cell count

Gene, transcript and protein nomenclature

Where possible the HUGO Gene Nomenclature Committee (HGNC) gene symbols and guidelines have been used to describe genes and gene products.

Human gene and transcript symbols are capitalised and italicised (e.g., *ITGA2B*) and proteins are represented in standard fonts (e.g., TPO).

TABLE OF CONTENTS

SUMMARY……I

DECLARATION……II

ACKNOWLEDGEMENTS……III

ABBREVIATIONS AND CONVENTIONS……IV

Abbreviations……IV

Gene, transcript and protein nomenclature……V

TABLE OF CONTENTS…… VI

CHAPTER 1……1

INTRODUCTION

1.1 Megakaryopoiesis……1

1.1.1 Haematopoietic stem cells…… 1

1.1.2 Early megakaryocyte lineage commitment from the HSC……... 3

1.1.3 Megakaryocyte-biased HSC population…… 4

1.1.4 Similarities between HSCs and megakaryocytes and interaction within the stem cell niche……..6

1.1.5 Terminal megakaryopoiesis……..9

1.1.6 Polyploidisation by endomitosis……... 9

1.1.7 Cytoplasmic maturation…… 11

1.1.8 Platelet formation…… 11

1.1.9 Regulation of megakaryocyte maturation……... 13

1.1.9.1 Cytokines…… 13

1.1.9.2 Transcription factors…… 14

1.1.10 Megakaryocyte maturation *in vitro*…… 15

1.1.11 The megakaryocyte transcriptome……15

1.2 Platelet biology…… 19

1.2.1 Platelet function……19

1.2.2 Platelet physical characeristics…… 20

1.2.2.1 Mean platelet volume…… 20

1.2.2.2 Immature platelet fraction…… 22

1.2.3 The platelet transcriptome……22

1.3 Megakaryocyte-platelet axis in states of increased haemostatic demand……25

1.4 Megakaryocyte-platelet axis in acute coromary syndromes…… 30

1.4.1 Acute coronary syndromes…… 30

1.4.2 Platelet and megakaryocyte characteristics in ACS....30
1.4.3 Alteration of megakaryocyte-platelet activity in ACS....32
1.4.4 Megakaryocyte-platelet transcriptome studies in ACS....33
1.5 Application of next generation sequencing in transcriptome analysis....34
1.6 Project aims....35

CHAPTER 2....37
MATERIALS AND METHODS
2.1 Sample collection....37
2.1.1 Patient recruitment....37
2.1.2 Adult blood for haematological and platelet measurement....38
2.1.3 Adult bone marrow for megakaryocyte and HSC isolation....38
2.1.3.1 Optimised method....38
2.1.3.2 Other trialled methods....39
2.1.3.3 Human bone marrow CD34+ cells for HSC isolation....39
2.1.4 Adult blood for CD34+ isolation....39
2.1.5 Animals used for megakaryocyte viability experiments....39
2.1.6 Animals used for xenotransplantation studies....40
2.2 Whole blood measurements....40
2.2.1 Platelet physical parameters....40
2.2.2 Platelet function....40
2.3 Cell isolation....41
2.3.1 Isolation of platelets from peripheral blood....41
2.3.2 Isolation of mononuclear cells from peripheral blood....42
2.3.3 Isolation of megakaryocytes from human bone marrow....42
2.3.3.1 Optimisation method: Isolation of megakaryocytes by FACS..42
2.3.3.2 Other trialled methods....43
2.3.3.3 Megakaryocyte imaging....44
2.3.3.4 Megakaryocyte viability experiments using mouse bone marrow....45
2.3.4 Isolation of HSCs from human bone marrow....46
2.3.4.1 HSCs from clinical human bone marrow samples....46
2.3.4.2 HSCs from purchased human bone marrow CD34+ cells (Lonza)....46
2.3.5 Cultured cells....47
2.3.5.1 K562 culture....47
2.3.5.2 Megakaryocyte culture from adult CD34+ cells....47
2.3.6 RNA extraction....47
2.4 Analysis of gene expression....48
2.4.1 Poly(A) RNA-seq library preparation for bulk cultured megakaryocytes...48

2.4.2 Poly(A) RNA-seq library preparation for Smart-seq 2..... 49
2.4.3 Poly(A) RNA-seq library preparation for G&T-seq.....50
2.4.4 Ribosomal RNA depleted RNA-seq library preparation for platelets.....51
2.4.5 RNA-sequencing..... 52
2.4.6 Taqman Real-time qPCR..... 52
2.5 Xenotransplantation..... 53
2.6 Clinical statistics.....53
2.7 RNA-seq analysis..... 54
2.7.1 Single cell RNA-seq analysis: HSCs..... 54
2.7.2 Single cell RNA-seq analysis: Megakaryocytes..... 55
2.7.3 Differential expression analysis: megakaryocytes and platelets..... 56
2.7.4 Functional gene list analysis..... 56

CHAPTER 3..... 57
Single cell RNA-seq reveals transcriptional variation and megakaryocyte biased subsets within the human bone marrow HSC compartment
3.1 Preface……57
3.2 Results……59
3.2.1 HSC phenotype……..... 59
3.2.2 Single cell RNA-seq library preparation:quality measures…….. 61
3.2.2.1 Choice of reverse transcriptase……61
3.2.2.2 Expression of GAPDH as a house-keeping gene..... 61
3.2.2.3 Cell selection for RNA sequencing.....61
3.2.2.4 Correlation between surface and transcript expression of VWF by qPCR..... 62
3.2.2.5 Batch effects.....62
3.2.3 Single cell RNA sequencing data: Quality measures.....67
3.2.3.1 Quality control data.....67
3.2.3.2 Cell filtering.....67
3.2.3.3 Gene filtering..... 68
3.2.4 Gene expression analysis.....74
3.2.4.1 Gene expression profiling reveals 5 clusters within the HSC population.....79
3.2.4.2 Identification of marker genes..... 79
3.2.4.3 Gene ontology of cluster marker genes..... 81
3.2.4.4 Expression of VWF.....81
3.2.5 Flow cytometric index sorting analysis.....95
3.2.6 Functional validation of HSC reconstitution capacity.....98
3.3 Discussion.....99

CHAPTER 4.....109

The Megakaryocyte Transcriptome

4.1 Preface.....109

4.2 Results.....110

4.2.1 Protocol optimisation for analysis of megakaryocyte transcriptome.....110

4.2.1.1 Optimisation of megakaryocyte isolation.....110

4.2.1.2 Megakaryocytes take up live/dead stains.....113

4.2.1.3 Use of G&T-seq protocol for RNA-seq library preparation.....113

4.2.1.4 RNA sequencing data: Quality measures.....118

4.2.1.5 Sequencing approach for megakaryocytes.....120

4.2.2 Defining the primary human megakaryocyte transcriptome.....128

4.2.2.1 Sample preparation and RNA-seq library preparation.....128

4.2.2.2 RNA sequencing data: Quality measures.....129

4.2.2.3 Gene expression analysis.....134

4.2.3 Ploidy associated megakaryocyte transcriptional signatures.....150

4.2.3.1 Sample preparation and RNA-seq library preparation.....150

4.2.3.2 RNA sequencing data: Quality measures.....152

4.2.3.3 Gene expression analysis: Megakaryocyte pools.....156

4.2.3.4 Differential expression analysis: Megakaryocyte pools.....163

4.2.3.5 Identification of transcripts encoding megakaryocyte transmembrane proteins.....164

4.2.3.6 Gene expression analysis: Megakaryocyte single cell.....170

4.2.3.7 Expression of megakaryocyte ploidy signatures in HSC clusters.....174

4.3 Discussion.....176

CHAPTER 5.....186

Platelet and megakaryocytes in patients with myocardial infarction

5.1 Preface.....186

5.2 Results: Megakaryocyte and platelet phenotype.....187

5.2.1 Patient characteristics.....187

5.2.2 Platelet volume and count.....188

5.2.3 Platelet activity.....193

5.2.3.1 Platelet activation.....193

5.2.3.2 Platelet function.....195

5.2.4 Megakaryocyte activity.....200

5.2.4.1 Platelet production.....200

5.2.4.2 Megakaryocyte frequency and ploidy.....201

5.3 Results: Disease associated transcriptional signatures.....203

5.3.1 Disease associated megakaryocyte transcriptional signatures.....203
5.3.1.1 Patient characteristics..... 203
5.3.1.2 RNA-seq library preparation: sample selection..... 203
5.3.1.3 RNA sequencing data: Quality measures.....207
5.3.1.4 Gene expression analysis..... 210
5.3.1.5 Expression of megakaryocyte disease signatures in HSC clusters..... 217
5.3.2 Disease associated platelet transcriptional signatures..... 219
5.3.2.1 Patient characteristics..... 219
5.3.2.2 RNA-seq library preparation.....219
5.3.2.3 RNA sequencing data: Quality measures..... 222
5.3.2.4 Gene expression analysis..... 225
5.4 Discussion.....232

CHAPTER 6.....239

Conclusions

6.1 The first demonstration of transcriptionally distinct subpopulations within the human bone marrow HSC compartment, using single cell RNA-seq.....239
6.2 The first interrogation of the human bone marrow megakaryocyte transcriptome demonstrating regulation of gene expression according to ploidy level..... 240
6.3 The first comparison of megakaryocyte transcriptome data in patients with AMI and severe coronary disease compared with a control group.....242
6.4 Summary of future work.....243

REFERENCES..... 244

APPENDICES.....279
Appendix 1: Study design……..279
Appendix 2: Log of patients recruited and sample use……280
Appendix 3: Chapter 3 supplementary data……CD
Appendix 4: Chapter 4 supplementary data……CD
Appendix 5: Chapter 5 supplementary data……CD

CHAPTER 1
Introduction

Megakaryocytes are highly specialised haematopoietic progenitor cells with the primary function of producing and releasing platelets into the circulation, a process known as thrombopoiesis. Like all haemopoietic cells, megakaryocytes derive from the haematopoietic stem cell (HSC) and growing evidence suggests the existence of a direct differentiation route from the HSC, together with the classically described route through a number of progressively committed progenitors (Notta et al. 2016). Megakaryocytes comprise <0.01% of the total number of nucleated cells in the bone marrow (Harker and Finch 1969). They are unique among mammalian cells in that they are the largest cell in the bone marrow with diameters up to 150μm and undergo multiple genomic DNA replications without cell division, known as endomitosis, resulting in cells with an enlarged cytoplasm and ploidy levels that can reach 128N (Odell and Jackson 1968; Tomer, Harker and Burstein 1988; Tomer et al. 1989). Platelet production can occur at any level of polyploidisation with each megakaryocyte releasing up to 6,000 platelets into the circulation (Harker and Finch 1969).

Platelets are small circulating anucleate cytoplasmic units (2-5 μm) that play a pivotal role in haemostasis. Their primary function is to maintain vascular integrity by sensing and responding to endothelial damage by activation, aggregation and thrombus formation and as such are important contributors to atherothrombotic disease (Monaco, Mathur, and Martin 2005).

The highly complex megakaryocyte-platelet haemostatic system is unique to mammals, having no precedent in the lower orders where nucleated thrombocytes circulate and control haemostasis (Ratnoff 1987). The benefit and evolutionary advantage of such a complex haemostatic system, and its role in health and disease remains unclear.

1.1 Megakaryopoiesis

1.1.1 Haematopoietic stem cells

HSCs are an extremely rare cell population that resides within a highly regulated stem cell niche in the bone marrow; although HSCs can also be isolated from foetal liver, cord blood and adult peripheral blood (Bluteau et al. 2013). HSCs sit at the apex of a developmental hierarchy, whereby they undergo long-term self-renewal, as well as, generate ~10^{11} new cells every day of all blood cell types (Orkin and Zon 2008). Current knowledge of HSCs and haematopoiesis originates predominantly from mouse and is a direct result of distinguishing cells by their surface markers, coupled with either *in vitro* clonal assays or functional, transplant-based tracking of their repopulation activity (Osawa et al. 1996). The biology of

HSCs, particularly in humans, remains poorly understood because of their rarity and due to the difficulty in obtaining pure populations in high numbers for molecular analysis. While the HSC population in the mouse is thought to reside within the $Lin^{-}Sca\text{-}1^{+}c\text{-}Kit^{hi}(LSK)CD34^{-}Flt3^{-}$ (±$CD48^{-}CD150^{+}$) subset (Adolfsson et al. 2001; Kent et al. 2009; Weksberg et al. 2008), human HSCs are defined by surface marker expression of $Lin^{-}CD34^{+}CD38^{-}CD45RA^{-}CD90^{+}CD49f^{+}$. However, only 1 in 10 of these cells will go on to give long term repopulation in serial transplantation experiments, which are considered the gold standard to prove self-renewal and ability to form all blood lineages (Notta et al. 2011).

HSCs were originally thought to be a homogeneous population, however heterogeneity of cellular states within the HSC compartment has in fact been demonstrated (Guo et al. 2013; Muller-Sieburg et al. 2012; Moignard et al. 2013). Recent studies have shown that individual HSCs exhibit different reconstitution patterns with distinct lineage bias and long term self-renewal that are stably inherited by their HSC offspring (Muller-Sieburg et al. 2012; Sieburg et al. 2006; Dykstra et al. 2007). HSC compartments have been defined in terms of their self-renewal ability into short term (more abundant) and long term (rarer) HSCs, ST-HSCs and LT-HSCs respectively (Muller-Sieburg et al. 2004; Reya et al. 2001; Weissman, Anderson, and Gage 2001). Novel technologies have now been developed that use in situ inducible labelling techniques, to enable the study of physiological haematopoiesis in healthy bone marrow in physiological settings, rather than in the inflammatory environment of the transplantation setting. These have demonstrated that traditionally defined ST-HSCs and multipotent progenitors together make up a reservoir of thousands of multipotent progenitor clones that maintain steady-state haematopoiesis, with the rarer LT-HSCs having only a limited contribution (Busch et al. 2015; Sun et al. 2014). HSCs have also been distinguished by their lineage differentiation potential (Müller-Sieburg et al. 2002; Muller-Sieburg et al. 2004, 2012; Dykstra et al. 2007) into: myeloid-biased, lymphoid-biased or balanced (Muller-Sieburg et al. 2004, 2012) or as lymphoid-deficient (α), myeloid-deficient (γ and δ) and balanced (β) (Dykstra et al. 2007). There is also support for myeloid-biased and lymphoid-biased HSC subsets in humans (W. W. Pang et al. 2011; Cavazzana-Calvo et al. 2011). Furthermore, there is evidence to show an increase in myeloid biased HSCs with age, which is thought to be epigenetically regulated (W. W. Pang, Schrier, and Weissman 2017).

Therefore, the HSC population, pure in terms of expression of surface markers, and previously considered homogeneous, is in fact made up of functionally distinct sub populations. However, while the different subsets may be defined by their progeny *in vivo*; due to the retrospective nature of these assays, individual cells shown to possess HSC properties are no longer available for molecular analyses. An effective method to prospectively distinguish HSC subsets at molecular level is not yet available. Moreover, it is increasingly evident that the presence of cell surface markers does not directly reflect the transcriptional state of a cell (Schroeder 2010). A number of studies have assessed

transcriptional variation within the HSC population either using single cell RT-qPCR (Guo et al. 2013; Adolfsson et al. 2005; Forsberg et al. 2006; Månsson et al. 2007; Akashi et al. 2003; Hu et al. 1997; Miyamoto et al. 2002) and more recently 2 studies have performed single cell RNA-sequencing (RNA-seq) of mouse HSCs (Wilson et al. 2015; Grover et al. 2016) and 1 study has used this technology to study the zebrafish haematopoietic system (Macaulay et al. 2016). Single cell RNA-seq of HSCs in mouse in combination with index FACS analysis and transplantation has enabled identification of genetic programs associated with the true repopulating HSC (Wilson et al. 2015) as well as proportional changes in transcriptionally distinct HSC subpopulations with age (Grover et al. 2016). In the zebrafish model single cell RNA-seq in HSCs has uncovered several transcriptionally distinct differentiation states as cells undergo lineage commitment (Macaulay et al. 2016).

1.1.2 Early megakaryocyte lineage commitment from the HSC

In the classical model of haematopoiesis, hierarchical differentiation begins from HSCs and progresses through a cascade of increasingly committed and lineage restricted progenitors, regarded as transient intermediates that ultimately generate all mature circulating myeloid and lymphoid cell types (Figure 1.1a). In this model HSCs produce short-lived multipotent progenitors (MPPs) which, in turn generate the oligopotent common myeloid (Akashi et al. 2000) and lymphoid progenitors (Kondo, Weissman, and Akashi 1997) (CMP, CLP respectively). In myeloid differentiation CMPs undergo further commitment to either the granulocyte-macrophage progenitor (GMP) or the megakaryocyte-erythroid progenitor (MEP) which ultimately produces megakaryocytes (and platelets) and erythrocytes.

With improvement in cell purification and functional clonal assays several findings have challenged this conventional paradigm, as further resolution has been made possible. It is now becoming evident that commitment to the megakaryocyte lineage may occur early and from the HSC directly, therefore independent of the intermediate CMP and MEP progenitors (Notta et al. 2016).

Importantly, the original model for thrombopoiesis proposed that ST-HSCs and MPPs maintain a full repertoire of lympho-myeloid lineage potential, comparable to the LT-HSC (Reya et al. 2001). However, the finding of an early loss of megakaryocyte/erythroid (MegE) potential upstream of the lympho-myeloid bifurcation has led to a key revision to this model. In mouse, LSKCD34+Flt3+ cells have been identified as lymphoid-primed multipotent progenitors (LMPPs) with combined granulocyte-macrophage and lymphoid potential but very limited MegE potential compared with LT-HSCs (LSKCD34-Flt3-) and ST-HSCs (LSKCD34+Flt3-) subsets both in *in vitro* clonal assays and *in vivo* transplantation (Adolfsson et al. 2005). This is supported by the finding that single HSCs were able to produce lineage committed bipotent MegE cells after just one cell division *in vitro* (Takano et al. 2004). The

concept of the LMPP and loss of MegE potential with progressive differentiation after the level of the HSC has been supported by a number of other studies (Forsberg et al. 2006; Lai and Kondo 2006; Yoshida et al. 2006; Luc et al. 2008; Arinobu et al. 2007) and following transplantation experiments, platelet reconstitution has been shown to be the first to be detected amongst all other terminally differentiated cells (Forsberg et al. 2006). More recently, paired daughter cell analysis has introduced the possibility of unipotent megakaryocyte progenitors arising directly from the LT-HSC independently of the erythroid lineage (Yamamoto et al. 2013; Haas et al. 2015) and in *in vitro* tracking experiments megakaryocyte colonies are the first to emerge from HSC cultures (Guo et al. 2013).

While human hematopoiesis is generally thought to follow the same differentiation model as in mouse, there has been very little certainty regarding megakaryocyte development from the human HSC, as standard *in vitro* clonal assays cannot be reliably interpreted (Notta et al. 2016). Recently, a new model for human haematopoiesis has been proposed based on improved cell sorting schemes, using additional cell surface markers and an optimised single cell clonal assay able to support simultaneous lineage commitment to myeloid, erythroid and megakaryocytic fates (Notta et al. 2016). Single cell clonal assays were performed on 11 populations to determine cell fate: HSCs, MPPs, CMPs, MEPs and GMPs; the classically defined MPPs, CMPs and MEPs were further subdivided into 3 populations based on their surface expression of CD71 and the thrombopoietin (TPO) receptor: MPL. Interestingly, MEPs from adult bone marrow fate was predominantly restricted to erythroid and to a lesser extent myeloid lineage, with less than 3% of cells giving rise to bipotent MegE progenitors. Rather megakaryocyte lineage potential was independent of intermediate progenitors (CMP and MEP) and branched directly from the multipotent HSC compartment, becoming the first major lineage bifurcation (Figure 1.1b). The originally defined CMPs were found to be a highly heterogeneous population (Notta et al. 2016).

1.1.3 Megakaryocyte-biased HSC population

Single cell RT-qPCR and more recently single cell RNA-seq has provided the opportunity to assess transcriptional variation within the HSC and other progenitor cell compartments and has thereby been used to decipher the organisation of haematopoietic hierarchies. Transcriptional profiling also supports the revised model for haematopoiesis described above. Unlike committed progenitors, HSCs and MPPs are shown to express multiple lineage associated genetic programs(Akashi et al. 2003; Hu et al. 1997; Miyamoto et al. 2002). A number of studies in mouse, have demonstrated megakaryocyte specific gene expression in the HSC compartment, a gradual down regulation of MegE transcriptional priming from LT-HSCs to LMPPs (Adolfsson et al. 2005; Månsson et al. 2007; Notta et al. 2016) and a lack of co-expression between MegE genes (e.g., *GATA-1*, *VWF*, *MPL*) and lymphoid specific genes in single HSCs suggesting a subset of LT-HSCs specifically primed towards a MegE lineage

(Månsson et al. 2007). Furthermore, when murine haematopoietic progenitor cell populations are clustered on the basis of single cell gene expression, the MegE lineage branch is found to be closely associated with LT-HSCs. This further substantiates previous evidence to show that the MegE lineage separates early from lympho-myeloid lineages. Indeed, a MegE genetic program has been demonstrated within the most primitive subset of mouse HSCs – the LSKCD34$^-$Flt3$^-$CD48$^-$CD150$^+$ HSC population (Guo et al. 2013).

More recently expression of larger repertoires of genes has been simultaneously assessed at the single cell level using RT-qPCR or single cell RNA-seq. Single cell gene expression analysis has allowed a number of megakaryocyte specific genes (including *VWF*, *GP1BA*, *ITGA2B* and *MPL* to be detected in populations within the most primitive HSC subset in both mouse (Wilson et al. 2015) and zebrafish (Macaulay et al. 2016) models which could reflect a HSC population biased towards the megakaryocyte lineage.

A platelet-primed HSC population has been prospectively identified on the basis of *VWF* gene expression which is the most highly expressed megakaryocytic gene in mouse HSCs (Sanjuan-Pla et al. 2013). Using a transgenic mouse model with a GFP (green fluorescent protein) reporter to the *VWF* gene, approximately 60% of HSCs (LSKCD34-CD48-CD150+) were shown to express *VWF*. This subpopulation was TPO dependent and demonstrated higher expression of other megakaryocyte specific genes using both microarrays and qPCR. When *VWF* expressing HSCs were serially transplanted in limiting numbers they led to both short and long-term reconstitution with a marked platelet bias, a contribution to the myeloid lineage and only limited lymphoid potential. While the *VWF*+ HSC population was able to give rise to *VWF*-ve HSCs, the opposite was not true hierarchically placing this platelet-biased HSC population at the apex of the haematopoietic tree. More recently, single cell RNA-seq of HSCs from young and aged mice has revealed that megakaryocyte specific genes are among the most upregulated in aged HSCs (Grover et al. 2016). Furthermore, single cell RNA-seq has made it possible to demonstrate that on a genome wide level both the number and level of expression of megakaryocyte and platelet specific genes increases with age. While this was true in both *VWF*+ and *VWF*- HSCs the *VWF*+ HSC population was seen to expand markedly in relation to the *VWF*- HSC population with age.

Further evidence for a megakaryocyte/platelet biased HSC population that could be analogous to that expressing *VWF* has also been reported on the basis of differences in cell surface marker expression. The megakaryocyte/platelet cell surface marker and part of the glycoprotein IIb/IIIa fibrinogen receptor, CD41, had been known to be thought to only be expressed in embryonic HSCs being switched off after birth (Ferkowicz 2003; Debili et al. 2001; Rhodes et al. 2008; Robin, Ottersbach and Boisset 2011). However, its surface expression has now been demonstrated on a subset of adult mouse HSCs (LSKCD34-CD48-CD150+) which show a platelet-bias on serial transplantation with the knock out of CD41

resulting in reduced levels of all mature blood cell lineages; again, a population that expands with age (Gekas and Graf 2013). Another platelet-biased subset within the mouse (LSKCD34-CD150+) HSC population has been proposed on the basis of expression of c-Kit (Grinenko et al. 2014; Shin et al. 2014). Early platelet production was shown to be restricted to the highest c-Kit expressing HSCs both transplantation and *in vitro* clonal assays and this was also correlated with megakaryocyte specific gene expression.

1.1.4 Similarities between HSCs and megakaryocytes and interaction within the stem cell niche

It has become increasingly clear that HSCs and megakaryocytes have a more complex relationship than previously appreciated. As described above, improved cell purification techniques and single cell transcriptional resolution have demonstrated expression of megakaryocyte lineage specific genes in murine and human HSCs (Wilson et al. 2015; Macaulay et al. 2016; Notta et al. 2016). Furthermore, in unsupervised clustering human HSCs and megakaryocyte progenitors have been shown to cluster and display more similarity than any other progenitor pairs in the haematopoietic tree using either single cell methods (Wilson et al. 2015) or bulk RNA-seq (Novershtern et al. 2011). Megakaryocytes and HSCs share dependence on a number of transcription factors including RUNX1 (Sun and Downing 2004; Growney et al. 2005; Gilliland et al. 1999), GATA2 (Rodrigues et al. 2005; Chang et al. 2002), the Ets family of transcription factors (Hock et al. 2004) and MEIS1 (Azcoitia et al. 2005; Hisa et al. 2004).

HSCs and megakaryocytes also share cell surface markers the most notable being the TPO receptor MPL. TPO was first cloned in 1994 and is known to be the major regulator of megakaryopoiesis and megakaryocyte maturation (Gurney et al. 1994). MPL expression is critical for HSC expansion with HSCs from MPL deficient mice being ineffective in long term mature blood cell reconstitution of irradiated recipients (Kimura et al. 1998). TPO signalling has also been shown to play an important regulatory role in maintaining HSC quiescence (Qian et al. 2007; Yoshihara et al. 2007). Other shared cell surface markers include CD150 (Kiel et al. 2005; Pronk et al. 2007), the fibrinogen receptor CD41 (Gekas and Graf 2013) and the CXCL12 receptor, CXCR4, which plays a role in homing both megakaryocytes and HSCs to the vascular sinusoidal space and in their retention within the bone marrow (Möhle et al. 1998; Z. Wang et al. 2008).

Regulation of HSC proliferation, quiescence and long-term repopulating potential occurs in the vascular sinusoidal space, at the endosteal surface of trabecular bone marrow, termed the stem cell niche (Morrison and Scadden 2014). The importance of this microenvironment in regulation of HSC function has been widely reported. A number of cell types within the niche, all of non-haematopoietic origin, have been implicated in HSC maintenance. These include

endothelial cells, osteoblasts, mesenchymal stromal cells, Schwann cells and sympathetic nerves either directly or via the production of factors such as CXCL12 (Ding and Morrison 2013; Greenbaum et al. 2013), angiopoietin (Arai et al. 2004), kit ligand (Thorén et al. 2008), and TGFB (Yamazaki et al. 2011). Megakaryocytes reside in direct contact with the sinusoidal endothelium, where platelet production occurs and are therefore in close proximity with HSCs in the stem cell niche (Morrison and Scadden 2014). A role for megakaryocytes in the regulation of HSC function has been proposed (Morrison and Scadden 2014; Heazlewood et al. 2013; Olson et al. 2013) suggesting a regulatory feedback mechanism between megakaryocytes and the HSC compartment. Recent reports demonstrate a critical role for megakaryocytes in maintaining HSC quiescence, whereby in mouse, acute depletion of megakaryocytes lead to HSC expansion and proliferation, with no effect on other cell types (Bruns et al. 2014; Zhao et al. 2014). It has also been suggested that this effect may be mediated through release of CXCL4 (PF4 or platelet factor 4) (Bruns et al. 2014) as injection of PF4 into megakaryocyte depleted mice has been shown to cause partial HSC quiescence. By contrast PF4 knockout mice showed increased HSC cell cycle activity and proliferation. TGFβ release from megakaryocytes has also been implicated (Zhao et al. 2014), whereby injection of TGFB in megakaryocyte depleted mice restored HSC quiescence and in mice with conditional deletion of TGFB an expansion in the HSC population is observed. Taken together these observations suggest that megakaryocytes serve as HSC-derived niche cells to dynamically regulate HSC function. It is of course possible that due to the HSC heterogeneity described above, specific molecular cues within the niche may instruct HSC fate at different levels depending on the targeted subpopulation. Moreover, platelets may also regulate HSC quiescence through effects on circulating TPO concentrations as described in Section 1.3 (de Graaf et al. 2010).

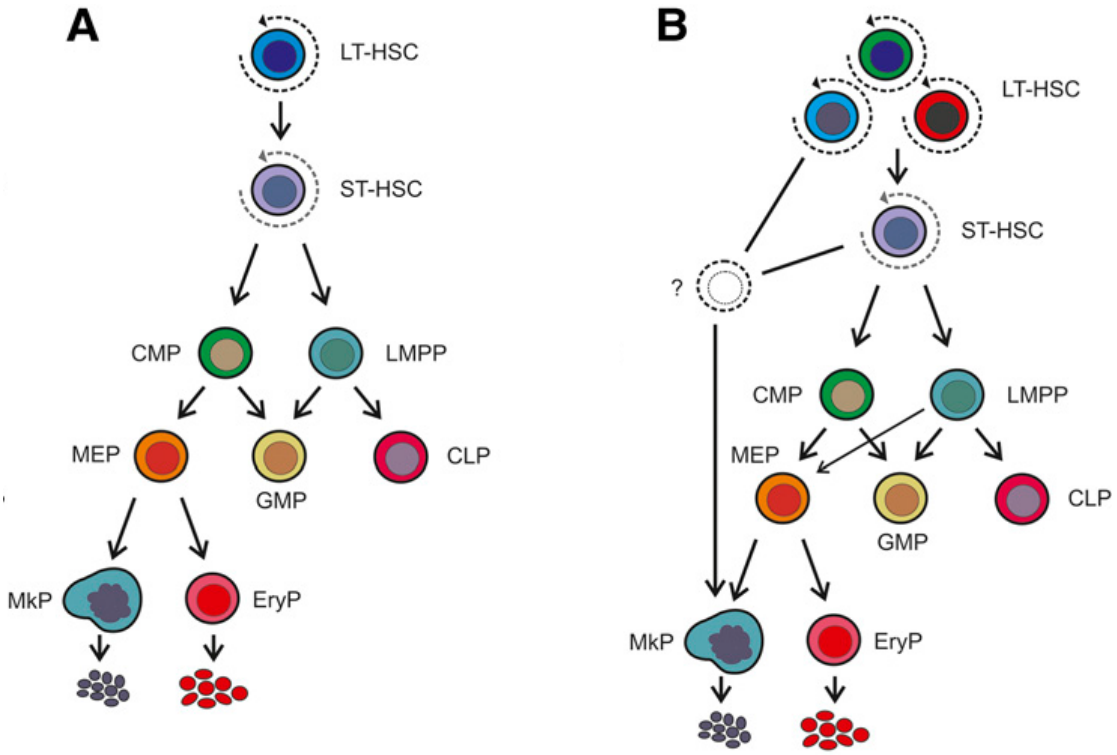


Figure 1.1 Classical and redefined blood cell differentiation pathways.
A. Classical haematopoietic tree with megakaryocyte differentiation occuring through a number of progressively committed progenitors; B. Redefined haematopoietic tree with direct differentiation from HSC to megakaryocyte independent of intermediate precursors. LT-HSC: long term HSC; ST-HSC: short term HSC; CMP: common myeloid progenitor; MEP: megakaryocyte-erythroid progenitor; MkP: megakaryocyte progenitor, EryP: erythroid progenitor, LMPP: lymphoid-primed multipotent progenitor; GMP: granulocyte-macrophage progenitor; CLP: common lymphoid progenitor. (Adapted from Woolthuis and Park 2016).

1.1.5 Terminal megakaryopoiesis

Despite recent evidence presented above to support an early lineage commitment to the megakaryocyte lineage from the HSC, our current understanding of the terminal stages of megakaryopoiesis is based on the classical model of haematopoietic differentiation where there is an established evidence base to support the common origin of the megakaryocyte and the erythroblast from the bi-potent MEP. Although much of these evidence to support an MEP population is from cell cultures (Debili et al. 1996), the MEP has been observed in murine bone marrow (Vannucchi et al. 2000) and there is an overlap in the transcription factors that regulate both megakaryocyte and erythroblast differentiation including GATA-1 (Lemarchandel et al. 1993), GATA-2 (Orkin 1992), NFE-2 (Romeo et al. 1990) and FOG-1 (Chang et al. 2002) and therefore similarities in gene expression (L. Chen et al. 2014). However, recent work using transcriptional profiling at the single cell level has shown that conventionally defined MEPs are a mixed population with the majority of cells transcriptionally primed to generate exclusively either megakaryocytes or erythroblasts (Psaila et al. 2016; Notta et al. 2016).

MEPs have traditionally been described to give rise to the burst forming unit megakaryocyte (BFU-MK) with high proliferation capacity with the ability to develop into 50-500 megakaryocytes in a week which in turn gives rise to the more mature colony forming megakaryocytes (CFU-MK) which can only give rise to 3-50 megakaryocytes (Briddell et al. 1989). Both of these early progenitors express the surface antigens CD34, CD33, surface glycoproteins IIb (GPIIb/CD41) and IIIa (GPIIIa/CD61), leukocyte-associated immunoglobulin-like receptor 1 (LAIR-1) and MPL. The promegakaryoblast is the most primitive morphologically distinct megakaryocyte precursor which gives rise to the megakaryoblast. The megakaryoblast (15-50um) has a high nuclear to cytoplasmic ratio, with elevated RNA levels and synthesis of alpha and dense granules. The pro-megakaryocyte is larger (20-80um) with a more mature cytoplasm, with developed granules and a complex demarcation membrane system, but not yet polyploid (Italiano and Shivdasani 2003).

1.1.6 Polyploidisation by endomitosis

While megakaryocyte progenitors proliferate through 2N-4N stages (mitosis), during which their progression through the cell cycle is indistinguishable from any other haematopoietic cell, the mature megakaryocyte is an obligate polyploid cell (Odell and Jackson 1968). Megakaryocytes maturation is accompanied by successive rounds of DNA replication without cytokinesis, known as endomitosis. This process results in large polyploid cells containing 2 (4N) to 64 (128N) times the DNA content of a cell that replicates by mitosis, within a lobulated nucleus. Although the endomitotic process can stop at any ploidy level, during steady state platelet production the mean ploidy of mammalian megakaryocytes in the bone marrow is

16N (Nagl 1978; Tomer, Harker, and Burstein 1988) (Figure 1.2).

It was initially thought that endomitosis corresponded to a failure of mitosis in anaphase, as the late stages of mitosis were not observed (Nagata, Muro, and Todokoro 1997). However, studies using time-lapse microscopy have shown that megakaryocytes complete anaphase, but there is regression of cleavage furrow formation leading to failure of the completion of cytokinesis (Lordier et al. 2008) (Figure 1.3). Although the mechanism of polyploidisation is still not well understood, altered expression of cyclins D and E, spindle checkpoint proteins and chromosome passenger proteins, such as Aurora A kinase (AURKA) and Aurora B kinase (AURKB), key regulators of cytokinesis, have been implicated (Wen et al. 2012; Lordier et al. 2010). AURKA is known to regulate microtubule-organising center localisation, chromosome dynamics and histone H3 phosphorylation in oocytes (Ding, Swain, and Smith 2011) and is required for bipolar spindle development (Cowley et al. 2009). AURKB, although with a less well understood role in polyploidisation, is a component of the chromosome passenger complex that is known to control mitotic and endomitotic cells cycles (Lordier et al. 2010). AURKA and AURKB inhibition has been found to lead to increased ploidy in cultured megakaryocytes as well as *in vivo* (Wen et al. 2012). It has recently been shown that polyploidy and thrombopoiesis may also occur via endoreplication which like endomitosis leads to uniform multiple genomic DNA replication rounds but achieves this by successive DNA synthesis and gap phases rather than aberrant mitotic cycles (Trakala et al. 2015). Endomitosis is certainly the predominant polyploidisation mechanism *in vivo* as evidenced by severe thrombocytopaenia caused by genetic ablation of Cdc20 (cell division cycle 20), which is known to lead to mitotic arrest (Trakala et al. 2015). However, this may be rescued by concomitant ablation of the main cell cycle kinases, Cdk1 and Cdk2, effectively facilitating polyploidisation via the unconventional route of endoreplication which is compatible with megakaryocyte function and thrombopoiesis (Trakala et al. 2015). This supports the hypothesis that the primary function of polyploidisation is to increase cell size for thrombopoiesis (Thon and Italiano 2012).

While it is assumed that the role of polyploidisation is to increase cell volume and synthetic capacity while minimising the energy required for cytokinesis (Zimmet and Ravid 2000), the precise purpose of acquiring multiple functional chromosome complements is unclear. Polyploidisation has been shown to be associated with functional gene amplification, where there is an increase in the number of copies of a gene without a proportional increase in other genes, rather than epigenetic silencing (Raslova et al. 2003). The concept of gene amplification with increases in ploidy level and size is supported by the observation that megakaryocytes derived from adult blood have an increased capacity to produce platelets in culture conditions than megakaryocytes derived from cord blood which tend to have a relatively lower ploidy (Mattia et al. 2002). Furthermore, the extent to which ploidy level and platelet formation are linked is not well established; high levels of polyploidy are thought not to

be essential for platelet production (Mattia et al. 2002; Miyazaki et al. 2000). It has been shown that there is no difference in cytoplasmic extensions or platelet production in culture between megakaryocytes cultured from adult blood of low ploidy compared with cultured megakaryocytes driven to higher ploidy levels with nicotinamide (vitamin B3) (Giammona et al. 2006). However, the authors note that it is not known to what extent the method used for measuring platelet production in culture recapitulates physiological platelet release. Furthermore, while nicotinamide increased cell ploidy it did not induce megakaryocyte cytoplasmic maturation (Giammona et al. 2009), emphasising the importance of cell maturation as an independent factor to megakaryocyte ploidy level in platelet production.

1.1.7 Cytoplasmic maturation

As the megakaryocyte undergoes polyploidisation it matures to contain all the machinery necessary for platelet release and platelet function. Secretory granules are formed in the cytosol including α granules and dense granules. α granules originate from the trans-Golgi network and contain either endogenously synthesised proteins such as PF4, β thromboglobulin, von Willebrand factor and platelet membrane proteins or endocytosed proteins such as albumin and fibrinogen although the latter may also be synthesised within the megakaryocyte (Podolak-Dawidziak et al. 1995). Dense granules contain haemostatically active substances that are released at platelet activation including serotonin, ADP, ATP, catecholamines and calcium. Electron microscopy studies have shown that multivescicular bodies are an intermediary stage of both α granule and dense granule maturation (Heijnen et al. 1998; Youssefian and Cramer 2000). As the megakaryocyte matures an extensive network of demarcation membrane composed of tubulin and microtubules develops in the cytoplasm (Yamada 1957), it is thought that these tubular membranes originate from invaginations of the megakaryocyte plasma membrane. This is a dynamic system and is thought to organise the cytoplasm into platelet territories and therefore may function as a membrane reserve in platelet formation. The mature megakaryocyte and its daughter platelets express the following cell surface markers, some of which are early markers and are also present on committed megakaryocyte progenitors (Section 1.1.5): GPIIb/IIIa (CD41/CD61), GPIb (CD42b), CXCR4 and GPVI.

1.1.8 Platelet formation

The primary function and the defining feature of a mature megakaryocyte is the release of platelets. Several models of *in vivo* platelet production have been proposed. The widely accepted model of thrombopoiesis is the proplatelet model, where megakaryocytes in the bone marrow develop multiple long complex cytoplasmic extensions known as proplatelets that extend into the marrow sinusoids to release platelets into the circulation (Radley and Scurfield 1980). These platelet territories, within the megakaryocyte, have been suggested to

predetermine platelet size (Paulus, Bury, and Grosdent 1979). This model is supported by the documentation of a network of branching proplatelets in a culture system (Italiano et al. 1999) and imaging of this process in the bone marrow compartment *in vivo*, the release of proplatelets into the circulation and proplatelet fragmentation in the circulation itself (Junt et al. 2007). *In vitro* derived megakaryocytes have been used to study pro-platelet formation (Patel et al. 2005) and variations of this model have been used to produce functional cultured platelets for transfusion (Moreau et al. 2016). However, the extent to which current *in vitro* methods of platelet production reflect true physiology is unknown.

Another model of platelet production suggests megakaryocyte fragmentation in the peripheral circulation (Behnke and Forer 1998). This is supported by the presence of megakaryocytes and megakaryocyte fragments in the peripheral circulation and the fact that it is possible to culture human megakaryocytes from cord blood and peripheral blood *in vitro* with the production of small numbers of functioning “platelets” means that the bone marrow niche may not be essential for platelet release.

There is now mounting evidence for the lung as a major site for thrombopoiesis. Mature megakaryocytes have been shown in the pulmonary arterial circulation, with bare megakaryocyte nuclei found in the pulmonary capillary bed (Slater, Trowbridge, and Martin 1983; Pedersen 1978). The pulmonary production model suggests that platelet production from circulating megakaryocytes occurs in the lungs where large circulating megakaryocytes, with significant kinetic energy from the right ventricle, impact on the pulmonary capillaries and fragment to platelets within the lung vasculature. It has also been noted that in disorders where megakaryocytes or megakaryocyte fragments could bypass the lung capillary network (e.g., right-to-left intracardiac shunts, lung disease with abnormal circulation) these large particles could reach the fingertips in axial vascular streams and impact there, releasing PDGF and other platelet-derived growth factors (Fox, Day, and Gatter 1991; Silveri et al. 1996) which would leads to the clubbing observed in these conditions (Dickinson and Martin 1987). A recent study has provided direct evidence of the pulmonary circulation as a major site for platelet production, proposing that beyond the mechanical model, the lung may contain unique signals to promote platelet release (Lefrançais et al. 2017). Platelets produced in the lung have been shown to be functional in an *in vivo* thrombosis model (Fuentes et al. 2010).

It is possible that all three models: thrombopoiesis in the bone marrow, peripheral circulation and lungs, play a physiological role. The relative contribution of each to platelet number is not clear, however it has been argued that pulmonary circulation could provide total platelet need based on quantitative data (Trowbridge, Martin, and Slater 1982).

1.1.9 Regulation of megakaryocyte maturation

1.1.9.1 Cytokines

Megakaryocyte differentiation from the HSC, as well as megakaryocyte maturation are primarily driven by TPO (Bartley et al. 1994; de Sauvage et al. 1994; Lok et al. 1994). TPO regulates both polyploidisation and cytoplasmic maturation by binding to its receptor on the megakaryocyte cell surface, MPL (CD110) (Bartley et al. 1994; de Sauvage et al. 1994). Binding and dimerisation of this receptor lead to the autophosphorylation of Janus kinase 2 (JAK2) and downstream activation of further signalling pathways including mitogen-activated protein kinases (MAPK), phosphoinositol-3 kinase (PI3K) and signal transducers and activators of transcription (STATs) (Dorsch et al. 1997). Ultimately this drives each stage of megakaryocyte differentiation by modifying transcription. TPO is synthesised in the liver and binds to its receptor on circulating platelets. Regulation of circulating TPO levels have been described to be via platelet TPO/MPL mediated adsorption (Kaushansky 2005). However, a recent report shows that abnormal circulating platelets influence the Janus kinase 2 signalling and activation of transcription in the liver leading to increased TPO production (Grozovsky et al. 2015). Megakaryocytes from mice that express no TPO or MPL receptor are small and low in ploidy with significant reductions in circulating platelet levels (Alexander et al. 1996; Gurney et al. 1994). Moreover, human congenital MPL deficiency manifests in severe thrombocytopaenia (Ihara et al. 1999). Interestingly, TPO may be dispensable for platelet release or even inhibit platelet release as suggested by experiments in cultured megakaryocytes (Choi et al. 1995; Ng et al. 2014). Along with TPO, megakaryopoiesis is influenced by a network of cytokines including interleukin1B (IL1B), IL3 IL6, IL11 and stem cell factor (SCF), chemokines and extracellular matrix proteins (Zheng et al. 2008).

Environmental signals are also crucial for megakaryocyte maturation and platelet formation. It has been shown that movement of the megakaryocyte into the vascular sinusoids and contact with the bone marrow endothelium induces maturation and platelet formation (Avecilla et al. 2004; Hamada et al. 1998). This interaction is promoted by the chemokine CXCL12 (SDF1) and growth factor FGF4 (Avecilla et al. 2004) and enhanced by the pro-inflammatory cytokine IL1β (Josefsson et al. 2011; White et al. 2012). Furthermore, *in vivo*, the pro-inflammatory VEGFR1 pathway, through upregulation of the CXCL12 receptor on megakaryocyte surface (CXCR4), also increases platelet production (Pitchford, Lodie, and Rankin 2012), possibly accounting for the increased platelet count observed in inflammatory conditions. There is also evidence that the extracellular matrix influences platelet production, with type I collagen inhibiting proplatelet formation (Sabri et al. 2004) and fibrinogen (Bury et al. 2012) and VWF (Balduini et al. 2009) promoting proplatelet formation through their respective receptors: GPIIb/IIIa and GPIb/IX/V.

1.1.9.2 Transcription factors

Haematopoietic differentiation is driven by transcription factors that direct lineage commitment by regulating gene expression (Iwasaki et al. 2006). MEP differentiation into the megakaryocyte lineage is known to be driven by the transcription factors FLI1, TAL1, GATA1, RUNX1 and MEIS1 (Watkins et al. 2009). The expression of this combination of transcription factors was identified as specific to megakaryocytes compared to other cells in the haematopoietic tree (L. Chen et al. 2014). Furthermore, lentiviral transduction with FLI1, GATA1, TAL1 have been used to forward program induced pluripotent stem cells (iPSCs) directly to megakaryocytes in culture (Moreau et al. 2016).

Several transcription factors have been also implicated as drivers of megakaryocyte maturation (endomitosis, cytoplasmic maturation and platelet formation). GATA1 has been shown to be essential at all three points of maturation. By directly enhancing expression of Cyclin D1, GATA1 plays a crucial role in polyploidisation (Muntean et al. 2007). GATA1 deficient megakaryocytes remain small and immature, with low ploidy levels (Shivdasani et al. 1997). GATA1 directly enhances *GP1BB* expression, thereby is critical in platelet production and function, demonstrated by the Bernard-Soulier Syndrome phenotype of thrombocytopaenia and abnormal platelet function caused by a point mutation in the GATA1 binding site in the *GP1BB* promoter region (Ludlow et al. 1996). RUNX1 plays an essential role in endomitosis and regulates constituents of the megakaryocyte cytoskeleton by the enhancing expression of *MPL* and *MYL9* (myosin light chain 9) (Sun et al. 2007) and silencing *MYH10* (non-muscle myosin IIb which specifically localises to the contractile ring in cytokinesis) (Lordier et al. 2012). FLI1 regulates the expression of *MPL*, *GP1BA*, *GPIX* and *PF4* thereby influencing both polyploidisation and cytoplasmic maturation (L. Pang et al. 2006). Knockout of transcription factor TAL1 leads to small immature low ploidy megakaryocytes with reduced platelet count but platelet function is preserved. This effect is in part due to its regulation of *CDKN1A* (cyclin-dependent kinase inhibitor 1A) (Chagraoui et al. 2011) and likely other targets essential in cytoplasmic maturation which may include *MEF2C* (myocyte enhancer factor 2C) (Gekas et al. 2009). The transcription factor EVI1 promotes *CDK2* (cyclin-dependent kinase 2) (Kilbey et al. 2005) and therefore implicated in endomitosis. The transcription factor NFE2 has been shown to be essential for platelet formation, and in turn is transactivated by GATA1 (Nakamura et al. 2014; Takayama et al. 2010). NFE2 knockout mice completely lack circulating platelets (Zimmet, Toselli and Ravid 1998). This may be due to the function of the NFE2 transcriptional targets: *TUBB1* (β tubulin) (Lecine et al. 2000), *RAB27β* (Tiwari et al. 2003), *CASP12* (Kerrigan et al. 2004) and *HSD3B1* (Nagata et al. 2003). Clearly the control system of both megakaryopoiesis and later differentiation is highly complex and it is reasonable to suggest these transcription factors may act interdependently. In support of this, it has been shown that the five transcription

factors GATA1, GATA2, RUNX1, FLI1 and TAL1 work as a network and bind simultaneously, in cultured megakaryocytes, at a number of binding sites in the genome (Tijssen et al. 2011).

1.1.10 Megakaryocyte maturation *in vitro*

TPO is also a potent stimulator of megakaryocyte maturation *in vitro*. It increases the size, ploidy and number of megakaryocytes and is associated with the expression of maturity markers such as CD41, CD61 and CD42. The discovery of TPO significantly advanced the ability to study megakaryocytes in the culture system, enabling the *in vitro* production of mature megakaryocytes differentiated from cord and adult blood or fetal liver CD34+ cells.

Despite this, it is not yet possible, in cultured human cells, to achieve ploidy levels that are as high as those observed *in vivo*. Cultured megakaryocytes from adult blood have a larger representation of the higher ploidy classes than cord blood or fetal liver derived megakaryocytes, but mean ploidy remains 2N-4N (Figure 1.4). This is very different to primary human bone marrow megakaryocytes where mean ploidy is 16N (Figure 1.2). Therefore, in order to study ploidy effectively, compounds that increase megakaryocyte ploidy, as well as their maturation are required. A number of compounds have previously been studied for this purpose. Phorbol 12-myristate 13-acetate (PMA) has been known for many years to promote megakaryocyte differentiation and polyploidisation in megakaryocyte like cell lines, including Dami cells (Greenberg et al. 1988) and K562 (X. Huang et al. 2004). Nicotinamide, as mentioned above, has also been shown to increase ploidy levels, this time in megakaryocytes cultured from human CD34+ cells (Giammona et al. 2006). However this is not accompanied by signs of megakaryocyte cytoplasmic maturation (Giammona et al. 2009). More recently, AURKA and AURKB inhibitors have been shown to increase ploidy and megakaryocyte differentiation. These include dimethylfasudil (diMF), an inhibitor of both AURKA and AURKB and MLN8237 (alisertib) that selectively inhibits AURKA. Alisertib is currently being used in clinical trials for a variety of tumours, including acute leukaemias and is thought to confer clinical benefit by inducing terminal cellular differentiation (Wen et al. 2012).

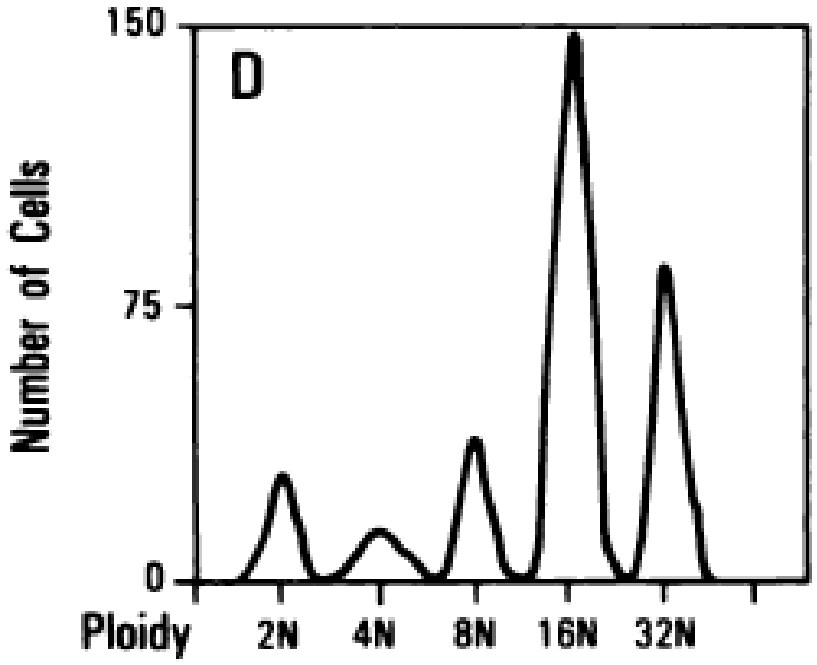
1.1.11 The megakaryocyte transcriptome

The transcriptome of human primary megakaryocytes has not yet been defined due to the requirement of bone marrow sampling, the relative infrequency of megakaryocytes in the marrow and the difficulty to obtain a pure cell population.

A small number of studies have profiled gene expression in megakaryocytes that have been cultured from CD34+ progenitor cells using whole genome expression arrays. Gene expression data that currently exists for human megakaryocytes cultured from progenitors

obtained from cord blood (Watkins et al. 2009), fetal liver (Bluteau et al. 2013) and adult blood (Shim et al. 2004; Raslova et al. 2007; Bluteau et al. 2013). These data provides a catalogue of genes that are specifically expressed at the transcript level in megakaryocytes. Apart from the differential expression of transcription factors, discussed above (Section 1.1.9), in adult blood derived megakaryocytes compared to haematopoietic progenitors, one third of the most highly upregulated genes are shown to participate in haemostasis (Shim et al. 2004). Comparison of megakaryocytes derived from fetal liver cells, cord blood and adult blood has shown that genes important in megakaryocyte maturation, platelet formation and function are switched on during ontogeny. This is reflected in the maturity and ploidy level of the megakaryocytes cultured from each source (Figure 1.4) as well as their ability to produce proplatelets *in vitro* (Bluteau et al. 2013). Comparing expressed genes between 2N-4N cultured megakaryocytes and 8N-16N cultured megakaryocytes gene transcripts were identified that were down regulated during polyploidisation, included those involved in DNA replication, cytokinesis and proliferation. Those genes that were upregulated during polyploidisation included transcripts for platelet glycoproteins, other important proteins with a role in platelet function e.g., VWF, P-selectin, thrombospondin and proteins involved in the actin, myosin and tubulin cytoskeleton (Raslova et al. 2007).

A



B

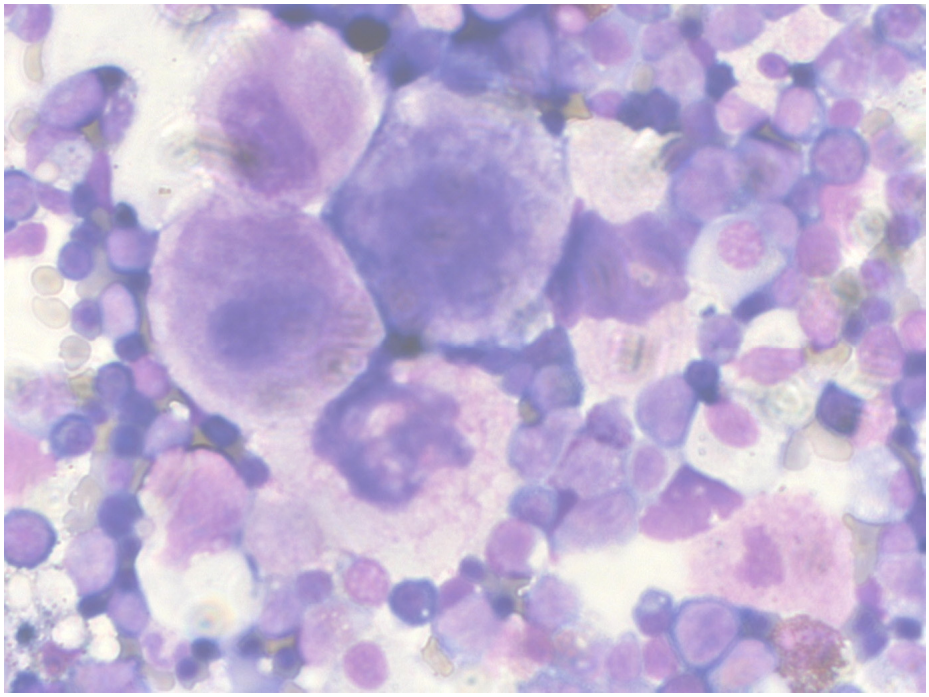


Figure 1.2 Highly polyploid mature bone marrow megakaryocytes.
A. Flow cytometry plot of enriched healthy human bone marrow megakaryocyte ploidy distribution measured by a DNA dye (figure adapted from Tomer, Harker and Burstein 1988); **B.** Morphology of healthy human bone marrow megakaryocytes in an enriched population after staining with Rapid Romanowsky stain.

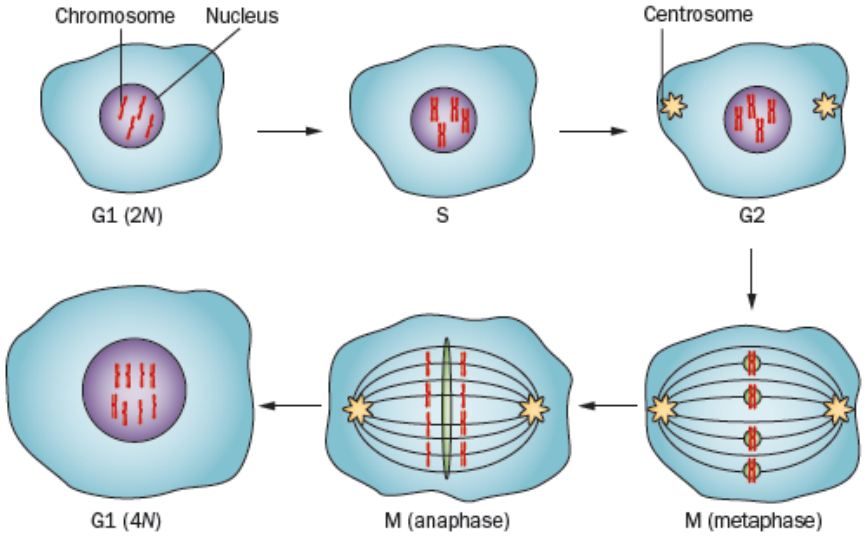


Figure 1.3 Megakaryocyte endomitosis.
A 2N megakaryocyte undergoes endomitosis to produce a 4N cell, with replication of its DNA without cell division. Continuation of this process to produce megakaryocytes up to 64N. (G1, Gap 1 of interphase; G2, Gap 2 of interphase; M, mitosis; S, synthesis of interphase). (Figure from Martin et al. 2012).

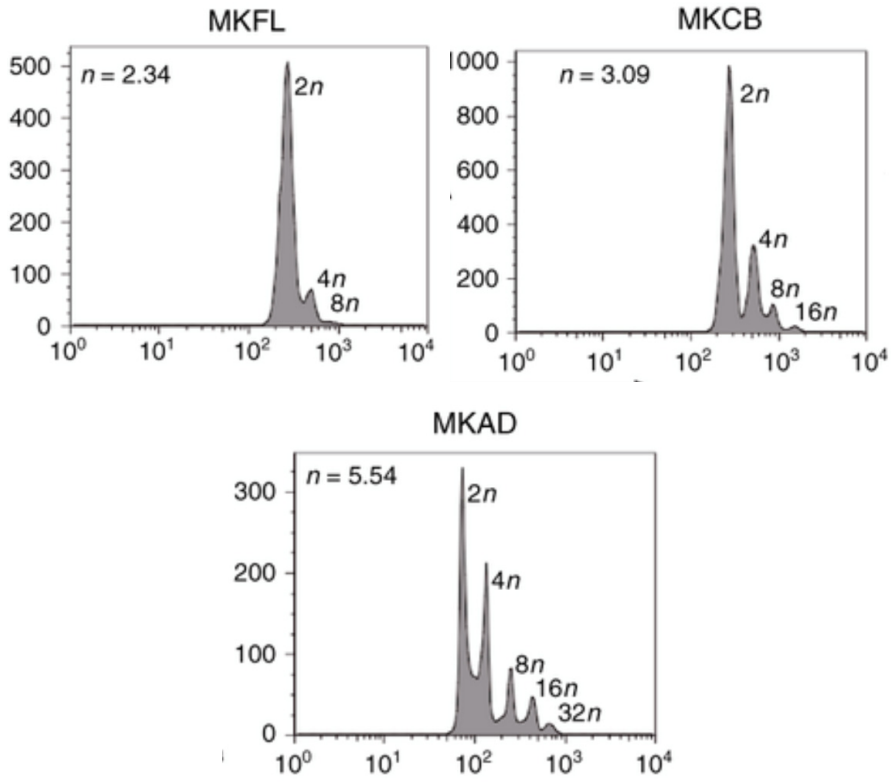


Figure 1.4 Ploidy distribution of megakaryocytes differentiated from CD34+ cells.
Flow cytometry plots of ploidy distribution measured by a DNA dye; (MKFL: megakaryocytes derived from fetal liver CD34+ cells, MKCB: megakaryocytes derived from cord blood CD34+ cells, MKAD: megakaryocytes derived from adult blood CD34+ cells. n: calculated mean ploidy). (Figure adapted from Bluteau et al. 2013).

1.2 Platelet Biology

1.2.1 Platelet function

Platelets released from the megakaryocyte are small anucleate cytoplasmic units typically measuring 2 to 5μm in diameter with a mean volume of 6 to 10fl. The average platelet lifespan is approximately ten days. Under steady-state conditions humans must produce at least 1 $x10^{11}$ platelets each day to maintain a normal platelet count (150-400 $x10^{9}$/L) and this level of production increases under conditions of increased demand. The primary role of platelets is in haemostasis and as such platelets play a pathological role in acute atherothrombotic disease. Furthermore, it is increasingly recognised that platelets have less well understood roles in inflammation, atherosclerosis, tumor growth, metastasis and angiogenesis (Nurden 2011; Bertozzi, Hess and Kahn 2010; Watson, Herbert and Pollitt 2010; Feng et al. 2011). They lack genomic DNA but contain megakaryocyte-derived mRNA and the translational machinery required for protein synthesis. Platelets contain numerous organelles, a circumferential microtubular system and a metabolically active membrane with a surface canalicular connection system giving a large surface area.

In the bloodstream, platelets circulate close to the luminal surface of the endothelium due to their small size in relation to other blood cells. Platelets are kept in a quiescent state by prostaglandin I2 receptor (PGI2) acting through the inhibitory G-protein coupled receptor PTGI2R and nitric oxide (NO) that diffuses across the platelet membrane. PGI2 and NO are released by endothelial cells and act to inhibit intracellular calcium release (Gibbins 2004). Platelets respond rapidly to changes in endothelial integrity and exposure of sub-endothelial structures, by binding the damaged surface via glycoprotein receptors that richly populate their outer surface, spreading and further amplifying the platelet response by the release of activating factors. The end result is the formation of thrombus or platelet plug that preserves vascular integrity and therefore minimises blood loss (Gibbins 2004).

The initial activation occurs when exposed sub-endothelial VWF and collagen interact with the receptor complexes GPIb/IX/V (López 1994) and GPVI-GPIa/IIa (Nieswandt and Watson 2003) respectively on the surface of resting platelet. Tissue factor from damaged endothelium further activates factor X to generate thrombin, which is a potent activator of platelet protease activated receptors 1 and 4 (PAR) (Bouchard et al. 1997) This initial activation causes a shift in the platelet surface integrins from a low to high affinity binding state (Liddington and Ginsberg 2002), leading to high affinity binding of GPIa/IIa to collagen and GPIIb/IIIa to VWF which promotes further activation and aggregation. Activation results in a unique cytoskeleton rearrangement transforming the resting oblate spheroid platelet to a more stellate form with long pseudopodia with the exteriorisation or secretion of the contents of platelet storage granules. Alpha granules contain haemostatically active factors - fibrinogen, thrombospondin,

VWF, platelet-derived growth factor (PDGF), PF4, transforming growth factor beta (TGFβ), fibronectin and coagulation factors as well as new GPIIb/IIIa and GPIb/IX/V complexes which allow for crosslinking between platelets, platelet aggregation and thrombotic propagation (King and Reed 2002). Furthermore, P-selectin which is transferred to the platelet membrane by the alpha granules on activation facilitates the interaction between platelets and leucocytes (Michelson 2012). The release of ADP, calcium and serotonin from dense granules causes further platelet activation through the G-protein coupled receptors P2Y1 and P2Y12 (Michelson 2012). Platelet activation also initiates the arachidonic acid pathway to produce thromboxane A_2 which activates nearby platelets via the TXA2 receptor as well as markedly increasing local concentrations of thrombin and Factor X (Michelson 2012). Overall, this cascade of events leads to formation of a platelet plug supported by a matrix including fibrin and leucocytes.

Platelet function can be measured by stimulating either whole blood or platelet rich plasma with agonists such as: ADP binding to P2Y1 and P2Y12 receptors, collagen binding to GPIa/IIa and GPVI and activating GPIIb/IIIa receptors; thrombin binding to PAR1 and PAR4 (Garner et al. 2017). Assessment of the resulting activation can then be made using either flow cytometry or aggregation using light transmission aggregometry.

1.2.2 Platelet physical characteristics

Commonly measured platelet characteristics by haematology analysers are: platelet count (PLT) by both electrical impedance and optical fluorescence and mean platelet volume (MPV) derived from electrical impedance. These are highly heritable traits (Evans, Frazer, and Martin 1999; Bertin et al. 2007; Mahaney et al. 2005; Bray et al. 2007; Garner et al. 2000) and highly reproducible within an individual while varying widely within the population. Plateletcrit (PCT) – a measure of platelet total mass, the platelet distribution width (PDW) and platelet large cell ratio (P-LCR), are also derived from electrical impedance. Immature platelet fraction (IPF), is a measure of reticulated platelets that contain mRNA using flow cytometry by detection of a RNA dye.

1.2.2.1 Mean platelet volume

Platelet size, measured as MPV is, together with platelet count, the most frequently studied platelet physical variable. MPV is calculated by dividing plateletcrit (a measure of platelet mass) by number of platelets. Platelet volume can vary from 2fl to 60fl with a typical MPV of 6-10fl. Unlike all other mammalian cells that undergo mitotic division, which have a Gaussian cell volume distribution, platelets fit instead a log Gaussian distribution curve (Paulus 1975) (Figure 1.5a). This means that the platelet volume distribution within an individual at any given time point demonstrates a greater heterogeneity or variation than in any other circulating cell

type. This has given rise to significant controversy as to how this unique volume distribution curve originates and what its physiological relevance is. As the platelet volume distribution changes, so do PDW and P-LCR. The PDW is the width of the size distribution curve in fl at the 20% level of the peak and the P-LCR is the number of cells falling above an arbitrary threshold of 12fl divided by the total platelet count.

It was originally thought that the unique platelet volume distribution curve was explained by platelet ageing, because each platelet volume decreased during its lifespan in the circulation (Karpatkin and Charmatz 1969). However, methodological flaws in the data have since questioned this conclusion (Martin et al. 1982). Evidence from a variety of species, where the size and density of labeled platelets have been measured throughout platelet lifespan, indicates that platelet heterogeneity does not arise from ageing (Thompson et al. 1983; Schulman I 1971; Martin and Penington 1983; Savage et al. 1986; Boughton et al. 1990). The implication of these findings is that the platelet volume distribution curve arises from the unique non-mitotic manner of platelet formation from megakaryocytes and is defined at the time of thrombopoiesis. During steady state thrombopoiesis an inverse relationship exists between MPV and platelet count within an individual (Karpatkin 1978) to conserve a constant circulating functional platelet mass. This is known to be true at a population level where the same inverse relationship exists at both the population level with individuals with high platelet counts having a low MPV and vice versa (Bessman 1984) and between different species (von Behrens 1972). Therefore, it is reasonable to suggest that platelet volume is regulated at the point of production according to physiological haemostatic demand. When equilibrium between platelet production and destruction is disturbed, with drop in platelet count, the primary response is the rapid production of larger platelets to maintain overall circulating platelet mass. The pulmonary production model provides an explanation for the observed log Gaussian distribution of platelet size, because the fragmentation pattern is a series of binary divisions (Trowbridge and Harley 1984).

Interestingly, a recent study has shown the generation of “platelet progeny” from platelets stored *ex vivo* either in culture media or within whole blood. The new “platelets” that are produced are in fact not smaller in size or volume and are shown to be functional (Schwertz et al. 2010). However the relevance of this phenomenon to the *in vivo* physiological setting is still unknown.

The importance of MPV as a biological variable has been extensively studied. There is a convincing body of evidence suggesting a positive correlation between platelet size and platelet function measured as aggregation and granular content release. Preferential aggregation of large platelets is observed after addition of the agonist ADP to platelet suspensions (Karpatkin 1978; Haver and Gear 1981). Larger platelets are denser, contain more granules and mitochondria per unit volume (Martin et al. 1983), cause reduced bleeding

time (Kristensen, Bath, and Martin 1990; Martin et al. 1983), have a higher capacity for thromboxane A_2 production (Martin et al. 1983) and TGFβ secretion (Thompson et al. 1982; Thompson et al. 1983) and express more GPIIb/IIIa receptors per unit area of plasma membrane (Giles, Smith, and Martin 1994) than smaller platelets. Platelet activation is significantly increased in larger compared to smaller platelets when these subpopulations are sorted by flow cytometry (Mangalpally et al. 2010). A negative correlation has also been observed between MPV and platelet activation in a large cohort of healthy individuals which has disputed the association between MPV and prothrombotic risk (Soranzo et al. 2009). However, more recently genome wide association studies have shown genetic loci associated with MPV that also have an effect on platelet function (Gieger et al. 2011).

1.2.2.2 Immature platelet fraction

"Immature platelets" and "reticulated platelets", (terms that are commonly used interchangeably) are newly released platelets from the megakaryocyte and are identifiable on the basis of their RNA content (Ault et al. 1992). In animal studies, where platelet lifespan is shorter (4-5 days compare to 9-10 in humans), RNA usually degrades within 24 hours (Dale et al. 1995). As these platelets have just been released from megakaryocytes, IPF, as measured by a haematology analyser, is used as a surrogate marker for megakaryocyte activity and is increased in individuals with activated thrombopoiesis and increased platelet turnover (Ault and Knowles 1995).

IPF is a quantification of reticulated platelets as a proportion of the total platelet population. The current method of measuring the IPF is based on the RNA content of a platelet detected by staining with an RNA dye as well as platelet size (Figure 1.5b) (Briggs et al. 2009). This method suggests that all newly formed platelets (reticulated platelets) are large. However, a major limitation of the method is that newly produced smaller platelets contain less RNA due to their size and may be below the limit of detection of the analyser. Therefore, the sensitivity of the test is low and it is likely that there are reticulated platelets not detected. The specificity is high, as all the platelets with high levels of mRNA are detected and these are by definition all reticulated platelets. These limitations however do not impact on the utility of the test, as it is used in the setting of activated thrombopoiesis where relative values are adequate, where a high IPF indicates increased megakaryocyte activity and platelet production. The term "immature" can therefore be somewhat confusing. Here the measured parameter will be referred to as IPF, as by convention and the newly formed platelets referred to as "reticulated platelets".

1.2.3 The platelet transcriptome

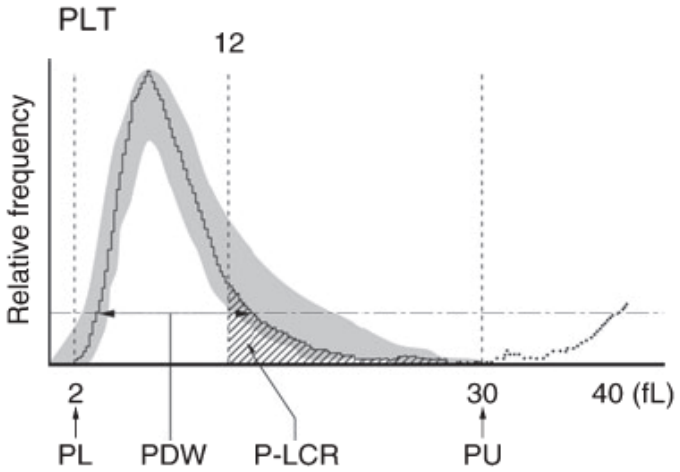
Being anucleate, platelets are incapable of transcription of nuclear genes. Therefore all platelet proteins, except plasma proteins taken up by endocytosis, are derived from the

megakaryocyte. Platelets contain mRNA derived from the megakaryocyte and translational machinery for protein synthesis. Through a number of studies, that have analysed and defined the transcriptome of platelets (Gnatenko et al. 2003; Bugert et al. 2003; McRedmond et al. 2004; Hillmann et al. 2006; Rowley et al. 2011), it is now clear that platelets contain a rich repertoire of RNA although in quantity, a platelet will contain significantly less RNA than a leukocyte (Fink et al. 2003). As platelets are anucleate and unable to transcribe RNA, the platelet RNA profile provides an insight into the transcriptional state of megakaryocytes and the bone marrow signals involved in thrombopoiesis. While much of the platelet transcriptome data comes from microarrays, arguably the most useful data is from RNA-seq experiments (Rowley et al. 2011).

It is generally assumed that mRNA expression in megakaryocytes and platelets reflect each other. However, differential packaging of mRNAs for matrix metalloproteases (MMPs) and tissue inhibitors of metalloproteases into platelets (TIMPs) has been shown (Cecchetti et al. 2011) which supports the regulation of mRNA transfer at the point of platelet production. The study analyses transcript and protein expression of these 2 families of proteins in the megakaryocyte and the platelet. *MMP-2* is expressed at mRNA and protein level in the megakaryocyte but only the protein is observed in platelets. Further, *TIMP-2* mRNA and protein have been identified in megakaryocytes but only the mRNA is seen in platelets.

Transcriptome analysis has been used in platelets to identify novel platelet proteins. Many of the most abundant platelet transcripts are products of known platelet genes *PF4*, *PPBP*, *ITGA2B*, *GP1BB*. However, there are a number of limitations to the study of the platelet transcriptome whereby it may not necessarily reflect the proteome of the platelet. The encoding transcript for proteins synthesised in the megakaryocyte and packaged into pro-platelets may be absent from the platelet transcriptome. Due to the difference in amount of RNA in platelets compared to leucocytes the potential for contamination is substantial, despite the fact that this is usually well addressed experimentally. Finally the purified population of platelets will be at differing ages with much of the RNA originating from newer platelets.

A



B

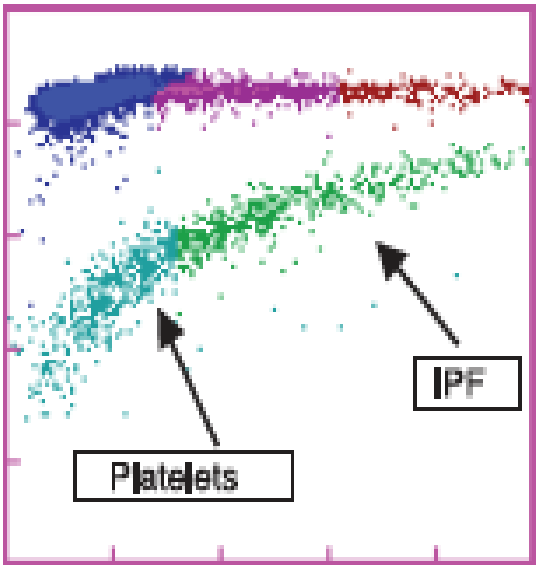


Figure 1.5 Platelet indices.
A. Log Gaussian platelet volume distribution curve. Frequency distribution of platelet volumes in a representative human platelet population measured using electrical impedance. Shaded section: range of results, continuous line: predicted log Gaussian curve (PL, lower discrimination; PDW, platelet distribution width; P-LCR, platelets-large cell ratio; PU, upper discrimination); **B.** Representative IPF plot of a thrombocytopaenic patient with raised IPF, flow cytometry plot: y: forward scatter (size), x: RNA content. IPF is defined by an arbitrary cut off of RNA content. (Figure adapted from Briggs et al. 2009).

1.3 Megakaryocyte-platelet axis in states of increased haemostatic demand

As discussed above, during steady state thrombopoiesis, a constant circulating platelet mass is maintained by an inverse relationship between platelet count and platelet volume at the point of thrombopoiesis. Therefore, in the setting of platelet consumption, such as acute blood loss or chronic immune thrombocytopaenia, the physiological response is the production of large platelets. As such, in these clinical scenarios, MPV and IPF values are increased. This phenomenon is demonstrated in animal models, where antiplatelet serum induced thrombocytopaenia results in an increase in MPV as well as platelet reactivity (Martin, Trowbridge, et al. 1983; Kuter, Beeler, and Rosenberg 1994), as well as in the clinical setting of immune thrombocytopaenia purpura where giant platelets are visualised on blood smear (Levin and Bessman 1983). Furthermore, in surgical patients on cardiopulmonary bypass, where platelets are acutely depleted in the extracorporeal circulation, a similar response is observed (Martin, Daniel, and Trowbridge 1987) (Figure 1.6). As shown in Figure 1.6, MPV remains unchanged immediately after the drop in platelet count, but then increases, arguably in response to the reduction in platelet number.

There is also evidence to suggest that in settings where there is an acute or chronic decrease in circulating platelet mass, the megakaryocyte response is an increase in ploidy level and cytoplasmic volume in both animals (Stenberg et al. 1991; Corash and Levin 1990; Harris and Penington 1984) and humans (Mazur et al. 1988). For example, the distribution of megakaryocyte ploidy in the bone marrow is seen to shift to higher values in patients with immune thrombocytopaenia purpura (Tomer et al. 1989) which is also associated with larger platelets (Penington, Streatfield, and Roxburgh 1976; Bessman 1984). In acute thrombocytopaenia a number of studies have demonstrated that an acute fall in platelet count is associated with higher megakaryocyte ploidy (Martin et al. 1982; Harker and Finch 1969). However there appears to be a temporal delay of 24-48 hours between drop in platelet count and maximal megakaryocyte ploidy (Kuter, Beeler, and Rosenberg 1994; Kuter and Rosenberg 1995) indicating the need for DNA replication and megakaryocyte maturation.

Taken together these evidence indicate that there is a physiological homeostatic control system that feeds back to the megakaryocyte and even possibly to the HSC to promote an increase in platelet production in states of increased haemostatic demand, when functional circulating platelet mass drops and results in the acute release of platelets of greater volume.

This complex feedback system is in part mediated by TPO which binds to receptors on platelets, megakaryocytes and even HSCs and has been implicated in the maintenance of steady state peripheral platelet mass. The exact mechanism by which TPO regulates megakaryopoiesis, megakaryocyte maturation and subsequent thrombopoiesis, however, remains unclear. TPO levels have been shown to be inversely proportional to platelet mass

(Nichol et al. 1995; Shinjo et al. 1998; Engel et al. 1999; Kuter and Rosenberg 1995), with peripheral TPO concentrations rising to a maximum approximately 24 hours after acute platelet loss, followed by an increase in megakaryocyte ploidy levels (Kuter and Rosenberg 1990; Kuter, Beeler, and Rosenberg 1994; Hirayama et al. 1998). A major component of TPO regulation is by the “sponge model” in which TPO production from the liver is constant (Stoffel, Wiestner, and Skoda 1996; Ulich et al. 1995) and the concentration of circulating TPO is determined by the level of TPO binding to MPL, its internalisation and degradation within circulating platelets. Thereby loss of platelet mass leads to increased circulating TPO levels and result in activated platelet production.

However, regulation of platelet mass is likely to be more complex as this model does not hold true in all scenarios. Firstly, TPO levels are not increased in chronic thrombocytopaenias (Kosugi et al. 1996; Ichikawa et al. 1996). Secondly, in states of reactive thrombocytosis, TPO concentrations are higher than expected, with an increase in hepatic *TPO* mRNA levels secondary to increased concentrations of the acute phase cytokine IL-6 (Hsu et al. 1999; Griesshammer et al. 1999; Wolber et al. 2001; Kaser et al. 2001). Thirdly, in severe thrombocytopaenia bone marrow stromal cells dramatically increase *TPO* mRNA levels (Sungaran, Markovic, and Chong 1997). Furthermore, there is evidence to show that TPO is not indispensible for platelet production (Ng et al. 2014) with genetic ablation of *TPO* or *MPL* receptor leading to severe but not absolute thrombocytopaenia (Gainsford et al. 1998; Q. Chen et al. 1998). CXCL12 has been conceptualised to be responsible for TPO independent thrombopoiesis. It enhances megakaryocyte colony formation *in vitro* (Hodohara et al. 2000) and administration of CXCL12 in combination with FGF4 can nearly normalise the platelet count of MPL deficient mice (Avecilla et al. 2004). To add to this complexity, as described above, megakaryocytes within the stem cell niche influence HSC quiescense through PF4 (Bruns et al. 2014) and TGFB (Zhao et al. 2014).

In states of acute increased haemostatic demand such as haemorrhage or acute severe thrombocytopaenia where the rapid generation of large reactive platelets is required it has been suggested that this may occur via a slow or fast response (Martin et al. 2012). The evidence presented above demonstrates an increase in megakaryocyte size and ploidy in settings of acute thrombocytopaenia but this occurs 24-48 hours after the drop in platelet mass in response to rising TPO levels which could represent the slow response, part of the physiological feedback that maintains steady state thrombopoiesis. However, there have been studies to show immediate increases in platelet mass after induced acute thrombocytopaenia without appreciable increases in megakaryocyte ploidy (Corash et al. 1987; Stenberg and Levin 1989) which suggests that in a pathological setting a fast response is activated for rapid thrombopoiesis. At the level of the megakaryocyte this could be from all levels of megakaryocyte ploidy or as a result of recruitment of a subset of megakaryocytes with a specialised transcriptional program or those of higher ploidy only. This feedback may

also occur at the level of the HSC. As described above it is now evident that megakaryocyte-biased HSC compartment can give rise to megakaryocyte progenitors directly bypassing intermediate differentiation steps as well as through the bipotent MEP via the classical pathway which could represent the fast and slow responses respectively (Figure 1.7). Recruitment of this subset of megakaryocyte lineage primed HSCs, that express high levels of CD41 and *VWF* and bypass conventional lineage commitment steps, has been demonstrated to function as part of the fast response described as “emergency” or “stress” megakaryopoiesis upon acute inflammatory stress (Haas et al. 2015).

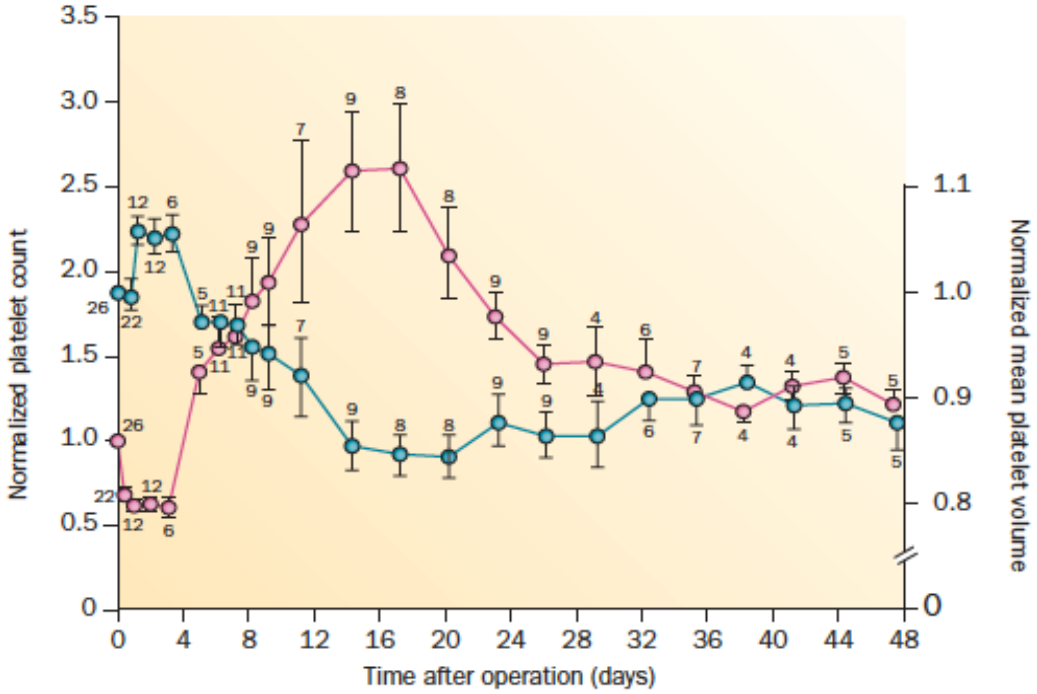


Figure 1.6 Changes in normalised platelet count (pink) and mean platelet volume (blue) following cardiopulmonary bypass surgery. The numbers above each value indicate the number of patients studied at each interval (mean ± SD). (Figure from Martin et al. 2012).

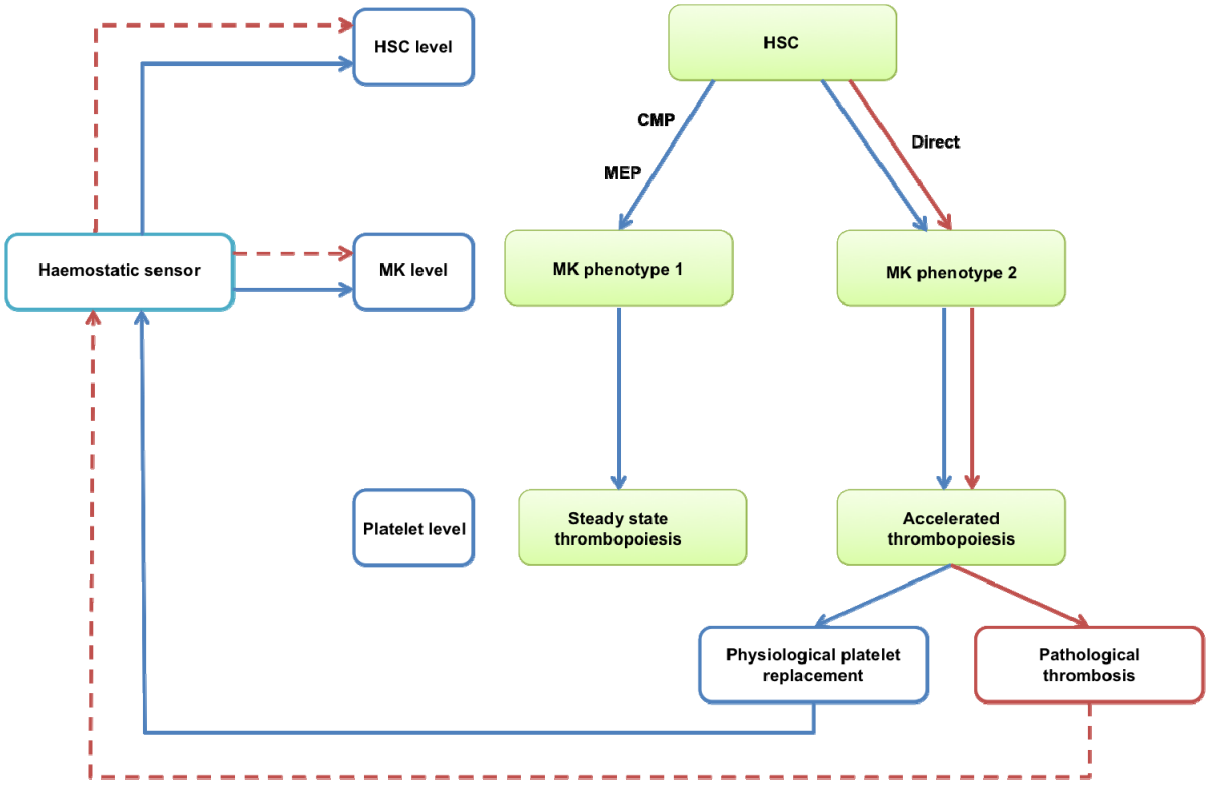


Figure 1.7 Steady state and accelerated thrombopoiesis.
A haemostatic stimulus is hypothesised to lead to thrombopoiesis via a haemostatic sensor. This may be in the physiological steady state where or in states of increased haemostatic demand where accelerated thrombopoiesis is required to restore physiological platelet mass. It could also occur in pathological states of thrombosis where an inappropriate stimulus leads to accelerated thrombopoiesis, which could then go on to produce an inappropriate positive feedback loop. It is hypothesised that feedback could occur at the level of either the HSC or megakaryocyte. There is now evidence to suggest that there may be 2 separate pathways for megakaryocyte differentiation from the HSC, one through the classical differentiation pathway via intermediate progenitors (CMP, MEP) which could be suggested to be the "slow" or steady state pathway, and the other a direct differentiation from the HSC which bypasses intermediate progenitor states which could be suggested to be the "fast" or accelerated pathway. It is possible that the 2 differentiation pathways produce 2 different megakaryocyte phenotypes, one which is primed for emergency thrombopoiesis.
Here red arrows refer to pathological pathways while blue to physiological pathways.

1.4 Megakaryocyte-Platelet axis in acute coronary syndrome

1.4.1 Acute coronary syndromes

Acute coronary syndromes (ACS) is an umbrella term referring to a spectrum of clinical diagnoses based on symptoms, electrocardiographic change and increased circulating cardiac markers (Hamm et al. 2009). It encompasses unstable angina, where there is partial coronary occlusion, and acute myocardial infarction (AMI), defined by total coronary occlusion and thereby myocardial necrosis. AMI is split into non-ST elevation myocardial infarction (NSTEMI), where coronary occlusion leads to partial thickness infarction and ST elevation myocardial infarction (STEMI), where coronary occlusion results in a full thickness infarction. ACS is both common and confers significant morbidity and mortality, despite improvement in prognosis with antiplatelet drugs and primary percutaneous coronary intervention (pPCI). According to the most recent figures approximately 30,000 people still die from AMI per year in the UK; this equates to over one third of AMI patients (Smolina et al. 2012).

Pathological platelet thrombosis resulting in coronary artery occlusion defines ACS. ACS does not represent a single disease with a distinct cellular change or pathophysiological cause, rather several disease processes that lead to formation of intracoronary thrombus as the final common pathway. These include inflammatory cell activation, abnormal lipid metabolism and endothelial cell dysfunction as well as platelet hyperactivity. Platelets are known to be a major constituent of coronary thrombus and have been found histologically in patients who died from ACS. In approximately 70% of cases, observed at post mortem, coronary thrombi develop at the site of ruptured atherosclerotic plaques. In 20-25% cases thrombus occludes the vessel with no plaque rupture (Libby 2013; Falk et al. 2013). This is known as plaque erosion and although it is uncertain exactly what initiates thrombosis, it may well be a result of increased systemic thrombogenicity. The most convincing evidence for the participation of platelets in coronary arterial thrombosis comes from trials of antiplatelet drugs that have shown significant reductions in morbidity and mortality and that remain the cornerstone of ACS treatment (Steg et al. 2012).

1.4.2 Platelet and megakaryocyte characteristics in ACS

A number of clinical observational studies have shown that MPV is increased in ACS. Historical data from the pre pPCI era and the use of P2Y12 or GPIIb/IIIa antagonists shows that MPV is increased within 12 hours after AMI and remains raised for the following four months compared with controls (Martin, et al. 1983; Cameron et al. 1983; Kishk, Trowbridge, and Martin 1985). This relationship has also been observed in the era of routine dual antiplatelet and GPIIb/IIIa inhibitor use (Ranjith et al. 2009; Lippi et al. 2009; H. Chu et al. 2011; S. G. Chu et al. 2010; Mathur et al. 2001; Sansanayudh et al. 2014). The functional

relevance of this well documented finding remains unclear. Platelet function based on bleeding time (Milner and Martin 1985; Kristensen, Bath, and Martin 1990) and surface expression of GPIIb/IIIa receptors (Giles, Smith, and Martin 1994) have been correlated with large platelets in AMI, indicating that the increase in volume may be pathologically important. In support of this, a number of large studies have shown an association between MPV and prognosis; namely mortality and repeat infarction which is independent of other risk factors, both prior to and since the initiation of pPCI, dual antiplatelet and GPIIb/IIIa agents (Taglieri et al. 2011; Aksu et al. 2009; Azab et al. 2011; Goncalves et al. 2011). MPV has been associated with other morbidity endpoints including coronary restenosis (Smyth et al. 1993; A. Yang, Pizzulli, and Lüderitz 2006) and heart failure (Estévez-Loureiro et al. 2009; Pereg, Berlin, and Mosseri 2010; Yazici et al. 2011). Interestingly, in large prospective studies of healthy individuals MPV is found to be the strongest independent risk factor for development of first AMI (Klovaite et al. 2011) and also predict mortality due to coronary artery disease (Slavka et al. 2011). Increased MPV has also been seen in groups with low HDL (Varol et al. 2014), diabetes (Guthikonda et al. 2008; Han et al. 2013) and obesity (Coban et al. 2005).

There is also some evidence to suggest that megakaryocyte activity is altered in the setting of ACS although little direct evidence in the bone marrow. First, IPF is increased with MPV in patients with AMI compared with controls (Cesari et al. 2008; Guthikonda et al. 2008), also in patients with in-stent thrombosis on dual antiplatelet therapy (Würtz et al. 2010), suggesting that the number of reticulated platelets in this disease setting are increased reflecting increased megakaryocyte activity. Furthermore, the level of IPF is shown to be independently predictive of mortality in the same way as MPV (Cesari et al. 2013). Second, not only is the numerical value for MPV increased in AMI, the entire platelet volume distribution curve is shifted to the right, suggesting a change in thrombopoiesis (Trowbridge and Martin 1987; Khandekar et al. 2006). Third, increased circulating levels of TPO associated with increases in MPV have been demonstrated in AMI (Senaran et al. 2001). Finally, while megakaryocyte ploidy has not been measured in AMI, an early study shows increased megakaryocyte volume in AMI patients compared with controls, correlating with increased MPV (12,227fl, n=7; 7,875fl, n=6 respectively; $p<0.001$, t-test). In addition, megakaryocyte volume was increased in bone marrow from post mortem studies of patients dying with sudden cardiac death compared with death through road traffic accident (10,293fl, n=10; 7,368fl, n=11 respectively; $p<0.001$, t-test) (Trowbridge et al. 1984). Increased volume of megakaryocytes in AMI also lends support to increased megakaryocyte activity in this setting.

It is not known if the increase in MPV is an effect of the state of acute thrombosis or if it is causally related to the development of ACS. The possibility that increased platelet volume may be a downstream effect of the inflammatory changes that are seen in ACS, is supported by the fact that pro-inflammatory cytokines also promote megakaryocyte maturation and platelet release. Expression of toll-like receptor 2 (TLR2) has been shown in megakaryocyte

and platelets. These are receptors that recognise molecules released in the circulation from a ruptured atherosclerotic plaque. TLR2 stimulation in the megakaryocyte can influence maturation and potentially influence platelet production and function (Beaulieu et al. 2011). Moreover, binding of HDL to its receptor ABCG4, highly expressed on megakaryocytes, leads to reduced platelet release whereas its absence analogous to low HDL levels seen in atherosclerosis promote an increase in total platelets, reticulated platelets, enhanced activation, and increased platelet–leucocyte aggregates (Murphy et al. 2013). These data suggest that inflammation and processes specific to atherosclerotic plaque rupture might influence megakaryocyte phenotype and MPV. However, given that the lifespan of a platelet is 10 days, and that >90% of platelets measured shortly after AMI would have been circulating before the occlusive event, a causal relationship between ACS and platelet density or MPV has been suggested. This is supported by the prognostic data presented above (Section 1.4.2) that show that a raised MPV increases likelihood of death and re-infarction post event but also is an independent marker of the risk of a primary event in a healthy population.

1.4.3 Alteration of megakaryocyte-platelet activity in ACS

What triggers the upregulation of the megakaryocyte-platelet system to produce large platelets is unknown, however, it is hypothesised that there maybe both genetic and environmental factors.

Genome wide association studies (GWAS) aim to find associations between particular genetic variants and a chosen phenotypic trait. These studies have dramatically increased the number of known disease-associated genetic variants (Hindorff et al. 2009). The largest GWAS for platelet count and MPV is a meta-analysis of 66,867 individuals that identified 43 genetic variants associated with platelet count and 25 associated with MPV (Gieger et al. 2011). One third of these variants were identified within or near to genes encoding proteins of known functional roles in megakaryopoiesis, thrombopoiesis, platelet survival or general haematopoiesis. A number of these genetic loci were also found to reach genome wide significance levels for association with platelet function (Johnson et al. 2010), supporting the link between MPV and platelet function.

A number of GWAS have been carried out to identify genetic variants associated with the endpoint of coronary artery disease (CAD) which is combination of AMI and chronic stable angina/chronic coronary stenosis. In total 83 genetic loci have been identified as reaching genome wide significance (Howson et al. 2017; Deloukas et al. 2013; Kathiresan et al. 2009; IBC 50K CAD Consortium 2011; Samani et al. 2007; Schunkert et al. 2011; Erdmann et al. 2009; Nikpay et al. 2015) including 10 from the recent GWAS meta-analysis of 185,000 CAD cases and controls using phased haplotypes from the 1000 Genomes Project (McVean et al.

2012; Auton et al. 2015) which has expanded coverage of lower frequency variants (Nikpay et al. 2015; Deloukas et al. 2013). These genetic loci have been linked to either a lipid trait, blood pressure or vascular integrity. Only 2 loci are found to overlap with those associated with the platelet phenotypes. This does not appear consistent with the idea of platelets being integral to the development of ACS. Furthermore, the possibility of a causal link between MPV and CAD has recently been challenged by large scale imputational GWAS study of 36 blood cell indices in 173,480 individuals (Astle et al. 2016). Here, through two sample Mendelian randomisation analysis the genetic traits that lead to reduction in MPV rather than increase are common to the development of CAD based on CAD GWAS data (Deloukas et al. 2013). This suggests that previously reported associations between MPV and cardiovascular disease may be non-causal, however, this analysis takes into account genetic signals only.

It must be noted that the variants linked to CAD only account for 10.6% of CAD heritability (Deloukas et al. 2013) indicating that the GWAS data is far from complete. Moreover, the loci discovered in the platelet GWAS were estimated to account for 4.8% of the phenotypic variance observed in platelet count within the population and 9.9% seen in MPV (Gieger et al. 2011). Taken together this supports a more prominent role for environmental determinants in platelet count and MPV as well as the development of CAD. Interestingly in the case of platelet indices, the majority of the GWAS signals are in noncoding regions, supporting the assertion that the principle mechanisms by which sequence variation alters phenotype are at the level of transcriptional regulation and/or splicing events (Gieger et al. 2011).

While regulation of platelet indices is highly heritable (Evans, Frazer, and Martin 1999; Bertin et al. 2007; Mahaney et al. 2005; Bray et al. 2007; C. Garner et al. 2000), there are a number of non-genetic factors that may in part explain trait variance of platelet indices within the population, although the understanding of the non genetic regulation is limited. Both platelet count and MPV vary according to age, gender and ethnicity (Bain 1996; Segal and Moliterno 2006; Saxena et al. 1987; Biino et al. 2011). In the largest study to date, MPV is shown to increase with age (Santimone et al. 2011) and in an elderly population MPV is found to increase with percentage body fat and fasting blood glucose (Muscari et al. 2008), which are also risk factors for acute coronary syndromes. It is certainly possible that the upregulation of the megakaryocyte-platelet system is the result of low grade increases in platelet consumption over a period of time caused by endothelial dysfunction such that is seen in chronic stable atherosclerotic disease or due to environmental coronary artery disease risk markers such as aging, smoking, hypertension or hypercholesterolaemia.

1.4.4 Megakaryocyte-platelet transcriptome studies in ACS

There are a few studies that have investigated platelet gene expression profiles in ACS. Transcriptional profiling of platelets is a novel opportunity to understand alterations in gene

expression preceding the acute coronary event without the confounding possibility that the event itself has provoked those. However, all of these studies are small and underpowered and therefore a robust association has not been confirmed.

In one study 54 transcripts were found to be differentially expressed between 16 patients with STEMI and 44 controls using gene expression arrays (Healy et al. 2006). The genes strongly associated with STEMI were *CD69* and myeloid-related protein *MRP-14*. In an independent validation study, *MRP-8/14* was associated with first cardiovascular event rate, independent of standard risk factors. The same protein has also been found to be predictive of mortality after an acute event (Morrow et al. 2008). In another study, 45 differentially expressed transcripts were identified between 15 ACS patients and 14 controls. Here *GP1BB* was found to be strongly associated with ACS (Colombo et al. 2011). These studies suggest that in ACS patients, platelets are preconditioned at the transcriptional level from the megakaryocyte to a more prothrombotic phenotype.

In another study 63 platelet transcripts were found to be associated with platelet function by assessing differential expression between healthy individuals grouped according to their platelet function (Goodall et al. 2010). These transcripts were then cross-referenced to genetic variants found to be associated with CAD in a larger case control study and expression of the genes *COMMD7* and *LRRFIP1* were found to be associated with CAD. Moreover, when these genes were silenced in zebrafish, thrombus formation was reduced suggesting that these particular transcripts are prothrombotic and potentially pathogenic. Racial difference in risk of atherothrombosis mediated by differences in the thrombin receptor PAR4 activity have also been linked to particular differentially expressed transcripts using transcriptome analysis (Edelstein et al. 2013).

More recently, RNA-seq has been used to identify differentially expressed genes between patients with STEMI compared with NSTEMI. Although none of the transcripts were found to be significantly differentially expressed there were a number that were increased in STEMI compared with NSTEMI with enrichment in platelet related pathways and platelet aggregation pathways (Eicher et al. 2016).

1.5 Application of next generation sequencing in transcriptome analysis

Until recently the majority of studies that have analysed the transcriptome of megakaryocytes and platelet have been based on whole transcriptome expression arrays rather than transcriptional profiling using next generation sequencing. Expression arrays can only give information about known genes by hybridisation to probes designed to detect known transcripts. RNA-seq, however allows the assessment of the full repertoire of known and

novel transcript isoforms and splice junctions. Here RNA-seq libraries are prepared by the conversion of RNA to cDNA, sequencing adaptors are subsequently added to each cDNA fragment and a short sequence is obtained from each cDNA using high-throughput sequencing technology. The resulting sequence reads are aligned with the reference transcriptome, thereby allowing all parts of the transcriptome to be analysed including multiple transcripts for each gene. RNA-seq also allows for detection of alternative splicing which is a post-transcriptional process used in up to 94% of human genes that produces multiple distinct transcripts from a single gene (E. T. Wang et al. 2008; Pan et al. 2008) and may have an important role in disease causation (Singh and Cooper 2012). All considered, RNA-seq provides a more complete assessment of the transcriptome. In order to perform RNA-seq, a pure cell population is required and at the time of starting this thesis within our laboratory approximately 10,000 cells was the number required for robust transcript quantification.

More recent advances in whole-transcriptome amplification and single cell isolation have allowed the high resolution sequencing of the minute amounts of RNA present in a single cell (less than 1 pg of mRNA (Kawasaki 2004)). This offers the opportunity to sequence rare cell populations and allows an insight into trancriptional heterogeneity within a phenotypically homogeneous cell population (S. Huang 2009). The transcriptome of an individual cell reflects its functionality and its responses to acute stimuli. Moreover, in conventional bulk sequencing of cell populations, genes thought to be co-expressed may in fact be mutually exclusive when single cell resolution is applied.

1.6 Project aims

The primary aim of this project was to gain insights into megakaryopoiesis and thrombopoiesis by utilising next generation sequencing to analyse the transcriptome of human HSCs, megakaryocytes and platelets in both the steady state and in states of accelerated thrombopoiesis that occurs due to haemostatic demand.

Single cell RNA sequencing was used to investigate transcriptional heterogeneity within the primary human bone marrow HSC population. It was anticipated that differentially expressed genetic programmes would provide insight into lineage potential of HSC subpopulations and the early stages of megakaryocyte lineage potential and differentiation.

Using low input and single cell RNA sequencing the transcriptome profile of the human bone marrow megakaryocyte was interrogated for the first time. In order to identify novel drivers of megakaryocyte maturation, human bone marrow megakaryocyte transcriptome was compared to that of their *in vitro* counterparts: megakaryocytes cultured from human CD34+ cells, a process known to generate immature megakaryocytes. Furthermore, transcriptional

signatures associated with increasing ploidy were then identified, again in primary bone marrow megakaryocytes.

The pathological setting of acute coronary thrombosis was then used as a model for accelerated thrombopoiesis. Here megakaryocyte and platelet transcriptomes were compared between controls and patients with recent acute coronary thrombosis with the objective to identify transcripts related to thrombosis and accelerated thrombopoiesis. Finally the gene signatures relating to increasing ploidy and accelerated thrombopoiesis were correlated with differentially expressed genetic programmes within the HSC population in order to investigate if thrombopoiesis in different physiological states could be controlled at the level of the HSC.

CHAPTER 2
Materials and Methods

2.1 Sample collection

2.1.1 Patient recruitment

Sixty-six patients admitted either acutely or electively to Barts Health NHS Trust between June 2014 and February 2016 gave written informed consent to enter the ethically approved BAMI Platelet Sub-study (REC London City & East 13/LO/1760, PI Prof. Anthony Mathur, Queen Mary University of London, Appendix 1).

The aim of the study was to obtain blood and bone marrow samples from 2 cohorts of patients: 1. Cases: individuals with recent acute myocardial infarction; 2. Controls: individuals with no evidence of coronary artery disease. Cohort 1: Cases were recruited through the multicentre phase III BAMI trial (The effect of intracoronary reinfusion of bone marrow derived mononuclear cells on all-cause mortality in acute myocardial infarction). Therefore, the main inclusion criteria for this arm were the inclusion criteria for BAMI, but only the individuals randomised to have bone marrow aspiration were recruited to the Sub-study. Briefly, the inclusion criteria were patients within 3-5 days of successfully reperfused STEMI, with a resultant left ventricular ejection fraction of <45%. STEMI was defined as ischaemic chest pain in the context of an ECG demonstrating >1-mm ST segment elevation in at least two contiguous limb leads or >2mm ST segment elevation in at least two contiguous chest leads or new left bundle branch block. Blood and the first 5ml bone marrow aspirate from the posterior iliac crest were collected from these individuals. Cohort 2: Controls were recruited as elective patients attending for non-coronary cardiac surgery. The main inclusion criteria were: no evidence of coronary disease requiring coronary artery bypass grafting on pre-operative invasive or CT coronary angiography, normal pre-operative platelet count, Hb >8.5g/dl and no history or evidence on blood count of haematological malignancy. Blood and bone marrow scraping from sternotomy at the time of surgery were collected from these individuals. Out of 66 patients recruited in total 4 patients were recruited from the multicentre BAMI trial. These 4 individuals were excluded from analysis as slow recruitment to the BAMI trial meant that the numbers required for platelet and megakaryocyte transcriptional analysis were not achievable.

As a result of this, the study design was modified, Cohort 1: Cases, these were recruited as patients admitted acutely with AMI defined as ischaemic chest pain and cardiac troponin rise, who then went on to have coronary artery bypass grafting. Therefore, these patients had an acute event on the background of progressive coronary artery disease. Only patients with normal pre-operative platelet count, Hb >8.5g/dl and no history or evidence on blood count of

haematological malignancy were included. Blood and bone marrow scraping from sternotomy at the time of surgery were collected from these individuals.

In total 66 patients provided informed consent for the BAMI Platelet Sub-study. 4 patients were recruited as BAMI trial patients and therefore excluded. 7 patients gave consent but samples were not taken due to logistical issues. There was not enough bone marrow material from 2 patients and these were therefore excluded. In 7 further patients, bone marrow material was used to optimise the protocol as described in Chapter 4. Blood and bone marrow material was collected from 46 individuals. Of these 19 were from Cohort 1: Cases and 27 were from Cohort 2: Controls. Baseline clinical characteristics were collected from each patient and all data was stored in an anonymised manner.

2.1.2 Adult blood for haematological and platelet measurement

Adult peripheral blood for haematological and platelet measurement was collected from all patients from whom bone marrow was also collected after informed consent and with ethical approval (as above). These were patients undergoing either urgent coronary artery bypass grafting following AMI or non-coronary cardiac surgery at Barts Health NHS Trust. Blood was collected after the patient had been anaesthetised into Vacutainer tubes: 4ml EDTA tube containing 1.8mg/ml K_2EDTA (BD), 4ml serum tube (BD), 22ml into citrate tubes containing 3.2% sodium citrate (BD). The samples were transported to the University of Cambridge at room temperature for further analysis.

Adult peripheral blood from healthy volunteers was collected with ethical approval (REC East of England – Cambridge Central, 12/EE/0040, PI Prof. Willem Ouwehand, University of Cambridge) from Cambridge University Hospitals NHS Trust. This was used either for healthy platelet function as controls for the study patients or for isolation of mononuclear cells for optimisation of single cell RNA-seq library preparation.

2.1.3 Adult bone marrow for megakaryocyte and HSC isolation

2.1.3.1 Optimised method

Adult bone marrow for megakaryocyte and HSC isolation was collected from patients undergoing either urgent coronary artery bypass grafting following AMI or non-coronary cardiac surgery at Barts Health NHS Trust after informed consent and with ethical approval REC (London – City & East) 13/LO/1760, PI Prof. Anthony Mathur, Queen Mary University of London). A sternal bone marrow scraping was taken after median sternotomy using a Volkmann's spoon. The sample was collected into an EDTA Vacutainer tube containing 1.8mg/ml EDTA. 4mL of Dulbecco's phosphate buffered saline (PBS, Sigma) containing 10%

human serum albumin (HSA, Gemini Bio Products) was added and the whole volume was resuspended by pipetting 2-3 times. The sample was then put on metallic thermal beads (ThermoFisher Scientific) at a temperature between 0-4°C and transported to the University of Cambridge for further processing.

2.1.3.2 Other trialled methods

1. Bone marrow was taken as a 5mL aspirate from the posterior iliac crest from 4 patients who were originally recruited from the phase III multicentre BAMI trial. The 5mL aspirate was collected into an EDTA Vacutainer tube containing 1.8mg/ml EDTA. The sample was transported at room temperature to the University of Cambridge for further processing. These samples were removed from analysis as described in Section 2.1.1, as well as the poor yield of megakaryocytes from aspirates compared with bone marrow scrapings.

2. Sternal bone marrow scraping was collected into an EDTA Vacutainer containing 1.8mg/ml EDTA and 4mL PBS with varying concentrations of HSA: 2%, 5% and 10%.

3. Sternal bone marrow scrapings were collected into RNAlater (Ambion) and transported to the University of Cambridge for further processing. These samples could not be processed further due to cellular breakdown.

2.1.3.3 Human bone marrow CD34+ cells for HSC isolation

Cryopreserved human bone marrow CD34+ cells for HSC sorting and use in xenotransplantation experiments were purchased from Lonza, cat. 2M-101C (Lot no: *0000500272, 0000312253, 0000312415*).

2.1.4 Adult blood for CD34+ cell isolation

Adult peripheral blood cells from apheresis filter of healthy platelet donors belonging to the NHS Blood and Transplant, Cambridge were collected for CD34+ cell isolation. The blood donor consent procedure included the provision for such surplus materials to be used in an anonymised manner for research. After full platelet donation, the apheresis filter typically contained approximately the same number of blood mononuclear cells as 1 unit of blood (2-5x10^9 mainly leucocytes).

2.1.5 Animals used for megakaryocyte viability experiments

Nbeal2$^{+/-}$ mice (heterozygous) (Wellcome Trust Sanger Institute) were used for harvesting of bone marrow for megakaryocyte viability experiments. As with human heterozygous carriers

of the NBEAL2 mutations, no phenotypic differences have been observed between wild-type mice and Nbeal2$^{+/-}$ mice. The mice were provided by Dr C Ghevaert's Group, University of Cambridge (Nbeal2$^{+/-}$ mouse model was generated by the Wellcome Trust Sanger Institute). Animal experiments were performed in accordance to institutional guidelines approved by University of Cambridge Animal care committee.

2.1.6 Animals used for xenotransplantation studies

NOD/LtSz-scidIL2Rg^{null} mice (NSG) (Jackson Laboratory) were used for xenotransplant experiments. They were bred and housed at the Cancer Research Institute/Cambridge Institute for Medical Research animal care facility, Addenbrooke's Hospital, University of Cambridge. Animal experiments were performed in accordance to institutional guidelines approved by University of Cambridge Animal care committee.

2.2 Whole blood measurements

2.2.1 Platelet physical parameters

Peripheral blood samples were collected into EDTA containing 1.8mg/ml K_2EDTA. Platelet characteristics and haematological parameters were measured using an XE-5000 haematology analyser (Sysmex) within 3 hours of blood collection. Along with the standard haematolgy parameters measured, the platelet parameters obtained from the analyser by measures of impedance included: platelet count (PLT, 10^9/L), plateletcrit (PCT, %) a measure of total platelet mass, mean platelet volume (MPV, fL) calculated as plateletcrit divided by platelet count, platelet distribution width (PDW, fL) the width of platelet size distribution, platelet large cell ratio (P-LCR, %) the ratio of platelets falling above the 12fL size cut-off. Finally, immature platelet fraction (IPF, %) was also measured. The analyser utilises automated flow cytometry and polymethine and oxazine fluorescent dyes that label nucleic acids thereby identifying RNA-containing (reticulated) platelets.

2.2.2 Platelet function

Platelet activation and function work was performed in collaboration with Dr K Downes group, Department of Haematology, University of Cambridge.

Platelet activation and function were measured by flow cytometry in whole blood collected in 3.2% sodium citrate. Blood was processed at exactly 4 hours after collection. 5uL 10mM Aspirin (Sigma) and 5uL 1000U/mL hirudin (Sigma) were added to 500uL citrated whole blood. 5uL of this was added to 45uL of Hepes Buffered Saline (HBS) containing either 1.

3ug/ml crosslinked collagen related peptide (CRP-XL, gift from Richard Farndale, Biochemistry, University of Cambridge NHS Trust)/4U/mL apyrase (Sigma) or 2. 5uM adenosine diphosphate (ADP, Sigma) or 3. 8uM thrombin receptor activating peptide (TRAP-6, Tocris)/4U/mL apyrase and either FITC–anti fibrinogen (Dako) or PE–anti-CD62P (P-selectin, IBGRL Research Products, Bristol). Samples were incubated for 20 min at room temperature with and the reaction was stopped by 100-fold dilution in formyl saline (0.2% formaldehyde/0.9% NaCl, Sigma). Negative controls for the P-selectin antibodies were set using an appropriate isotype control and for the anti fibrinogen using samples incubated with the antibody in the presence of 10mM EDTA. [HBS: 150mM NaCl (Sigma), 5mM KCl (AnalaR BDH), 1mM MgSO4.7H20 (AnalaR BDH), 10mM Hepes (Sigma)].

Flow cytometric analysis was carried out on a FC500 flow cytometer (Beckman-Coulter Ltd). Platelets were identified by light scatter, and data was displayed as a histogram of fluorescence intensity against count and a threshold was set on the flow cytometer at 1% using platelets labelled with the respective PE and FITC negative control. Platelet activation was measured by resting surface expression of P-selectin. Platelet function results were recorded as the percentage of platelets positive for the relevant activation marker after incubation normalised for resting activation levels. All tests were performed in duplicate, and the average result were used for analysis.

2.3 Cell isolation

2.3.1 Isolation of platelets from peripheral blood

To isolate platelets free from contaminating white blood cells a leucodepletion strategy was used, involving centrifugation steps to isolate platelet rich plasma (PRP) and anti-CD45 magnetic beads were applied.

22ml whole blood taken in 3.2% citrate vacutainer tubes. These were centrifuged at 150g for 20 minutes with the brake off. The upper layer (PRP) was harvested and transferred to a fresh 15ml Falcon tube and centrifuged at 150g for 10minutes. The PRP layer was removed and transferred again to fresh 15ml Falcon tube and centrifuged for a third time at 150g for 10minutes. The PRP layer was removed and mixed at room temperature for 20 minutes with 50ul CD45 Dynabeads (Life technologies) per 2ml PRP. The Dynabeads had been washed in 2ml PBS. White cells bound to the CD45 beads were then captured by placing the 15ml Falcon tube in a magnetic rack. The leucodepleted PRP was then harvested into a new 15ml Falcon tube and the platelets were pelleted by centrifugation at 1500g for 10 minutes. The supernatant was carefully discarded and the platelets were lysed with the addition of 1ml

Trizol reagent (Life Technologies) and stored at -80°C ahead of RNA extraction (Section 2.3.5).

2.3.2 Isolation of mononuclear cells from peripheral blood

This was performed for 2 purposes: 1. For isolating CD34+ cells for megakaryocyte cultures (from apheresis filters of platelet donors), 2. For testing and optimising the single cell RNA-seq library preparation method (Whole blood from volunteers).

Blood was processed within 18 hours of collection. Adult blood from apheresis filters was diluted 1:20 with PBS, containing 1% Human Serum Albumin (HSA, Gemini Bioproducts) and 100mM trisodium citrate (Sigma). Whole peripheral blood was diluted 1:1 with PBS containing 1% HSA and 100mM trisodium citrate. This diluted blood was layered over Ficoll-Paque Plus (GE Healthcare), centrifuged at 800g for 15 minutes and the mononuclear cell layer recovered and washed.

2.3.3 Isolation of megakaryocytes from human bone marrow

2.3.3.1 Optimised method: Isolation of megakaryocytes by FACS

Processing of clinical bone marrow took place 2-3 hours after harvest. The cellular content was flushed out of the bone marrow using PBS containing 1.2% HSA, 2mM EDTA (Sigma) and the red cells were lysed using ammonium chloride lysis.

The cells were stained for surface megakaryocyte markers with mouse anti-human CD41a APC conjugated antibody (BD) and mouse anti-human CD42b PE conjugated antibody (BD) and for ploidy analysis with 1ug/ml Hoechst 33342 (Invitrogen). Hoechst DNA stain was used for ploidy quantification rather than other DNA stains such as 4',6-Diamidino-2-Phenylindole (DAPI) due to its ability to stain nuclear DNA without disrupting the cell or nuclear membrane and without the requirement for cell fixation which would prohibit the transcriptional analysis. After incubation at 37°C for 30 minutes the cells were kept at 4°C before sorting using a FACS Aria Fusion flow sorter (BD). Due to fragility cells were only centrifuged once at 120g for 8 minutes with the brake off. Megakaryocytes were FACS sorted approximately 5 hours after harvest.

Single cells and megakaryocyte pools of 20-100 cells (hereafter multicellular pools) were sorted by FACS using a 100uM nozzle directly into individual wells of a 96-well plate each containing 2.5ul lysis buffer/well: 950uL RLT Plus buffer (Qiagen) and 50ul 20U/ul SUPERase In RNase inhibitor (Ambion) and stored at -80°C until library preparation. Due the time taken for single cell FACS and 100uM nozzle FAC sorting took place over approximately 3 hours.

Flow cytometry data for all samples was collected including frequency of megakaryocytes (defined as CD41+CD42+Hoechst+), frequency of 2n, 4n, 8n 16n, 32n cells. FACS sorting also allowed both single cell and multicellular pools to be collected according to ploidy level.

2.3.3.2 Other trialled methods

1. Density centrifugation of megakaryocyte population from bone marrow aspirates/scraping
Whole bone marrow either aspirate or cellular content of scraping was diluted 1:1 with PBS containing 3% HSA, 75mM trisodium citrate, 5mM EDTA (Sigma) and mixed with Percoll (Sigma) to a final density of 1.020g/ml. This was then layered over Percoll at a density of 1.050g/ml and over layered with PBS containing 0.15% HSA, 75mM trisodium citrate, 5mM EDTA. Following centrifugation at 400g for 20 minutes, the upper layer was removed and discarded. The middle Percoll layer and the cells at the lower interface were recovered and washed at 120g for 8 minutes.
While this provided an enriched population of megakaryocytes, it was not pure. Following Percoll separation it was then possible to sort cells by FACS into lysis buffer and store at -80°C until library preparation, however the time delay in sorting cells into lysis buffer and the cell loss due to processing was not thought to be ideal.

2. Cell picking
Individual megakaryocytes were picked under a x40 objective inverted light microscope using a glass pipette mounted on a micromanipulator and deposited into lysis buffer and stored at -80°C until library preparation. This was trialled from whole bone marrow and from Percoll enriched megakaryocyte populations.
The limitation of this method was low cell output. Despite competency and speed in cell picking it could take 2-4 hours to pick 96 single megakaryocytes. Furthermore, megakaryocytes were identified only on the basis of cell size relative to neighbouring cells and it was not possible to gain any ploidy information.

3. Staining whole bone marrow for FACS sorting without red cell lysis
A side-by-side analysis was made between whole bone marrow FACS with or without red cell lysis. The benefits of not performing red cell lysis including one less processing step thereby leaving the cells as undisturbed as possible and saving approximately 10minutes in processing time before taking bone marrow to be sorted. However, the overwhelming numbers of contaminating red cells at the sort prolonged the sort time by approximately 100% compared with the bone marrow that underwent red cell lysis. The overall numbers of megakaryocytes were the same in both, the ploidy distribution was the same and there were no differences in the quality of cDNA reverse transcribed at the single cell RNA-seq library preparation stage. Therefore, a red cell lysis step remained in the optimised method.

4. Isolation of megakaryocytes 1 day or more after bone marrow harvest

Due to logistical issues this was trialled, however it was felt that due to low input rather than bulk RNA-seq, a time delay in sorting cells into lysis buffer might compromise sequencing quality. Therefore, bone marrow was processed as soon as logistically possible after harvest (within 2-3 hours, cells sorted into lysis buffer between 5-8 hours).

5. Use of different cell lysis buffers

A number of different lysis buffers were trialled to optimise quality of cDNA produced at the single cell RNA-seq library preparation stage. A total volume of 2.5uL/well was used.

1. Smart-seq2 (Picelli2014): 950uL 0.2% Triton X-100 (Sigma) and 50ul 2U/ul Recombinant RNase inhibitor (Clontech). Good quality cDNA from peripheral blood mononuclear cells, K562 cell line and cultured megakaryocytes at the single cell RNA-seq library preparation stage. cDNA from bone marrow megakaryocytes was degraded.
2. Smartseq2 lysis buffer modification: 950uL 0.2% Triton X-100 (Sigma) and 50ul 20U/ul SUPERase In RNase inhibitor (Ambion). Good quality cDNA from peripheral blood mononuclear cells, K562 cell line and cultured megakaryocytes at the single cell RNA-seq library preparation stage. cDNA from bone marrow megakaryocytes was degraded.
3. Smartseq2 lysis buffer modification: 950uL 0.5% NP-40 (Sigma) and 50ul 20U/ul SUPERase In RNase inhibitor (Ambion). Poor quality cDNA produced for all cell types trialled at the single cell RNA-seq library preparation stage.
4. G&T-seq (Macaulay et al. 2015): 950uL RLT Plus buffer (Qiagen) and 50ul 20U/ul SUPERase In RNase inhibitor (Ambion). Good quality cDNA produced from bone marrow megakaryocytes at the single cell RNA-seq library preparation stage. This was therefore used in the optimised method.

2.3.3.3 Megakaryocyte imaging

Bone marrow smears were produced for each bone marrow sample. To analyse the enriched bone marrow megakaryocyte population, Cytospins were produced on glass slides (Thermo Scientific) or the cell suspension was allowed to settle for 1 hour at 37°C on Polysine microscope adhesion slides (Thermo Scientific) within an area drawn using a water-repellent pen (Dako). To analyse cultured megakaryocyte morphology, Cytospins were generated on glass slides on the day of the culture harvest.

To assess cell morphology, slides were stained with Rapid Romanowsky stain (TCS Biosciences) and observed by light microscopy.

Immunofluorescent staining of the enriched bone marrow megakaryocyte population was performed using a VWF surface stain and a 4',6-Diamidino-2-Phenylindole (DAPI) DNA stain. The slide was fixed with 2% paraformaldehyde, washed with PBS containing 0.01% tween

and then stained with a 1/200 dilution of the primary antibody: unconjugated rabbit anti-human VWF antibody (Abcam). The slide was washed and stained with a 1/500 dilution of the secondary antibody: FITC conjugated donkey anti-rabbit antibody (Abcam). After a further washing the slide was mounted with DAPI mounting media (VectaShield) and a coverslip. The slide was then analysed on an immunofluorescent microscope.

2.3.3.4 Megakaryocyte viability experiments using mouse bone marrow

Viability of *ex-vivo* megakaryocytes was examined in mouse bone marrow from Nbeal2$^{+/-}$ (heterozygous) male mice, 2-4 months old. 2 mice were sacrificed for harvesting of bone marrow. Bone marrow was flushed out of 2 femurs from each mouse into either: CATCH medium (Calcium and magnesium free Hanks balanced salt solution (Gibco), 2.9mM sodium citrate (Sigma), 1.4mM adenosine (Sigma), 2.7mM theophylline (Sigma), 3.5% bovine serum albumin (Sigma), 1ug/ml apyrase (Sigma), 1ug/ml PGE1 (Sigma)) or PBS with 10% HSA at either room temperature or on ice. Conditions of the experiment included room temperature/on ice and readouts at different time points up to 10 hours post bone marrow harvests. The following cell stains were used to assess viability:

1. 0.4% Trypan Blue (Sigma): cell aliquot stained 1:1, observed under x10 objective light microscope, live cells counted, megakaryocytes were identified by size. Impermeable dye excluded by live cells.
2. 7-AAD (Life Technologies): cells stained with 7AAD (cross over with PerCPCy5.5 488/695), and APC conjugated anti-CD41(BD)/Hoechst to identify megakaryocytes. Impermeable dye excluded by live cells. Flow cytometric analysis was carried out on a Gallios flow cytometer (Beckman-Coulter).
3. PI (BD): cells stained with PI (cross over with PerCPCy5.5 488/695), and APC conjugated anti-CD41/Hoechst to identify megakaryocytes. Impermeable dye excluded by live cells. Flow cytometric analysis was carried out on a Gallios flow cytometer (Beckman-Coulter).
4. Zombie Far Red (Life Technologies): cells stained with Zombie Far Red (cross over with APC 633/660), and PE conjugated anti-CD41(BD)/Hoechst to identify megakaryocytes. Impermeable dye excluded by live cells. Flow cytometric analysis was carried out on a Gallios flow cytometer.
5. Calcein (Life Technologies) cells stained with Calcein (cross over with FITC 488/525), and APC conjugated anti-CD41/Hoechst to identify megakaryocytes. Actively taken up by living cells due to intracellular esterase activity, excluded by dead cells. Flow cytometric analysis was carried out on a Gallios flow cytometer.

2.3.4 Isolation of HSCs from human bone marrow

2.3.4.1 HSCs from clinical human bone marrow samples

Processing of clinical bone marrow occurred 2-3 hours after harvest. The cellular content was flushed out of the bone marrow using PBS containing 1.2% HSA, 2mM EDTA (Sigma) and the red cells were lysed using ammonium chloride lysis.

In order to identify the HSC population as defined as: Lineage-, CD34+, CD38-, CD45RA-, CD90+, CD49f+ and measure surface expression of VWF, cells were stained with the following antibody cocktail:

PECy5 conjugated anti-lineage specific antibodies: CD2 (BD), CD3 (BD), CD10 (BD), CD11b (BD), CD11c (BD), CD19 (BD), CD20 (BD), CD56 (BD), biotinylated CD42b (Pab5, NHS Blood and Transplant, International Blood Group Reference Laboratory [IBGRL]) and biotinylated GP6 (Pab5, NHS Blood and Transplant, International Blood Group Reference Laboratory [IBGRL]) both used in combination with PECy5 conjugated streptavidin (Biolegend).
Alexa Fluor 700 conjugated anti-CD34 (BD)
PerCP-Cy5.5 conjugated anti-CD38 (BD)
Pacific Blue conjugated anti-CD45RA (Invitrogen)
PECy7 conjugated anti-CD90 (BD)
PE conjugated anti-CD49f (BD)
FITC conjugated anti-VWF (Abcam)

Stained cells were kept at 4°C and sorted using a FACS Aria instrument (BD). Single HSCs defined as Lineage-, CD34+, CD38-, CD45RA-, CD90+, CD49f+ cells were sorted by FACS directly into individual wells of a 96-well plate each containing 2.5ul lysis buffer/well: 950uL RLT Plus buffer (Qiagen) and 50ul 20U/ul SUPERase In RNase inhibitor (Ambion) and stored at -80°C until library preparation. Index sort data was collected for each single cell (Osborne 2011).

2.3.4.2 HSCs from purchased human bone marrow CD34+ cells (Lonza)

Cryopreserved vials of human bone marrow CD34+ cells purchased from Lonza were used to isolate HSCs for xenotransplantation experiments. These were thawed in a waterbath set at 37°C and thawing media consisting of PBS with 5% foetal calf serum (FCS, Gibco) at 37°C was added drop by drop to make up to a 1:10 cell dilution. Cell from 3 vials were combined at this stage and centrifuged at 500g for 5minutes at room temperature followed by resuspension in 150ul PBS/5%FCS. The cells were then stained as described in Section

2.3.4.1 with the addition of the viability stain Zombie Aqua (LifeTechnologies): excitation/emission: 405/516.

Stained cells were kept at 4°C and sorted using a FACS Aria instrument (BD). Single HSCs defined as viable, Lineage-, CD34+, CD38-, CD45RA-, CD90+, CD49f+ cells were sorted by FACS directly into 2 Eppendorf tubes each containing 500ul PBS/5%FCS. VWF+ HSCs were FACS sorted into 1 Eppendorf tube and VWF- HSCs were sorted into the other ahead of transplantation into 10 sub lethally irradiated mice.

2.3.5 Cultured cells

2.3.5.1 K562 culture

Cells from the lymphoblastic cell line K562 (ATCC) were maintained at 37°C in RPMI media 1640 (Sigma) containing 10% foetal bovine serum (Sigma), 500µl Tylosin solution (Sigma) and 6ml MEM Amino acids with L-Glutamine (Sigma). For the purposes of testing and optimising the single cell RNA-seq library preparation method, they were washed resuspended in media and were manually picked into lysis buffer and stored at -80°C until library preparation.

2.3.5.2 Megakaryocyte culture from adult blood CD34+ cells

CD34+ cells were isolated from adult blood mononuclear cells using the human CD34 isolation kit (Miltenyi) with an autoMACS separator (Miltenyi). To generate megakaryocytes, CD34+ cells were cultured for 10 days in CellGro (CellGenix) and supplemented with 100ng/ml TPO (CellGenix) and 10ng/ml IL1β(R&D). For each experiment, cells were seeded in suspension plates at 1 $x10^5$ cells/ml. Media and cytokines were rejuvenated at days 3 and 6. At day 10 cells were counted and harvested. Mature megakaryocytes were immunopurified using a mouse anti-human CD42b PE conjugated antibody (Pab5, NHS Blood and Transplant, International Blood Group Reference Laboratory [IBGRL]) and a PE-positive selection kit. Cells were lysed using TRIzol reagent (Life Technologies) and stored at -80°C ahead of RNA extraction, library preparation and RNA-seq. A small sample was also taken for morphological characterisation and flow cytometry.

2.3.6 RNA extraction

For bulk RNA-seq of using Poly-A selection for both cultured megakaryocytes and platelets as well as rRNA depletion for platelets, cells were lysed in TRIzol reagent and stored at -80°C prior to library preparation. RNA was then extracted from TRIzol reagent as per manufacturer's instructions. Briefly, each sample was brought to room temperature and the

RNA was separated using 0.2ml chloroform (Sigma) per 1ml TRIzol and centrifugation in a phase-lock tube (Eppendorf). RNA was precipitated by adding 10ug RNase-free glycogen (Invitrogen) and 0.5ml 100% isopropanol (Sigma) and centrifugation. The RNA was washed twice with 1ml 70% ethanol (Sigma), allowed to air-dry and then resuspended in RNase-free water. The sample was then stored at -80°C until library preparation. 1µl was used for accurate RNA quantification using the Qubit 2.0 Fluorometer (Invitrogen) and Qubit RNA assay kit (Invitrogen). 1µl was used for checking the quality of the sample using the Agilent 2100 Bioanalyzer (Agilent Technologies) and the Agilent RNA 6000 Pico Kit (Agilent Technologies) using the RNA integrity number (RIN).

In the process of optimising the single cell RNA-seq library preparation method, the Single Cell RNA Purification Kit (Norgen Biotek) was trialled as per manufacturer's instructions for single, 20, 50 and 100 peripheral blood mononuclear cells that had been FACS sorted. Briefly, each sample was sorted into wells containing 100ul lysis buffer provided with the kit. A further 100ul of 70% ethanol was added and this volume applied to 2 columns with a final elution step.

2.4 Analysis of gene expression

2.4.1 Poly(A) RNA-seq library preparation for bulk cultured megakaryocytes

Non-strand specific poly(A) RNA libraries were prepared using the SMARTer Ultra Low Input RNA for Illumina Sequencing – HV kit (Clontech) following manufacturer's instructions. This was used for cultured megakaryocytes.

The input RNA quantity was 10ng. ERCC Spike-In RNA (Ambion) was added to provide artificial transcripts of known concentrations for RNA-seq analysis (Loven et al. 2012) at a concentration of 1:1000.

First strand cDNA synthesis was performed by adding 5xfirst strand buffer, 100mM DTT, 20mM dNTP mix, the Smarter transcript switching oligo, RNase inhibitor and Smartscribe reverse transcriptase according to the manufacturer's instructions and incubating at 42°C for 90min and the 70°C for 10min. A 1:1 Ampure XP (Beckman Coulter) bead clean-up was performed. The Advantage 2 PCR kit was used to perform global amplification of the cDNA by PCR by adding 10x Advantage 2 PCR buffer, Smart PCR primer, 10nM dNTP mix and the 50x Advantage 2 polymerase mix according to manufacturer's instructions. The PCR reaction consisted of heating to 95°C for 1 min, 14 PCR cycles (95°C/15 s, 65°C/30 s, 68°C/6 min) and incubation at 72°C for 10 min. A 1:1 Ampure XP (Beckman Coulter) bead clean-up was performed on the PCR product. The resultant product was quantified using the Qubit 2.0

Fluorometer and a Qubit High Sensitivity DNA assay kit (Invitrogen). This was followed by assessment of size distribution using an Agilent high-sensitivity DNA chip (Agilent technologies).

RNA-seq libraries were prepared from 1ng of amplified cDNA using the Nextera XT DNA sample preparation kit (Illumina). This included a tagmentation step and 14 cycle PCR amplification of adapter-ligated fragments. Thereafter, a final 1:1 Ampure XP bead clean-up was performed. Final library concentrations were quantified by real-time quantitative-polymerase chain reaction (RT qPCR) using a KAPA Library Quantification kit (KAPA Biosystems) according to manufacturer's instructions. Each library was checked for fragment length using an Agilent High Sensitivity DNA chip prior to RNA-seq.

2.4.2 Poly(A) RNA-seq library preparation for Smart-seq 2

RNA libraries were prepared according to the published Smart-seq 2 protocol (Picelli et al. 2014). This method was designed to produce full length non-strand specific poly(A) RNA libraries from single cells and low input purified RNA and is a modification of the SMARTer Ultra Low Input RNA for Illumina Sequencing kit described in Section 2.4.1.

The input RNA quantity was 1ng and the volume made up to 2.5ul. ERCC Spike-In RNA was added to this in a dilution of in a dilution of 1:4,000,000. After hybridisation with 1ul oligo-dT primer: 5'-biotin-triethyleneglycol-AAGCAGTGGTATCAACGCAGAGTACT$_{30}$VN-3' where V is A/C/G and N is any base (IDT) and 1ul 10mM dNTP (Thermo Scientific) at 72°C for 3 min, the reverse transcription mix was added for first strand cDNA synthesis. The reverse transcription mix included: 0.5ul 200U/µl SuperScript II reverse transcriptase (Life technologies), 0.25ul 20U/ul SUPERase In RNase inhibitor, 2ul 5x Superscript II First-Strand Buffer (Life technologies), 0.5ul 100mM DTT (Life technologies), 2ul 5M betaine (Sigma), 0.06ul 1M $MgCl_2$ (Life technologies), 0.1ul 100mM Template-Switching Oligo: 5'-AAGCAGTGGTAT CAACGCAGAGTACrGrG+G-3' where rG is a ribo-guanosine and +G is a locked nucleic modified guanosine (Exiqon) and 5.70ul nuclease free water. cDNA first strand synthesis was performed by incubating at 42°C for 90 min, followed by 10 cycles (50°C/2 min, 42°C /2 min) and 70°C for 15 min.

Global amplification of the cDNA by PCR was then performed by adding PCR mastermix consisting of: 12.5ul 2x KAPA HiFi HotStart ReadyMix (KAPA Biosystems), 0.25ul 10mM Smart PCR primer: 5′ -AAGCAGTGGTATCAACGCAGAGT-3′ (Biomers) and 2.25ul nuclease-free water to the reverse transcription reaction mixture. The PCR reaction consisted of heating to 98°C for 3 min, 20 PCR cycles (98°C/20 s, 67°C/15 s, 72°C/6 min) and incubation at 72°C for 5 min. A 1:1 Ampure XP (Beckman Coulter) bead clean up was

performed on the PCR product. This was followed by assessment of size distribution using an Agilent high-sensitivity DNA chip (Agilent technologies).

RNA seq libraries were prepared from between 0.5 and 10ng of amplified cDNA using the Nextera XT DNA sample preparation kit (Illumina). Including a tagmentation step and 12 cycle PCR amplification of adapter-ligated fragments. Thereafter, a final 1:1 Ampure XP bead clean up was performed on each sample. Final library concentrations were quantified by real-time quantitative-polymerase chain reaction (RT qPCR) using a KAPA Library Quantification kit (KAPA Biosystems) according to manufacturer's instructions on the StepOnePlus system (Applied Biosystems). Each library was checked for fragment length using an Agilent High Sensitivity DNA chip prior to RNA-seq.

2.4.3 Poly(A) RNA-seq library preparation for G&T-seq

Single cells or mini-bulk samples (1-100 cells) were sorted into individual wells of a 96-well plate each containing 2.5ul lysis buffer/well: 950uL RLT Plus buffer (Qiagen) and 50ul 20U/ul SUPERase In RNase inhibitor (Ambion) and stored at -80°C until library preparation.

ERCC spike-in RNA (Ambion) was added to the lysis buffer in a dilution of 1:4,000,000. cDNA synthesis and poly(A) enrichment was performed following the G&T-seq protocol (Macaulay et al. 2015), a variation of the Smart-seq2 protocol (Picelli et al. 2014). Using this protocol genomic DNA and mRNA were separated; the genomic DNA stored and the mRNA amplified. Briefly, streptavidin-coupled magnetic beads (Dynabeads, Life technologies) were conjugated to the biotinylated oligo-dT primer: 5'-biotin-triethyleneglycol-AAGCAGTGGTATCAACGC AGAGTACT_{30}VN-3' where V is A/C/G and N is any base (IDT) according to manufacturer's instructions. Initial separation of DNA and RNA was performed using a Biomek FXP Laboratory Automation Workstation (Beckman Coulter). 10ul conjugated beads were added to the cell lysate and incubated for 20minutes at room temperature with continual mixing. This allowed the pull down of the polyadenylated mRNA. The mRNA bound to streptavidin beads was then magnetised to the side of the well and the genomic DNA containing supernatant was transferred to a different plate. The beads were further washed 4 times at room temperature with a wash buffer consisting of 50 mM Tris-HCl pH 8.3, 75 mM KCl, 3 mM MgCl2, 10 mM DTT (Life technologies), 0.5% Tween-20, 0.1x 20U/ul SUPERase In RNase inhibitor.

After this either the Superscript II or the Smartscribe reverse transcription (RT) mix (10ul) was added to the beads. Common to both mixes were: 0.25ul 20U/ul SUPERase In RNase inhibitor, 2ul 5M betaine (Sigma), 0.06ul 1M $MgCl_2$ (Life technologies), 1ul 10mM dNTP mix (Thermo Scientific) and 1ul 100mM Template-Switching Oligo: 5'-AAGCAGTGGTAT CAACGCAGAGTACrGrG+G-3' where rG is a ribo-guanosine and +G is a locked nucleic

modified guanosine (Exiqon). The Superscript II RT mix also contained: 0.5ul 200U/ul Superscript II reverse transcriptase (Life technologies), 2ul 5x Superscript II First-Strand Buffer (Life technologies), 0.5ul 100mM DTT (Life technologies) and 3.59ul nuclease-free water (Life Technologies). The Smartscribe RT mix also contained: 1ul Smartscribe reverse transcriptase (Clontech), 2ul 5x Smartscribe First-Strand Buffer (Clontech), 1ul 20nM DTT (Clontech) and 2.59ul nuclease-free water. cDNA first strand synthesis was performed by incubating at 42°C for 60 min, followed by 50°C for 30 min and 60°C for 10 min with mixing on a Thermomixer (Eppendorf).

Global amplification of the cDNA by PCR was then performed by adding PCR mastermix consisting of: 12.5ul 2x KAPA HiFi HotStart ReadyMix (KAPA Biosystems), 0.25ul 10mM Smart PCR primer: 5′ -AAGCAGTGGTATCAACGCAGAGT-3′ (Biomers) and 2.25ul nuclease-free water to the reverse transcription reaction mixture. The PCR reaction consisted of heating to 98°C for 3 min, 22 PCR cycles (98°C/20 s, 67°C/15 s, 72°C/6 min) and incubation at 72°C for 5 min. A 1:1 Ampure XP (Beckman Coulter) bead clean-up was performed on the PCR product. This was followed by assessment of size distribution using an Agilent high-sensitivity DNA chip (Agilent technologies) and expression of selected transcripts was also examined at this stage using Taqman real-time PCR (Section 2.4.6).

RNA seq libraries were prepared from between 0.5 and 10ng of amplified cDNA using the Nextera XT DNA sample preparation kit (Illumina). Including a tagmentation step and 12 cycle PCR amplification of adapter-ligated fragments. Thereafter, samples from each 96 well plate were pooled using 3ul of PCR product from each sample and a 1:1 Ampure XP bead clean-up was performed. Final library concentrations were quantified by real-time quantitative-polymerase chain reaction (RT qPCR) using a KAPA Library Quantification kit (KAPA Biosystems) according to manufacturer's instructions on the StepOnePlus system (Applied Biosystems). Each library was checked for fragment length using an Agilent High Sensitivity DNA chip prior to RNA-seq.

2.4.4 Ribosomal RNA depleted RNA-seq library preparation for platelets

Strand specific ribosomal RNA depleted libraries were prepared using the KAPA Stranded RNA-seq kit with Riboerase (Kapa Biosystems) following manufacturer's instructions.

The input RNA quantity was 50ng in a volume of 10ul. After hybridisation with 10ul hybridisation oligonucleotides designed to bind to rRNA at 95°C for 2 min the reaction mix was treated with 5ul RNaseH mix to remove rRNA, a 2.2x Ampure XP bead clean up and 22ul DNase mix to remove hybridisation oligonucleotides followed by a second 2.2x Ampure XP bead clean up. The RNA was then eluted from the beads and fragmented by incubation at high temperature before first strand synthesis with 9ul first strand synthesis buffer and 1ul

KAPA script provided in the kit. cDNA first strand synthesis was performed by incubating at 25°C for 10 min, followed by 42°C for 15 min and 70°C for 15 min. 30ul second strand synthesis buffer was then added for a further incubation at 16°C for 60 min converting the cDNA:RNA hybrid to double stranded cDNA followed by a 1.8x Ampure XP bead clean up. 30ul A-tailing master mix including the A-tailing enzyme was then added to the beads with cDNA and incubated at 30°C for 30 min and 60°C for 30 min in order to add dAMP to the 3' ends of the cDNA library fragments. 30ul adapter ligation master mix including DNA ligase enzyme and adapters (TrueSeq LT) was then added where double stranded adapters with a 3' dTMP overhang are ligated to the A-tailed library. This was followed by two 1:1 Ampure XP bead clean ups.

Global amplification of the cDNA by PCR was then performed by adding PCR mastermix consisting of: 25ul 2x KAPA HiFi HotStart ReadyMix, 5ul 10x KAPA Library Amplification Primer mix. The PCR reaction consisted of heating to 98°C for 45 s, 14 PCR cycles (98°C/15 s, 60°C/30 s, 72°C/30 s) and incubation at 72°C for 5 min. A 1:1 Ampure XP (Beckman Coulter) bead clean-up was performed on the PCR product. This was followed by assessment of size distribution using an Agilent high-sensitivity DNA chip (Agilent technologies).

Final library concentrations were quantified by real-time quantitative-polymerase chain reaction (RT qPCR) using a KAPA Library Quantification kit (KAPA Biosystems) according to manufacturer's instructions. Each library was checked for fragment length using an Agilent High Sensitivity DNA chip prior to RNA-seq.

2.4.5 RNA-sequencing

The libraries were sequenced at the Cancer Research Institute sequencing facility on the Addenbrooke's campus. 150 base pair (bp) paired-end sequencing was performed on the Illumina HiSeq 4000 instrument using TruSeq reagents (Illumina), according to manufacturer's instructions.

2.4.6 Taqman Real-time qPCR

The expression of selected transcripts was quantified from 1 to 5ng of amplified cDNA PCR product prior to Nextera XT library preparation in methods 2.4.2 and 2.4.4 by performing Taqman Real-time qPCR. TaqMan Fast Advanced Master Mix (Applied Biosystems) and the listed transcript Taqman probes (Applied Biosystems) were used to perform real-time PCR reactions over 40 cycles according to manufacturer's instructions on the StepOnePlus system. At least 2 technical replicates for each sample were performed.

Selected transcript probes included: Glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*, Hs02758991_g1), Beta-2-microglobulin (*B2M*, Hs00187842_m1), von Willebrand factor (*VWF*, Hs01109446_m1), Mitochondrial genome encoded 16S RNA (*MT-RNR2*, Hs02596860_s1), Pro-platelet basic protein (*PPBP*, Hs00234077_m1), Platelet factor 4 (*PF4*, Hs00427220_g1).

2.5 Xenotransplantation

As described in Section 2.3.4.2 cryopreserved vials of human bone marrow CD34+ cells purchased from Lonza were used to isolate HSCs for xenotransplantation experiments using FACS. Single HSCs defined as viable, Lineage-, CD34+, CD38-, CD45RA-, CD90+, CD49f+ cells were sorted as either VWF+ or VWF- prior to transplantation. 1653 VWF+ HSCs were sorted by FACS, 138,386 VWF- HSCs were sorted by FACS. Cells were centrifuged at 600xg for 5min.

All mouse work was performed in collaboration with Dr E Laurenti's group, University of Cambridge. 12-week-old mice were sub lethally irradiated (2.40 Gy) 24 hr before transplant. Prior to transplantation, mice were temporarily sedated with isoflurane. A 27g needle was used to drill the right femur and cells were transplanted in a 25mL volume using a 28.5g insulin needle. 8 mice were transplanted with 400 cells (4 mice: VWF+ HSCs; 4 mice: VWF- HSCs), 2 mice were transplanted with 68,000 VWF-HSCs, 1 mouse was not transplanted and kept as a control.

At 16 weeks after transplantation all mice were sacrificed and the right and left femur and tibiae and peripheral blood were analysed for human cell engraftment. Cells were stained in PBS/2% FBS and analysed by flow cytometry (LSRII, Becton Dickinson). The marrow and blood was analysed with GlyA PE (Beckman Coulter, clone 11E4B-7-6 [KC16]): marker of erythroid cells; CD45 PeCy5 (Beckman Coulter, clone J33), CD45 BV510 (BD): marker of leucocytes; CD3 APCCy7 (BD): marker of T cells; CD19 FITC (BD), CD19 AF700 (BD): marker of B cells; CD14 PeCy7 (Beckman Coulter, clone RMO52), CD33 APC (BD, clone P67.6): markers of myeloid cells. A small aliquot was also stained with CD41 APC (BD), CD42 PE (BD) and Hoechst to identify megakaryocytes in the bone marrow.

2.6 Clinical statistics

Clinical data were analysed using Students T-test for comparison of means between two groups and ANOVA was used for multiple means testing. Correlations were performed using

the Spearman rank method. All calculations were performed using Graphpad Prism software (version 5.0). Statistical significance was taken to be $p<0.05$.

2.7 RNA-seq analysis

Up to the point of functional analysis, single cell and low input RNA-seq data was analysed by Dr Frederik Bagger, European Bioinformatics Institute, Hinxton and platelet RNA-seq data was analysed by Dr Romina Petersen and Dr Luigi Grassi, Department of Haematology, University of Cambridge.

Paired-end reads were trimmed of both PCR and sequencing adapters <32bp using TrimGalore!(http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/ 2014). The trimmed reads were then mapped to the human reference genome (GRCh37) using STAR (Dobin et al. 2013) creating an expression matrix. Quality metrics were assessed for each sample at this stage. These included: total reads: total sum of all features including spike-in transcripts, alignment rate: rate of reads mapped to the reference genome using STAR, ERCC ratio: the proportion of reads mapped to spike-in transcripts (an enrichment of reads mapping to ERCCs indicative of loss of endogenous RNA), number of called genes: number of expressed features in each cell with non-zero counts, ratio of exonic reads to total reads, ratio of reads mapping to the mitochondrial genome to total reads. These metrics were illustrated using: http://servers.binf.ku.dk:8890/sinaplot/.

2.7.1 Single cell RNA-seq analysis: HSCs

Cell filtering to distinguish high-quality single cells from low-quality cells was performed using a machine learning model able to simultaneously detect subtle patterns across a number of quality metrics based on a previously described method (Ilicic et al. 2016). Here the list of technical and biological variables were modified and used to train a support vector machine to recognise cells that shared features with cells expressing the house-keeping gene, GAPDH. These cells were then taken forward for downstream analysis.

Transcript expression was normalised across single cells using Scater package (McCarthy et al. 2017) in order to correct for technical and biological variance between cells using previously described methods (Lun, McCarthy, and Marioni 2016). This was based on the assumption that most genes are not differentially expressed between cells, therefore systemic differences are removed by scaling. Since ERCC spike-ins were added in the same quantity to all cells, and therefore exhibit only technical and no biological variability, variance in spike-in expression was used to model the trend in technical variability. Genes that exhibited variance that did not fit this trend were identified as genes that drive biological heterogeneity.

Unsupervised clustering of the cells according to similarity of gene expression was performed firstly by constructing pairwise comparisons between the first 3 principal components of variation in transcript expression (PCA: principal component analysis) and secondly using consensus clustering of RNA-seq data known as SC3 (Kiselev et al. 2017) which is a heirrachical clustering based on correlation in the PCA space. The robustness of the clusters was determined by the Silhuoette index (Rousseeuw 1987). Using SC3, cluster marker genes were identified by a value to each gene denoting how well it is a predictor for each cluster. A p-value was then assigned to each gene using the Wilcoxen signed-rank test. Genes with $P<0.001$ were defined as marker genes.

Cells were then clustered using the Monocle 2 single-cell analysis toolset (Trapnell et al. 2014; Qiu et al. 2017). Here the cells were ordered in terms of differential gene expression incorporating correlation distances, whereby each cell was placed along a differentiation trajectory, with branch points indicating where cells exhibit alternative cell differentiation programs. Cluster marker genes were those that were consistently differentially expressed compared to all other genes. These were identified using the Wilcoxon Rank test.

In order to assess expression of known haematopoietic gene signatures within the single cells clusters, gene signatures for each cell type in the DMAP dataset (Novershtern et al. 2011) were created using the LIMMA package (Ritchie et al. 2015). Here a one vs all comparison was made for all cell types. P values were adjusted for multiple hypothesis testing with a false discovery rate <0.05. Genes that were considered to be part of the gene signature were defined as those with log2 fold change of >1. The expression of genes from each DMAP gene signature was then assessed in each individual single cell and this information was superimposed on the Monocle 2 plots. Furthermore, expression of cluster marker genes within haematopoietic progenitor populations in the Blueprint dataset (L. Chen et al. 2014) was assessed using the Blueprint progenitors tools portal: https://blueprint.haem.cam.ac.uk.

2.7.2 Single cell RNA-seq analysis: Megakaryocytes

The lower quality metrics of the single cell megakaryocytes samples meant that the single cell data for this cell type was used for validation of and comparison with low input multicell megakaryocyte pools only. Therefore the machine learning model that was used to predict high quality cells for downstream analysis was a random forest classifier (Breiman 2001). The training dataset that was used were the multicell megakaryocyte pools and 5 rounds of training were used to identify cells taken forward for differential expression analysis. Normalisation of gene expression and gene filtering was performed as described in Section 2.6.1.1.

2.7.3 Differential expression analysis: megakaryocytes and platelets

DE-Seq2 (Love, Huber, and Anders 2014) was used to find genes that were differentially expressed between different datasets in single megakaryoyctes, megakaryocyte multicellular pools, cultured bulk megakaryocytes and platelets. Here the Wald test is used for hypothesis testing. Differentially expressed genes were defined as $p<0.05$. For more than one comparison correlations were performed using the Spearman rank method.

2.7.4 Functional gene list analysis

Gene ontology (GO) analysis was performed and its visual representation generated using the online tool, Fidea (D'Andrea et al. 2013): http://circe.med.uniroma1.it/fidea. Here enrichment analysis is performed by statistically assessing whether a pathway or process is enriched in the specific gene list. This is achieved by the hypergeometric test with the resulting p-values corrected using the Benjamini and Yekutieli FDR method (Yekutieli and Benjamini 2001). This method was used for highly expressed gene lists, cluster marker genes and differentially expressed genes.

CHAPTER 3

Single cell RNA-seq reveals transcriptional variation and megakaryocyte biased subsets within the human bone marrow HSC compartment

3.1 Preface

The first evidence of heterogeneity in the HSC compartment came from mouse spleen colony forming unit (CFU) assays, that showed high degrees of variability in numbers and types of daughter cell colonies produced (Siminovitch, McCulloch, and Till 1963; Till, McCulloch, and Siminovitch 1964). This challenged the traditional concept of a single homogenous population. However, direct evidence of HSC heterogeneity came from methods that allow assessment of mature cell outputs from limiting numbers and even single HSCs (Kiel et al. 2005; Osawa et al. 1996). These include the ability of purified single HSCs to repopulate a secondary myelo-ablated host and cellular barcoding whereby lentiviral transfer is used to uniquely label individual HSCs allowing their progenies to be tracked within the transplanted host (Lu et al. 2011; Gerrits et al. 2010). With improvement in cell purification techniques and functional clonal assays, as well as transplantation experiments, human HSCs are now thought to reside within the $Lin^-CD34^+CD38^-CD45RA^-CD90^+CD49f^+$ surface markers defined compartment (Notta et al. 2011). Functionally distinct HSC subsets within these gates have been distinguished, as defined by their mature cell output, as myeloid-biased, lymphoid-biased or balanced (Muller-Sieburg et al. 2004, 2012; Dykstra et al. 2007; Pang et al. 2011; Cavazzana-Calvo et al. 2011). However, the retrospective nature of these assays has meant that cells are not available for further analysis. There remains no reliable method for prospectively distinguishing these subpopulations. Therefore, it is evident that despite having the same cell surface markers subsets exist within the HSC population with distinct transcriptional programs.

Almost all the accrued knowledge of the HSC transcriptome is derived from studies at population level that do not take into account the heterogeneity within the population, the extent of which remains largely unknown. Therefore, a study of HSCs at single cell resolution is essential to define the heterogeneity that underpins many biological processes including distinct cell fate decision.

A human cell is predicted to contain less that 1pg of mRNA with the majority of transcripts being expressed at less than 100 transcript copies per cell (Kawasaki 2004). Therefore, it can be envisaged that genes co-expressed at population level are in fact mutually exclusive at the single-cell level. This implies that subsets within a cell population could be functionally distinct, for example in the HSC population, it would mean separate subgroups of cells with different mature cell lineage output. Recent technological advances have allowed single cell transcriptome profiling using methods ranging from the probe dependent fluorescent in situ

hybridization to single-cell reverse transcription and transcriptome wide amplification methods including qPCR, microarray, and more recently, NGS analyses of the transcriptomes of single cells; previously reviewed (Kalisky, Blainey, and Quake 2011; Tang, Lao, and Surani 2011; Saliba et al. 2014; Etzrodt, Endele, and Schroeder 2014). Human HSCs have been studied at the single cell level using qPCR targeting specific transcripts (Guo et al. 2013; Adolfsson et al. 2005; Forsberg et al. 2006; Månsson et al. 2007; Akashi et al. 2003; Hu et al. 1997; Miyamoto et al. 2002). At the time of beginning this project single cell RNA-seq had not been used in human HSCs to investigate HSC heterogeneity.

As described, a new model of human haematopoiesis using functional clonal assays and improved cell purification suggests that megakaryocyte lineage commitment can bypass intermediate differentiation steps (CMP and MEP), becoming completely independent of these steps in adult bone marrow (Notta et al. 2016). Furthermore, a megakaryocyte-biased HSC subset has been identified in the mouse (Sanjuan-Pla et al. 2013; Gekas and Graf 2013). VWF gene expression has been described as a marker of a megakaryocyte-biased HSC subset, which defines a population leading to long term hematopoietic reconstitution with platelet bias when serially transplanted in limiting numbers or single cell (Sanjuan-Pla et al. 2013). VWF is a large multimeric plasma glycoprotein that is synthesised by megakaryocytes and are packaged into the platelet alpha granule to be released on activation. There is some evidence that VWF is also found independently of the alpha granule release on the surface of platelets activated by ADP (Parker, Shafer, and Gralnick 1987; Parker, Rick, and Gralnick 1985). It is also synthesised in endothelial cells, packaged into specialised vesicles called Weibel-Palade bodies and released into the extracellular space. Its transcripts are expressed in HSC, megakaryocytes and platelets (Novershtern et al. 2011; Rowley et al. 2011; L. Chen et al. 2014). It has a major role in platelet aggregation by forming a molecular bridge between the subendothelial collagen matrix and the platelet surface receptor: GP1b-IX-V.

The work described in this chapter, includes the use of single cell RNA-seq combined with fluorescence-activated cell index sorting to analyse transcriptional variation within the human bone marrow HSC population and identify subpopulations biased to different cell fates with a particular focus on defining a megakaryocyte-biased HSC subset. Since *VWF* transcript expression has been proposed as a marker of a megakaryocyte or platelet-biased HSC population, differential expression analysis was also performed between VWF+ and VWF- HSCs. Finally, although VWF is not known to be a cell surface protein, it has been hypothesised that it could play a physiological regulatory role in HSCs and therefore its surface expression at the single cell level was also measured and correlated to transcript expression. HSCs sorted according to presence or absence of surface VWF were transplanted into sub lethally irradiated mice to compare reconstitution patterns.

3.2 Results

3.2.1 HSC phenotype

Single HSCs were flow sorted from sternal bone marrow from 5 individuals for the purpose of single cell RNA-seq (Section 2). Table 3.1 shows the baseline characteristics for these individuals, they were either undergoing mitral (n=3) or aortic valve surgery (n=2), 2 had mild to moderately impaired LV systolic function and there was no evidence of coronary artery disease (on pre-operative coronary angiography). Mean age was 65, with haemoglobin concentration, white cell count and platelet count all within normal limits.

Figure 3.1 illustrates the experimental design. Bone marrow samples were prepared for FACS sorting as described in Section 2.2.4.1. The HSC population was identified as Lineage negative, CD34+, CD38-, CD45RA-, CD90+, CD49f+. The following antibodies were used as lineage cocktail: CD2, CD3, CD10, CD11b, CD11c, CD19, CD20, CD56, in order to gate out any mature cells as well platelets with CD42b and GP6. An antibody to VWF was also added to measure the surface expression of VWF. Figure 3.2 shows the gating strategy used after doublets had been excluded from analysis.

In total, 884 HSCs were FACS sorted and index data collected (Osborne 2011). For each single cell to enable post-sort review of where each cell was located within the gates as well as the exact flow cytometric phenotype including information on forward/side scatter and the 7 additional surface markers. Specifically, the surface expression of VWF was analysed using index data for each individual HSC. Approximately 50% HSCs were found to be VWF+ by surface marker expression. Figure 3.1 shows the cell numbers sorted from each individual with mean fluorescent intensities across the sorted populations for forward and side scatter as well as all surface markers used. There were no significant differences in index data, or VWF expression within the HSC population between individuals.

Table 3.1 Baseline characteristics

	Healthy controls (n=5)
Gender M	3
Age (mean ± SD)	65 ± 6.8
Hypertension	2
Hypercholesterolaemia	2
Diabetes	3
Smoking	0
Coronary disease	0
Antiplatelets	0
Left ventricular dysfunction	2
Aortic valve surgery	2
Mitral valve surgery	3
Hb (mean ± SD)	12.18 ± 1.8
WC (mean ± SD)	8.0 ± 0.89
Plt (mean ± SD)	219 ± 43

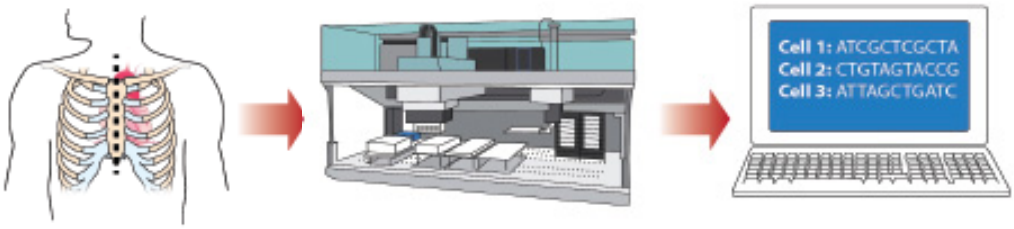


Figure 3.1 Experimental design
Sternal bone marrow was harvested from 5 patients undergoing non-coronary cardiac surgery. It was stained with antibodies for surface markers that define the human HSC population and sorted as single cells using index FACS into cell lysis buffer. Single cell RNA sequencing libraries were then prepared using a modified Smart-seq2 method and libraries were sequenced on a HiSeq 4000 instrument. The data was then analysed as described in Chapter 2.

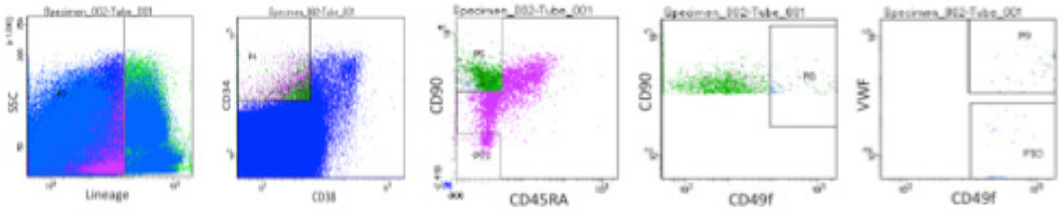


Figure 3.1 HSC flow and raw index data.
Fluorescence activated cell sorting of primary HSCs from clinical samples of whole human bone marrow. Cells were sorted into lysis buffer within 6 to 8 hours after bone marrow harvest. Whole bone marrow was stained with PECy5 conjugated anti-lineage antibodies: CD2/CD3/CD10/CD11b/CD11c/CD19/CD20/CD56, biotinylated CD42b/GP6 plus PECy5 conjugated streptavidin, AlexaFluor 700 conjugated anti-CD34, PerCP- Cy5.5 conjugated anti-CD38, Pacific Blue conjugated anti-CD45RA, PECy7 conjugated anti-CD90, PE conjugated anti-CD49f, FITC conjugated anti-VWF. The data shown is from Pt ID FX7. NS: non significant.

Pt ID	60Z	G1U	FX7	GBA	GPJ	p value
HSCs sorted	192	192	288	96	116	NS
MFI						
FS	59276.4	72028.2	51447.3	42393.9	50659.6	NS
SS	22733.5	34762.7	26194.1	24324.1	24098.6	NS
Lineage	6150.77	11750.3	6583.63	2300.40	5881.84	NS
CD34	13552.8	17624.1	24386.1	13493.7	39248.5	NS
CD38	-2235.28	388.24	-498.83	736.85	2720.25	NS
CD45RA	-2355.41	-268.26	-640.36	-476.45	-559.81	NS
CD90	6524.80	7286.85	10805.6	8909.76	5842.78	NS
CD49f	43608.5	17600.3	13224.9	16380.7	6845.60	NS
VWF	21090.8	12405.2	12718.3	11619.9	18585.8	NS

3.2.2 Single cell RNA-seq library preparation: quality measures

3.2.2.1 Choice of reverse transcriptase

The Smart-seq2 protocol was used to generate a high yield of full length cDNA for single cell transcriptome analysis. The recommended Superscript II reverse transcriptase was used for first strand synthesis in the initial 116 single HSCs processed (ID GPJ). However, the resulting BioAnalyser traces showed suboptimal cDNA quality with a wide distribution of fragments size. Upon testing alternative reverse transcriptases, Smartscribe demonstrated improved efficiency in producing full length cDNA from both purified RNA and other cell types as well as HSCs and was therefore used instead of Superscript II for the remaining 768 single HSCs (Figure 3.3).

3.2.2.2 Expression of *GAPDH* as a house-keeping gene

In order to detect high quality single cell cDNA for the first 4 HSC plates amplified, Taqman real-time PCR assays for the presence of the house-keeping gene *GAPDH* were used. Figure 3.4 shows the number of cells expressing *GAPDH* in the first 4 plates. The mean no. of *GAPDH* expressing cells per 96 well plate was 19 ± 3.2 (mean ± SD). It is noted that despite suboptimal BioAnalyser traces for the cells reverse transcribed using Superscript II (GPJ P1) rather than Smartscribe (G1U P1, G1U P2, FX7 P1) the numbers of cells expressing *GAPDH* were comparable.

As high expression of mitochondrial transcripts led to their overrepresentation the RNA-seq data for megakaryocytes (Chapter 4), the expression of the most highly expressed mitochondrial gene: *MT-RNR2* was measured by Taqman real-time PCR alongside *GAPDH* expression in the same 4 HSC plates and compared to the same measurement in a plate of megakaryocytes. Figure 3.5 shows that the ΔCT between *GAPDH* and *MT-RNR2* expression in *GAPDH* positive megakaryocytes was significantly higher than the more favourable ΔCT in HSCs. That is, MT-RNR2 was detected a mean of 4 PCR cycles before *GAPDH* in megakaryocytes compared with HSCs where both *GAPDH* and *MT-RNR2* transcripts were detected at approximately the same PCR cycle number.

3.2.2.3 Cell selection for RNA sequencing

In order to investigate whether GAPDH expression was an adequate marker for cDNA quality and therefore a criterion to select cells for RNA-seq, expression of *B2M* was analysed alongside *GAPDH* in a single 96 well plate of HSCs. Figure 3.6 shows that although many cells co-expressed both housekeeping transcripts, there were cells that expressed either one or the other. The data shown is from patient and plate ID FX7 P1. Here *GAPDH* alone

identified 16 cells, whereas if high quality cDNA were to be defined by expression of either *GAPDH* or *B2M* this would identify 27 cells in total.

Next, a set of 48 cells *GAPDH* selected cells from plates G1U P1 and G1U P2 and a single unselected 96 well plate (FX7 P1) were sequenced. Preliminary quality checks were made to compare the 2 datasets. This included alignment rate, exon/exon+intron ratio and ERCC ratio. Figure 3.7 shows that the majority of *GAPDH* selected cells had higher quality by these criteria compared with unselected cells. Using preliminary and arbitrary cut-offs based on the performance of the *GAPDH* selected cells, 26 cells from the unselected plate (FX7 P1) were defined by sequencing as high quality. Figure 3.8 shows that, by using this definition of quality, 6 more cells were deemed good quality than would have been selected on the basis of a combination of *GAPDH* and *B2M* expression alone. To maximise the number of good quality cells sequenced it was therefore decided to unselectively sequence all 884 cells and filter the data post RNA-seq rather than prospectively selecting cells according to single gene expression using Taqman real-time qPCR.

3.2.2.4 Correlation between surface and transcript expression of VWF by qPCR

To investigate if expression of VWF on the surface of HSCs could be related to the presence of *VWF* transcript, Taqman real-time qPCR was used to quantify expression of the *VWF* transcript in cDNA from all single HSCs. This was correlated with MFI from FACS index data for VWF. In total *VWF* transcript was identified in 43 out of 884 single cells. Figure 3.9 shows that although there is a weakly significant r value there is no clear correlation between surface marker expression and transcript expression of *VWF*.

3.2.2.5 Batch effects

Batch effects, have been widely reported as a major challenge in high-throughput technologies (Leek et al. 2010) and especially in single cell RNA-seq as it is more difficult to randomise batches across sequencing lanes or flow cells (Stegle, Teichmann, and Marioni 2015). Systematic error has been shown to explain a substantial percentage of observed cell-to-cell expression variability in a number of previous studies (Hicks, Teng, and Irizarry 2015). Table 3.2 shows each HSC 96 well plate and which sequencing lane ID whereas Table 3.3 outlines the batches in which samples were collected, cDNA was reverse transcribed/amplified and Nextera libraries were prepared and sequenced.

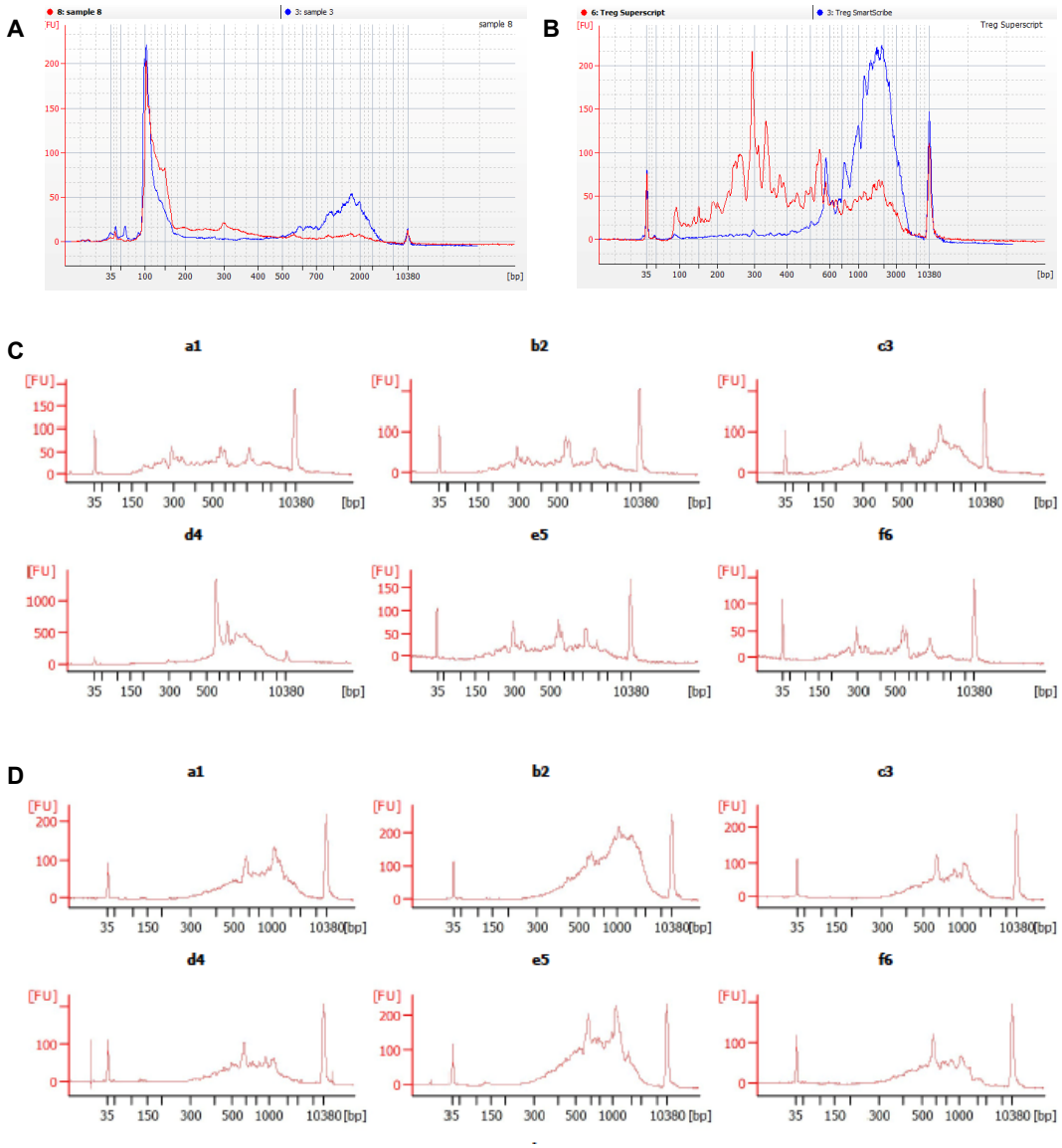


Figure 3.3 Smartscribe improves generation of full length cDNA.
BioAnalyser electropherograms of cDNA libraries; x axis shows fragment size (bp), y axis shows quantity. A. cDNA reverse transcribed from purified RNA using Superscript II (red) compared with Smartscribe (blue). BioAnalyser trace prior to PCR clean-up, showing large peak of primer-dimers. (Performed by Ricardo Miragaia, WTSI/EBI, Hinxton). B. cDNA reversed transcribed from regulatory T cells using Superscript II (red) compared with Smartscribe (blue). (Performed by Ricardo Miragaia, WTSI/EBI, Hinxton, UK). C. cDNA reverse transcribed from six randomly selected single HSCs from Patient ID GPJ using Superscript II. D. cDNA reverse transcribed from six randomly selected single HSCs from Patient ID G1U using Smartscribe.

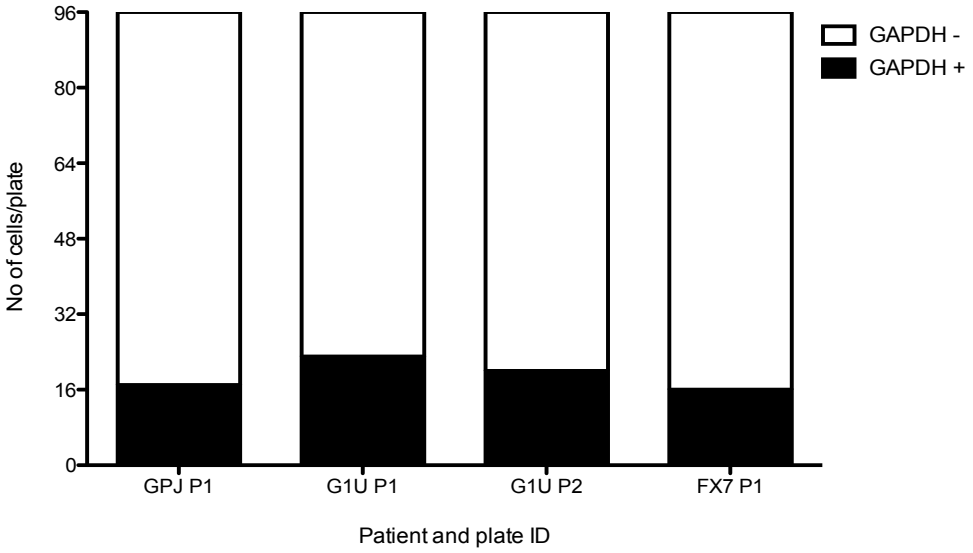


Figure 3.4 Number of *GAPDH* expressing cells over the first 4 HSC plates amplified. Taqman real-time PCR was used to investigate the expression of *GAPDH* in cDNA from 96 cells in each plate. Data is from GPJ P1, G1U P1, G1U P2, FX7 P1.

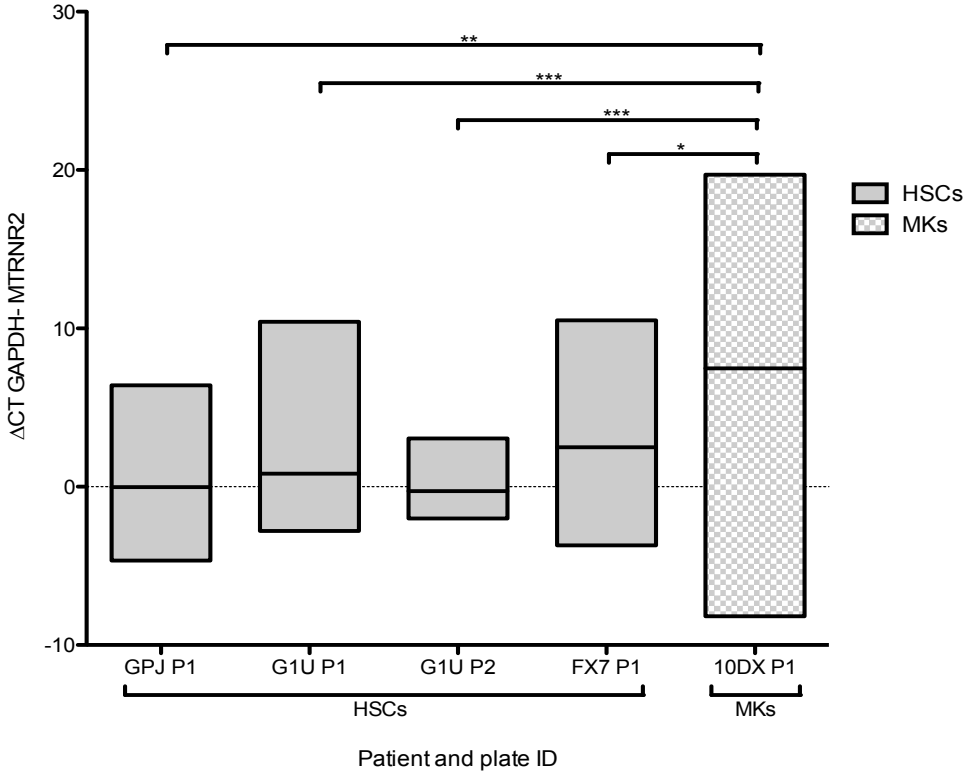


Figure 3.5 Expression of *GAPDH* compared to *MT-RNR2* in HSCs and megakaryocytes. Taqman real-time PCR was used to investigate the expression of *GAPDH* and *MT-RNR2* in HSCs and megakaryocytes. Data is presented as ΔCT between CT value for *GAPDH* and CT value for *MT-RNR2*. Each bar represents the range of values for ΔCT in each plate of cells with the line at the mean. A more negative value indicates a higher *GAPDH* expression than *MT-RNR2*, whereas a more positive value indicates higher *MT-RNR* expression. (*p<0.05, **p<0.001, ***p<0.0001

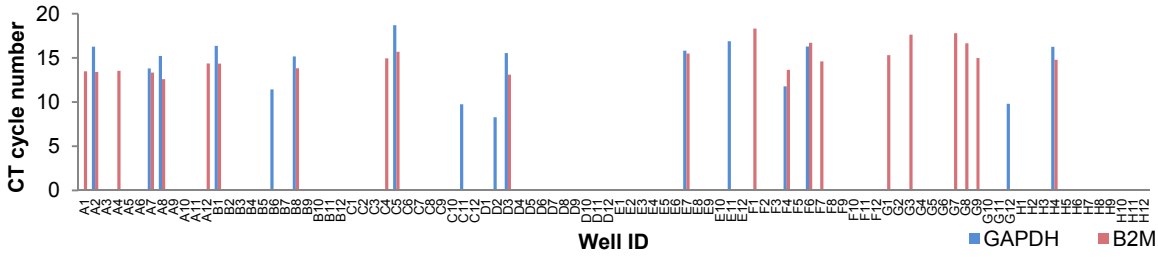


Figure 3.6 Expression of *GAPDH* and *B2M* in a single 96 well plate of HSCs. Taqman real-time PCR was used to investigate the expression of *GAPDH* and *B2M* in cDNA from single HSCs. Data shown is from FX7 P1. Data is presented as CT i.e., the cycle number at which the transcript was detected; the lower the CT the more transcript.

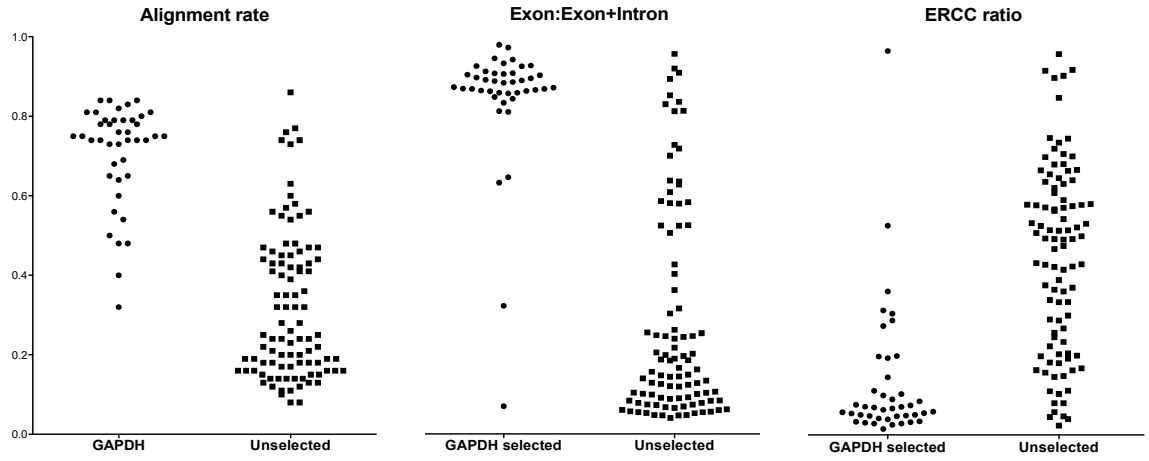


Figure 3.7 Preliminary RNA sequencing quality. Data for 48 *GAPDH* selected HSCs from G1U P1 and G1U P2 and unselected HSCs from FX7P1 is presented. *GAPDH* selected HSCs had a significantly higher alignment rate, ratio of exons to introns and lower levels of ERCC.

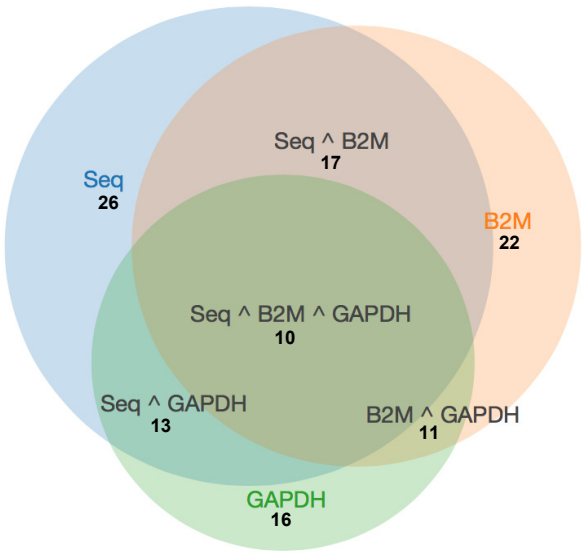


Figure 3.8 Overlap of good quality as defined by sequencing quality and expression of housekeeping genes using Taqman real-time PCR. Sequencing quality was defined as high alignment rate and exon/intron ratio as well as low ERCC ratio. Arbitrary cut offs were used based on the performance of *GAPDH* selected cells. Data shown is from FX7 P1.

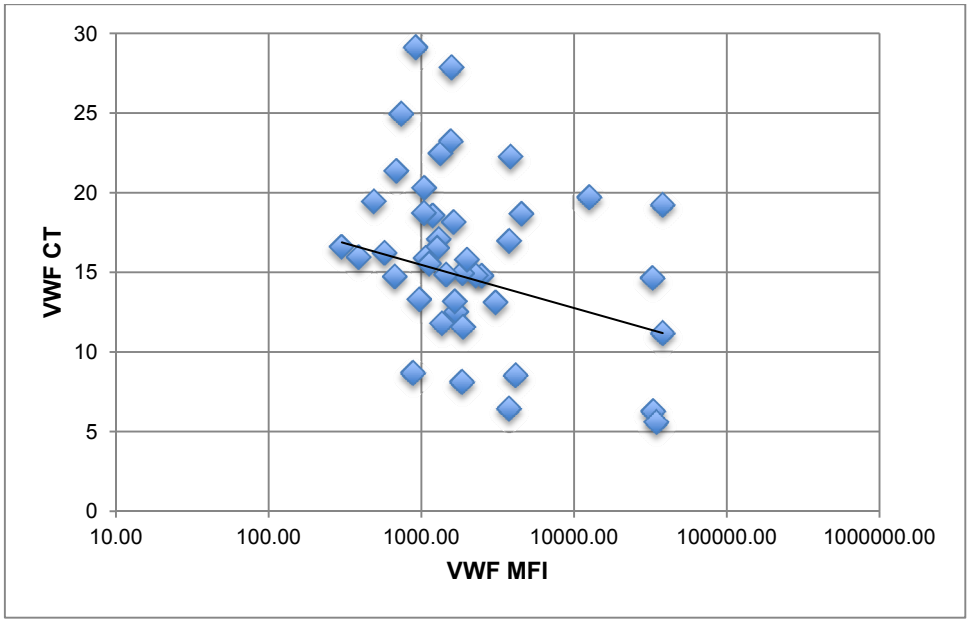


Figure 3.9 Correlation between VWF surface and transcript expression by qPCR. VWF surface expression is in MFI measured by FACS index sorting for each individual cell. Transcript expression was measured by Taqman real-time PCR. Data is presented as CT value, the lower the cycle number or CT at which the transcript was detected, the higher the expression. R value= -0.302 which is weakly significant (p=0.0443).

Plate ID	Lane sequenced
60Z P1	SLX-11217
60Z P2	SLX-11218
FX7 P1	SLX-10542
FX7 P2	SLX-11219
FX7 P3	SLX-11220
G1U P1 GAPDH+	SLX-10548
G1U P1 GAPDH-	SLX-11215
G1U P2 GAPDH+	SLX-10548
G1U P2 GAPDH-	SLX-11216
GBA P1	SLX-10543
GPJ P1	SLX-11221
GPJ P2	SLX-11216

Table 3.2 Patient and plate ID on each sequencing lane

Sample collection	Amplification	Nextera Library prep / RNA-seq
1	1	1
60Z P1	GPJ P2	SLX-10548
60Z P2	2	2
2	GPJ P1	SLX-10542
FX7 P1	3	SLX-10543
FX7 P2	G1U P1	3
FX7 P3	4	SLX-11215
3	G1U P1	SLX-11216
G1U P1	60Z P1	SLX-11217
G1U P2	5	SLX-11218
4	FX7 P1	SLX-11219
GBA P1	GBA P1	SLX-11220
5	6	SLX-11221
GPJ P1	FX7 P2	
GPJ P2	FX7 P3	
	60Z P2	

Table 3.3 HSC collection, amplification, and library preparation/RNA-seq batches

3.2.3 Single cell RNA sequencing data: Quality measures

3.2.3.1 Quality control data

Figure 3.10 shows metrics per sequencing lane for total read count, alignment to the reference transcriptome (GRCh37), proportion of reads mapped to external spike-in ERCCs compared to endogenous transcripts, number of genes called, proportion of reads mapping to annotated exonic regions and the proportion of reads mapping to the mitochondrial genome.

All lanes (with the exception of SLX-10548, SLX-11217, SLX-11218) demonstrate similar quality metrics. The majority of cells have a total read count $<5x10^6$, alignment rate <50%, ERCC ratio >50%, number of called genes <1000, exonic ratio <50%. Lanes SLX-11217 and SLX-11218 demonstrate particularly poor quality with relatively lower alignment rates, genes called, and exonic reads and higher proportion of reads mapped to ERCC spike-in transcripts. These 2 lanes were FACS sorted on the same day and therefore the similarity in overall quality metrics may be representative of batch effect.

As previously discussed in 3.2.2.3, lane SLX-10548, consisting of cells selected on the basis of *GAPDH* expression, had relatively better performance with significantly higher total read count, alignment rate, number of called genes, and reads mapping to exonic regions and significantly lower ERCC ratio and reads mapping to the mitochondrial genome. However, a smaller proportion of *GAPDH* negative cells from the same individuals (Lanes SLX-11215 and SLX11216) also shared similar favourable metrics demonstrating a lower specificity than sensitivity of *GAPDH* as a quality marker.

3.2.3.2 Cell filtering

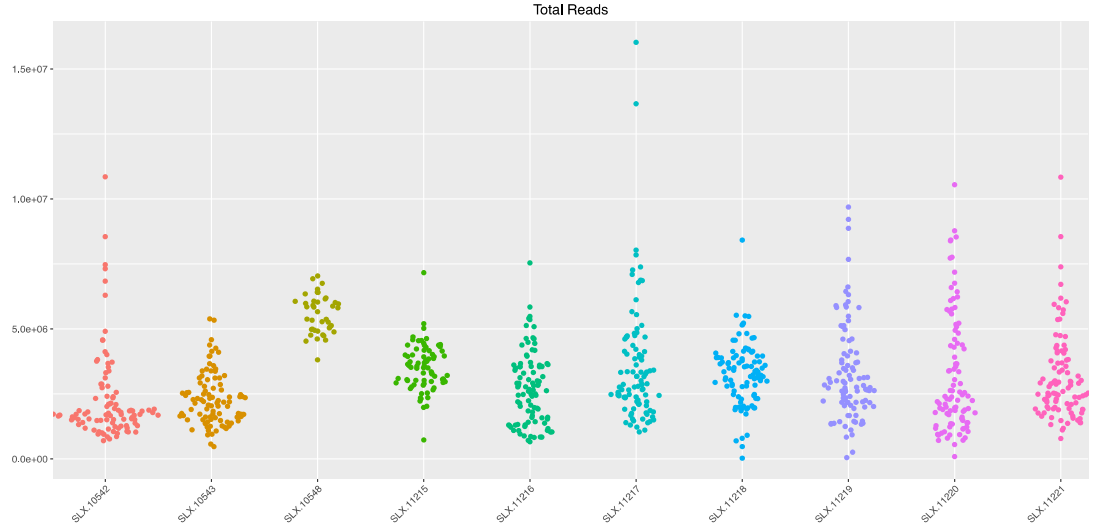
Several approaches have been proposed to filter out low quality single cell data at the point of data analysis, but they require arbitrarily set filtering thresholds which perhaps only capture part of the entire landscape and are ideally suited to bulk sequencing (Hartley and Mullikin 2015; Okonechnikov, Conesa, and García-Alcalde 2015). To ensure that only high quality cells were taken forward for downstream analysis, a recently developed classification approach for identifying low-quality cells was used (Ilicic et al. 2016). This takes into account both technical and biological measures of cell quality. The principle component analysis (PCA) performed and the specific variables driving the PCA are shown in Figure 3.11a. However, the model failed to identify cells with high alignment rates and low ERCC ratios as high quality, instead filtering these out (Figure 3.11b). It has to be noted that this machine learning model used mouse data which may explain its limitations in analysing primary human bone marrow HSCs.

The list of technical and biological variables used was then modified (Figure 3.12a) and used to train a support vector machine (SVM) model to identify cells that resembled *GAPDH* +ve cells (lane SLX-10548). The variables included mapping statistics, gene expression and its variance as well as gene expression signatures based on gene ontology. *GAPDH* expression was excluded as a feature. The cells identified as of good quality share features with *GAPDH* +ve cells, including low ERCC ratio, and mitochondrial mapping as well as higher number of reads mapping to exonic regions and number of genes called (Figure 3.12b). In this way 119 out of 884 cells were taken forward to downstream analysis.

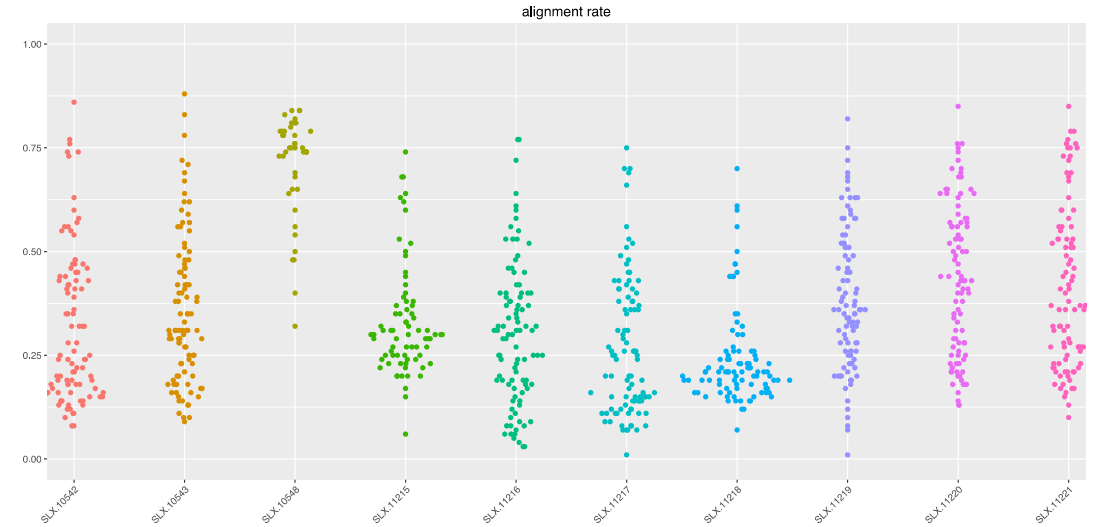
3.2.3.3 Gene filtering

Normalisation of overall transcript expression within a population of single cells is a critical step in single cell RNA-seq analysis to correct for cell to cell differences and ensure that downstream relative expression analysis between cells is valid. Gene filtering allows identification of a subset of genes that drive heterogeneity across the population of cells. In order to identify these highly variable genes, variance in expression of each gene was plotted against its expression level within the single cell population using a method described by Lun and Marioni (Lun, McCarthy, and Marioni 2016). This included the expression level of the ERCC spike-in transcripts, that should exhibit no biological variability and therefore represents the technical component of the expression variance observed. 2000 genes that demonstrated a higher variance than that of the ERCC curve fit were used for gene expression analysis (Figure 3.13).

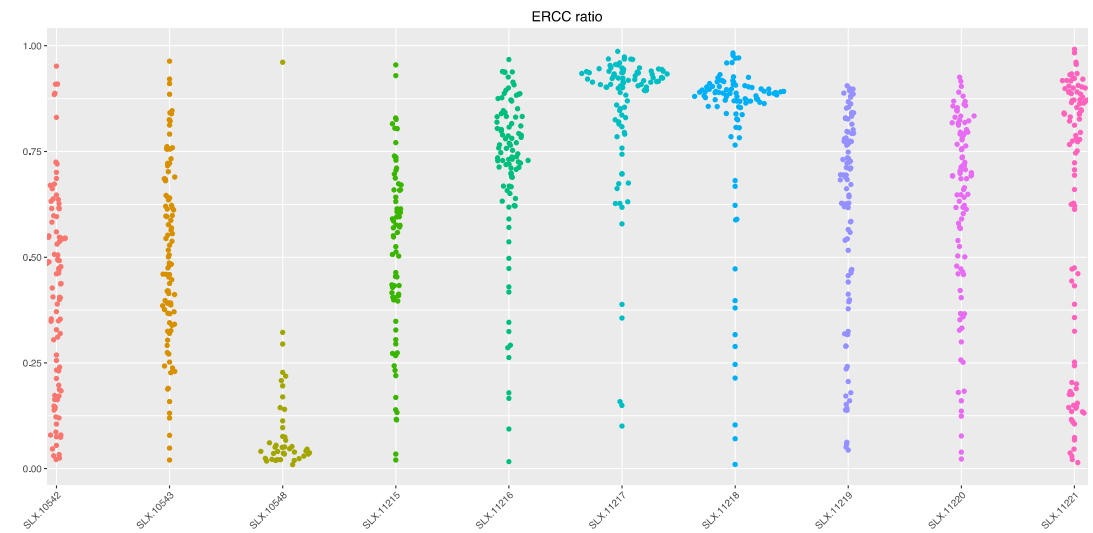
A


Total Reads
1.5e+07
1.0e+07
5.0e+06
0.0e+00
SLX.10542
SLX.10543
SLX.10548
SLX.11215
SLX.11216
SLX.11217
SLX.11218
SLX.11219
SLX.11220
SLX.11221

B


alignment rate
1.00
0.75
0.50
0.25
0.00
SLX.10542
SLX.10543
SLX.10548
SLX.11215
SLX.11216
SLX.11217
SLX.11218
SLX.11219
SLX.11220
SLX.11221

C


ERCC ratio
1.00
0.75
0.50
0.25
0.00
SLX.10542
SLX.10543
SLX.10548
SLX.11215
SLX.11216
SLX.11217
SLX.11218
SLX.11219
SLX.11220
SLX.11221

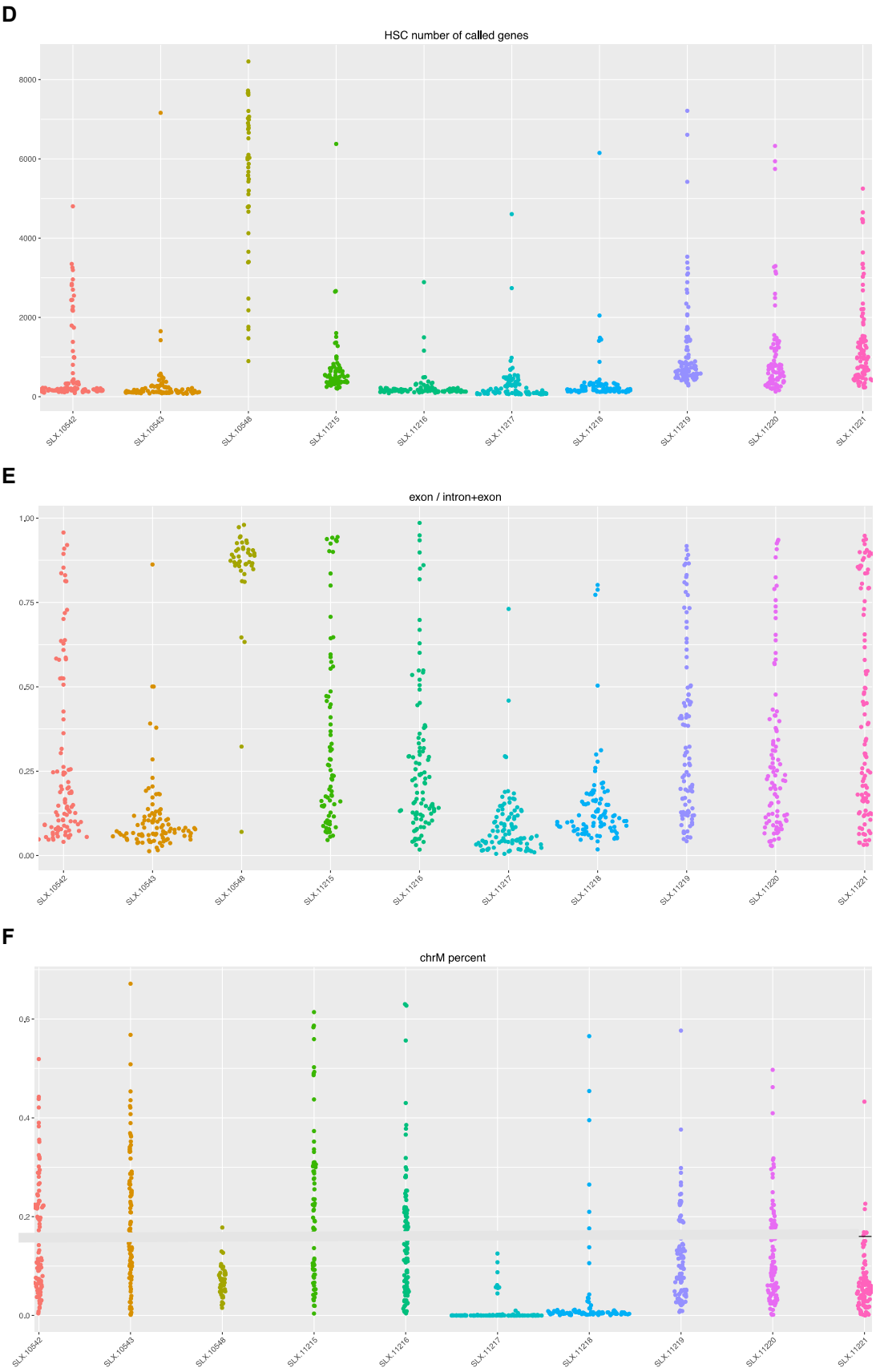


Figure 3.10 Quality control metrics/sequencing lane. A. Total reads; B. Rate of alignment to transcriptome; C. Ratio of reads mapping to ERCCs; D. No. genes called; E. Ratio of exonic reads to exonic+intronic; F. Ratio of reads mapping to mitochondrial genes.

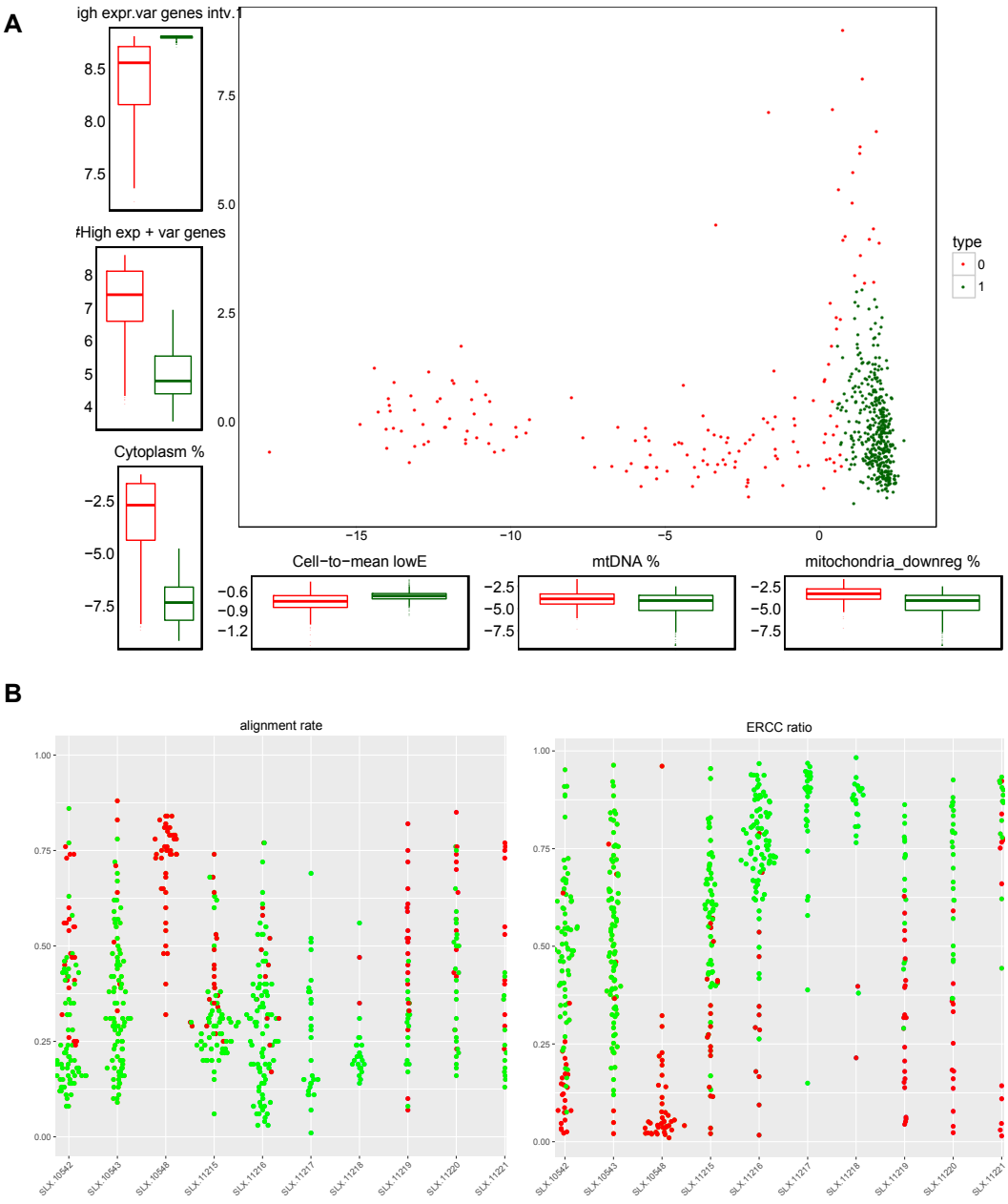


Figure 3.11 Filtering cells: Method 1.
Method 1 uses the SVM model described by Ilicic et al 2016. A. PCA performed on the technical and biological quality variables on which the model is based. The most explanatory variables for high and low quality are plotted; B. High and low-quality cells shown by alignment rate and ERCC ratio. Green denotes high quality, red denotes low quality.

A

```
Technical features:
  Cell-to-Mean-Correlation(gene expression)
  Cell-to-Mean-Correlation(gene expression:lowly expressed)
  Proportion of mapped reads
  Proportion intragenic reads
  Proportion of exonic reads
  Proportion of intronic reads
  Proportion of ambiguous reads
  #Detected genes
  #Highly expressed and highly variable genes
  #Very highly expressed variable genes (Interval 1)
  #Highly expressed variable genes (Interval 2)
  #Moderately expressed variable genes (Interval 3)
  #Lowly expressed variable genes (Interval 4)
  #Very lowly expressed variable genes (Interval 5)
  Transcriptome variance

Biological features:
  Apoptotic proportion (GO:0006915)
  Metabolic proportion (GO:0008152)
  Ribosome proportion (GO:0005840)
  Membrane proportion (GO:0016020)
  Cytoplasm proportion (GO:0005737)
  Extracellular region proportion (GO:0005576)
  mtDNA proportion (37 genes)
  Mitochondria upregulated (~700 genes, GO:0005739)
  Mitochondria downregulated (~700 genes, GO:0005739)
  Actb gene expression
```

B

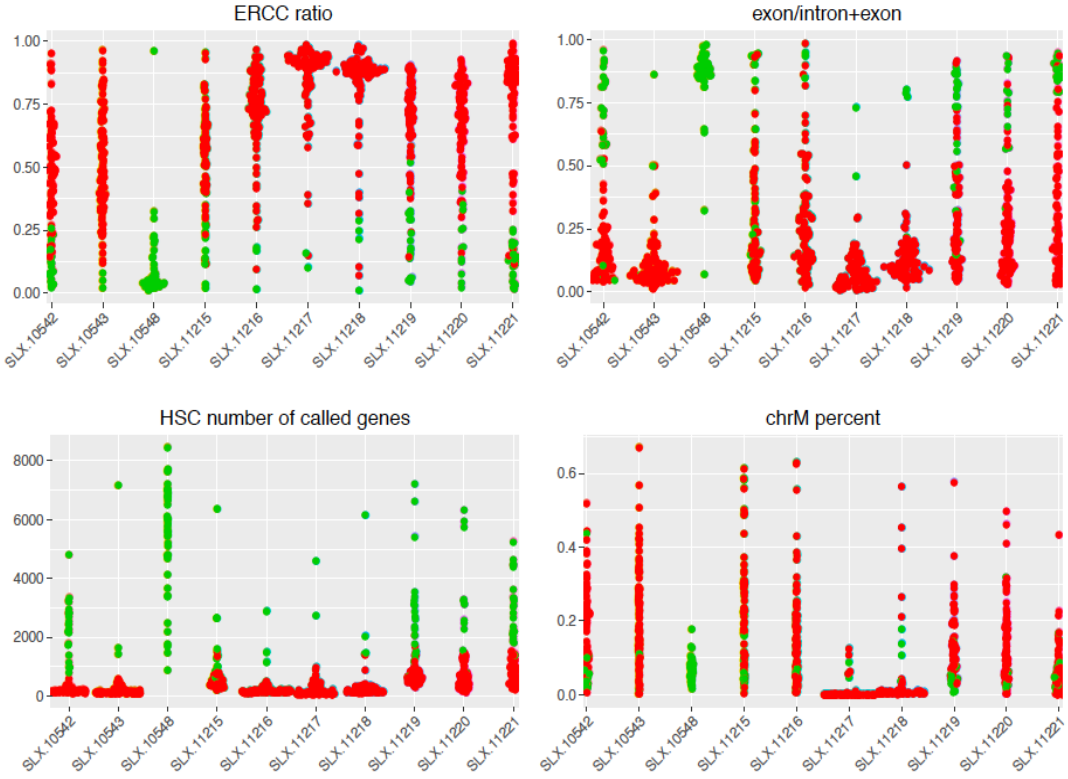


Figure 3.12 Filtering cells: Method 2.
Method 2 is a modification of the SVM model described by Ilicic et al 2016, using *GAPDH* positive cells as “good quality”. A. Modified technical and biological features on which the model is based; B. High and low-quality cells shown by ERCC ratio, exonic ratio, number of called genes and mitochondrial mapping. Green denotes high quality, red denotes low quality.

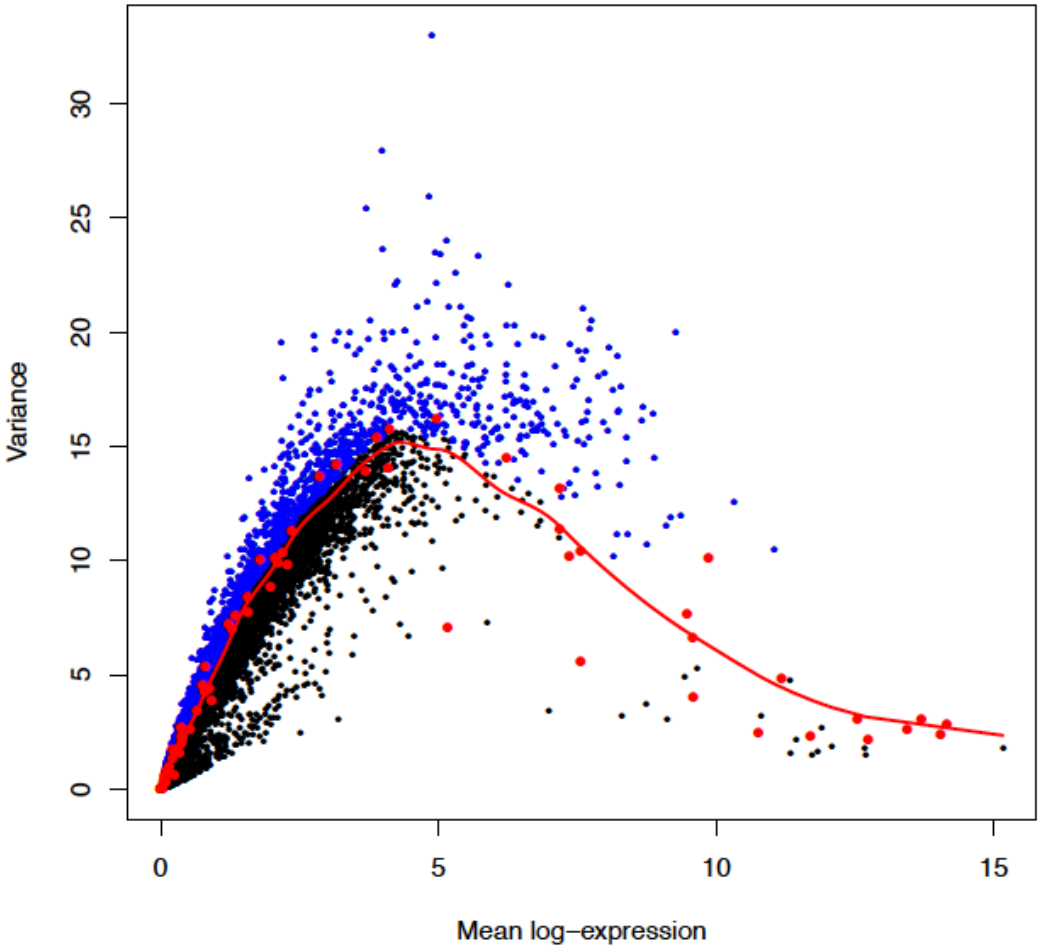


Figure 3.13 Filtering genes: Lun & Marioni Method.
Variance of normalized log-expression values for each gene in the HSC dataset, plotted against the mean log-expression. Variance estimates for ERCC spike-in transcripts and curve fit are highlighted in red. Blue dots represent highly varying genes (n=2000).

3.2.4 Gene expression analysis

The 200 most highly expressed genes in the HSC dataset (n=119) with their raw total read counts are shown in Table 3.4. The most highly expressed genes are mitochondrial (*MT-RNR2, MT-RNR1*) and ribosomal subunits. There are also a number of lineage specific genes that feature as highly expressed, particularly apparent are those of the megakaryocyte/platelet lineage: *PPBP* (ranked 19), *PF4* (ranked 29), *PLEK* (ranked 56), *SDPR* (ranked 107), *ITGA2B* (ranked 119), *VWF* (ranked 134), *CD36* (ranked 138), *CD63* (ranked 199).

GO classification is a systematic way in which large gene lists from RNA-seq are analysed to group genes according to their biological process, molecular function and/or site of cellular expression. This method highlights functionally related clusters of genes in a hierarchical manner and allows the identification of biologically relevant groups. In order to group the highly expressed HSC genes into biological processes, the data was analysed using the online tool Fidea (http://circe.med.uniroma1.it/fidea/) (D'Andrea et al. 2013). While the full GO analysis can be found in Appendix 3, Table S3.1, Figure 3.14a shows a visual representation of the GO biological processes over-represented in the top 200 genes. The most significant are related to translation and protein localisation within the cell, in fact 64 out of the 200 genes encoded for ribosomal subunits. Figure 3.14b shows the enriched GO biological processes with the ribosomal transcripts removed in order to reveal the next most significant functional classes: platelet degranulation/activation, exocytosis/cell secretion, wound healing, coagulation, haemostasis.

No	ENSEMBL_GENE_ID	HGNC_symbol	Gene_name	Total_reads
1	ENSG00000210082	*MT-RNR2*	Mitochondrially Encoded 16S RNA (MT-RNR2)	10021645
2	ENSG00000198804	*COX1*	cytochrome c oxidase subunit I(COX1)	1939185
3	ENSG00000198886	*ND4*	NADH dehydrogenase, subunit 4 (complex I)(ND4)	1509724
4	ENSG00000205542	*TMSB4X*	thymosin beta 4, X-linked(TMSB4X)	1452232
5	ENSG00000231500	*RPS18*	ribosomal protein S18(RPS18)	1235180
6	ENSG00000198938	*COX3*	cytochrome c oxidase III(COX3)	1188536
7	ENSG00000166710	*B2M*	beta-2-microglobulin(B2M)	1125812
8	ENSG00000198727	*CYTB*	cytochrome b(CYTB)	1035874
9	ENSG00000211459	*MT-RNR1*	Mitochondrially Encoded 12S RNA (MT-RNR1)	967361
10	ENSG00000198712	*COX2*	cytochrome c oxidase subunit II(COX2)	938267
11	ENSG00000198899	*ATP6*	ATP synthase F0 subunit 6(ATP6)	861381
12	ENSG00000156508	*EEF1A1*	eukaryotic translation elongation factor 1 alpha 1(EEF1A1)	721606
13	ENSG00000137154	*RPS6*	ribosomal protein S6(RPS6)	631701
14	ENSG00000198763	*ND2*	MTND2(ND2)	625597
15	ENSG00000112306	*RPS12*	ribosomal protein S12(RPS12)	623679
16	ENSG00000198888	*ND1*	NADH dehydrogenase, subunit 1 (complex I)(ND1)	613192
17	ENSG00000164587	*RPS14*	ribosomal protein S14(RPS14)	595091
18	ENSG00000089157	*RPLP0*	ribosomal protein lateral stalk subunit P0(RPLP0)	575666
19	ENSG00000163736	*PPBP*	pro-platelet basic protein(PPBP)	575218
20	ENSG00000198786	*ND5*	NADH dehydrogenase, subunit 5 (complex I)(ND5)	573750
21	ENSG00000122406	*RPL5*	ribosomal protein L5(RPL5)	568956
22	ENSG00000122026	*RPL21*	ribosomal protein L21(RPL21)	554861
23	ENSG00000142541	*RPL13A*	ribosomal protein L13a(RPL13A)	543703
24	ENSG00000137818	*RPLP1*	ribosomal protein lateral stalk subunit P1(RPLP1)	538550
25	ENSG00000174748	*RPL15*	ribosomal protein L15(RPL15)	517654
26	ENSG00000100316	*RPL3*	ribosomal protein L3(RPL3)	498646
27	ENSG00000177954	*RPS27*	ribosomal protein S27(RPS27)	488591
28	ENSG00000075624	*ACTB*	actin beta(ACTB)	485622
29	ENSG00000163737	*PF4*	platelet factor 4(PF4)	484140
30	ENSG00000142937	*RPS8*	ribosomal protein S8(RPS8)	483196
31	ENSG00000187514	*MIR1244-1*	microRNA 1244-1(MIR1244-1)	460569
32	ENSG00000080824	*HSP90AA1*	heat shock protein 90 alpha family class A member 1(HSP90AA1)	439388
33	ENSG00000140988	*RPS2*	ribosomal protein S2(RPS2)	426732
34	ENSG00000071082	*RPL31*	ribosomal protein L31(RPL31)	425537
35	ENSG00000197756	*RPL37A*	ribosomal protein L37a(RPL37A)	413026
36	ENSG00000163220	*S100A9*	S100 calcium binding protein A9(S100A9)	400980
37	ENSG00000108298	*RPL19*	ribosomal protein L19(RPL19)	397777
38	ENSG00000170323	*FABP4*	fatty acid binding protein 4(FABP4)	397273
39	ENSG00000147604	*RPL7*	ribosomal protein L7(RPL7)	394005
40	ENSG00000204592	*HLA-E*	major histocompatibility complex, class I, E(HLA-E)	386203
41	ENSG00000143947	*RPS27A*	ribosomal protein S27a(RPS27A)	376764
42	ENSG00000105372	*RPS19*	ribosomal protein S19(RPS19)	355547
43	ENSG00000251562	*MALAT1*	metastasis associated lung adenocarcinoma transcript 1(MALAT1)	354207
44	ENSG00000110700	*RPS13*	ribosomal protein S13(RPS13)	346524
45	ENSG00000149273	*RPS3*	ribosomal protein S3(RPS3)	346231
46	ENSG00000109475	*RPL34*	ribosomal protein L34(RPL34)	334276
47	ENSG00000167526	*RPL13*	ribosomal protein L13(RPL13)	316593
48	ENSG00000198034	*RPS4X*	ribosomal protein S4, X-linked(RPS4X)	314848
49	ENSG00000008988	*RPS20*	ribosomal protein S20(RPS20)	312428
50	ENSG00000133112	*TPT1*	tumor protein, translationally-controlled 1(TPT1)	310585
51	ENSG00000144713	*RPL31*	ribosomal protein L31(RPL31)	306935
52	ENSG00000142534	*RPS11*	ribosomal protein S11(RPS11)	303491
53	ENSG00000115268	*RPS15*	ribosomal protein S15(RPS15)	293437
54	ENSG00000143546	*S100A8*	S100 calcium binding protein A8(S100A8)	285860
55	ENSG00000277048		ENSG00000277048	285203
56	ENSG00000115956	*PLEK*	pleckstrin(PLEK)	278887
57	ENSG00000142676	*RPL11*	ribosomal protein L11(RPL11)	273439
58	ENSG00000145592	*RPL37*	ribosomal protein L37(RPL37)	270502
59	ENSG00000158710	*TAGLN2*	transgelin 2(TAGLN2)	263545
60	ENSG00000204628	*RACK1*	receptor for activated C kinase 1(RACK1)	263178
61	ENSG00000184009	*ACTG1*	Actin Gamma 1(ACTG1)	260016
62	ENSG00000234745	*HLA-B*	major histocompatibility complex, class I, B(HLA-B)	259346
63	ENSG00000212907	*ND4L*	NADH dehydrogenase, subunit 4L (complex I)(ND4L)	255164
64	ENSG00000143933	*CALM2*	calmodulin 2(CALM2)	253703
65	ENSG00000117632	*MIR3917*	microRNA 3917(MIR3917)	251082
66	ENSG00000105640	*RPL18A*	ribosomal protein L18a(RPL18A)	248305
67	ENSG00000182004	*SNRPE*	small nuclear ribonucleoprotein polypeptide E(SNRPE)	247799
68	ENSG00000135486	*HNRNPA1*	heterogeneous nuclear ribonucleoprotein A1(HNRNPA1)	235022
69	ENSG00000092841	*MYL6*	myosin light chain 6(MYL6)	232341
70	ENSG00000189403	*HMGB1*	high mobility group box 1(HMGB1)	232190
71	ENSG00000181163	*NPM1*	nucleophosmin(NPM1)	231304
72	ENSG00000148303	*RPL7A*	ribosomal protein L7a(RPL7A)	228175
73	ENSG00000083845	*RPS5*	ribosomal protein S5(RPS5)	225977
74	ENSG00000198755	*RPL10A*	ribosomal protein L10a(RPL10A)	225131
75	ENSG00000163041	*H3F3A*	H3 histone family member 3A(H3F3A)	221907
76	ENSG00000087086	*FTL*	ferritin light chain(FTL)	219103

Table 3.4 The most highly expressed transcripts in the HSC dataset (n=119)

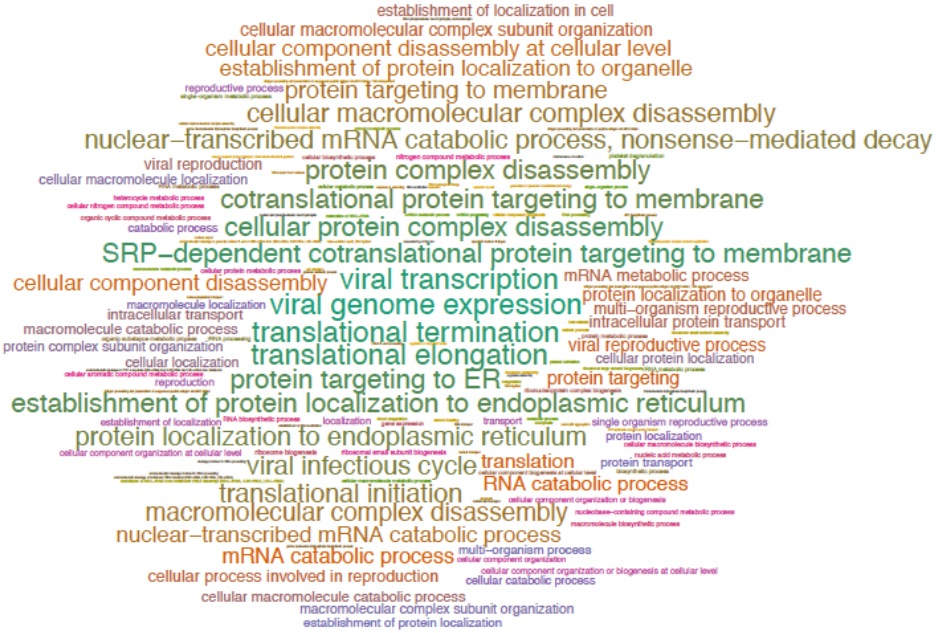
77	ENSG00000196262	*PPIA*	peptidylprolyl isomerase A(PPIA)	217041
78	ENSG00000162244	*RPL29*	ribosomal protein L29(RPL29)	214130
79	ENSG00000138326	*RPS24*	ribosomal protein S24(RPS24)	213403
80	ENSG00000164032	*H2AFZ*	H2A histone family member Z(H2AFZ)	212982
81	ENSG00000145425	*RPS3A*	ribosomal protein S3A(RPS3A)	208431
82	ENSG00000137076	*MIR6852*	microRNA 6852(MIR6852)	205803
83	ENSG00000245532	*NEAT1*	nuclear paraspeckle assembly transcript 1 (NEAT1)	205233
84	ENSG00000161016	*MIR6850*	microRNA 6850(MIR6850)	204611
85	ENSG00000182899	*RPL35A*	ribosomal protein L35a(RPL35A)	204500
86	ENSG00000127314	*RAP1B*	RAP1B, member of RAS oncogene family(RAP1B)	201747
87	ENSG00000130255	*RPL37*	ribosomal protein L37(RPL37)	200917
88	ENSG00000063177	*RPL18*	ribosomal protein L18(RPL18)	199397
89	ENSG00000147403	*RPL10*	ribosomal protein L10(RPL10)	198665
90	ENSG00000111640	*GAPDH*	glyceraldehyde-3-phosphate dehydrogenase(GAPDH)	195029
91	ENSG00000136942	*RPL35*	ribosomal protein L35(RPL35)	194164
92	ENSG00000105193	*RPS16*	ribosomal protein S16(RPS16)	193968
93	ENSG00000163682	*RPL9*	ribosomal protein L9(RPL9)	191854
94	ENSG00000070756	*MIR7705*	microRNA 7705(MIR7705)	186143
95	ENSG00000168028	*RPSA*	ribosomal protein SA(RPSA)	185500
96	ENSG00000177600	*RPLP2*	ribosomal protein lateral stalk subunit P2(RPLP2)	184892
97	ENSG00000118181	*RPS25*	ribosomal protein S25(RPS25)	184461
98	ENSG00000111669	*TPI1*	triosephosphate isomerase 1(TPI1)	184247
99	ENSG00000229117	*RPL41*	ribosomal protein L41(RPL41)	183968
100	ENSG00000135535	*CD164*	CD164 molecule(CD164)	182743
101	ENSG00000197958	*RPL12*	ribosomal protein L12(RPL12)	180717
102	ENSG00000171863	*RPS7*	ribosomal protein S7(RPS7)	179684
103	ENSG00000127920	*GNG11*	G protein subunit gamma 11(GNG11)	179018
104	ENSG00000233276	*GPX1*	glutathione peroxidase 1(GPX1)	176487
105	ENSG00000096384	*HSP90AB1*	heat shock protein 90 alpha family class B member 1(HSP90AB1)	169829
106	ENSG00000132475	*MIR4738*	microRNA 4738(MIR4738)	168437
107	ENSG00000168497	*SDPR*	serum deprivation response(SDPR)	168342
108	ENSG00000131469	*RPL27*	ribosomal protein L27(RPL27)	163219
109	ENSG00000114942	*EEF1B2*	eukaryotic translation elongation factor 1 beta 2(EEF1B2)	159002
110	ENSG00000163687	*DNASE1L3*	deoxyribonuclease 1 like 3(DNASE1L3)	158764
111	ENSG00000166441	*RPL27A*	ribosomal protein L27a(RPL27A)	157719
112	ENSG00000187109	*NAP1L1*	nucleosome assembly protein 1 like 1(NAP1L1)	157465
113	ENSG00000198840	*ND3*	NADH dehydrogenase, subunit 3 (complex I)(ND3)	156704
114	ENSG00000122566	*HNRNPA2B1*	heterogeneous nuclear ribonucleoprotein A2/B1(HNRNPA2B1)	156375
115	ENSG00000125691	*RPL23*	ribosomal protein L23(RPL23)	154183
116	ENSG00000169567	*HINT1*	histidine triad nucleotide binding protein 1(HINT1)	152518
117	ENSG00000233927	*RPS28*	ribosomal protein S28(RPS28)	152123
118	ENSG00000084207	*GSTP1*	glutathione S-transferase pi 1(GSTP1)	151430
119	ENSG00000005961	*ITGA2B*	integrin subunit alpha 2b(ITGA2B)	150685
120	ENSG00000169756	*LIMS1*	LIM zinc finger domain containing 1(LIMS1)	150341
121	ENSG00000204525	*HLA-C*	major histocompatibility complex, class I, C(HLA-C)	150220
122	ENSG00000174444	*RPL4*	ribosomal protein L4(RPL4)	149572
123	ENSG00000196531	*NACA*	nascent polypeptide-associated complex alpha subunit(NACA)	148095
124	ENSG00000124172	*ATP5E*	ATP synthase, mitochondrial F1 complex, epsilon subunit(ATP5E)	147228
125	ENSG00000170889	*RPS9*	ribosomal protein S9(RPS9)	145452
126	ENSG00000109971	*HSPA8*	heat shock protein family A (Hsp70) member 8(HSPA8)	143774
127	ENSG00000111348	*ARHGDIB*	Rho GDP Dissociation Inhibitor Beta (ARHGDIB)	143114
128	ENSG00000166913	*YWHAB*	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase	141312
129	ENSG00000089009	*RPL6*	ribosomal protein L6(RPL6)	139467
130	ENSG00000104904	*OAZ1*	ornithine decarboxylase antizyme 1(OAZ1)	139057
131	ENSG00000075415	*SLC25A3*	solute carrier family 25 member 3(SLC25A3)	136085
132	ENSG00000188846	*RPL14*	ribosomal protein L14(RPL14)	135401
133	ENSG00000167658	*EEF2*	eukaryotic translation elongation factor 2(EEF2)	133525
134	ENSG00000110799	*VWF*	von Willebrand factor(VWF)	132184
135	ENSG00000019582	*CD74*	CD74 molecule(CD74)	132044
136	ENSG00000189043	*HMGB1*	high mobility group box 1(HMGB1)	131818
137	ENSG00000198695	*ND6*	NADH dehydrogenase, subunit 6 (complex I)(ND6)	131303
138	ENSG00000135218	*CD36*	CD36 molecule(CD36)	130215
139	ENSG00000164104	*HMGB2*	high mobility group box 2(HMGB2)	128054
140	ENSG00000163466	*ARPC2*	actin related protein 2/3 complex subunit 2(ARPC2)	127487
141	ENSG00000206503	*HLA-A*	major histocompatibility complex, class I, A(HLA-A)	127440
142	ENSG00000111716	*LDHB*	lactate dehydrogenase B(LDHB)	125782
143	ENSG00000171858	*RPS21*	ribosomal protein S21(RPS21)	124758
144	ENSG00000173812	*EIF1*	eukaryotic translation initiation factor 1(EIF1)	122384
145	ENSG00000074800	*ENO1*	enolase 1(ENO1)	117956
146	ENSG00000140319	*SRP14*	signal recognition particle 14(SRP14)	117504
147	ENSG00000125356	*NDUFA1*	NDUFA4, mitochondrial complex associated(NDUFA4)	117367
148	ENSG00000198830	*HMGN2*	high mobility group nucleosomal binding domain 2(HMGN2)	117319
149	ENSG00000198918	*RPL39*	ribosomal protein L39(RPL39)	116076
150	ENSG00000115310	*RTN4*	reticulon 4(RTN4)	115917
151	ENSG00000125743	*SNRPD2*	small nuclear ribonucleoprotein D2 polypeptide(SNRPD2)	115695
152	ENSG00000197746	*PSAP*	prosaposin(PSAP)	114857
153	ENSG00000118680	*MYL12B*	myosin light chain 12B(MYL12B)	111497

Table 3.4 The most highly expressed transcripts in the HSC dataset (n=119) Continued

154	ENSG00000113648	*H2AFY*	H2A histone family member Y(H2AFY)	110153
155	ENSG00000131143	*COX4I1*	cytochrome c oxidase subunit 4I1(COX4I1)	108912
156	ENSG00000108654	*MIR3064*	microRNA 3064(MIR3064)	107373
157	ENSG00000213741	*LOC100288910*	uncharacterized LOC100288910(LOC100288910)	107018
158	ENSG00000087460	*GNAS*	GNAS complex locus(GNAS)	106704
159	ENSG00000172809	*RPL38*	ribosomal protein L38(RPL38)	105433
160	ENSG00000123975	*CKS2*	CDC28 protein kinase regulatory subunit 2(CKS2)	105232
161	ENSG00000127184	*MIR3607*	microRNA 3607(MIR3607)	104800
162	ENSG00000169714	*CNBP*	CCHC-type zinc finger nucleic acid binding protein(CNBP)	104638
163	ENSG00000070081	*LOC105376575*	uncharacterized LOC105376575(LOC105376575)	103918
164	ENSG00000196924	*FLNA*	filamin A(FLNA)	103811
165	ENSG00000114391	*RPL24*	ribosomal protein L24(RPL24)	102943
166	ENSG00000213719	*CLIC1*	chloride intracellular channel 1(CLIC1)	102908
167	ENSG00000132341	*RAN*	RAN, member RAS oncogene family(RAN)	102761
168	ENSG00000021355	*SERPINB1*	serpin family B member 1(SERPINB1)	102195
169	ENSG00000143549	*TPM3*	tropomyosin 3(TPM3)	101822
170	ENSG00000109332	*UBE2D3*	ubiquitin conjugating enzyme E2 D3(UBE2D3)	101208
171	ENSG00000164919	*COX6C*	cytochrome c oxidase subunit 6C(COX6C)	101060
172	ENSG00000145741	*BTF3*	basic transcription factor 3(BTF3)	98668
173	ENSG00000101182	*PSMA7*	proteasome subunit alpha 7(PSMA7)	98618
174	ENSG00000142168	*SOD1*	serine and arginine rich splicing factor 11(SRSF11)	97214
175	ENSG00000136156	*ITM2B*	superoxide dismutase 1, soluble(SOD1)	97021
176	ENSG00000116754	*SRSF11*	integral membrane protein 2B(ITM2B)	95966
177	ENSG00000131747	*TOP2A*	topoisomerase (DNA) II alpha(TOP2A)	95964
178	ENSG00000152795	*HNRNPDL*	heterogeneous nuclear ribonucleoprotein D like(HNRNPDL)	94057
179	ENSG00000122545	*SEPT7*	septin 7(SEPT7)	93553
180	ENSG00000154518	*ATP5G3*	ATP synthase,mitochondrial Fo complex subunit C3 (ATP5G3)	93362
181	ENSG00000112081	*SRSF3*	serine and arginine rich splicing factor 3(SRSF3)	93253
182	ENSG00000139644	*TMBIM6*	transmembrane BAX inhibitor motif containing 6(TMBIM6)	93032
183	ENSG00000265972	*TXNIP*	thioredoxin interacting protein(TXNIP)	91404
184	ENSG00000255302	*EID1*	EP300 interacting inhibitor of differentiation 1(EID1)	91241
185	ENSG00000135390	*ATP5G2*	ATP synthase, mitochondrial Fo complex subunit C2(ATP5G2)	91141
186	ENSG00000175063	*UBE2C*	ubiquitin conjugating enzyme E2 C(UBE2C)	89829
187	ENSG00000049323	*LTBP1*	latent transforming growth factor beta binding protein 1(LTBP1)	88971
188	ENSG00000120885	*MIR6843*	microRNA 6843(MIR6843)	88778
189	ENSG00000131171	*SH3BGRL*	SH3 domain binding glutamate rich protein like(SH3BGRL)	88746
190	ENSG00000034510	*TMSB10*	thymosin beta 10(TMSB10)	88240
191	ENSG00000068796	*KIF2A*	kinesin family member 2A(KIF2A)	87845
192	ENSG00000142669	*SH3BGRL3*	SH3 domain binding glutamate rich protein like 3(SH3BGRL3)	87166
193	ENSG00000149925	*ALDOA*	aldolase, fructose-bisphosphate A(ALDOA)	86756
194	ENSG00000185201	*IFITM2*	interferon induced transmembrane protein 2(IFITM2)	86729
195	ENSG00000100650	*SRSF5*	serine and arginine rich splicing factor 5(SRSF5)	85729
196	ENSG00000140264	*SERF2*	small EDRK-rich factor 2(SERF2)	85443
197	ENSG00000110955	*ATP5B*	ATP synthase, mitochondrial F1 complex, beta	85203
198	ENSG00000167978	*SRRM2*	serine/arginine repetitive matrix 2(SRRM2)	84996
199	ENSG00000135404	*CD63*	CD63 molecule(CD63)	84190
200	ENSG00000198668	*CALM1*	calmodulin 1(CALM1)	84131

Table 3.4 The most highly expressed transcripts in the HSC dataset (n=119) Continued

A



B

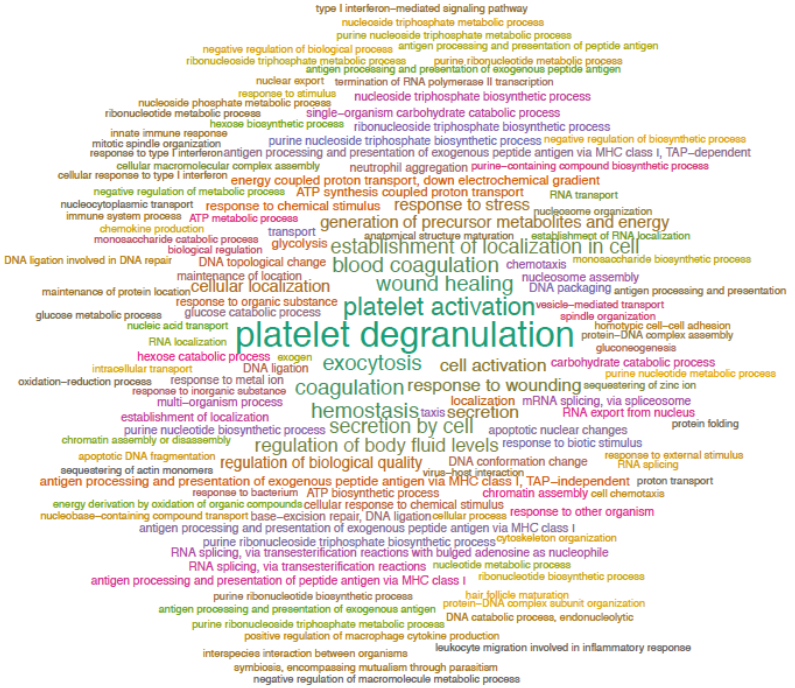


Figure 3.14 Over-represented gene ontologies in the highly expressed HSC gene set

The full GO analysis may be found in Appendix 1. The word cloud is a visual representation of the over-represented GO biological processes with their size an indicator of significance for the highly expressed HSC gene set (A) and the same gene set with ribosomal transcripts removed (B).

3.2.4.1 Gene expression profiling reveals 5 clusters within the HSC population

To identify subpopulations within the HSC population to explain heterogeneity, unsupervised clustering was performed based on Pearson's correlation of the cells within the PCA space (i.e., a quantification of correlation between the cells based on the principal components of variation in gene expression). This revealed 5 clusters (Figure 3.15a). The silhouette plot demonstrates how well each cell fits into its cluster (Figure 3.15b). Cluster 1 (n=46), cluster 2 (n=39) and cluster 5 (n=10) are the most homogeneous whereas cluster 4 (n=10) appears to have 2 subpopulations and cluster 3 (n=14) is the most heterogeneous. The PCA plots shown in Figure 3.15c demonstrate the separation of clusters 1-5 according to the first 4 components that contribute most to the variance with cells with similar expression profiles located closer together. Figure 3.15d demonstrates that PC1 and PC2 are not explained by patient or plate ID and shows a good spread of cells from all patients/plates across the clusters.

In order to further characterise the relationship between distinct single cell clusters the Monocle 2 single-cell analysis toolset (Qiu et al. 2017) was used to apply an order to cells in terms of differential gene expression. Here, the sequence of changes in gene expression is analysed from one cluster to another in an unsupervised manner, thereby allowing each single cell to be placed at the appropriate position along that trajectory depending on differing gene expression. This model builds trajectories via which differentiation may occur between cell populations where trajectory branches are considered as the point at which cell fate decisions are made. The resulting unbiased trajectory has ordered cells as shown in Figure 3.16 where clusters 2 and 5 are furthest apart with clusters 1 and 4 branching from the same trajectory. Cells in cluster 3 appear to remain near the root of the trajectory.

3.2.4.2 Identification of marker genes

Each cluster was first assessed in terms of the most highly expressed transcripts. The most striking observation was that there was a higher expression of megakaryocyte/platelet associated genes in cluster 4 compared to other clusters (16%; 8 out of 50 most highly expressed, Appendix 3, Table S3.2). Indeed, when examined in terms of GO categories (Appendix 3, Figure S3.1) the GO term platelet degranulation was the most significantly enriched category in cluster 4. However, this analysis was superseded by the identification of genes specific to each cluster, or "cluster marker genes".

Marker genes for each cluster were identified as those that were consistently differentially expressed in one cluster compared to the others (Wilcoxon Rank test $p<0.001$ (Kiselev et al. 2017)). Figure 3.17 shows the top 20 marker genes for each cluster visualised by heatmap indicating distinct expression profiles and functional states. Thus, cells from the cluster of interest can be identified by their marker genes that are systematically up-regulated. No

marker gene set was identified for cluster 3, as there were no genes that met the set threshold for significance, which is consistent with the Pearson correlation plot (Figure 3.15).

The function of the 20 marker genes for each cluster as well as the platelet expression of the protein encoded by each gene was assessed using the online GeneCards tool (www.genecards.org). This is shown in Table 3.4. The full list of marker genes for each cluster is presented in Appendix 3, Table S3.3; cluster 1: n=313, cluster 2: n=108, cluster 4: n=19, cluster 5: n=62. Cluster 1 marker genes include several known megakaryocyte and platelet associated genes: *ANGPT1, IL1b, HEMGN, GUCY1A3, NPM1, TSPAN32, NRIP1*. Other megakaryocyte lineage associated genes present in the cluster 1 marker gene list (not top 20) include *SELP, SERPINB1, Gfl1B, NFE2*. Of the 19 cluster 4 marker genes, 13 show high protein expression in platelets and megakaryocytes including *TUBB1, CCL5, CDKN1A, CDKN2D, TUBA4A, TMEM40, F13A1, GNAZ, ZNF185, SNN, HIST1H2AC, PTCRA, GRAP2*. Cluster 2 marker genes have functions primarily related to angiogenesis, with none showing significant protein expression in platelets. Cluster 5 marker genes are related to the immune response although 2 of the marker genes give rise to protein expression in platelets: *SERPINA1, S100P*.

The data was then compared to the most recently published haematopoietic cell RNA-seq gene set, the Blueprint progenitor database, where RNA was sequenced from eight human primary haematopoietic progenitor populations including erythroblasts and megakaryocytes both cultured from primary CD34+ cells (L. Chen et al. 2014). Expression of the putative marker genes for each cluster within the Blueprint populations is demonstrated in heatmap form in Figure 3.18. Marker genes for HSC cluster 1 are seen to be highly expressed in megakaryocytes as well as the megakaryocyte precursors according to the classical model for hierarchical haematopoietic differentiation: MEP, GMP, CMP, MPP and HSC, but not in the erythroblast population which also arises from the MEP or the CLP. Cluster 4 marker genes are also expressed at high levels in megakaryocytes within the Blueprint dataset with significantly lower expression in the other progenitor populations.

To assess the expression of known haematopoietic cell type gene signatures within the clustered single cell HSCs and how this was ordered in terms of differentiation trajectories, Monocle 2 was used to overlay different cell type gene signatures from the recently published DMAP database, where expression arrays were performed on 38 haematopoietic cell populations (Novershtern et al. 2011), onto the previously constructed unbiased trajectory in Figure 3.16. The megakaryocyte signature is largely restricted to clusters 1 and 4 which branch from the same trajectory (Figure 3.19a). In keeping with the data shown in Figure 3.18 the MEP signature is restricted to cluster 1 only while these genes are downregulated in the other megakaryocyte like cluster 4 which has a similar expression of the MEP signature to

clusters 2, 3 and 5 (Figure 3.19b). The monocyte signature is expressed most highly in cluster 5 and B cell signature most highly in cluster 2 (Figure 3.19c&d).

3.2.4.3 Gene ontology of cluster marker genes

To investigate the biological themes that are differentially represented according to HSC cluster a GO analysis was performed. A comparison based on the biological processes of the cluster marker genes highlights the differences in lineage potential of these HSC subsets. The full datasets may be found in Appendix 1 with over-represented GO categories identified using a low stringency statistical cut-off (<0.1), therefore the categories with lower significance levels should be regarded as qualitative only. The data shown in the representative word clouds in Figures 3.20 and 3.21 are significant ($p<0.05$).

Cluster 1 transcriptome appears to have an ontological bias towards translation, ribosomal function and protein localisation with 25% (77 out of 313) of cluster 1 marker genes encoding ribosomal proteins (Figure 3.20, Appendix 3, Table S3.4). Apart from ribosomal functionality other enriched GO biological processes in cluster 1 are related to cellular metabolism, coagulation and haemostasis. The cluster 2 marker genes appear to have an ontological bias towards angiogenesis, endothelial cell function and wound healing (Figure 3.20, Appendix 3, Table S3.5). The over-represented GO categories for cluster 4 are related to negative regulation of cell death/apoptosis, cell cycle, protein folding and exocytosis (Figure 3.20, Appendix 3, Table S3.6). The cluster 5 gene signature showed over-representation of categories related to immunity and leukocyte function (Figure 3.20, Appendix 3, Table S3.7). Specific categories relating to a spectrum of immune cells are present: neutrophil chemotaxis/aggregation, phagocytosis, fungal response, antigen processing and presentation, mast cell degranulation.

3.2.4.4 Expression of VWF

Taqman RT-qPCR was previously used to correlate *VWF* transcript expression with surface expression of VWF by flow cytometry demonstrating a weak correlation only (Figure 3.8). However, this was based on all HSCs prior to sequencing. In order to assess true correlation at the single cell level, *VWF* transcript expression was correlated with index FACS data in only high-quality cells post cell filtering (n=119). Figure 3.21 shows VWF surface expression plotted against and its transcription within the HSC as well as cluster information. According to compensation controls on flow cytometry, positive VWF surface expression was defined as MFI>4000. Interestingly, within clusters 1 and 2 which both show high expression of *VWF* (ranked: 238 cluster 1; 22 cluster 2) there is some correlation between surface and transcript expression. However, this is not the case across all cells. It should be noted that despite correlation within clusters 1 and 2 very few of the cells would be defined as showing positive

surface expression by flow cytometry (MFI>4000).

As described above gene expression of *VWF* has been identified as a marker for a platelet/megakaryocyte biased HSC population in the mouse (Sanjuan-Pla et al. 2013). While *VWF* ranked highly within the most abundantly expressed genes within the HSC population (rank: 134) as well as in clusters 1 and 2, it did not emerge as a cluster marker gene. To assess if the transcription of *VWF* could define a functional subpopulation a differential expression analysis was performed between HSCs with transcript expression of *VWF* and those without. This analysis demonstrated no significant differences in gene expression between *VWF* transcript positive and negative cells (data not shown).

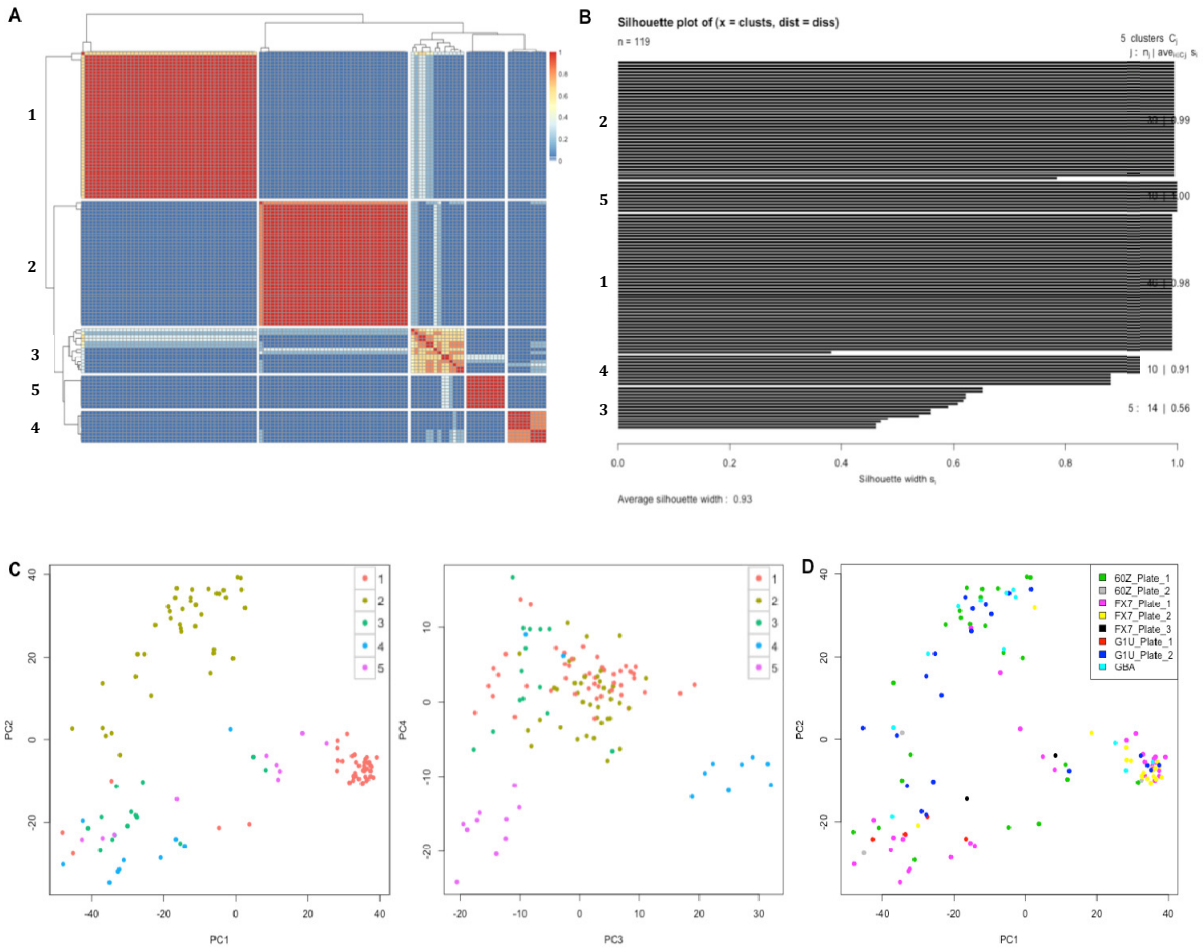


Figure 3.15 Unsupervised clustering of HSCs based on highly variable gene expression.
A. Pearson's correlation map of HSCs in unsupervised clusters by distance within the PCA space, dendrograms are formed by hierarchical clustering on the Euclidean distances between cells; B. Silhouette plot showing how similar each cell is to cells in its cluster, compared to other clusters, Silhouette widths -1 to 1; C. PCA plot constructed from normalized log-expression values of correlated highly variable genes, where each point represents a cell in the HSC dataset. First, second third and fourth components are shown, along with the percentage of variance explained; D. PC1/PC2 plot shown with patient and plate ID identified for each cell.

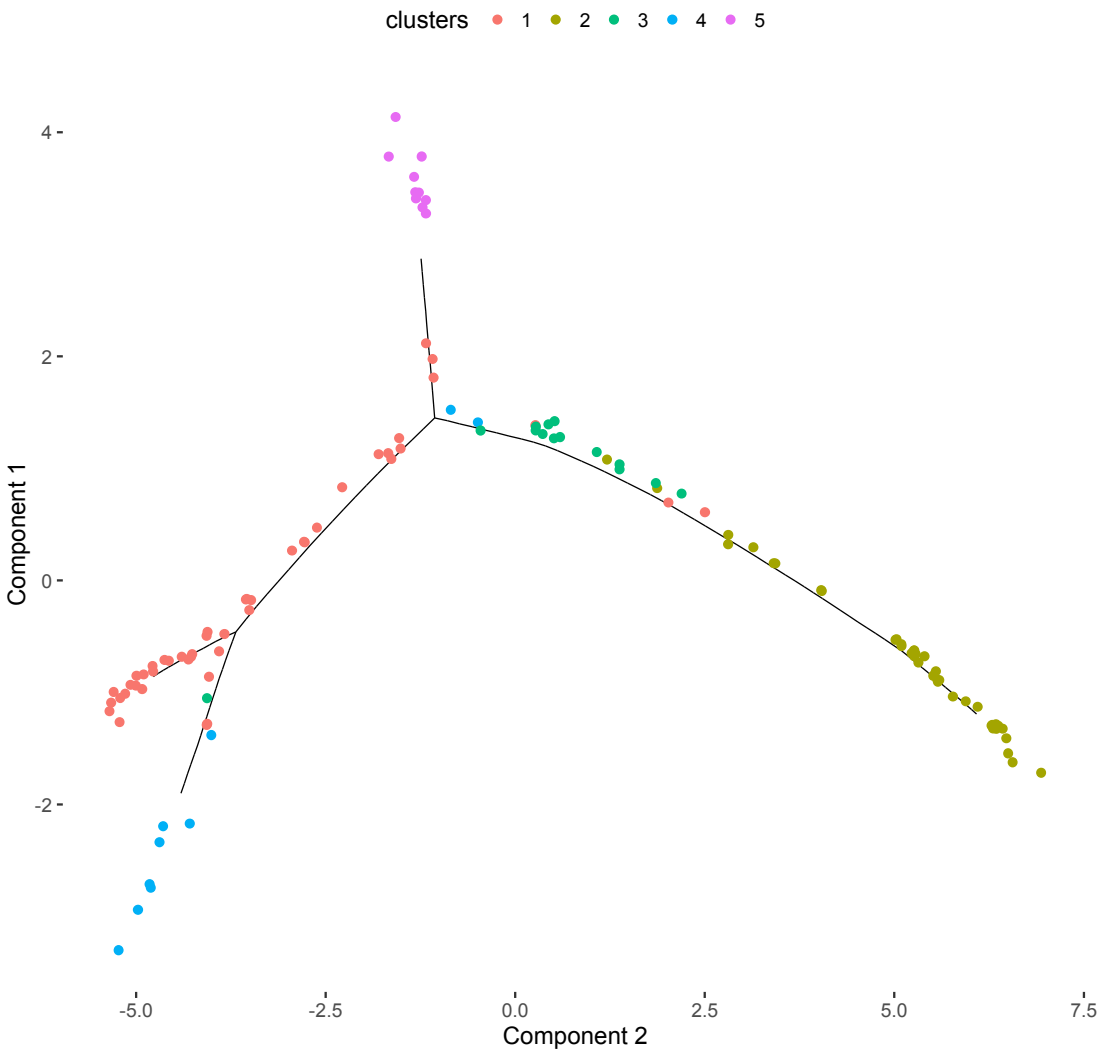


Figure 3.16 Ordering single cell differentiation between cell clusters using Monocle 2. Individual cells are connected by a minimum spanning tree with branch points (thin lines), representing the differentiation trajectory and clusters are differentiated by colour.

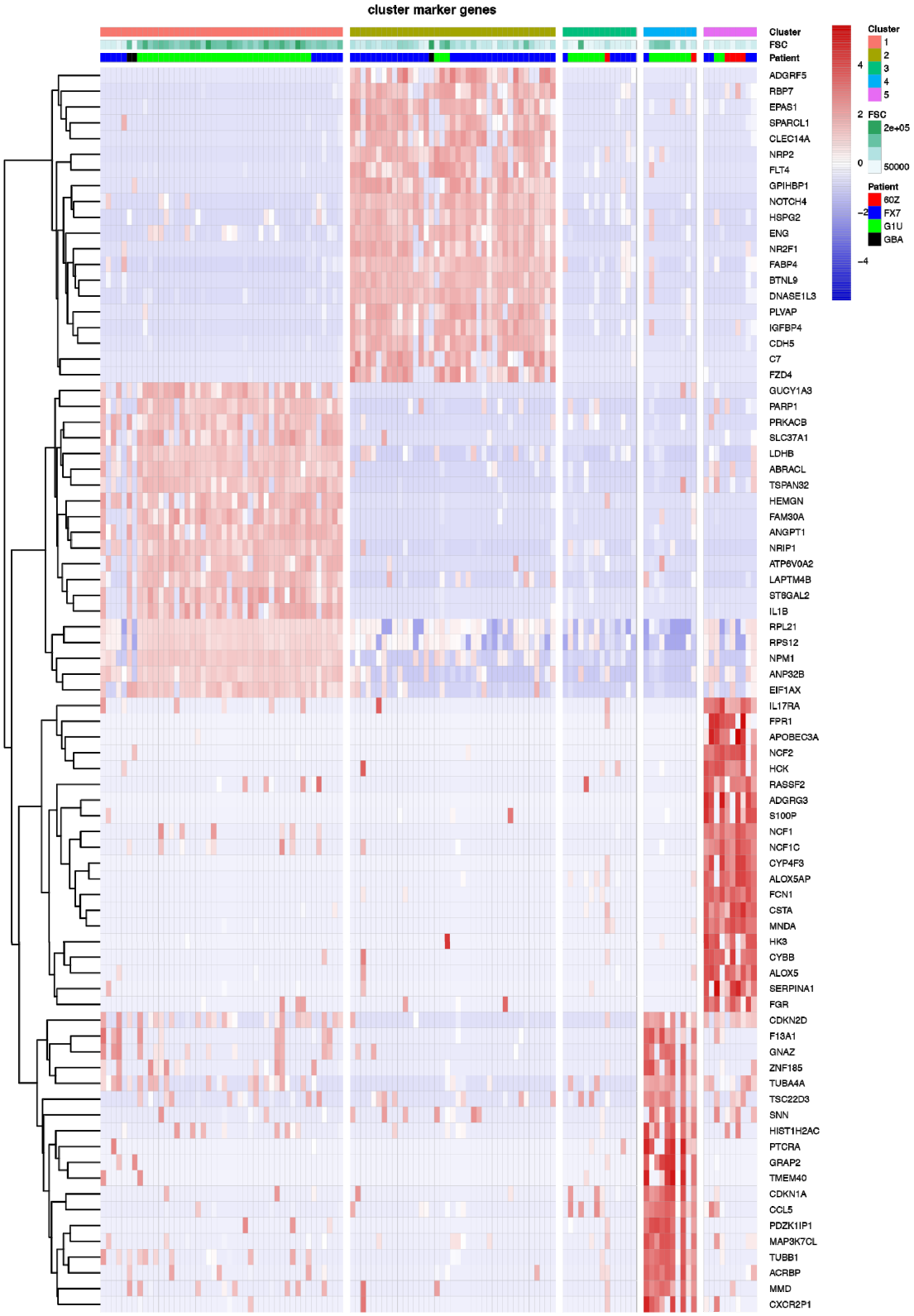


Figure 3.17 Heatmap to show top marker gene expression according to clusters. Heatmap of mean-centred normalized and corrected log-expression values for the top 20 marker genes for each cluster. Dendrograms are formed by hierarchical clustering on the Euclidean distances between genes (row). Column colours represent the cluster to which each cell is assigned. Each cell is indexed by size (forward scatter) and patient and plate ID.

Cluster	ENSEMBL_GENE_ID	HGNC_symbol	Gene_name	P_value	Platelet Protein	Gene_ biotype	Function
1	ENSG00000154188	*ANGPT1*	angiopoietin 1(ANGPT1)	6.09E-18	x	Protein coding	Mediates cAMP-dependent signalling triggered by receptor binding to GPCRs. Regulates diverse cellular processes such as cell proliferation, the cell cycle, differentiation and regulation of microtubule dynamics, chromatin condensation and decondensation, nuclear envelope disassembly and reassembly, regulation of intracellular transport mechanisms and ion flux.
1	ENSG00000226777	*FAM30A*	Family With Sequence Similarity 30 Member A (FAM30A)	1.37E-15		RNA gene	Affiliated with the non-coding RNA class.
1	ENSG00000180530	*NRIP1*	nuclear receptor interacting protein 1(NRIP1)	3.10E-15	xs	Protein coding	Modulates transcriptional activation by steroid receptors such as NR3C1, NR3C2 and ESR1. Also modulates transcriptional repression by nuclear hormone receptors. Positive regulator of the circadian clock gene expression.
1	ENSG00000146386	*ABRACL*	ABRA C-terminal like(ABRACL)	6.73E-15	x	Protein coding	
1	ENSG00000164116	*GUCY1A3*	guanylate cyclase 1 soluble subunit alpha(GUCY1A3)	1.11E-14	x	Protein coding	Guanylyl Cyclases (GC) are a group of enzymes that, in the presence of a metal ion co-factor such as Mg2+ or Mn2+, convert guanosine-5'-triphosphate (GTP) into 3',5'-guanosine monophosphate (cGMP) and pyrophosphate.
1	ENSG00000144057	*ST6GAL2*	ST6 beta-galactoside alpha-2,6-sialyltransferase 2(ST6GAL2)	2.43E-14		Protein coding	Transfers sialic acid from the donor of substrate CMP-sialic acid to galactose containing acceptor substrates. Has alpha-2,6-sialyltransferase activity toward oligosaccharides that have the Gal-beta-1,4-GlcNAc sequence at the non-reducing end of their carbohydrate groups, but it has weak or no activities toward glycoproteins and glycolipids.
1	ENSG00000136929	*HEMGN*	hemogen(HEMGN)	6.60E-14	x	Protein coding	Regulates the proliferation and differentiation of hematopoietic cells. Overexpression block the TPA-induced megakaryocytic differentiation in the K562 cell model. May also prevent cell apoptosis through the activation of the nuclear factor-kappa B
1	ENSG00000125538	*IL1B*	interleukin 1 beta(IL1B)	6.99E-14	x	Protein coding	Potent proinflammatory cytokine. Initially discovered as the major endogenous pyrogen, induces prostaglandin synthesis, neutrophil influx and activation, T/B-cell activation and cytokine production and antibody production, and fibroblast proliferation and collagen production.
1	ENSG00000142875	*PRKACB*	protein kinase cAMP-activated catalytic subunit beta(PRKACB)	1.43E-13	x	Protein coding	Mediates cAMP-dependent signaling triggered by receptor binding to GPCRs. PKA activation regulates diverse cellular processes such as cell proliferation, the cell cycle, differentiation and regulation of microtubule dynamics, chromatin condensation and decondensation, nuclear envelope disassembly and reassembly, as well as regulation of intracellular transport mechanisms and ion flux.
1	ENSG00000111716	*LDHB*	lactate dehydrogenase B(LDHB)	1.60E-13	x	Protein coding	This gene encodes the B subunit of lactate dehydrogenase enzyme, which catalyses the interconversion of pyruvate and lactate with concomitant interconversion of NADH and NAD+ in a post-glycolysis process.
1	ENSG00000160190	*SLC37A1*	solute carrier family 37 member 1(SLC37A1)	2.99E-13	x	Protein coding	Inorganic phosphate and glucose-6-phosphate antiporter. May transport cytoplasmic glucose-6-phosphate into the lumen of the endoplasmic reticulum and translocate inorganic phosphate into the opposite direction. Independent of a luminal glucose-6-phosphatase.
1	ENSG00000185344	*ATP6V0A2*	ATPase H+ transporting V0 subunit a2(ATP6V0A2)	3.04E-13		Protein coding	Part of the proton channel of V-ATPases. Essential component of the endosomal pH-sensing machinery. May play a role in maintaining the Golgi functions, such as glycosylation maturation, by controlling the Golgi's pH.
1	ENSG00000181163	*NPM1*	nucleophosmin(NPM1)	3.96E-13		Protein coding	Involved in diverse cellular processes such as ribosome biogenesis, centrosome duplication, protein chaperoning, histone assembly, cell proliferation, and regulation of tumour suppressors p53/TP53 and ARF. Binds ribosome presumably to drive ribosome nuclear export. Associated with nucleolar ribonucleoprotein structures and bind single-stranded nucleic acids.
1	ENSG00000136938	*ANP32B*	acidic nuclear phosphoprotein 32 family member B(ANP32B)	6.40E-13		Protein coding	Multifunctional protein working as a cell cycle progression factor as well as a cell survival factor. Required for the progression from the G1 to the S phase. Anti-apoptotic protein which functions as a caspase-3 inhibitor. Exhibits histone chaperone properties, stimulating core histones to assemble into a nucleosome.
1	ENSG00000064201	*TSPAN32*	tetraspanin 32(TSPAN32)	8.22E-13	x	Protein	Member of the tetraspanin superfamily, one of several tumour-suppressing sub

						coding	transferable fragments located in the imprinted gene domain of chromosome 11p15.5
1	ENSG00000122026	*RPL21*	ribosomal protein L21(RPL21)	1.07E-12		Protein coding	Encodes a ribosomal protein that is a component of the 60S subunit. The protein belongs to the L21E family of ribosomal proteins. It is located in the cytoplasm.
1	ENSG00000143799	*PARP1*	poly(ADP-ribose) polymerase 1(PARP1)	1.13E-12		Protein coding	Involved in the base excision repair (BER) pathway, by catalysing the poly(ADP-ribosyl)ation of a limited number of acceptor proteins involved in chromatin architecture and in DNA metabolism.
1	ENSG00000104341	*LAPTM4B*	lysosomal protein transmembrane 4 beta(LAPTM4B)	1.15E-12		Protein coding	
1	ENSG00000112306	*RPS12*	ribosomal protein S12(RPS12)	1.18E-12	x	Protein coding	Encodes a ribosomal protein that is a component of the 40S subunit. The protein belongs to the S12E family of ribosomal proteins. It is located in the cytoplasm.
1	ENSG00000173674	*EIF1AX*	eukaryotic translation initiation factor 1A, X-linked(EIF1AX)	1.92E-12		Protein coding	Required for maximal rate of protein biosynthesis. Enhances ribosome dissociation into subunits and stabilizes the binding of the initiator Met-tRNA(I) to 40 S ribosomal subunits.
2	ENSG00000165810	*BTNL9*	butyrophilin like 9(BTNL9)	9.22E-19		Protein coding	Related pathway: Innate Immune System
2	ENSG00000277494	*GPIHBP1*	glycosylphosphatidylinositol anchored high density lipoprotein binding protein 1(GPIHBP1)	2.45E-18		Protein coding	Plays a key role in the lipolytic processing of chylomicrons. Required for the transport of lipoprotein lipase LPL into the capillary lumen
2	ENSG00000175745	*NR2F1*	nuclear receptor subfamily 2 group F member 1(NR2F1)	4.55E-18		Protein coding	The protein encoded by this gene is a nuclear hormone receptor and transcriptional regulator. The encoded protein acts as a homodimer and binds to 5'-AGGTCA-3' repeats.
2	ENSG00000130300	*PLVAP*	plasmalemma vesicle associated protein(PLVAP)	2.49E-17		Protein coding	Involved in the formation of stomatal and fenestral diaphragms of caveolae. May function in microvascular permeability.
2	ENSG00000163687	*DNASE1L3*	deoxyribonuclease 1 like 3(DNASE1L3)	4.45E-17		Protein coding	This gene encodes a member of the deoxyribonuclease I family. The encoded protein hydrolyses DNA, is not inhibited by actin, and mediates the breakdown of DNA during apoptosis.
2	ENSG00000141753	*IGFBP4*	insulin like growth factor binding protein 4(IGFBP4)	1.12E-16		Protein coding	IGF-binding proteins prolong the half-life of the IGFs and have been shown to either inhibit or stimulate the growth promoting effects of the IGFs on cell culture. They alter the interaction of IGFs with their cell surface receptors
2	ENSG00000204301	*NOTCH4*	notch 4(NOTCH4)	1.42E-16		Protein coding	Functions as a receptor for membrane-bound ligands Jagged1, Jagged2 and Delta1 to regulate cell-fate determination. Upon ligand activation through the released notch intracellular domain (NICD) it forms a transcriptional activator complex with RBPJ/RBPSUH and activates genes of the enhancer of split locus. Affects the implementation of differentiation, proliferation and apoptotic programs.
2	ENSG00000179776	*CDH5*	cadherin 5(CDH5)	6.38E-16		Protein coding	Cadherins are calcium-dependent cell adhesion proteins. They preferentially interact with themselves in a homophilic manner in connecting cells; cadherins may thus contribute to the sorting of heterogeneous cell types. This cadherin may play a important role in endothelial cell biology through control of the cohesion and organization of the intercellular junctions. It associates with alpha-catenin forming a link to the cytoskeleton.
2	ENSG00000152583	*SPARCL1*	SPARC like 1(SPARCL1)	1.03E-15		Protein coding	GO annotations related to this gene include calcium ion binding.
2	ENSG00000170323	*FABP4*	fatty acid binding protein 4(FABP4)	1.34E-15		Protein coding	Lipid transport protein in adipocytes. Binds both long chain fatty acids and retinoic acid. Delivers long-chain fatty acids and retinoic acid to their cognate receptors in the nucleus
2	ENSG00000118257	*NRP2*	neuropilin 2(NRP2)	2.12E-15		Protein coding	High affinity receptor for semaphorins 3C, 3F, VEGF-165 and VEGF-145 isoforms of VEGF, and the PLGF-2 isoform of PGF
2	ENSG00000162444	*RBP7*	retinol binding protein 7(RBP7)	2.16E-15		Protein coding	Intracellular transport of retinol.
2	ENSG00000112936	*C7*	complement C7(C7)	2.38E-15		Protein	Constituent of the membrane attack complex (MAC) that plays a key role in the innate and adaptive immune response by forming pores in the plasma membrane of target cells. C7

						coding	serves as a membrane anchor.
2	ENSG00000176435	*CLEC14A*	C-type lectin domain family 14 member A(CLEC14A)	2.70E-15		Protein coding	Member of the C-type lectin/C-type lectin-like domain (CTL/CTLD) superfamily. Members of this family share a common protein fold and have diverse functions, such as cell adhesion, cell-cell signalling, glycoprotein turnover, and roles in inflammation and immune response. This family member plays a role in cell-cell adhesion and angiogenesis.
2	ENSG00000116016	*EPAS1*	endothelial PAS domain protein 1(EPAS1)	2.83E-15		Protein coding	Transcription factor involved in the induction of oxygen regulated genes. Binds to core DNA sequence 5-[AG]CGTG-3 within the hypoxia response element (HRE) of target gene promoters. Regulates the vascular endothelial growth factor (VEGF) expression
2	ENSG00000142798	*HSPG2*	heparan sulfate proteoglycan 2(HSPG2)	3.53E-15		Protein coding	Integral component of basement membranes. Component of the glomerular basement membrane (GBM), responsible for the fixed negative electrostatic membrane charge, and which provides a barrier which is both size- and charge-selective. It serves as an attachment substrate for cells. Plays essential roles in vascularization. Critical for normal heart development and for regulating the vascular response to injury. Also required for avascular cartilage development.
2	ENSG00000106991	*ENG*	endoglin(ENG)	8.59E-15		Protein coding	Major glycoprotein of vascular endothelium. Involved in the regulation of angiogenesis. May play a critical role in the binding of endothelial cells to integrins and/or other RGD receptors. Acts as TGF-beta co-receptor and is involved in the TGF-beta/BMP signalling cascade.
2	ENSG00000037280	*FLT4*	fms related tyrosine kinase 4(FLT4)	1.12E-14		Protein coding	This gene encodes a tyrosine kinase receptor for vascular endothelial growth factors C and D. The protein is thought to be involved in lymphangiogenesis
2	ENSG00000069122	*ADGRF5*	adhesion G protein-coupled receptor F5(ADGRF5)	1.16E-14		Protein coding	Receptor that plays a critical role in lung surfactant homeostasis. May play a role in controlling adipocyte function.
2	ENSG00000174804	*FZD4*	frizzled class receptor 4(FZD4)	1.16E-14		Protein coding	Receptor for Wnt proteins. Most of frizzled receptors are coupled to the beta-catenin (CTNNB1) canonical signalling pathway, which leads to the activation of dishevelled proteins, inhibition of GSK-3 kinase, nuclear accumulation of beta-catenin and activation of Wnt target genes.
4	ENSG00000100351	*GRAP2*	GRB2-related adaptor protein 2(GRAP2)	7.66E-12	x	Protein coding	Interacts with SLP-76 to regulate NF-AT activation. Binds to tyrosine-phosphorylated shc.
4	ENSG00000162366	*PDZK1IP1*	PDZK1 interacting protein 1(PDZK1IP1)	6.50E-11	xs	Protein coding	
4	ENSG00000088726	*TMEM40*	transmembrane protein 40(TMEM40)	2.58E-10	xs	Protein coding	
4	ENSG00000171611	*PTCRA*	pre T-cell antigen receptor alpha(PTCRA)	2.69E-09	xs	Protein coding	The pre-T-cell receptor complex (composed of PTCRA, TCRB and the CD3 complex) regulates early T-cell development.
4	ENSG00000124762	*CDKN1A*	cyclin dependent kinase inhibitor 1A(CDKN1A)	3.25E-09	xs	Protein coding	May be involved in p53/TP53 mediated inhibition of cellular proliferation in response to DNA damage. Binds to and inhibits cyclin-dependent kinase activity, preventing phosphorylation of critical cyclin-dependent kinase substrates and blocking cell cycle progression.
4	ENSG00000101162	*TUBB1*	tubulin beta 1 class VI(TUBB1)	1.82E-08	xs	Protein coding	Tubulin is the major constituent of microtubules. It binds two moles of GTP, one at an exchangeable site on the beta chain and one at a non-exchangeable site on the alpha chain (By similarity).
4	ENSG00000111644	*ACRBP*	acrosin binding protein(ACRBP)	6.63E-08		Protein coding	May be involved in packaging and condensation of the acrosin zymogen in the acrosomal matrix via its association with proacrosin.
4	ENSG00000271503	*CCL5*	C-C motif chemokine ligand 5(CCL5)	1.01E-07	x	Protein coding	Chemoattractant for blood monocytes, memory T-helper cells and eosinophils. Causes the release of histamine from basophils and activates eosinophils. May activate several chemokine receptors
4	ENSG00000108960	*MMD*	monocyte to macrophage differentiation associated(MMD)	3.39E-07		Protein coding	This protein is expressed by *in vitro* differentiated macrophages but not freshly isolated monocytes. Although sequence analysis identifies seven potential transmembrane domains, this protein has little homology to G-protein receptors and it has not been positively identified as a receptor.

4	ENSG00000229754	*CXCR2P1*	C-X-C Motif Chemokine Receptor 2 Pseudogene 1 (CXCR2P1)	1.79E-06		Pseudo gene	
4	ENSG00000180573	*HIST1H2AC*	histone cluster 1 H2A family member c(HIST1H2AC)	9.93E-06		Protein coding	Core component of nucleosome. Nucleosomes wrap and compact DNA into chromatin, limiting DNA accessibility to the cellular machineries which require DNA as a template. Histones thereby play a central role in transcription regulation, DNA repair, DNA replication and chromosomal stability.
4	ENSG00000124491	*F13A1*	coagulation factor XIII A chain(F13A1)	1.46E-05	x	Protein coding	Factor XIII is activated by thrombin and calcium ion to a transglutaminase that catalyses the formation of gamma-glutamyl-epsilon-lysine cross-links between fibrin chains, thus stabilizing the fibrin clot. Also cross-link alpha-2-plasmin inhibitor, or fibronectin, to the alpha chains of fibrin. highly expressed in platelet
4	ENSG00000156265	*MAP3K7CL*	MAP3K7 C-terminal like(MAP3K7CL)	2.49E-05		Protein coding	
4	ENSG00000147394	*ZNF185*	zinc finger protein 185 (LIM domain)(ZNF185)	0.000107 303	x	Protein coding	May be involved in the regulation of cellular proliferation and/or differentiation.
4	ENSG00000127824	*TUBA4A*	tubulin alpha 4a(TUBA4A)	0.001087 649	xs	Protein coding	Tubulin is the major constituent of microtubules. It binds two moles of GTP, one at an exchangeable site on the beta chain and one at a non-exchangeable site on the alpha chain.
4	ENSG00000184602	*SNN*	stannin(SNN)	0.001106 284		Protein coding	Plays a role in endosomal maturation
4	ENSG00000128266	*GNAZ*	G protein subunit alpha z(GNAZ)	0.001416 816	x	Protein coding	Guanine nucleotide-binding proteins (G proteins) are involved as modulators or transducers in various transmembrane signalling systems.
4	ENSG00000129355	*CDKN2D*	cyclin dependent kinase inhibitor 2D(CDKN2D)	0.004843 985	x	Protein coding	This protein has been shown to form a stable complex with CDK4 or CDK6, and prevent the activation of the CDK kinases, thus function as a cell growth regulator that controls cell cycle G1 progression. The abundance of the transcript of this gene was found to oscillate in a cell-cycle dependent manner with the lowest expression at mid G1 and a maximal expression during S phase.
4	ENSG00000157514	*TSC22D3*	TSC22 domain family member 3(TSC22D3)	0.007346 32	x	Protein coding	Protects T-cells from IL2 deprivation-induced apoptosis through the inhibition of FOXO3A transcriptional activity that leads to the down-regulation of the pro-apoptotic factor BCL2L11. In macrophages, plays a role in the anti-inflammatory and immunosuppressive effects of glucocorticoids and IL10. In T-cells, inhibits anti-CD3-induced NFKB1 nuclear translocation.
5	ENSG00000186529	*CYP4F3*	cytochrome P450 family 4 subfamily F member 3(CYP4F3)	3.91E-19		Protein coding	This gene, CYP4F3, encodes a member of the cytochrome P450 superfamily of enzymes. The cytochrome P450 proteins are monooxygenases which catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids.
5	ENSG00000182885	*ADGRG3*	adhesion G protein-coupled receptor G3(ADGRG3)	1.46E-18		Protein coding	Orphan receptor that regulates migration of lymphatic endothelial cells *in vitro* via the small GTPases RhoA and CDC42 (PubMed:24178298). Regulates B-cell developmen
5	ENSG00000085265	*FCN1*	ficolin 1(FCN1)	2.29E-17		Protein coding	Extracellular lectin functioning as a pattern-recognition receptor in innate immunity. Binds the sugar moieties of pathogen-associated molecular patterns (PAMPs) displayed on microbes and activates the lectin pathway of the complement system.
5	ENSG00000121552	*CSTA*	cystatin A(CSTA)	2.69E-17		Protein coding	This is an intracellular thiol proteinase inhibitor. Has an important role in desmosome-mediated cell-cell adhesion in the lower levels of the epidermis.
5	ENSG00000128383	*APOBEC3A*	apolipoprotein B mRNA editing enzyme catalytic subunit 3A(APOBEC3A)	2.20E-16		Protein coding	DNA deaminase (cytidine deaminase) with restriction activity against viruses, foreign DNA and mobility of retrotransposons. Exhibits antiviral activity against adeno-associated virus (AAV) and human T-cell leukemia virus type 1 (HTLV-1) and may inhibit the mobility of LTR and non-LTR retrotransposons.
5	ENSG00000163563	*MNDA*	myeloid cell nuclear differentiation antigen(MNDA)	2.99E-16		Protein coding	May act as a transcriptional activator/repressor in the myeloid lineage. Plays a role in the granulocyte/monocyte cell-specific response to interferon. Stimulates the DNA binding of the transcriptional repressor protein YY1.
5	ENSG00000116701	*NCF2*	neutrophil cytosolic factor 2(NCF2)	6.79E-16		Protein coding	NCF2, NCF1, and a membrane bound cytochrome b558 are required for activation of the latent NADPH oxidase

5	ENSG00000165168	*CYBB*	cytochrome b-245 beta chain(CYBB)	1.12E-15		Protein coding	Critical component of the membrane-bound oxidase of phagocytes that generates superoxide. It is the terminal component of a respiratory chain that transfers single electrons from cytoplasmic NADPH across the plasma membrane to molecular oxygen on the exterior.
5	ENSG00000171051	*FPR1*	formyl peptide receptor 1(FPR1)	1.32E-13		Protein coding	This gene encodes a G protein-coupled receptor of mammalian phagocytic cells that is a member of the G-protein coupled receptor 1 family. The protein mediates the response of phagocytic cells to invasion of the host by microorganisms and is important in host defense and inflammation.
5	ENSG00000012779	*ALOX5*	arachidonate 5-lipoxygenase(ALOX5)	1.60E-13		Protein coding	Catalyzes the first step in leukotriene biosynthesis, and thereby plays a role in inflammatory processes.
5	ENSG00000101336	*HCK*	HCK proto-oncogene, Src family tyrosine kinase(HCK)	2.84E-13		Protein coding	The protein encoded by this gene is a member of the Src family of tyrosine kinases. This protein is primarily hemopoietic, particularly in cells of the myeloid and B-lymphoid lineages. It may help couple the Fc receptor to the activation of the respiratory burst. In addition, it may play a role in neutrophil migration and in the degranulation of neutrophils.
5	ENSG00000197249	*SERPINA1*	serpin family A member 1(SERPINA1)	3.27E-13	x	Protein coding	The protein encoded by this gene is secreted and is a serine protease inhibitor whose targets include elastase, plasmin, thrombin, trypsin, chymotrypsin, and plasminogen activator.
5	ENSG00000158517	*NCF1*	neutrophil cytosolic factor 1(NCF1)	7.11E-13		Protein coding	NCF2, NCF1, and a membrane bound cytochrome b558 are required for activation of the latent NADPH oxidase
5	ENSG00000160883	*HK3*	hexokinase 3(HK3)	9.70E-13		Protein coding	Hexokinases phosphorylate glucose to produce glucose-6-phosphate, the first step in most glucose metabolism pathways. This gene encodes hexokinase 3.
5	ENSG00000101265	*RASSF2*	Ras association domain family member 2(RASSF2)	2.16E-12		Protein coding	Potential tumor suppressor. Acts as a KRAS-specific effector protein. May promote apoptosis and cell cycle arrest.
5	ENSG00000177663	*IL17RA*	interleukin 17 receptor A(IL17RA)	5.61E-12		Protein coding	Interleukin 17A (IL17A) is a proinflammatory cytokine secreted by activated T-lymphocytes. It is a potent inducer of the maturation of CD34-positive hematopoietic precursors into neutrophils. The transmembrane protein encoded by this gene (interleukin 17A receptor; IL17RA) is a ubiquitous type I membrane glycoprotein that binds with low affinity to interleukin 17A. Interleukin 17A and its receptor play a pathogenic role in many inflammatory and autoimmune diseases such as rheumatoid arthritis.
5	ENSG00000132965	*ALOX5AP*	arachidonate 5-lipoxygenase activating protein(ALOX5AP)	7.40E-12		Protein coding	Required for leukotriene biosynthesis by ALOX5 (5-lipoxygenase). Anchors ALOX5 to the membrane. Binds arachidonic acid, and could play an essential role in the transfer of arachidonic acid to ALOX5. Binds to MK-886, a compound that blocks the biosynthesis of leukotrienes.
5	ENSG00000165178	*NCF1C*	neutrophil cytosolic factor 1C pseudogene(NCF1C)	1.00E-11		Pseudo gene	May be required for activation of the latent NADPH oxidase
5	ENSG00000000938	*FGR*	FGR proto-oncogene, Src family tyrosine kinase(FGR)	3.82E-11		Protein coding	Non-receptor tyrosine-protein kinase that transmits signals from cell surface receptors devoid of kinase activity and contributes to the regulation of immune responses, including neutrophil, monocyte, macrophage and mast cell functions, cytoskeleton remodeling in response to extracellular stimuli, phagocytosis, cell adhesion and migration. Promotes mast cell degranulation, release of inflammatory cytokines and IgE-mediated anaphylaxis.
5	ENSG00000163993	*S100P*	S100 calcium binding protein P(S100P)	6.50E-11	x	Protein coding	May function as calcium sensor and contribute to cellular calcium signaling. In a calcium-dependent manner, functions by interacting with other proteins, such as EZR and PPP5C, and indirectly plays a role in physiological processes like the formation of microvilli in epithelial cells.

Table 3.4 HSC cluster top 20 marker genes

Top 20 marker genes for HSC clusters 1,2,4,5, showing significance level (Pvalue), Protein expression in platelets, gene biotype and functional summary from the online tool: Genecards. (Platelet protein: x indicates expressed, xs indicates cell specific expression)

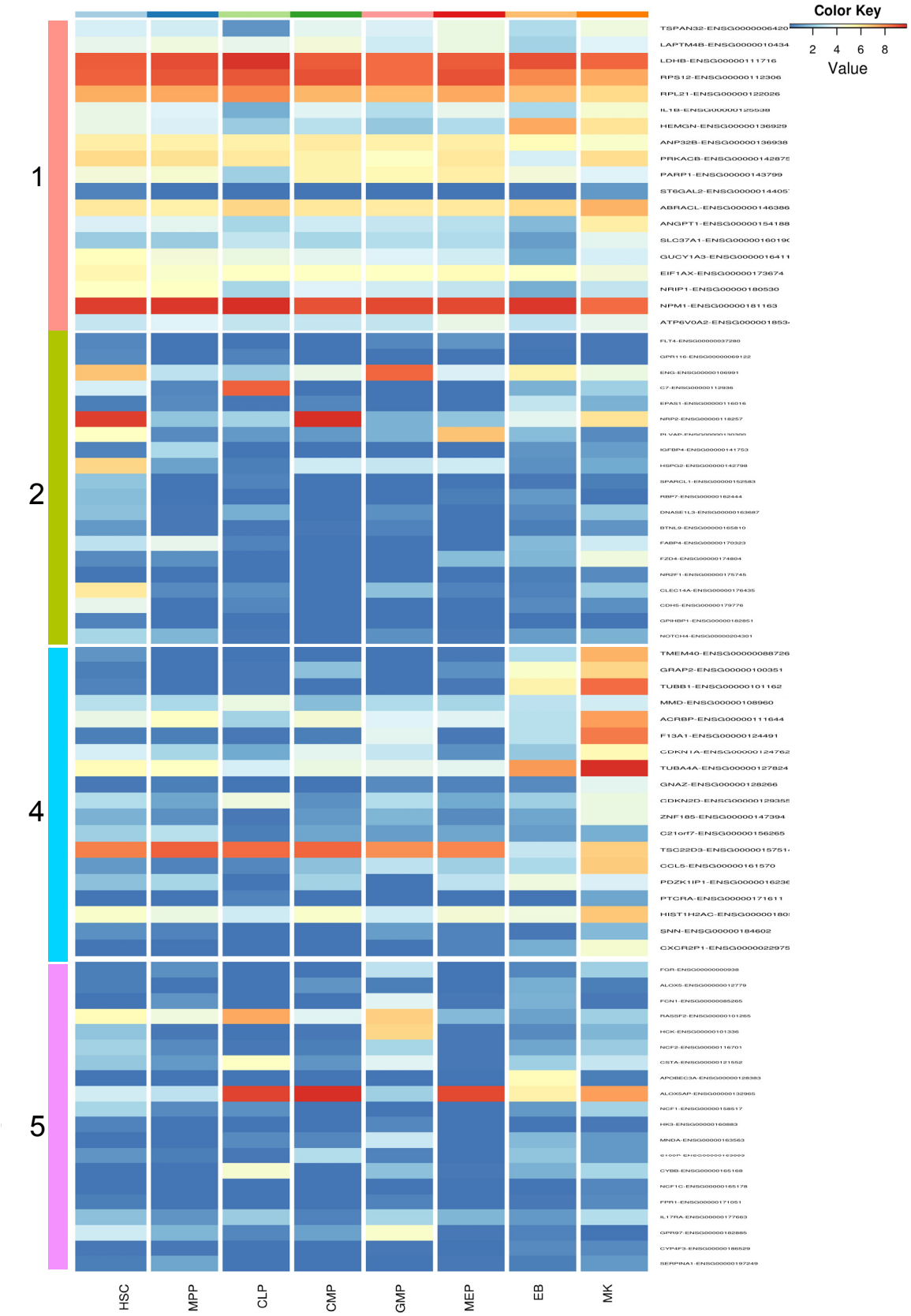


Figure 3.18 Heatmap to show expression of cluster marker genes in Blueprint dataset
Heatmap of mean-centred normalized and corrected log-expression values from the Blueprint progenitors' dataset for the top 20 marker genes for each HSC cluster (rows). Column colours represent 8 progenitor cell populations studied in Blueprint.

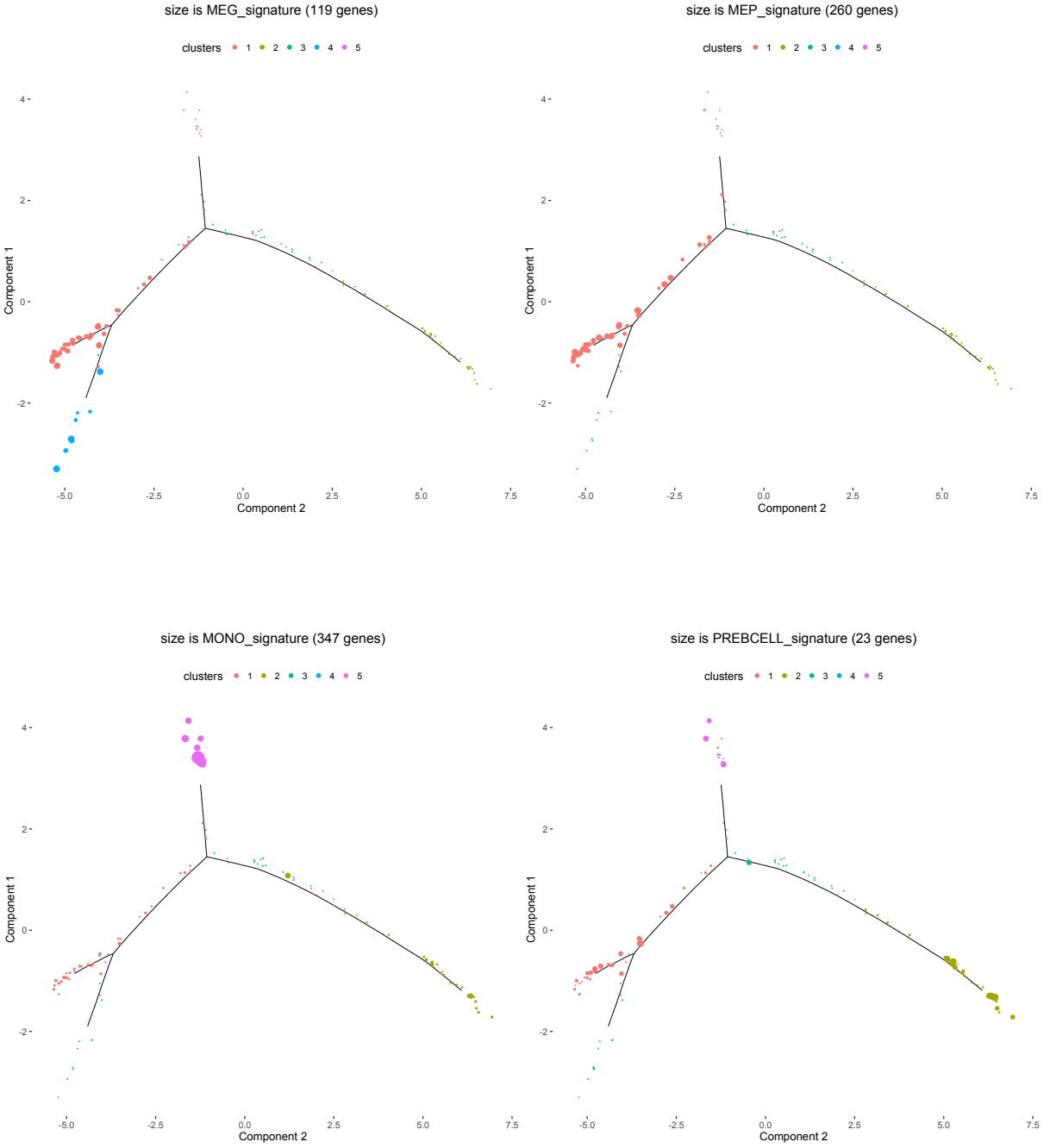


Figure 3.19 Single cell differentiation trajectories using Monocle 2 showing expression of cell type gene signatures from DMAP dataset.
Individual cells are connected by a minimum spanning tree with branch points (thin lines), representing the differentiation trajectory and clusters are differentiated by colour. Expression of gene signature is indicated by the size of each single cell. a. Megakaryocyte signature; b. MEP signature; c. Monocyte signature; d. B cell signature.

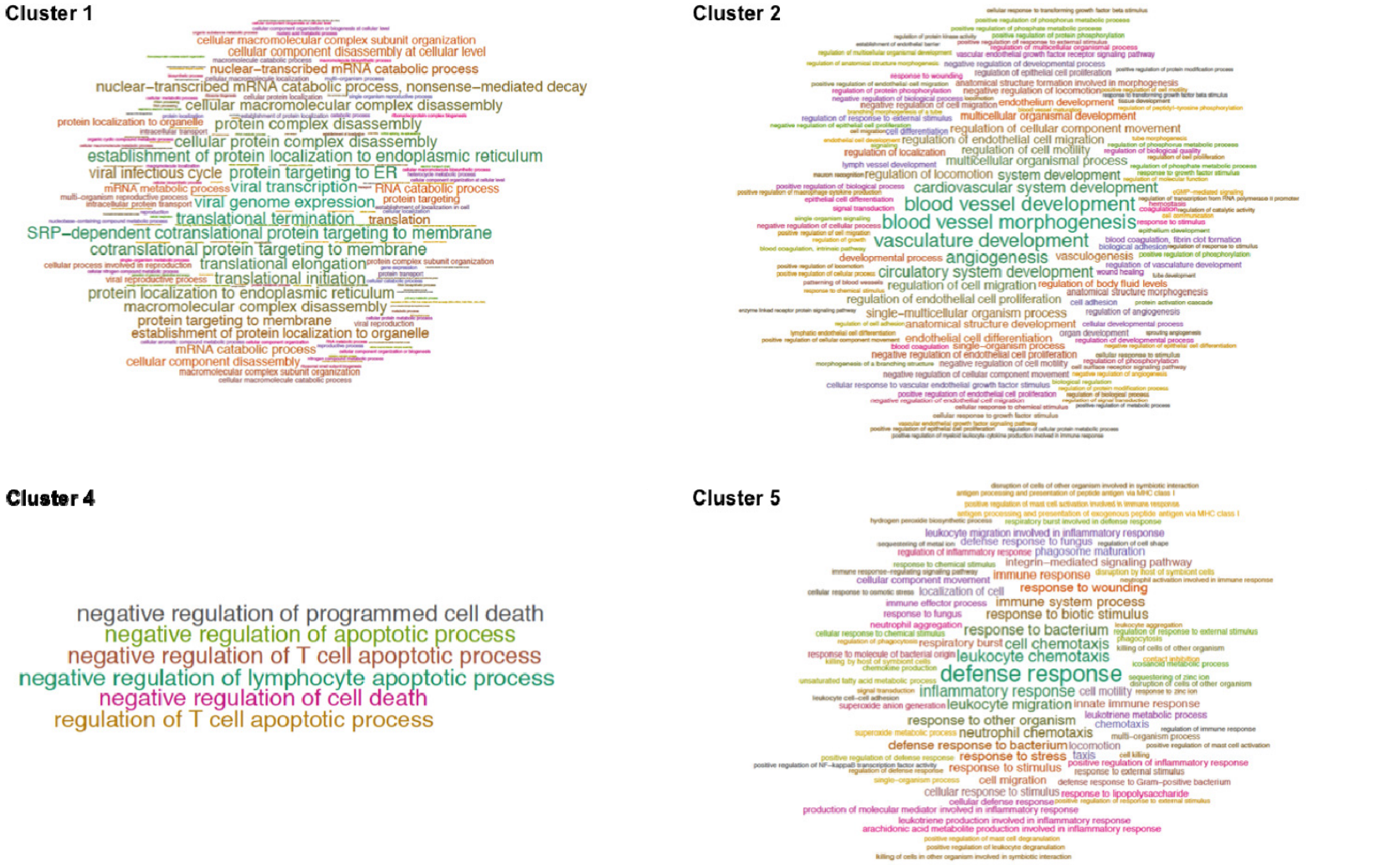


Figure 3.20 Over-represented gene ontologies within each HSC cluster marker genes

The full GO analysis may be found in Appendix 1. The word cloud is a visual representation of the over-represented GO biological processes with their size an indicator of significance for clusters 1,2,4 and 5.

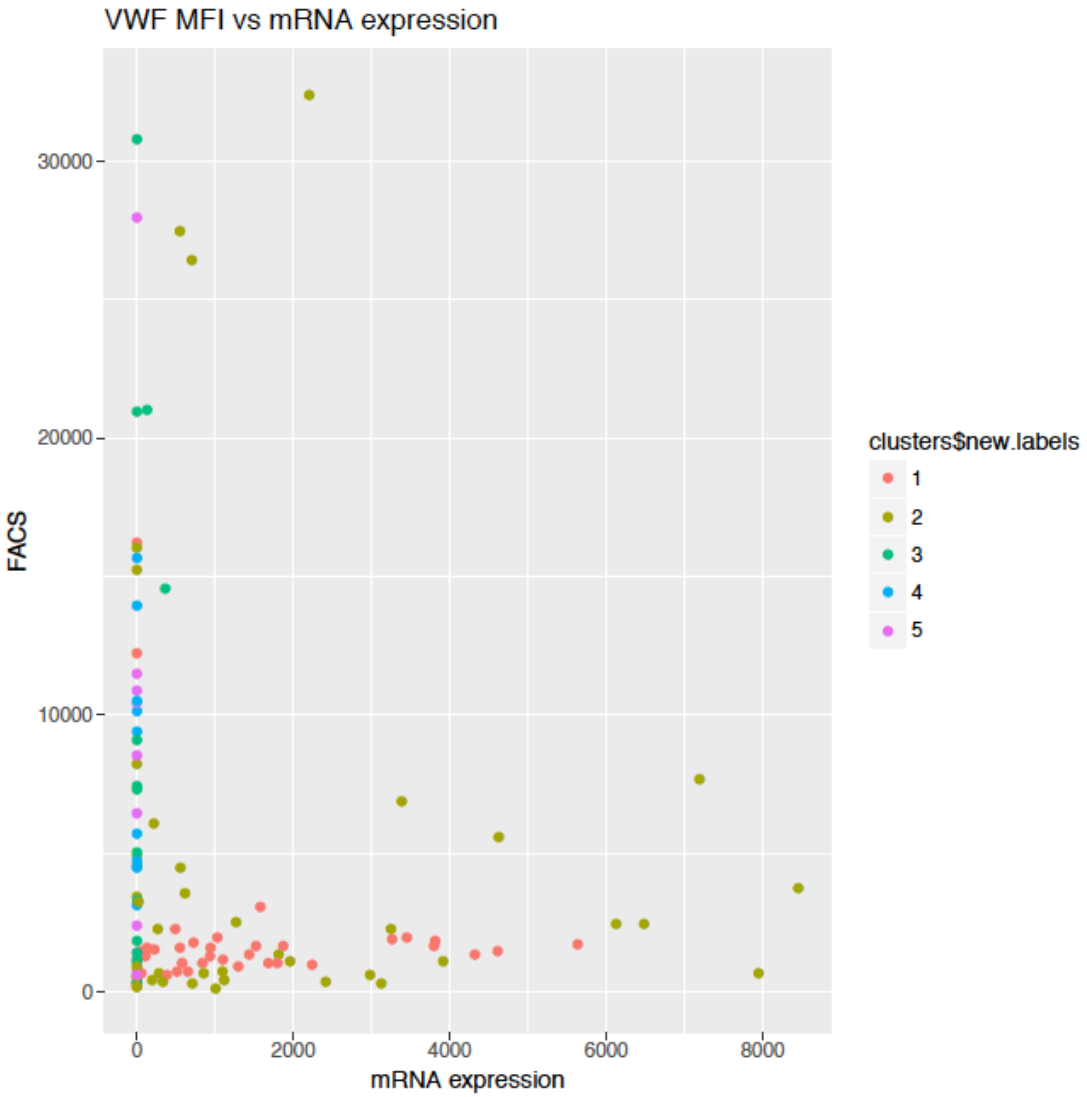


Figure 3.21 Correlation between VWF surface and transcript expression by RNA-seq. VWF surface expression measured by FACS index sorting (MFI, Yaxis) for each HSC plotted against transcript expression by RNA-seq (Xaxis).

3.2.5 Flow cytometric index sorting analysis

In order to prospectively sort the different HSC clusters, a unique pattern of surface marker expression would need to be identified for cells belonging to each cluster. To investigate if there was a correlation between cell surface markers that had been used to FACS sort the HSCs, the indexed flow cytometry data was retrospectively analysed within each gate that was sorted, including VWF expression that was also recorded but not used as criteria on which to sort (Figure 3.22). This shows that the cells taken forward for gene expression analysis on the basis of high quality sequencing data were representative of all cells sorted (low quality cells marked in grey, Figure 3.22). Furthermore, cells from each subset also tend to cluster on the basis of surface marker expression.

Figure 3.23 shows normalised surface marker expression for each individual cluster. Cluster 1 and 4 which are both megakaryocyte-like HSC subsets are similar in expression of CD38, CD45RA, and CD90. However, cluster 4 has higher expression of CD49f and VWF than cluster 1 and lower expression of CD34. On the basis of the data presented it is not possible to prospectively distinguish all clusters based solely on the surface markers used here.

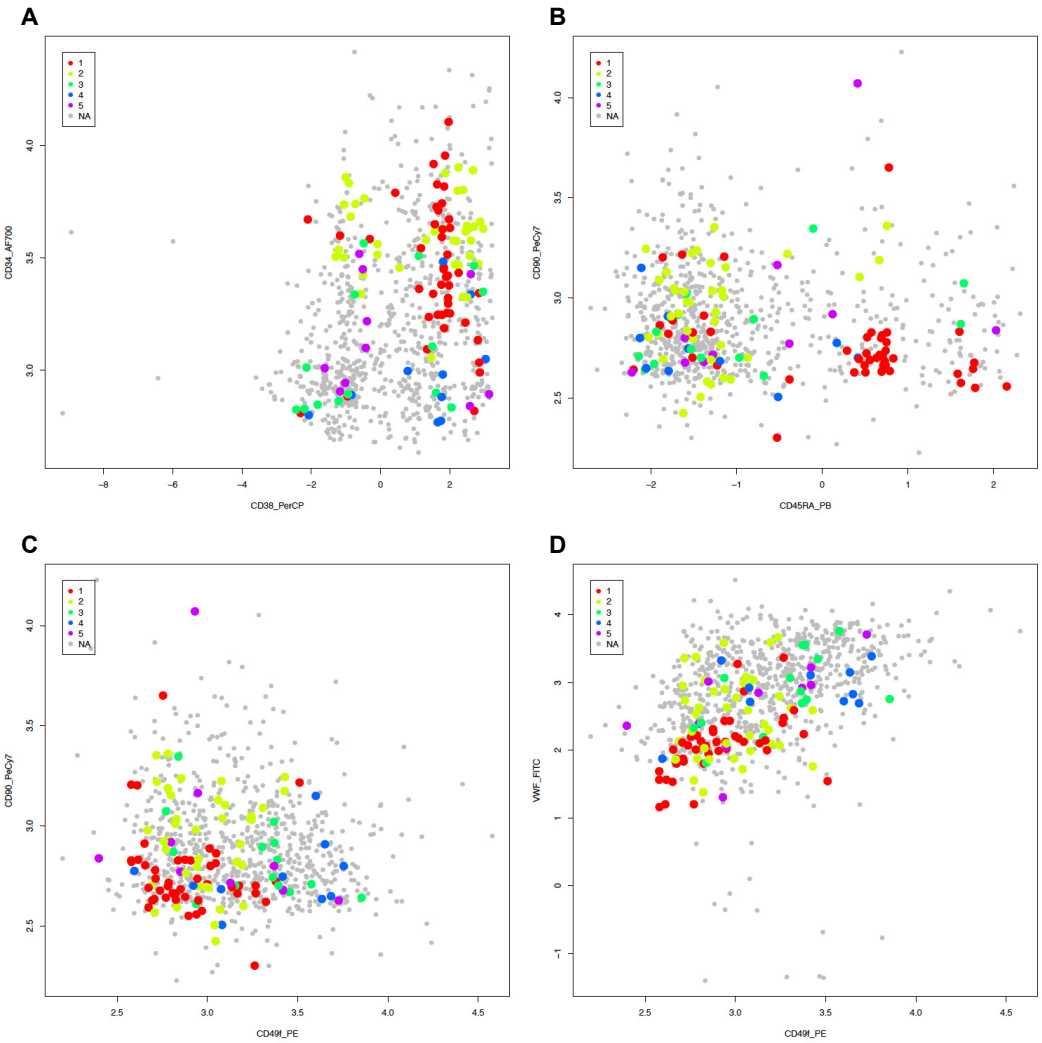


Figure 3.22 Flow cytometry plots for HSCs sorted for RNA-seq.
Retrospective analysis of flow FACS analysis for each individual HSC sorted showing: A. CD34 vs CD38; B. CD90 vs CD45RA; C. CD90 vs CD49f; D. VWF vs CD49f. Individual surface marker expression was normalised. Cluster 1: red, cluster 2: yellow, cluster 3: green, cluster 4: blue, cluster 5: magenta, cells filtered due to poor quality: grey.

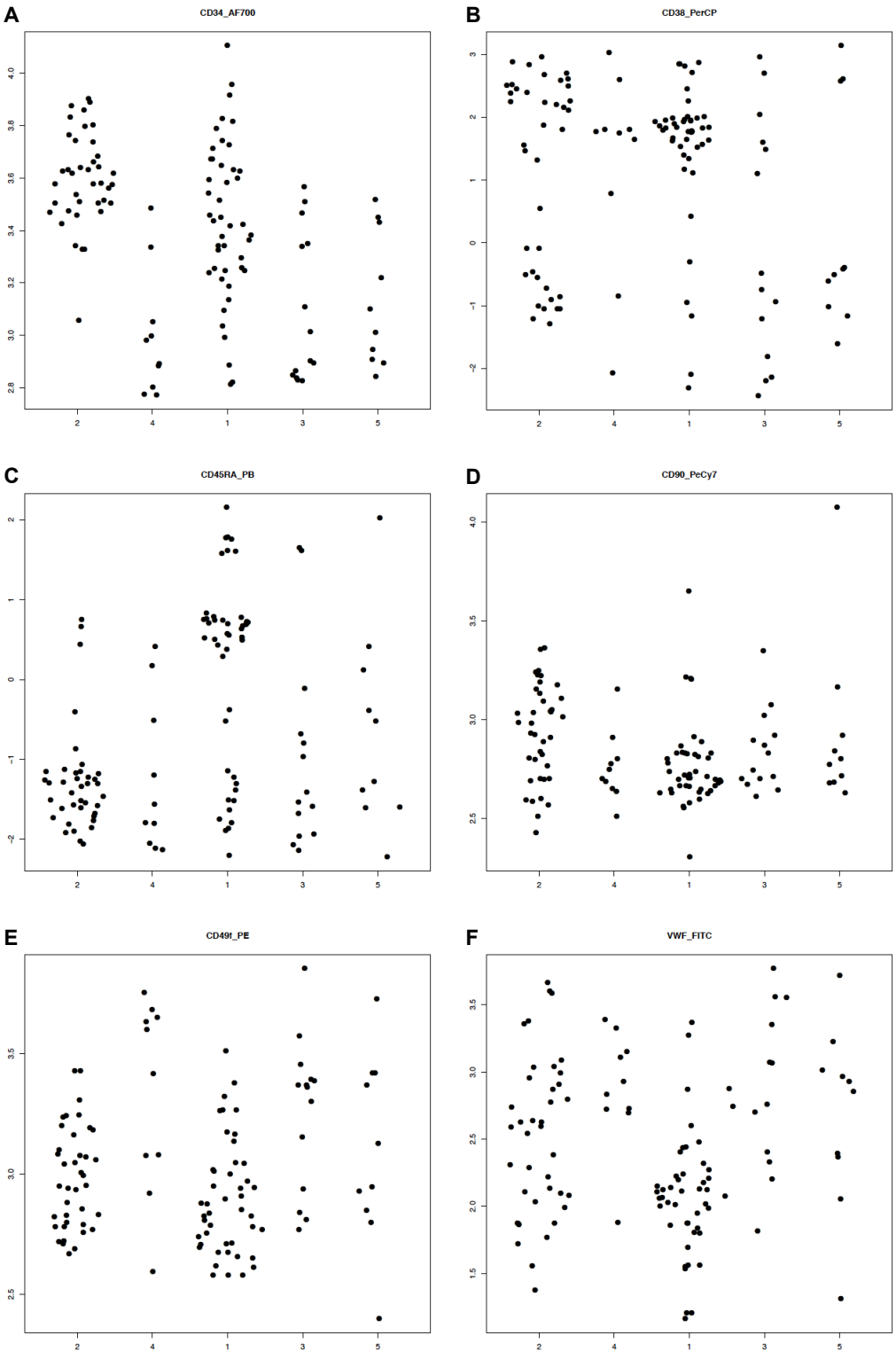


Figure 3.23 Flow cytometry index data by HSC cluster.
Difference in normalised surface marker expression (y axis) between HSC clusters 1, 2, 3, 4, 5. Mean and SD are shown. A. CD34; B. CD38; C. CD45RA; D. CD90; E. CD49f; F. VWF.

3.2.6 Functional validation of HSC reconstitution capacity

To verify the functional capacity of the sorted HSC population in terms of multilineage reconstitution, human bone marrow HSCs were transplanted into sub-lethally irradiated NSG mice by methods detailed in Section 2.5. Cells from cryopreserved human bone marrow CD34+ cells were FACS sorted using the gating strategy shown in Figure 3.1. They were also sorted according to VWF surface expression positive or negative. 400 VWF+ HSCs per mouse were transplanted by right femoral injection into 4 mice; 400 VWF- HSCs per mouse were transplanted into 4 mice, 68,000 VWF- HSCs per mouse were transplanted into 2 mice, 1 mouse was kept as a negative control. After 16 weeks, the mice were sacrificed and their peripheral blood and bone marrow were harvested and stained with antibodies to detect engraftment of human lymphoid, myeloid and megakaryocytic cells by flow cytometry.

Table 3.5 shows the level of engraftment in each animal, with the right femur and rest of bone marrow analysed separately. There was no engraftment seen in the peripheral blood and no megakaryocytes were identified in the bone marrow for any of the mice by flow cytometry. The 2 mice with 68,000 VWF- HSCs transplanted both engrafted showing both lymphoid and myeloid engraftment. From the 8 mice with only 400 cells transplanted, 3 mice engrafted, 2 mice that were transplanted with VWF+ HSCs and 1 mouse transplanted with VWF- HSCs. All 3 of the mice demonstrated both myeloid and lymphoid reconstitution.

Animal	Transplanted HSCs	RIGHT FEMUR			REST OF BONE MARROW			Engraftment
		CD45+	CD33+	CD19+	CD45+	CD33+	CD19+	
1	VWF+ 400 cells	3	2	1	0	0	0	0
2	VWF+ 400 cells	0	0	0	1	0	1	0
3	VWF+ 400 cells	114	6	98	1	0	0	1
4	VWF+ 400 cells	3	0	3	12	0	8	1
5	VWF- 400 cells	0	0	0	0	0	0	0
6	VWF- 400 cells	11	0	9	0	0	0	1
7	VWF- 400 cells	1	0	0	0	0	0	0
8	VWF- 400 cells	0	0	0	0	0	0	0
9	VWF- 68K cells	60	5	45	3	0	2	1
10	VWF- 68K cells	89	3	76	1186	28	1050	1
11	Control	0	0	0	4	0	3	0

Table 3.5 Blood cell reconstitution 16 weeks post xenotransplantation with VWF+/- human HSCs in sub lethally irradiated NSG mice.
Cell counts are given. Engraftment is defined as cell count >5 (shaded).

3.3 Discussion

The aim of this work was to uncover transcriptional variation and identify subpopulations within the human bone marrow HSC compartment in order to gain insights into the heterogeneity displayed by the HSC as well as insights into early megakaryopoiesis. The first studies using single cell RNA-seq to demonstrate HSCs as a transcriptionally heterogeneous population have recently been published in mouse (Wilson et al. 2015; Grover et al. 2016) and zebrafish (Macaulay et al. 2016). During the course of this project, human bone marrow HSCs were simultaneously sequenced at the single cell level by another group (Velten et al. 2017). However, the work presented here represents the first single cell RNA-seq analysis in human bone marrow HSCs to define transcriptionally distinct cell populations within the HSC compartment itself.

In this study, *ex-vivo* primary human bone marrow cells collected as sternal samples from five individuals undergoing aortic/mitral valve replacement. Fresh cells were stained with a panel of 16 surface markers and HSCs (defined as $Lin^{-}CD34^{+}CD38^{-}CD45RA^{-}CD90^{+}CD49f^{+}$) were index sorted for transcriptome analysis. To ensure that the cells that were sorted and studied were in fact HSCs with both multilineage engraftment potential, cells sorted within these gates were transplanted into sub lethally irradiated NSG mice and indeed demonstrated both myeloid and lymphoid engraftment at 16 weeks.

Whole transcriptome amplification was performed on the index sorted HSCs using a modified Smart-seq2 protocol and a total of 884 cells were sequenced. Previous transcriptome studies on human HSCs have been performed on CD34+ fraction of peripheral and cord blood or on preserved/frozen bone marrow samples (Notta et al. 2016). Here fresh human bone marrow was used in an attempt to provide the most physiological transcriptional representation of HSCs that reside in the bone marrow stem cell niche. The challenges of working with *ex-vivo* primary human cells necessitated rigorous quality filtering of the RNA-seq data using a machine learning cell filtering method (Section 2.6). Finally, 119 cells were used for gene expression analysis.

Unsupervised clustering revealed 5 transcriptionally distinct cell population “clusters” within the HSC compartment which separated out according to Pearson’s correlation. When the cells were clustered using a different method by which each single cell was placed according to differing gene expression creating differentiation trajectories, the cells clustered into the similar groupings adding confidence to the robustness of the original clusters. Cluster 3, as the most heterogeneous group lay at the root/branch point of the trajectories with the more homogenous clusters 1, 2 and 5 furthest apart on separate trajectories. Cluster 4 was seen to branch from the same trajectory as cluster 1 indicating a closer relationship.

A number of genes of megakaryocyte lineage featured in the top 200 most highly expressed genes in the HSC population as a whole including *PPBP, PF4, SDPR, ITGA2B and VWF*. Transcription of genes that are known to be expressed at both transcript and protein level in published megakaryocyte and platelet datasets demonstrates that the genetic programs required to produce the megakaryocyte lineage and platelets are switched on early in the HSC in agreement with studies in the mouse (Wilson et al. 2015; Grover et al. 2016; Sanjuan-Pla et al. 2013; Gekas and Graf 2013).

Identification of marker genes for each cluster revealed a number of megakaryocyte lineage genes as markers for cluster 1 and cluster 4. Comparison of cluster marker genes with recently published progenitor RNA-seq datasets from the Blueprint (L. Chen et al. 2014) and DMAP (Novershtern et al. 2011) databases added a further degree of confidence in these observations (Figure 3.18, 3.19). Again, this showed that cluster 1 and 4 marker genes were also expressed in megakaryocytes derived *in vitro* from blood CD34+ cells. While cluster 4 marker genes were specific to megakaryocytes, cluster 1 markers genes were expressed in megakaryocyte lineage population from the HSC population, through CMP, GMP, MPP to MEP populations (Figure 3.18). This observation was confirmed when placing the DMAP gene signatures onto the Monocle 2 differentiation trajectories (Figure 3.19). This showed that while the megakaryocyte signature was over-represented in cluster 1 and 4, the MEP signature was over-represented in only cluster 1 cells. The monocyte signature was over-represented in cluster 5 whilst cluster 2 expressed genes from the B cell signature. Cluster 3 had no significantly expressed marker genes which is in keeping with its position on the Monocle 2 differentiation trajectories.

A GO analysis of the cluster 1 marker genes indicated that the most significantly enriched GO categories for biological process were related to ribosomal function and protein localisation. While genes encoding for ribosomal or translational proteins made up only two of the top 20 marker genes, 77 out of the total 313 significant cluster 1 marker genes were ribosomal. This was a much greater proportion than any other cluster and it would indicate that the processes involved in translation, protein packaging are functionally important in this HSC cluster and its lineage committed progeny. After translation, the most significantly enriched GO biological processes were related to cellular metabolism, coagulation, haemostasis, translation and RNA splicing. 313 marker genes were identified for cluster 1 which was the largest HSC cluster comprising 46 single cells. However, only 19 marker genes were identified for cluster 4 which was made up of only 10 cells. The GO categories significantly enriched in cluster 4 pertained to the negative regulation of programmed cell death, cell cycle, protein folding and exocytosis. While it did not meet significance the GO category platelet degranulation also ranked highly. As suggested by the correlation with DMAP data, cluster 5 was significantly enriched for GO categories related to the immune response particularly with innate immunity

and GO analysis for cluster 2 suggested that the genes upregulated in this cluster were related to blood vessel development and endothelial cell function.

When the marker genes for each cluster are compared with published transcriptome datasets for megakaryocytes (cultured *in vitro* from CD34+cells) and other haematopoietic cells (L. Chen et al. 2014; Novershtern et al. 2011; Watkins et al. 2009) and platelets (Rowley et al. 2011; Goodall et al. 2010), cluster 1 and 4 marker genes show high levels of expression in megakaryocytes and platelets. In cluster 1, 7 out of 20 marker genes are highly expressed in both megakaryocytes and platelets with a further 11 out of 20 highly expressed in megakaryocytes with low to moderate expression in platelets. All 19 of cluster 4 marker genes show at least moderate expression in megakaryocytes and high expression in platelets. The evidence presented suggests that while both HSC cluster 1 and cluster 4 may be biased to the megakaryocyte lineage, HSCs in cluster 4 are already primed for platelet production and haemostatic functionality. In contrast marker genes from clusters 2 and 5 have much lower expression levels within the megakaryocyte lineage. Out of 20 marker genes in cluster 5 only 7 are seen to be expressed in megakaryocytes but with little or no expression in platelets and only 1 out of 20 of cluster 2 marker genes show transcript expression in megakaryocytes.

Of the cluster 1 marker genes, three encode proteins that are directly involved in platelet function. *ANGPT1*, located to chromosome 8, encodes the glycoprotein growth factor angiopoietin 1. It is stored in platelet alpha granules and released on platelet activation with thrombin (Y. Q. Huang, Li, and Karpatkin 2000; J. J. Li et al. 2001), as such it is present at the protein level in platelets with high quantity (Genecards.org). It is essential for blood vessel development where it functions by binding to a tyrosine kinase receptor, however it also has high transcript expression in the HSC (Forsberg et al. 2010; Cabezas-Wallscheid et al. 2014) as well as in stromal cells within the stem cell niche (Méndez-Ferrer et al. 2010; Ding et al. 2012) and therefore may also have an important role in haematopoiesis. *IL1B*, located on chromosome 2, encodes for the pro-inflammatory cytokine interleukin1B and is also a cluster 1 marker gene that has been shown to be important in platelet function (Beaulieu et al. 2014). While it is known to be released from activated macrophages as part of the inflammatory response platelet activation has also been shown to induce the synthesis and release of IL-1β via shedding of membrane-bound vesicles or release of exosomes (Denis et al. 2005; Weyrich et al. 2009). It is also well known to promote megakaryopoiesis and platelet release both *in vivo* and *in vitro* (Kimura et al. 1990; Jiang et al. 1994) and recently been shown to play a regulatory role in HSC (Espín-Palazón et al. 2014). It is highly expressed at the transcript level in HSCs, megakaryocytes (Novershtern et al. 2011; L. Chen et al. 2014) and platelet (Rowley et al. 2011). Another cluster 1 marker gene, *TSPAN32*, located on chromosome 11, that encodes a transmembrane protein of the tetraspanin family has also been shown to have an essential role in platelet function. TSPAN32 is a positive regulator of

the GPIIb/IIIa signalling and thrombus formation with *TSPAN32* knockout mice demonstrating bleeding phenotypes (Goschnick et al. 2006; Orlowski et al. 2009). *TSPAN32* has been shown to be expressed in megakaryocytes (Novershtern et al. 2011; L. Chen et al. 2014) as well as platelets (Rowley et al. 2011). A fourth cluster 1 marker gene has also been implicated in the regulation of platelet activation although as a negative regulator. *GUCY1A3*, a gene located on chromosome 4, encodes for the soluble alpha-3 subunit of guanylate cyclase which is an important signalling pathway in VEGF-induced angiogenesis, however the same pathway has been shown to inhibit platelet activation (Moro et al. 1996). The transcript has been seen to have an increased expression in HSCs and is upregulated in megakaryocytes compared to other mature cells lineages (Watkins et al. 2009). It is expressed at high transcript (Rowley et al. 2011) and protein levels in platelets (Genecards.org).

A number of significant marker genes for cluster 1 have also been shown to play a role in megakaryopoiesis, in addition to *IL1B* described above. *HEMGN*, a gene located on chromosome 9, encodes a nuclear protein that has been shown to be expressed in immature erythroid and megakaryocyte precursors and down regulated with further differentiation (L. V. Yang et al. 2001; Krüger et al. 2002). From DMAP and Blueprint data it is expressed throughout the megakaryocyte precursors (Novershtern et al. 2011; L. Chen et al. 2014) as well as in platelets at both transcript (Rowley et al. 2011) and protein level (Genecards.org). Interestingly in recent GWAS data *HEMGN* has been associated with platelet traits PLT, PCT, PDW. NPM1, located on chromosome 3, which encodes the protein nucleophosmin 1 and can either be nuclear or cytoplasmic is expressed highly in HSCs and megakaryocytes (Novershtern et al. 2011; L. Chen et al. 2014) but at lower levels in the platelet (Rowley et al. 2011). It is known to be integral to a number of cellular processes including cell cycle, translation and histone assembly. Mutation in the *NPM1* gene (which is commonly seen in acute myeloid leukaemia), leads to expansion of the megakaryocyte compartment with impaired megakaryocyte function and resultant reduced platelet counts (Sportoletti et al. 2013). Another of the cluster 1 marker genes *PRKACB*, located on chromosome 1, encodes the catalytic B subunit of cAMP dependent protein kinase which mediates cAMP signalling. It is integral to a number of cellular processes including cell proliferation and differentiation and it has recently been shown that TAL1 activity in megakaryocytes is likely to be mediated through this pathway (Kuvardina et al. 2017). TAL1 is an important transcription factor in determining megakaryocyte cell fate (Tijssen et al. 2011). Accordingly, *PRKACB* is shown to be upregulated in megakaryocytes compared to other mature cells. Although lower transcript levels are expressed in the platelet (Rowley et al. 2011), high levels of the protein are seen (Genecards.org), indeed PRKACB has been positively correlated with platelet function (Goodall et al. 2010).

NRIP1 (chromosome 21) and *PARP1* (chromosome 1) encode two nuclear proteins that were significantly upregulated in cluster 1 and while they are also expressed in the megakaryocyte lineage, they have regulatory effects on the HSC itself. *NRIP1* encodes for a nuclear protein described as a transcriptional co-regulator that modulates transcription by activation of steroid receptors. Although it is ubiquitously expressed, it is thought to be an important factor required for the maintenance and function of normal quiescent HSCs as it has shown high HSC transcript expression (T. S. Huang et al. 2008) and with downregulation with further differentiation (Forsberg et al. 2010). NRIP1 protein is also expressed in platelets although with unclear function. *PARP1* encodes a nuclear binding protein that mediates post-translational protein modification. It also has an important role in DNA repair which is demonstrated by its upregulation in cancer cells and the emerging therapeutic role for PARP1 inhibition (Rouleau et al. 2010). In a similar way it is thought to protect HSC DNA from self-renewal defects (X. Li et al. 2012; Du et al. 2016). Its particular significance here is that it may also have a role in megakaryopoiesis. ETS-1 is a transcription factor known to switch on megakaryocyte differentiation and as PARP1 DNA repair can be mediated through ETS-1 (Legrand et al. 2013) it could be important in fulfilling cell fate decisions. In accordance with this *PARP1* has been shown to be expressed in HSCs, other progenitors and at higher levels in megakaryocytes (Novershtern et al. 2011; L. Chen et al. 2014).

Several other genes involved in cellular metabolism were upregulated in cluster 1 compared to the other HSC subsets include *LDHB* (chromosome 12), *SLC37A1* (chromosome 21) and *ATP6V0A2* (chromosome 12) and have previously been reported to be expressed in the megakaryocyte lineage (Novershtern et al. 2011; L. Chen et al. 2014; Rowley et al. 2011). While *LDHB* is expressed in all cells it is highly enriched in HSCs suggesting that anaerobic metabolism plays an important role in HSCs (Cabezas-Wallscheid et al. 2014). *SLC37A1* encodes an endoplasmic reticulum transporter protein and is specifically upregulated in megakaryocytes (Watkins et al. 2009) with protein expression in platelets (Genecards.org). *ATP6V0A2* encodes an enzyme transporter ubiquitous to all endomembrane organelles. *LAPTM4B*, located on chromosome 8 also encodes an endomembrane protein, it has unknown function but is expressed highly in HSCs and megakaryocytes. Another transmembrane protein specific to the golgi membrane is encoded by another significant cluster 1 marker gene, *ST6GAL2* located on chromosome 2. Interestingly this gene has not previously been reported to be expressed in haematopoietic cells.

Other nuclear genes significantly upregulated in cluster 1 compared with the other HSC subsets include: *ABRACL* (or *C6orf115*), a gene located on chromosome 6, which encodes a costar family protein shown to be one of the top 50 platelet proteins (Gevaert et al. 2003) although its function is unknown. *ANP32B* (chromosome 9) which encodes a protein highly expressed in HSCs, megakaryocytes and platelets that is anti-apoptotic and involved in transcriptional regulation. *EIF1AX*, located on the X chromosome, is expressed highly in

HSCs and megakaryocytes and involved in protein translation, as are *RPS12* (chromosome 6) and *RPL21* (chromosome 13) both encoding ribosomal subunits highly expressed through from HSC to megakaryocyte and in platelets. The final cluster 1 marker gene, *FAM30A* (chromosome 14) is a non-coding RNA gene.

In summary, while three of cluster 1 marker genes encode proteins that play a direct role in platelet activation, a number of genes are involved in megakaryopoiesis from the HSC. Although all but 2 (*ST6GAL2, FAM30A*) of the cluster genes have previously reported high levels of transcript expression in megakaryocytes, only *ANGPT1, GUCY1A3* and *SLC37A1* are seen to be upregulated in megakaryocytes compared with other differentiated haematopoietic cells, and none are solely expressed in megakaryocytes compared with the other progenitors. This is in keeping with the known expression of cluster 1 marker genes in all intermediate progenitors from HSC to megakaryocyte in the Blueprint data (Figure 3.18) and the overlap between expression of cluster 1 marker genes with both megakaryocyte and MEP from the DMAP dataset (Figure 3.19).

Conversely as discussed below, 7 out of 19 of cluster 4 marker genes are directly involved in platelet activation and platelet function and 12 of the genes are upregulated or expressed specifically in the megakaryocyte lineage. Furthermore all 19 genes have previously reported high transcript levels in platelets (Rowley et al. 2011).

Two of the cluster 4 markers genes that are known to play an important role in platelet function are tubulin B1 (*TUBB1*, located on chromosome 20) and tubulin A4 (*TUBA4A*, located on chromosome 2) which are highly specific to the megakaryocyte/platelet lineage. They are expressed at low levels in HSCs (Novershtern et al. 2011; L. Chen et al. 2014) but highly expressed in megakaryocytes as well as having high platelet transcript and protein expression (Rowley et al. 2011; Genecards.org). These genes encode proteins integral to organisation of the cytoskeleton and both are thereby directly involved in proplatelet formation and platelet conformational change. Transcripts of important proteins released by the platelet during activation were also upregulated in the cluster 4 markers genes. *F13A1*, a gene located on chromosome 6, encodes for platelet coagulation factor XIII alpha chain. It is known to reside in the platelet alpha granule and is released at the time of platelet degranulation. While previous datasets have demonstrated low expression in HSCs, *F13A1* is specifically expressed at high levels in megakaryocytes (Novershtern et al. 2011; L. Chen et al. 2014) and platelets at both the transcript and protein levels (Rowley et al. 2011; Genecards.org). The gene for the chemokine CCL5 was also a cluster 4 marker gene and is one of the chemokine genes found on chromosome 17. *CCL5* encodes a chemokine known to be highly expressed in the platelet and is released upon platelet activation. Recently it has been suggested to have regulatory feedback activity on the megakaryocyte through its receptor CCR5 (Machlus et al. 2016). *CCL5* is known to be specifically expressed in megakaryocytes

with low levels of expression in the HSC (Novershtern et al. 2011; L. Chen et al. 2014). Cluster 4 marker genes also included three genes involved in signal transduction during platelet activation. The most significantly upregulated gene in cluster 4 was *GRAP2* (chromosome 22), that encodes an adaptor involved in protein tyrosine kinase signalling on activation of the GPVI collagen receptor, that results in a cascade of integrin activation, platelet shape change and platelet degranulation. In previous data sets *GRAP2* is expressed at relatively low levels in HSCs (Novershtern et al. 2011; L. Chen et al. 2014). The levels of transcription are much higher levels in megakaryocytes, with expression increasing with megakaryocyte differentiation (Kim et al. 2002; Shim et al. 2004). Both transcript and protein expression are high in platelets (Rowley et al. 2011; Genecards). Another transcript involved in signalling within the cell, found to be a significant marker gene of cluster 4, is *ZNF185* (located on the X chromosome). It encodes a LIM and zinc finger domain protein that mediates protein-protein interactions and thereby regulates a number of cellular processes, including cytoskeletal change. Current data suggests that while the transcript is only expressed at low levels in the HSC, there are high transcript levels in the megakaryocyte (Novershtern et al. 2011; L. Chen et al. 2014) and platelet (Rowley et al. 2011) and is thereby specific to the megakaryocyte lineage. *GNAZ* (chromosome 22), is also upregulated in cluster 4 but with expression at lower levels. It encodes the signal transduction G protein subunit alpha Z which has been shown to be present in human megakaryocytes and platelets, and is phosphorylated during platelet activation (Gagnon et al. 1991). It is present at transcript level in HSCs, other progenitors as well as megakaryocytes and platelets (Novershtern et al. 2011; L. Chen et al. 2014; Rowley et al. 2011).

Three out of 19 cluster 4 marker genes encode for transmembrane proteins that are specific to the megakaryocyte/platelet lineage. *PDZK1IP1* (also known as *MAP17*) was the second most highly significant cluster 4 marker gene. It is located on chromosome 1 in close proximity to *TAL1* (regulator of megakaryocyte/erythroid differentiation) and encodes a small plasma membrane associated protein with a PDZ-binding domain and two transmembrane domains. It has been shown to be specifically expressed at high levels in platelets both at the transcript and protein level (Rowley et al. 2011), although levels of transcript expression in HSCs and megakaryocytes are far lower (Novershtern et al. 2011; L. Chen et al. 2014). It has however been shown to be expressed at higher levels in the murine HSC (Cabezas-Wallscheid et al. 2014; Forsberg et al. 2010; Wilson et al. 2015) with a major contribution of PDZK1IP1 expressing cells to all lineages after transplantation especially platelets (Sawai et al. 2016). In humans, however, its physiological role is unclear. *TMEM40* was also found to be a highly significant marker gene for cluster 4. It is found on chromosome 3 and encodes a transmembrane protein that has been localised to the exosome within the cell. *TMEM40* demonstrates moderate levels of transcript expression in the HSC (Novershtern et al. 2011; L. Chen et al. 2014), it has been shown to be upregulated in megakaryocytes compared with erythroblasts (Macaulay et al. 2007) and shows high levels of expression in the platelet

(Rowley et al. 2011) as well as high levels of platelet protein expression (Genecards.org). *PTCRA*, another significant marker gene for cluster 4, is pre-T cell receptor alpha on chromosome 6. It encodes a plasma membrane protein which along with TCR and CD3 form the immature T cell receptor complex. What its function is on the surface of the HSC, megakaryocyte or platelet is unknown. It is seen to be expressed at low levels in HSCs and megakaryocytes (Novershtern et al. 2011; L. Chen et al. 2014). However, it is upregulated at the transcript as well as protein level in platelets as well as the lymphoid lineage (Genecards.org).

Two of the cluster 4 marker genes encoded the negative regulators of proliferation: cyclin dependent kinase inhibitors *CDKN1A* and *CDKN2D* (chromosome 6, 19 respectively). They have both been shown to be expressed in HSCs and megakaryocytes at the transcript level (Novershtern et al. 2011; L. Chen et al. 2014) as well as platelets as transcript and protein (Rowley et al. 2011; Genecards.org). A further two of the cluster 4 marker genes encode proteins related to transcriptional regulation. *HIST1H2AC* (chromosome 6) was found to a be highly significant cluster 4 marker gene. It encodes a chromosome organisational protein, a core component of the nucleosome and through its own post-translational modification it regulates transcription. While this gene is expressed across all progenitors it has a higher expression in megakaryocytes (Novershtern et al. 2011; L. Chen et al. 2014) and platelets (Rowley et al. 2011), also being part of the platelet proteome (Genecards.org). Another cluster 4 marker gene, *TSC22D3* (X chromosome) encodes for a DNA binding protein with transcription factor activity and is a known to be a down regulator of pro-apoptotic factors. It is upregulated in megakaryocytes compared with HSCs (Novershtern et al. 2011; L. Chen et al. 2014) and remains highly expressed in platelets (Rowley et al. 2011; Genecards.org).

Of the remaining cluster 4 marker genes, three encode organelle membrane proteins, *ACRBP, SNN, MMD* (chromosomes 12, 16 and 17 respectively) and although all are expressed highly in megakaryocytes although not-specifically (Novershtern et al. 2011; L. Chen et al. 2014) and platelets (Rowley et al. 2011) their function in this setting is unclear. Other cluster 4 marker genes include the pseudogene *CXCR2P1* (chromosome 2) and *C21orf7* (chromosome 21) both of uncertain significance.

As previously discussed, to date megakaryocyte transcriptome data which is available is from cultured megakaryocytes that have been derived *in vitro* from peripheral blood CD34+ cells, under the influence of TPO and IL1B. Therefore, they are not representative of *in vivo* megakaryocytes. In an attempt to analyse the transcriptome of higher ploidy cells one study has assessed differential transcript expression between *in vitro* derived 2-4N megakaryocytes and 8-16N megakaryocytes which are found in much smaller quantities (Raslova et al. 2007). Interestingly, of the cluster 1 marker genes only *ANGPT1* was seen to be upregulated with increasing ploidy, 17 of the cluster 1 marker genes were not differentially expressed and

expression of *NPM1* was downregulated. Of the cluster 4 marker genes, 10 genes were upregulated with increasing ploidy: *TUBB1, TUBA4A, F13A1, CCL5, GRAP2, PTCRA, CDKN2D, HIST1H2AC, ACRBP, SNN* with the remaining not being differentially expressed.

Taken together it is hypothesised that while HSC subsets 1 and 4 are megakaryocyte biased, HSCs in cluster 1 are primed to produce HSCs through the classical differentiation pathway and, whereas cluster 4 HSCs already express many of the necessary transcripts to produce megakaryocytes directly bypassing intermediate progenitors, and even have many of the genes required for platelet function. In keeping with this the genes upregulated in cluster 1 compared to other HSC subsets have previously been shown to be expressed in megakaryocyte progenitors as well as megakaryocytes but often not in platelets, a number of them are related to HSC function and megakaryopoiesis, while fewer encode proteins involved in platelet function. Moreover, cluster 4 upregulated genes are more specifically expressed in megakaryocytes, with low expression in megakaryocyte progenitors and high expression at both transcript and protein level in platelets. Many of the upregulated cluster 4 genes encode proteins directly involved in platelet activation and function. Therefore cluster 4 could be analogous to the lineage-restricted HSC subset primed to produce emergency or "stress" megakaryocytes as previously described (Haas et al. 2015).

In an attempt to investigate whether it was possible to prospectively identify clusters 1 and 4 the index FACS sorting data was analysed. While cluster 1 and 4 differed in terms of their surface expression of CD34 and CD49F. While cluster 1 had high CD34, and lower CD49F expression, cluster 4 had expressed lower CD34 and higher CD49F on the surface. It would also be possible to differentiate some cluster 1 cells from clusters 2, 3 and 5 on the basis of CD45RA expression. However, prospectively differentiation cluster 4 from clusters 3 and 5 on this basis would be difficult. Further work is required to assess whether differentiating clusters 1 and 4 from the whole HSC population is indeed possible with the existing surface markers or if work is required on adding new markers. From the work on cluster 1 and 4 marker genes It may also be possible to target cluster 1 and 4 marker genes that are known to be encode proteins expressed on the cell surface such as TSPAN32 for cluster 1 and PTCRA and PDZK1IP1 for cluster 4. If these could be prospectively differentiated, further *in vitro* differentiation, transplantation or RNA sequencing or larger cell numbers could be attempted to gain further understanding of the role of these 2 HSC subpopulations.

The main limitations of the methods used are the limitations of any single cell RNA-seq analysis. There remains no standardised analysis for normalising for technical dropouts or "zeros", although in this study the Lun method (Lun, McCarthy, and Marioni 2016) was chosen for normalisation. Furthermore, as with any PCR amplification process there is a heavy bias towards amplification of more abundant transcripts which may result in overrepresentation of these transcripts in the resulting data. This is of course may be more

apparent in single cell analysis as it leads to a disparity between overrepresented abundant transcripts and technical dropouts which may be false negatives. Nevertheless, the clustering methods used here are robust and previously validated in published work (Kiselev et al. 2017; Trapnell et al. 2014; Qiu et al. 2017). Another limitation was number of cells with only 119 deemed to be of high quality and taken forward for gene expression analysis from a total of 884 cells. The proportion of high quality cells amongst the number of cells sorted may be related to the time taken from cell harvest to sorting into lysis buffer which was at least 3 hours. Further work would include increasing the number of cells to further validate the transcriptional differences found in this study.

The gene signatures presented in this chapter relating to the two megakaryocytes biased clusters have been compared to previously published cultured megakaryocyte gene expression. In the following chapters the transcriptome of *ex-vivo* human bone marrow megakaryocytes is analysed at different levels of polyploidisation and in the setting of coronary thrombosis which is hypothesised to produce stress megakaryopoiesis. HSC cluster 1 and 4 signatures are then correlated with the bone marrow megakaryocyte transcriptome to gain further understanding of their functional significance within the HSC population.

CHAPTER 4
The Megakaryocyte Transcriptome

4.1 Preface

The molecular mechanisms controlling megakaryopoiesis, including megakaryocyte maturation, remain largely unknown due to a number of technical challenges that have precluded the study of cellular and molecular events. These include obtaining human bone marrow samples from patients who do not clinically require bone marrow aspiration, the rarity of the cell population, estimated to make up less than 0.05% of the nucleated bone marrow cell compartment and the size, and fragility of the megakaryocyte. This has so far prevented transcriptome analysis of primary human bone marrow megakaryocytes. However, with the ability to interrogate single cell transcriptomes it has now become possible. Our current megakaryocyte knowledge is largely based on gene expression array data of megakaryocytes cultured from haematopoietic progenitors obtained from foetal liver, cord blood and adult blood (Bluteau et al. 2013; Watkins et al. 2009; Shim et al. 2004; Raslova et al. 2007; L. Chen et al. 2014), as well as knowledge of the platelet transcriptome (Gnatenko et al. 2003; Bugert et al. 2003; McRedmond et al. 2004; Hillmann et al. 2006; Rowley et al. 2011). When comparing megakaryocytes cultured from adult blood with their precursor CD34+ cells one third of the upregulated genes have been shown to be directly involved in haemostasis or platelet function (Shim et al. 2004). Many of the transcripts that were upregulated are known to play key roles in thrombus formation such as glycoprotein V and IIIa, von Willebrand factor, integrins B3 and a6, P-selectin, coagulation factor XIII (Shim et al. 2004).

Gene expression profiles of cultured megakaryocytes of differing ploidies have also been analysed, with comparison made between 2-4N and 8-16N megakaryocytes to gain further understanding of megakaryocyte maturation (Raslova et al. 2007). This has shown further upregulation of haemostasis and platelet related genes. Concomitantly, with increasing ploidy, there was a downregulation of genes participating in cell cycle functions. The ploidy profile of megakaryocytes cultured from human CD34+ cells is known to be dramatically different to that found physiologically within the bone marrow, with mean ploidy levels of 2N and 16N respectively. It is therefore likely that there are putative drivers of *in-vivo* megakaryocyte differentiation and maturation that have yet to be identified. Consequently, the small numbers of 8N and 16N megakaryocytes cultured *in vitro* from CD34+ cells may functionally differ from those found in bone marrow.

Substantial work has gone into generating platelets *in vitro* on a large scale, for transfusion purposes (Moreau et al. 2016). Despite the ability to produce large numbers of megakaryocytes from iPSCs, the platelet output is relatively small and remains inadequate to

meet the needs of platelet transfusion. This may be a direct result of the differences between megakaryocytes produced in culture and megakaryocytes in the bone marrow niche in terms of ploidy level, cytoplasmic maturation or extracellular signals. There is currently no evidence that explains a direct function for mean megakaryocyte ploidy levels of 16N, whether ploidy level is a marker of maturity, or if it is associated with the capability to produce platelets. Studies of the transcriptome of bone marrow megakaryocytes may provide important insights, including the identification of novel drivers of megakaryocyte maturation that could enable more physiological levels of *in vitro* platelet production.

In this chapter single cell and low input RNA sequencing techniques are applied and optimised to investigate the transcriptome of primary human bone marrow megakaryocytes. Bone marrow megakaryocyte transcriptomes are compared to cultured megakaryocytes and gene signatures related to ploidy are investigated.

4.2 Results

4.2.1 Protocol optimisation for analysis of megakaryocyte transcriptome

Out of a total of 53 patients recruited to this study, bone marrow material for 7 patients was used solely for protocol optimisation. This was due to initial technical difficulties in isolating megakaryocytes and optimal preparation of single cell or low-input pooled RNA-seq libraries. Of the remaining 46 patients, megakaryocytes were FACS sorted from bone marrow material from 34 patients for transcriptome analysis. Study design and samples used are shown in Appendix 1 and 2 respectively.

4.2.1.1 Optimisation of megakaryocyte isolation

A number of methods to obtain pure megakaryocyte populations from whole bone marrow scrapings for the purposes of RNA-seq were trialled. Figure 4.1a shows a bone marrow smear with megakaryocytes located in particles within whole bone marrow. In an attempt to optimise isolation of megakaryocytes a Percoll gradient was trialled (Section 2.3.3.2). Percoll enrichment has been reported to increase megakaryocyte frequency in a cell suspension from 0.05% to 1.5% (30 fold increase) and has only been carried out in a small number of studies, with different methodologies (Tomer, Harker, Burstein 1988; Brown et al. 1997; Ishibashi et al. 1986). The method used here is described in Chapter 2. Figure 4.1b shows a cytospin slide from this enriched population; direct observation demonstrated a much higher density of megakaryocytes than in whole bone marrow but a large number of contaminating cells remained. Megakaryocyte identity was verified with by immunofluorescence staining using a DAPI (4',6-diamidino-2-phenylindole) dye to stain DNA and VWF surface staining to

identify megakaryocytes (Figure 4.1c).

Two methods for collecting purified megakaryocytes were then tested, as outlined in Section 2.3.3. These were: 1. Cell picking, where individual megakaryocytes were picked under a 40x objective inverted light microscope, using a glass pipette which was trialled on both whole bone marrow and from Percoll enriched populations (Section 2.3.3.2). However, this was limited by time to cell lysis, even when only small cell numbers were picked and reliance only on relative cell size in distinguishing megakaryocytes from other cell types. 2. FACS cell sorting, which became possible with use of a large bore FACS nozzle (100uM). Ultimately, in order to decrease time from bone marrow collection to cell lysis and therefore reduce the chance of cell/RNA degradation, whole bone marrow was FACS sorted without Percoll density separation. Therefore, megakaryocytes were isolated according to ploidy level from whole bone marrow scrapings using FACS with an initial treatment with ammonium chloride red cell lysis solution and then staining for the markers of megakaryocyte maturity CD41 and CD42 as well as Hoechst 33342 DNA stain to allow for quantification of DNA content without the requirement for cell permeabilisation (as described in Section 2.3.3). Figure 4.1d shows the sorting strategy used.

As described above, the aim of this project had been to perform single cell megakaryocyte RNA-seq. Therefore, much of the FACS sorted megakaryocyte material was collected as single cell according to ploidy (27/34 patients), 6/34 patients had megakaryocyte pools FACS sorted according to ploidy and 16/34 patients had megakaryocyte pools FACS sorted irrespective of ploidy (Appendix 1).

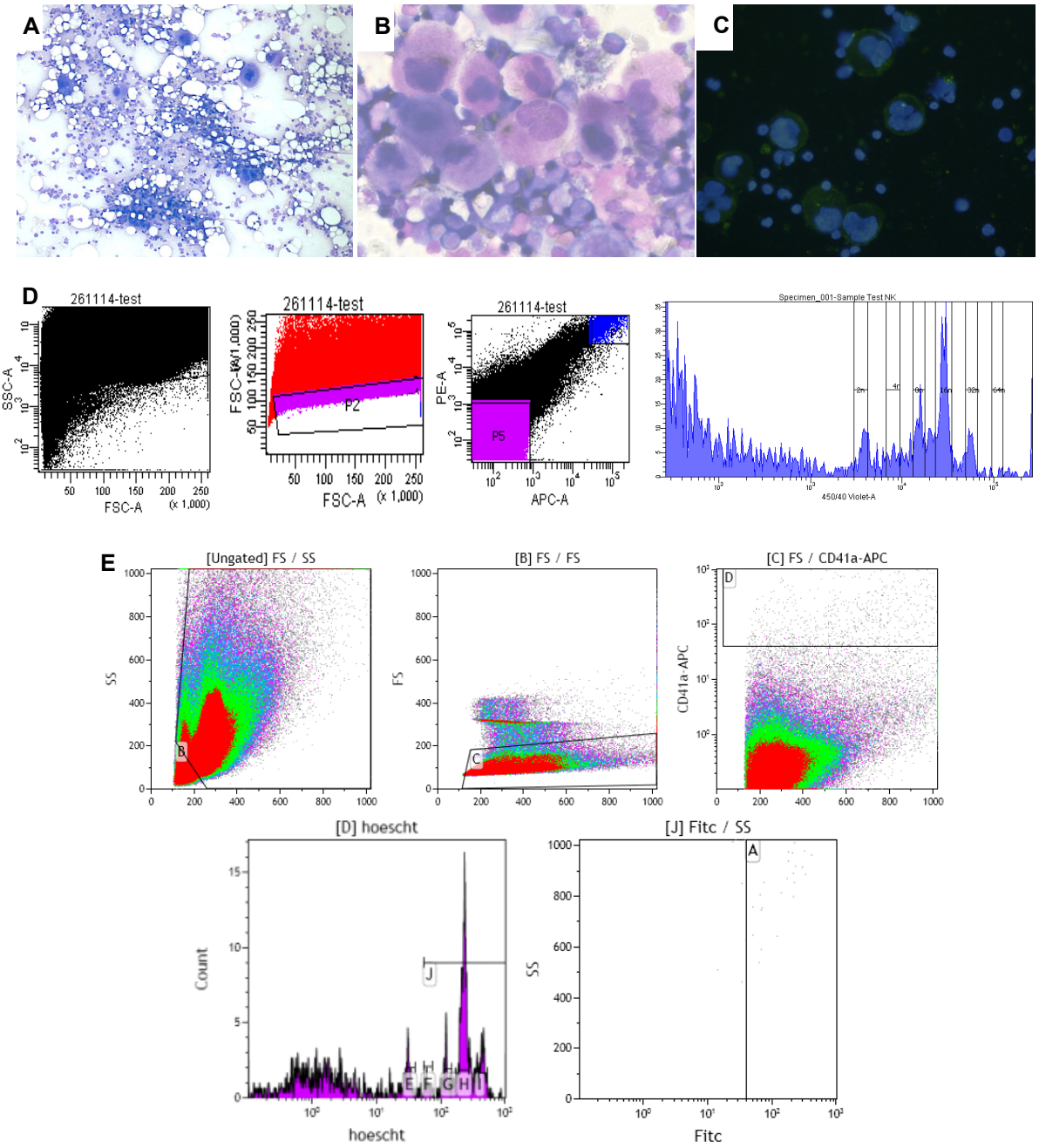


Figure 4.1 Bone marrow megakaryocyte phenotype

A. Human bone marrow smear stained with Rapid Romanowsky, 10x objective (Control); B. Cytocentrifugation of enriched megakaryocyte population stained with Roberts stain, 40x objective (Control); C. Immunofluorescent staining of enriched megakaryocyte population with DAPI DNA stain (blue) and VWF surface stain (green) (Control); D. Fluorescence activated cell sorting strategy for primary megakaryocytes from clinical samples of whole human bone marrow. Cells were sorted into lysis buffer within 6-8 hours of bone marrow harvest. Whole bone marrow was stained with APC conjugated anti-CD41a, PE conjugated anti-CD42a and Hoechst staining. Ploidy data was collected from all samples (Control); E. Flow cytometry data of mouse megakaryocytes at 10 hours after collection, kept at room temperature, stained with APC conjugated anti-CD41a, Hoechst staining and FITC conjugated Calcein. 98% megakaryocytes are seen to take up this live cell stain.

4.2.1.2 Megakaryocytes take up live/dead stains

In order to ensure that collected cells were viable a number of live/dead stains were tested. However, it was apparent that even in the absence of any processing all megakaryocytes were positive for basic live/dead stains. This included trypan blue stain which was added to clinical bone marrow material as early after sampling as 10 minutes and it was confirmed to hold true for *ex-vivo* mouse megakaryocytes. This was seen to be the case for all stains based on dye uptake by dead cells due to loss of plasma membrane integrity, such as 7AAD and propidium iodide. No changes in the experimental procedure resulted in an amelioration of this phenomenon. It was postulated that these findings could be explained by the extensive demarcation membrane system that is a system of invaginated plasma membrane that fills the megakaryocyte cytoplasm and therefore it may in fact be an artifactual result. To test this theory, mouse bone marrow cells were stained with fluorescent calcein dye under varying conditions and time points followed by flow cytometric analysis. Calcein relies on intracellular esterase activity and active uptake in living cells whereby live cells take up the dye and dead cells exclude it (King 2000; Fritzsche and Mandenius 2010). Calcein was shown to be taken up by 98% megakaryocytes even up to 10 hours after sampling demonstrating that the "uptake" of loss of plasma membrane integrity stains were indeed artifactual (Figure 4.1e).

4.2.1.3 Use of G&T-seq protocol for RNA-seq library preparation

The ability to generate full length cDNA using the Smart-seq2 protocol (Picelli et al. 2014) was trialled on a range of cell types at single cell level to ensure it was robust before beginning to use it on clinical bone marrow samples recruited for the study. Here cells were flow sorted into the Smart-seq2 lysis buffer, a mild hypotonic lysis buffer containing a ribonuclease inhibitor that blocks RNA degradation and stabilizes the RNA. Figure 4.2a-c shows BioAnalyser traces of cDNA generated from single cells of the K562 cell line, megakaryocytes cultured from adult blood CD34+ cells and primary peripheral blood mononuclear cells. In contrast, much shorter fragments of cDNA were generated from megakaryocytes and other nucleated cells from clinical bone marrow (Figure 4.2d, e, f).

The generation of fragmented cDNA has been suggested to indicate RNA degradation prior to collection of cells within lysis buffer related to cell condition (Picelli et al. 2014). Therefore, a number of different variables were trialled in an attempt to optimise the condition of the cells at the point of flow sorting into lysis buffer. These conditions included: 1. Bone marrow scraping or aspirate as starting material; 2. Transportation at room temperature or on ice; 3. Different strengths of human serum albumin in transportation buffer (2%, 5%, 10%); 4. Collection into RNAlater solution (Florell et al. 2001); 5. Processing of bone marrow harvest immediately or following day; 6. Manual cell picking either on same day of harvest, up to 5 days post-harvest or flow sorting; 7. Collection of cells from whole bone marrow, after red cell lysis or after Percoll separation. Cells collected under these conditions generated fragmented

cDNA with consistent BioAnalyser traces and therefore did not reveal the underlying cause for the discrepancy between primary bone marrow cells compared to primary blood cells. While this experience formed the basis of the fully optimised bone marrow collection, transportation and processing protocol that was taken forward (Chapter 2), it also suggested that RNA degradation may not play a major role in generating poor quality cDNA.

To investigate this possibility further, total RNA was extracted from >1x10^6 nucleated bone marrow cells using TRIzol reagent 5 days after bone marrow harvest. 1ng of purified RNA was then reverse transcribed and amplified using the Smart-seq2 protocol revealing high quality full length cDNA on BioAnalyser traces (Figure 4.3a). This suggested that modifications in RNA lysis buffer might improve the quality of cDNA generated. In order to assess the effects of different methods of preserving RNA the Norgen single cell RNA purification kit was trialled on peripheral blood mononuclear cells with single cells, 20 cells, 50 cells and 100 cells. This kit did not yield any appreciable cDNA, likely due to loss of material within magnetic columns and large buffer volumes used. Therefore, it was not tested in clinical bone marrow samples. Figure 4.3b-d shows BioAnalyser traces for cDNA from bone marrow megakaryocytes collected in Triton x-100 based lysis buffer, NP-40 based lysis buffer and Qiagen RLT Plus buffer. RLT Plus buffer is the lysis buffer used in G&T-seq (Macaulay et al. 2015), described in Section 2.4.4, allowing analysis of both RNA as well as genomic DNA. It contains the chaotropic agent guanidinium thiocyanate which lyses the cell and is able to denature RNase and DNase. Use of this lysis buffer consistently lead to full length cDNA amplification in 20x megakaryocytes (Figure 4.3d) and improved amplification in single cells, although in single cells with more variable results (Figure 4.4). This suggested a more complete or efficient lysis of bone marrow cells using Qiagen RLT Plus buffer compared to the SMART-seq2 Triton-based hypotonic lysis buffer.

The G&T-seq protocol was therefore employed for RNA-seq library preparation for all bone marrow cells, both single cell and multicellular pools (1-100cells). The aim was to perform mainly single cell RNA-seq, therefore much of the sample collection was of single cells according to ploidy. Single megakaryocytes were FACS sorted for transcriptome analysis in 27 out of 34 patients. Multicellular pools were sorted from 16/34 patients and pools formed according to ploidy were sorted from 6 patients.

Initial RNA-seq libraries were prepared for 1160 single megakaryocytes with a ploidy level ranging from 2N to 32N in 5 patients and 55 multicellular pools between 20-100 megakaryocytes from 7 control individuals. 125bp paired-end sequencing was performed at an average sequencing depth of 64 either single cells or multicellular pools per Hi-seq lane. In terms of batch effects Table 4.1 shows each sample plate ID and corresponding sequencing lane ID whereas Table 4.2 outlines the batches in which samples were collected, cDNA reverse transcribed/amplified and Nextera libraries prepared and sequenced.

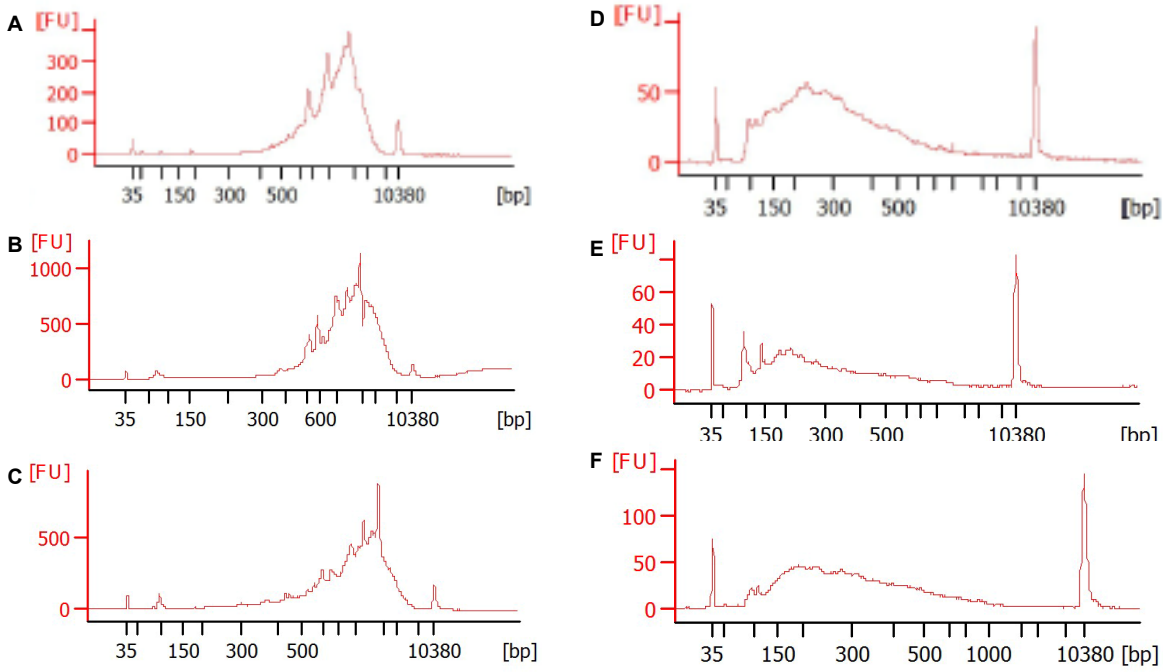


Figure 4.2 cDNA generated using Smart-seq2
BioAnalyser electropherograms of cDNA using Smart-seq2. cDNA generated from: A. single cell: K562 cell line; B. 20x megakaryocytes cultured from adult blood CD34+ cells; C. 20x primary human peripheral blood mononuclear cells; D. 20x bone marrow megakaryocytes; E. single cell: bone marrow megakaryocyte; F. 20x nucleated bone marrow mononuclear cells.

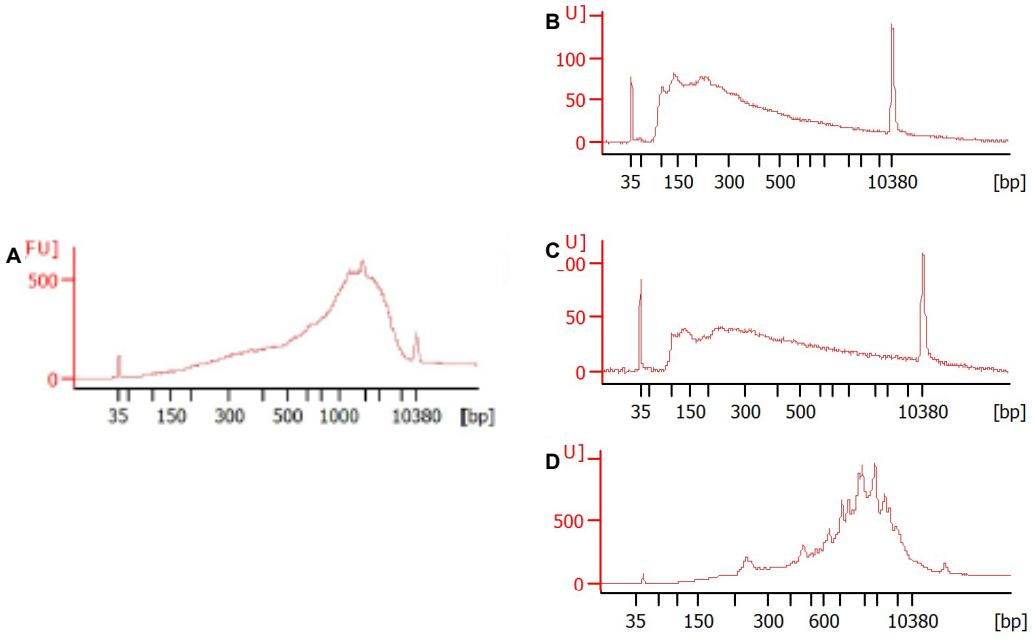


Figure 4.3 cDNA generated using Smart-seq2: cell lysis
BioAnalyser electropherograms of cDNA generated from: A. 1ng purified RNA extracted with TRIzol from >1 million nucleated bone marrow mononuclear cells; B. 20x bone marrow megakaryocytes lysed in Triton-based lysis buffer; C. 20x bone marrow megakaryocytes lysed in NP-40 based lysis buffer; D. 20x bone marrow megakaryocytes lysed in Qiagen Buffer RLT Plus, amplified with G&T-seq.

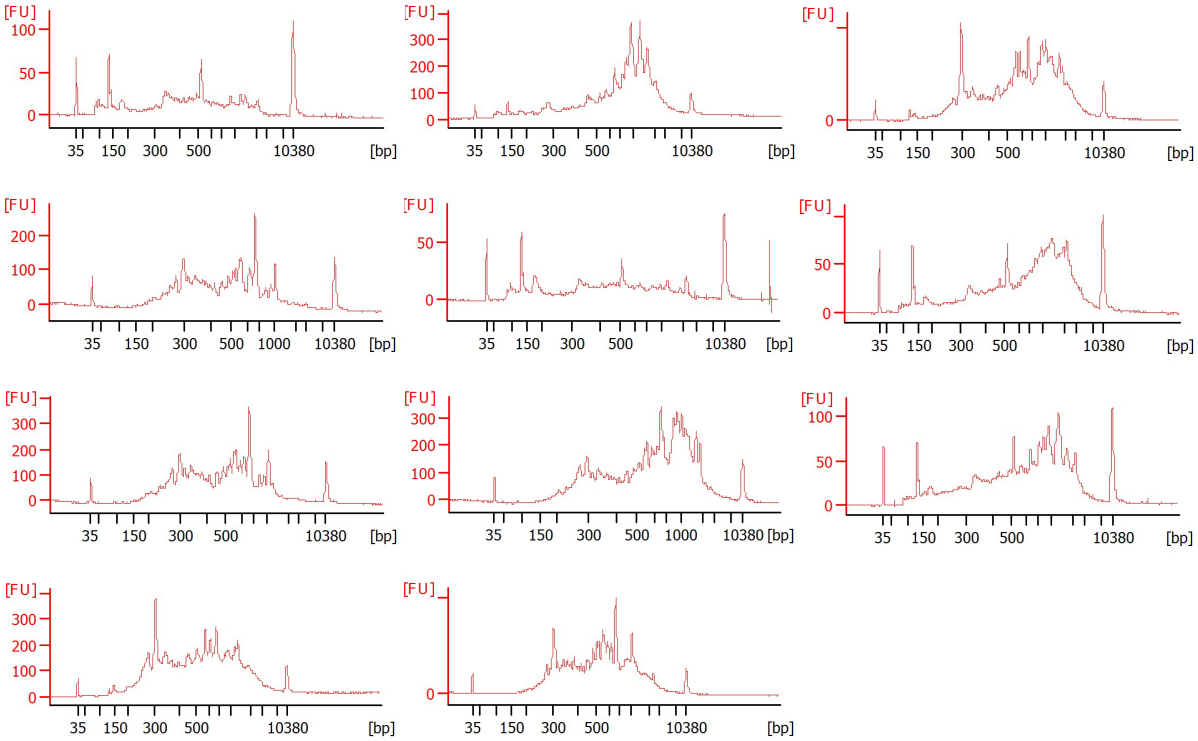


Figure 4.4 Quality of cDNA amplified from single cell bone marrow megakaryocytes
BioAnalyser electropherograms of cDNA generated from 11 single cell bone marrow megakaryocytes, showing variable quality.

Plate ID	Sequencing Lane ID
03G	SLX-8757
60Z 16N	SLX-9491
60Z 32N	SLX-9492
60Z 8N	SLX-9493
60Z 4N/2N	SLX-9494
7TA 16N	SLX-9495
7TA 32N	SLX-9496
CYH 16N	SLX-9490
CYH 32N	SLX-9497
CYH 8N	SLX-9488
CYH 4N	SLX-9489
CYH 2N	SLX-10160
CUP 8N	SLX-10154
CUP 16N	SLX-10155
CUP 32N	SLX-10156
CUP 4N/2N	SLX-10157
CST 2N	SLX-10158
318	SLX-10208
3XI	SLX-10208
ATZ	SLX-10208
DSP	SLX-10208
H4K	SLX-10208
M8T	SLX-10482

Table 4.1 Patient and plate ID on each sequencing lane

Sample collection	Amplification	Nextera Library prep / RNA-seq
1	1	1
03G	03G	SLX-8757
2	2	2
318	60Z 16N	SLX-9491
3	60Z 32N	SLX-9492
3XI	60Z 8N	SLX-9493
4	60Z 4N/2N	SLX-9494
60Z 16N	3	SLX-9495
60Z 32N	7TA 16N	SLX-9496
60Z 8N	7TA 32N	3
60Z 4N/2N	4	SLX-9490
5	CYH 16N	SLX-9497
7TA 16N	CYH 32N	SLX-9488
7TA 32N	CYH 8N	SLX-9489
7TA 8N	CYH 4N	SLX-10160
6	CYH 2N	4
ATZ	5	SLX-10154
7	CUP 8N	SLX-10155
CUP 8N	CUP 16N	SLX-10156
CUP 16N	CUP 32N	SLX-10157
CUP 32N	CUP 4N/2N	SLX-10158
CUP 4N/2N	CST 2N	5
8	6	SLX-10208 (multicellular pools)
CST 2N	318	6
9	7	SLX-10482 (multicellular pools)
CYH 16N	H4K	
CYH 32N	ATZ	
CYH 8N	8	
CYH 4N	3XI	
CYH 2N	DSP	
10	9	
DSP	M8T	
11		
H4K		
12		
M8T		

Table 4.2 Megakaryocyte collection, amplification, library preparation/RNA-seq batches

4.2.1.4 RNA sequencing data: Quality measures

Figure 4.5 shows metrics per sequencing lane for total read count, alignment to the reference transcriptome (GRCh37), proportion of reads mapped to external spike-in ERCCs compared to endogenous transcripts, no of genes called, proportion of reads mapping to annotated exonic regions and the proportion of reads mapping to mitochondria transcripts.

It is clearly evident that the quality data for the 2 lanes formed of megakaryocyte pools rather than single cells (SLX-10208, SLX-10482) was better with higher total read count, alignment rate, and lower ERCC ratio than the single cell lanes. In approximately 10% of all single cell and multicellular samples >50% reads were mapped to mitochondrial transcripts. While this may be a reflection of the biology of these cells, it effectively reduced the sequencing depth for each sample. Mitochondrial transcripts were found to form the top 10 genes expressed in all sequencing lanes (including: *MT-RNR2, MT-CO1, MT-RNR1, MT-ND4, MT-CO2, MT-CO3*). There was a higher proportion of reads mapping to the mitochondria in the multicellular samples compared with the single cells (samples with >50% reads mapping to mitochondrial transcripts: 36% and 8% respectively) (Figure 4.5) which might partially explain the better alignment rates.

While the vast majority of single cell sequencing demonstrated low quality metrics, the multicellular samples performed better. Despite this, saturation of sequencing capacity by reads mapping to mitochondrial transcripts was problematic in both single cell and multicellular samples. While posing an unexpected and substantial technical challenge this is in keeping with previously reported human platelet RNA-seq datasets (Rowley et al. 2011; Eicher et al. 2016) and the knowledge that both megakaryocytes and platelets are densely populated with mitochondria. In an attempt to overcome this challenge, BioAnalyser traces and Taqman real time qPCR were used in an attempt to identify samples to take forward for downstream sequencing analysis as sequencing all samples with such low yield of useful transcriptome data would not be financially viable. Therefore, the aim was to identify features at the cDNA stage of library preparation that would allow discrimination of samples (single-cell or multicellular pools) for RNA-seq.

As described above cDNA from multicellular pools were consistently of good quality i.e., full length cDNA, and therefore for multicellular pools BioAnalyser would be unable to predict which samples would perform favourably in gene expression analysis. The quality of cDNA according to BioAnalyser traces in single cells was more variable. Taqman real-time qPCR assays were used to assess levels of the following transcripts in low and high quality single cell cDNA (according to BioAnalyser trace): *MT-RNR2* (most highly expressed transcript in sequencing data, *PPBP* (most highly expressed genes in megakaryocyte/platelet). As demonstrated in Figure 4.6 while higher quality single cell cDNA expresses higher levels of

MT-RNR2 than lower quality single cell cDNA, *PPBP* levels are undetectable irrespective of cDNA quality by BioAnalyser. This indicates that quality of cDNA reverse transcribed from RNA from either multicellular pools or single megakaryocytes does not help to differentiate between samples with higher numbers of called genes/complexity of gene expression. Moreover, it reflects a more efficient reverse transcription of the most abundant signal – which in this case is mitochondrial reads.

Taqman real-time qPCR assays for transcripts known to be highly expressed in the megakaryocyte/platelet lineage (*PPBP, PF4, SDPR*) were then correlated with their RNA-seq expression levels in both single cell and multicellular samples to ascertain whether qPCR data might be a useful predictor of samples that would be more useful to take forward for downstream RNA-seq analysis. cDNA from single cells in sequencing lane SLX-9489 was selected as either highly expressed in RNA-seq results or not expressed. The corresponding level on using qPCR Taqman assays was then correlated (Table 4.3). No correlation was demonstrated between qPCR and RNA-seq transcript expression. Therefore, it was concluded that it is not possible to prospectively identify cDNA from single cells for further transcriptome analysis.

In the multicellular pools however, a clear correlation was found between RNA-seq expression of the highly expressed megakaryocyte/platelet transcript PPBP and detection by Taqman qPCR probes as demonstrated in Figure 4.7a (r^2=0.507). While PPBP and PF4 were consistently detected in multicellular pools, the level varied with only detection at high levels correlating with detection of reads at RNA-seq (Figure 4.7b and c). Figure 4.7b shows the typical amplification profile for *MT-RNR2, PPBP* and *PF4* seen for cells with low megakaryocyte/platelet specific transcript counts on RNA-seq. Figure 4.7c conversely shows the same typical amplification profile for cells with higher megakaryocyte/platelet specific transcript counts on RNA-seq. Interestingly the latter group have similar *MT-RNR2* and *PPBP* CT detection levels on qPCR, compared with the former in which *MT-RNR2* is detected significantly earlier on qPCR.

On the basis that there is no reliable method to prospectively identify single cells prior to sequencing that could provide informative RNA-seq data either by qPCR or BioAnalyser traces, all further megakaryocyte sequencing was restricted to multicellular pools only from 20-100cells.

4.2.1.5 Sequencing approach for megakaryocytes

Given the technical challenges described above, particularly with single cells, in order to optimise the acquisition of informative RNA-seq data from bone marrow megakaryocytes in a financially viable manner, the following approach was taken.

As there was no way to prospectively identify high quality single megakaryocytes prior to sequencing, no further single megakaryocytes were sequenced and the single cell RNA-seq data shown above (5 patients) were subject to stringent quality filtering at the time of analysis to ascertain the value and reliability of further downstream analysis.

All megakaryocyte pools collected by ploidy, were sequenced irrespective of prospective qPCR data due to the small numbers (samples from 6/34 patients only, 1-2 wells collected per ploidy group, per patient).

16/34 patients had megakaryocyte pools sorted irrespective of ploidy level, with between 8-96 wells collected per patient depending on the quantity of the starting bone marrow material. From these multicellular pools, samples were be prospectively identified and selected for RNA-seq on the basis of cDNA gene expression using Taqman probes as described below.

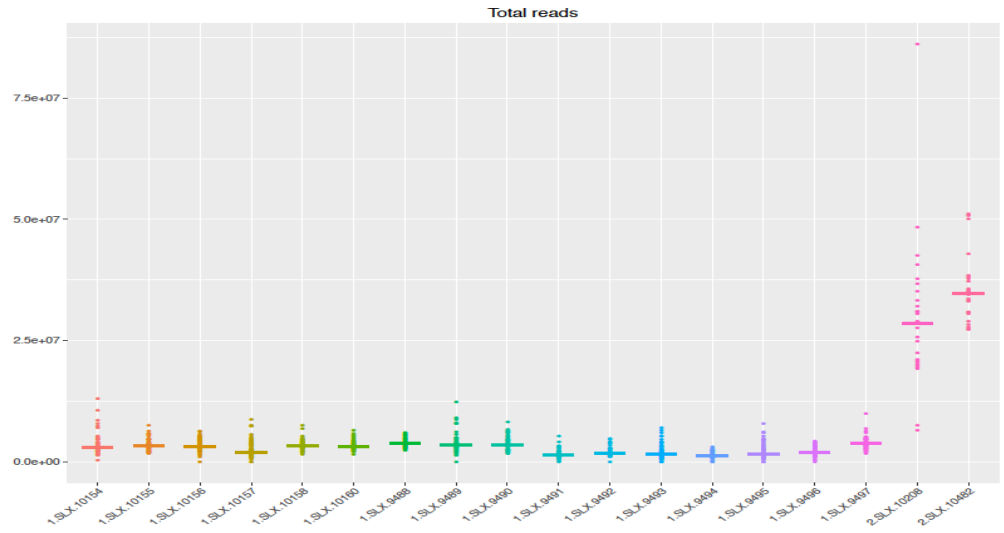
PPBP was selected as the most useful marker gene to identify high quality multicellular pools. This was on the basis that it is the most highly expressed platelet specific transcript based on reported RNA-seq data (Rowley et al. 2011), therefore it is the most likely of all highly expressed megakaryocyte/platelet transcripts to be detected at qPCR. Furthermore, *PPBP* CT was seen to be highly correlated with expression at RNA-seq (Figure 4.7a). To investigate if expression levels of *PPBP* could act as a quality marker its level of expression was correlated with no of called genes in those samples. No. of called genes was found to correlate positively with *PPBP* expression at RNA-seq level, r^2=0.261 (Figure 4.8a) and no. called genes was significantly higher in samples with a *PPBP* read count above 300 reads, p=0.0003 (t-test) (Figure 4.8b). Furthermore, expression of the housekeeping gene *B2M* at RNA-seq was only found in the samples with a *PPBP* read count of >300 reads.

However, the main challenge identified with RNA-seq analysis in megakaryocytes was the saturation of sequencing capacity with mitochondrial transcripts, seen clearly in the quality metrics of RNA-seq data (Figure 4.5F) and in the typical qPCR amplification plots with large ΔCTs observed between *MT-RNR2* and *PPBP*/*PF4* (Figure 4.7b). To investigate if ΔCT *PPBP – MT-RNR2* based on qPCR using *MT-RNR2* and *PPBP* Taqman probes might be the ideal marker of quality, ΔCT *PPBP – MT-RNR2* was correlated with *PPBP* gene expression in the multicellular RNA-seq data and no. of called genes. Figure 4.9a shows a strong inverse correlation between ΔCT *PPBP – MT-RNR2* and *PPBP* read count (r^2=0.610, more significant

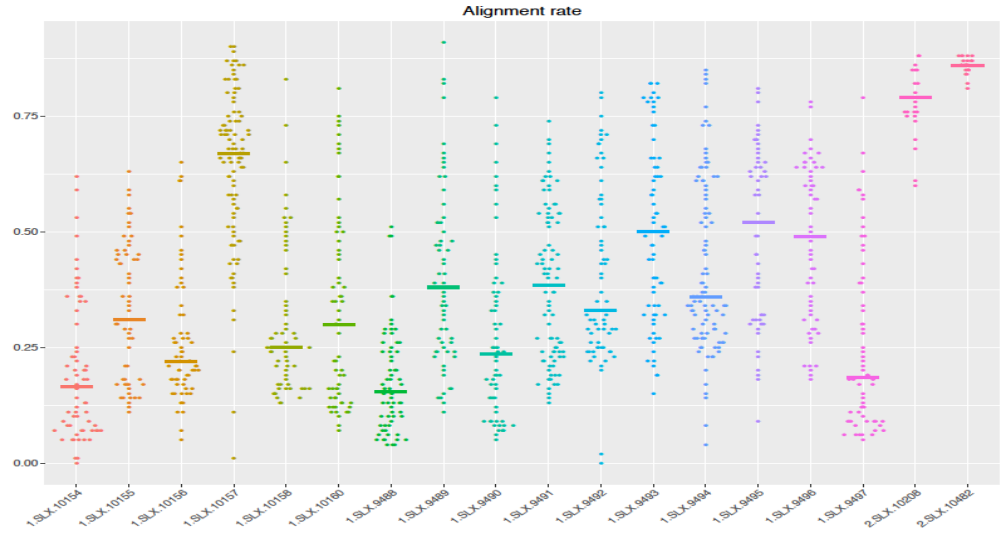
than *PPBP* CT alone, Figure 4.7a). Figure 4.10a also shows an inverse correlation between ΔCT *PPBP* – *MT-RNR2* and no. of called genes (r^2=0.298).

Various arbitrary ΔCT *PPBP* – *MT-RNR2* cut-offs for inclusion were trialled. A ΔCT level between expression of *PPBP* and *MT-RNR2* by Taqman qPCR assay of 7 or less was used to prospectively identify multicellular megakaryocyte samples for RNA-seq. Figure 4.9b demonstrates significantly higher *PPBP* read counts in samples with ΔCT *PPBP* – *MT-RNR2* <7 compared with >7 (p=0.0109), whereas there is no significant difference in *PPBP* read counts between samples with ΔCT *PPBP* – *MT-RNR2* <8 compared with >8 (Figure 4.9c). Similar observations were made with no. of called genes: Figure 4.10b shows significantly greater no. of called genes in samples with ΔCT *PPBP* – *MT-RNR2* <7 compared with >7 (p=0.0005) (t-test). Setting the cut-off as a ΔCT *PPBP* – *MT-RNR2* of 8 reduced this significance to an order of magnitude less (p=0.006, (t-test) Figure 4.10c).

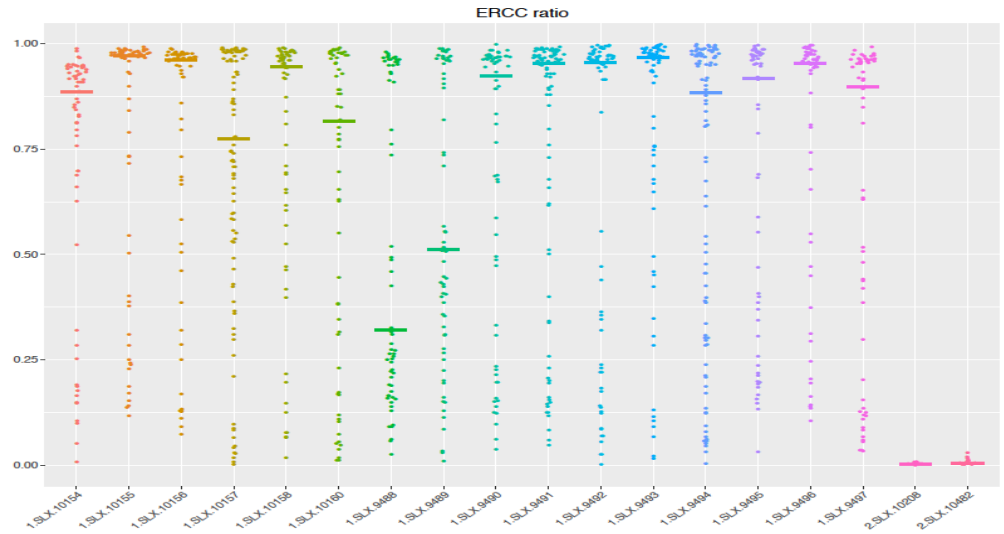
A


Total reads
7.5e+07
5.0e+07
2.5e+07
0.0e+00
1.SLX.10154
1.SLX.10155
1.SLX.10156
1.SLX.10157
1.SLX.10158
1.SLX.10160
1.SLX.9488
1.SLX.9489
1.SLX.9490
1.SLX.9491
1.SLX.9492
1.SLX.9493
1.SLX.9494
1.SLX.9495
1.SLX.9496
1.SLX.9497
2.SLX.10208
2.SLX.10482

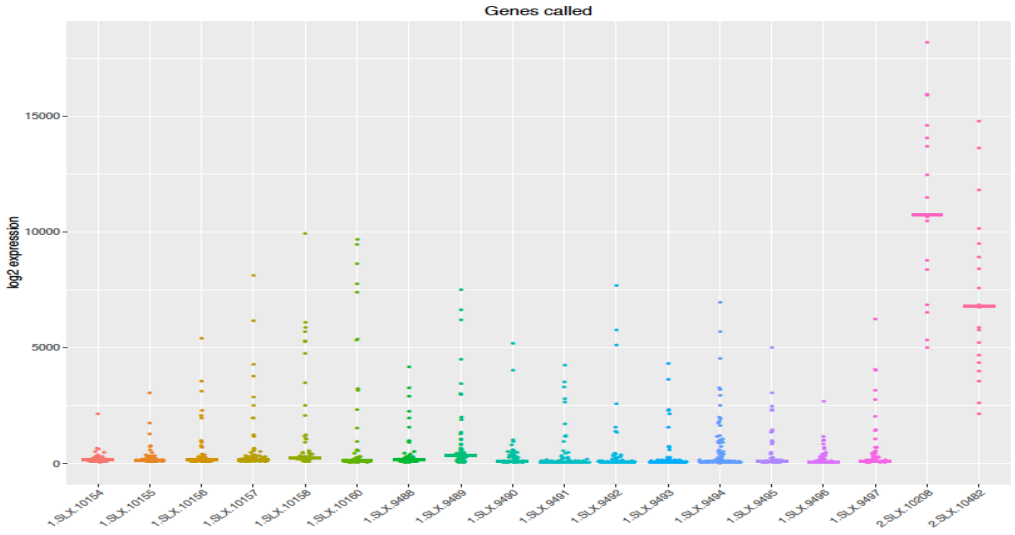
B


Alignment rate
0.75
0.50
0.25
0.00
1.SLX.10154
1.SLX.10155
1.SLX.10156
1.SLX.10157
1.SLX.10158
1.SLX.10160
1.SLX.9488
1.SLX.9489
1.SLX.9490
1.SLX.9491
1.SLX.9492
1.SLX.9493
1.SLX.9494
1.SLX.9495
1.SLX.9496
1.SLX.9497
2.SLX.10208
2.SLX.10482

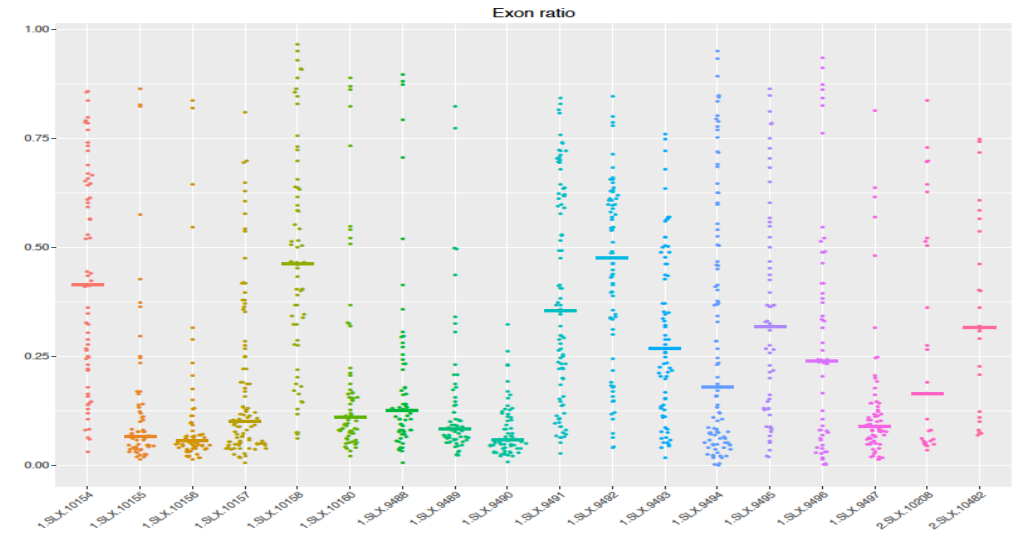
C


ERCC ratio
1.00
0.75
0.50
0.25
0.00
1.SLX.10154
1.SLX.10155
1.SLX.10156
1.SLX.10157
1.SLX.10158
1.SLX.10160
1.SLX.9488
1.SLX.9489
1.SLX.9490
1.SLX.9491
1.SLX.9492
1.SLX.9493
1.SLX.9494
1.SLX.9495
1.SLX.9496
1.SLX.9497
2.SLX.10208
2.SLX.10482

D



E



F

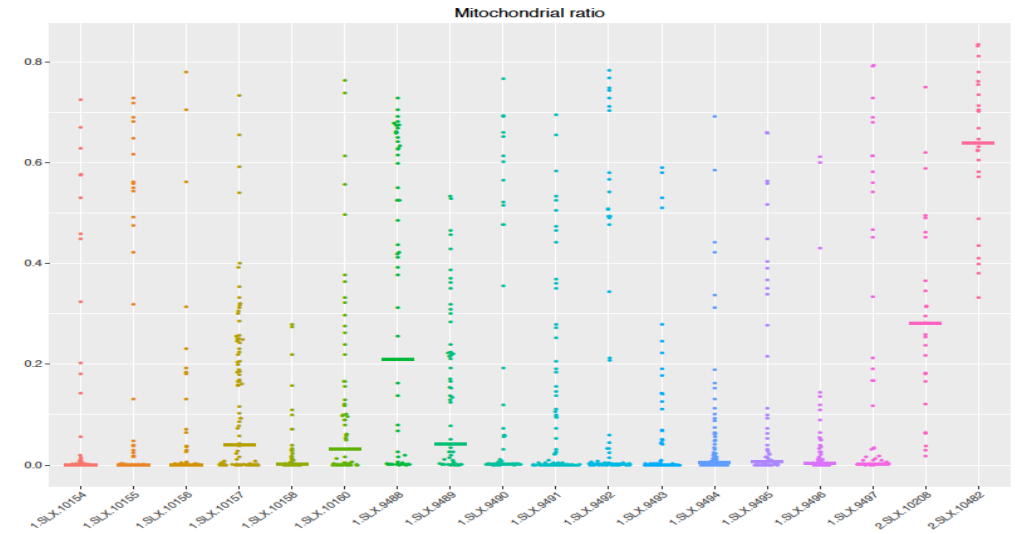


Figure 4.5 Quality control metrics/sequencing lane. A. Total reads; B. Rate of alignment to transcriptome; C. Ratio of reads mapping to ERCCs; D. No. genes called; E. Ratio of exonic reads to exonic+intronic; F. Ratio of reads mapping to mitochondrial genes.

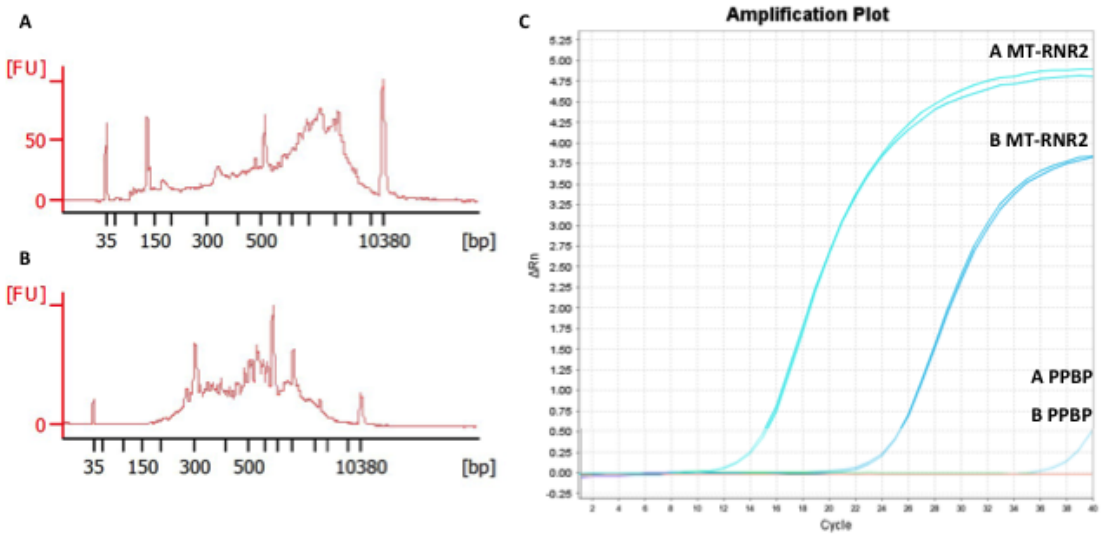


Figure 4.6 Full length cDNA from single cell megakaryocytes correlates with *MT-RNR2* expression.
Representative high quality(A) and low quality(B) single megakaryocyte BioAnalyser electropherograms. C. Taqman qPCR amplification plot for cell A and cell B showing the amplification of 2 transcripts: *MT-RNR2* and *PPBP*, X axis shows the cycle number or CT at which amplification begins.

Single cell RNAseq ID	PPBP		PF4		SDPR	
	Taqman assay (CT)	RNA-seq (read count)	Taqman assay (CT)	RNA-seq (read count)	Taqman assay (CT)	RNA-seq (read count)
SLX.9489.N707_N502	Undetermined	30077	35.41544724	15651	Undetermined	1887
SLX.9489.N706_N502	Undetermined	21139	Undetermined	9101	Undetermined	976
SLX.9489.N704_N501	Undetermined	493	Undetermined	898	Undetermined	1691
SLX.9489.N703_N506	Undetermined	25858	Undetermined	8042	Undetermined	866
SLX.9489.N706_N501	Undetermined	393	Undetermined	876	Undetermined	2180
SLX.9489.N705_N501	Undetermined	247	Undetermined	1393	Undetermined	336
SLX.9489.N706_N503	Undetermined	0	Undetermined	0	Undetermined	0
SLX.9489.N705_N508	Undetermined	0	Undetermined	2	Undetermined	5
SLX.9489.N703_N508	Undetermined	1	36.77412033	0	Undetermined	0
SLX.9489.N711_N506	Undetermined	0	Undetermined	0	Undetermined	0
SLX.9489.N712_N503	37.05277252	0	36.98429108	0	Undetermined	0
SLX.9489.N703_N504	Undetermined	0	Undetermined	0	Undetermined	0

Table 4.3 Transcript detection in single megakaryocytes. Correlation of detection of *PPBP*, *PF4* and *SDPR* by qPCR and RNA-seq in single cells from SLX-9489. Samples were selected on the basis of expression of *PPBP*, *PF4* and *SPDR* by RNA-seq (n=6 high expression - shaded orange, n=6 low expression). Taqman real-time qPCR was then performed on cDNA for these samples for the expression of *PPBP*, *PF4* and *SDPR*.

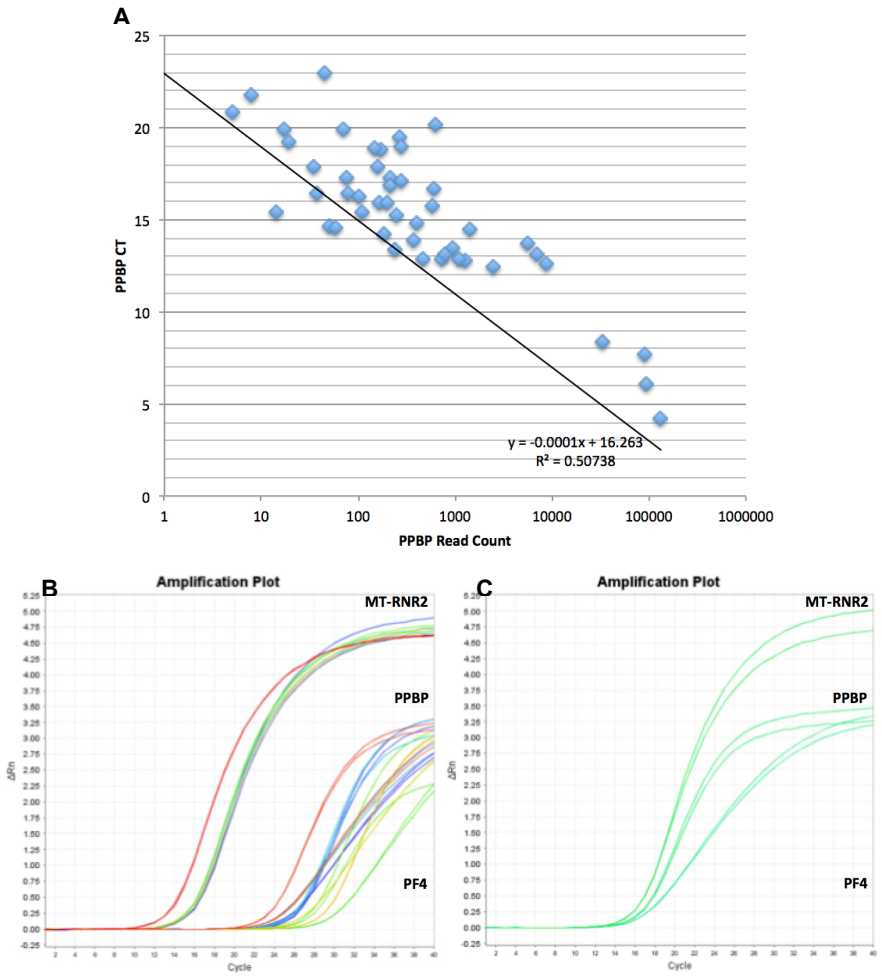


Figure 4.7 Transcript detection in multicellular megakaryocytes. A. Inverse correlation between CT expression of *PPBP* by Taqman real-time qPCR and *PPBP* read count on RNA-seq, r^2=0.507. Representative amplification plots showing cycle no at which *MT-RNR2*, *PPBP* and *PF4* are detected using Taqman real-time qPCR for multicellular samples with low *PPBP* CTs/low *PPBP* reads on RNA-seq (B), high *PPBP* CTs/high *PPBP* reads on RNA-seq (C).

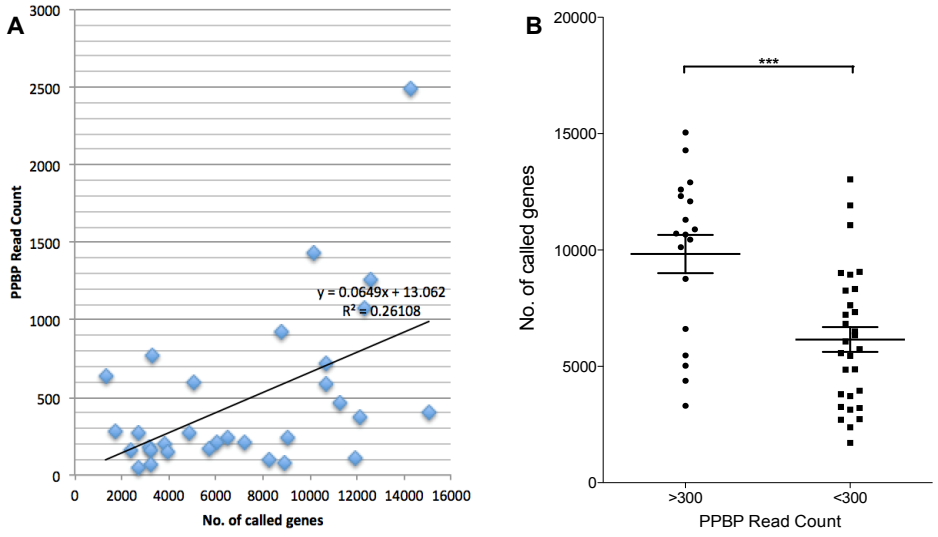


Figure 4.8 No. of called genes according to *PPBP* Read Count in multicellular pool megakaryocyte RNA-seq. A. Correlation between *PBPP* read count by RNA-seq and no. of called genes in that sample, r2=0.261. Read counts >3000 were removed; B. No. of called genes according to *PPBP* read count above or below 300 as an abitrary cut off shows that samples with >300 *PPBP* read counts have significantly higher numbers of called genes (t-test, p=0.0003).

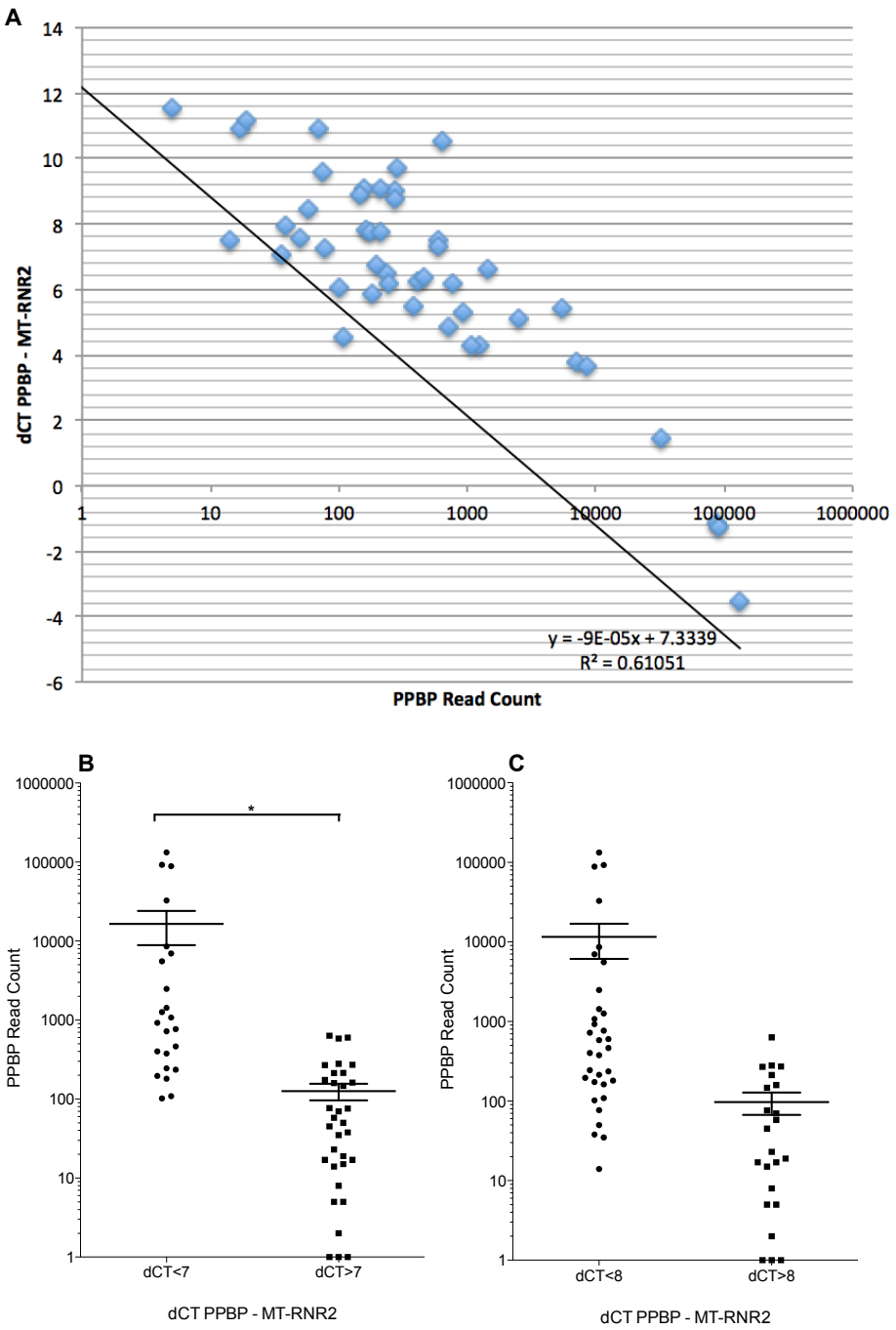


Figure 4.9 Relationship between ΔCT *PPBP – MT-RNR2* and *PPBP* read count. A. Inverse correlation between ΔCT *PPBP – MT-RNR2* by Taqman real-time qPCR and *PPBP* read count on RNA-seq, r^2=0.610, in multicellular megakaryocyte samples. B. Significantly higher *PPBP* read count in samples with ΔCT *PPBP – MT-RNR2* <7 compared to >7, p=0.0109. C. No significant difference in *PPBP* read count in samples with ΔCT *PPBP – MT-RNR2* <8 compared to >8.

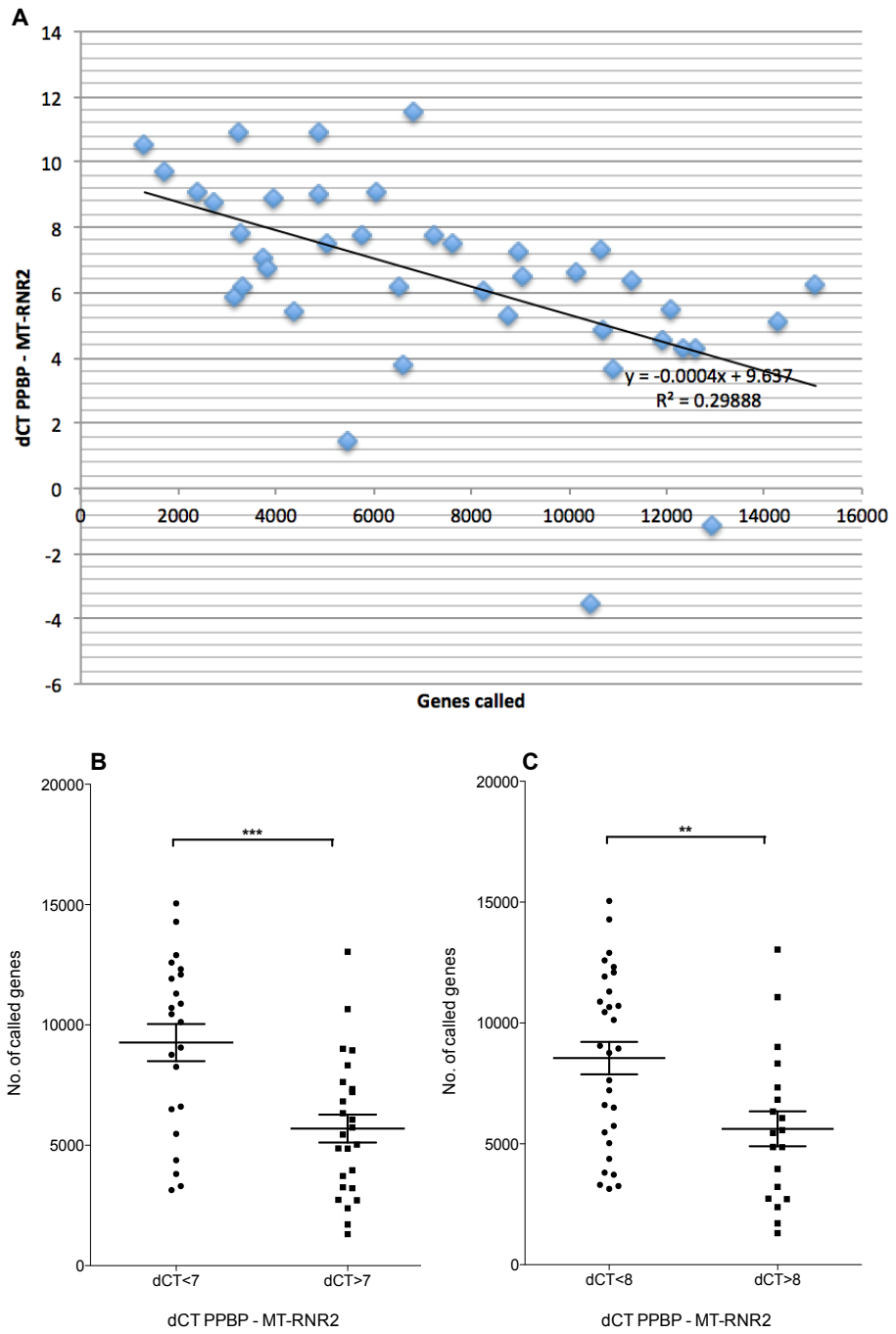


Figure 4.10 Relationship between ΔCT *PPBP – MT-RNR2* and No. of called genes. A. Inverse correlation between ΔCT *PPBP – MT-RNR2* by Taqman real-time qPCR and No. of called genes on RNA-seq, r^2=0.298, in multicellular megakaryocyte pools. B. Significantly higher no. of called genes in samples with ΔCT *PPBP – MT-RNR2* <7 compared to >7, p=0.0005. C. Significantly higher no. of called genes in samples with ΔCT *PPBP – MT-RNR2* <8 compared to >8, p=0.006.

4.2.2 Defining the primary human megakaryocyte transcriptome

4.2.2.1 Sample preparation and RNA-seq library preparation

To define the primary human megakaryocyte transcriptome in the context of current evidence, the objective was to make comparisons with the known megakaryocyte transcriptome from megakaryocytes cultured from human CD34+ cells.

Bone marrow megakaryocytes were flow sorted from sternal bone marrow collected from 6 individuals without coronary artery disease as multicellular pools (20-100 multicellular pools) for the purpose of defining for the first time the primary human megakaryocyte transcriptome. The modified Smart-seq2 protocol (G&T-seq) was used, as described above, to generate RNA-seq libraries for transcriptome analysis from a total of 85 megakaryocyte multicellular pools that were FACS sorted irrespective of megakaryocyte ploidy. Table 4.4 shows the baseline characteristics for these individuals. All individuals were undergoing non-coronary cardiac surgery and had no evidence of coronary artery disease (by pre-operative coronary angiography).

To enable a comparison with cultured megakaryocytes, megakaryocytes were derived from human adult blood CD34+ cells as described in Section 2. The phenotypic appearance of the cultured megakaryocytes (Figure 4.11) was in contrast to the bone marrow megakaryocytes (Figure 4.1) due to size and nuclear appearance. 4 biological replicates were used and RNA-seq libraries were generated using the SMARTer Ultra Low Input RNA for Illumina Sequencing kit which utilises the same poly A pulldown as the Smart-seq2 protocol. Of note ERCC spike-ins were added in a concentration of 1:1000, compared with a concentration of $1:4x10^6$ in the Smartseq2 protocol for bone marrow megakaryocytes.

85 megakaryocyte multicellular pools sorted irrespective of ploidy were prospectively identified on the basis of ΔCT *PPBP* – *MT-RNR2* as defined by Taqman qPCR gene expression as described above (data presented in Section 5.2.2.2). 125bp paired-end sequencing was performed at an average sequencing depth of approximately 50 million reads/sample. 3 multicellular pools with the lowest ΔCT *PPBP* – *MT-RNR2* were also selected from each individual to take forward for deeper 125bp paired-end sequencing at a depth of 500 million reads/sample (data presented in Section 5.2.2.2). All 4 biological replicates of the megakaryocytes derived from CD34+ cells were sequenced over a single 125bp paired end High-seq lane (~500 million reads/sample).

In terms of batch effects Table 4.5 shows each sample plate ID and corresponding sequencing lane ID whereas Table 4.6 outlines the batches in which samples were collected, cDNA reverse transcribed/amplified and Nextera libraries prepared and sequenced.

4.2.2.2 RNA sequencing data: Quality measures

Figure 4.12 shows metrics per sequencing lane for total read count, alignment to the reference transcriptome (GRCh37), proportion of reads mapped to external spike-in ERCCs compared to endogenous transcripts, no of genes called, proportion of reads mapping to annotated exonic regions and the proportion of reads mapping to the mitochondria. The most marked differences are in comparison of the low input sequencing with bulk cultured megakaryocytes where there are a higher number of genes called and a significantly greater representation of ERCC spike-ins. There are also markedly lower numbers of reads mapping to mitochondrial transcripts.

Table 4.4 Baseline patient characteristics

	Non-coronary disease controls (n=6)
Gender M	2
Age (mean ± SD)	57 ± 20.3
Hypertension	4
Hypercholesterolaemia	2
Diabetes	1
Smoking	0
Coronary disease	0
Antiplatelets	2
Left ventricular dysfunction	3
Aortic valve surgery	3
Mitral valve surgery	3
Hb (mean ± SD)	13.6 ± 1.2
WC (mean ± SD)	6.9 ± 1.9
Plt (mean ± SD)	250 ± 38.8

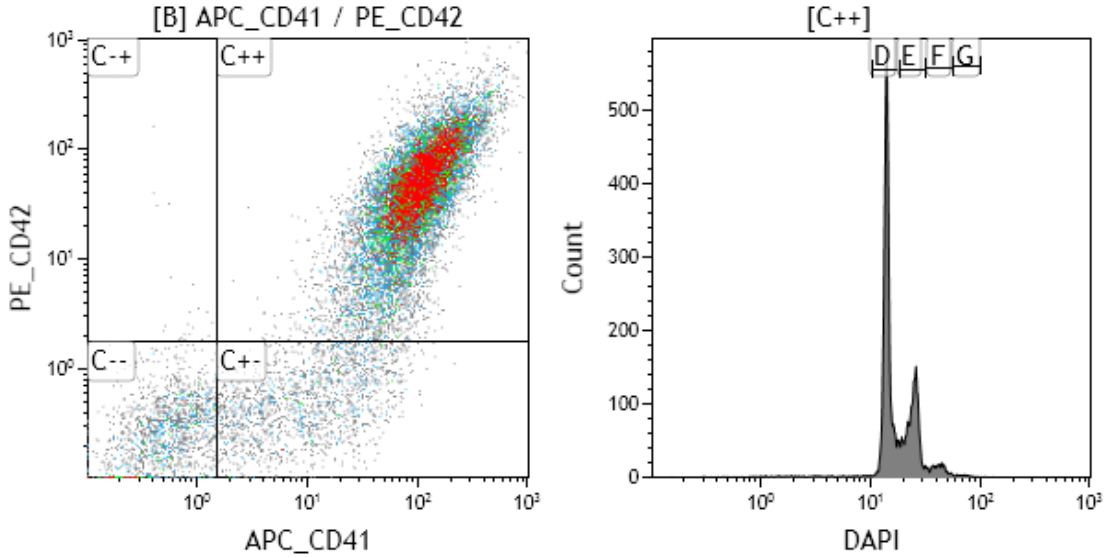
Plate ID	Sequencing Lane ID
Megakaryocytes:	
3XI	SLX-10208, SLX-10546, SLX-11601
DSP	SLX-10208, SLX-10546, SLX-11602
ATZ	SLX-10208, SLX-10546, SLX-11603
H4K	SLX-10208, SLX-10546, SLX-11604
M8T	SLX-10482, SLX-11605
Q6H	SLX-10546, SLX-11605
Cultured megakaryocytes	SLX-7719

Table 4.5 Patient and plate ID on each sequencing lane

Sample collection	Amplification	Nextera Library prep / RNA-seq
1	1	1
Cultured megakaryocytes	Cultured megakaryocytes	SLX-7719
2	2	2
3XI	H4K	SLX-10208
3	ATZ	3
ATZ	3	SLX-10482
4	3XI	4
DSP	DSP	SLX-10546
5	4	5
H4K	M8T	SLX-11601
6	5	SLX-11602
M8T	Q6H	SLX-11603
7		SLX-11604
Q6H		SLX-11605

Table 4.6 Megakaryocyte and platelet collection, amplification, and library preparation/RNA-seq

A



Gate	Number	%Total	%Gated
All	16,846	68.03	100.00
C--	1,398	5.65	8.30
C-+	23	0.09	0.14
C+-	1,647	6.65	9.78
C++	13,778	55.64	81.79

Gate	Number	%Total	%Gated
All	13,778	55.64	100.00
D	8,480	34.24	61.55
E	4,119	16.63	29.90
F	746	3.01	5.41
G	108	0.44	0.78

B

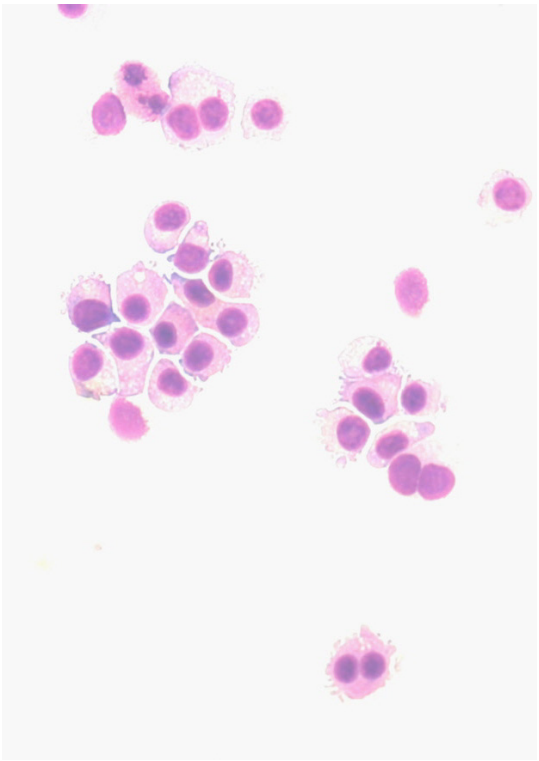
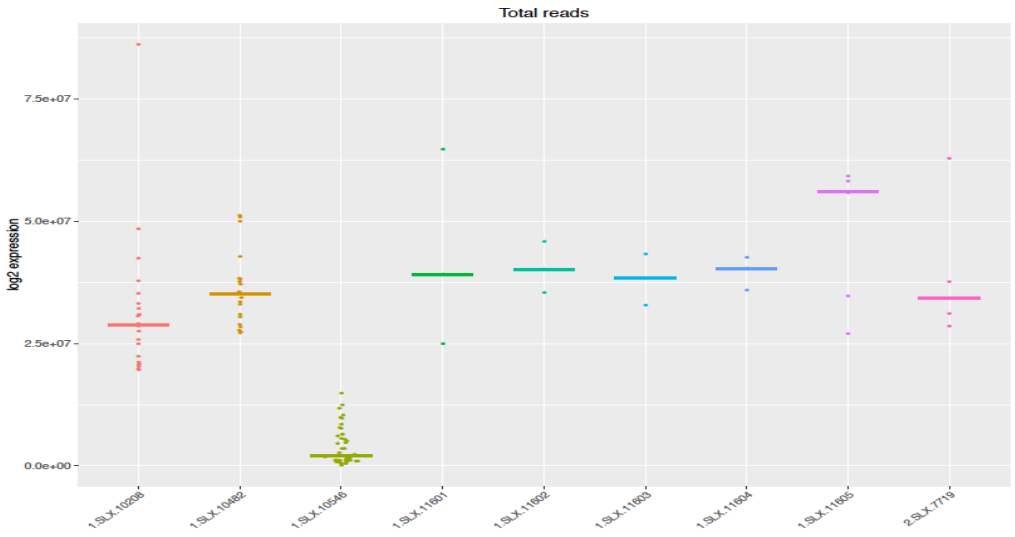


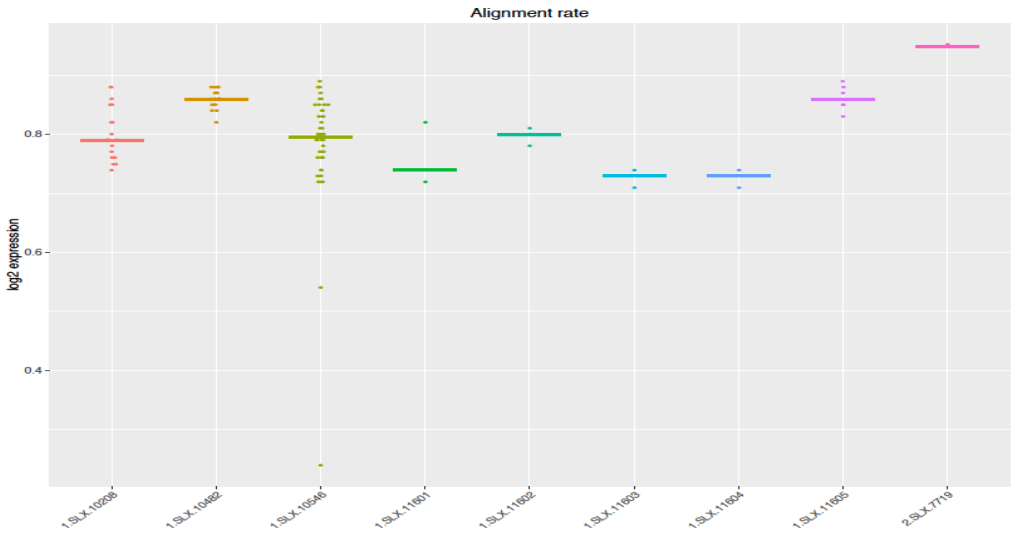
Figure 4.11 Cultured megakaryocyte phenotype

A. Harvested cell suspension was stained with APC conjugated anti-CD41a, PE conjugated anti-CD42a and Hoechst. Flow cytometry data of cultured megakaryocytes showing 81% to be CD41+CD42+, Hoechst staining demonstrates 61% of these are 2N, 30% are 4N, 5% are 8N and less than 1% are 16N. B. Cytocentrifugation of enriched megakaryocyte population with Rapid Romanowsky, x40 objective.

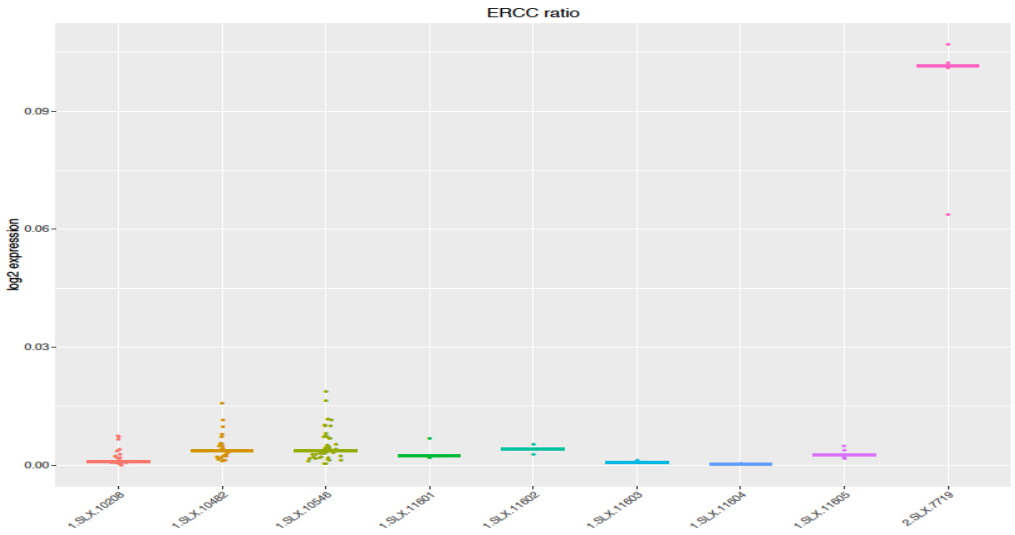
A


Total reads
log2 expression
7.5e+07
5.0e+07
2.5e+07
0.0e+00
1.SLX.10208
1.SLX.10482
1.SLX.10546
1.SLX.11601
1.SLX.11602
1.SLX.11603
1.SLX.11604
1.SLX.11605
2.SLX.7719

B


Alignment rate
log2 expression
0.8
0.6
0.4
1.SLX.10208
1.SLX.10482
1.SLX.10546
1.SLX.11601
1.SLX.11602
1.SLX.11603
1.SLX.11604
1.SLX.11605
2.SLX.7719

C


ERCC ratio
log2 expression
0.09
0.06
0.03
0.00
1.SLX.10208
1.SLX.10482
1.SLX.10546
1.SLX.11601
1.SLX.11602
1.SLX.11603
1.SLX.11604
1.SLX.11605
2.SLX.7719

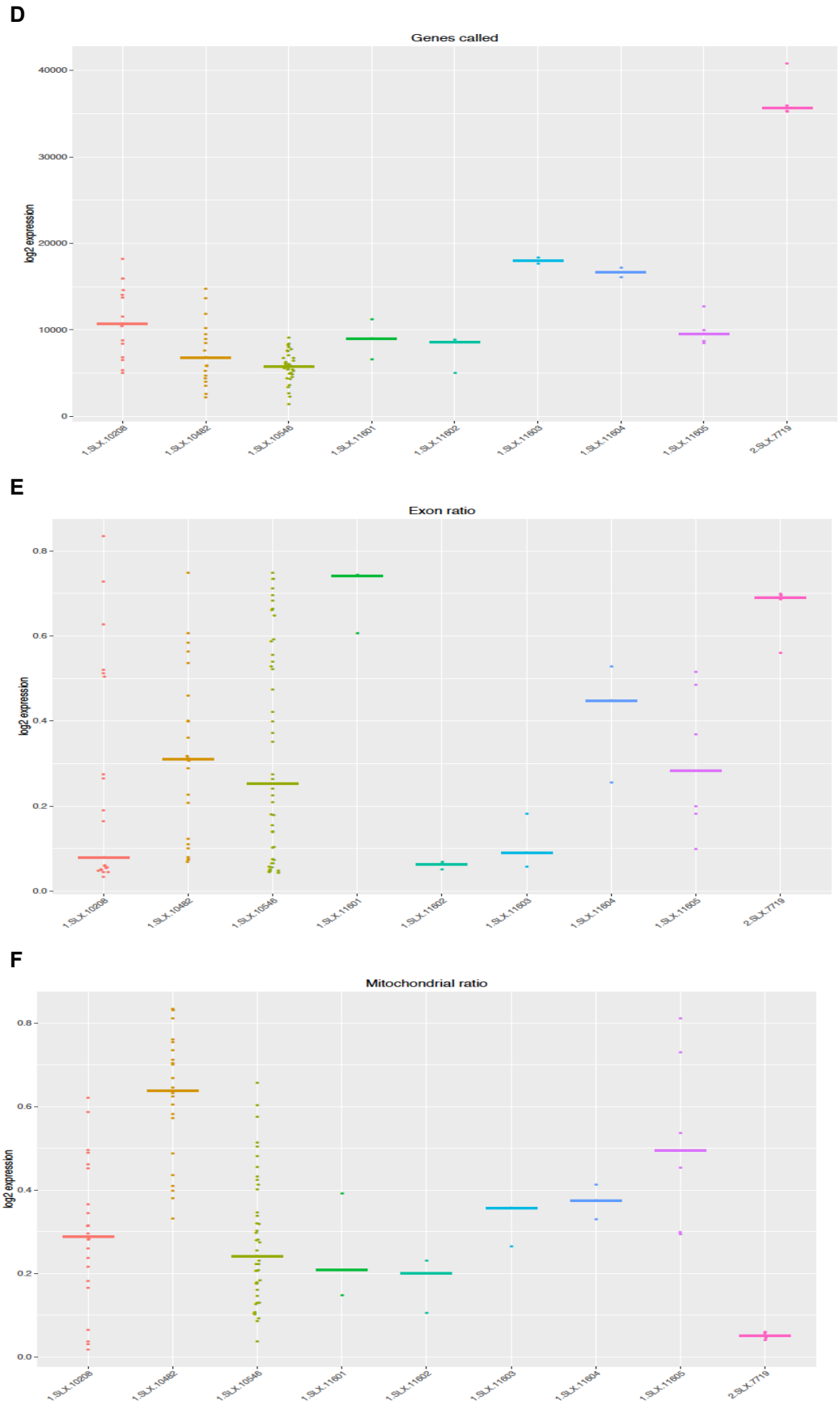


Figure 4.12 Quality control metrics/sequencing lane. A. Total reads; B. Rate of alignment to transcriptome; C. Ratio of reads mapping to ERCCs; D. No. genes called; E. Ratio of exonic reads to exonic+intronic; F. Ratio of reads mapping to mitochondrial genes.

4.2.2.3 Gene expression analysis

The 100 most highly expressed genes in cultured megakaryocytes and bone marrow megakaryocytes are shown in Table 4.7. In total, there were 30 genes that overlapped between the most highly expressed genes between the two lists (Figure 4.13). The most highly expressed transcripts in cultured megakaryocytes appear to be ERCC spike-ins (18 out of 100) likely to be a result of the high concentrations at which were added in RNA-seq library preparation (also seen in QC data; Figure 4.12c). As expected there is a similarity between the 2 lists, with a number of megakaryocyte/platelet lineage genes expressed highly. Both cultured and bone marrow megakaryocytes expressed high levels of *PPBP, PF4, PLEK, TUBA4A, ACTB, RAP1B* and *TMSB4X* which have important functions in platelet activation. Furthermore, bone marrow megakaryocytes also expressed high levels of the platelet glycoproteins *ITGA2B* (CD41) and *CD36*. Cultured megakaryocytes also had high expression of the platelet activation markers *CD9* and *CD63*. There were a number of other genes with shared high expression between the 2 datasets including those encoding proteins involved in cellular respiration and metabolism such as cytochrome c oxidase subunits, NADH dehydrogenase subunits and genes encoding ribosomal subunits. *MT-RNR1* and *MT-RNR2* were the most highly expressed genes overall in bone marrow megakaryocytes.

A GO analysis of the two lists and their overlapped 30 genes shows the main biological themes significantly over-represented (Figure 4.14, 4.15) and full GO lists are found in Appendix 4, Tables S4.1, S4.2, S4.3. The main biological processes enriched in bone marrow megakaryocytes were involved in cellular metabolism and respiration, coagulation, haemostasis and wound healing, then platelet degranulation and cytoskeleton, followed by translation and protein targeting. These were also important biological themes in cultured megakaryocytes, and shared in the overlapping GO analysis (Figure 4.15). However, defence response was unique to the bone marrow megakaryocytes with the presence of the following transcripts: *S100A9, CAMP, LYZ, DEFA3, LTF, S100A8, DEFA4, S100A12*.

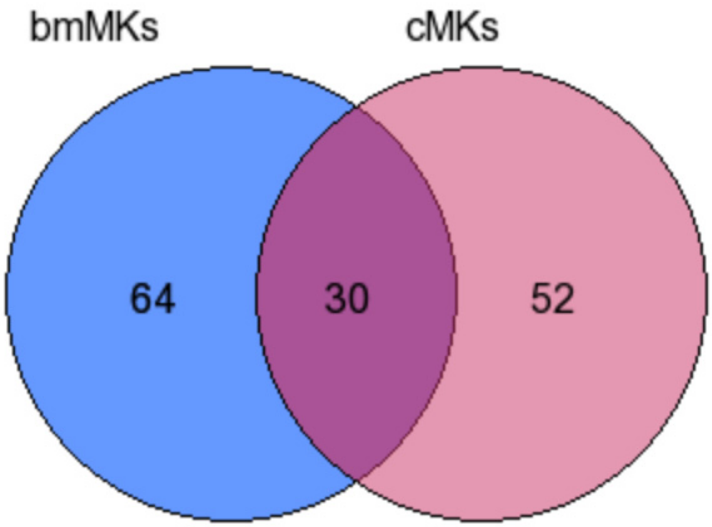
A differential expression analysis was then performed between the normalised read count data from both cultured and bone marrow megakaryocytes in order to discover quantitative changes in gene expression levels between the two data sets. Of the 521 differentially expressed features, 51 were shown to be significantly upregulated in bone marrow megakaryocytes relative to cultured megakaryocytes ($P<0.05$, Wald test). The full list is shown in Table 4.8 with levels of significance. Interestingly 9 out of 51 transcripts identified as upregulated in bone marrow megakaryocytes encoded for regions of immunoglobulins, 8 encode for constant domains of the immunoglobulin heavy chains alpha, gamma, kappa and lambda; 1 encoded the variable domain of the immunoglobulin kappa light chain. Apart from these, other upregulated genes included those involved in the immune response: *S100A9, AZU1, CAMP, LYZ, DEFA3, LTF, S100A8, DEFA4, S100A12*, as well as a number of highly

significant differentially expressed genes that map to the mitochondrial transcriptome. To reflect this the most highly expressed GO biological processes were defence response to fungus, bacteria, neutrophil aggregation and inflammatory response (Figure 4.15, Appendix 4, Table S4.5). The top 50 genes downregulated in bone marrow megakaryocytes (or upregulated in cultured megakaryocytes) are shown with significance levels in Table 4.9 and the full list of 470 genes are shown in Appendix 4, Table S4.4. The significance levels for the downregulated gene are markedly lower ($p>1x10^{-4}$, compared with $p>1x10^{-39}$) (Wald test). By contrast the most highly enriched GO biological processes were translation, protein targeting/transport, platelet degranulation and cell cycle shown in Figure 4.16 and Appendix 4, S4.6.

No	Cultured megakaryocytes		Bone marrow megakaryocytes	
	HGNC_ID	Gene_name	HGNC_ID	Gene_name
1	*ERCC-00096*	ERCC-00096	*MT-RNR2*	mitochondrially Encoded 16S RNA (MT-RNR2)
2	*ERCC-00002*	ERCC-00002	*COX1*	cytochrome c oxidase subunit I(COX1)
3	*ERCC-00074*	ERCC-00074	*MT-RNR1*	mitochondrially Encoded 12S RNA (MT-RNR1)
4	*COX1*	cytochrome c oxidase subunit I(COX1)	*ND4*	NADH dehydrogenase, subunit 4 (complex I)(ND4)
5	*ERCC-00130*	ERCC-00130	*COX2*	cytochrome c oxidase subunit II(COX2)
6	*ACTB*	actin beta(ACTB)	*ATP6*	ATP synthase F0 subunit 6(ATP6)
7	*MALAT1*	metastasis associated lung adenocarcinoma transcript 1 (MALAT1)	*ND1*	NADH dehydrogenase, subunit 1 (complex I)(ND1)
8	*COX2*	cytochrome c oxidase subunit II(COX2)	*CYTB*	cytochrome b(CYTB)
9	*MT-RNR2*	mitochondrially Encoded 16S RNA (MT-RNR2)	*ND2*	MTND2(ND2)
10	*CYB*	cytochrome b(CYTB)	*COX3*	cytochrome c oxidase III(COX3)
11	*PPBP*	pro-platelet basic protein(PPBP)	*ND5*	NADH dehydrogenase, subunit 5 (complex I)(ND5)
12	*ND4*	NADH dehydrogenase, subunit 4 (complex I)(ND4)	*ND4L*	NADH dehydrogenase, subunit 4L (complex I)(ND4L)
13	*TMSB4X*	thymosin beta 4, X-linked(TMSB4X)	*MTATP6P1*	MTATP6P1
14	*FTL*	ferritin light chain(FTL)	*ND3*	NADH dehydrogenase, subunit 3 (complex I)(ND3)
15	*PLEK*	pleckstrin(PLEK)	*HBB*	haemoglobin subunit beta(HBB)
16	*COX3*	cytochrome c oxidase III(COX3)	*ND6*	NADH dehydrogenase, subunit 6 (complex I)(ND6)
17	*ERCC-00004*	ERCC-00004	*ERCC-00074*	ERCC-00074
18	*ERCC-00113*	ERCC-00113	*IGHG2*	IGHG2
19	*ERCC-00046*	ERCC-00046	*ACTB*	actin beta(ACTB)
20	*B2M*	beta-2-microglobulin(B2M)	*S100A8*	S100 calcium binding protein A8(S100A8)
21	*ATP6*	ATP synthase F0 subunit 6(ATP6)	*S100A9*	S100 calcium binding protein A9(S100A9)
22	*ERCC-00136*	ERCC-00136	*TMSB4X*	thymosin beta 4, X-linked(TMSB4X)
23	*ERCC-00009*	ERCC-00009	*JCHAIN*	joining chain of multimeric IgA and IgM(JCHAIN)
24	*ERCC-00108*	ERCC-00108	*MALAT1*	metastasis associated lung adenocarcinoma transcript 1 (MALAT1)
25	*ND5*	NADH dehydrogenase, subunit 5 (complex I)(ND5)	*PPBP*	pro-platelet basic protein(PPBP)
26	*RPL3*	ribosomal protein L3(RPL3)	*B2M*	beta-2-microglobulin(B2M)
27	*TPI1*	triosephosphate isomerase 1(TPI1)	*PF4*	platelet factor 4(PF4)
28	*MIR4738*	microRNA 4738(MIR4738)	*MTCO1P12*	MTCO1P12
29	*EEF1A1*	eukaryotic translation elongation factor 1 alpha 1(EEF1A1)	*FTL*	ferritin light chain(FTL)
30	*ACTG1*	actin gamma 1(ACTG1)	*DEFA3*	defensin alpha 3(DEFA3)
31	*GAPDH*	glyceraldehyde-3-phosphate dehydrogenase(GAPDH)	*ATP8*	ATP synthase F0 subunit 8(ATP8)
32	*MT-RNR1*	mitochondrially Encoded 12S RNA (MT-RNR1)	*ERCC-00002*	ERCC-00002
33	*ARHGDIB*	Rho GDP dissociation inhibitor beta(ARHGDIB)	*ACTG1*	actin gamma 1(ACTG1)
34	*HBD*	Haemoglobin subunit delta(HBD)	*MTCO1P12*	MTCO1P12
35	*MIR1244-1*	microRNA 1244-1(MIR1244-1)	*IGLC2*	IGLC2
36	*SH3BGRL3*	SH3 domain binding glutamate rich protein like 3(SH3BGRL3)	*SRGN*	serglycin(SRGN)
37	*PF4*	platelet factor 4(PF4)	*IGHV3-53*	IGHV3-53
38	*LDHB*	lactate dehydrogenase B(LDHB)	*ERCC-00096*	ERCC-00096
39	*ND2*	MTND2(ND2)	*MTND1P23*	MTND1P23
40	*ERCC-00171*	ERCC-00171	*IGHA1*	IGHA1
41	*RPLP0*	ribosomal protein lateral stalk subunit P0(RPLP0)	*LYZ*	lysozyme(LYZ)
42	*ALDOA*	aldolase, fructose-bisphosphate A(ALDOA)	*NEAT1*	nuclear paraspeckle assembly transcript 1 (non-protein coding)(NEAT1)
43	*LIMS1*	LIM zinc finger domain containing 1(LIMS1)	*CTTN*	cortactin(CTTN)
44	*ERCC-00003*	ERCC-00003	*MTND2P28*	MTND2P28
45	*ERCC-00043*	ERCC-00043	*LTF*	lactotransferrin(LTF)
46	*H2AFZ*	H2A histone family member Z(H2AFZ)	*HNRNPH1*	heterogeneous nuclear ribonucleoprotein H1 (H)(HNRNPH1)
47	*RPS6*	ribosomal protein S6(RPS6)	*ERCC-00130*	ERCC-00130
48	*RACK1*	receptor for activated C kinase 1(RACK1)	*HLA-E*	major histocompatibility complex, class I, E(HLA-E)
49	*HSPA8*	heat shock protein family A (Hsp70) member 8(HSPA8)	*STOM*	stomatin(STOM)
50	*ATP5B*	ATP synthase, mitochondrial F1 complex, betapolypeptide(ATP5B)	*CAMP*	cathelicidin antimicrobial peptide(CAMP)
51	*TAGLN2*	transgelin 2(TAGLN2)	*LIMS1*	LIM zinc finger domain containing 1(LIMS1)
52	*MIR6843*	microRNA 6843(MIR6843)	*DAAM1*	dishevelled associated activator of morphogenesis 1(DAAM1)
53	*RPL15*	ribosomal protein L15(RPL15)	*ITGA2B*	integrin subunit alpha 2b(ITGA2B)
54	*CD63*	CD63 molecule(CD63)	*MYL6*	myosin light chain 6(MYL6)
55	*PKM*	pyruvate kinase, muscle(PKM)	*ERCC-00046*	ERCC-00046
56	*ENO1*	enolase 1(ENO1)	*HBA2*	haemoglobin subunit alpha 2(HBA2)

57	*CLIC1*	chloride intracellular channel 1(CLIC1)	*PLEK*	pleckstrin(PLEK)
58	*RSU1*	Ras suppressor protein 1(RSU1)	*EEF1A1*	eukaryotic translation elongation factor 1 alpha 1(EEF1A1)
59	*RPL5*	ribosomal protein L5(RPL5)	*MYL12A*	myosin light chain 12A(MYL12A)
60	*LDHA*	lactate dehydrogenase A(LDHA)	*CAPZA1*	capping actin protein of muscle Z-line alpha subunit 1(CAPZA1)
61	*RPS3*	ribosomal protein S3(RPS3)	*HBD*	haemoglobin subunit delta(HBD)
62	*NPM1*	nucleophosmin(NPM1)	*SSR4*	signal sequence receptor subunit 4(SSR4)
63	*UBC*	ubiquitin C(UBC)	*MT-TF*	MT-TF
64	*EIF1*	eukaryotic translation initiation factor 1(EIF1)	*GNG11*	G protein subunit gamma 11(GNG11)
65	*RPL7*	ribosomal protein L7(RPL7)	*OAZ1*	ornithine decarboxylase antizyme 1(OAZ1)
66	*TALDO1*	transaldolase 1(TALDO1)	*CLC*	Charcot-Leyden crystal galectin(CLC)
67	*GATA1*	GATA binding protein 1(GATA1)	*IGLV2-14*	IGLV2-14
68	*ND1*	NADH dehydrogenase, subunit 1 (complex I)(ND1)	*PPP1CB*	protein phosphatase 1 catalytic subunit beta(PPP1CB)
69	*FDFT1*	farnesyl-diphosphate farnesyltransferase 1(FDFT1)	*IGLC3*	IGLC3
70	*MIR3064*	microRNA 3064(MIR3064)	*RPLP1*	ribosomal protein lateral stalk subunit P1(RPLP1)
71	*HMGB1*	high mobility group box 1(HMGB1)	*RAP1B*	RAP1B, member of RAS oncogene family(RAP1B)
72	*RPL10*	ribosomal protein L10(RPL10)	*S100A12*	S100 calcium binding protein A12(S100A12)
73	*LAPTM5*	lysosomal protein transmembrane 5(LAPTM5)	*TAGLN2*	transgelin 2(TAGLN2)
74	*RAP1B*	RAP1B, member of RAS oncogene family(RAP1B)	*RPS27A*	ribosomal protein S27a(RPS27A)
75	*RPL19*	ribosomal protein L19(RPL19)	*RPL9*	ribosomal protein L9(RPL9)
76	*ACAT2*	acetyl-CoA acetyltransferase 2(ACAT2)	*SAT1*	spermidine/spermine N1-acetyltransferase 1(SAT1)
77	*KPNA2*	karyopherin subunit alpha 2(KPNA2)	*SAR1A*	secretion associated Ras related GTPase 1A(SAR1A)
78	*CAP1*	adenylate cyclase associated protein 1(CAP1)	*MIR4738*	microRNA 6852(MIR6852)
79	*TUBA4A*	tubulin alpha 4a(TUBA4A)	*RPL21*	ribosomal protein L21(RPL21)
80	*STOM*	stomatin(STOM)	*RPS20*	ribosomal protein S20(RPS20)
81	*MIR3917*	microRNA 3917(MIR3917)	*RPS25*	ribosomal protein S25(RPS25)
82	*GPX1*	glutathione peroxidase 1(GPX1)	*CD36*	CD36 molecule(CD36)
83	*ERCC-00145*	ERCC-00145	*IGKC*	IGKC
84	*HSP90AB1*	heat shock protein 90 alpha family class B member 1(HSP90AB1)	*YWHAZ*	tyrosine 3-monooxygenase activation protein zeta(YWHAZ)
85	*HSP90AA1*	heat shock protein 90 alpha family class A member 1(HSP90AA1)	*HLA-B*	major histocompatibility complex, class I, B(HLA-B)
86	*ATP5A1*	ATP synthase, mitochondrial F1 complex, alpha subunit 1(ATP5A1)	*C6orf25*	chromosome 6 open reading frame 25(C6orf25)
87	*ERCC-00111*	ERCC-00111	*H3F3A*	H3 histone family member 3A(H3F3A)
88	*ERCC-00042*	ERCC-00042	*TUBA4A*	tubulin alpha 4a(TUBA4A)
89	*CD9*	CD9 molecule(CD9)	*FERMT3*	fermitin family member 3(FERMT3)
90	*TMBIM6*	transmembrane BAX inhibitor motif containing 6(TMBIM6)	*DEFA4*	defensin alpha 4(DEFA4)
91	*SRSF3*	serine and arginine rich splicing factor 3(SRSF3)	*HSP90AA1*	heat shock protein 90 alpha family class A member 1(HSP90AA1)
92	*RPS18*	ribosomal protein S18(RPS18)	*SRF3*	serine and arginine rich splicing factor 3(SRSF3)
93	*ERCC-00116*	ERCC-00116	*TPM4*	tropomyosin 4(TPM4)
94	*HNRNPA1*	heterogeneous nuclear ribonucleoprotein A1(HNRNPA1)	*CDC42*	cell division cycle 42(CDC42)
95	*HNRNPA2B1*	heterogeneous nuclear ribonucleoprotein A2/B1(HNRNPA2B1)	*ERCC-00004*	ERCC-00004
96	*RPS14*	ribosomal protein S14(RPS14)	*PTP4A2*	protein tyrosine phosphatase type IVA, member 2(PTP4A2)
97	*CALM2*	calmodulin 2(CALM2)	*SRSF6*	serine and arginine rich splicing factor 6(SRSF6)
98	*TPM3*	tropomyosin 3(TPM3)	*RPS13*	ribosomal protein S13(RPS13)
99	*ERCC-00092*	ERCC-00092	*MIR1244-1*	microRNA 1244-1(MIR1244-1)
100	*HMGN2*	high mobility group nucleosomal binding domain 2(HMGN2)	*RPL15*	ribosomal protein L15(RPL15)

Table 4.7 The 100 highly expressed transcripts in cultured and bone marrow megakaryocytes



Commonly expressed genes:
ACTB MIR4738 MIR1244-1 ATP6 TUBA4A COX2 LIMS1 ND2 STOM TMSB4X ND1 MT-RNR2 PPBP PLEK ACTG1 COX3 RAP1B COX1 B2M HBD MALAT1 ND5 ND4 TAGLN2 EEF1A1 MT-RNR1 HSP90AA1 FTL PF4 RPL15

Figure 4.13 Venn diagram to show commonly highly expressed genes between cultured megakaryocytes (cMK) and bone marrow megakaryocytes (bmMK)

Bone marrow megakaryocytes

Cultured megakaryocytes (derived from CD34+ cells)

Figure 4.14 Over-represented gene ontologies in most highly expressed megakaryocyte genes
The full GO analysis may be found in Appendix 4, Table 4.1, Table 4.2. The word cloud is a visual representation of the over-represented GO biological processes with their size an indicator of significance for most abundant genes in bone marrow megakaryocytes and cultured megakaryocytes derived from CD34+ cells.

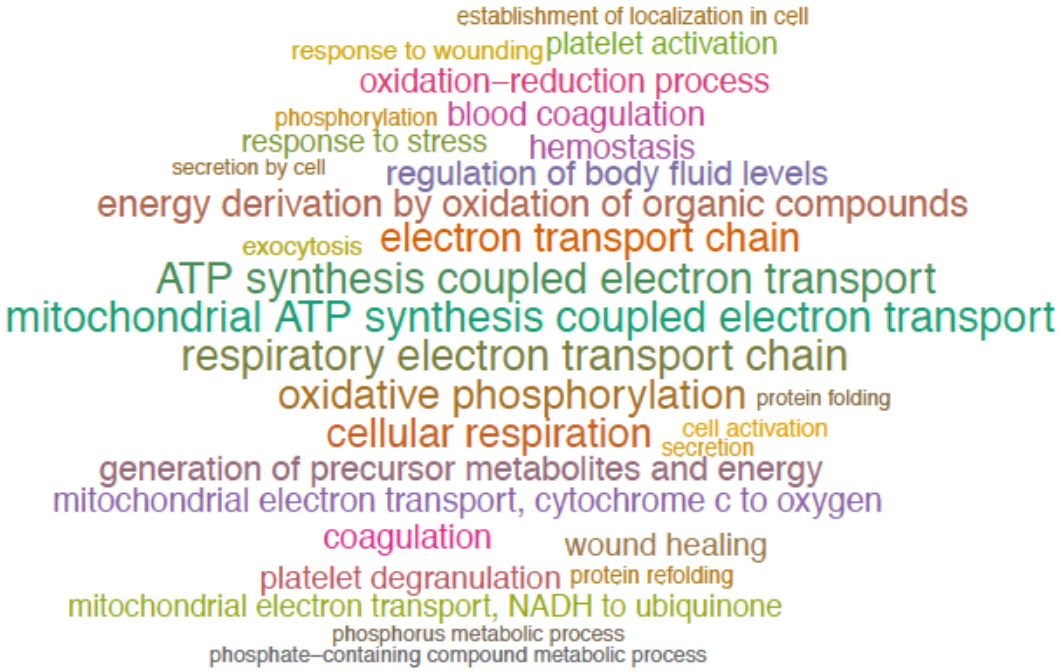


Figure 4.15 Over-represented gene ontologies common highly expressed genes between cultured megakaryocytes and bone marrow megakaryocytes

ENSEMBL_GENE_ID	HGNC_symbol	Gene_name	log2_FoldChange	p_adjusted	Gene_biotype	Function
ENSG00000211896	*IGHG1*	Immunoglobulin Heavy Constant Gamma 1 (G1m Marker)	-28.46272793	6.79E-39	Protein coding	Constant region of immunoglobulin heavy chains. Immunoglobulins, also known as antibodies, are membrane-bound or secreted glycoproteins produced by B lymphocytes. In the recognition phase of humoral immunity, the membrane-bound immunoglobulins serve as receptors which, upon binding of a specific antigen, trigger the clonal expansion and differentiation of B lymphocytes into immunoglobulins-secreting plasma cells. Secreted immunoglobulins mediate the effector phase of humoral immunity, which results in the elimination of bound antigens
ENSG00000211892	*IGHG4*	Immunoglobulin Heavy Constant Gamma 4 (G4m Marker)	-23.1315936	9.80E-15	Protein coding	Constant region of immunoglobulin heavy chains. Immunoglobulins, also known as antibodies, are membrane-bound or secreted glycoproteins produced by B lymphocytes. In the recognition phase of humoral immunity, the membrane-bound immunoglobulins serve as receptors which, upon binding of a specific antigen, trigger the clonal expansion and differentiation of B lymphocytes into immunoglobulins-secreting plasma cells. Secreted immunoglobulins mediate the effector phase of humoral immunity, which results in the elimination of bound antigens
ENSG00000143546	*S100A8*	S100 calcium binding protein A8(S100A8)	-14.38809424	9.80E-15	Protein coding	S100A8 is a calcium- and zinc-binding protein which plays a prominent role in the regulation of inflammatory processes and immune response.
ENSG00000210082	*MTRNR2*	Mitochondrially Encoded 16S RNA	-2.226873538	2.80E-14	RNA gene	Plays a role as a neuroprotective factor. Protects against death induced by multiple different familial Alzheimer disease genes and beta amyloid proteins in Alzheimer disease. Suppresses apoptosis by binding to BAX and preventing the translocation of BAX from the cytosol to mitochondria. Binds to IGFBP3 and specifically blocks IGFBP3-induced cell death Induces chemotaxis of mononuclear phagocytes via FPR2. Reduces the aggregation and fibrillary formation by suppressing the effect of APP on mononuclear phagocytes and acts by competitively inhibiting the access of FPRL1 to APP.
ENSG00000210049	*MT-TF*	Mitochondrially Encoded TRNA Phenylalanine	-9.43289167	7.33E-13	RNA gene	
ENSG00000163220	*S100A9*	S100 calcium binding protein A9(S100A9)	-11.97443679	3.19E-12	Protein coding	S100A9 is a calcium- and zinc-binding protein which plays a prominent role in the regulation of inflammatory processes and immune response.
ENSG00000198840	*ND3*	NADH dehydrogenase, subunit 3 (complex I)(ND3)	-4.059957895	4.96E-10	Protein coding	Core subunit of the mitochondrial membrane respiratory chain NADH dehydrogenase (Complex I) that is believed to belong to the minimal assembly required for catalysis. Complex I functions in the transfer of electrons from NADH to the respiratory chain. The immediate electron acceptor for the enzyme is believed to be ubiquinone (By similarity).
ENSG00000210194	*MT-TE*	Mitochondrially Encoded TRNA Glutamic Acid	-10.00891869	5.71E-09	RNA gene	
ENSG00000209082	*MT-TL1*	Mitochondrially Encoded TRNA Leucine 1 (UUA/G)	-8.151580487	1.07E-08	RNA gene	
ENSG00000239839	*DEFA3*	defensin alpha 3(DEFA3)	-12.46779432	9.03E-08	Protein coding	Defensin 2 and defensin 3 have antibiotic, fungicide and antiviral activities. Has antimicrobial activity against Gram-negative and Gram-positive bacteria. Defensins are thought to kill microbes by permeabilizing their plasma membrane. Defensins are a family of antimicrobial and cytotoxic peptides thought to be involved in host defense. They are abundant in the granules of neutrophils and also found in the epithelia of mucosal surfaces such as those of the intestine, respiratory tract, urinary tract, and vagina. Members of the defensin family are highly similar in protein sequence and distinguished by a conserved cysteine motif. The protein encoded by this gene, defensin, alpha 3, is found in the microbicidal granules of neutrophils and likely plays a role in phagocyte-mediated host defense. Several alpha defensin genes are clustered on chromosome 8.
ENSG00000210077	*MT-TV*	Mitochondrially Encoded TRNA Valine	-8.253682594	1.39E-07	RNA gene	
ENSG00000211459	*MTRNR1*	Mitochondrially Encoded 12S RNA	-2.843322805	1.39E-07	RNA gene	Regulates insulin sensitivity and metabolic homeostasis. Inhibits the folate cycle, thereby reducing de novo purine biosynthesis which leads to the accumulation of the de novo purine synthesis intermediate 5-aminoimidazole-4-carboxamide (AICAR) and the activation of the metabolic regulator 5-AMP-activated protein kinase (AMPK).
ENSG00000163221	*S100A12*	S100 calcium binding protein A12(S100A12)	-10.0211077	8.48E-07	Protein coding	S100A12 is a calcium-, zinc- and copper-binding protein which plays a prominent role in the regulation of inflammatory processes and immune response.
ENSG00000210112	*MT-TM*	Mitochondrially Encoded TRNA Methionine	-7.373255564	9.77E-07	RNA gene	

ENSG00000211592	*IGKC*	Immunoglobulin Kappa Constant	-11.02678863	1.20E-06	Protein coding	Constant region of immunoglobulin heavy chains. Immunoglobulins, also known as antibodies, are membrane-bound or secreted glycoproteins produced by B lymphocytes. In the recognition phase of humoral immunity, the membrane-bound immunoglobulins serve as receptors which, upon binding of a specific antigen, trigger the clonal expansion and differentiation of B lymphocytes into immunoglobulins-secreting plasma cells. Secreted immunoglobulins mediate the effector phase of humoral immunity, which results in the elimination of bound antigens
ENSG00000210164	*MT-TG*	Mitochondrially Encoded TRNA Glycine	-8.966751581	1.59E-05	RNA gene	
ENSG00000247627	*MTND4P12*	Mitochondrially Encoded NADH:Ubiquinone Oxidoreductase Core Subunit 4 Pseudogene 12	-4.71975486	1.87E-05	Pseudogene	
ENSG00000164821	*DEFA4*	defensin alpha 4(DEFA4)	-13.05753487	7.19E-05	Protein coding	Has antimicrobial activity against Gram-negative bacteria, and to a lesser extent also against Gram-positive bacteria and fungi. Defensins are a family of antimicrobial and cytotoxic peptides thought to be involved in host defense. They are abundant in the granules of neutrophils and also found in the epithelia of mucosal surfaces such as those of the intestine, respiratory tract, urinary tract, and vagina. Members of the defensin family are highly similar in protein sequence and distinguished by a conserved cysteine motif. Several alpha defensin genes are clustered on chromosome 8.
ENSG00000225972	*MTND1P23*	Mitochondrially Encoded NADH:Ubiquinone Oxidoreductase Core Subunit 1 Pseudogene 23	-6.736727971	1.09E-4	Pseudogene	
ENSG00000211893	*IGHG2*	Immunoglobulin Heavy Constant Gamma 2 (G2m Marker)	-11.5887655	1.62E-4	Protein coding	Constant region of immunoglobulin heavy chains. Immunoglobulins, also known as antibodies, are membrane-bound or secreted glycoproteins produced by B lymphocytes. In the recognition phase of humoral immunity, the membrane-bound immunoglobulins serve as receptors which, upon binding of a specific antigen, trigger the clonal expansion and differentiation of B lymphocytes into immunoglobulins-secreting plasma cells. Secreted immunoglobulins mediate the effector phase of humoral immunity, which results in the elimination of bound antigens
ENSG00000008438	*PGLYRP1*	peptidoglycan recognition protein 1(PGLYRP1)	-11.19335288	3.30E-4	Protein coding	Pattern receptor that binds to murein peptidoglycans (PGN) of Gram-positive bacteria. Has bactericidal activity towards Gram-positive bacteria. May kill Gram-positive bacteria by interfering with peptidoglycan biosynthesis. Binds also to Gram-negative bacteria, and has bacteriostatic activity towards Gram-negative bacteria. Plays a role in innate immunity.
ENSG00000210176	*MT-TH*	Mitochondrially Encoded TRNA Histidine	-7.666519016	5.48E-4	RNA gene	
ENSG00000269028	*MTRNR2L2*	MT-RNR2-like 2(MTRNR2L2)	-2.712601455	6.04E-4	Protein coding	Plays a role as a neuroprotective and antiapoptotic factor.
ENSG00000164047	*CAMP*	cathelicidin antimicrobial peptide(CAMP)	-8.308808438	7.03E-4	Protein coding	Binds to bacterial lipopolysaccharides (LPS), has antibacterial activity.
ENSG00000237973	*MTCO1P12*	Mitochondrially Encoded Cytochrome C Oxidase I Pseudogene 12	-2.956706078	7.65E-4	Pseudogene	
ENSG00000198888	*ND1*	NADH dehydrogenase, subunit 1 (complex I)(ND1)	-2.37505821	9.46E-4	Protein coding	Core subunit of the mitochondrial membrane respiratory chain NADH dehydrogenase (Complex I) that is believed to belong to the minimal assembly required for catalysis. Complex I functions in the transfer of electrons from NADH to the respiratory chain. The immediate electron acceptor for the enzyme is believed to be ubiquinone (By similarity).
ENSG00000005381	*MPO*	myeloperoxidase(MPO)	-11.87105356	1.02E-3	Protein coding	Part of the host defense system of polymorphonuclear leukocytes. It is responsible for microbicidal activity against a wide range of organisms. In the stimulated PMN, MPO catalyzes the production of hypohalous acids, primarily hypochlorous acid in physiologic situations, and other toxic intermediates that greatly enhance PMN microbicidal activity.
ENSG00000210100	*MT-TI*	Mitochondrially Encoded TRNA Isoleucine	-7.078475242	2.00E-3	RNA gene	

ENSG00000211677	*IGLC2*	Immunoglobulin Lambda Constant 2	-8.04273591	2.27E-3	Protein coding	Constant region of immunoglobulin heavy chains. Immunoglobulins, also known as antibodies, are membrane-bound or secreted glycoproteins produced by B lymphocytes. In the recognition phase of humoral immunity, the membrane-bound immunoglobulins serve as receptors which, upon binding of a specific antigen, trigger the clonal expansion and differentiation of B lymphocytes into immunoglobulins-secreting plasma cells. Secreted immunoglobulins mediate the effector phase of humoral immunity, which results in the elimination of bound antigens
ENSG00000105374	*NKG7*	natural killer cell granule protein 7(NKG7)	-10.20666474	2.85E-3	Protein coding	
ENSG00000210174	*MT-TR*	Mitochondrially Encoded TRNA Arginine	-5.665161981	3.88E-3	RNA gene	
ENSG00000104918	*RETN*	resistin(RETN)	-9.824865804	4.26E-3	Protein coding	Hormone that seems to suppress insulin ability to stimulate glucose uptake into adipose cells.
ENSG00000272398	*CD24*	CD24	-7.857743349	4.35E-3	Protein coding	This gene encodes a sialoglycoprotein that is expressed on mature granulocytes and B cells and modulates growth and differentiation signals to these cells. The precursor protein is cleaved to a short 32 amino acid mature peptide which is anchored via a glycosyl phosphatidylinositol (GPI) link to the cell surface. This gene was missing from previous genome assemblies, but is properly located on chromosome 6.Modulates B-cell activation responses. Signaling could be triggered by the binding of a lectin-like ligand to the CD24 carbohydrates, and transduced by the release of second messengers derived from the GPI-anchor. Promotes AG-dependent proliferation of B-cells, and prevents their terminal differentiation into antibody-forming cells.
ENSG00000169385	*RNASE2*	ribonuclease A family member 2(RNASE2)	-10.33236174	6.08E-3	Protein coding	This is a non-secretory ribonuclease. It is a pyrimidine specific nuclease with a slight preference for U. Cytotoxin and helminthotoxin. Selectively chemotactic for dendritic cells. Possesses a wide variety of biological activities.
ENSG00000211679	*IGLC3*	Immunoglobulin Lambda Constant 3 (Kern-Oz+ Marker)	-8.186319547	6.26E-3	Protein coding	Constant region of immunoglobulin heavy chains. Immunoglobulins, also known as antibodies, are membrane-bound or secreted glycoproteins produced by B lymphocytes. In the recognition phase of humoral immunity, the membrane-bound immunoglobulins serve as receptors which, upon binding of a specific antigen, trigger the clonal expansion and differentiation of B lymphocytes into immunoglobulins-secreting plasma cells. Secreted immunoglobulins mediate the effector phase of humoral immunity, which results in the elimination of bound antigens
ENSG00000197561	*ELANE*	elastase, neutrophil expressed(ELANE)	-8.647384526	8.48E-3	Protein coding	Modifies the functions of natural killer cells, monocytes and granulocytes. Inhibits C5a-dependent neutrophil enzyme release and chemotaxis.
ENSG00000051523	*CYBA*	cytochrome b-245 alpha chain(CYBA)	-5.44612098	9.10E-3	Protein coding	Critical component of the membrane-bound oxidase of phagocytes that generates superoxide. Associates with NOX3 to form a functional NADPH oxidase constitutively generating superoxide.
ENSG00000163993	*S100P*	S100 Calcium Binding Protein P	-7.979285986	9.51E-3	Protein coding	The protein encoded by this gene is a member of the S100 family of proteins containing 2 EF-hand calcium-binding motifs. S100 proteins are localized in the cytoplasm and/or nucleus of a wide range of cells, and involved in the regulation of a number of cellular processes such as cell cycle progression and differentiation. May function as calcium sensor and contribute to cellular calcium signaling.
ENSG00000211895	*IGHA1*	Immunoglobulin Heavy Constant Alpha 1	-7.048854891	1.00E-2	Protein coding	Constant region of immunoglobulin heavy chains. Immunoglobulins, also known as antibodies, are membrane-bound or secreted glycoproteins produced by B lymphocytes. In the recognition phase of humoral immunity, the membrane-bound immunoglobulins serve as receptors which, upon binding of a specific antigen, trigger the clonal expansion and differentiation of B lymphocytes into immunoglobulins-secreting plasma cells. Secreted immunoglobulins mediate the effector phase of humoral immunity, which results in the elimination of bound antigens
ENSG00000012223	*LTF*	lactotransferrin(LTF)	-6.666379141	1.25E-2	Protein coding	This gene is a member of the transferrin family of genes and its protein product is found in the secondary granules of neutrophils. The protein is a major iron-binding protein in milk and body secretions with an antimicrobial activity, making it an important component of the non-specific immune system. The protein demonstrates a broad spectrum of properties, including regulation of iron homeostasis, host defense against a broad range of microbial infections, anti-inflammatory activity, regulation of cellular growth and differentiation and protection against cancer development and metastasis.
ENSG00000085265	*FCN1*	ficolin 1(FCN1)	-6.436003316	1.71E-2	Protein coding	Extracellular lectin functioning as a pattern-recognition receptor in innate immunity. Binds the sugar moieties of pathogen-associated molecular patterns (PAMPs) displayed on microbes and activates the lectin pathway of the complement system.

ENSG00000163563	*MNDA*	myeloid cell nuclear differentiation antigen(MNDA)	-5.847511327	1.72E-2	Protein coding	May act as a transcriptional activator/repressor in the myeloid lineage. Plays a role in the granulocyte/monocyte cell-specific response to interferon. Stimulates the DNA binding of the transcriptional repressor protein YY1.
ENSG00000243466	*IGKV1-5*	Immunoglobulin Kappa Variable 1-5	-8.535460173	2.32E-2	Protein coding	V region of the variable domain of immunoglobulin light chains that participates in the antigen recognition. Immunoglobulins, also known as antibodies, are membrane-bound or secreted glycoproteins produced by B lymphocytes. In the recognition phase of humoral immunity, the membrane-bound immunoglobulins serve as receptors which, upon binding of a specific antigen, trigger the clonal expansion and differentiation of B lymphocytes into immunoglobulins-secreting plasma cells. Secreted immunoglobulins mediate the effector phase of humoral immunity, which results in the elimination of bound antigens
ENSG00000229314	*ORM1*	orosomucoid 1(ORM1)	-8.479300583	3.07E-2	Protein coding	Functions as transport protein in the blood stream. Binds various ligands in the interior of its beta-barrel domain. Also binds synthetic drugs and influences their distribution and availability in the body. Appears to function in modulating the activity of the immune system during the acute-phase reaction.
ENSG00000172232	*AZU1*	azurocidin 1(AZU1)	-7.056676321	3.43E-2	Protein coding	Azurophil granules, specialized lysosomes of the neutrophil, contain at least 10 proteins implicated in the killing of microorganisms. This gene encodes a preproprotein that is proteolytically processed to generate a mature azurophil granule antibiotic protein, with monocyte chemotactic and antimicrobial activity. It is also an important multifunctional inflammatory mediator. This encoded protein is a member of the serine protease gene family but it is not a serine proteinase, because the active site serine and histidine residues are replaced. The genes encoding this protein, neutrophil elastase 2, and proteinase 3 are in a cluster located at chromosome 19pter.
ENSG00000228253	*ATP8*	ATP synthase F0 subunit 8(ATP8)	-2.258116418	3.70E-2	Protein coding	Mitochondrial membrane ATP synthase (F(1)F(0) ATP synthase or Complex V) produces ATP from ADP in the presence of a proton gradient across the membrane which is generated by electron transport complexes of the respiratory chain. F-type ATPases consist of two structural domains, F(1) – containing the extramembraneous catalytic core and F(0) – containing the membrane proton channel, linked together by a central stalk and a peripheral stalk. During catalysis, ATP synthesis in the catalytic domain of F(1) is coupled via a rotary mechanism of the central stalk subunits to proton translocation. Part of the complex F(0) domain. Minor subunit located with subunit a in the membrane (By similarity).
ENSG00000132465	*JCHAIN*	joining chain of multimeric IgA and IgM(JCHAIN)	-5.751792155	4.29E-2	Protein coding	Serves to link two monomer units of either IgM or IgA. In the case of IgM, the J chain-joined dimer is a nucleating unit for the IgM pentamer, and in the case of IgA it induces larger polymers. It also help to bind these immunoglobulins to secretory component.
ENSG00000255823	*MTRNR2L8*	MT-RNR2-like 8(MTRNR2L8)	-2.215898828	4.60E-2	Protein coding	Plays a role as a neuroprotective and antiapoptotic factor.
ENSG00000096006	*CRISP3*	cysteine rich secretory protein 3(CRISP3)	-5.911851146	4.78E-2	Protein coding	This gene encodes a member of the cysteine-rich secretory protein (CRISP) family within the CRISP, antigen 5 and pathogenesis-related 1 proteins superfamily. The encoded protein has an N-terminal CRISP, antigen 5 and pathogenesis-related 1 proteins domain, a hinge region, and a C-terminal ion channel regulator domain. This protein contains cysteine residues, located in both the N- and C-terminal domains, that form eight bonds, a distinguishing characteristic of this family.
ENSG00000221869	*CEBPD*	CCAAT/enhancer binding protein delta(CEBPD)	-7.634229384	4.89E-2	Protein coding	The protein encoded by this intronless gene is a bZIP transcription factor which can bind as a homodimer to certain DNA regulatory regions. It can also form heterodimers with the related protein CEBP-alpha. The encoded protein is important in the regulation of genes involved in immune and inflammatory responses, and may be involved in the regulation of genes associated with activation and/or differentiation of macrophages.
ENSG00000211897	*IGHG3*	Immunoglobulin Heavy Constant Gamma 3 (G3m Marker)	-7.019552749	4.95E-2	Protein coding	Constant region of immunoglobulin heavy chains. Immunoglobulins, also known as antibodies, are membrane-bound or secreted glycoproteins produced by B lymphocytes. In the recognition phase of humoral immunity, the membrane-bound immunoglobulins serve as receptors which, upon binding of a specific antigen, trigger the clonal expansion and differentiation of B lymphocytes into immunoglobulins-secreting plasma cells. Secreted immunoglobulins mediate the effector phase of humoral immunity, which results in the elimination of bound antigens

Table 4.8 Transcripts upregulated in bone marrow megakaryocytes

ENSEMBL_GENE_ID	HGNC_symbol	Gene_name	log2_FoldChange	p_adjusted	Gene_biotype	Function
ENSG00000111716	*LDHB*	lactate dehydrogenase B(LDHB)	6.981085767	1.20E-4	Protein coding	This gene encodes the B subunit of lactate dehydrogenase enzyme, which catalyzes the interconversion of pyruvate and lactate with concomitant interconversion of NADH and NAD+ in a post-glycolysis process.
ENSG00000152234	*ATP5A1*	ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit 1, cardiac muscle(ATP5A1)	6.156129178	1.20E-4	Protein coding	Mitochondrial membrane ATP synthase (F(1)F(0) ATP synthase or Complex V) produces ATP from ADP in the presence of a proton gradient across the membrane which is generated by electron transport complexes of the respiratory chain.
ENSG00000079459	*FDFT1*	farnesyl-diphosphate farnesyltransferase 1(FDFT1)	8.852331677	1.29E-4	Protein coding	This gene encodes a membrane-associated enzyme located at a branch point in the mevalonate pathway.
ENSG00000148484	*RSU1*	Ras suppressor protein 1(RSU1)	7.039094583	2.77E-4	Protein coding	This gene encodes a protein that is involved in the Ras signal transduction pathway, growth inhibition, and nerve-growth factor induced differentiation processes, as determined in mouse and human cell line studies.
ENSG00000132967	*HMGB1P5*	High Mobility Group Box 1 Pseudogene 5	6.604671437	3.09E-4	Pseudogene	
ENSG00000142669	*SH3BGRL3*	SH3 domain binding glutamate rich protein like 3(SH3BGRL3)	6.281939497	4.10E-4	Protein coding	Could act as a modulator of glutaredoxin biological activity.
ENSG00000173812	*EIF1*	eukaryotic translation initiation factor 1(EIF1)	5.69146454	4.72E-4	Protein coding	Necessary for scanning and involved in initiation site selection. Promotes the assembly of 48S ribosomal complexes at the authentic initiation codon of a conventional capped mRNA.
ENSG00000099797	*TECR*	microRNA 639(MIR639)	7.177161877	6.26E-4	Protein coding	Catalyzes the last of the four reactions of the long-chain fatty acids elongation cycle. This endoplasmic reticulum-bound enzymatic process, allows the addition of 2 carbons to the chain of long- and very long-chain fatty acids/VLCFAs per cycle.
ENSG00000111669	*TPI1*	triosephosphate isomerase 1(TPI1)	7.282297504	1.09E-3	Protein coding	This gene encodes an enzyme, consisting of two identical proteins, which catalyzes the isomerization of glyceraldehydes 3-phosphate (G3P) and dihydroxy-acetone phosphate (DHAP) in glycolysis and gluconeogenesis.
ENSG00000132475	*H3F3B*	microRNA 4738(MIR4738)	5.747596478	1.13E-3	Protein coding	Variant histone H3 which replaces conventional H3 in a wide range of nucleosomes in active genes. Constitutes the predominant form of histone H3 in non-dividing cells and is incorporated into chromatin independently of DNA synthesis.
ENSG00000175567	*UCP2*	uncoupling protein 2(UCP2)	8.060929314	1.13E-3	Protein coding	UCP are mitochondrial transporter proteins that create proton leaks across the inner mitochondrial membrane, thus uncoupling oxidative phosphorylation from ATP synthesis. As a result, energy is dissipated in the form of heat.
ENSG00000142227	*EMP3*	epithelial membrane protein 3(EMP3)	6.431466011	1.37E-3	Protein coding	The protein encoded by this gene belongs to the PMP-22/EMP/MP20 family of proteins. The protein contains four transmembrane domains and two N-linked glycosylation sites. It is thought to be involved in cell proliferation, cell-cell interactions and function as a tumor suppressor.
ENSG00000115956	*PLEK*	pleckstrin(PLEK)	6.606771366	1.40E-3	Protein coding	Major protein kinase C substrate of platelets.
ENSG00000100316	*RPL3*	ribosomal protein L3(RPL3)	5.981957468	1.49E-3	Protein coding	The L3 protein is a component of the large subunit of cytoplasmic ribosomes.
ENSG00000150991	*UBC*	ubiquitin C(UBC)	6.024907408	1.49E-3	Protein coding	This gene represents a ubiquitin gene, ubiquitin C. The encoded protein is a polyubiquitin precursor. Conjugation of ubiquitin monomers or polymers can lead to various effects within a cell, depending on the residues to which ubiquitin is conjugated. Ubiquitination has been associated with protein degradation, DNA repair, cell cycle regulation, kinase modification, endocytosis, and regulation of other cell signaling pathways.
ENSG00000251562	*MALAT1*	metastasis associated lung adenocarcinoma transcript 1 (non-protein coding)(MALAT1)	5.074887685	1.59E-3	RNA gene	This gene produces a precursor transcript from which a long non-coding RNA is derived by Rnase P cleavage of a tRNA-like small ncRNA (known as mascRNA) from its 3' end. The resultant mature transcript lacks a canonical poly(A) tail but is instead stabilized by a 3' triple helical structure. This transcript is retained in the nucleus where it is thought to form molecular scaffolds for ribonucleoprotein complexes. It may act as a transcriptional regulator for numerous genes, including some genes involved in cancer metastasis and cell migration, and it is involved in cell cycle regulation.
ENSG00000111348	*ARHGDIB*	Rho GDP dissociation inhibitor beta(ARHGDIB)	5.924163811	1.82E-3	Protein coding	Regulates the GDP/GTP exchange reaction of the Rho proteins by inhibiting the dissociation of GDP from them, and the subsequent binding of GTP to them (PubMed:8356058, PubMed:7512369). Regulates reorganization of the actin cytoskeleton mediated by Rho family members (PubMed:8262133).
ENSG00000087086	*FTL*	ferritin light chain(FTL)	4.407116736	2.27E-3	Protein coding	Stores iron in a soluble, non-toxic, readily available form. Important for iron homeostasis. Iron is taken up in the ferrous form and deposited as ferric hydroxides after oxidation. Also plays a role in delivery of iron to cells. Mediates iron uptake in capsule cells of the developing kidney (By similarity).

ENSG00000167996	*FTH1*	ferritin heavy chain 1(FTH1)	4.996277272	2.65E-3	Protein coding	Stores iron in a soluble, non-toxic, readily available form. Important for iron homeostasis. Has ferroxidase activity. Iron is taken up in the ferrous form and deposited as ferric hydroxides after oxidation. Also plays a role in delivery of iron to cells. Mediates iron uptake in capsule cells of the developing kidney (By similarity).
ENSG00000149925	*ALDOA*	aldolase, fructose-bisphosphate A(ALDOA)	5.832348323	2.85E-3	Protein coding	Plays a key role in glycolysis and gluconeogenesis. In addition, may also function as scaffolding protein (By similarity).
ENSG00000075624	*ACTB*	actin beta(ACTB)	5.214051666	3.03E-3	Protein coding	This gene encodes one of six different actin proteins. Actins are highly conserved proteins that are involved in cell motility, structure, and integrity. This actin is a major constituent of the contractile apparatus and one of the two nonmuscle cytoskeletal actins.
ENSG00000091164	*TXNL1*	thioredoxin like 1(TXNL1)	7.167904918	3.07E-3	Protein coding	Active thioredoxin with a redox potential of about -250 mV.
ENSG00000149485	*FADS1*	microRNA 1908(MIR1908)	7.348958389	3.07E-3	Protein coding	The protein encoded by this gene is a member of the fatty acid desaturase (FADS) gene family. Desaturase enzymes regulate unsaturation of fatty acids through the introduction of double bonds between defined carbons of the fatty acyl chain. FADS family members are considered fusion products composed of an N-terminal cytochrome b5-like domain and a C-terminal multiple membrane-spanning desaturase portion, both of which are characterized by conserved histidine motifs.
ENSG00000140374	*ETFA*	electron transfer flavoprotein alpha subunit(ETFA)	5.742945282	3.17E-3	Protein coding	The electron transfer flavoprotein serves as a specific electron acceptor for several dehydrogenases, including five acyl-CoA dehydrogenases, glutaryl-CoA and sarcosine dehydrogenase. It transfers the electrons to the main mitochondrial respiratory chain via ETF-ubiquinone oxidoreductase (ETF dehydrogenase).
ENSG00000165629	*ATP5C1*	ATP synthase, H+ transporting, mitochondrial F1 complex, gamma polypeptide 1(ATP5C1)	6.528620171	3.23E-3	Protein coding	This gene encodes a subunit of mitochondrial ATP synthase. Mitochondrial ATP synthase catalyzes ATP synthesis, utilizing an electrochemical gradient of protons across the inner membrane during oxidative phosphorylation. ATP synthase is composed of two linked multi-subunit complexes: the soluble catalytic core, F1, and the membrane-spanning component, Fo, comprising the proton channel. The catalytic portion of mitochondrial ATP synthase consists of 5 different subunits (alpha, beta, gamma, delta, and epsilon) assembled with a stoichiometry of 3 alpha, 3 beta, and a single representative of the other 3.
ENSG00000130816	*DNMT1*	DNA methyltransferase 1(DNMT1)	5.22759368	3.36E-3	Protein coding	Protein arginine methyltransferases are enyzmes that catalyze the transfer of methyl groups from S-adenosylmethionine (SAM) to the arginine residues on histones and other proteins.
ENSG00000059377	*TBXAS1*	thromboxane A synthase 1(TBXAS1)	7.462968009	3.88E-3	Protein coding	This gene encodes a member of the cytochrome P450 superfamily of enzymes. The cytochrome P450 proteins are monooxygenases which catalyze many reactions involved in drug metabolism and synthesis of cholesterol, steroids and other lipids. However, this protein is considered a member of the cytochrome P450 superfamily on the basis of sequence similarity rather than functional similarity. This endoplasmic reticulum membrane protein catalyzes the conversion of prostglandin H2 to thromboxane A2, a potent vasoconstrictor and inducer of platelet aggregation.
ENSG00000127314	*RAP1B*	RAP1B, member of RAS oncogene family(RAP1B)	4.349835715	4.01E-3	Protein coding	GTP-binding protein that possesses intrinsic GTPase activity. Contributes to the polarizing activity of KRIT1 and CDH5 in the establishment and maintenance of correct endothelial cell polarity and vascular lumen. Required for the localization of phosphorylated PRKCZ, PARD3 and TIAM1 to the cell junction. Plays a role in the establishment of basal endothelial barrier function.
ENSG00000054118	*THRAP3*	thyroid hormone receptor associated protein 3(THRAP3)	5.823109092	4.35E-3	Protein coding	Involved in pre-mRNA splicing. Remains associated with spliced mRNA after splicing which probably involves interactions with the exon junction complex (EJC). Can trigger mRNA decay which seems to be independent of nonsense-mediated decay involving premature stop codons (PTC) recognition. May be involved in nuclear mRNA decay. Involved in regulation of signal-induced alternative splicing.
ENSG00000143226	*FCGR2A*	Fc fragment of IgG receptor IIa(FCGR2A)	5.306413589	4.51E-3	Protein coding	Binds to the Fc region of immunoglobulins gamma. Low affinity receptor. By binding to IgG it initiates cellular responses against pathogens and soluble antigens. Promotes phagocytosis of opsonized antigens.
ENSG00000198242	*RPL23A*	ribosomal protein L23a(RPL23A)	4.76496201	4.93E-3	Protein coding	This protein binds to a specific region on the 26S rRNA.
ENSG00000110934	*BIN2*	bridging integrator 2(BIN2)	5.503489991	5.06E-3	Protein coding	Promotes cell motility and migration, probably via its interaction with the cell membrane and with podosome proteins that mediate interaction with the cytoskeleton. Modulates membrane curvature and mediates membrane tubulation. Plays a role in podosome formation. Inhibits phagocytosis.
ENSG00000100353	*EIF3D*	eukaryotic translation initiation factor 3 subunit D(EIF3D)	6.955622621	5.68E-3	Protein coding	Eukaryotic translation initiation factor-3 (eIF3), the largest of the eIFs, is a multiprotein complex composed of at least ten nonidentical subunits. The complex binds to the 40S ribosome and helps maintain the 40S and 60S ribosomal subunits in a dissociated state. It is also thought to play a role in the formation of the 40S initiation complex by interacting with the ternary complex of eIF2/GTP/methionyl-tRNA, and by promoting mRNA binding.

ENSG00000160752	*FDPS*	farnesyl diphosphate synthase(FDPS)	8.531009093	5.68E-3	Protein coding	Key enzyme in isoprenoid biosynthesis which catalyzes the formation of farnesyl diphosphate (FPP), a precursor for several classes of essential metabolites including sterols, dolichols, carotenoids, and ubiquinones. FPP also serves as substrate for protein farnesylation and geranylgeranylation. Catalyzes the sequential condensation of isopentenyl pyrophosphate with the allylic pyrophosphates, dimethylallyl pyrophosphate, and then with the resultant geranylpyrophosphate to the ultimate product farnesyl pyrophosphate.
ENSG00000161203	*AP2M1*	adaptor related protein complex 2 mu 1 subunit(AP2M1)	6.66057589	5.68E-3	Protein coding	This gene encodes a subunit of the heterotetrameric coat assembly protein complex 2 (AP2), which belongs to the adaptor complexes medium subunits family. The encoded protein is required for the activity of a vacuolar ATPase, which is responsible for proton pumping occurring in the acidification of endosomes and lysosomes.
ENSG00000235655	*H3F3AP4*	H3 Histone, Family 3A, Pseudogene 4	4.769962652	5.68E-3	Pseudogene	
ENSG00000110955	*ATP5B*	ATP synthase, H+ transporting, mitochondrial F1 complex, beta polypeptide(ATP5B)	6.880049327	6.01E-3	Protein coding	This gene encodes a subunit of mitochondrial ATP synthase. Mitochondrial ATP synthase catalyzes ATP synthesis, utilizing an electrochemical gradient of protons across the inner membrane during oxidative phosphorylation. ATP synthase is composed of two linked multi-subunit complexes: the soluble catalytic core, F1, and the membrane-spanning component, Fo, comprising the proton channel. The catalytic portion of mitochondrial ATP synthase consists of 5 different subunits (alpha, beta, gamma, delta, and epsilon) assembled with a stoichiometry of 3 alpha, 3 beta, and a single representative of the other 3. The proton channel consists of three main subunits (a, b, c).
ENSG00000198898	*CAPZA2*	capping actin protein of muscle Z-line alpha subunit 2(CAPZA2)	5.249095992	7.03E-3	Protein coding	The protein encoded by this gene is a member of the F-actin capping protein alpha subunit family. It is the alpha subunit of the barbed-end actin binding protein Cap Z. By capping the barbed end of actin filaments, Cap Z regulates the growth of the actin filaments at the barbed end.
ENSG00000108671	*PSMD11*	proteasome 26S subunit, non-ATPase 11(PSMD11)	5.570801143	7.33E-3	Protein coding	Component of the 26S proteasome, a multiprotein complex involved in the ATP-dependent degradation of ubiquitinated proteins. This complex plays a key role in the maintenance of protein homeostasis by removing misfolded or damaged proteins, which could impair cellular functions, and by removing proteins whose functions are no longer required. Therefore, the proteasome participates in numerous cellular processes, including cell cycle progression, apoptosis, or DNA damage repair.
ENSG00000149357	*LAMTOR1*	late endosomal/lysosomal adaptor, MAPK and MTOR activator 1(LAMTOR1)	5.949555526	7.70E-3	Protein coding	As part of the Ragulator complex it is involved in amino acid sensing and activation of mTORC1, a signaling complex promoting cell growth in response to growth factors, energy levels, and amino acids. Activated by amino acids through a mechanism involving the lysosomal V-ATPase, the Ragulator functions as a guanine nucleotide exchange factor activating the small GTPases Rag. Activated Ragulator and Rag GTPases function as a scaffold recruiting mTORC1 to lysosomes where it is in turn activated. LAMTOR1 is directly responsible for anchoring the Ragulator complex to membranes. Also required for late endosomes/lysosomes biogenesis it may regulate both the recycling of receptors through endosomes and the MAPK signaling pathway through recruitment of some of its components to late endosomes.
ENSG00000135486	*HNRNPA1*	heterogeneous nuclear ribonucleoprotein A1(HNRNPA1)	5.436961839	7.90E-3	Protein coding	Involved in the packaging of pre-mRNA into hnRNP particles, transport of poly(A) mRNA from the nucleus to the cytoplasm and may modulate splice site selection.
ENSG00000116288	*PARK7*	Parkinsonism associated deglycase(PARK7)	5.571874119	7.94E-3	Protein coding	The product of this gene belongs to the peptidase C56 family of proteins. It acts as a positive regulator of androgen receptor-dependent transcription. It may also function as a redox-sensitive chaperone, as a sensor for oxidative stress, and it apparently protects neurons against oxidative stress and cell death.
ENSG00000254999	*BRK1*	BRICK1, SCAR/WAVE actin nucleating complex subunit(BRK1)	6.163344524	8.01E-3	Protein coding	Involved in regulation of actin and microtubule organization. Part of a WAVE complex that activates the Arp2/3 complex. As component of the WAVE1 complex, required for BDNF-NTRK2 endocytic trafficking and signaling from early endosomes (By similarity).
ENSG00000168003	*SLC3A2*	solute carrier family 3 member 2(SLC3A2)	6.94514985	8.25E-3	Protein coding	This gene is a member of the solute carrier family and encodes a cell surface, transmembrane protein. The protein exists as the heavy chain of a heterodimer, covalently bound through di-sulfide bonds to one of several possible light chains. The encoded transporter plays a role in regulation of intracellular calcium levels and transports L-type amino acids.
ENSG00000131876	*SNRPA1*	small nuclear ribonucleoprotein polypeptide A'(SNRPA1)	5.686022399	8.44E-3	Protein coding	This protein is associated with sn-RNP U2. It helps the A protein to bind stem loop IV of U2 snRNA.

ENSG00000204628	*RACK1*	receptor for activated C kinase 1(RACK1)	5.573898576	8.44E-3	Protein coding	Involved in the recruitment, assembly and/or regulation of a variety of signaling molecules. Interacts with a wide variety of proteins and plays a role in many cellular processes. Component of the 40S ribosomal subunit involved in translational repression
ENSG00000233276	*GPX1*	glutathione peroxidase 1(GPX1)	5.092046956	8.44E-3	Protein coding	Protects the haemoglobin in erythrocytes from oxidative breakdown.
ENSG00000085662	*AKR1B1*	aldo-keto reductase family 1 member B(AKR1B1)	7.799142359	8.48E-3	Protein coding	Catalyzes the NADPH-dependent reduction of a wide variety of carbonyl-containing compounds to their corresponding alcohols with a broad range of catalytic efficiencies.
ENSG00000174748	*RPL15*	ribosomal protein L15(RPL15)	4.599841155	8.68E-3	Protein coding	This gene encodes a member of the L15E family of ribosomal proteins and a component of the 60S subunit.
ENSG00000111640	*GAPDH*	glyceraldehyde-3-phosphate dehydrogenase(GAPDH)	5.738486429	8.935E-3	Protein coding	Has both glyceraldehyde-3-phosphate dehydrogenase and nitrosylase activities, thereby playing a role in glycolysis and nuclear functions, respectively. Participates in nuclear events including transcription, RNA transport, DNA replication and apoptosis. Nuclear functions are probably due to the nitrosylase activity that mediates cysteine S-nitrosylation of nuclear target proteins such as SIRT1, HDAC2 and PRKDC. Modulates the organization and assembly of the cytoskeleton.

Table 4.9 Top 50 transcripts upregulated in cultured megakaryocytes

Upregulated in bone marrow megakaryocytes

Upregulated in cultured megakaryocytes (derived from CD34+cells)

Figure 4.16 Over-represented gene ontologies in upregulated genes in bone marrow megakaryocytes and in cultured megakaryocytes
The full GO analysis may be found in Appendix 4, Table 4.5, Table 4.6. The word cloud is a visual representation of the over-represented GO biological processes with their size an indicator of significance for upregulated genes in bone marrow megakaryocytes and cultured megakaryocytes derived from CD34+ cell

4.2.3 Ploidy associated megakaryocyte transcriptional signatures

4.2.3.1 Sample preparation and RNA-seq library preparation

In order to investigate changes in gene expression associated with endomitotic replication and subsequent increase in megakaryocyte ploidy level, bone marrow megakaryocytes were FACS sorted for RNA-seq according to ploidy level. The majority of cells collected were as single cell. As discussed above the low quality of single cell megakaryocyte sequencing compared to multicellular pools as a result of oversaturation by mitochondrial transcripts meant that after an initial round of sequencing, no further single cells were amplified or sequenced. Megakaryocytes from 10 individuals undergoing non-coronary cardiac surgery were collected as single cells and megakaryocyte pools according to ploidy level from 2N to 32N. Table 4.10 shows the baseline characteristics for these individuals.

32 megakaryocyte pools (20-50 cells) FACS sorted according to ploidy were collected from 5 individuals from the control group: 2N: n=3, 4N: n=6, 8N: n=7, 16N: n=8, 32N: n=8. 125bp paired-end sequencing was performed at a depth of approximately 500 million reads/sample. In total 1106 single megakaryocytes had also been sorted from a further 5 individuals from the control group, 2N: n=220, 4N: n=158, 8N: n=200, 16N: n=272, 32N: n=256. Here again 125bp paired-end sequencing was performed but at an approximate depth of 50 million reads/sample.

In terms of batch effects Table 4.11 shows each sample plate ID and corresponding sequencing lane ID whereas Table 4.12 outlines the batches in which samples were collected, cDNA reverse transcribed/amplified and Nextera libraries prepared and sequenced.

Table 4.10 Baseline patient characteristics

	Healthy controls (n=10)
Gender M	6
Age (mean ± SD)	64.7 ± 14.3
Hypertension	4
Hypercholesterolaemia	4
Diabetes	2
Smoking	0
Coronary disease	0
Antiplatelets	2
Left ventricular dysfunction	3
Aortic valve surgery	7
Mitral valve surgery	4
Hb (mean ± SD)	13.12 ± 2.54
WC (mean ± SD)	7.12 ± 1.90
Plt (mean ± SD)	240 ± 45.3

Plate ID	Sequencing Lane ID
60Z 16N	SLX-9491
60Z 32N	SLX-9492
60Z 8N	SLX-9493
60Z 4N/2N	SLX-9494
7TA 16N	SLX-9495
7TA 32N	SLX-9496
CYH 16N	SLX-9490
CYH 32N	SLX-9497
CYH 8N	SLX-9488
CYH 4N	SLX-9489
CYH 2N	SLX-10160
CUP 8N	SLX-10154
CUP 16N	SLX-10155
CUP 32N	SLX-10156
CUP 4N/2N	SLX-10157
CST 2N	SLX-10158
M8T	SLX-10482
H4K	SLX-10208, SLX-11606
1FP	SLX-11607
0DX	SLX-11608
3XI	SLX-11609

Table 4.11 Patient and plate ID on each sequencing lane

Sample collection	Amplification	Nextera Library prep / RNA-seq
1	1	1
0DX	60Z 16N	SLX-9491
2	60Z 32N	SLX-9492
1FP	60Z 8N	SLX-9493
3	60Z 4N/2N	SLX-9494
3XI	2	SLX-9495
4	7TA 16N	SLX-9496
60Z 16N	7TA 32N	2
60Z 32N	3	SLX-9490
60Z 8N	CYH 16N	SLX-9497
60Z 4N/2N	CYH 32N	SLX-9488
5	CYH 8N	SLX-9489
7TA 16N	CYH 4N	SLX-10160
7TA 32N	CYH 2N	3
7TA 8N	4	SLX-10154
6	CUP 8N	SLX-10155
CUP 8N	CUP 16N	SLX-10156
CUP 16N	CUP 32N	SLX-10157
CUP 32N	CUP 4N/2N	SLX-10158
CUP 4N/2N	CST 2N	4
7	5	SLX-10208
CST 2N	H4K	5
8	6	SLX-10482
CYH 16N	M8T	6
CYH 32N	7	SLX-11606
CYH 8N	0DX	SLX-11607
CYH 4N	8	SLX-11608
CYH 2N	3XI	SLX-11609
9	1FP	
H4K		
10		
M8T		

Table 4.12 Megakaryocyte collection, amplification, library preparation/RNA-seq batches

4.2.3.2 RNA sequencing data: Quality measures

1. Quality control

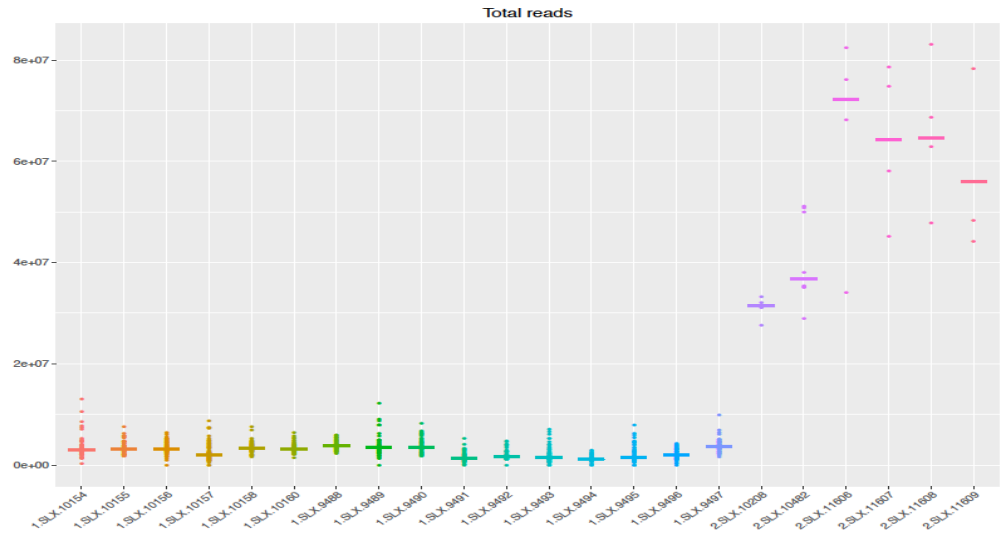
Figure 4.17 shows metrics per sequencing lane for total read count, alignment to the reference transcriptome (GRCh37), proportion of reads mapped to external spike-in ERCCs compared to endogenous transcripts, no of genes called, proportion of reads mapping to annotated exonic regions and the proportion of reads mapping to mitochondrial transcripts.

The single cells have lower total read counts more variable alignment rates to the reference transcriptome, higher ERCC to mapped transcript ratio and lower number of genes called in comparison with the multicellular megakaryocyte pools and therefore have overall lower quality. The exon/(exon+intron) ratio is variable throughout and there is a high proportion of reads taken up by mitochondrial reads in all sequencing lanes.

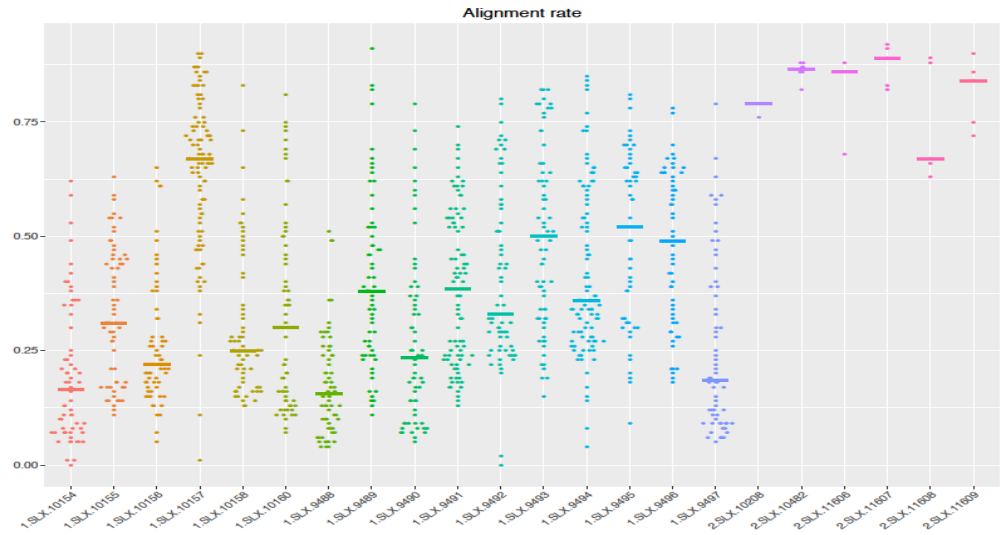
2. Cell filtering

In order to filter out higher quality single cells for downstream gene expression analysis the machine learning approach (random forest) described in Section 2.6.2 was used. Good quality cells were those that had low ERCC ratios and mitochondrial mapping, higher number of reads mapping to exonic regions and genes called. Figure 4.18 shows the cells deemed to be of low quality were taken out of further analysis. From a total of 1106 cells sequenced, 282 were taken forward for differential expression analysis.

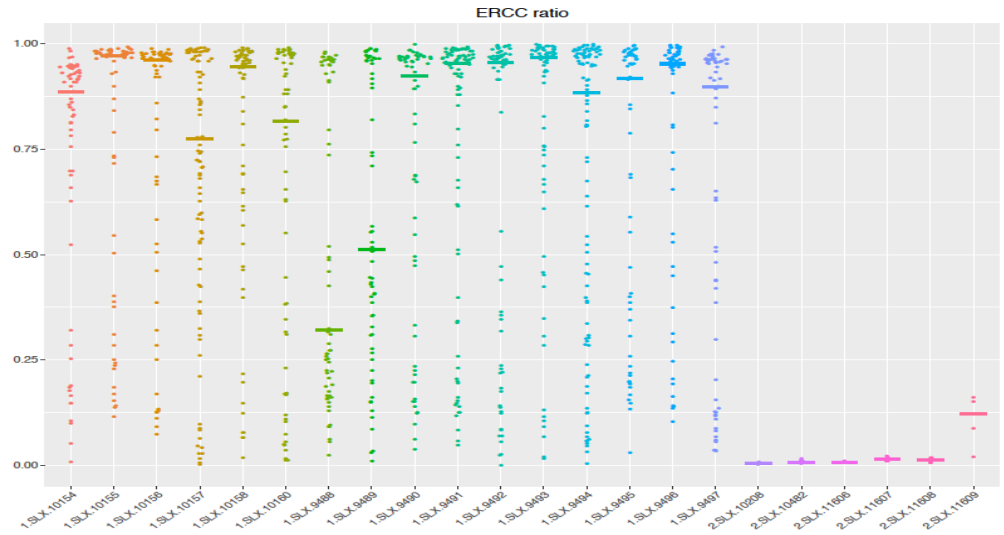
A


Total reads
8e+07
6e+07
4e+07
2e+07
0e+00
1.SLX.10154
1.SLX.10155
1.SLX.10156
1.SLX.10157
1.SLX.10158
1.SLX.10160
1.SLX.9488
1.SLX.9489
1.SLX.9490
1.SLX.9491
1.SLX.9492
1.SLX.9493
1.SLX.9494
1.SLX.9495
1.SLX.9496
1.SLX.9497
2.SLX.10208
2.SLX.10482
2.SLX.11606
2.SLX.11607
2.SLX.11608
2.SLX.11609

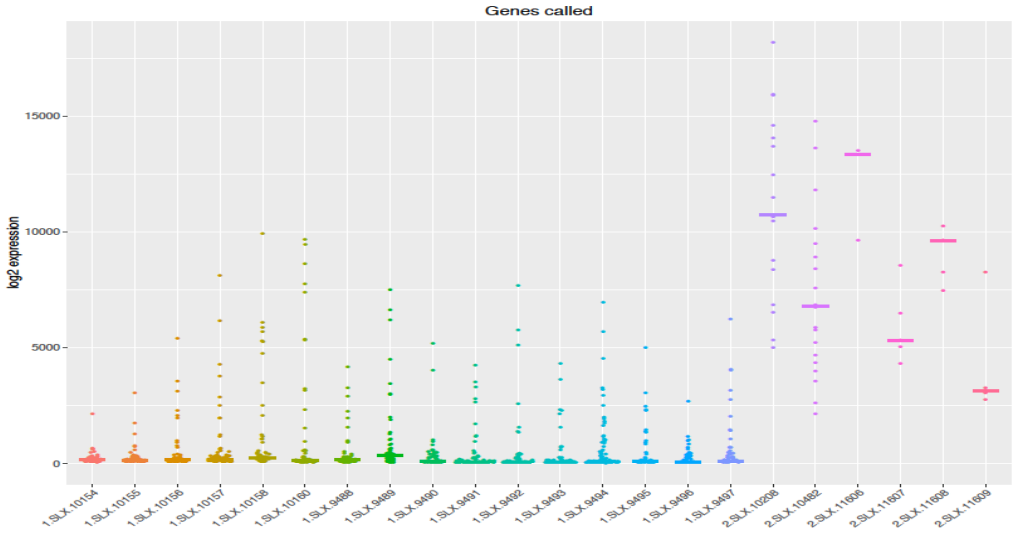
B


Alignment rate
0.75
0.50
0.25
0.00
1.SLX.10154
1.SLX.10155
1.SLX.10156
1.SLX.10157
1.SLX.10158
1.SLX.10160
1.SLX.9488
1.SLX.9489
1.SLX.9490
1.SLX.9491
1.SLX.9492
1.SLX.9493
1.SLX.9494
1.SLX.9495
1.SLX.9496
1.SLX.9497
2.SLX.10208
2.SLX.10482
2.SLX.11606
2.SLX.11607
2.SLX.11608
2.SLX.11609

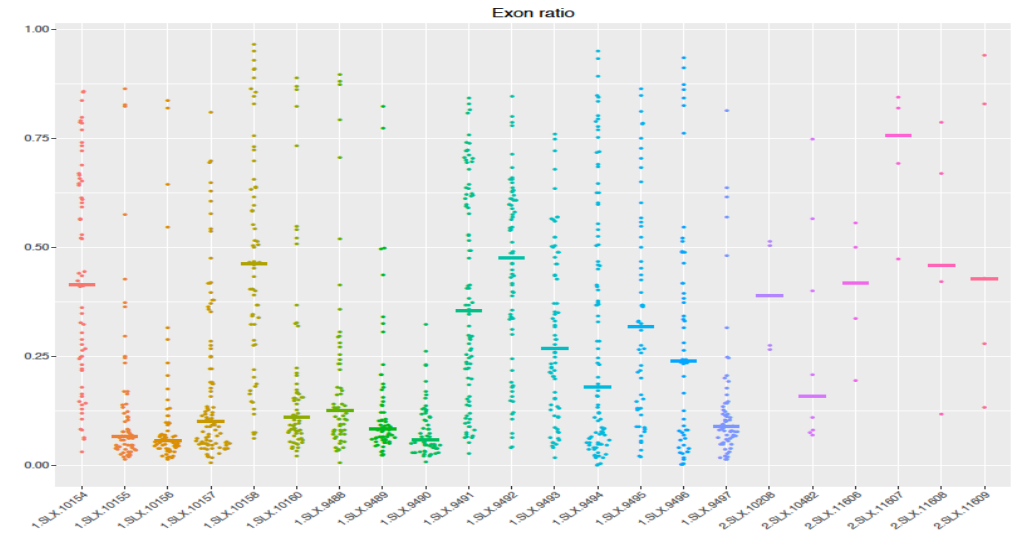
C


ERCC ratio
1.00
0.75
0.50
0.25
0.00
1.SLX.10154
1.SLX.10155
1.SLX.10156
1.SLX.10157
1.SLX.10158
1.SLX.10160
1.SLX.9488
1.SLX.9489
1.SLX.9490
1.SLX.9491
1.SLX.9492
1.SLX.9493
1.SLX.9494
1.SLX.9495
1.SLX.9496
1.SLX.9497
2.SLX.10208
2.SLX.10482
2.SLX.11606
2.SLX.11607
2.SLX.11608
2.SLX.11609

D



E



F

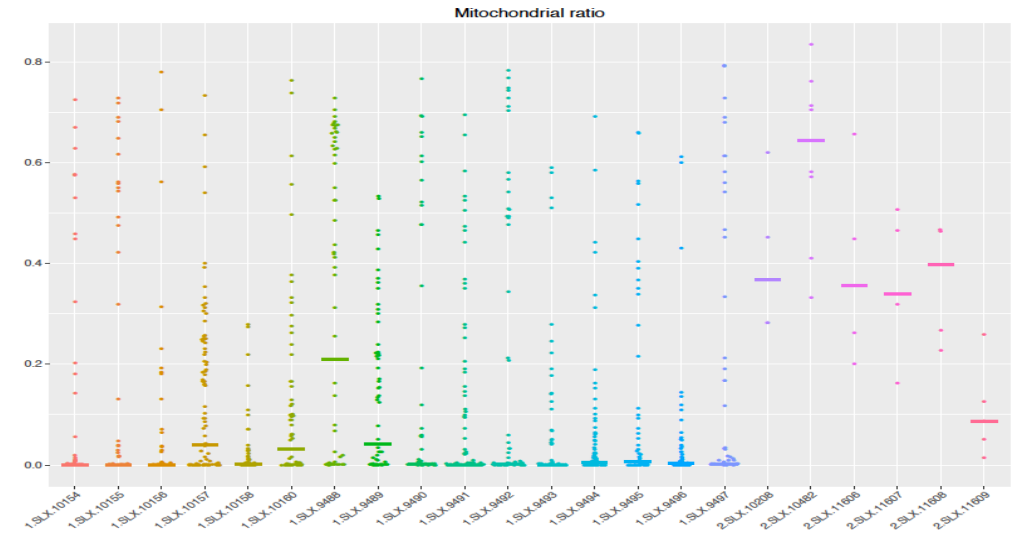


Figure 4.17 Quality control metrics/sequencing lane. A. Total reads; B. Rate of alignment to transcriptome; C. Ratio of reads mapping to ERCCs; D. No. genes called; E. Ratio of exonic reads to exonic+intronic; F. Ratio of reads mapping to mitochondrial genes.

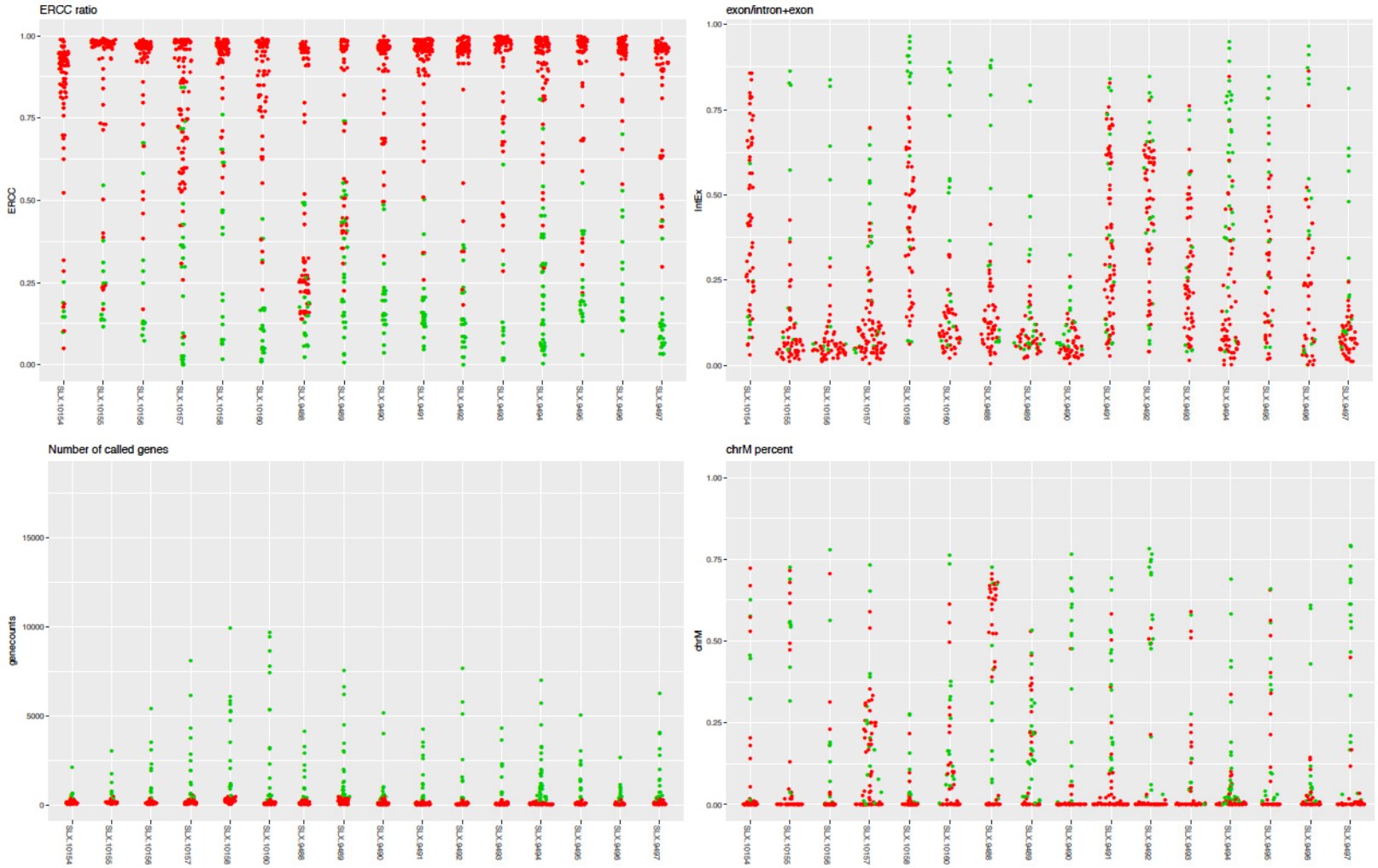


Figure 4.18 Filtering single cells.
High and low-quality cells shown by ERCC ratio, exonic ratio, number of called genes and mitochondrial mapping after cell filtering using a random forest classifier (Breiman2001). Green denotes high quality, red denotes low quality.

4.2.3.3 Gene expression analysis: Megakaryocyte pools

Gene expression analysis was performed on megakaryocyte pools before single cells due to their higher sequencing quality (4.2.3.2). The 100 most highly expressed genes in each ploidy group: 2N, 4N, 8N, 16N, 32N are shown in Table 4.13. In all groups, there was a high expression of transcripts mapping to mitochondrial transcripts and a number of genes involved in cellular metabolism as seen in the megakaryocyte pools sorted irrespective of ploidy level in section 4.2.2 above. Genes that are known to be involved in platelet function are highlighted orange in Table 4.13. Platelet transcripts highly expressed in 2N cells these included *PPBP* (growth factor released by platelets), *PLEK* (a major protein kinase C substrate in platelets), *TMSB4X* and *ACTB* (important in cytoskeletal change), *RAP1B* (located in the plasma membrane and regulator of outside-in signalling in platelets), *SRGN* (Granule proteoglycan released on platelet activation), *CD47* (thrombospondin adhesion receptor on platelets). In addition to these platelet genes, the following were also highly expressed in 4N and 8N cells: *ITGA2B* (CD41), *PF4* (chemokine released from platelet alpha granules), *CD36* (platelet glycoprotein that binds collagen and thrombospondin on platelet activation), *F13A1* (Coagulation factor XIII), *ACTG1* (important in cytoskeletal change). Other genes expressed in 4N and 8N cells included those involved in immune defense and translation. By contrast, in 16N and 32N cells, far fewer platelet transcripts are present in the most highly expressed genes, only *TMSB4X, SGRN, ACTB.* There are highly levels of expression of genes involved in cellular metabolism, translation and protein transport and localisation (Table 4.13). All ploidy groups show high expression of the housekeeping gene *B2M* and a number of immunoglobulin regions.

Shared expression of highly expressed genes in each ploidy group shown in Venn diagram format (Figure 4.20a) demonstrated that there were number of shared features between 2N, 4N and 8N cells, and separate shared features between 16N and 32N with fewer genes shared between the other pairs. Overall all ploidy groups shared 28 highly expressed genes involved in cellular metabolism and respiration. A similar pattern was seen when assessing shared GO categories (Figure 4.20b) with an overlap of 20 GO categories between all ploidy groups pertaining to cellular metabolism, mitochondrial transcripts. The shared GO categories are listed in Table 4.14. While 4N and 8N shared 15 additional GO categories related to platelet activation, cell secretion and defense response only 6 additional categories were shared between 2N, 4N and 8N. 16N and 32N shared 40 additional GO categories all related to processes involved in translation and protein localisation. There were very few shared GO categories between other pairs. The transition in biological processes based on the 100 most highly expressed genes in each ploidy group can be seen in the individual GO word clouds in Figures 4.21, 4.22 and the full GO analysis may be found in Appendix 4, Tables S4.7, S4.8, S4.9, S4.10, S4.11.

No	2N	4N	8N	16N	32N
1	*MT-RNR2*	*MT-RNR2*	*MT-RNR2*	*MT-RNR2*	*MT-RNR2*
2	*HBB*	*COX1*	*COX1*	*COX1*	*COX1*
3	*COX2*	*HBB*	*MT-RNR1*	*HBB*	*MT-RNR1*
4	*ND4*	*ND4*	*ND4*	*MT-RNR1*	*ND4*
5	*COX3*	*MT-RNR1*	*COX2*	*ND4*	*ERCC-00074*
6	*COX1*	*COX2*	*CYTB*	*COX2*	*COX2*
7	*CYTB*	*CYTB*	*COX3*	*COX3*	*CYTB*
8	*MT-RNR1*	*COX3*	*ATP6*	*CYTB*	*COX3*
9	*ATP6*	*ND2*	*ND2*	*ATP6*	*ND1*
10	*ERCC-00074*	*ATP6*	*ND1*	*ND1*	*ATP6*
11	*IGHG2*	*ND1*	*ERCC-00074*	*ND5*	*ND2*
12	*ND1*	*ND5*	*ND5*	*ND2*	*HBB*
13	*MALAT1*	*MALAT1*	*HBB*	*ERCC-00074*	*ND5*
14	*IGHG1*	*ERCC-00074*	*ERCC-00046*	*ERCC-00046*	*ERCC-00046*
15	*ND2*	*HBA2*	*ND4L*	*S100A9*	*S100A8*
16	*ND5*	*IGHG1*	*ND3*	*S100A8*	*S100A9*
17	*ERCC-00046*	*ND4L*	*IGHG1*	*HBD*	*CLC*
18	*H2AFZ*	*PPBP*	*ERCC-00096*	*ND4L*	*PRG3*
19	*B2M*	*NEAT1*	*ERCC-00130*	*HBA2*	*ERCC-00004*
20	*HBA*	*ND3*	*MALAT1*	*ERCC-00004*	*ND4L*
21	*HSP90AA1*	*PF4*	*MTATP6P1*	*ND3*	*ERCC-00002*
22	*TMSB4X*	*TMSB4X*	*ERCC-00002*	*ERCC-00002*	*ERCC-00130*
23	*ITM2B*	*ERCC-00046*	*ND6*	*ERCC-00130*	*B2M*
24	*HMGB1*	*IGHA1*	*ERCC-00004*	*MTATP6P1*	*ND3*
25	*SNRPE*	*MTATP6P2*	*S100A8*	*ERCC-00096*	*IGHG1*
26	*DEFA3*	*HBD*	*ERCC-00042*	*ND6*	*HBD*
27	*IGHA1*	*DEFA3*	*S100A9*	*PNRC2*	*RNASE2*
28	*SRSF3*	*B2M*	*IGLC2*	*TNPO1*	*HBA2*
29	*MIR1244-1*	*S100A8*	*PPBP*	*S100A12*	*ERCC-00096*
30	*S100A9*	*ND6*	*ASNSD1*	*MTCO1P12*	*ND6*
31	*CALM2*	*ACTB*	*HBA2*	*ARFIP1*	*IGKV3-20*
32	*LOC105376575*	*S100A9*	*IGHM*	*TMSB4X*	*MTATP6P1*
33	*HNRNPH1*	*ERCC-00002*	*B2M*	*FTL*	*TMSB4X*
34	*PPP1CB*	*HSP90AA1*	*PF4*	*CAMP*	*LOC105376575*
35	*S100A8*	*ERCC-00004*	*UHMK1*	*B2M*	*RPLP1*
36	*JCHAIN*	*ERCC-00096*	*SDCBP*	*LYZ*	*HSP90AA1*
37	*ERCC-00004*	*ERCC-00130*	*ERCC-00003*	*AHSP*	*LTF*
38	*LIPA*	*GNG11*	*ERCC-00113*	*MNDA*	*MALAT1*
39	*NAP1L1*	*LOC105376575*	*MTND2P28*	*CALM2*	*MTCO1P12*
40	*MIR3917*	*SRGN*	*TMSB4X*	*CA1*	*EPX*
41	*HBD*	*MTCO1P13*	*ERCC-00136*	*ERCC-00136*	*RPL21*
42	*MIR4738*	*DEFA4*	*ACTB*	*ACTG1*	*FTL*
43	*ND4L*	*SAT1*	*HBD*	*SAR1A*	*HNRNPH1*
44	*WTAP*	*LIMS1*	*MTCO1P12*	*USP12*	*EEF1A1*
45	*RPL21*	*RAP1B*	*ATP8*	*ERCC-00043*	*SRGN*
46	*RPS13*	*SUB1*	*DEFA3*	*ERCC-00113*	*RNASE3*
47	*FTL*	*C6orf25*	*PPARGC1A*	*HNRNPH1*	*S100A12*
48	*ND3*	*IGKC*	*FTL*	*ATP8*	*RPS25*
49	*PTGES3*	*PLEK*	*NEAT1*	*UBE2D1*	*ZNF442*
50	*RAP1B*	*ACTG1*	*ERCC-00043*	*ARPC5*	*RPS20*
51	*CDC42*	*CTTN*	*TMEM200C*	*ACTB*	*MNDA*
52	*CAPZA1*	*HBA1*	*IGKC*	*LTF*	*RPS27A*
53	*MARCKS*	*ITGA2B*	*IGLV3-21*	*HSPA8*	*ERCC-00171*
54	*SAR1A*	*ATP8*	*MYL12B*	*PPP1CB*	*DEFA3*
55	*ACTB*	*YWHAZ*	*ERCC-00108*	*MALAT1*	*RPL5*
56	*SNRPG*	*EEF1A1*	*STGN*	*FNTA*	*SUB1*
57	*HNRNPA2B1*	*TPT1*	*F13A1*	*MTND2P28*	*ACTB*
58	*SAT1*	*HBM*	*ITGA2B*	*LCN2*	*ERCC-00113*
59	*TPT1*	*ABCC3*	*ERCC-00171*	*METTL9*	*RPS13*
60	*DNAJB9*	*DAAM1*	*LIMS1*	*TLK2*	*CD63*
61	*SUB1*	*FTL*	*GNG11*	*BTG1*	*ERLEC1*
62	*PTP4A2*	*SLC25A37*	*INIP*	*H2AFZ*	*IGHG2*
63	*UBE2D3*	*RPL21*	*ERCC-00092*	*RPL21*	*HMGB1*
64	*VCAM1*	*MIR3064*	*JCHAIN*	*RANBP2*	*H2AFZ*
65	*HMGN2*	*HNRNPH1*	*CDC42EP3*	*EEF1A1*	*TPT1*
66	*NDUFA4*	*MTND2P29*	*RNASE2*	*RPL15*	*RPLP2*
67	*RPL9*	*TLK1*	*C22orf39*	*SNX3*	*LCN2*
68	*TM9SF2*	*NAP1L1*	*RPSP1*	*TMEM9B*	*RPS2*
69	*IGKV2D-28*	*MIR6852*	*SDPR*	*HNRNPA2B1*	*HSPA8*
70	*PPA2*	*CD47*	*RNASE3*	*RPS18*	*RPL19*
71	*SMC4*	*ACRBP*	*MTND1P23*	*ERCC-00171*	*ERCC-00043*
72	*THUMPD1*	*TPM4*	*C6orf25*	*ERCC-00003*	*RPL9*
73	*PPBP*	*TTK*	*GRN*	*UBE2W*	*RPS12*
74	*MNDA*	*HLA-E*	*CLC*	*RPL7*	*RPS16*
75	*AK6*	*SNRPE*	*PLEK*	*RPL9*	*CALM2*
76	*CNBP*	*DNM3*	*EEF1A1*	*H3F3A*	*ERCC-00003*
77	*SRGN*	*ITM2B*	*ERCC-00095*	*HSP90AA1*	*RPLP0*
78	*RPL5*	*HNRNPA2B1*	*MNDA*	*DCLRE1C*	*ATP8*

79	*CD47*	*MAX*	*LYZ*	*RPL13A*	*SSR4*
80	*CD74*	*AHSP*	*ZBTB34*	*RPL19*	*RPS27*
81	*BNIP3L*	*F13A1*	*DAD1*	*MITD1*	*RPS6*
82	*EAF2*	*GPX1*	*HLA-E*	*SLC25A37*	*RPS18*
83	*MTCO1P12*	*IFNAR1*	*RPLP0*	*HMGB1*	*EEF1B2*
84	*LYZ*	*SDPR*	*RPL9*	*MIR1244-1*	*SH3BGRL*
85	*UGP2*	*SSR3*	*S100A12*	*TRMT10C*	*WTAP*
86	*ND6*	*RPLP1*	*EIF4G2*	*HBM*	*IGHA1*
87	*SRP14*	*NORAD*	*ACTG1*	*MSRB1*	*MIR3607*
88	*PLEK*	*RPLP0*	*IGHV3-33*	*WTAP*	*RPL27*
89	*APOE*	*SSR4*	*CALM2*	*LINC00920*	*PRTG*
90	*ARL8B*	*CD36*	*SSR4*	*CPNE3*	*COX7B*
91	*RNF13*	*CDC42*	*RAD54L*	*NDC1*	*FCF13*
92	*POMP*	*PPP1R12A*	*RPS20*	*RGS2*	*RPL7*
93	*MORF4L1*	*ERCC-00113*	*RPUSD3*	*GYPA*	*CA1*
94	*HSPA8*	*SRSF3*	*RPL21*	*MIR3917*	*ERCC-00136*
95	*MYL12B*	*MIR3652*	*MCL1*	*HMGN2*	*CAPZA1*
96	*MIR3064*	*SARAF*	*CD36*	*YPEL5*	*RPL13A*
97	*SET*	*SNX3*	*IGKV1D-39*	*SRGN*	*TKT*
98	*CKS2*	*NCOA4*	*MIR4738*	*ERCC-00108*	*SSR3*
99	*SNX3*	*EAF2*	*LOC101927345*	*RPS2*	*ABI1*
100	*EEF1A1*	*STOM*	*PSMA3*	*SFT2D1*	*RPL11*

Table 4.13 The 100 most highly expressed genes in each megakaryocyte ploidy group. (Platelet function associated genes are highlighted in orange)

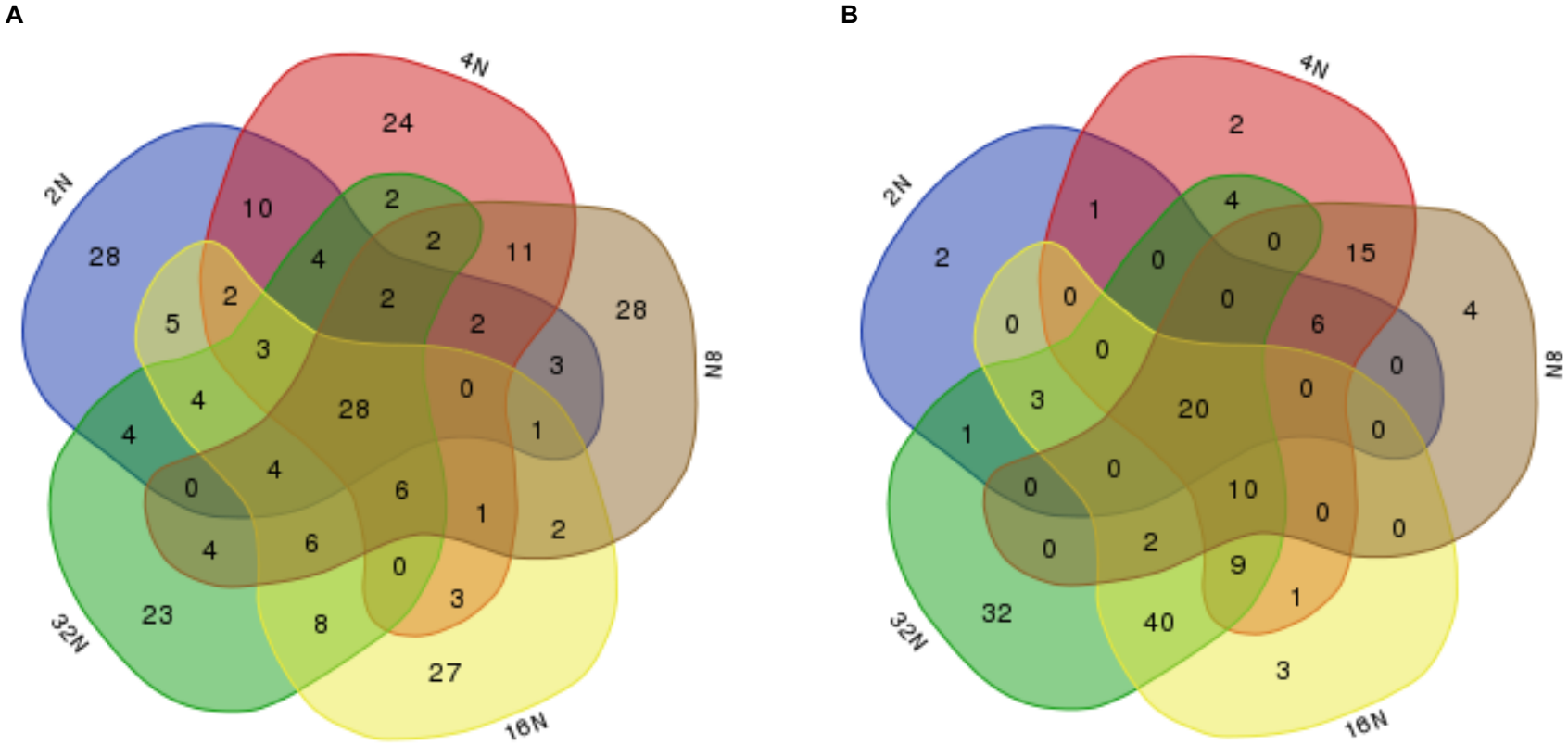


Figure 4.20 Venn diagrams to show commonality between ploidy groups.
Gene expression (A) and over-represented GO biological processes (B). Data shown based on 100 most abundantly expressed genes in each group.

	GO_BP_ID	GO_BP_NAME
2N, 4N, 8N, 16N, 32N	GO:0055114	oxidation-reduction process
	GO:0006796	phosphate-containing compound metabolic process
	GO:0010035	response to inorganic substance
	GO:0042775	mitochondrial ATP synthesis coupled electron transport
	GO:0050832	defense response to fungus
	GO:0015980	energy derivation by oxidation of organic compounds
	GO:0006091	generation of precursor metabolites and energy
	GO:0032119	sequestering of zinc ion
	GO:0022904	respiratory electron transport chain
	GO:0006979	response to oxidative stress
	GO:0006793	phosphorus metabolic process
	GO:0070488	neutrophil aggregation
	GO:0016310	phosphorylation
	GO:0006123	mitochondrial electron transport, cytochrome c to oxygen
	GO:0042773	ATP synthesis coupled electron transport
	GO:0006950	response to stress
	GO:0006119	oxidative phosphorylation
	GO:0022900	electron transport chain
	GO:0006120	mitochondrial electron transport, NADH to ubiquinone
	GO:0045333	cellular respiration
2N, 4N, 8N	GO:0042060	wound healing
	GO:0007596	blood coagulation
	GO:0007599	hemostasis
	GO:0050817	coagulation
	GO:0050878	regulation of body fluid levels
	GO:0009611	response to wounding
4N, 8N	GO:0042330	taxis
	GO:0051707	response to other organism
	GO:0065008	regulation of biological quality
	GO:0042542	response to hydroperoxide
	GO:0006935	chemotaxis
	GO:0009607	response to biotic stimulus
	GO:0009617	response to bacterium
	GO:0034109	homotypic cell-cell adhesion
	GO:0042221	response to chemical stimulus
	GO:0046903	secretion
	GO:0001775	cell activation
	GO:0002576	platelet degranulation
	GO:0030168	platelet activation
	GO:0032940	secretion by cell
	GO:0006887	exocytosis
16N, 32N	GO:0022411	cellular component disassembly
	GO:0006605	protein targeting
	GO:0019058	viral infectious cycle
	GO:0006402	mRNA catabolic process
	GO:0009057	macromolecule catabolic process
	GO:0006415	translational termination
	GO:0033214	iron assimilation by chelation and transport
	GO:0006414	translational elongation
	GO:0006413	translational initiation
	GO:0022415	viral reproductive process
	GO:0016071	mRNA metabolic process
	GO:0015031	protein transport
	GO:0000184	nuclear-transcribed mRNA catabolic process, nonsense-mediated decay
	GO:0006412	translation
	GO:0000956	nuclear-transcribed mRNA catabolic process
	GO:0015701	bicarbonate transport
	GO:0033036	macromolecule localization
	GO:0019080	viral genome expression
	GO:0045184	establishment of protein localization
	GO:0071822	protein complex subunit organization
	GO:0009056	catabolic process
	GO:0034621	cellular macromolecular complex subunit organization
	GO:0046907	intracellular transport
	GO:0006886	intracellular protein transport
	GO:0019083	viral transcription
	GO:0008104	protein localization
	GO:0033365	protein localization to organelle
	GO:0043624	cellular protein complex disassembly
	GO:0070727	cellular macromolecule localization
	GO:0032984	macromolecular complex disassembly
	GO:0033212	iron assimilation
	GO:0044265	cellular macromolecule catabolic process
	GO:0016032	viral reproduction
	GO:0034613	cellular protein localization
	GO:0071845	cellular component disassembly at cellular level
	GO:0043241	protein complex disassembly
	GO:0034623	cellular macromolecular complex disassembly
	GO:0051704	multi-organism process
	GO:0044248	cellular catabolic process
	GO:0006401	RNA catabolic process

Table 4.14 Shared GO categories between ploidy groups (most highly expressed genes).

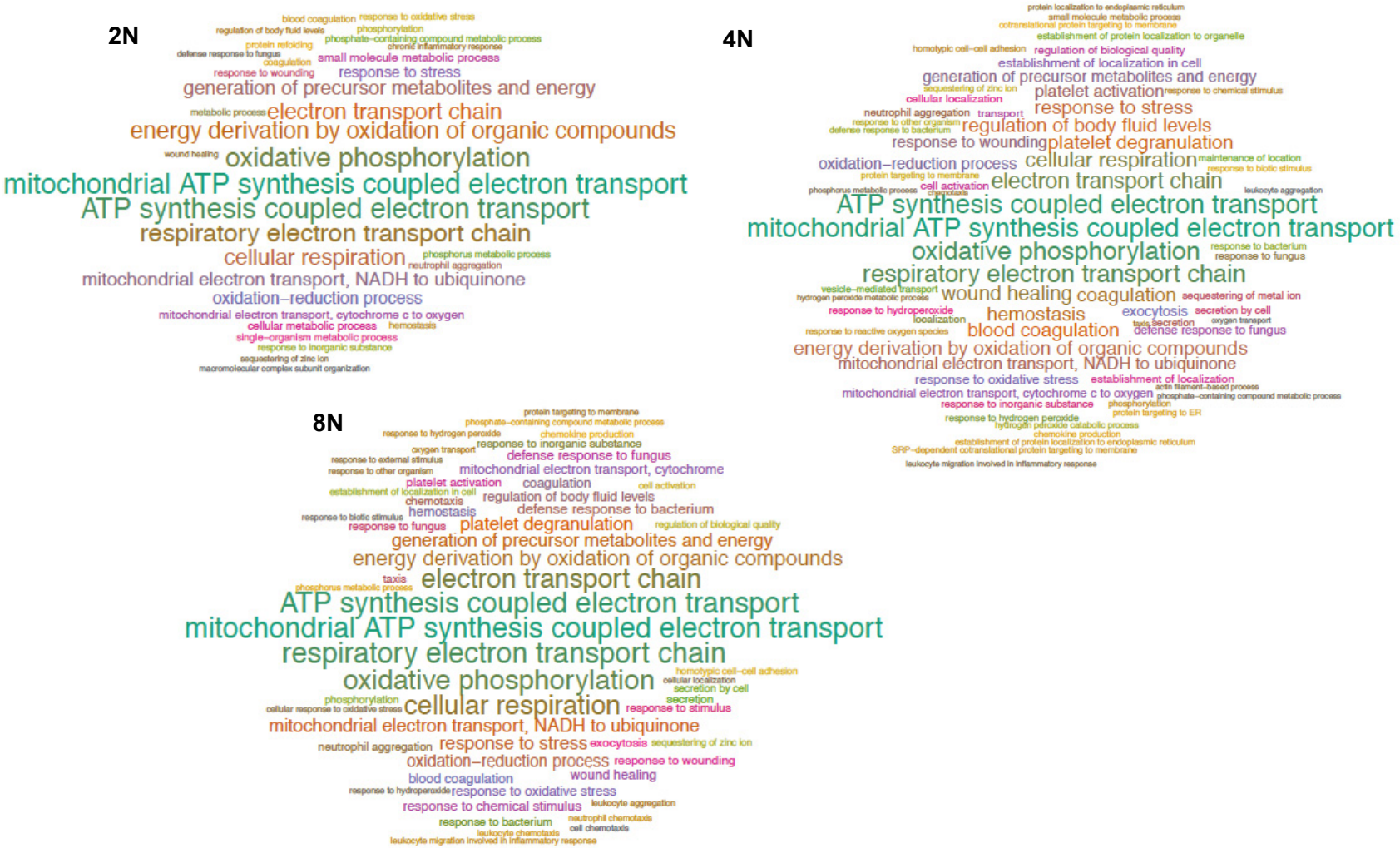
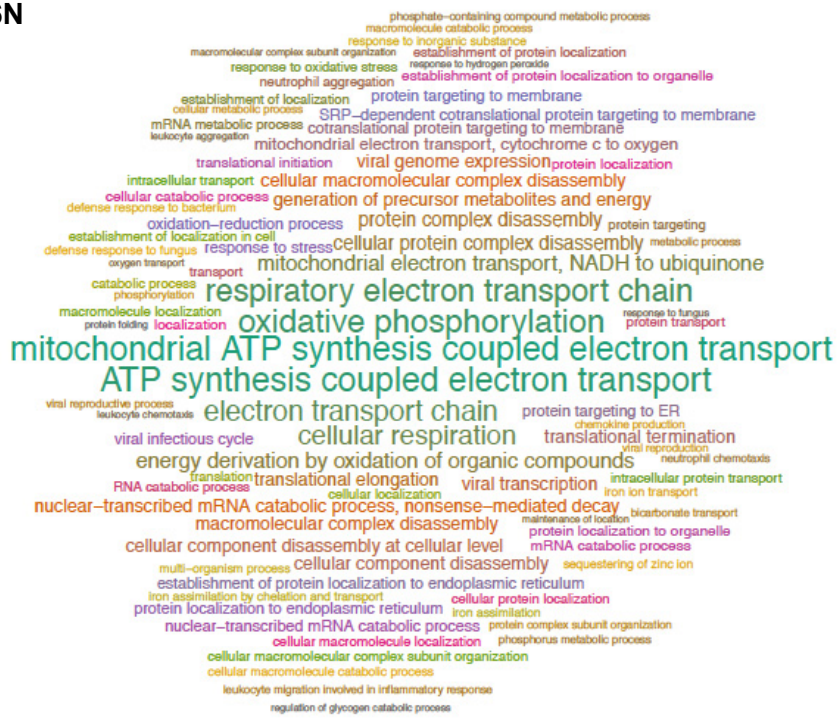


Figure 4.21 Over-represented gene ontologies in highly expressed genes in 2N, 4N and 8N ploidy groups. The full GO analysis may be found in Appendix 2. The word cloud is a visual representation of the over-represented GO biological processes with their size an indicator of significance.

16N



32N

Figure 4.22 Over-represented gene ontologies in highly expressed genes in 16N and 32N ploidy groups. The full GO analysis may be found in Appendix 2. The word cloud is a visual representation of the over-represented GO biological processes with their size an indicator of significance.

4.2.3.4 Differential expression analysis: Megakaryocyte pools

The initial differential expression analysis of megakaryocyte ploidy was focused on comparing the extreme ends of ploidy to ascertain if there is differential gene expression as polyploidisation occurs. Further pairwise expression analysis is currently in progress.

There was an experimental concern regarding the purity of the 2N cells. They were collected on the basis of the same FACS gating as shown in Figure 4.1 using surface markers of CD41 and CD42 antibodies. Therefore, by definition these are megakaryocytes. However, given that the entirety of the nucleated bone marrow cell population from which megakaryocytes were sorted were also 2N and large megakaryocytes may be fragile and break down into smaller platelet-like particles particularly during the sorting process, there was a concrete and nor experimentally controllable risk that the 2N population may be contaminated with non-megakaryocytes conjugated to platelet-like particles. The highly expressed gene analysis presented above (Section 4.2.3.3) shows the 2N cells to express a number of genes that are also expressed in other ploidy groups, however apart from the genes that are shared between all groups, mainly related to cell metabolism, there is little other overlap with other groups. Furthermore, there were only 3 2N ploidy megakaryocyte pools collected from 3 individuals compared with 6-8 megakaryocyte pools of other ploidy groups collected from 5 individuals. As there was some theoretical uncertainty, differential expression analysis was performed between 2N and 32N cells as well as 4N and 32N cells. While both of these analyses were performed and are presented here, the focus has been on the quantitative differences in gene expression between 4N and 32N.

A differential expression analysis was then performed between 32N ploidy group and the 4N ploidy group. Of the 944 differentially expressed features, 373 were shown to be significantly upregulated in 32N megakaryocytes compared with 4N ($P<0.05$, Wald test), full list is shown in Appendix 4, Table S4.12 with levels of significance. The top 50 upregulated genes are shown in Table 4.15 along with their localisations within the cell. Of the top 50, 17 transcripts encode proteins that localise to the cell membrane of these are the three G-protein coupled receptors: prostaglandin E2 receptor, pyrimidinergic receptor P2Y and the opioid receptor mu 1. Others include two low density lipoprotein receptors (*LRP6, LRP1B*), a receptor from the tumour necrosis factor receptor superfamily (*TNFRSF1B*) and a number of ion channels: *UNC80, KCNK5, CACNA2D1 and TRPM2*. The 373 upregulated genes were annotated for GO based on biological process (Appendix 4, Table S4.14; Figure 4.23). 20 GO categories were significantly enriched within the upregulated gene list and all of these were related to the same 17 genes consisting of a number of ribosomal transcripts. The over-represented GO categories all related to translational initiation, elongation and termination, protein localisation, cellular protein complex disassembly and inflammatory response.

571 transcripts were found to be significantly downregulated with increasing ploidy. The full list is shown in Appendix 4, Table S4.13 and the top 50 downregulated transcripts are presented in Table 4.16 below, with their localisation within the cell. One of the most significantly downregulated genes was tubulin folding cofactor E (*TBCE*) which is involved with cytoskeletal change and protein folding. There were 10 genes in the downregulated list that encode proteins involved in signalling including: *CYP7B1, PIK3IP1, ING2, RAB38, LEFTY1, XAF1*. 2 of the genes were related to cell-cell adhesion, encoding secreted proteins: *SRPX2* and *EPDR1* and 2 encode proteins involved in chromatin remodelling and gene expression as well as DNA repair: *KAT5, SWI5*. Performing GO analysis on the full 571 genes revealed the most over-represented biological themes included platelet degranulation, coagulation, haemostasis, wound healing, vesicle mediated transport and exocytosis (Appendix 4, Table S4.15; Figure 4.23). A number of genes related to platelet function were significantly downregulated with increased ploidy including: *PF4, ITGA2B, ITGB3, PDGFB, PDGFA, PLEK, F13A1 and RAP1B*. To give further confidence to these results, this dataset was compared with the differential expression analysis between 32N and 2N megakaryocytes by way of a Venn diagram, it showed 352 genes were shared between downregulated gene lists in the two analyses, and 48 genes were shared between upregulated gene lists (Appendix 4, Figure S4.1; Table S4.19, for reference the differential expression analysis is presented in Appendix 4, Table S4.20, S4.21, S4.22, S4.23).

4.2.3.5 Identification of transcripts encoding megakaryocyte transmembrane proteins

Using the Ensembl database (Zerbino et al. 2017) the genes upregulated with ploidy were annotated in terms of their localisation within the cell. Of all upregulated genes 78 contained transmembrane domains and localised to the cell membrane. These were further broken down into structural and functional categories, shown in Table 4.17. Table 4.18 shows all localisations for the upregulated genes, 77 were annotated as encoding cytoplasmic proteins, 57 as nuclear proteins, 27 as internal membrane proteins, 18 as mitochondrial proteins, 18 and secreted proteins and for 104 transcripts there was no localisation information available.

As seen in Table 4.18, plasma proteins upregulated in 32N cells have a wide range of functionality with 6 belonging to the immunoglobulin superfamily and 6 being G-protein coupled receptors. 2 of the immunoglobulin superfamily are immunoglobulin variable chains. Of the G-protein coupled receptors, PTGER2 is known to be expressed on the platelet membrane. This is also true of the tetraspanin CD63, the TNF alpha receptor TNFRSF1B and the lysophosphatidic acid receptor PPAP2A. A number of genes encode ion transporter channels were also upregulated.

ENSEMBL_GENE_ID	HGNC_symbol	Gene_name	log2_ FoldChange	p_ adjusted	Localisation
ENSG00000144406	*UNC80*	unc-80 homolog (C. elegans) [26582]	50.30615971	5.17E-69	Cell membrane
ENSG00000176485	*PLA2G16*	phospholipase A2, group XVI [17825]	51.86601985	1.28E-50	Cytoplasm
ENSG00000138792	*ENPEP*	glutamyl aminopeptidase (aminopeptidase A) [3355]	50.66893343	1.02E-47	Cell membrane
ENSG00000154678	*PDE1C*	phosphodiesterase 1C, calmodulin-dependent 70kDa [8776]	45.76676733	1.10E-44	Cytoplasm
ENSG00000125384	*PTGER2*	prostaglandin E receptor 2 (subtype EP2), 53kDa [9594]	48.86620804	1.29E-43	Cell membrane
ENSG00000074964	*ARHGEF10L*	Rho guanine nucleotide exchange factor (GEF) 10-like [25540]	52.0539922	1.02E-40	Cytoplasm
ENSG00000115295	*CLIP4*	CAP-GLY domain containing linker protein family, member 4 [26108]	45.25946636	4.16E-40	Cytoplasm
ENSG00000176040	*TMPRSS7*	transmembrane protease, serine 7 [30846]	51.97533467	2.92E-39	Cell membrane
ENSG00000178445	*GLDC*	glycine dehydrogenase (decarboxylating) [4313]	47.04831453	3.03E-38	Mitochondria
ENSG00000066813	*ACSM2B*	acyl-CoA synthetase medium-chain family member 2B [30931]	49.96867869	3.02E-37	Mitochondria
ENSG00000198963	*RORB*	RAR-related orphan receptor B [10259]	43.43630612	3.67E-37	Nucleus
ENSG00000213753	*CENPBD1P1*	CENPB DNA-binding domains containing 1 pseudogene 1 [28421]	47.06938471	8.86E-36	NA
ENSG00000171163	*ZNF692*	zinc finger protein 692 [26049]	45.3983006	5.40E-35	Nucleus
ENSG00000100473	*COCH*	cochlin [2180]	47.63080837	2.38E-34	Secreted
ENSG00000161944	*ASGR2*	asialoglycoprotein receptor 2 [743]	53.68317198	6.93E-34	Cell membrane
ENSG00000272754	*ENSG00000272754*	RNA gene	42.94039269	6.23E-31	NA
ENSG00000169040	*PMCHL2*	pro-melanin-concentrating hormone-like 2, pseudogene [9111]	47.85078026	1.14E-29	NA
ENSG00000165238	*WNK2*	WNK lysine deficient protein kinase 2 [14542]	48.56998379	7.34E-29	Cytoplasm
ENSG00000184967	*NOC4L*	nucleolar complex associated 4 homolog [28461]	46.26348016	9.21E-29	Nucleus
ENSG00000164626	*KCNK5*	potassium channel, two pore domain subfamily K, member 5 [6280]	47.44254629	4.63E-27	Cell membrane
ENSG00000264825	*LOC105372017*	RNA gene	47.31231763	1.68E-22	NA
ENSG00000173572	*NLRP13*	NLR family, pyrin domain containing 13 [22937]	49.89010587	1.81E-22	Cytoplasm
ENSG00000153956	*CACNA2D1*	calcium channel, voltage-dependent, alpha 2/delta subunit 1 [1399]	28.19752661	2.54E-22	Cell membrane
ENSG00000204348	*DXO*	decapping exoribonuclease [2992]	45.82662906	2.58E-22	Nucleus
ENSG00000196979	*LOC401554*	RNA gene	50.41775517	4.54E-22	NA
ENSG00000174473	*GALNTL6*	polypeptide N-acetylgalactosaminyltransferase-like 6 [33844]	36.85291755	7.88E-22	Golgi
ENSG00000116761	*CTH*	cystathionine gamma-lyase [2501]	35.42696832	1.64E-21	Cytoplasm
ENSG00000260742	*ENSG00000260742*	RNA gene	44.5450049	2.85E-21	NA
ENSG00000280120	*ENSG00000280120*		41.2001224	3.21E-21	NA
ENSG00000104413	*ESRP1*	epithelial splicing regulatory protein 1 [25966]	38.80339998	4.02E-21	Nucleus
ENSG00000171631	*P2RY6*	pyrimidinergic receptor P2Y, G-protein coupled, 6 [8543]	50.96246308	5.08E-21	Cell membrane
ENSG00000151611	*MMAA*	methylmalonic aciduria (cobalamin deficiency) cblA type [18871]	43.65871028	1.65E-20	Mitochondria
ENSG00000140538	*NTRK3*	neurotrophic tyrosine kinase, receptor, type 3 [8033]	26.68937579	1.26E-19	Cell membrane
ENSG00000258628	*RHPN2P1*	Pseudogene	23.88306708	2.20E-19	NA
ENSG00000279078	*SND1-IT1*	SND1 intronic transcript 1 [24158]	43.2920309	2.89E-19	NA
ENSG00000254226	*ENSG00000254226*	uncharacterized LOC101927115 [101927115]	43.70424897	2.55E-18	NA
ENSG00000101276	*SLC52A3*	solute carrier family 52 (riboflavin transporter), member 3 [16187]	48.62643273	2.58E-18	Cell membrane
ENSG00000173175	*ADCY5*	adenylate cyclase 5 [236]	41.65120701	5.75E-18	Cell membrane
ENSG00000251022	*THAP9-AS1*	THAP9 antisense RNA 1 [44172]	31.51449173	9.40E-18	NA
ENSG00000028137	*TNFRSF1B*	tumor necrosis factor receptor superfamily, member 1B [11917]	24.71550514	1.16E-17	Cell membrane
ENSG00000112038	*OPRM1*	opioid receptor, mu 1 [8156]	35.81585473	1.74E-17	Cell membrane
ENSG00000070018	*LRP6*	low density lipoprotein receptor-related protein 6 [6698]	23.81938493	1.94E-17	Cell membrane
ENSG00000184414	*ENSG00000184414*		38.33320975	2.35E-17	NA
ENSG00000042304	*C2orf83*	chromosome 2 open reading frame 83 [25344]	41.65507917	2.50E-17	Nucleus
ENSG00000168702	*LRP1B*	low density lipoprotein receptor-related protein 1B [6693]	20.71952594	4.68E-17	Cell membrane
ENSG00000119888	*EPCAM*	epithelial cell adhesion molecule [11529]	36.24207499	1.10E-16	Cell membrane
ENSG00000159761	*C16orf86*	chromosome 16 open reading frame 86 [33755]	44.06890648	1.72E-16	Nucleus
ENSG00000242516	*LINC00960*	long intergenic non-protein coding RNA 960 [48710]	33.15482787	1.80E-16	NA
ENSG00000130208	*APOC1*	apolipoprotein C-I [607]	22.03458584	5.28E-16	Secreted
ENSG00000242641	*LINC00971*	long intergenic non-protein coding RNA 971 [48737]	24.6489451	7.12E-16	NA

Table 4.15 Top 50 upregulated genes in 32N compared with 4N megakaryocytes

ENSEMBL_GENE_ID	HGNC_symbol	Gene_name	log2_ FoldChange	p_ adjusted	Localisation
ENSG00000134548	*SPX*	spexin hormone [28139]	-54.21517882	5.35E-76	Secreted
ENSG00000116957	*TBCE*	tubulin folding cofactor E [11582]	-52.98486843	4.44E-76	Cytoplasm
ENSG00000267575	*ENSG00000267575*		-50.31497884	6.75E-57	NA
ENSG00000246889	*ENSG00000246889*		-48.79508752	1.36E-48	NA
ENSG00000163637	*PRICKLE2*	prickle homolog 2 [20340]	-49.54589448	4.23E-40	Nucleus
ENSG00000119965	*C10orf88*	chromosome 10 open reading frame 88 [25822]	-48.5087549	1.79E-37	Nucleus
ENSG00000243709	*LEFTY1*	left-right determination factor 1 [6552]	-50.27240975	1.91E-37	Secreted
ENSG00000132530	*XAF1*	XIAP associated factor 1 [30932]	-52.25072835	1.94E-37	Cytoplasm
ENSG00000086289	*EPDR1*	ependymin related 1 [17572]	-50.40423414	3.67E-37	Secreted
ENSG00000100100	*PIK3IP1*	phosphoinositide-3-kinase interacting protein 1 [24942]	-48.8193452	1.01E-36	Cell membrane
ENSG00000224805	*LINC00853*	long intergenic non-protein coding RNA 853 [43716]	-53.57223474	6.20E-36	NA
ENSG00000172927	*MYEOV*	myeloma overexpressed [7563]	-46.43016992	7.67E-35	Nucleus
ENSG00000102359	*SRPX2*	sushi-repeat containing protein, X-linked 2 [30668]	-46.79981641	2.83E-34	Secreted
ENSG00000069509	*FUNDC1*	FUN14 domain containing 1 [28746]	-48.6174538	3.45E-33	Mitochondria
ENSG00000259869	*ENSG00000259869*		-51.81349731	1.16E-32	NA
ENSG00000132109	*TRIM21*	tripartite motif containing 21 [11312]	-45.34389945	2.22E-32	Cytoplasm
ENSG00000168994	*PXDC1*	PX domain containing 1 [21361]	-49.53984716	7.91E-32	Mitochondria
ENSG00000259225	*ENSG00000259225*		-41.55628328	4.48E-31	NA
ENSG00000120526	*NUDCD1*	NudC domain containing 1 [24306]	-39.31853818	2.74E-30	Cytoplasm
ENSG00000249274	*PDLIM1P4*	PDZ and LIM domain 1 pseudogene 4 [48947]	-51.7483801	5.38E-30	NA
ENSG00000172977	*KAT5*	K(lysine) acetyltransferase 5 [5275]	-51.20319815	1.05E-29	Nucleus
ENSG00000279476	*ENSG00000279476*		-41.17247665	1.22E-28	NA
ENSG00000154146	*NRGN*	neurogranin (protein kinase C substrate, RC3) [8000]	-24.92441388	1.70E-28	Mitochondria
ENSG00000170627	*GTSF1*	gametocyte specific factor 1 [26565]	-40.88939062	4.34E-28	Nucleus
ENSG00000172817	*CYP7B1*	cytochrome P450, family 7, subfamily B, polypeptide 1 [2652]	-51.90226179	6.66E-28	Internal membrane
ENSG00000175294	*CATSPER1*	cation channel, sperm associated 1 [17116]	-46.62517637	1.12E-27	Cytoplasm
ENSG00000156239	*N6AMT1*	N-6 adenine-specific DNA methyltransferase 1 (putative) [16021]	-51.88504515	3.22E-27	Cytoplasm
ENSG00000237181	*ENSG00000237181*		-52.08954422	4.52E-27	NA
ENSG00000168556	*ING2*	inhibitor of growth family, member 2 [6063]	-50.32067709	1.38E-26	Nucleus
ENSG00000182795	*C1orf116*	chromosome 1 open reading frame 116 [28667]	-49.61671706	2.38E-26	Cytoplasm
ENSG00000224490	*TTC21B-AS1*	TTC21B antisense RNA 1 [41115]	-43.81020591	2.88E-25	NA
ENSG00000118939	*UCHL3*	ubiquitin carboxyl-terminal esterase L3 (ubiquitin thiolesterase) [12515]	-40.20461126	1.04E-24	Cytoplasm
ENSG00000275964	*ENSG00000275964*		-51.64776396	1.89E-24	NA
ENSG00000236320	*SLFN14*	schlafen family member 14 [32689]	-53.47697412	3.13E-24	Nucleus
ENSG00000173626	*TRAPPC3L*	trafficking protein particle complex 3-like [21090]	-42.56900468	1.74E-23	Internal membrane
ENSG00000088356	*PDRG1*	p53 and DNA-damage regulated 1 [16119]	-43.89524747	9.31E-23	Cytoplasm
ENSG00000269220	*LINC00528*	long intergenic non-protein coding RNA 528 [26875]	-42.27916854	2.76E-22	NA
ENSG00000123892	*RAB38*	RAB38, member RAS oncogene family [9776]	-48.60336386	1.78E-21	Cell membrane
ENSG00000232667	*ENSG00000232667*		-52.37893194	2.42E-21	NA
ENSG00000166189	*HPS6*	Hermansky-Pudlak syndrome 6 [18817]	-50.05052468	3.85E-21	Internal membrane
ENSG00000212747	*FAM127C*	family with sequence similarity 127, member C [33156]	-46.77627963	1.71E-20	Mitochondria
ENSG00000175854	*SWI5*	SWI5 homologous recombination repair protein [31412]	-24.22649855	4.64E-20	Nucleus
ENSG00000105388	*CEACAM5*	carcinoembryonic antigen-related cell adhesion molecule 5 [1817]	-46.73755876	9.86E-20	Cell membrane
ENSG00000078177	*N4BP2*	NEDD4 binding protein 2 [29851]	-25.67727207	1.12E-19	Cytoplasm
ENSG00000156127	*BATF*	basic leucine zipper transcription factor, ATF-like [958]	-30.58506166	1.36E-19	Nucleus
ENSG00000224616	*RTCA-AS1*	RTCA antisense RNA 1 [50573]	-48.21465088	3.57E-19	NA
ENSG00000270061	*ENSG00000270061*		-49.82983008	3.33E-18	NA
ENSG00000165983	*PTER*	phosphotriesterase related [9590]	-24.19139896	4.62E-17	Cytoplasm
ENSG00000126464	*PRR12*	proline rich 12 [29217]	-36.7527534	4.71E-17	Nucleus
ENSG00000206605	*RNU6-946P*	RNA, U6 small nuclear 946, pseudogene [47909]	-41.63265249	1.03E-16	NA

Table 4.16 Top 50 downregulated genes in 32N compared with 4N megakaryocytes

Upregulated

Downregulated

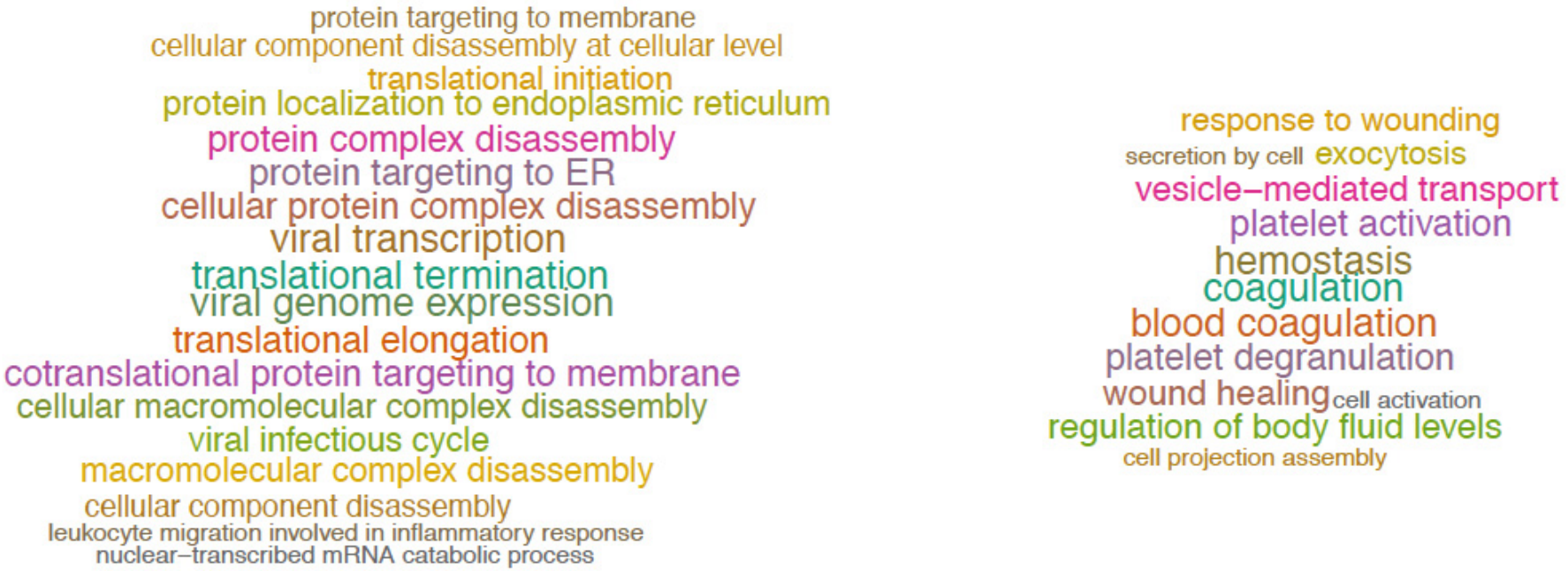


Figure 4.23 Over-represented gene ontologies in upregulated and downregulated genes in 32N compared with 4N megakaryocytes.
The full GO analysis may be found in Appendix 2. The word cloud is a visual representation of the over-represented GO biological processes with their size an indicator of significance.

ENSEMBL_GENE_ID	HGNC_symbol	Gene_name	p_ adjusted
		Integrins	
ENSG00000137809	*ITGA11*	integrin, alpha 11 [6136]	5.49E-07
		Immunoglobulin superfamily	
ENSG00000211962	*IGHV1-46*	immunoglobulin heavy variable 1-46 [5554]	9.73E-10
ENSG00000211659	*IGLV3-25*	immunoglobulin lambda variable 3-25 [5908]	2.80E-04
ENSG00000164520	*RAET1E*	retinoic acid early transcript 1E [16793]	4.03E-04
ENSG00000124731	*TREM1*	triggering receptor expressed on myeloid cells 1 [17760]	2.31E-03
ENSG00000081237	*PTPRC*	protein tyrosine phosphatase, receptor type, C [9666]	1.43E-02
ENSG00000186818	*LILRB4*	leukocyte immunoglobulin-like receptor, subfamily B (with TM and ITIM domains), member 4 [6608]	3.49E-02
		G-protein coupled receptors	
ENSG00000125384	*PTGER2*	prostaglandin E receptor 2 (subtype EP2), 53kDa [9594]	1.29E-43
ENSG00000171631	*P2RY6*	pyrimidinergic receptor P2Y, G-protein coupled, 6 [8543]	5.08E-21
ENSG00000112038	*OPRM1*	opioid receptor, mu 1 [8156]	1.74E-17
ENSG00000107518	*ATRNL1*	attractin-like 1 [29063]	9.05E-07
ENSG00000157542	*KCNJ6*	potassium channel, inwardly rectifying subfamily J, member 6 [6267]	8.47E-04
ENSG00000171049	*FPR2*	formyl peptide receptor 2 [3827]	9.22E-03
		Leucine Rich Repeat proteins	
ENSG00000185158	*LRRC37B*	leucine rich repeat containing 37B [29070]	9.14E-03
		Tetraspanins	
ENSG00000135404	*CD63*	CD63 molecule [1692]	2.83E-02
		Cytokine/chemokine/complement receptors	
ENSG00000028137	*TNFRSF1B*	tumor necrosis factor receptor superfamily, member 1B [11917]	1.16E-17
ENSG00000203710	*CR1*	complement component (3b/4b) receptor 1 (Knops blood group) [2334]	5.34E-15
ENSG00000134470	*IL15RA*	interleukin 15 receptor, alpha [5978]	6.19E-11
ENSG00000131724	*IL13RA1*	interleukin 13 receptor, alpha 1 [5974]	2.24E-08
ENSG00000054967	*RELT*	RELT tumor necrosis factor receptor [13764]	2.79E-03
		Low density lipoprotein receptors	
ENSG00000168702	*LRP1B*	low density lipoprotein receptor-related protein 1B [6693]	4.68E-17
ENSG00000070018	*LRP6*	low density lipoprotein receptor-related protein 6 [6698]	1.94E-17
		Lectins	
ENSG00000161944	*ASGR2*	asialoglycoprotein receptor 2 [743]	6.93E-34
ENSG00000184293	*CLECL1*	C-type lectin-like 1 [24462]	7.43E-03
		Transporter proteins	
ENSG00000144406	*UNC80*	unc-80 homolog (C. elegans) [26582]	5.17E-69
ENSG00000164626	*KCNK5*	potassium channel, two pore domain subfamily K, member 5 [6280]	4.63E-27
ENSG00000153956	*CACNA2D1*	calcium channel, voltage-dependent, alpha 2/delta subunit 1 [1399]	2.54E-22
ENSG00000101276	*SLC52A3*	solute carrier family 52 (riboflavin transporter), member 3 [16187]	2.58E-18
ENSG00000142185	*TRPM2*	transient receptor potential cation channel, subfamily M, member 2 [12339]	1.58E-15
ENSG00000181333	*HEPHL1*	hephaestin-like 1 [30477]	2.73E-11
ENSG00000148408	*CACNA1B*	calcium channel, voltage-dependent, N type, alpha 1B subunit [1389]	3.54E-11
ENSG00000091138	*SLC26A3*	solute carrier family 26 (anion exchanger), member 3 [3018]	8.78E-10
ENSG00000167614	*TTYH1*	tweety family member 1 [13476]	3.93E-08
ENSG00000164124	*TMEM144*	transmembrane protein 144 [25633]	1.37E-07
ENSG00000137571	*SLCO5A1*	solute carrier organic anion transporter family, member 5A1 [19046]	3.13E-04
ENSG00000033867	*SLC4A7*	solute carrier family 4, sodium bicarbonate cotransporter, member 7 [11033]	3.61E-03
ENSG00000153253	*SCN3A*	sodium channel, voltage gated, type III alpha subunit [10590]	1.30E-02
ENSG00000154864	*PIEZO2*	piezo-type mechanosensitive ion channel component 2 [26270]	1.41E-02
ENSG00000144452	*ABCA12*	ATP-binding cassette, sub-family A (ABC1), member 12 [14637]	1.55E-02
ENSG00000182389	*CACNB4*	calcium channel, voltage-dependent, beta 4 subunit [1404]	1.82E-02
ENSG00000141485	*SLC13A5*	solute carrier family 13 (sodium-dependent citrate transporter), member 5 [23089]	3.03E-02
ENSG00000180773	*SLC36A4*	solute carrier family 36 (proton/amino acid symporter), member 4 [19660]	3.35E-02
		Enzymatic activity	
ENSG00000140538	*NTRK3*	neurotrophic tyrosine kinase, receptor, type 3 [8033]	1.26E-19
ENSG00000067113	*PPAP2A*	phosphatidic acid phosphatase type 2A [9228]	1.80E-10
ENSG00000167261	*DPEP2*	dipeptidase 2 [23028]	1.93E-09
ENSG00000168594	*ADAM29*	ADAM metallopeptidase domain 29 [207]	2.31E-08
ENSG00000257335	*MGAM*	maltase-glucoamylase [7043]	4.30E-03
ENSG00000178568	*ERBB4*	erb-b2 receptor tyrosine kinase 4 [3432]	6.98E-03
ENSG00000151651	*ADAM8*	ADAM metallopeptidase domain 8 [215]	4.09E-02
		Other	
ENSG00000176040	*TMPRSS7*	transmembrane protease, serine 7 [30846]	2.92E-39
ENSG00000119888	*EPCAM*	epithelial cell adhesion molecule [11529]	1.10E-16
ENSG00000162542	*TMCO4*	transmembrane and coiled-coil domains 4 [27393]	6.07E-12
ENSG00000141574	*SECTM1*	secreted and transmembrane 1 [10707]	5.12E-11
ENSG00000156738	*MS4A1*	membrane-spanning 4-domains, subfamily A, member 1 [7315]	5.42E-06
ENSG00000181143	*MUC16*	mucin 16, cell surface associated [15582]	1.18E-03

ENSG00000235568	*NFAM1*	NFAT activating protein with ITAM motif 1 [29872]	1.83E-03
ENSG00000178075	*GRAMD1C*	GRAM domain containing 1C [25252]	2.92E-03
ENSG00000038945	*MSR1*	macrophage scavenger receptor 1 [7376]	5.32E-03
ENSG00000092421	*SEMA6A*	sema domain, transmembrane domain (TM), and cytoplasmic domain, (semaphorin) 6A [10738]	9.17E-03
ENSG00000146839	*ZAN*	zonadhesin (gene/pseudogene) [12857]	1.59E-02
ENSG00000147872	*PLIN2*	perilipin 2 [248]	2.23E-02
ENSG00000002933	*TMEM176A*	transmembrane protein 176A [24930]	2.60E-02
ENSG00000155719	*OTOA*	otoancorin [16378]	4.67E-02
		Signalling	
ENSG00000138792	*ENPEP*	glutamyl aminopeptidase (aminopeptidase A) [3355]	1.02E-47
ENSG00000173175	*ADCY5*	adenylate cyclase 5 [236]	5.75E-18
ENSG00000031081	*ARHGAP31*	Rho GTPase activating protein 31 [29216]	5.34E-15
ENSG00000101384	*JAG1*	jagged 1 [6188]	2.44E-08
ENSG00000185527	*PDE6G*	phosphodiesterase 6G, cGMP-specific, rod, gamma [8789]	4.24E-06
ENSG00000133818	*RRAS2*	related RAS viral (r-ras) oncogene homolog 2 [17271]	1.18E-04
ENSG00000126709	*IFI6*	interferon, alpha-inducible protein 6 [4054]	4.21E-04
ENSG00000134594	*RAB33A*	RAB33A, member RAS oncogene family [9773]	6.58E-04
ENSG00000198010	*DLGAP2*	discs, large (Drosophila) homolog-associated protein 2 [2906]	5.23E-03
ENSG00000266028	*SRGAP2*	SLIT-ROBO Rho GTPase activating protein 2 [19751]	8.76E-03
ENSG00000196975	*ANXA4*	annexin A4 [542]	2.44E-02
ENSG00000131381	*RBSN*	rabenosyn, RAB effector [20759]	2.53E-02
ENSG00000130475	*FCHO1*	FCH domain only 1 [29002]	4.16E-02
		Scaffold proteins	
ENSG00000180900	*SCRIB*	scribbled planar cell polarity protein [30377]	4.23E-07
ENSG00000215045	*GRID2IP*	glutamate receptor, ionotropic, delta 2 (Grid2) interacting protein [18464]	4.16E-05

Table 4.17 Upregulated transmembrane proteins in high ploidy megakaryocytes.
Transcripts have been arranged into platelet associated protein families and functional groups and ranked according to level of differential expression between 32N and 4N ploidy.

Localisation	Significantly upregulated genes in 32N megakaryocyte pools compared with 4N
Cell membrane	*UNC80, ENPEP, PTGER2, TMPRSS7, ASGR2, KCNK5, CACNA2D1, P2RY6, NTRK3, SLC52A3, ADCY5, TNFRSF1B, OPRM1, LRP6, LRP1B, EPCAM, TRPM2, ARHGAP31, CR1, TMCO4, HEPHL1, CACNA1B, SECTM1, IL15RA, PPAP2A, SLC26A3, IGHV1-46, DPEP2, MYO1B, IL13RA1, ADAM29, JAG1, TTYH1, TMEM144, SCRIB, ITGA11, ATRNL1, PDE6G, MS4A1, GRID2IP, RRAS2, IGLV3-25, SLCO5A1, RAET1E, IFI6, RAB33A, KCNJ6, MUC16, NFAM1, TREM1, RELT, GRAMD1C, SLC4A7, MGAM, DLGAP2, MSR1, ERBB4, CLECL1, SRGAP2, LRRC37B, SEMA6A, FPR2, SCN3A, PIEZO2, PTPRC, ABCA12* *ZAN, CACNB4, PLIN2, ANXA4, RBSN, TMEM176A, CD63, SLC13A5, SLC36A4, LILRB4, ADAM8, FCHO1, OTOA*
Cytoplasm	*PLA2G16, PDE1C, ARHGEF10L, CLIP4, WNK2, NLRP13, CTH, USP13, KRT23, MYH4, GDPGP1, PRG3, TTLL4, GYS1, GSDMA, NAF1, MYO7A, UNC13C, SPATA21, CAMSAP1, PLEKHS1, ARHGEF26, PRKAA2, OAS3, GAS2, ASZ1, RGS7, KIF20A, APAF1, KLHL22, GALK1, MYO10, SAC3D1, SCLT1, FRY, DNAH8, CHORDC1, SPHK2, RPL38, CNKSR2, TANC2, CDC45, SACS, SEPT14, CCNG2, HS3ST1, PTPN13, DIAPH3, FAM49B, RPLP2, EPB41L4A, GPALPP1, RPS21, PIWIL3, RPLP1, SLFN13, HOOK2, RPL27, PSMB7, UBA6, SWAP70, FRMPD1, PSMA6, DOCK3, PLBD1, TTC6, HINT3, FEZ1, RPS12, PSMB9, CHODL, SLC24A4, RPL27A, RPS17, MTHFD1, RPS26*
Internal membrane	*GALNTL6, SNX24, CCZ1B, CFTR, YIF1A, CCDC51, POMT2, LDAH, ACAP1, STEAP2, TMEM38B, NCS1, PIPOX, ST6GALNAC1, PIGK, RETSAT, FAM21C, UBIAD1, SMPD4, SNX14, EMC10, RYR1, GCNT1, VPS18, TRAPPC8, AGPAT9, SNX13*
Mitochondria	*GLDC, ACSM2B, MMAA, GRAMD4, OMA1, SFXN4, TFB1M, MCCC1, GOT2, COX7B, IDH3B, ATP5G2, ABAT, NDUFB11, COX7C, MARC1, SURF1, NDUFC2, CENPBD1P1*
Nucleus	*RORB, ZNF692, NOC4L, DXO, ESRP1, C2orf83, C16orf86, NUF2, NLE1, ZNF234, KIAA0556, ZNF763, ASF1B, FIGN, SH2D6, ZNF347, ZCCHC4, NUP35, NOVA1, POLR3H, ZNF343, PELO, CCDC155, TATDN1, ZNF780A, PWP2, IQCA1, CHTF18, KIAA0101, AIM1, CDCA5, RHNO1, ZNF174, WDHD1, ZNF543, WDR12, WDFY3, PRIM2, COMMD6, ZMYM5, LRRIQ1, FBXO48, DNASE1L3, NCAPD2, PPARG, ZWILCH, ATAD2, MAP2K6, ZNF154, LYRM1, BRAT1*
Secreted	*COCH, APOC1, TIMP2, TLL2, CDNF, AGR3, GDF11, OVCH1, ADAMTS1, CES4A, CPAMD8, HSD11B1L, LAMA3, S100A8, OSCAR, VCAN, CFD, S100A9*

Table 4.18 Upregulated transcripts in 32N vs 4N ploidy groups according to localisation within the cell

4.2.3.6 Gene expression analysis: Megakaryocyte single cell

Based on single cell quality filtering 282 single cell megakaryocytes were taken forward for gene expression analysis (Section 4.2.3.2). A PCA was performed to identify correlation between the cells based on the principal components of variation in gene expression. Figure 4.24 shows separation of cell according to the first two principal components and indicate that the ploidy groups do not cluster together.

A differential expression analysis was then performed to identify genes that were consistently upregulated with increasing levels of ploidy from 2N cells up to 32N cells. Further analysis of their data is in progress with particular attention to pairwise comparisons between individual ploidy classes.

In total 328 genes were found to be differentially expressed. Of these 28 genes were found to be consistently upregulated with increasing ploidy, shown in Table 4.19 with their localisation. The main group of genes significantly upregulated across increasing ploidy in the single cell analysis was ribosomal transcripts and others involved in all parts of translation and protein targeting as seen before adding strength to this observation. Interestingly there were also a number of upregulated genes encoding proteins involved in platelet function including: *ACTB, RAP1B, TMSB4X* and *TUBA4A* that was not seen in the megakaryocyte pool ploidy analysis where these genes were found to be downregulated in 32N cells. Out of the 28 upregulated genes, 8 encoded for proteins localised to the nucleus, 14 encoded for cytoplasmic proteins, 2 genes encoded plasma membrane proteins, 1 localised to internal membrane and 1 to the mitochondria. The cell membrane encoding genes included *HLA-E* and *RAP1B*.

320 genes were significantly downregulated with increasing ploidy. The top 50 genes are shown in Table 4.20 with their localisations and the full list can be found in Appendix 4, Table S4.16. Of the downregulated genes, those associated with platelet function include: *PLEK, PECAM1* (platelet endothelial cell adhesion molecule 1), *TPM1, ANGPT1, PDGFA, PDGFB, PTGIR2* (prostacyclin receptor).

GO analysis performed to ascertain which biological processes were enriched with increasing ploidy based on the single cell expression data. In keeping with the multicellular ploidy pools, the enriched GO terms for the genes upregulated with increasing ploidy were related to translation, protein targeting and cellular mRNA disassembly. By contrast genes downregulated with increasing ploidy had an ontological bias towards coagulation, haemostasis, platelet degranulation, exocytosis and wound healing as was seen in the megakaryocyte multicellular pool ploidy analyses. The GO analysis is presented in word cloud format in Figure 4.25 and in full in Appendix 4, Tables S4.17, S4.18.

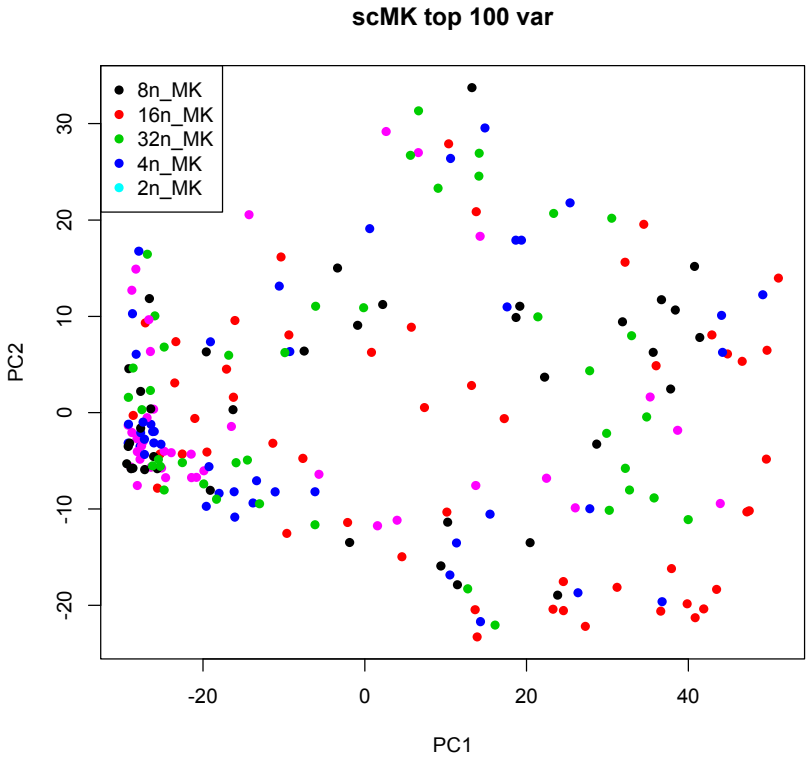


Figure 4.24 Unsupervised clustering of single cell megakaryocytes: highly variable gene expression.
PCA plot constructed from normalized log-expression values of correlated highly variable genes, where each point represents a cell in the HSC dataset. First and second principal components are shown, along with the percentage of variance explained and ploidy level identified for each cell.

ENSEMBL_GENE_ID	HGNC_symbol	Gene_name	log2_ FoldChange	p_ adjusted	Localisation
ENSG00000137154	*RPS6*	ribosomal protein S6 [10429]	3.183026171	0	Nucleus
ENSG00000174444	*RPL4*	ribosomal protein L4 [10353]	3.137861447	0	Nucleus
ENSG00000142676	*RPL11*	ribosomal protein L11 [10301]	2.622808412	0	Nucleus
ENSG00000100316	*RPL3*	ribosomal protein L3 [10332]	2.620738572	0	Nucleus
ENSG00000110700	*RPS13*	ribosomal protein S13 [10386]	2.598318566	0	Nucleus
ENSG00000177954	*RPS27*	ribosomal protein S27 [10416]	2.539357652	0	Nucleus
ENSG00000109475	*RPL34*	ribosomal protein L34 [10340]	2.461253922	0	Cytoplasm
ENSG00000122566	*HNRNPA2B1*	heterogeneous nuclear ribonucleoprotein A2/B1 [5033]	2.33894906	0	Nucleus
ENSG00000156508	*EEF1A1*	eukaryotic translation elongation factor 1 alpha 1 [3189]	2.264242007	0	Cytoplasm
ENSG00000144713	*RPL32*	ribosomal protein L32 [10336]	2.216490508	0	Cytoplasm
ENSG00000109971	*HSPA8*	heat shock 70kDa protein 8 [5241]	2.212457967	0	Cytoplasm
ENSG00000108298	*RPL19*	ribosomal protein L19 [10312]	2.142127333	0	Cytoplasm
ENSG00000251562	*MALAT1*	Metastasis-associated lung adenocarcinoma transcript 1	2.062369743	0	NA
ENSG00000136167	*LCP1*	lymphocyte cytosolic protein 1 (L-plastin) [6528]	2.030492545	0	Cytoplasm
ENSG00000180879	*SSR4*	signal sequence receptor, delta [11326]	1.784311093	0	Internal membrane
ENSG00000204592	*HLA-E*	major histocompatibility complex, class I, E [4962]	1.65275616	0	Cell membrane
ENSG00000075624	*ACTB*	actin, beta [132]	1.629435694	0	Cytoplasm
ENSG00000196262	*PPIA*	peptidylprolyl isomerase A (cyclophilin A) [9253]	1.528351964	0	Cytoplasm
ENSG00000142669	*SH3BGRL3*	SH3 domain binding glutamate-rich protein like 3 [15568]	1.471993812	0	Cytoplasm
ENSG00000127314	*RAP1B*	RAP1B, member of RAS oncogene family [9857]	1.356731701	0	Cell membrane
ENSG00000205542	*TMSB4X*	thymosin beta 4, X-linked [11881]	0.654205228	1.83E-165	Cytoplasm
ENSG00000127824	*TUBA4A*	tubulin, alpha 4a [12407]	1.015278925	1.82E-93	Cytoplasm
ENSG00000196924	*FLNA*	filamin A, alpha [3754]	0.781610326	5.21E-81	Cytoplasm
ENSG00000233276	*GPX1*	glutathione peroxidase 1 [4553]	0.588387748	2.39E-53	Cytoplasm
ENSG00000101608	*MYL12A*	myosin, light chain 12A, regulatory, non-sarcomeric [16701]	0.403757228	1.26E-16	Cytoplasm
ENSG00000249119	*MTND6P4*	MT-ND6 pseudogene 4 [39467]	0.088909274	5.51E-03	Mitochondria
ENSG00000112576	*CCND3*	cyclin D3 [1585]	0.132308239	8.35E-03	Nucleus
ERCC-00044	*ERCC-00045*		0.041129269	4.00E-02	NA

Table 4.19 Upregulated genes with increasing ploidy in single cell megakaryocytes

ENSEMBL_GENE_ID	HGNC_symbol	Gene_name	log2_ FoldChange	p_ adjusted	Localisation
ENSG00000172889	*EGFL7*	EGF-like-domain, multiple 7 [20594]	-1.405588023	2.13E-109	Secreted
ENSG00000149781	*FERMT3*	fermitin family member 3 [23151]	-0.55228455	2.04E-36	Cytoplasm
ENSG00000175538	*KCNE3*	potassium channel, voltage gated subfamily E regulatory beta subunit 3 [6243]	-0.483025625	3.06E-17	Cell membrane
ENSG00000169224	*GCSAML*	germinal center-associated, signaling and motility-like [29583]	-0.565362592	1.22E-14	Nucleus
ENSG00000081041	*CXCL2*	chemokine (C-X-C motif) ligand 2 [4603]	-0.478482101	6.73E-12	Secreted
ENSG00000267279	*ENSG00000267280*		-0.529413699	2.25E-11	NA
ENSG00000115956	*PLEK*	pleckstrin [9070]	-0.259756908	3.11E-10	Cell membrane
ENSG00000174233	*ADCY6*	adenylate cyclase 6 [237]	-0.431146164	6.23E-10	Cell membrane
ENSG00000129993	*CBFA2T3*	core-binding factor, runt domain, alpha subunit 2; translocated to, 3 [1537]	-0.388466758	9.44E-10	Nucleus
ENSG00000261371	*PECAM1*	platelet/endothelial cell adhesion molecule 1 [8823]	-0.259713049	3.92E-09	Cell membrane
ENSG00000144893	*MED12L*	mediator complex subunit 12-like [16050]	-0.298571295	6.85E-08	Nucleus
ENSG00000160190	*SLC37A1*	solute carrier family 37 (glucose-6-phosphate transporter), member 1 [11024]	-0.303833455	1.33E-07	Cell membrane
ENSG00000223551	*TMSB4XP4*	thymosin beta 4, X-linked pseudogene 4 [11886]	-0.157090467	1.51E-07	NA
ENSG00000106077	*ABHD11*	abhydrolase domain containing 11 [16407]	-0.322363877	6.17E-07	Mitochondria
ENSG00000187653	*TMSB4XP8*	thymosin beta 4, X-linked pseudogene 8 [11885]	-0.139808966	1.04E-06	NA
ENSG00000154146	*NRGN*	neurogranin (protein kinase C substrate, RC3) [8000]	-0.255505957	3.81E-06	Mitochondria
ENSG00000160013	*PTGIR*	prostaglandin I2 (prostacyclin) receptor (IP) [9602]	-0.268074194	8.68E-06	Cell membrane
ENSG00000047648	*ARHGAP6*	Rho GTPase activating protein 6 [676]	-0.305996489	1.50E-05	Cytoplasm
ENSG00000137449	*CPEB2*	cytoplasmic polyadenylation element binding protein 2 [21745]	-0.269642079	1.75E-05	Cytoplasm
ENSG00000167460	*TPM4*	tropomyosin 4 [12013]	-0.214371351	2.40E-05	Cytoplasm
ENSG00000154188	*ANGPT1*	angiopoietin 1 [484]	-0.297951969	3.44E-05	Secreted
ENSG00000010671	*BTK*	Bruton agammaglobulinemia tyrosine kinase [1133]	-0.234362074	3.67E-05	Cytoplasm
ENSG00000081087	*OSTM1*	osteopetrosis associated transmembrane protein 1 [21652]	-0.2660356	4.63E-05	Internal membrane
ENSG00000181104	*F2R*	coagulation factor II (thrombin) receptor [3537]	-0.306239198	4.87E-05	Cell membrane
ENSG00000035403	*VCL*	vinculin [12665]	-0.169847357	6.50E-05	Cytoplasm
ENSG00000197563	*PIGN*	phosphatidylinositol glycan anchor biosynthesis, class N [8967]	-0.242488409	6.50E-05	Internal membrane
ENSG00000205038	*PKHD1L1*	polycystic kidney and hepatic disease 1 (autosomal recessive)-like 1 [20313]	-0.267146195	7.15E-05	Cell membrane
ENSG00000150867	*PIP4K2A*	phosphatidylinositol-5-phosphate 4-kinase, type II, alpha [8997]	-0.210377296	7.96E-05	Cell membrane
ENSG00000017260	*ATP2C1*	ATPase, Ca++ transporting, type 2C, member 1 [13211]	-0.215356598	9.88E-05	Internal membrane
ENSG00000177697	*CD151*	CD151 molecule (Raph blood group) [1630]	-0.264987752	1.65E-04	Cell membrane
ENSG00000277048	*ENSG00000277048*		-0.101174062	1.79E-04	NA
ENSG00000185052	*SLC24A3*	solute carrier family 24 (sodium/potassium/calcium exchanger), member 3 [10977]	-0.271097886	1.90E-04	Cell membrane
ENSG00000071127	*WDR1*	WD repeat domain 1 [12754]	-0.161861544	2.10E-04	Cytoplasm
ENSG00000068383	*INPP5A*	inositol polyphosphate-5-phosphatase, 40kDa [6076]	-0.260094512	3.22E-04	Cell membrane
ENSG00000172757	*CFL1*	cofilin 1 (non-muscle) [1874]	-0.124005504	3.47E-04	Nucleus
ENSG00000151702	*FLI1*	Fli-1 proto-oncogene, ETS transcription factor [3749]	-0.175687065	3.56E-04	Nucleus
ENSG00000175701	*LINC00116*	long intergenic non-protein coding RNA 116 [27339]	-0.2472789	3.86E-04	Cell membrane
ENSG00000108839	*ALOX12*	arachidonate 12-lipoxygenase [429]	-0.265892496	3.91E-04	Cytoplasm
ENSG00000158769	*F11R*	F11 receptor [14685]	-0.222992927	3.91E-04	Cell membrane
ENSG00000104805	*NUCB1*	nucleobindin 1 [8043]	-0.193667087	5.92E-04	Internal membrane
ENSG00000241553	*ARPC4*	actin related protein 2/3 complex, subunit 4, 20kDa [707]	-0.156279405	5.94E-04	Cytoplasm
ENSG00000128294	*TPST2*	tyrosylprotein sulfotransferase 2 [12021]	-0.223439941	6.01E-04	Internal membrane
ENSG00000101290	*CDS2*	CDP-diacylglycerol synthase (phosphatidate cytidylyltransferase) 2 [1801]	-0.195695857	7.33E-04	Mitochondria
ENSG00000131236	*CAP1*	CAP, adenylate cyclase-associated protein 1 (yeast) [20040]	-0.148400922	7.38E-04	Cell membrane
ENSG00000169062	*UPF3A*	UPF3 regulator of nonsense transcripts homolog A (yeast) [20332]	-0.173921756	7.79E-04	Nucleus
ENSG00000125534	*PPDPF*	pancreatic progenitor cell differentiation and proliferation factor [16142]	-0.186288949	8.88E-04	Nucleus
ENSG00000157637	*SLC38A10*	solute carrier family 38, member 10 [28237]	-0.179342179	9.17E-04	Cell membrane
ENSG00000128739	*SNRPN*	small nuclear ribonucleoprotein polypeptide N [11164]	-0.212731259	1.02E-03	Nucleus
ENSG00000173230	*GOLGB1*	golgin B1 [4429]	-0.176591809	1.02E-03	Internal membrane
ENSG00000213719	*CLIC1*	chloride intracellular channel 1 [2062]	-0.142037678	1.14E-03	Nucleus

Table 4.20 Top 50 downregulated genes with increasing ploidy in single cell megakaryocytes

Upregulated

Downregulated

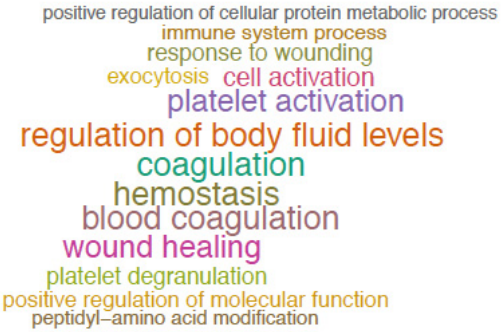


Figure 4.25 Over-represented gene ontologies in upregulated and downregulated genes with increasing ploidy in single cell megakaryocytes.
The full GO analysis may be found in Appendix 2. The word cloud is a visual representation of the over-represented GO biological processes with their size an indicator of significance.

4.2.3.7 Expression of megakaryocyte ploidy signatures in HSC clusters

To assess the expression of the megakaryocyte ploidy gene signature within the clustered single cell HSCs from Chapter 3, the expression of the genes upregulated in increasing and those upregulated in decreasing ploidy was overlaid onto the Monocle 2 plot in Figure 3.16. This not only shows the HSC clusters but also the differentiation trajectories that the single cells follow. In keeping with previously published datasets of *in vitro* derived megakaryocytes the expression of megakaryocyte ploidy signatures was mainly restricted to cluster 1. There was little or no expression of ploidy associated gene signatures in cluster 4 which was thought to be another megakaryocyte-biased subset from Chapter 3, Figure 4.26. Here the gene signatures were from the megakaryocyte multicellular pools, rather than single cells.

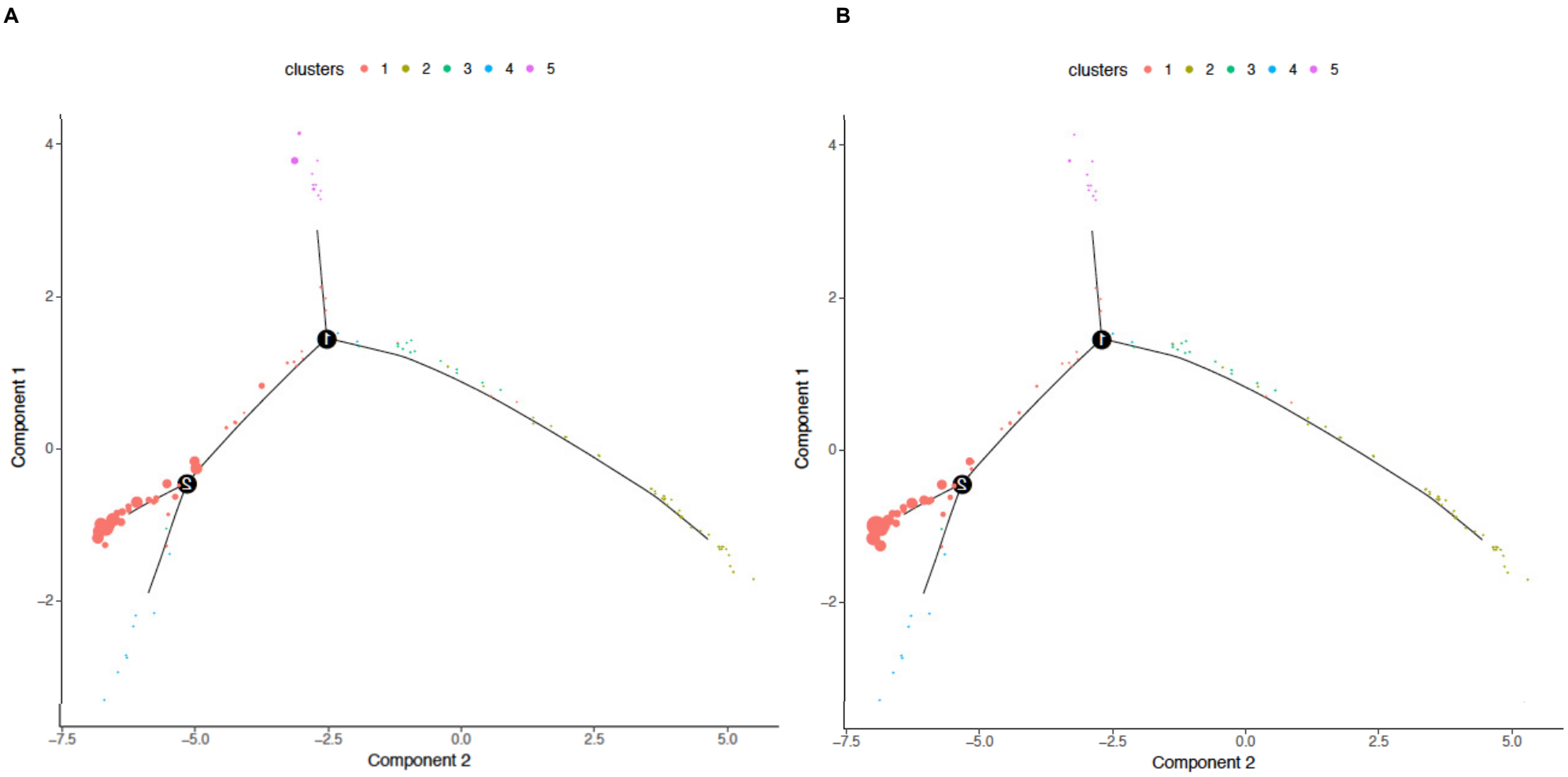


Figure 4.26 Single cell HSC differentiation trajectories using Monocle 2 showing expression of megakaryocyte ploidy signatures: multicellular pools. Individual HSCs are connected by a minimum spanning tree with branch points (thin lines), representing the differentiation trajectory and clusters are differentiated by colour. Expression of gene signature is indicated by the size of each single cell. a. Increasing ploidy signature; b. Decreasing ploidy signature.

4.3 Discussion

The aim of this chapter was to obtain transcriptome data primary human bone marrow megakaryocytes for the first time, to compare them to their *in vitro* counterparts cultured from human CD34+ cells and to examine differences in gene expression with increasing ploidy class. Current knowledge on the megakaryocyte transcriptome is based on *in vitro* derived cells (L. Chen et al. 2014) which are phenotypically different from megakaryocytes found *in vivo*. They reach a mean ploidy of 2-4N and in conditions of prolonged culture are able to produce small numbers of platelets known as proplatelets with limited functionality (Moreau et al. 2016; Machlus and Italiano 2013). In contrast *in vivo*, megakaryocytes reach mean ploidy levels of 16N and with each cell releasing thousands of platelets into the circulation. An understanding of the transcriptional profile of megakaryocytes *in vivo* and if gene expression is regulated with changes in ploidy may provide insight the processes of megakaryocyte maturation and platelet production.

Until this study no transcriptome data existed for *ex-vivo* megakaryocytes. This pertains to a number of technical difficulties in isolating pure megakaryocyte populations which are required for meaningful transcriptomics. Here megakaryocytes were isolated from fresh sternal bone marrow scrapings from patients undergoing non-coronary cardiac surgery (aortic/mitral valve replacement). After substantial optimisation of protocols full length RNA-seq of megakaryocytes was achieved as single cells and multicellular pools of up to 100 cells. Whole bone marrow was stained using megakaryocyte markers CD41 and CD42 as well as the DNA dye Hoechst. Viable megakaryocytes were then sorted as single cells or multicellular pools; either by specific ploidy groups based on Hoechst staining or as whole populations in pools up to 100 cells with ploidy levels the same as *in vivo* proportions. The cells were sorted into a DNA and RNA lysis buffer and RNA-seq libraries were prepared using a recently published modified Smartseq2 protocol (G&T-seq (Macaulay et al. 2015)) from which both genome and transcriptome may be simultaneously amplified if required (megakaryocyte DNA was stored only). While the amplified cDNA produced was of good quality according to length it was shown to contain overwhelming levels of mitochondrial transcript according to both qPCR and RNA-seq, which saturated much of the sequencing capacity. Based on significantly low proportion of single cells containing house-keeping/megakaryocyte lineage genes on qPCR and RNA-seq as well as poor RNA-seq quality measures compared with multicellular pools, attention was then focused on multicellular pools only. Multicellular megakaryocyte pools that were collected irrespective of ploidy were then selected for RNA-seq based on low relative levels of mitochondrial transcript on qPCR and multicellular ploidy pools were all sequenced due to smaller numbers. The vast majority of multicellular pool sequencing data showed good quality statistics. The single cell RNA-seq data was of poorer quality but was then subjected to stringent cell selection on the basis of a number of quality measures using a machine learning model as described in Section 2.6.2.

In order to compare the transcriptome from bone marrow megakaryocytes with that of cultured megakaryocytes which forms the current knowledge base, megakaryocytes were derived *in vitro* from adult blood CD34+ cells. The main differences in library preparation was the input material. For cultured megakaryocytes, 1ng of RNA from >1x10^6cells was used, whereas for bone marrow megakaryocytes, all the RNA from 50-100 cells was used as input material into the Smart reaction. Therefore, when comparing data from cultured megakaryocytes which is bulk RNA-seq and bone marrow megakaryocytes certain caveats must be kept in mind. In single cell and low input RNA-seq there is a relative paucity of starting material giving rise to an abundance of both biological and technical zeros or gene 'drop-out' and increased variability. However, in this analysis the transcriptome data from a number of technical and biological replicates were combined in order to perform differential gene expression analysis between the two datasets. The phenotypic difference in terms of ploidy level of the cells should also be noted here; the megakaryocytes derived *in vitro* had a mean ploidy of 2N to 4N although appeared polyploid on microscopy whereas the bone marrow megakaryocytes in multicellular pools had a mean ploidy of 16N and were larger by size (Figures 4.1, 4.11).

A number of the most abundantly expressed transcripts in bone marrow megakaryocytes encoded for platelet proteins involved in activation and showed shared expression in cultured megakaryocytes. Of the most highly expressed genes in bone marrow megakaryocytes were the glycoproteins: *ITGA2B* (CD41) and *CD36* and *RAP1B* involved in integrin signalling. Also, highly expressed were *PPBP* and *PF4* which are potent chemokines released from platelets on activation and are also activators of neutrophils in the platelet response. A number of genes involved in the process of cytoskeletal rearrangement that occurs in proplatelet formation and platelet activation/spreading showed high gene expression in bone marrow megakaryocytes including *TUBA4A, ACTB, MYL12a* and *MYL6* encoding tubulin, actin and myosin light chain proteins. Furthermore, the structural regulators *PLEK* (a major protein kinase C substrate in platelets), the actin-capping protein *CAPZA1* and actin polymerisation regulator *TMSB4X* were also highly expressed.

Many of the highly expressed transcripts encoded genes important in cellular respiration. Transcripts mapping to the mitochondria were the most highly expressed genes in both cultured and bone marrow megakaryocytes. The overwhelming presence of mitochondrial transcripts in the RNA-seq libraries was a challenge but likely to be a biological phenomenon as the level of mitochondrial transcripts was orders of magnitude greater in megakaryocytes compared with HSCs (Figure 3.5) and these transcripts have previously been shown to be highly expressed in the platelet transcriptome (Eicher et al. 2016). This is likely to be functional as compared with other circulating leukocytes, platelets have demonstrated increased metabolic activity (Kramer et al. 2014). This may be related to their relatively small

size and high surface area associated with the extensive open canalicular system of the cell as well as ATP-dependent calcium and other ion channels that are required to prevent inadvertent platelet activation during circulation. Interestingly mitochondrial number within platelets, thought to be pre-determined in the megakaryocyte, is increased in patients with cancer (Wang et al. 2015; Zharikov and Shiva 2013) and separate to their role in cellular metabolism have also been shown to be released upon activation as microparticles (Boudreau et al. 2014).

A differential gene expression analysis identified 51 unique transcripts that were upregulated in bone marrow megakaryocytes compared with cultured megakaryocytes and 470 transcripts that were downregulated. Interestingly, GO analysis of the genes downregulated in bone marrow megakaryocytes showed a functional bias towards translation, protein targeting, platelet degranulation and cell cycle. In terms of platelet activation, a number of genes were downregulated including *PLEK, PPBP, CD63* (a surface marker of platelet activation), *RAP1B, TMSB4X, ACTB, CAPZA2, F13A1* (coagulation factor XIII), the platelet and endothelial cell adhesion molecule, *PECAM1* and the transcription factor *GATA1*, known to promote megakaryopoiesis. This may in part be due to recent differentiation of cultured megakaryocytes from CD34+ cells, having recently been exposed to high concentrations of TPO, which may not be in the same levels at in bone marrow. However, it should be noted here that while they are described as downregulated, these genes were expressed at high levels also in bone marrow megakaryocytes.

Only 51 genes were found to be significantly upregulated in bone marrow compared with cultured megakaryocytes with significance levels much higher than the downregulated genes ($p>1x10^{-39}$, compared with $p>1x10^{-4}$ respectively). These included genes involved in the defense response which reflects the terms enriched in GO analyses: defense response, response to bacterium, response to fungus, cell killing, leukocyte chemotaxis, immune response, inflammatory response.

It is known that bacteria can induce platelet activation both directly and indirectly (Fitzgerald, Foster, and Cox 2006). However, there is a body of evidence now emerging that platelets play an important role in immunity (Morrell et al. 2014). They are able to release microbicidal proteins termed collectively as platelet microbicidal proteins (PMPs), they express Toll-like receptors that detect pathogens (Aslam et al. 2006), also expressed on megakaryocytes (Beaulieu et al. 2011), and they are able to internalise bacteria and viruses (Youssefian et al. 2002). Platelets participate in the inflammatory response by recruitment of leukocytes and have influence over their maturation and activation (Gear and Camerini 2010), and contribute to antimicrobial neutrophil extracellular trap formation (Brinkmann et al. 2004). There is growing evidence to also support their role in adaptive immunity including the expression of CD40L on their surface which stimulates antigen presenting cells (Kroczek et al. 1998) and

the expression of MHC class I complexes for antigen presentation (Chapman et al. 2012) amongst others.

Of the upregulated genes were the defensins *DEFA3* and *DEFA4*. They encode proteins involved in the host defense against bacteria and fungus where they insert into cell membranes and lead to cell permeability (Aslam et al. 2006). They are found in neutrophil granules as well as epithelial surfaces but not previously identified in platelets. Another defensin *DEFB1* has more recently been found in the alpha granule of the platelet (Kraemer et al. 2011) as a class of PMPs. The genes *S100A8, S100A9* and *S100A12* that encode calcium binding proteins are also highly upregulated in bone marrow megakaryocytes. They have been observed at protein level within the platelet (Genecards.org). They play an important role in the inflammatory response and are involved in neutrophil chemotaxis and adhesion, which occurs in platelet. Other genes differentially expressed in bone marrow megakaryocytes are thought to be specific to neutrophils such as the cell surface receptor: peptidoglycan recognition protein 1 (*PGLYRP1*) which is a pattern receptor recognising and killing bacteria and neutrophil granular proteins: *AZU1, LTF, MPO*. They have not previously been identified megakaryocytes or platelets. A final set of 9 upregulated genes in bone marrow megakaryocytes encoded constant and variable regions of IgG, IgA, IgK (Ig Kappa), IgL (Ig Lambda) light and heavy chains some of which were highly significant. There is no current literature discussing the possible functionality of immunoglobulin transcripts or protein in megakaryocytes. Out of the 7 genes, *IGHG1, IGHG2, IGHG3, IGHG4, IGHA1, IGLC2, IGLC3, IGKC, IGKV1-5*, the encoded protein is seen in proteomic studies in platelets with the exception of *IGKV1-5* (Genecards.org). Platelet transcript has been seen at moderate levels for all but *IGKV1-5, IGHG2, IGHG4* (Rowley et al. 2011). There are also moderate levels of transcription in cultured megakaryocytes of *IGHG1, IGHG2, IGLC3, IGKC* (Novershtern et al. 2011).

Taken together, it would appear that the transcriptional profile of megakaryocytes cultured *in vitro*, derived from CD34+ cells and *ex-vivo* primary human bone marrow megakaryocytes are comparable. Genes significantly upregulated in bone marrow megakaryocytes which are represented at lower levels in cultured megakaryocytes are those pertaining to the immune and inflammatory response. The high levels at which genes involved in innate immunity and well as those encoding for immunoglobulins are expressed in *ex-vivo* megakaryocytes suggest an important role for them in the immune response therefore adding further to the repertoire of circulating platelet functionality. This finding supports recent evidence that megakaryocytes may themselves play a role in immunity where genes related to innate immunity and inflammation have been shown to be upregulated in *ex-vivo* lung megakaryocytes in mice compared with those found in the bone marrow (Lefrançais et al. 2017). Transcription of these genes may also be related to megakaryocyte interactions within the osteoclastic or vascular bone marrow niche.

Given that the gene expression in megakaryocytes *in vitro* and bone marrow megakaryocytes was relatively comparable, it is interesting that there is a discrepancy between platelets produced *in vitro* and *in vivo*. Proplatelets can be formed in suspension from megakaryocytes derived from either CD34+ cells (Choi et al. 1995; Miyazaki et al. 2000; Thon et al. 2010) or indeed iPSCs (Moreau et al. 2016; Feng et al. 2014; Nakamura et al. 2014). However, the efficiency remains poor with the numbers on average 1000-fold lower that *in vivo* (Moreau et al. 2016). This suggests that there remains insufficient understanding of the trigger required for platelet production. It is likely to be related to the complex interactions of megakaryocytes in the vascular niche whereby proplatelets are released into the sinusoidal space. The lack of difference in gene expression between cultured and bone marrow megakaryocytes may also be explained by the relative heterogeneity of the bone marrow megakaryocytes compared with cultured megakaryocytes, with cells at different stages of maturation and ploidy. Therefore, further investigation was then focused on the possibility of transcriptional variation between ploidy populations.

Megakaryocytes collected as 20 cell pools and single cells and distinguished according to ploidy level (2N to 32N) were sequenced for transcriptome analysis to gain insights into modulation of gene expression according to ploidy. Firstly the 100 most highly expressed genes were analysed for each ploidy group and shared expression by gene list and functional annotation using GO between ploidy groups. All groups expressed high levels of genes mapping to the mitochondria and those related to cellular metabolism (28 genes, 20 enriched GO biological processes shared between all ploidy groups) however both in terms of gene lists and over-represented GO categories there was seen to be a dichotomy with more overlap between ploidy classes 2N, 4N, 8N (particularly 4N and 8N) and then separately 16N and 32N.

The lower ploidy levels expressed increasing numbers of genes involved in platelet function within the 100 highly expressed genes, up to 8N. Very few platelet function associated genes featured in the top 100 gene lists for 16N or 32N ploidy groups, only genes encoding for structural proteins: actin, TMSB4X (an actin polymerisation regulator) and SRGN (a collagen binding proteoglycan released by platelets). Alternatively, 16N and 32N megakaryocytes demonstrated high expression of genes associated with mRNA disassembly, all parts of translation, protein packaging or targeting, localisation to organelle, membrane or endoplasmic reticulum. This GO categories were most significant in the 32N megakaryocytes. The differential expression analysis that was performed between 32N and 4N as well as the ploidy signature seen in the limited single cell analysis that was performed showed a similar pattern and therefore added a further degree of confidence to the data.

The differential expression analysis between 32N megakaryocytes and 4N megakaryocytes revealed 373 transcripts that were upregulated with increased ploidy and 571 transcripts that were downregulated with increased ploidy. Functional enrichment analysis indicated that while there were 373 upregulated transcripts increasing ploidy was associated with a limited number of biological functions. These included translational initiation, elongation and termination, protein localisation, protein complex disassembly and inflammatory response. The downregulated genes, again showed significant ontological bias to a limited number of biological processes. These were all related to coagulation, platelet degranulation, haemostasis and wound healing. This was reflected in the single cell analysis. Here only 28 genes were upregulated with increasing ploidy, the vast majority of these were ribosomal transcripts meaning that the main upregulated biological processes were the same and in the multicellular pools. However, interestingly there were some genes relating to platelet function that were also found to be upregulated. These included the cell membrane protein RAP1B and structural proteins: TMSB4X and TUBA4A. The 300 downregulated genes again showed significant ontological bias towards the same GO terms mentioned above (platelet activation, wound healing, coagulation and haemostasis).

It should be noted here that the multicellular pools differential expression analysis was performed between 32N and 4N to compare high with low ploidy classes. However, the single cell analysis was a differential expression analysis was performed to identify genes that were consistently upregulated with increasing levels of ploidy from 2N cells up to 32N cells, so genes that were consistently differentially expressed. This may account for the upregulation of genes that are involved in platelet activation which on the basis on the previous data presented could be differentially expressed due to increasing expression levels in intermediate ploidy groups for example 2N up to 8N. The next step in data analysis which will help to better understand some of this uncertainty will be pairwise comparisons between all ploidy groups in multicellular pools with corroboration from the single cell data.

Up until now the purpose of polyploidy in megakaryocytes and its association with platelet release remains unclear. There is evidence to support low levels of platelet production from low ploidy megakaryocytes both *in vitro* (Mattia et al. 2002; Miyazaki et al. 2000) and *in vivo* in myelodysplastic syndromes (Wong and Chan 2008) and it has been suggested that platelets released from different ploidy levels have different functionality (Martin et al. 2012; Hancock, Martin, and Lelchuk 1993). It has been suggested that polyploidisation is required to meet the requirement for large quantities of mRNA and protein to be packaged into platelets without energy used for mitosis and cytokinesis (Zimmet and Ravid 2000), however relationship between polyploidisation to achieve high nuclear content and terminal megakaryocyte differentiation related to efficient platelet release and whether these are linked or independent processes is still debated.

The aim of transcriptome analysis of different megakaryocyte ploidy classes was to better understand the processes occurring at different ploidy levels and how they relate to megakaryocyte maturity and platelet release. The data presented here suggests that at lower levels of megakaryocyte ploidy, a large number of platelet specific genes are transcribed. However, as ploidy level increases the transcription of platelet function associated genes is downregulated with and increase in transcription of ribosomal genes and genes encoding proteins involved directly in translation and protein packaging and targeting to membrane. It is reasonable to hypothesise, therefore, that ploidy level is in fact related to megakaryocyte maturity. While platelet function associated transcription occurs at lower ploidy levels, with increasing rounds of endomitosis or megakaryocyte maturity, these transcripts are then translated to proteins and packaged ready for platelet release. Therefore, with increasing ploidy there is an upregulation of transcripts related to translation and protein targeting. At high levels of ploidy where there is a high demand for protein synthesis it is entirely appropriate to significantly increase transcription of ribosomal subunits.

This fits with previous studies of megakaryocyte ultrastructure through different stages of maturity where the 4 stages of megakaryocyte maturation have been described (Bessis 1956, 1973) as: stage 1: megakaryoblasts; stage 2: basophilic megakaryocyte (2N/4N); stage 3: granular megakaryocyte (8N, 16N, 32N); stage 4: thrombocytogenic megakaryocyte. By electron microscopy, in stage 2 megakaryocytes the nucleus and cytoplasm develop synchronously with a number of large mitochondria, free ribosomes and smooth endoplasmic reticulum seen within the cytoplasm. Very few alpha or dense granules are seen (Kosaki and Kambayashi 2011; Ru et al. 2015). The ploidy transcriptome data presented in this chapter indicates an upregulation of transcription of platelet function genes at this stage. It also shows that in the next stage there is preferential upregulation of transcription of genes involved in translation, formation of ribosomal subunits and protein packaging which is in keeping with the electron microscopic appearances observed. Here, there are increased numbers of rough endoplasmic reticulum (large numbers of ribosomes now attached), mitochondria, Golgi and formed granules are seen within the cytoplasm (Kosaki and Kambayashi 2011; Ru et al. 2015). The demarcation membrane system is also seen at this stage which has been shown to be made in part by extension on rough endoplasmic reticulum and golgi apparatus (Ru et al. 2016). In keeping with some of the GO terms seen to be enriched in the higher ploidy megakaryocytes the ultrastructural appearance of the nucleus in stage 3 indicates a shutting-down of transcription in the maturing cell with an increase in the proportion of heterochromatin, and smaller nucleoli (Tanum and Engeset 1983). The megakaryocyte is then ready for the final stage where the cytoplasm shows global fragmentation into individual platelets via the demarcation membrane system and for platelet release (Kosaki and Kambayashi 2011; Ru et al. 2015). Therefore, the transcriptome data presented in this chapter shows that with megakaryocyte maturity there is an increase in translation related transcription which supports existing evidence from electron microscopy.

The only possible comparative transcriptome data that has been published is a study in which megakaryocytes where cultured from human CD34+ cells were sorted on the basis of ploidy level with the transcriptome compared between 2N/4N cells and 8N/16N cells (which due to small numbers were collected from a number of biological replicates). This data showed a number of upregulated and downregulated features with the overall pattern showing an upregulation of genes involved with platelet, coagulation and haemostatic pathways and a downregulation of cell cycle associated genes (Raslova et al. 2007). This is in direct contrast with the results presented in this chapter. This may at least in part be due to the fact that the cultured cells were all in the same conditions receiving the same extracellular cues from TPO and IL1B and the cell suspension media whereas the megakaryocytes in the bone marrow receive different external cues and signalling at different times in their development reflecting a complexity of the bone marrow niche form the osteoclastic niche to the vascular sinusoid. The fact that cultured megakaryocytes are unable to consistently produce functional platelets and certainly not in physiological numbers suggests that the extracellular growth environment in culture compared to *in vivo* is still very different and may account for the clear transcriptional difference between higher ploidy megakaryocytes grown in culture compared to their *ex-vivo* counterparts.

In order to investigate the extracellular signals that might be triggers for physiological platelet release in mature megakaryocytes 78 of the upregulated genes in higher ploidy megakaryocytes localising to the plasma membrane were characterised further by structure and function. It was envisaged that differentially expressed genes in higher ploidy megakaryocytes would encoding transmembrane receptors that might transduce the signals required for changes in functionality leading to platelet production from the surrounding extracellular environment. These include movement from the osteoclastic niche in which megakaryocytes are in close proximity with HSCs where the most abundant component is collagen 1 (Reddi et al. 1977) to the vascular niche polyploid megakaryocytes localise to where the later stages of megakaryocyte maturation and platelet release into vascular sinusoids can occur (Tavassoli and Aoki 1989; Avecilla et al. 2004). In this environment there may be a number of cues from the extracellular matrix made up of collagen IV, fibrinogen, fibronectin and VWF. Therefore, it is possible that signalling through the receptors that are seen to be upregulated with ploidy could be important in this migration of megakaryocytes or lead to important modifications in function and initiate release of platelets. It is also interesting that out of 373 genes that were upregulated in higher ploidy 55 remain unannotated with unknown functionality, which could provide an avenue for further investigation.

Of the 78 upregulated genes that localised to the plasma membrane 6 were found to encode G-protein coupled receptors. These included the transcript encoding the PGE2 receptor, known to be present on the platelet membrane. This can be either stimulatory by potentiating

responses to ADP and collagen or inhibitory to platelet aggregation. Its activity is mediated by stimulation of adenylate cyclase via G-proteins (Fabre et al. 2001). PGE2 is known to be important in regulation of the HSC expansion and function (North et al. 2007) and is therefore present in the bone marrow stem cell niche suggesting it could have a regulatory role in modulating megakaryocyte maturity and platelet production. Furthermore, PGE2 has been shown to specifically promote megakaryocyte lineage recovery after radiation injury in mouse (Niswander et al. 2015). Another of the G-protein coupled receptors is the pyrimidine nucleotide receptor P2Y6, whose transcript is upregulated in higher ploidy megakaryocytes. It comes from the same family as the P2Y12 ADP receptor but is not known to be present in transcript form or protein form in the platelet (Rowley et al. 2011; Genecards.org). It binds UDP and it is proposed that it mediates the inflammatory response.

Transcripts for a number of cytokine receptors are also upregulated with increasing ploidy, two from the TNF receptor superfamily *TNFRSF1B*. This is known to be present on the platelet membrane and enhance platelet response to collagen (Pignatelli et al. 2005). However, TNF binding can also suppress apoptosis so may have an important role within the stem cell niche as megakaryocytes mature. Moreover TNF alpha has recently been shown to induce megakaryopoiesis in haematopoietic progenitors which could in part account for increased platelet mass in inflammation (Haas et al. 2015). Transcripts encoding for other cytokine receptors include *RELT*, another TNF superfamily receptor, the anti-inflammatory *IL13RA1, IL15RA* as well as the complement receptor *CR1*.

Only 1 of the upregulated transcripts encoded a leucine rich repeat receptor, *LRRC37B*. This has no known function. A number of receptors from this family are found on the platelet surface including the GPIb-IX-V and toll-like receptors. Other transmembrane proteins that are known to be expressed on the surface of platelets with transcripts upregulated with ploidy are: *PPAP2A, CD63* a tetraspanin known as an activation marker for platelets and *SEMA6A*.

The increasing and decreasing ploidy signatures from the most robust dataset (multicellular pools) were then superimposed on the Monocle plot clustering the single cell HSCs from Chapter 3 according to differentiation trajectories (Figure 4.26). The expression of these ploidy associated gene signatures was observed in the different HSC clusters. HSC clusters 1 and 4 were hypothesised to be megakaryocyte-biased by comparison to previously published data. Both increasing and decreasing ploidy signatures were highly expressed in cluster 1 HSCs. There were low levels of expression of these gene signatures in all other HSC clusters. The MEP gene signature was also seen in cluster 1 and therefore in Chapter 3 cluster 1 was proposed as the HSC cluster that might produce megakaryocytes in steady state thrombopoiesis. While cluster 4 was the HSC cluster which might be recruited in the setting of accelerated thrombopoiesis. The observation that both increasing and decreasing ploidy gene signature if the HSC cluster hypothesis were to be true would suggest that

transcriptional modulation according to ploidy is not associated with the switch from steady state to accelerated thrombopoiesis in states of haemostatic demand. Whether megakaryocyte gene signatures from a clinical setting with anticipated accelerated thrombopoiesis might show increased expression in cluster 4 is investigated in the next chapter.

Limitations of the methods and relevance to the results presented in this chapter would include those relating to single cell and low input sequencing as described in the previous chapter. Technical dropouts or “zeros” lead to false negatives which are often transcripts that are expressed at low levels. Therefore, transcripts that play a role in regulating differentiation and cell fate decisions which exhibit burst expression (Suter et al. 2011) and perhaps at low levels may be missed. This may be compounded by a PCR amplification bias towards highly expressed genes and may lead to false similarity between cells states when assessing differential gene expression. The limitations involved in the comparisons between bulk and low input RNA-seq comparing cultured megakaryocytes and bone marrow megakaryocytes are highlighted above. The first section of this Chapter focuses on the technical challenges of working with this particular cell type, many of which were overcome, however the quality of the resulting single cell sequencing data was poorer than the HSC data and therefore only used for corroboration.

The data presented in this chapter represents the first interrogation of *ex-vivo* primary bone marrow megakaryocytes at the transcript level both as a whole population as well as separated by ploidy group. It demonstrates that gene expression is regulated by ploidy level reflecting the processes occurring within the cell and may reveal putative drivers in the understanding of in-vivo megakaryopoiesis. It also proposes 2 transcriptional states in the maturation of megakaryocytes; the first focused on transcription of genes associated with platelet function with a transition towards transcription of protein synthesis associated genes as the cell matures in preparation for platelet formation and release.

CHAPTER 5
Platelet and megakaryocytes in patients with myocardial infarction

5.1 Preface

Platelet turnover in acute coronary thrombosis is thought to be similar to that in states of increased haemostatic demand. It has been shown that patients with acute coronary syndromes have higher mean platelet volume (MPV), both historically (Martin et al. 1983; Cameron et al. 1983; Kishk, Trowbridge, and Martin 1985) and in the era of P2Y12 receptor antagonists (Ranjith et al. 2009; Lippi et al. 2009; S. G. Chu et al. 2010; H. Chu et al. 2011; Sansanayudh et al. 2014). MPV is increased in stable coronary artery disease (Guthikonda et al. 2008; A. Yang, Pizzulli, and Lüderitz 2006) and is shown to change with age, gender, ethnicity, obesity, changes in lipid levels and presence of diabetes (Bain 1996; Segal and Moliterno 2006; Saxena et al. 1987; Biino et al. 2011; Santimone et al. 2011; Muscari et al. 2008), factors that modulate risk of atherosclerosis and coronary thrombosis. MPV has been correlated with platelet function and increased expression of platelet activation markers in patients with AMI (Milner and Martin 1985; Kristensen, Bath, and Martin 1990; Giles, Smith, and Martin 1994) and has been shown to be a strong predictor of outcome (Huczek et al. 2005; Taglieri et al. 2011; Aksu et al. 2009; Azab et al. 2011; Goncalves et al. 2011) suggesting its pathological significance. Moreover, there is evidence to show increases in numbers of reticulated platelets in AMI and in the setting of in-stent thrombosis (Guthikonda et al. 2008; Cesari et al. 2008; Würtz et al. 2010) as well as increases in circulating TPO levels (Senaran et al. 2001). This suggests that there may be fundamental differences in megakaryocyte activity, platelet production and platelet function in patients with ACS. However, megakaryocytes from patients with acute coronary thrombosis have not been studied and compared with the healthy population yet.

In this chapter platelets and megakaryocytes are characterised in terms of function and phenotype in order to further explore the relationship between platelet phenotype and megakaryocyte activity in both the steady state and the pathological state of platelet thrombosis. Comparisons are made between patients with recent AMI and severe coronary artery disease, undergoing coronary artery bypass grafting and a control group undergoing non-coronary cardiac surgery.

In the pathological setting of thrombosis, there is an alteration in platelet activity thought to be modulated at the point of thrombopoiesis (Martin et al. 2012). Characterising the megakaryocyte and platelet transcriptomes in the setting of thrombosis may guide further understanding of the pathogenesis of thrombosis.

5.2 Results: Megakaryocyte and platelet phenotype

5.2.1 Patient characteristics

Platelets and megakaryocytes were collected from 46 individuals undergoing cardiac surgery: 27 from the control group and 19 patients with recent AMI. Of these 39 were paired samples with platelets and megakaryocytes from the same individuals (Controls: n=22, AMI n=17). In 3 individuals only blood samples were collected (platelet data only) and in 4 individuals only bone marrow samples were taken (megakaryocyte data only). The study design and sample use are shown in Appendix 1 and 2 respectively.

Table 5.1 shows the baseline characteristics of each study population and the differences found between them. The AMI group had a higher proportion of males compared with control ($p=0.013$). There was also a higher incidence of smoking ($p=0.0009$), hypertension ($p=0.0156$) and statin use ($p=0.0156$) in the AMI group, all coronary risk factors. Interestingly there was no difference in the incidence of diabetes between the 2 groups.

The level of residual antiplatelet activity at the time of surgery was higher for the AMI group with increased use of aspirin within 48 hours (n=19) and P2Y12 inhibitor within 5 days (n=9), (p value for both <0.0001). Pre-operative platelet count taken 1-14 days prior to surgery was significantly higher in patients with AMI compared to controls despite dual antiplatelet use. Samples were taken from patients in the AMI group a mean of 38 days post-infarction (range: 5 - 117 days). There was no correlation between time from infarct and pre-operative platelet count (data not shown).

Table 5.1 Baseline patient characteristics

	Controls n=27	AMI n=19	p value
Gender M	12	16	0.013*
Age (mean±SD)	65.1 ± 14.83	65.4 ± 8.79	0.9278
Hypertension	17	18	0.0156*
Statins	11	15	0.0156*
Diabetes	6	8	0.1988
Smoking	0	7	0.0009*
Aspirin within last 48 hours	4	19	<0.0001*
ADP antagonist within last 5 days	0	9	<0.0001*
Left ventricular dysfunction	7	8	0.3412
Coronary disease on angiography	-	19	-
History of AMI within 6 months	-	19	-
Time from infarction (days, mean±SD)	-	37.7 ± 33.61	-
Aortic valve surgery	16	2	0.0017*
Mitral valve surgery	12	0	0.0005*
Pre-operative Hb (mean±SD)	13.01 ± 1.95	12.64 ± 1.73	0.5118
Pre-operative WC (mean±SD)	7.53 ± 2.21	8.12 ± 2.89	0.4399
Pre-operative Plt (mean±SD)	222.3 ± 46.0	264.0 ± 70.8	0.0195*

Baseline data from patients with myocardial infarction and severe coronary disease (AMI) and a control group with normal coronary arteries undergoing valvular heart surgery (Controls).

5.2.2 Platelet volume and count

Platelet volume is known to be increased in patients with AMI compared to controls a parameter which is increased within 12 hours post-infarction (Martin, Bath, and Burr 1991). This has also been shown with the routine use of dual antiplatelet agents (Sansanayudh et al. 2014). In order to assess basic circulating blood cell parameters in the AMI and control groups blood was collected in EDTA just prior to bone marrow sampling and analysed on a Sysmex-5000 instrument. Platelets collected in EDTA are have been shown to swell with time in a predictable manner (Bath 1993). As demonstrated in Figure 5.1, all blood samples were analysed at a similar time interval from collection, allowing the groups to remain comparable (overall mean 153.5 ± 26.8 min) with no correlation between MPV and time from collection to analysis (r=0.09).

No difference was observed in MPV between the AMI and control group (11.06 ± 0.85fL, 11.16 ± 0.91fL respectively, Figure 5.2a). This observation, whilst conflicting with the known evidence base, could in part be explained by the use of younger healthier individuals as controls in previous studies as aging has been associated with increases in MPV (Santimone et al. 2011). In this dataset controls are age-matched and despite having no evidence of coronary disease, have severe aortic/mitral valve disease, which has also been shown to be

associated with elevated MPV albeit in small studies (Varol et al. 2011; Bilen et al. 2012; Yavuz et al. 2009). It has also been suggested that in the acute infarct population two groups may exist, one with normal platelet volume and count and one with increased MPV (Trowbridge and Martin 1987). This observation may also reflect the use of newer generation P2Y12 receptor antagonists such as ticagrelor. However, platelet counts were higher in the AMI group (235.6 ± 65.1 $x10^9$/L) compared to controls (198 ± 49.7 $x10^9$/L, p=0.0396; t-test; Figure 5.2b). Consequently, plateletcrit or platelet mass (product of platelet count and MPV) was also higher in the AMI group compared to controls (0.26 ± 0.06%, 0.22 ± 0.05% respectively, p=0.0206; t-test; Figure 4.2c). There is limited evidence for increase in platelet counts post-infarction, but it has been shown to correlate with poor prognosis (Ly et al. 2006). Despite no difference in MPV, patients with AMI exhibited an increased overall circulating platelet mass, suggesting an upregulation of platelet production. This would support the hypothesis of platelet consumption and increased haemostatic demand in AMI (Martin et al. 2012). A constant circulating platelet mass is physiologically maintained by an inverse relationship between MPV and platelet count. By plotting MPV against platelet count for each individual patient, this relationship is shown to be maintained in both control and AMI groups (Figure 5.3). This suggests that the same mechanisms controlling platelet homeostasis in normal physiology may also lead to the changes seen in AMI.

MPV has been shown to be elevated within 12 hours of AMI and remain elevated up to 4 months thereafter (Kishk, Trowbridge, and Martin 1985; López-Cuenca et al. 2012; Ranjith et al. 2009; Lippi et al. 2009; H. Chu et al. 2011). To establish whether time from infarction to sample collection might account for the observed between group differences in platelet parameters, time from infarction in the AMI group was plotted against MPV or platelet count for the same individual. No overall trend in MPV or platelet count was demonstrated over time (p=ns; Figure 5.4). Furthermore, use of aspirin or P2Y12 receptor antagonists within 48hours and 5 days respectively was not shown to be related to platelet volume (p=ns; Figure 5.5).

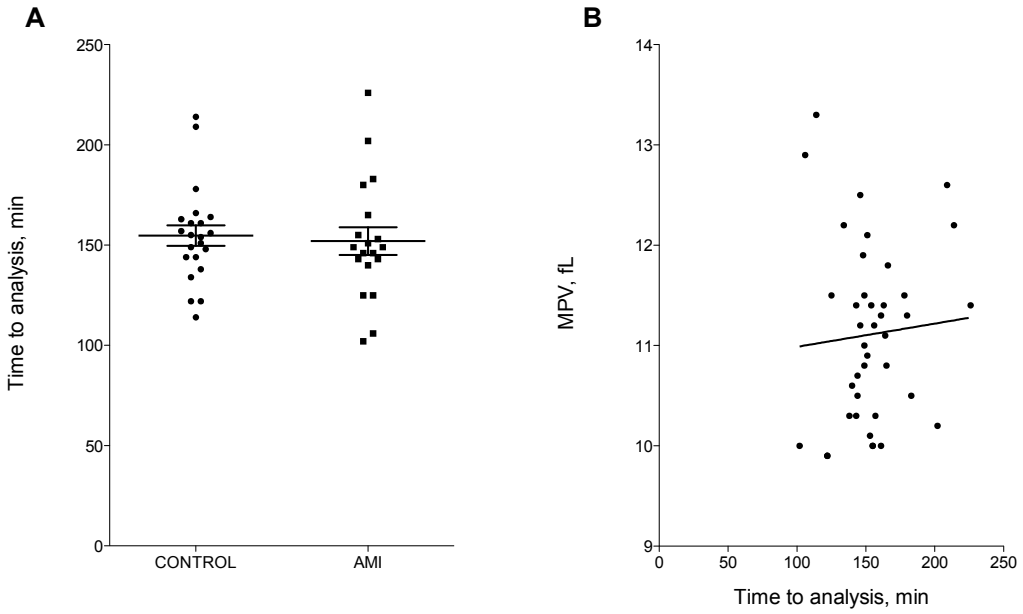


Figure 5.1 Time from blood collection to analysis on Sysmex-5000 instrument.
A. Time from blood collection to analysis in control and AMI groups (mean ±SEM shown); B. Spearman rank correlation between MPV and time from blood collection to analysis in the full cohort (both control and AMI groups) (r=0.09, p=0.5).

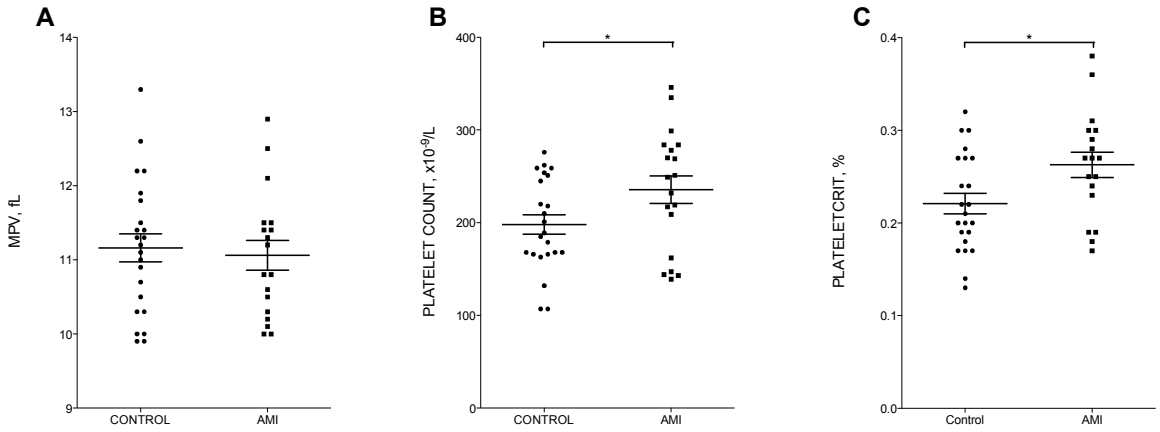


Figure 5.2 Platelet parameters
A. MPV in control and AMI groups; B. Platelet count in control and AMI groups (p=0.0396); C. Plateletcrit in control and AMI groups (p=0.0206). Mean ± SEM shown, * indicates p<0.05. (Full Sysmex data may be found in Appendix 2.)

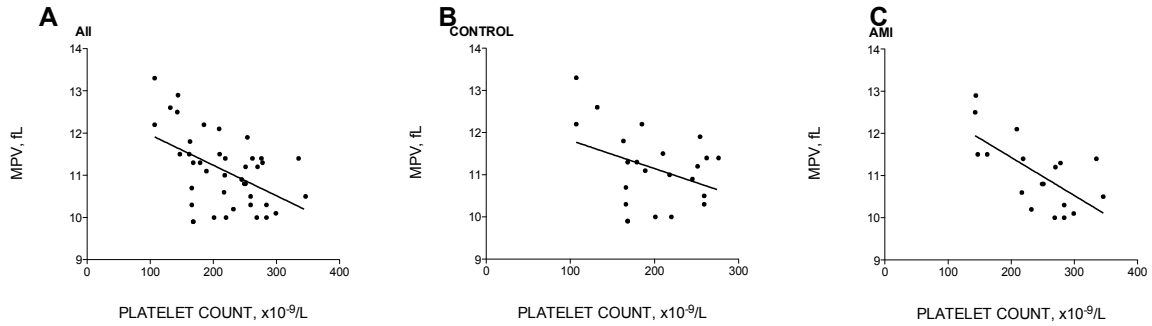


Figure 5.3 Inverse relationship between MPV and platelet count
Spearman rank correlations between MPV and platelet count for A. all individuals (r=-0.45, p=0.0032); B. Control group (r=-0.23, p=0.28), C. AMI group (r=-0.67, p=0.0023).

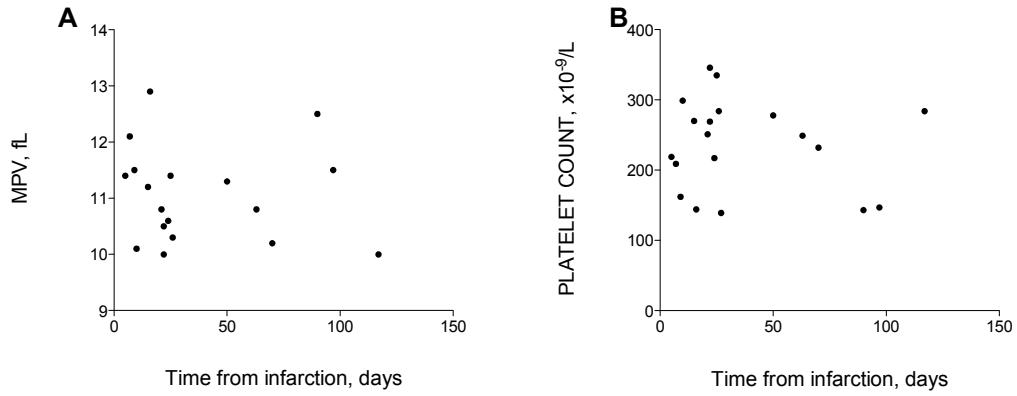


Figure 5.4 Platelet parameters according to time of sample post-infarction
A. Spearman rank correlation between MPV and time from infarction (p=ns); B. Spearman rank correlation between platelet count and time from infarction (p=ns).

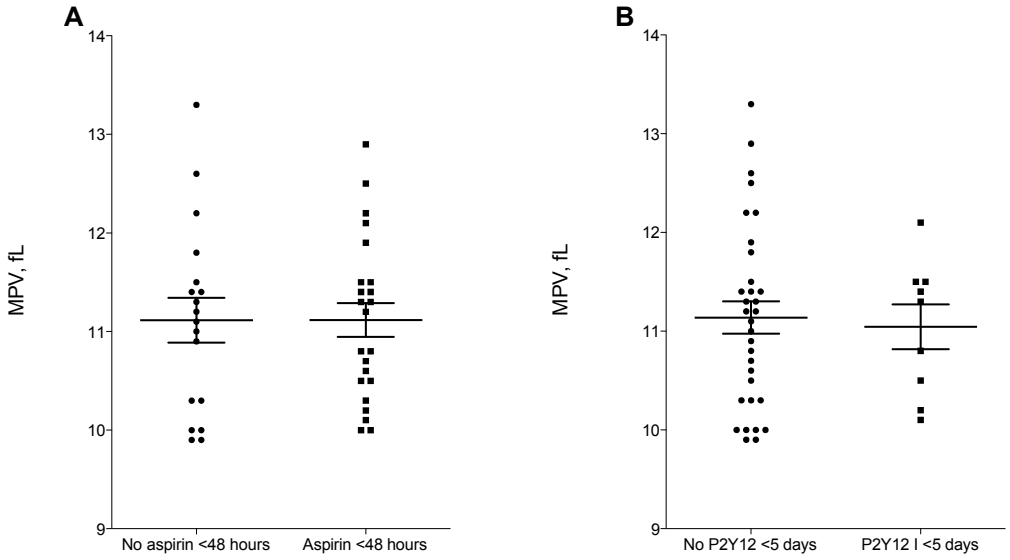


Figure 5.5 MPV according to antiplatelet usage

A. MPV in patients with aspirin use within 48 hours compared to without; B. MPV in patients with P2Y12 antagonist use within 5 days compared to without. Mean ± SEM shown. 48 hours and 5 days were used for aspirin and P2Y12 antagonists respectively according to documented off-set of antiplatelet activity (Levine et al. 2016).

5.2.3 Platelet activity

For platelet activity the 2 groups were compared to a group of anonymised healthy volunteers (n=20); see Chapter 2. These were individuals with no co-morbidities (see Section 2.1.2).

5.2.3.1 Platelet activation

Activation in resting or unstimulated platelets was assessed by measuring baseline P-selectin expression using flow cytometry. P-selectin is an α granule membrane glycoprotein in the resting platelet, it becomes externalised to the plasma membrane upon activation with various agonists. It is the most highly expressed surface marker of platelet degranulation (Ruf and Patscheke 1995). The proportion of platelets expressing P-selectin on the plasma membrane has been used as a measure of baseline platelet activation. P-selectin may be a useful marker of clinical outcome (Hollander et al. 1999; Itoh et al. 1995). In order to measure baseline platelet activity, resting levels of P-selectin was measured on the surface of platelets after blood collection into citrate tubes. All flow cytometry measurements were taken at exactly 4 hours from sampling to allow for results to be comparable between patients. Resting surface expression of P-selectin was increased in the AMI group (52.1 ± 7.23%), when compared to the control group (46.8 ± 7.36%, $p=0.044$) and to the healthy group (40.4 ± 7.19%, $p<0.0001$; t-test). The control group also showed increased resting activation level compared to the healthy group ($p=0.0088$; t-test; Figure 5.6a). Therefore, despite high levels of dual antiplatelet therapy, patients with AMI showed elevated resting platelet activation. This supports evidence that show increased surface expression of P-selectin in AMI compared with controls in previous studies (Ault et al. 1999; Shimomura et al. 1998; Michelson 2004). To investigate whether patients with larger platelets also had increased resting surface expression of P-selectin, correlation between these two markers was plotted and no relationship between resting levels of P-selectin and platelet volume was observed (Figure 5.6b,c).

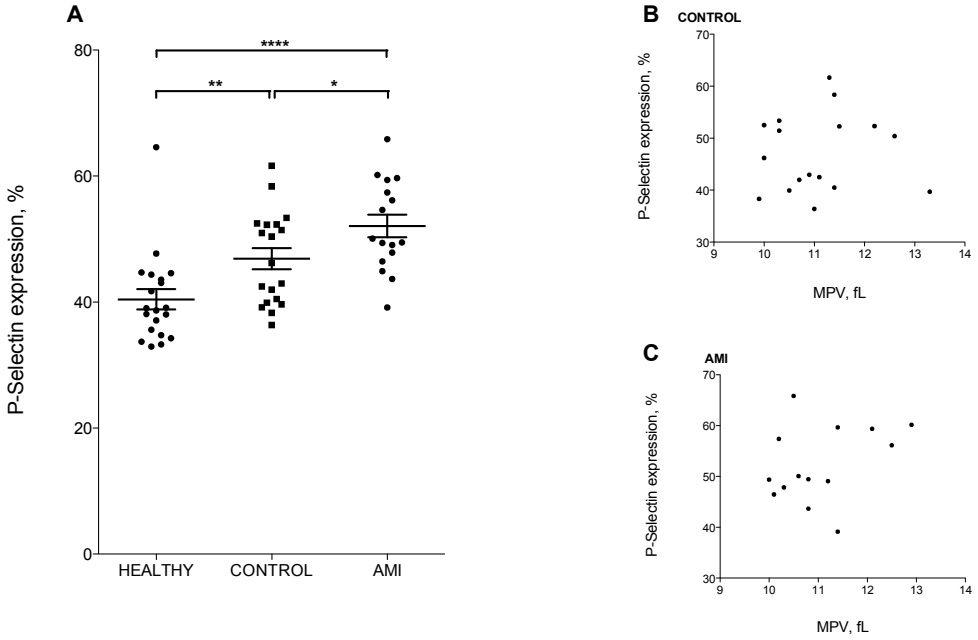


Figure 5.6 Resting platelet surface P-selectin expression

A. Proportion of platelets positive for surface P-selectin by flow cytometry in healthy, control and AMI groups (mean ± SEM shown, *indicates $p<0.05$, ** indicates $p<0.01$, ***indicates $p<0.001$, ****indicates $p<0.0001$); Spearman rank correlation between P-selectin expression and MPV in B. control group (p=ns); C. AMI group (p=ns).

5.2.3.2 Platelet function

Platelet function or reactivity in the AMI and control group were assessed by flow cytometry and compared to the healthy group, measuring change in platelet aggregatory response (fibrinogen binding) and platelet degranulation (P-selectin expression) after stimulation with the following agonists: adenosine diphosphate (ADP; P2Y12 receptor agonist), cross linked collagen related peptide (CRP-XL; GPVI receptor agonist) and thrombin receptor activating peptide (TRAP; PAR-1 receptor agonist). As described above, blood samples were collected in citrate and all flow cytometry measurements were taken in whole blood at exactly 4hours from sampling to allow for results to be comparable between individuals.

The AMI group showed a lower ADP mediated response as measured by fibrinogen binding (26.0 ± 21.92%) than the control (47.7 ± 16.1%, p=0.0019; t-test) or healthy (64.2 ± 17.2%, p<0.0001; t-test) groups (Figure 5.7a). In order to ascertain whether this was due to residual activity of P2Y12 inhibition, fibrinogen binding in response to ADP stimulation was plotted separately for patients in the AMI group who has received P2Y12 antagonist within the last 5 days compared to those that had not (Figure 5.7b). This revealed significant P2Y12 antagonist inhibition of ADP mediated fibrinogen binding (11.7 ± 11.6%) compared to the subgroup of AMI patients in which P2Y12 antagonists were stopped more than 5 days before surgery (38.3 ± 20.3%, p=0.013; t-test). This subgroup showed a similar platelet response to the control group. A similar pattern was observed when platelet degranulation was measured by P-selectin in response to ADP. Again, patients in the AMI group showed lower ADP mediated response than the control group (28.5 ± 12.5% compared with 36.0 ± 10.8%, p=0.06; t-test) and the healthy group (49.7 ±6.7, p<0.0001; t-test; Figure 5.8a). This was largely due to recent P2Y12 inhibition (Figure 5.8b), although the difference in ADP mediated P-selectin expression between AMI patients with and without recent P2Y12 antagonist use did not reach significance.

There were no significant differences between groups in CRP-XL mediated platelet activation both in terms of fibrinogen binding or expression of P-selectin (Figure 5.7c, 5.8c) although patients with recent use of P2Y12 antagonist showed a trend towards reduced responsiveness (Figure 5.7d, 5.8d). Similarly to ADP stimulation, fibrinogen binding mediated by TRAP-6 was lower in the AMI group (3.3 ± 3.8%) than healthy (12.1 ± 12.8%, p=0.013; t-test; Figure 4.7e) largely caused by inhibition of platelet aggregatory response by recent P2Y12 use (5.1 ± 3.2% compared with 1.0 ± 1.1%, p=0.0101; t-test; Figure 5.7f).

The relationship between MPV and markers of platelet reactivity was then studied (Figure 5.9, 5.10). No correlation was found between any marker of platelet reactivity and MPV for any of the groups.

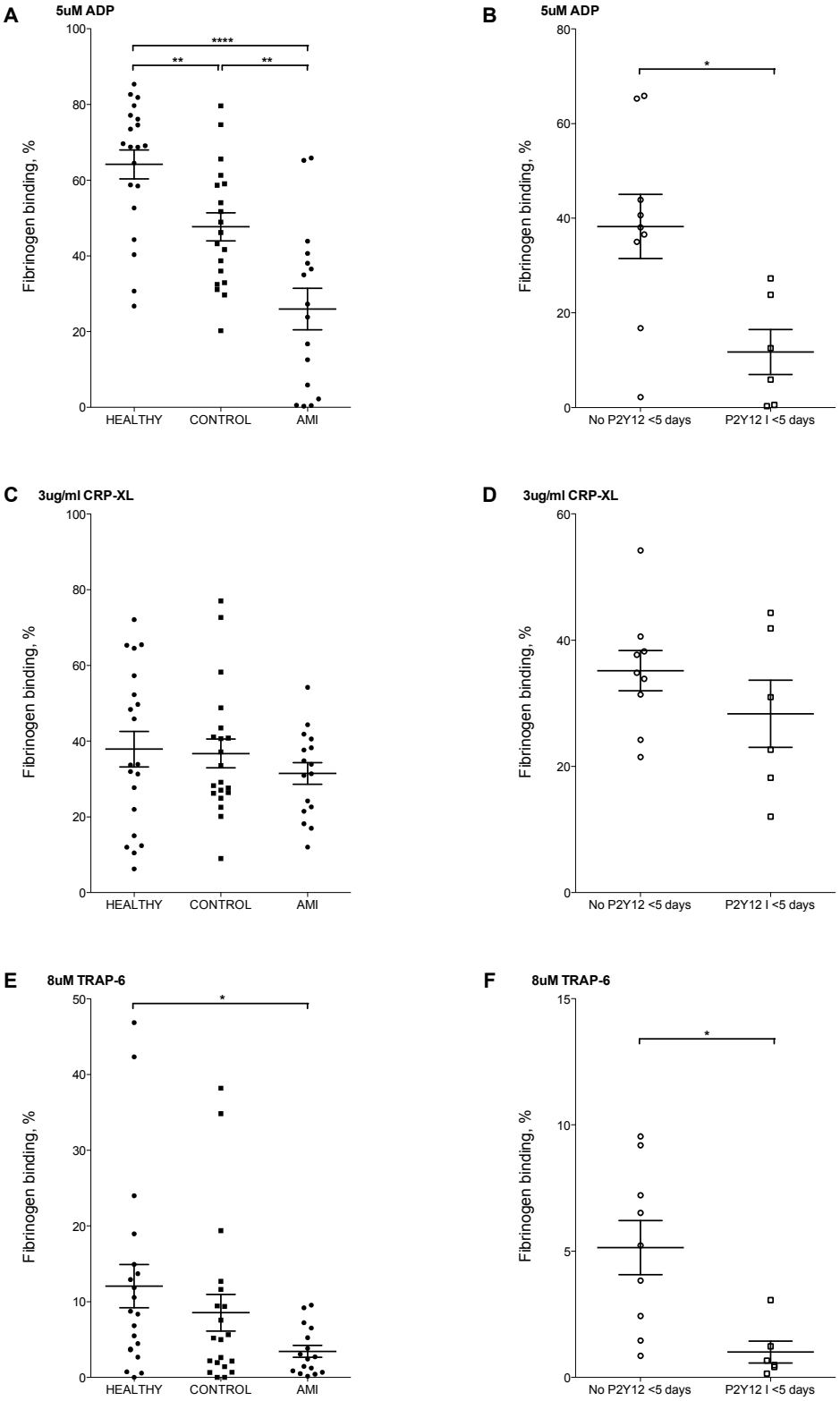


Figure 5.7 Platelet function measured by change in fibinogen binding

Fibrinogen binding response measured as % positive platelets in whole blood to 5uM ADP (A,B); 3ug/ml CRP-XL (C,D); 8uM TRAP-6 (E,F). Difference in response between healthy, control and AMI groups (A,C,E); Difference in response between patients with P2Y12 antagonist use within 5 days and without (B,D,F). Mean ± SEM shown. *indicates $p<0.05$, ** indicates $p<0.01$, ***indicates $p<0.001$, ****indicates $p<0.0001$.

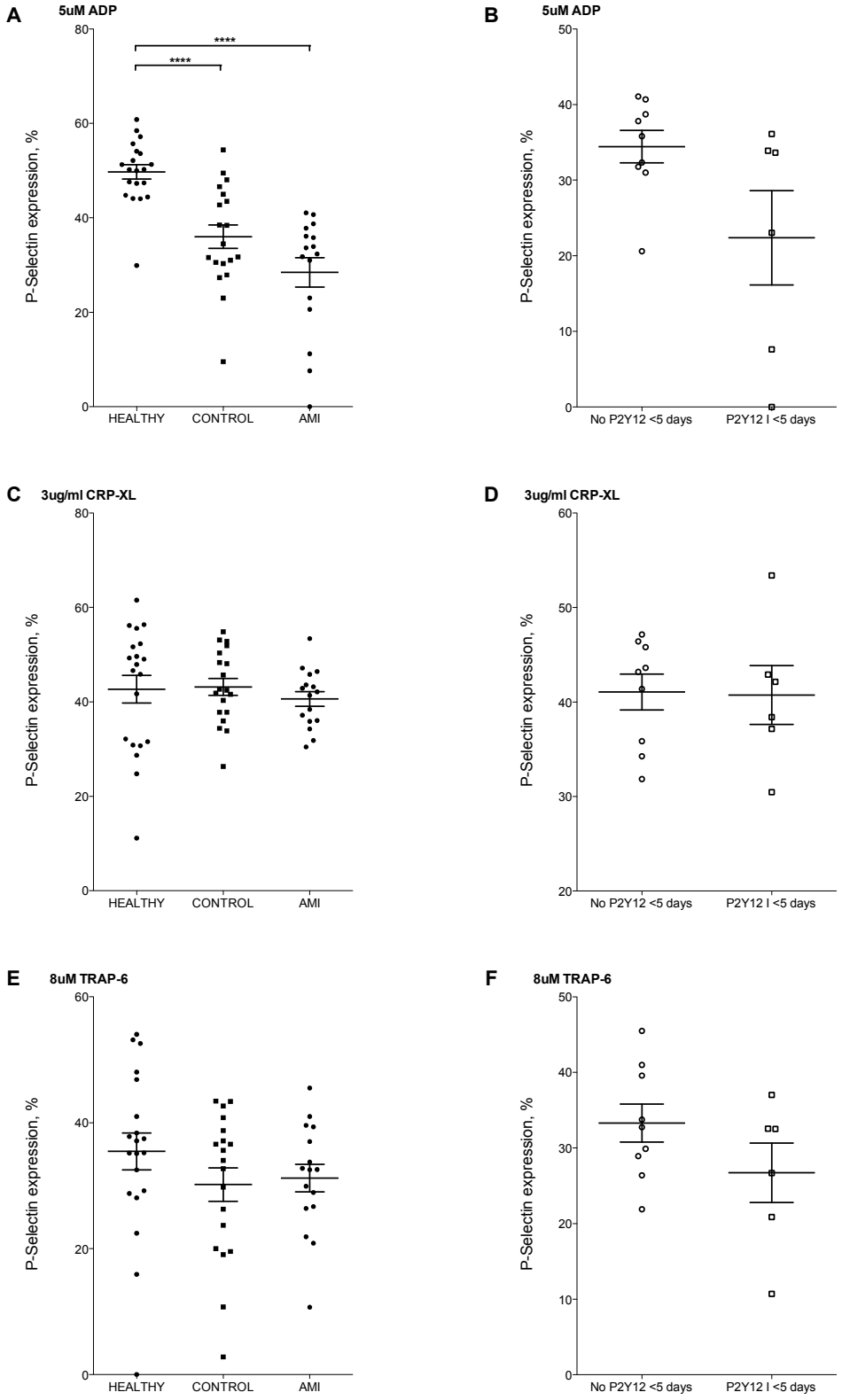


Figure 5.8 Platelet function measured by change in P-selectin expression
P-selectin expression response measured as % positive platelets in whole blood to 5uM ADP (A,B); 3ug/ml CRP-XL (C,D); 8uM TRAP-6 (E,F). Difference in response between healthy, control and AMI groups (A,C,E); Difference in response between patients with P2Y12 antagonist use within 5 days and without (B,D,F). Mean ± SEM shown. *indicates p<0.05, ** indicates p<0.01, ***indicates p<0.001, ****indicates p<0.0001.

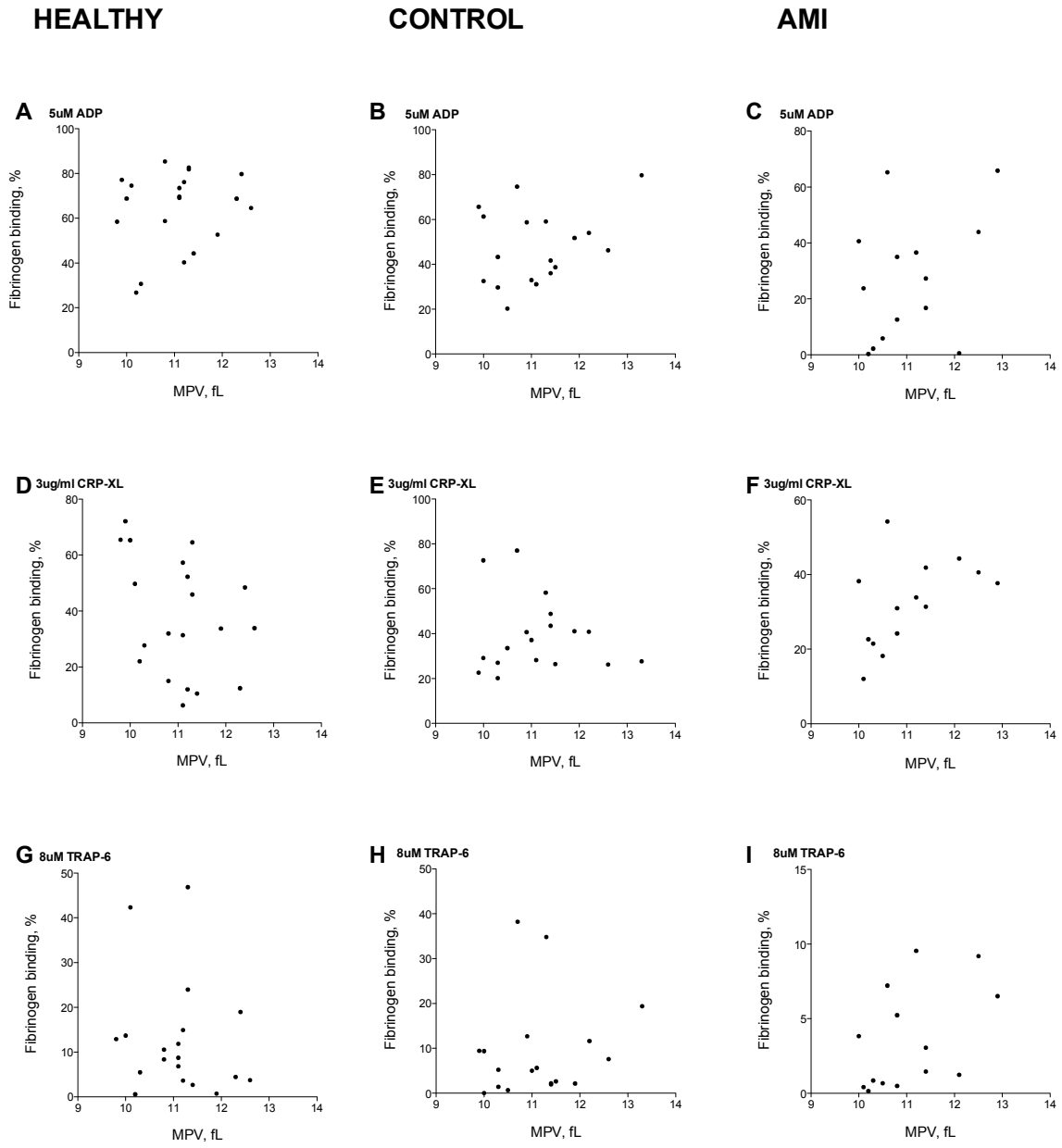


Figure 5.9 Relationship between MPV and platelet function measured by change in fibrinogen binding

Spearman rank correlations between MPV and fibrinogen binding platelet response after stimulation with 5uM ADP (A,B,C); 3ug/ml CRP-XL (D,E,F); 8uM TRAP-6 (E,F,G). Healthy group (A,D,G); Control group (B,E,H); AMI group (C,F,I). No significant relationship was found between MPV and platelet function in any of the groups. Pooling all groups together also showed no significant relationship.

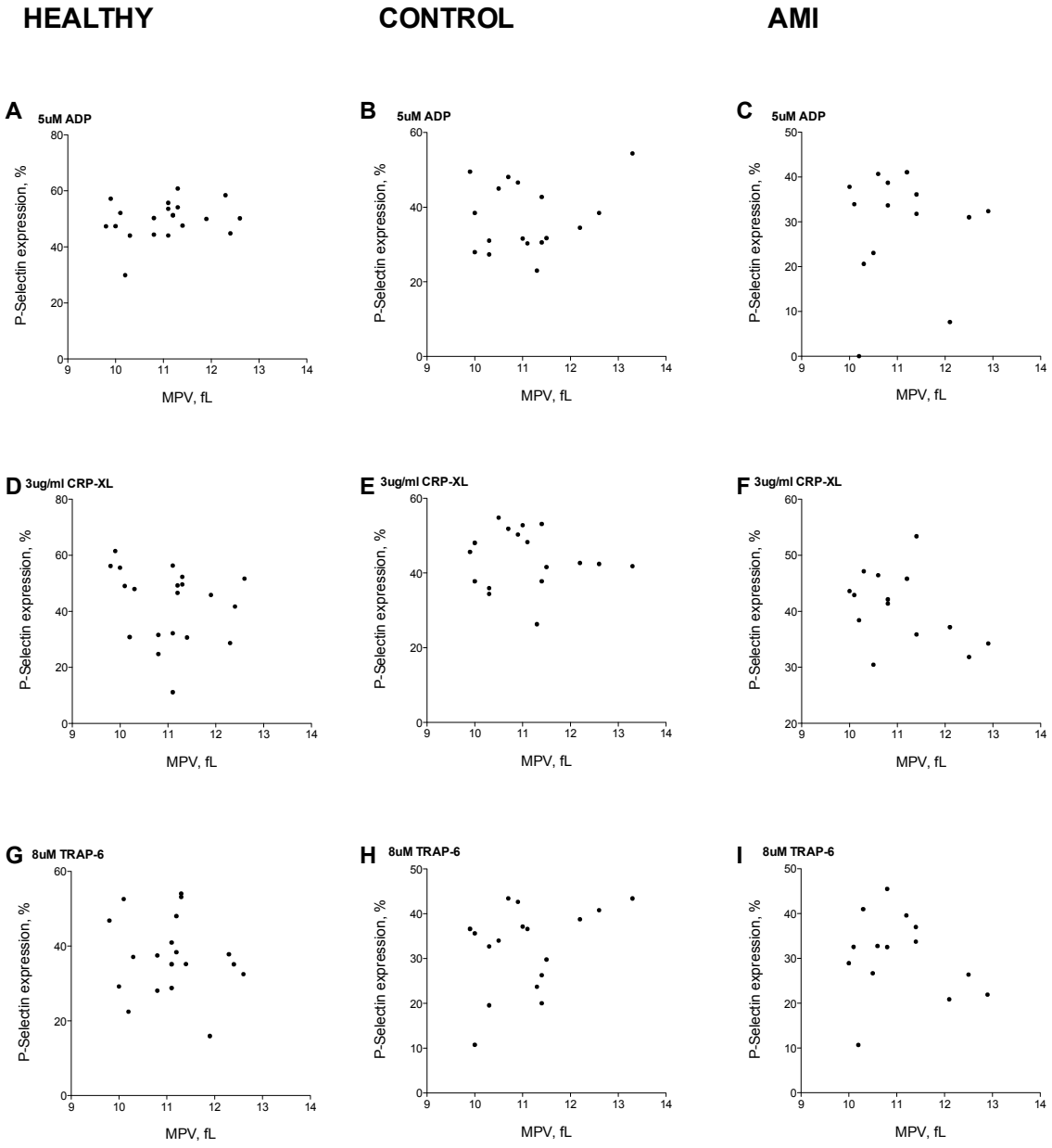


Figure 5.10 Relationship between MPV and platelet function measured by change in P-selectin expression

Spearman rank correlations between MPV and P-selectin platelet response after stimulation with 5uM ADP (A,B,C); 3ug/ml CRP-XL (D,E,F); 8uM TRAP-6 (E,F,G). Healthy group (A,D,G); Control group (B,E,H); AMI group (C,F,I). No significant relationship was found between MPV and platelet function in any of the groups. Pooling all groups together also showed no significant relationship.

5.2.4 Megakaryocyte activity

5.2.4.1 Platelet production

Platelet production and megakaryocyte activity were indirectly measured by the reticulated platelet fraction or “immature platelet fraction” in whole blood collected in EDTA by Sysmex. As previously described, this is the proportion of platelets that contain mRNA which is found in newly formed platelets and degrades with age. It has been used as a measure of thrombopoiesis and platelet turnover. As the technology is limited in terms of the amount of RNA that can be measured (Briggs et al. 2009) it has been suggested above that although newly formed platelets can be of any size, only large platelets contain enough RNA to enable quantification. Therefore, the reticulated platelet fraction gives an indication of number of newly formed large platelets. This explains elevated levels of reticulated platelets in increased thrombopoiesis (Briggs et al. 2004). Reticulated platelet fraction has also been shown to be increased in AMI indicating increased platelet turnover, correlated to increases in MPV (Guthikonda et al. 2008; Cesari et al. 2008) and associated with outcome in this setting (Cesari et al. 2013). Figure 5.11a demonstrates no significant difference in reticulated platelet fraction between the AMI group (3.24 ± 2.1%) and the control group (3.49 ± 2.3%, p=ns; t test). This was also the case when the AMI group was divided according to the use of P2Y12 antagonist use within 5 days (data not shown). Reticulated platelet fraction and MPV are shown to be highly correlated in this dataset (r=0.76, p<0.0001; Spearman Rank correlation; Figure 5.11b) as well as when divided into control group (r=0.59, p=0.0031) and AMI group (r=0.82, p<0.0001). This is in keeping with previous studies (Grove, Hvas, and Kristensen 2008).

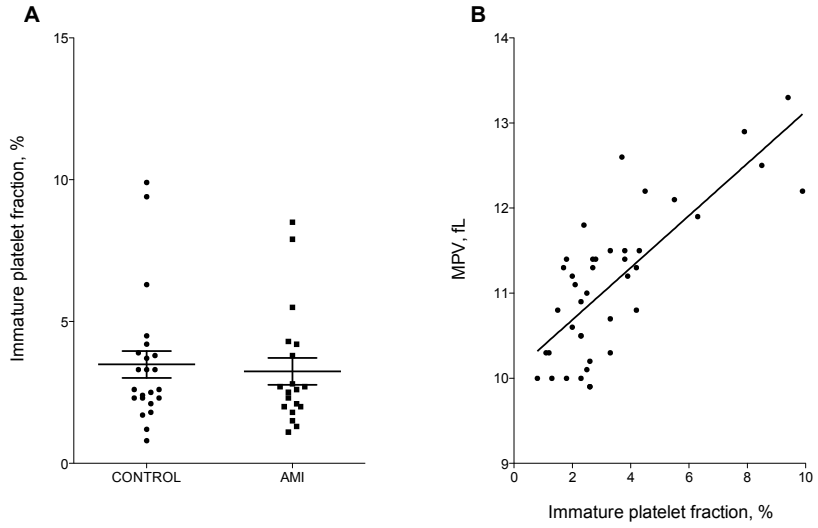


Figure 5.11 Reticulated platelet fraction (Immature platelet fraction)
A. Reticulated platelet fraction in control and AMI groups, mean ± SEM shown; B. Spearman rank correlation between reticulated platelet fraction and MPV for all individuals (r=-0.76, p<0.0001).

5.2.4.2 Megakaryocyte frequency and ploidy

No direct measurements of megakaryocyte ploidy have been made in AMI patients although there are data from few small studies in normal healthy controls and altered platelet counts (ref). These have shown a shift to higher ploidy levels in patients with low platelet counts and raised MPV, which is hypothesised to be the case in AMI if this state is associated with increased or altered haemostasis. MPV and reticulated platelet fraction were not found to be raised in the AMI group, although platelet count and overall platelet mass was significantly increased compared to controls (see above). Figure 5.13a shows that there was no significant difference in overall megakaryocyte fraction of nucleated bone marrow cells in the AMI group (0.1 ± 0.05%) compared with the control group (0.09 ± 0.05%, p=ns; t-test).

Ploidy levels were measured as proportion of cells with 4N, 8N, 16N, 32N genomic DNA content. 2N levels were not measured, as there were some questions as to whether this population was made up of purely megakaryocytes or that it could be contaminated with any other 2N bone marrow cell with megakaryocyte fragments or platelets stuck on the surface. Megakaryocytes above 32N were also not quantified as detection of these cells by flow was inconsistent suggesting that either these cells were fragmenting at some point either during staining or FACS sorting. The ploidy distribution observed in all individuals reflected previous studies with a mean ploidy of 16N and a normal distribution around this point (Figure 5.12d, 5.13b). Although there was a trend towards a ploidy shift to higher ploidy levels in the AMI group this did not reach statistical significance (p=ns) (Figure 5.13b), and it was also not affected by use of antiplatelet drugs (data not shown).

Furthermore, when investigating the relationship between fraction of 32N megakaryocytes and either MPV or platelet counts in the same individuals, no significant correlation was found (p=ns) (Figure 5.13c, 5.13d). This was also true if patients were subdivided by group, or comparisons were made with fraction of megakaryocytes in the nucleated bone marrow compartment, 4N, 8N or 16N cells (data not shown). Fraction of high ploidy megakaryocytes (32N or sum of 16N and 32N) was also not seen to correlate with resting platelet activation marker P-selectin or platelet reactivity as stimulated by ADP, CRP-XL or TRAP-6 (data not shown).

It is possible that this lack of difference in megakaryocyte phenotype between the AMI group and controls may be related to experimental limitations, due to the difficulty of working with megakaryocytes. This provided further rationale for study of both megakaryocyte and platelet transcriptome in disease and control groups.

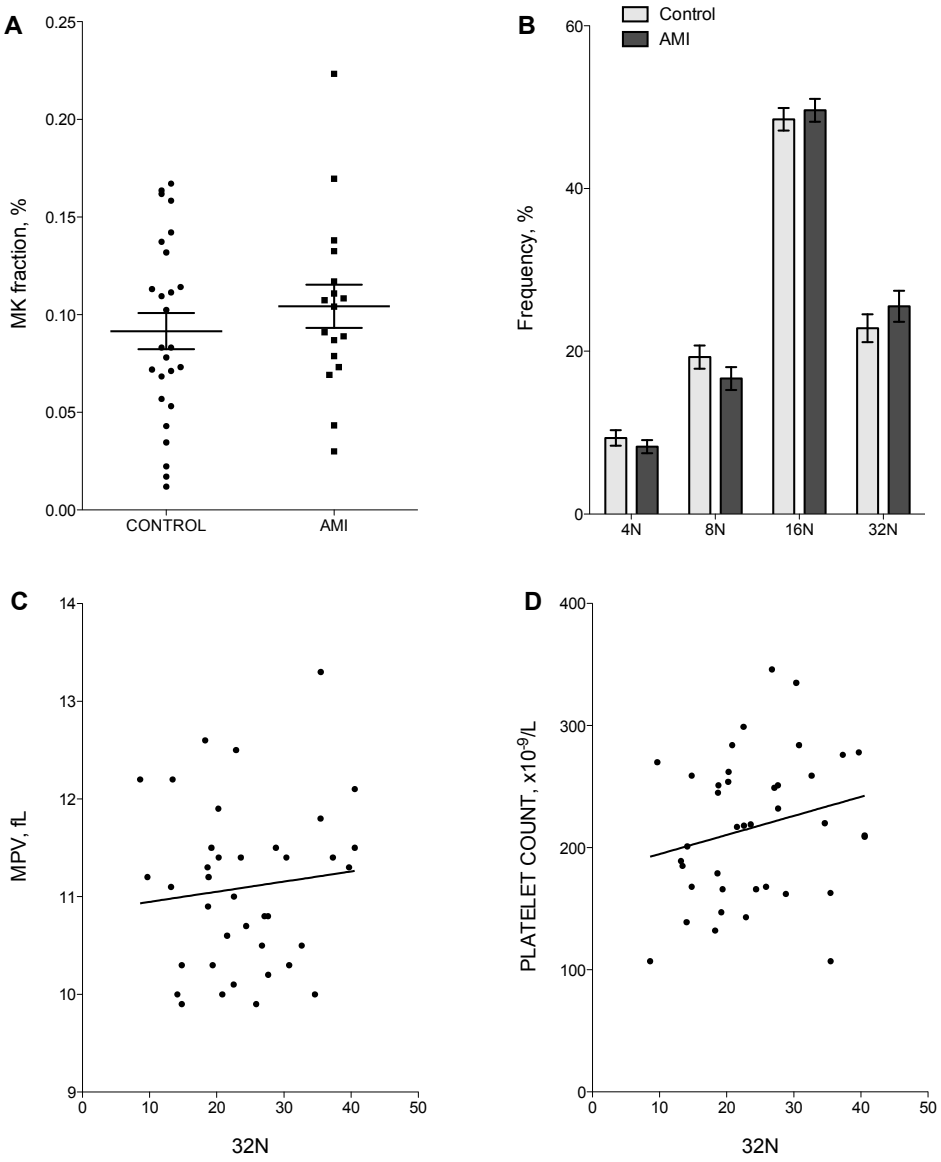


Figure 5.13 Megakaryocyte ploidy

A. Fraction of megakaryocytes within the nucleated bone marrow compartment in control and AMI groups, mean ± SEM shown; B. Megakaryocyte ploidy distribution in control and AMI groups, mean ± SEM shown; Spearman rank correlation between 32N ploidy fraction and C. MPV; D. Platelet count; for all individuals (No significant relationship was found).

5.3 Results: Disease associated transcriptional signatures

5.3.1 Disease associated megakaryocyte transcriptional signatures

5.3.1.1 Patient characteristics

Megakaryocyte pooled samples (multicellular pools) consisting of 20-100cells were collected from 15 individuals undergoing cardiac surgery, including 8 from the control group and 7 patients from with recent AMI (Appendix 1). Table 5.2 shows the baseline characteristics, platelet parameters and function for each study group, including differences between the groups.

5.3.1.2 RNA-seq library preparation: sample selection

Multicellular pools were collected for 3 patients as 20 megakaryocytes by ploidy: 2N, 4N, 8N, 16N, 32N (AMI: Patient ID 318, Control: Patients ID 0DX, 1FP). These were sequenced without qPCR based filtering or cell selection (as explained in Section 4.2.1.5). For the purposes of RNA-seq analysis the data from these ploidy multicellular pools was combined per patient ID.

For the remaining 12 patients: Control group n=6, AMI group n=6, multicellular pools were FACS sorted as 50-100 cells per well irrespective of megakaryocyte ploidy. Here Taqman real-time PCR assays for the presence of *MT-RNR2* and *PPBP* in cDNA of multicellular pools. This allowed the detection of multicellular pools with a ΔCT between *PPBP* and *MT-RNR2* of less than 7. As previously explained, the aim was to avoid overwhelming representation of mitochondrial transcripts in the resulting RNA-seq data. Figure 5.14 demonstrates the ΔCT levels for all multicellular pools for 6 individuals from the control group and 6 patients from the AMI group. A total of 156 multicell pools went on to be sequenced based on these criteria. To validate the use of these criteria in the identification of high quality samples for RNA-seq, Taqman real-time qPCR for the presence of *B2M* was used in 8/12 patients and correlated with ΔCT *PPBP* – *MT-RNR2*. Figure 5.15 demonstrates a correlation between these 2 measurements (r^2=0.1026).

For the selected 156 multicellular pools, 125bp paired-end sequencing was performed at an average sequencing depth of approximately 50 million reads/sample. However, it was felt that increasing the sequencing depth might improve quality of RNA-seq data. Therefore 3 multicell pools with the lowest ΔCT *PPBP* – *MT-RNR2* were selected from each of the 12 patients to take forward for deeper 125bp paired-end sequencing at a depth of approximately 500 million reads/sample (Figure 5.16). Although these were selected as the best performing samples from each individual, it is evident from the table that the average ΔCT *PPBP* – *MT-RNR2*

varied widely between samples. The multicellular pools collected by ploidy for the remaining 3 patients were also sequenced at this depth. In terms of batch effects, Table 5.3 shows each sample plate ID and corresponding sequencing lane ID, whereas Table 5.4 outlines the batches in which samples were collected, cDNA reverse transcribed/amplified, and Nextera libraries prepared and sequenced. Note is made that RNA-seq analysis also included RNA-seq from previously discussed lanes SLX-10208/10482 for the 15 individuals.

Table 5.2 Baseline patient characteristics

	Controls n=8	AMI n=7	p value
Gender M	4	6	ns
Age (mean±SD)	60.9±18.6	67.6±5.10	ns
Hypertension	5	7	ns
Statins	2	6	0.0406*
Diabetes	1	3	ns
Smoking	0	3	ns
Aspirin within last 48 hours	3	7	0.0256*
Left ventricular dysfunction	3	2	ns
Coronary disease on angiography	-	7	-
History of AMI within 6 months	-	7	-
Time from infarction (mean±SD)	-	49.7±41.4	-
Aortic valve surgery	5	1	ns
Mitral valve surgery	3	0	0.0083**
Pre-operative Hb (mean±SD)	13.6±1.18	12.5±1.63	ns
Pre-operative WC (mean±SD)	6.94±1.65	8.81±3.72	ns
Pre-operative Plt (mean±SD)	245±35.0	260±62.6	ns
Time to sysmex, min (mean±SD)	152.8±10.0	160.6±33.8	ns
PLT, x109/L (mean±SD)	214.5±55.5	228.1±74.0	ns
MPV, fL (mean±SD)	10.89±0.84	10.97±0.88	ns
Plateletcrit, % (mean±SD)	0.23±0.056	0.26±0.060	ns
IPF, % (mean±SD)	3.13±2.88	3.30±2.46	ns
MK fraction , % (mean±SD)	0.074±0.050	0.085±0.038	ns
MK 4N frequency, % (mean±SD)	12.2±7.57	10.2±3.93	ns
MK 8N frequency, % (mean±SD)	20.3±9.12	18.0±5.47	ns
MK 16N frequency, % (mean±SD)	45.1±8.26	49.4±7.64	ns
MK 32N frequency, % (mean±SD)	22.4±10.4	22.4±4.40	ns
Resting P-selectin expression, % (mean±SD)	48.9±7.01	49.9±8.89	ns
Platelet activation:			
Fibrinogen binding, % with ADP (mean±SD)	42.9±15.6	36.6±17.9	ns
Fibrinogen binding, % with CRP-XL (mean±SD)	40.8±18.4	36.0±11.9	ns
Fibrinogen binding, % with TRAP-6 (mean±SD)	4.35±5.32	4.52±2.93	ns
P-selectin expression, % with ADP (mean±SD)	36.6±7.96	34.8±5.98	ns
P-selectin expression, % with CRP-XL (mean±SD)	44.2±7.95	41.5±8.04	ns
P-selectin expression, % with TRAP-6 (mean±SD)	27.1±12.1	32.5±6.83	ns

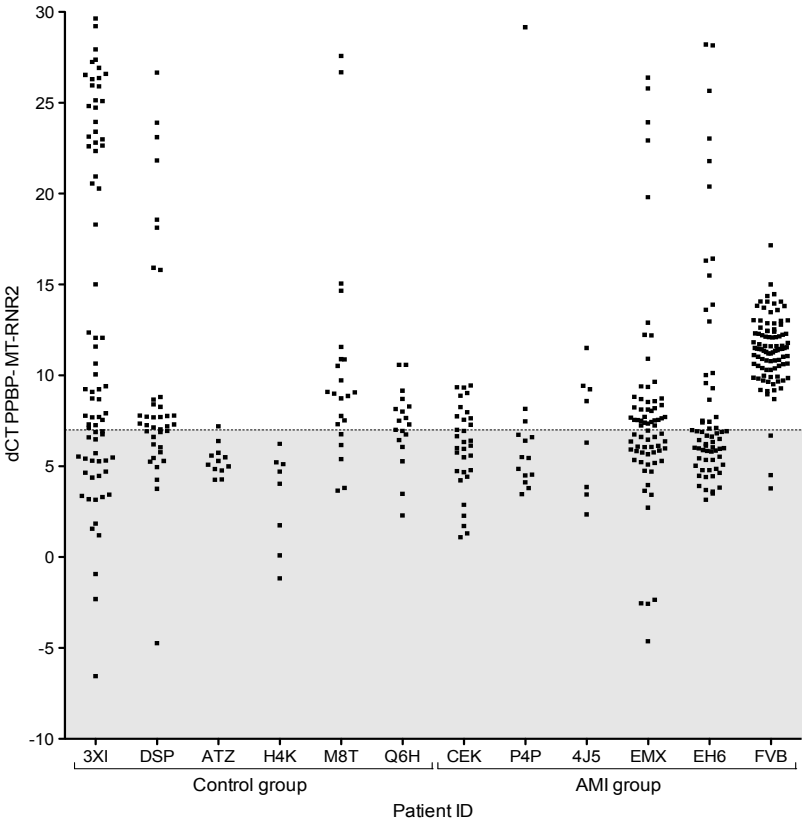


Figure 5.14 Sample selection based on ΔCT *PPBP – MT-RNR2*.
Taqman real-time PCR was used to investigate the expression of *PPBP* and *MT-RNR2* transcripts in cDNA from multicellular megakaryocyte pools of 50-100cells each from 12 individuals (Control group n=6, AMI group n=6). ΔCT *PPBP – MT-RNR2* was calculated for each sample. ΔCT *PPBP – MT-RNR2* <7 was the criteria used to identify samples for RNA-seq (marked by the horizontal gridline).

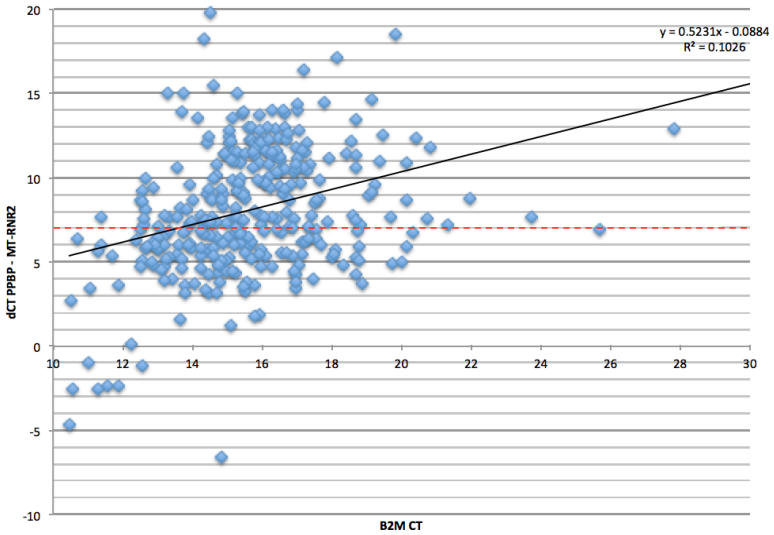


Figure 5.15 Relationship between ΔCT *PPBP – MT-RNR2* and expression of B2M.
Correlation between ΔCT *PPBP – MT-RNR2* by Taqman real-time qPCR and expression of B2M by Taqman probe in cDNA for corresponding multicellular megakaryocyte samples, r^2=0.103.

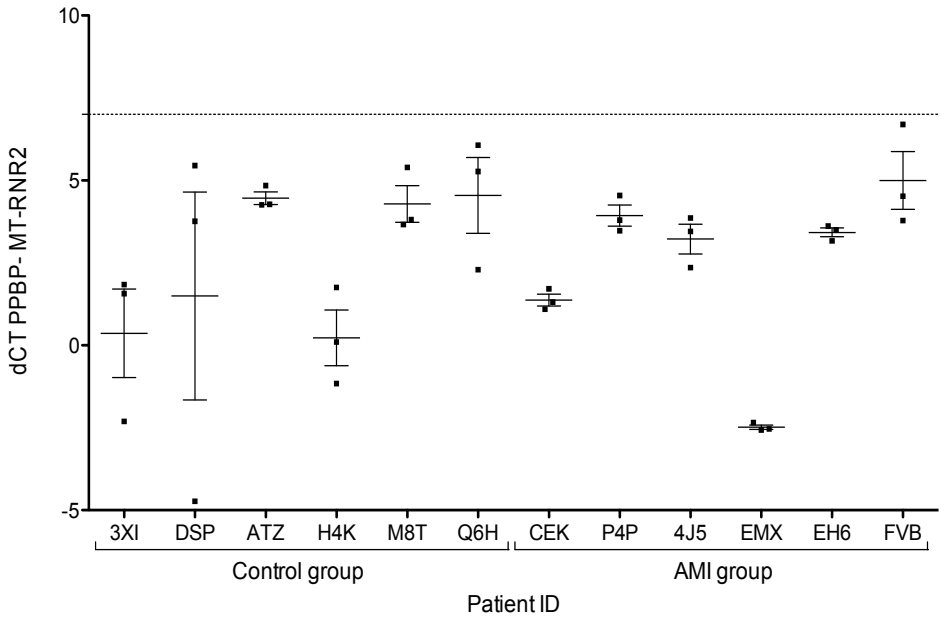


Figure 5.16 ΔCT *PPBP – MT-RNR2* for prospectively identified for deep RNA-seq. Multicell pools of 50-100 megakaryocytes each from 12 patients (Control group n=6, AMI group n=6).

Plate ID	Sequencing Lane ID
3XI	SLX-10208, SLX-10546, SLX-11601, SLX-11609
DSP	SLX-10208, SLX-10546, SLX-11602
ATZ	SLX-10208, SLX-10546, SLX-11603
H4K	SLX-10208, SLX-10546, SLX-11604, SLX-11606
M8T	SLX-10482, SLX-11605
Q6H	SLX-10546, SLX-11605
CEK	SLX-10546, SLX-10598
P4P	SLX-10547, SLX-11601
4J5	SLX-10546, SLX-11602
EMX	SLX-10547, SLX-10598
EH6	SLX-10547, SLX-11603
FVB	SLX-10546, SLX-11604
318	SLX-10208, SLX-11606, SLX-11607, SLX-11608, SLX-11609
0DX	SLX-11608
1FP	SLX-11607

Table 5.3 Patient and plate ID on each sequencing lane

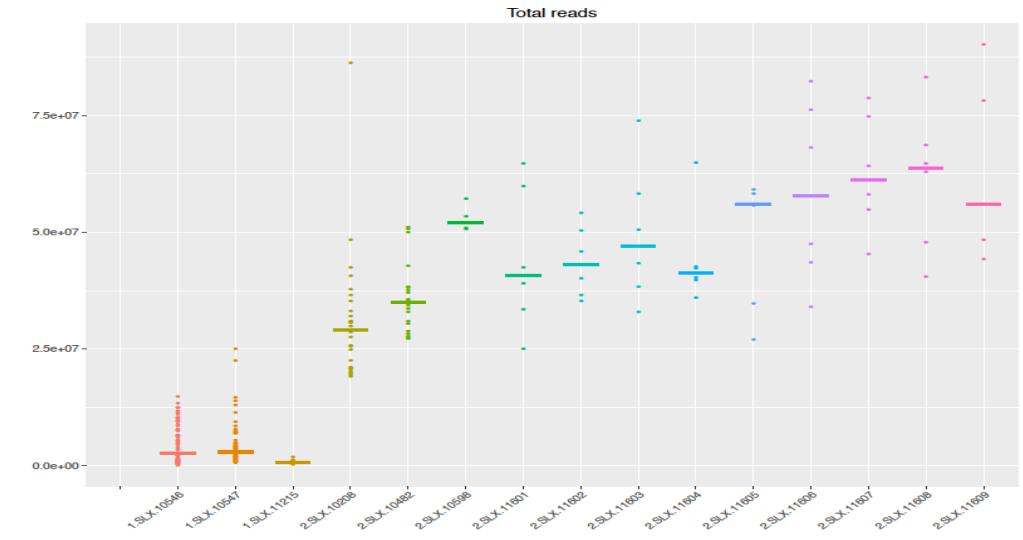
Sample collection	Amplification	Nextera Library prep / RNA-seq
1	1	1
0DX	318	SLX-10208
2	2	2
1FP	H4K	SLX-10482
3	ATZ	3
318	3	SLX-10546
4	3XI	SLX-10547
3XI	DSP	4
5	4	SLX-10598
4J5	M8T	5
6	5	SLX-11601
ATZ	0DX	SLX-11602
7	6	SLX-11603
CEK	FVB	SLX-11604
8	EMX	SLX-11605
DSP	EH6	SLX-11606
9	7	SLX-11607
EH6	CEK	SLX-11608
10	P4P	SLX-11609
EMX	Q6H	
11	8	
FVB	4J5	
12	9	
H4K	1FP	
13	3XI	
M8T		
14		
P4P		
15		
Q6H		

Table 5.4 Megakaryocyte collection, amplification, library preparation/RNA-seq batches

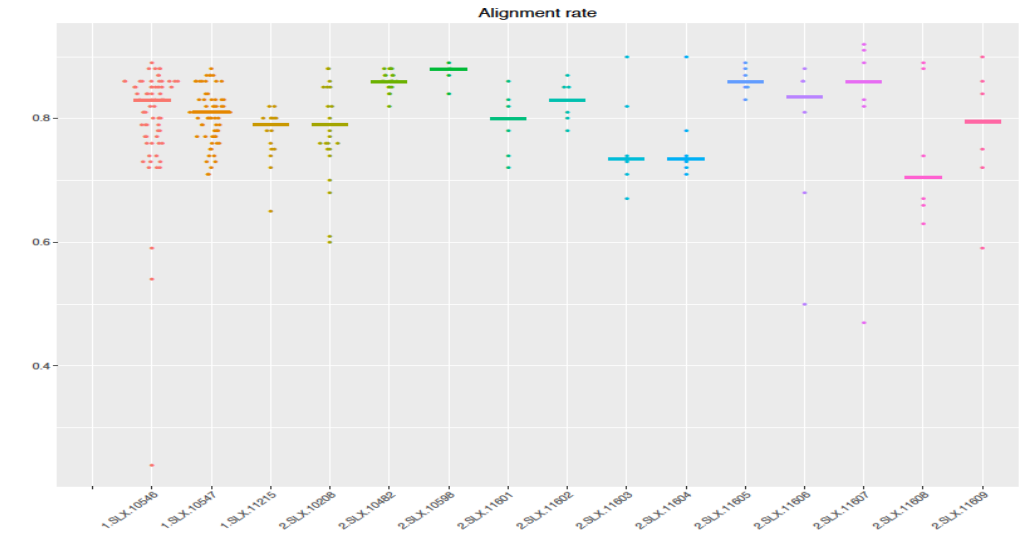
5.3.1.3 RNA sequencing data: Quality measures

Figure 5.17 shows metrics per sequencing lane for total read count, alignment to the reference transcriptome (GRCh37), proportion of reads mapped to external spike-in ERCCs compared to endogenous transcripts, no of genes called, proportion of reads mapping to annotated exonic regions and the proportion of reads mapping to the mitochondrial genome.

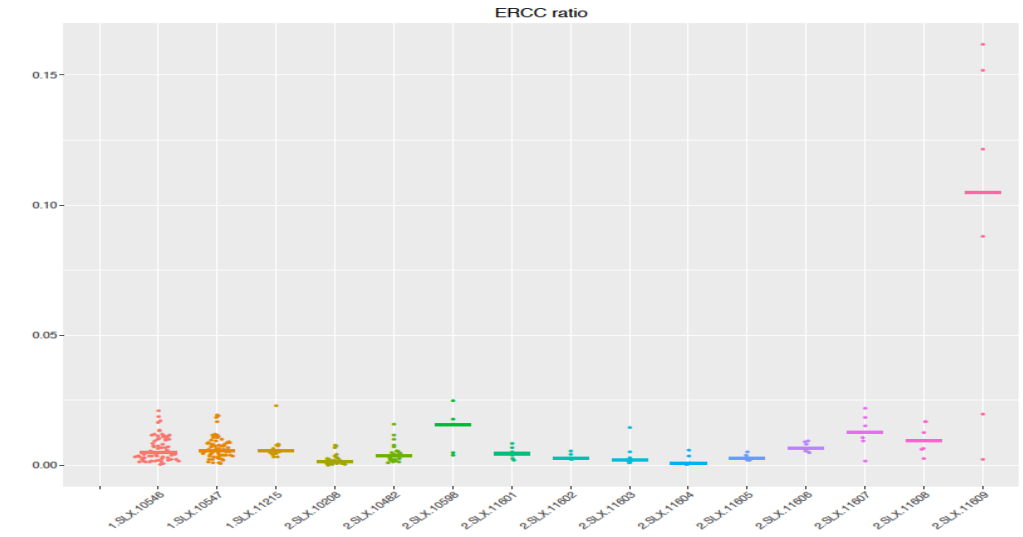
A


Total reads
7.5e+07
5.0e+07
2.5e+07
0.0e+00
1.SLX.10546
1.SLX.10547
1.SLX.11215
2.SLX.10208
2.SLX.10482
2.SLX.10598
2.SLX.11601
2.SLX.11602
2.SLX.11603
2.SLX.11604
2.SLX.11605
2.SLX.11606
2.SLX.11607
2.SLX.11608
2.SLX.11609

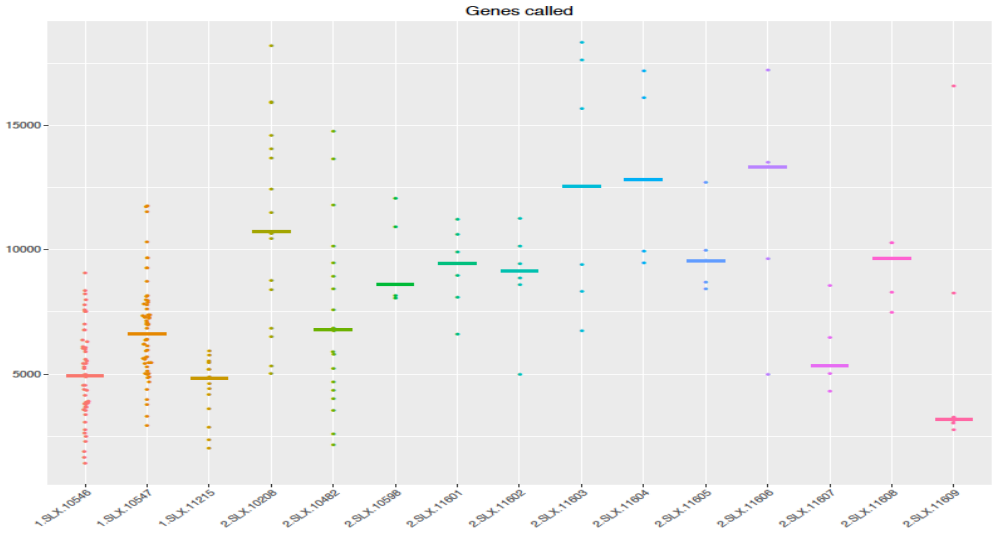
B


Alignment rate
0.8
0.6
0.4
1.SLX.10546
1.SLX.10547
1.SLX.11215
2.SLX.10208
2.SLX.10482
2.SLX.10598
2.SLX.11601
2.SLX.11602
2.SLX.11603
2.SLX.11604
2.SLX.11605
2.SLX.11606
2.SLX.11607
2.SLX.11608
2.SLX.11609

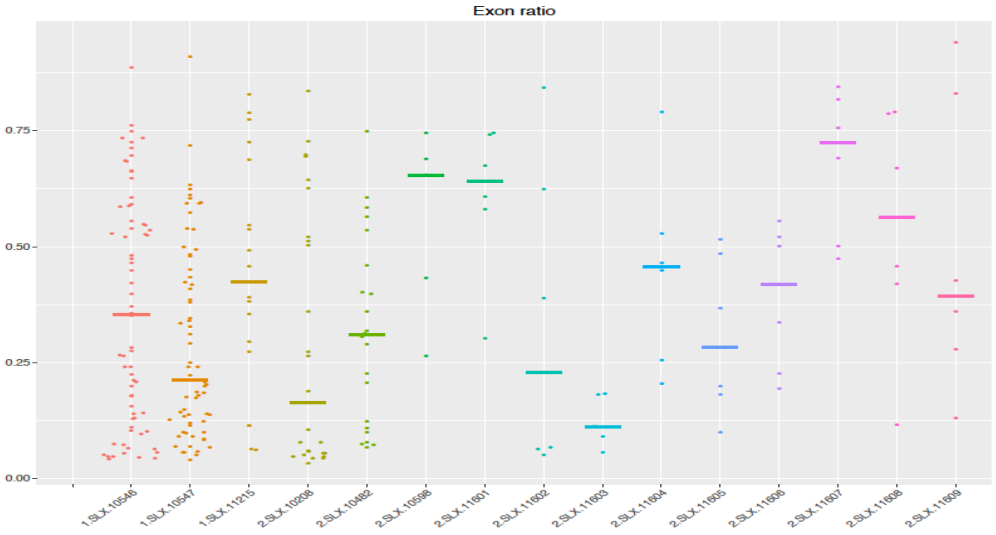
C


ERCC ratio
0.15
0.10
0.05
0.00
1.SLX.10546
1.SLX.10547
1.SLX.11215
2.SLX.10208
2.SLX.10482
2.SLX.10598
2.SLX.11601
2.SLX.11602
2.SLX.11603
2.SLX.11604
2.SLX.11605
2.SLX.11606
2.SLX.11607
2.SLX.11608
2.SLX.11609

D



E



F

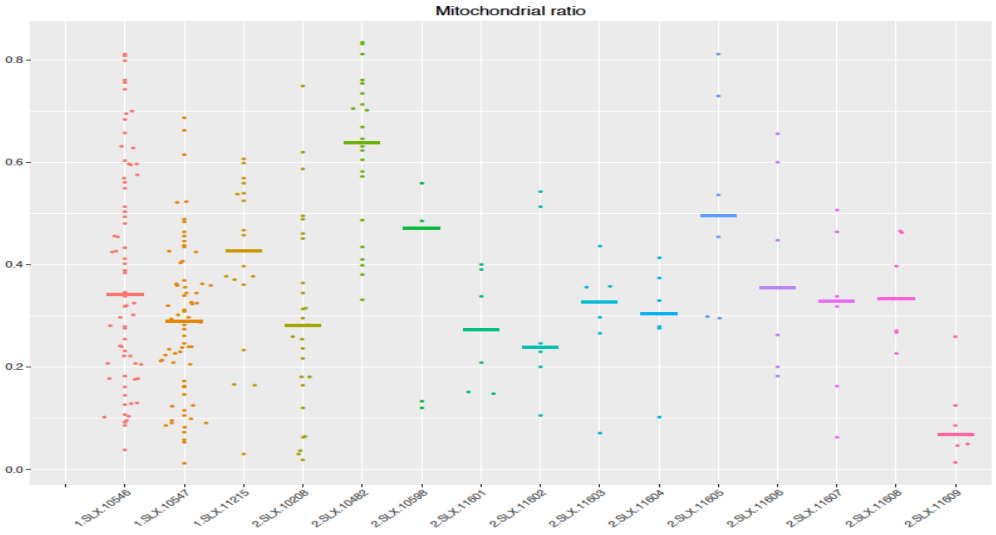


Figure 5.17 Quality control metrics/sequencing lane. A. Total reads; B. Rate of alignment to transcriptome; C. Ratio of reads mapping to ERCCs; D. No. genes called; E. Ratio of exonic reads to exonic+intronic; F. Ratio of reads mapping to mitochondrial genes.

5.3.1.4 Gene expression analysis

The top 100 highly expressed genes for the megakaryocytes in AMI group and in the control group are shown in Table 5.5. Both highly expressed gene lists showed similarities with 73 genes commonly seen between the groups (Figure 5.18). Similarly, megakaryocytes from the AMI and control groups highly expressed a number of megakaryocyte/platelet lineage genes. These included *PPBP, PF4, TMSB4X, PLEK* and its binding protein *SDPR* and *ACTB* and *CAPZA1*, both involved in cytoskeletal change. The most abundantly expressed gene list for the AMI group also contained *THBS1* and its receptor *CD47*. In keeping with the transcriptomes presented in Chapter 4, megakaryocytes from both groups highly expressed mitochondrial transcripts and genes involved in cellular ATP synthesis as well as translational machinery. There was also high expression of gene involved in the defense response including a host of immunoglobulin domains previously described in Chapter 4. These were represented in megakaryocytes from both groups. Genes in both lists were grouped together by function using GO analysis. This demonstrated similar enriched biological processes for megakaryocytes from both AMI and control groups. There was a strong over-representation of GO categories describing cellular metabolism, followed by coagulation, haemostasis, platelet degranulation and wound healing as well as translation and protein localisation. The GO results are in word cloud format in Figure 5.19 and Appendix 5, Tables S5.1, S5.2.

A differential expression analysis was then performed between the AMI and control groups to identify significantly upregulated megakaryocytes transcripts related to arterial thrombosis. In total 810 unique transcripts were differentially expressed between megakaryocytes from the AMI group and megakaryocytes from the control group. Of these 140 transcripts were shown to be significantly upregulated in recent AMI compared with control. The top 50 transcripts are shown in Table 5.6 (full list in Appendix 5, Table S5.3). Interestingly three of the significantly upregulated genes were related to platelet activation: *PPBP, THBS1 and RAP1B*. There were also a number of genes encoding transmembrane proteins: *BCAM* (basal cell adhesion molecule), *TSPAN10, TMEM130, TMEFF2, FITMB* (fat storage-inducing transmembrane protein 2). No enriched GO categories were found in genes upregulated in megakaryocytes from the AMI group, those that did not reach significance may be found in Appendix 5, Table S5.5. However, there were 11 upregulated genes encoding constant and variable immunoglobulin domains. Also represented were a number of genes involved in immunity and translation. 35 of the 140 transcripts (25%) were those that are as yet unannotated.

670 transcripts were downregulated in megakaryocytes from patients with recent AMI compared with controls (Top 50 genes in Table 5.7, Full list in Appendix 5, Table S5.4). The main functional categories that were enriched were translation and multi-organism processes including the immune response. The significantly downregulated GO terms are shown in Figure 5.20 and the full list is in Appendix 5, Table S5.6.

No	Megakaryocytes from AMI group		Megakaryocytes from control group	
	HGNC_symbol	Gene_name	HGNC_symbol	Gene_name
1	*MT-RNR2*	MT-RNR2	*MT-RNR2*	MT-RNR2
2	*ND4*	NADH dehydrogenase, subunit 4 (complex I)(ND4)	*COX1*	cytochrome c oxidase subunit I(COX1)
3	*COX1*	cytochrome c oxidase subunit I(COX1)	*MT-RNR1*	MT-RNR1
4	*MT-RNR1*	MT-RNR1	*ND4*	NADH dehydrogenase, subunit 4 (complex I)(ND4)
5	*COX2*	cytochrome c oxidase subunit II(COX2)	*COX2*	cytochrome c oxidase subunit II(COX2)
6	*CYTB*	cytochrome b(CYTB)	*ATP6*	ATP synthase F0 subunit 6(ATP6)
7	*COX3*	cytochrome c oxidase III(COX3)	*CYTB*	cytochrome b(CYTB)
8	*ATP6*	ATP synthase F0 subunit 6(ATP6)	*COX3*	cytochrome c oxidase III(COX3)
9	*ND1*	NADH dehydrogenase, subunit 1 (complex I)(ND1)	*ND1*	NADH dehydrogenase, subunit 1 (complex I)(ND1)
10	*ND2*	MTND2(ND2)	*ND2*	MTND2(ND2)
11	*HBB*	hemoglobin subunit beta(HBB)	*HBB*	hemoglobin subunit beta(HBB)
12	*ND5*	NADH dehydrogenase, subunit 5 (complex I)(ND5)	*ND5*	NADH dehydrogenase, subunit 5 (complex I)(ND5)
13	*ERCC-00074*		*ERCC-00074*	
14	*ND4L*	NADH dehydrogenase, subunit 4L (complex I)(ND4L)	*ERCC-00046*	
15	*ND3*	NADH dehydrogenase, subunit 3 (complex I)(ND3)	*ND4L*	NADH dehydrogenase, subunit 4L (complex I)(ND4L)
16	*MALAT1*	metastasis associated lung adenocarcinoma transcript 1 (MALAT1)	*S100A8*	S100 calcium binding protein A8(S100A8)
17	*S100A8*	S100 calcium binding protein A8(S100A8)	*ND3*	NADH dehydrogenase, subunit 3 (complex I)(ND3)
18	*PPBP*	pro-platelet basic protein(PPBP)	*S100A9*	S100 calcium binding protein A9(S100A9)
19	*S100A9*	S100 calcium binding protein A9(S100A9)	*MALAT1*	metastasis associated lung adenocarcinoma transcript 1(MALAT1)
20	*MTCO1P12*	MTCO1P12	*MTATP6P1*	MTATP6P1
21	*B2M*	beta-2-microglobulin(B2M)	*ND6*	NADH dehydrogenase, subunit 6 (complex I)(ND6)
22	*ERCC-00046*		*B2M*	beta-2-microglobulin(B2M)
23	*HBA2*	hemoglobin subunit alpha 2(HBA2)	*TMSB4X*	thymosin beta 4, X-linked(TMSB4X)
24	*MTATP6P1*	MTATP6P1	*HBD*	hemoglobin subunit delta(HBD)
25	*TMSB4X*	thymosin beta 4, X-linked(TMSB4X)	*ERCC-00002*	
26	*PF4*	platelet factor 4(PF4)	*IGHG2*	IGHG2
27	*ATP8*	ATP synthase F0 subunit 8(ATP8)	*ACTB*	actin beta(ACTB)
28	*PPP1CB*	protein phosphatase 1 catalytic subunit beta(PPP1CB)	*PPBP*	pro-platelet basic protein(PPBP)
29	*MTCO1P12*	MTCO1P12	*MTCO1P12*	MTCO1P12
30	*DEFA3*	defensin alpha 3(DEFA3)	*HBA2*	hemoglobin subunit alpha 2(HBA2)
31	*IGLC2*	IGLC2	*ERCC-00004*	
32	*ND6*	NADH dehydrogenase, subunit 6 (complex I)(ND6)	*ERCC-00130*	
33	*ACTB*	actin beta(ACTB)	*ERCC-00096*	
34	*HNRNPH1*	heterogeneous nuclear ribonucleoprotein H1 (H)(HNRNPH1)	*PF4*	platelet factor 4(PF4)
35	*SRGN*	serglycin(SRGN)	*JCHAIN*	joining chain of multimeric IgA and IgM(JCHAIN)
36	*GNG11*	G protein subunit gamma 11(GNG11)	*FTL*	ferritin light chain(FTL)
37	*CALM2*	calmodulin 2(CALM2)	*DEFA3*	defensin alpha 3(DEFA3)
38	*RAP1B*	RAP1B, member of RAS oncogene family(RAP1B)	*MTCO1P12*	MTCO1P12
39	*ERCC-00004*		*IGHA1*	IGHA1
40	*TLK1*	tousled like kinase 1(TLK1)	*NEAT1*	nuclear paraspeckle assembly transcript 1 (non-protein coding)(NEAT1)
41	*ERCC-00002*		*ATP8*	ATP synthase F0 subunit 8(ATP8)
42	*HBD*	hemoglobin subunit delta(HBD)	*CLC*	Charcot-Leyden crystal galectin(CLC)
43	*LIMS1*	LIM zinc finger domain containing 1(LIMS1)	*SRGN*	serglycin(SRGN)
44	*S100A12*	S100 calcium binding protein A12(S100A12)	*HSP90AA1*	heat shock protein 90 alpha family class A member 1(HSP90AA1)
45	*NEAT1*	nuclear paraspeckle assembly transcript 1 (non-protein coding)(NEAT1)	*ACTG1*	actin gamma 1(ACTG1)
46	*DAAM1*	dishevelled associated activator of morphogenesis 1(DAAM1)	*IGLC2*	IGLC2
47	*FTL*	ferritin light chain(FTL)	*SDPR*	serum deprivation response(SDPR)
48	*CAPZA1*	capping actin protein of muscle Z-line alpha subunit 1(CAPZA1)	*MTND2P28*	MTND2P28
49	*YWHAZ*	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein zeta(YWHAZ)	*H2AFZ*	H2A histone family member Z(H2AFZ)
50	*ERCC-00096*		*LYZ*	lysozyme(LYZ)
51	*JCHAIN*	joining chain of multimeric IgA and IgM(JCHAIN)	*PRG3*	proteoglycan 3, pro eosinophil major basic protein 2(PRG3)
52	*SRF3*	serine and arginine rich splicing factor 3(SRSF3)	*LOC105376575*	uncharacterized LOC105376575(LOC105376575)
53	*ERCC-00130*		*RPL21*	ribosomal protein L21(RPL21)
54	*IGLV3-21*	IGLV3-21	*CALM2*	calmodulin 2(CALM2)
55	*SAR1A*	secretion associated Ras related GTPase 1A(SAR1A)	*LTF*	lactotransferrin(LTF)
56	*MNDA*	myeloid cell nuclear differentiation antigen(MNDA)	*EEF1A1*	eukaryotic translation elongation factor 1 alpha 1(EEF1A1)
57	*HMGB1*	high mobility group box 1(HMGB1)	*PPP1CB*	protein phosphatase 1 catalytic subunit beta(PPP1CB)
58	*IGHG2*	IGHG2	*HMGB1*	high mobility group box 1(HMGB1)
59	*MIR1244-1*	microRNA 1244-1(MIR1244-1)	*CAMP*	cathelicidin antimicrobial peptide(CAMP)
60	*ACTG1*	actin gamma 1(ACTG1)	*RPLP1*	ribosomal protein lateral stalk subunit P1(RPLP1)
61	*NAP1L1*	nucleosome assembly protein 1 like 1(NAP1L1)	*ERCC-00113*	
62	*IGKC*	IGKC	*GNG11*	G protein subunit gamma 11(GNG11)
63	*H2AFZ*	H2A histone family member Z(H2AFZ)	*MTND1P23*	MTND1P23

64	*MT-TF*	MT-TF	*SAR1A*	secretion associated Ras related GTPase 1A(SAR1A)
65	*CDC42*	cell division cycle 42(CDC42)	*LIMS1*	LIM zinc finger domain containing 1(LIMS1)
66	*LYZ*	lysozyme(LYZ)	*ITM2B*	integral membrane protein 2B(ITM2B)
67	*TPT1*	tumor protein, translationally-controlled 1(TPT1)	*ERCC-00136*	
68	*IGHA1*	IGHA1	*MIR1244-1*	microRNA 1244-1(MIR1244-1)
69	*SAT1*	spermidine/spermine N1-acetyltransferase 1(SAT1)	*S100A12*	S100 calcium binding protein A12(S100A12)
70	*PLEK*	pleckstrin(PLEK)	*RPL9*	ribosomal protein L9(RPL9)
71	*HSP90AA1*	heat shock protein 90 alpha family class A member 1(HSP90AA1)	*SRF3*	serine and arginine rich splicing factor 3(SRSF3)
72	*ARPC5*	actin related protein 2/3 complex subunit 5(ARPC5)	*RPS13*	ribosomal protein S13(RPS13)
73	*RPL1*	ribosomal protein lateral stalk subunit P1(RPLP1)	*HLA-E*	major histocompatibility complex, class I, E(HLA-E)
74	*SRSF6*	serine and arginine rich splicing factor 6(SRSF6)	*MNDA*	myeloid cell nuclear differentiation antigen(MNDA)
75	*CTTN*	cortactin(CTTN)	*SAT1*	spermidine/spermine N1-acetyltransferase 1(SAT1)
76	*WTAP*	Wilms tumor 1 associated protein(WTAP)	*PLEK*	pleckstrin(PLEK)
77	*CAMP*	cathelicidin antimicrobial peptide(CAMP)	*CTTN*	cortactin(CTTN)
78	*STOM*	stomatin(STOM)	*CAPZA1*	capping actin protein of muscle Z-line alpha subunit 1(CAPZA1)
79	*THBS1*	thrombospondin 1(THBS1)	*STOM*	stomatin(STOM)
80	*HNRNPA2B1*	heterogeneous nuclear ribonucleoprotein A2/B1(HNRNPA2B1)	*ITGA2B*	integrin subunit alpha 2b(ITGA2B)
81	*UBE2D3*	ubiquitin conjugating enzyme E2 D3(UBE2D3)	*RPS27A*	ribosomal protein S27a(RPS27A)
82	*MIR4738*	microRNA 4738(MIR4738)	*SNPRE*	small nuclear ribonucleoprotein polypeptide E(SNRPE)
83	*RPL21*	ribosomal protein L21(RPL21)	*RAP1B*	RAP1B, member of RAS oncogene family(RAP1B)
84	*GNAQ*	G protein subunit alpha q(GNAQ)	*MYL6*	myosin light chain 6(MYL6)
85	*MTATP6P1*	MTATP6P1	*NAP1L1*	nucleosome assembly protein 1 like 1(NAP1L1)
86	*DEFA4*	defensin alpha 4(DEFA4)	*MIR4738*	microRNA 4738(MIR4738)
87	*SDCBP*	syndecan binding protein(SDCBP)	*ERCC-00042*	
88	*MYL12A*	myosin light chain 12A(MYL12A)	*IGHV3-53*	IGHV3-53
89	*MYL6*	myosin light chain 6(MYL6)	*DAAM1*	dishevelled associated activator of morphogenesis 1(DAAM1)
90	*RPL9*	ribosomal protein L9(RPL9)	*PNRC2*	proline rich nuclear receptor coactivator 2(PNRC2)
91	*SDPR*	serum deprivation response(SDPR)	*RPS20*	ribosomal protein S20(RPS20)
92	*RPS13*	ribosomal protein S13(RPS13)	*SUB1*	SUB1 homolog, transcriptional regulator(SUB1)
93	*HLA-E*	major histocompatibility complex, class I, E(HLA-E)	*RPL5*	ribosomal protein L5(RPL5)
94	*ITM2B*	integral membrane protein 2B(ITM2B)	*WTAP*	Wilms tumor 1 associated protein(WTAP)
95	*PIP4K2A*	phosphatidylinositol-5-phosphate 4-kinase type 2 alpha(PIP4K2A)	*RNASE2*	ribonuclease A family member 2(RNASE2)
96	*CD47*	CD47 molecule(CD47)	*SSR4*	signal sequence receptor subunit 4(SSR4)
97	*IGKV1-5*	IGKV1-5	*TPT1*	tumor protein, translationally-controlled 1(TPT1)
98	*NT5C3A*	5'-nucleotidase, cytosolic IIIA(NT5C3A)	*HNRNPA2B1*	heterogeneous nuclear ribonucleoprotein A2/B1(HNRNPA2B1)
99	*EEF1A1*	eukaryotic translation elongation factor 1 alpha 1(EEF1A1)	*CDC42*	cell division cycle 42(CDC42)
100	*RPL5*	ribosomal protein L5(RPL5)	*ERCC-00003*	

Table 5.5 The 100 most highly expressed transcripts in megakaryocytes from patients with AMI compared to control

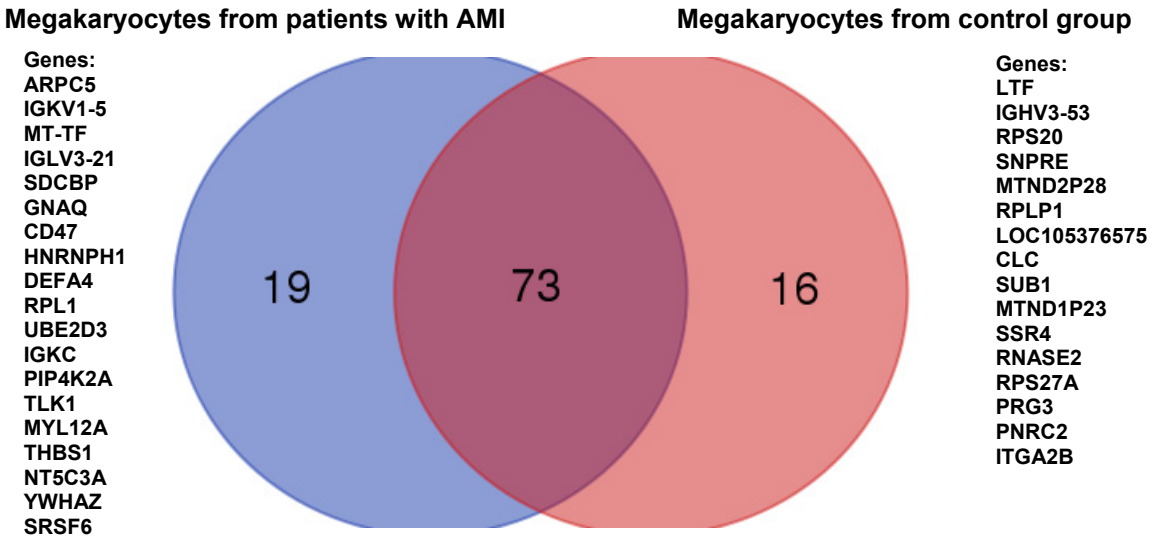


Figure 5.18 Venn diagram to show common expression of abundantly expressed genes in megakaryocytes from patients with AMI and from a control group

AMI group

Control group

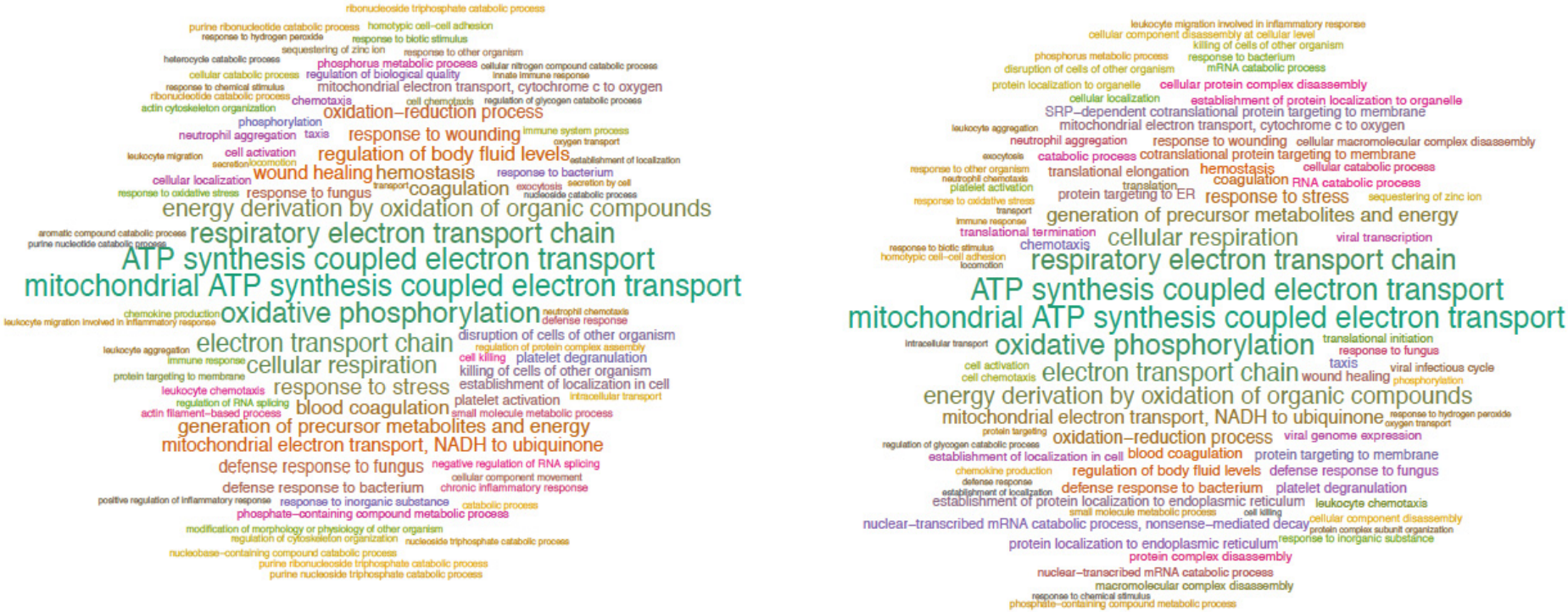


Figure 5.19 Over-represented gene ontologies in most highly expressed megakaryocyte genes from AMI and control groups
The full GO analysis may be found in Appendix 5, Table S5.1, Table S5.2. The word cloud is a visual representation of the over-represented GO biological processes with their size an indicator of significance for most abundant genes in megakaryocytes from AMI and control groups.

ENSEMBL_GENE_ID	HGNC_symbol	Gene_name	log2_ FoldChange	p_ adjusted
ENSG00000210184	*MT-TS2*	mitochondrially encoded tRNA serine 2 (AGU/C) [7498]	2.847334232	1.56E-24
ENSG00000269028	*MTRNR2L12*	MT-RNR2-like 12 [37169]	1.461315639	1.66E-14
ENSG00000210144	*MT-TY*	mitochondrially encoded tRNA tyrosine [7502]	1.761370411	2.03E-09
ENSG00000210077	*MT-TV*	mitochondrially encoded tRNA valine [7500]	1.606536821	2.28E-08
ENSG00000243264	*IGKV2D-29*	immunoglobulin kappa variable 2D-29 [5800]	6.397659303	4.13E-08
ENSG00000198744	*ENSG00000198744*		0.764145271	3.67E-06
ENSG00000210195	*MT-TT*	mitochondrially encoded tRNA threonine [7499]	1.857850981	4.78E-06
ENSG00000210112	*MT-TM*	mitochondrially encoded tRNA methionine [7492]	1.19984291	8.22E-06
ENSG00000211949	*IGHV3-23*	immunoglobulin heavy variable 3-23 [5588]	4.042359855	2.00E-05
ENSG00000211679	*IGLC3*	immunoglobulin lambda constant 3 (Kern-Oz+ marker) [5857]	2.591235725	2.60E-05
ENSG00000251546	*IGKV1D-39*	immunoglobulin kappa variable 1D-39 [5756]	4.155514756	2.96E-05
ENSG00000211659	*IGLV3-25*	immunoglobulin lambda variable 3-25 [5908]	4.492515864	4.37E-05
ENSG00000210154	*MT-TD*	mitochondrially encoded tRNA aspartic acid [7478]	1.731186971	6.68E-05
ENSG00000210174	*MT-TR*	mitochondrially encoded tRNA arginine [7496]	1.369858066	1.91E-04
ENSG00000280334	*ENSG00000280334*		4.551573876	2.25E-04
ENSG00000261654	*ENSG00000261654*		3.112157925	4.71E-04
ENSG00000211898	*IGHD*	immunoglobulin heavy constant delta [5480]	4.544680756	5.43E-04
ENSG00000148357	*HMCN2*	hemicentin 2 [21293]	2.898573868	7.00E-04
ENSG00000134463	*ECHDC3*	enoyl CoA hydratase domain containing 3 [23489]	3.953027839	8.73E-04
ENSG00000221995	*TIAF1*	TGFB1-induced anti-apoptotic factor 1 [11803]	4.11849775	9.35E-04
ENSG00000163736	*PPBP*	pro-platelet basic protein (chemokine (C-X-C motif) ligand 7) [9240]	1.513084446	9.72E-04
ENSG00000203875	*SNHG5*	small nucleolar RNA host gene 5 [21026]	1.625652587	1.09E-03
ENSG00000210100	*MT-TI*	mitochondrially encoded tRNA isoleucine [7488]	1.226486024	1.78E-03
ENSG00000142609	*CFAP74*	cilia and flagella associated protein 74 [29368]	2.922822246	2.09E-03
ENSG00000240382	*IGKV1-17*	immunoglobulin kappa variable 1-17 [5733]	4.412594197	2.36E-03
ENSG00000224650	*IGHV3-74*	immunoglobulin heavy variable 3-74 [5624]	4.087073797	2.36E-03
ENSG00000210194	*MT-TE*	mitochondrially encoded tRNA glutamic acid [7479]	1.124514457	2.57E-03
ENSG00000210196	*MT-TP*	mitochondrially encoded tRNA proline [7494]	0.821292279	3.01E-03
ENSG00000205622	*ENSG00000205622*		4.357099261	3.60E-03
ENSG00000020577	*SAMD4A*	sterile alpha motif domain containing 4A [23023]	2.966890395	3.87E-03
ENSG00000188783	*PRELP*	proline/arginine-rich end leucine-rich repeat protein [9357]	3.965613306	4.14E-03
ENSG00000185585	*OLFML2A*	olfactomedin-like 2A [27270]	2.927602289	4.54E-03
ENSG00000260293	*ENSG00000260293*		3.930519326	5.54E-03
ENSG00000203387	*ENSG00000203387*		4.245687526	5.73E-03
ENSG00000187244	*BCAM*	basal cell adhesion molecule (Lutheran blood group) [6722]	3.268477088	5.78E-03
ENSG00000144339	*TMEFF2*	transmembrane protein with EGF-like and two follistatin-like domains 2 [11867]	3.078674541	5.84E-03
ENSG00000268549	*ENSG00000268549*		3.33779745	6.25E-03
ENSG00000270986	*ENSG00000270986*		3.538589966	6.47E-03
ENSG00000211598	*IGKV4-1*	immunoglobulin kappa variable 4-1 [5834]	2.93182172	6.84E-03
ENSG00000211638	*IGLV8-61*	immunoglobulin lambda variable 8-61 [5931]	3.597917664	6.91E-03
ENSG00000127920	*GNG11*	guanine nucleotide binding protein (G protein), gamma 11 [4403]	1.73948149	7.13E-03
ENSG00000213753	*CENPBD1P1*	CENPB DNA-binding domains containing 1 pseudogene 1 [28421]	3.235478948	7.65E-03
ENSG00000211662	*IGLV3-21*	immunoglobulin lambda variable 3-21 [5905]	2.586296177	8.33E-03
ENSG00000229753	*ENSG00000229753*		5.304411343	8.65E-03
ENSG00000258741	*ENSG00000258741*		2.945056862	9.06E-03
ENSG00000140279	*DUOX2*	dual oxidase 2 [13273]	2.741809501	9.14E-03
ENSG00000197296	*FITM2*	fat storage-inducing transmembrane protein 2 [16135]	2.813817846	9.65E-03
ENSG00000152939	*MARVELD2*	MARVEL domain containing 2 [26401]	2.597888809	9.65E-03
ENSG00000116745	*RPE65*	retinal pigment epithelium-specific protein 65kDa [10294]	4.336426816	9.99E-03
ENSG00000232398	*TMPRSS11CP*	transmembrane protease, serine 11C, pseudogene [31934]	3.629439205	1.02E-02

Table 5.6 Transcripts upregulated in AMI compared with control group

ENSEMBL_GENE_ID	HGNC_symbol	Gene_name	log2_ FoldChange	p_ adjusted
ENSG00000225972	*MTND1P23*	MT-ND1 pseudogene 23 [42092]	-6.887850524	4.94E-109
ENSG00000198804	*MT-CO1*	mitochondrially encoded cytochrome c oxidase I [7419]	-1.212678956	2.77E-38
ENSG00000198786	*MT-ND5*	mitochondrially encoded NADH dehydrogenase 5 [7461]	-1.189259734	3.02E-27
ENSG00000211459	*MT-RNR1*	mitochondrially encoded 12S RNA [7470]	-0.864392322	1.06E-16
ENSG00000256618	*MTRNR2L1*	MT-RNR2-like 1 [37155]	-2.299632211	1.17E-11
ENSG00000198695	*MT-ND6*	mitochondrially encoded NADH dehydrogenase 6 [7462]	-0.581178597	4.21E-09
ENSG00000211666	*IGLV2-14*	immunoglobulin lambda variable 2-14 [5888]	-5.623089436	2.76E-08
ENSG00000125691	*RPL23*	ribosomal protein L23 [10316]	-2.095471862	3.06E-08
ENSG00000225630	*MTND2P28*	MT-ND2 pseudogene 28 [42129]	-1.210734612	4.13E-08
ENSG00000198763	*MT-ND2*	mitochondrially encoded NADH dehydrogenase 2 [7456]	-0.513826979	7.89E-08
ENSG00000136167	*LCP1*	lymphocyte cytosolic protein 1 (L-plastin) [6528]	-2.317792127	2.50E-06
ENSG00000232177	*MTND4P24*	MT-ND4 pseudogene 24 [42220]	-2.857699359	2.61E-06
ENSG00000166012	*TAF1D*	TATA box binding protein (TBP)-associated factor, RNA polymerase I, D, 41kDa [28759]	-3.035159937	3.99E-06
ENSG00000118181	*RPS25*	ribosomal protein S25 [10413]	-1.747228133	6.70E-06
ENSG00000070081	*NUCB2*	nucleobindin 2 [8044]	-2.332301955	6.87E-06
ENSG00000026025	*VIM*	vimentin [12692]	-2.127970766	1.61E-05
ENSG00000170802	*FOXN2*	forkhead box N2 [5281]	-2.559844929	2.96E-05
ENSG00000110203	*FOLR3*	folate receptor 3 (gamma) [3795]	-3.81752763	3.38E-05
ENSG00000156508	*EEF1A1*	eukaryotic translation elongation factor 1 alpha 1 [3189]	-1.614829991	3.46E-05
ENSG00000163682	*RPL9*	ribosomal protein L9 [10369]	-1.553754602	4.51E-05
ENSG00000187325	*TAF9B*	TAF9B RNA polymerase II, TATA box binding protein (TBP)-associated factor, 31kDa [17306]	-4.270522086	4.51E-05
ENSG00000140199	*SLC12A6*	solute carrier family 12 (potassium/chloride transporter), member 6 [10914]	-2.594290773	5.23E-05
ENSG00000122862	*SRGN*	serglycin [9361]	-1.934077687	5.57E-05
ENSG00000119314	*PTBP3*	polypyrimidine tract binding protein 3 [10253]	-2.245101726	6.78E-05
ENSG00000168461	*RAB31*	RAB31, member RAS oncogene family [9771]	-2.303785535	7.09E-05
ENSG00000100241	*SBF1*	SET binding factor 1 [10542]	-3.587127679	7.09E-05
ENSG00000122026	*RPL21*	ribosomal protein L21 [10313]	-1.428659958	1.32E-04
ENSG00000272211	*ENSG00000272211*		-3.89338801	1.39E-04
ENSG00000137076	*TLN1*	talin 1 [11845]	-1.96740421	1.48E-04
ENSG00000115457	*IGFBP2*	insulin-like growth factor binding protein 2, 36kDa [5471]	-4.554655775	1.60E-04
ENSG00000145425	*RPS3A*	ribosomal protein S3A [10421]	-1.586498056	1.65E-04
ENSG00000249119	*MTND6P4*	MT-ND6 pseudogene 4 [39467]	-1.387081985	1.78E-04
ENSG00000163428	*LRRC58*	leucine rich repeat containing 58 [26968]	-2.811236399	1.81E-04
ENSG00000127603	*MACF1*	microtubule-actin crosslinking factor 1 [13664]	-1.719927507	2.03E-04
ENSG00000138185	*ENTPD1*	ectonucleoside triphosphate diphosphohydrolase 1 [3363]	-2.423431939	2.25E-04
ENSG00000008086	*CDKL5*	cyclin-dependent kinase-like 5 [11411]	-3.016263468	3.01E-04
ENSG00000081237	*PTPRC*	protein tyrosine phosphatase, receptor type, C [9666]	-1.855896932	3.16E-04
ENSG00000119335	*SET*	SET nuclear proto-oncogene [10760]	-1.753945952	3.64E-04
ENSG00000021355	*SERPINB1*	serpin peptidase inhibitor, clade B (ovalbumin), member 1 [3311]	-1.846181192	4.04E-04
ENSG00000139218	*SCAF11*	SR-related CTD-associated factor 11 [10784]	-1.860512301	4.36E-04
ENSG00000008988	*RPS20*	ribosomal protein S20 [10405]	-1.497491206	5.24E-04
ENSG00000007944	*MYLIP*	myosin regulatory light chain interacting protein [21155]	-3.391709804	5.24E-04
ENSG00000168876	*ANKRD49*	ankyrin repeat domain 49 [25970]	-3.688781913	5.24E-04
ENSG00000108599	*AKAP10*	A kinase (PRKA) anchor protein 10 [368]	-2.534354621	5.51E-04
ENSG00000198830	*HMGN2*	high mobility group nucleosomal binding domain 2 [4986]	-1.779902529	5.86E-04
ENSG00000087086	*FTL*	ferritin, light polypeptide [3999]	-1.200532107	6.48E-04
ENSG00000173193	*PARP14*	poly (ADP-ribose) polymerase family, member 14 [29232]	-2.734470718	6.55E-04
ENSG00000159322	*ADPGK*	ADP-dependent glucokinase [25250]	-2.792167824	6.58E-04
ENSG00000110435	*PDHX*	pyruvate dehydrogenase complex, component X [21350]	-3.633520422	6.58E-04
ENSG00000047410	*TPR*	translocated promoter region, nuclear basket protein [12017]	-2.07250317	7.00E-04

Table 5.7 Transcripts downregulated in AMI compared with control group

Figure 5.20 Over-represented gene ontologies in downregulated genes megakaryocytes from the AMI compared with control group.
The full GO analysis may be found in Appendix 5, Table S5.5. The word cloud is a visual representation of the over-represented GO biological processes with their size an indicator of significance for downregulated genes in megakaryocytes from the AMI group. There were no significant GO categories upregulated in megakaryocytes from the AMI compared to control group.

5.3.1.5 Expression of megakaryocyte disease signatures in HSC clusters

To assess the expression of the megakaryocyte gene signatures associated with AMI (disease) and control within the clustered single cell HSCs from Chapter 3, the expression of the genes upregulated in increasing and those upregulated in decreasing ploidy was overlaid onto the Monocle 2 plot from Figure 3.16. The expression of both gene signatures from the AMI group and the control group were mainly restricted to cluster 1. There was little or no expression of ploidy associated gene signatures in cluster 4 which was thought to be another megakaryocyte-biased subset from Chapter 3, Figure 5.21.

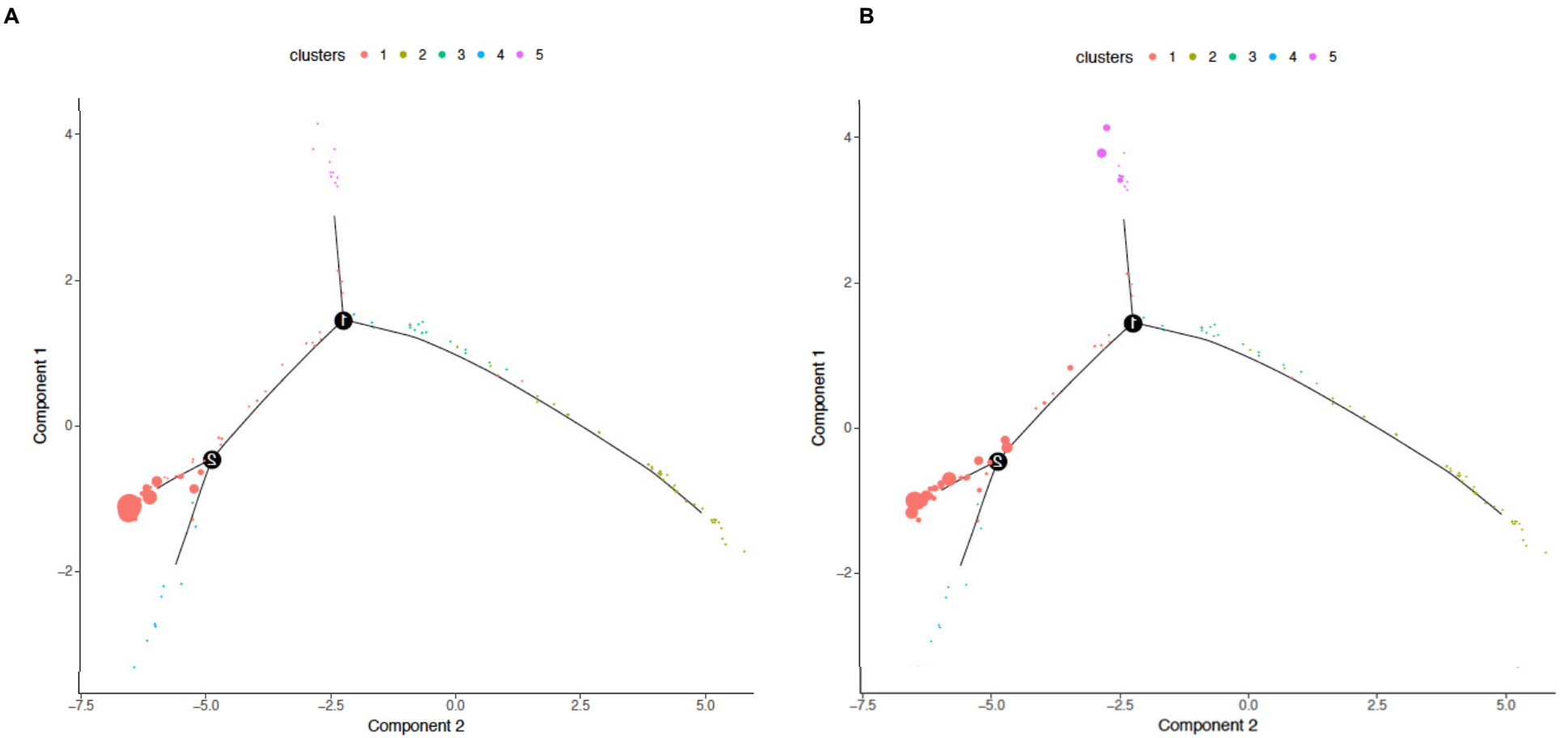


Figure 5.21 Single cell HSC differentiation trajectories using Monocle 2 showing expression of megakaryocyte disease signatures.
Individual HSCs are connected by a minimum spanning tree with branch points (thin lines), representing the differentiation trajectory and clusters are differentiated by colour. Expression of gene signature is indicated by the size of each single cell. a. AMI group megakaryocyte signature; b. Control group megakaryocyte signature.

5.3.2 Disease associated platelet transcriptional signatures

5.3.2.1 Patient characteristics

Platelets were isolated from the peripheral blood of 29 individuals undergoing cardiac surgery for the purposes of RNA-seq. This included 15 non-coronary artery disease control individuals and 14 patients with AMI in the past 6 months. Table 5.8 demonstrates the baseline characteristics, platelet parameters and function for each study group, including differences between the groups.

5.3.2.2 RNA-seq library preparation

RNA was extracted from platelets isolated from peripheral blood using Trizol reagent as described in Section 2.3.6. A head to head comparison between RNA-seq library preparation using the RiboErase protocol (Kapa) or the Smart-seq2 protocol was made by preparing libraries for 3 samples (60Z, 6SP, 6VA) using both methods (Figure 5.22). This comparison showed that the RiboErase libraries had higher read counts and higher numbers of reads mapping to the transcriptome (Figure 5.22a, b). There was a high correlation in terms of gene expression between the two methods (r=0.82-0.84) but the RiboErase kit was able to identify expression of greater numbers of protein-coding genes (Figure 5.22c). Furthermore, the overall ratio of mitochondrial RNA was significantly lower in the libraries prepared using the RiboErase kit compared with SmartSeq2 ($p=0.0019$; t-test) shown in Figure 5.22d. Therefore, all platelet RNA-seq libraries were prepared using the RiboErase kit if enough material was available (minimum total RNA input quantity: 50ng).

In terms of batch effects Table 5.6 shows each sample ID and corresponding sequencing lane ID whereas Table 5.7 outlines the batches in which samples were collected, and libraries prepared and sequenced.

Table 5.8 Baseline patient characteristics

	Controls n=15	AMI n=14	p value
Gender M	6	12	0.0183*
Age (mean±SD)	63.5±17.7	65.4±8.40	ns
Hypertension	8	13	0.0352*
Statins	5	11	0.0253*
Diabetes	2	7	ns
Smoking	0	5	0.0169*
Aspirin within last 48 hours	1	14	<0.0001***
Left ventricular dysfunction	3	5	ns
Coronary disease on angiography	-	14	-
History of AMI within 6 months	-	14	-
Time from infarction (mean±SD)	-	33.4±31.9	-
Aortic valve surgery	7	1	0.0372*
Mitral valve surgery	9	0	0.0007***
Pre-operative Hb (mean±SD)	12.75±2.14	12.52±1.68	ns
Pre-operative WC (mean±SD)	7.52±2.62	8.14±3.10	ns
Pre-operative Plt (mean±SD)	230±47.2	274±74.0	ns
Time to sysmex, min (mean±SD)	173.9±70.9	147.6±31.2	ns
PLT, x109/L (mean±SD)	197.5±50.0	242.4±70.6	ns
MPV, fL (mean±SD)	11.2±0.78	11.0±0.92	ns
Plateletcrit, % (mean±SD)	0.22±0.049	0.26±0.058	ns
IPF, % (mean±SD)	3.38±2.09	3.12±2.06	ns
MK fraction , % (mean±SD)	0.089±0.046	0.106±0.046	ns
MK 4N frequency, % (mean±SD)	10.1±5.52	8.26±3.44	ns
MK 8N frequency, % (mean±SD)	20.0±6.54	17.0±5.82	ns
MK 16N frequency, % (mean±SD)	48.8±7.29	50.2±5.44	ns
MK 32N frequency, % (mean±SD)	21.2±7.52	24.6±7.03	ns
Resting P-selectin expression, % (mean±SD)	47.9±7.98	51.4±7.43	ns
Platelet activation:			
Fibrinogen binding, % with ADP (mean±SD)	45.8±15.6	29.6±21.0	ns
Fibrinogen binding, % with CRP-XL (mean±SD)	37.9±15.4	31.6±11.1	ns
Fibrinogen binding, % with TRAP-6 (mean±SD)	9.74±12.0	3.65±3.34	ns
P-selectin expression, % with ADP (mean±SD)	35.7±8.87	31.2±10.7	ns
P-selectin expression, % with CRP-XL (mean±SD)	43.2±8.47	41.2±6.41	ns
P-selectin expression, % with TRAP-6 (mean±SD)	32.1±10.2	31.4±8.66	ns

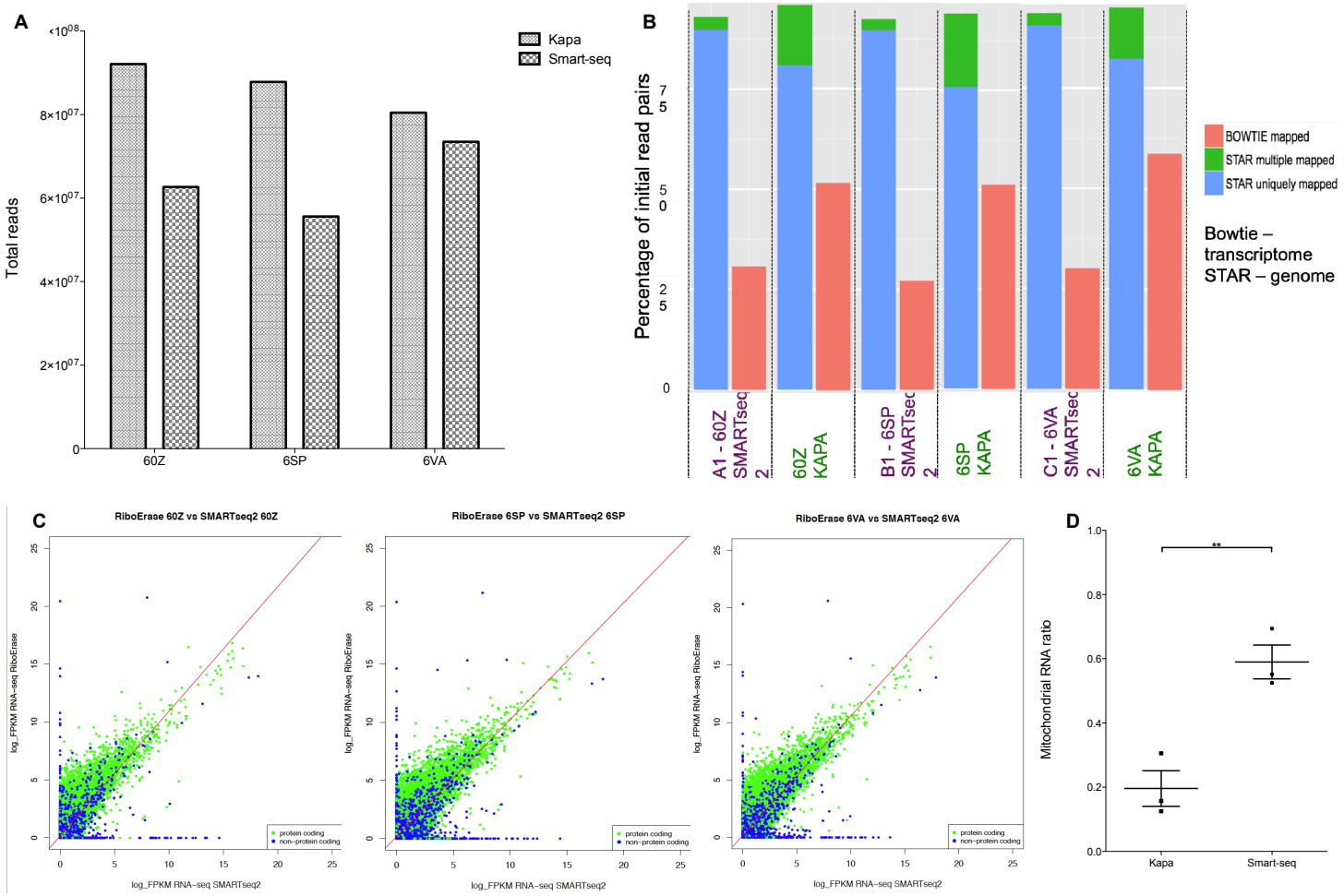


Figure 5.22 Direct comparison of RiboErase (Kapa) and Smart-seq2 protocols for platelet RNA-seq library preparation. RNA extracted from peripheral blood of 3 non-coronary disease controls (60Z, 6SP, 6VA); RNA-seq libraries were prepared using both the RiboErase kit and Smart-seq2. A. Total read counts; B. Alignment rate to transcriptome and genome; C. Spearman rank correlations between gene expression by RiboErase vs Smart-seq2 methods for samples 60Z, 6SP, 6VA (r=0.82, r=0.82, r=0.84 respectively), showing protein coding and non-protein coding genes; D. Mitochondrial RNA ratio, ** indicates p<0.01.

Patient ID	Sequencing Lane ID
60Z, 6VA, 6SP DM0, G6K, FVB, P4P R7B, GC8, ATZ, FBE CST, M8T, CUP, CEK CDM, 318, H4K, EH6 4J5, CYH, 0DX, F7M F5Q, 5Q0, EFA 46V, W4Y, WBK	SLX-10486 (SLX-10488 Smart-seq2) SLX-10591 SLX-10592 SLX-10593 SLX-10594 SLX-10595 SLX-10599 SLX-10600

Table 5.9 Patient ID on each sequencing lane

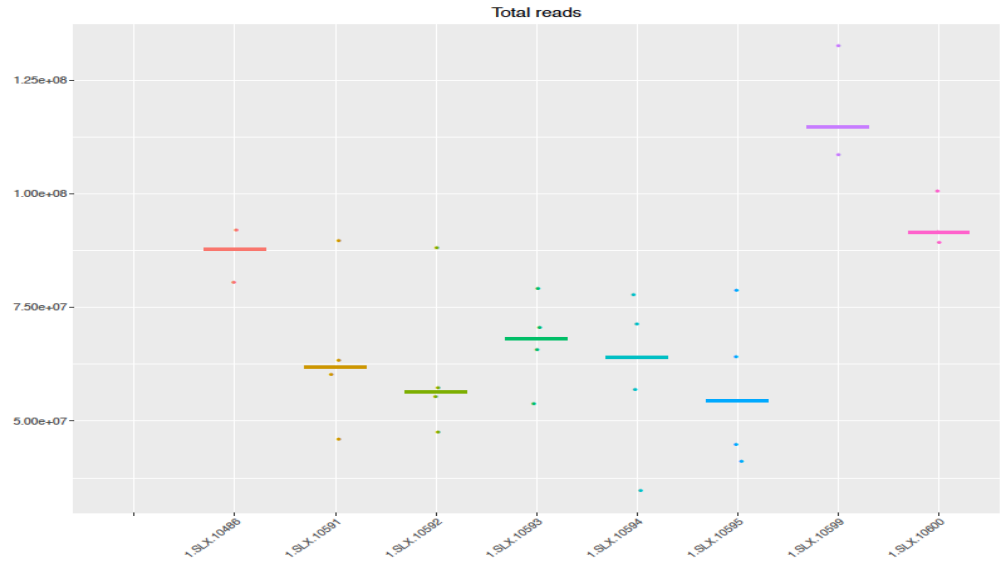
Sample collection	Amplification	Library prep / RNA-seq
0DX 318 46V 4J5 5Q0 60Z 6SP 6VA ATZ CDM, CEK CST, CUP, CYH DM0 EFA EH6 F5Q F7M FBE FVB G6K GC8 H4K M8T P4P R7B W4Y WBK	1 60Z, 6SP, 6VA 2 DM0, G6K, FVB, P4P R7B, GC8, ATZ, FBE CST, M8T, CUP, CEK CDM, 318, H4K, EH6 4J5, CYH, 0DX, F7M 3 F5Q, 5Q0, EFA 46V, W4Y, WBK	1 SLX-10486 2 SLX-10591 SLX-10592 SLX-10593 SLX-10594 SLX-10595 3 SLX-10599 SLX-10600

Table 5.10 Platelet collection and library preparation/RNA-seq batches

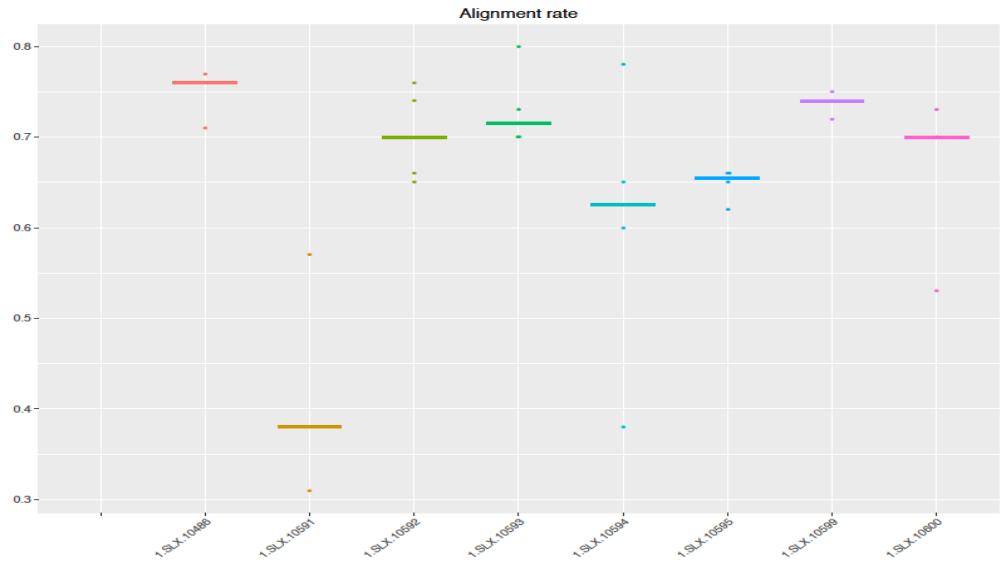
5.3.2.3 RNA sequencing data: Quality measures

Figure 5.23 shows metrics per sequencing lane for total read count, alignment to the reference transcriptome (GRCh37), no of genes called, proportion of reads mapping to annotated exonic regions and the proportion of reads mapping to the mitochondrial genome. One of the lanes SLX10591, appears to have lower alignment rate and lower exon ratio than other lanes, however, otherwise performs comparably with the other lanes.

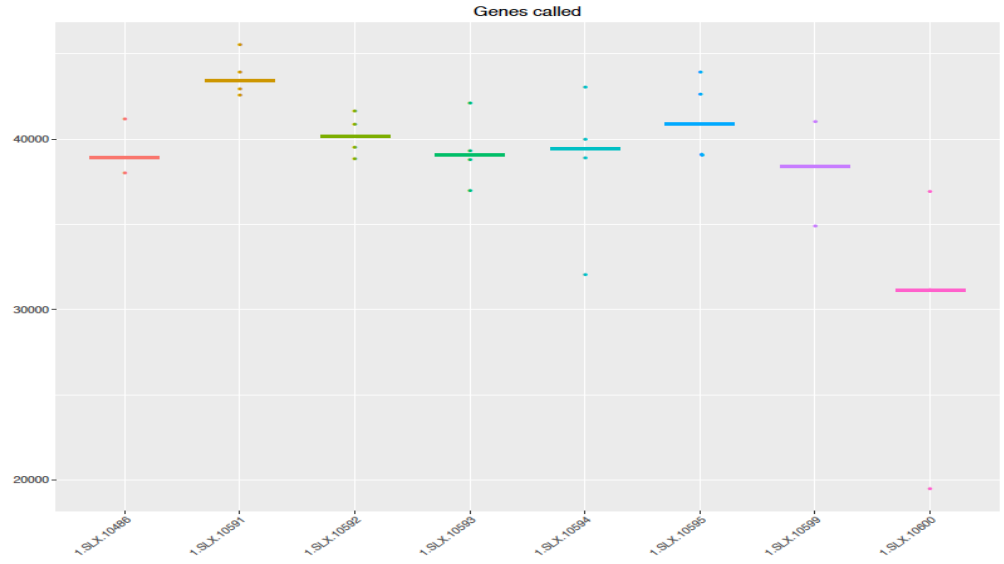
A


Total reads
1.25e+08
1.00e+08
7.50e+07
5.00e+07
1.SLX.10486
1.SLX.10591
1.SLX.10592
1.SLX.10593
1.SLX.10594
1.SLX.10595
1.SLX.10599
1.SLX.10600

B


Alignment rate
0.8
0.7
0.6
0.5
0.4
0.3
1.SLX.10486
1.SLX.10591
1.SLX.10592
1.SLX.10593
1.SLX.10594
1.SLX.10595
1.SLX.10599
1.SLX.10600

C


Genes called
40000
30000
20000
1.SLX.10486
1.SLX.10591
1.SLX.10592
1.SLX.10593
1.SLX.10594
1.SLX.10595
1.SLX.10599
1.SLX.10600

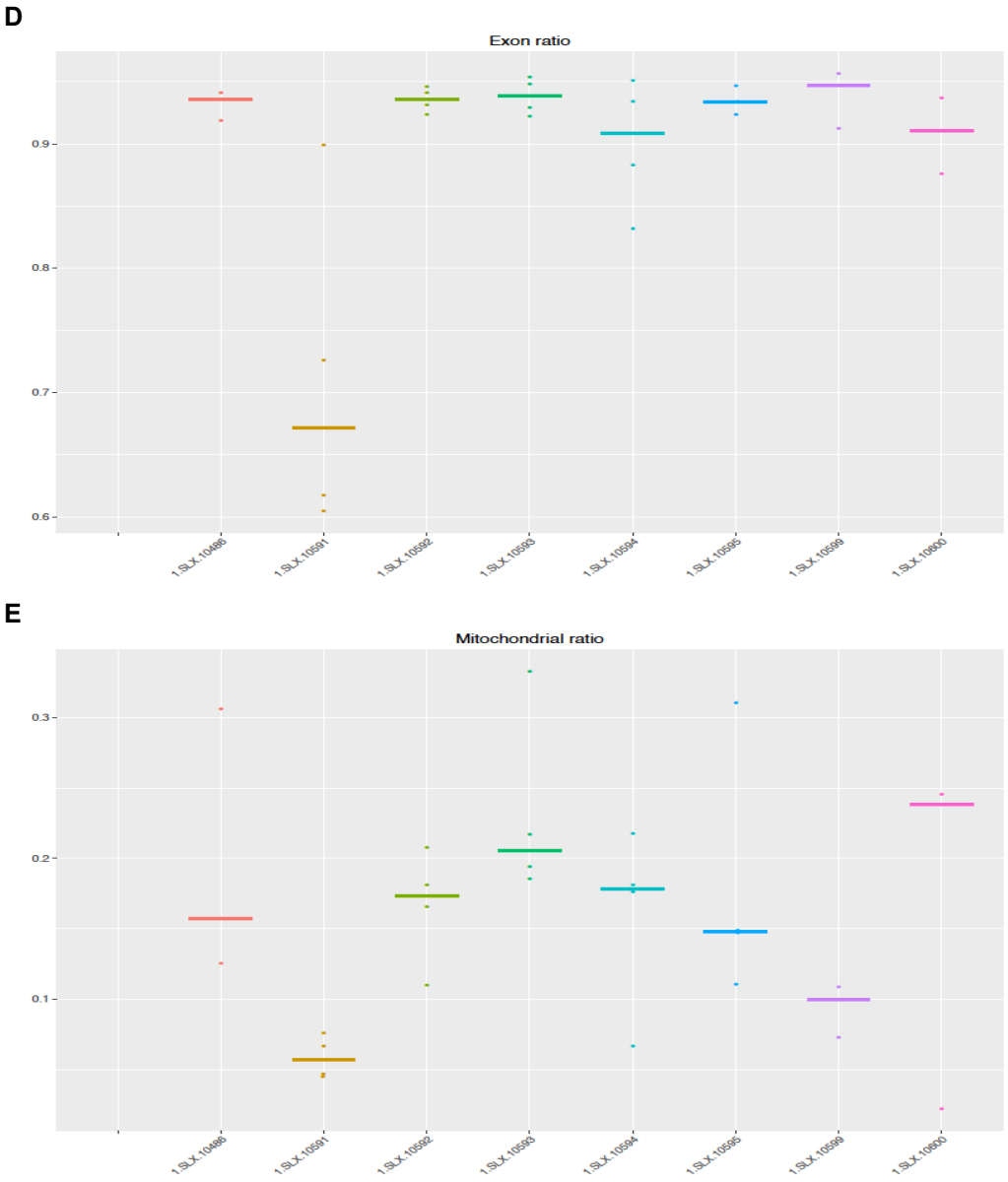


Figure 5.23 Quality control metrics/sequencing lane. A. Total reads; B. Rate of alignment to transcriptome; C. No. genes called; D. Ratio of exonic reads to exonic+intronic; E. Ratio of reads mapping to mitochondrial genes.

5.3.2.4 Gene expression analysis

The 100 most abundantly expressed genes in platelets from patients with recent AMI and the control group are shown in Table 5.11. There is a marked overlap in the top 100 platelet genes expressed in each group with 97 being common to both groups (Figure 5.24). Functional annotation of the gene list that is common to both groups using a GO analysis demonstrates a strong ontological bias towards platelet degranulation/activation, wound healing, coagulation and haemostasis as well as cellular metabolism; Figure 5.25 (full GO analysis Appendix 5, Table S5.7). Both sets of platelets contain many transcripts known to be highly expressed in platelets from previous datasets (Rowley et al. 2011) including *PPBP, PF4, SDPR, TMSB4X, TUBB, GRAP2, ITGA2B, ITG1B, MYL9, TPM4, CXCL5. PTGS1*, the gene that encodes the cyclooxygenase 1 enzyme that is the target for aspirin is also highly transcribed in both datasets.

To better understand differences in platelet transcriptome between the AMI and control groups hierarchical clustering was performed between samples (Figure 5.26a). Although some of the control samples appeared to cluster together there was no clear pattern separating the two groups. Similarly, the two groups did not separate according to any of the most important variance principal components in gene expression in the dataset (Figure 5.26b). To investigate if there was a correlation between gene expression and other clinical data, rather than disease group, further analysis of the clustering with categorical clinical data such as smoking, presence of diabetes and aspirin use was performed. These showed no clear effect on clustering (Figure 5.27).

A differential gene expression analysis identified 16 differentially expressed genes. Of these 10 were significantly upregulated in patients with recent AMI (Figure 5.28). 4 of the upregulated genes were Y-linked and therefore gender specific: *TMSN4Y, EIF1AY, KDM5D, DDX3Y*. The majority of the patients with recent AMI were male with the control group being more heterogeneous which is likely to have accounted for this differential expression. The other genes upregulated included *IGHA2* that encodes for the constant region of the IgA heavy chain, *RPS17L* encoding for the ribosomal S17 protein, part of the 40S subunit, *RHD* the Rh blood group D antigen and *LOXL3* which encodes lysyl oxidase like 3, that plays a role in crosslinking collagen and elastin. Of the remaining upregulated genes 1 was an RNA gene (*ENSG00000238201*) and one a ribosomal subunit pseudogene (*RPL21P120*).

No	Platelets from AMI group		Platelets from control group	
	HGNC_symbol	Gene_name	HGNC_symbol	Gene_name
1	*ND1*	NADH dehydrogenase, subunit 1 (complex I)	*ND1*	NADH dehydrogenase, subunit 1 (complex I)
2	*ND5*	NADH dehydrogenase, subunit 5 (complex I)	*ND5*	NADH dehydrogenase, subunit 5 (complex I)
3	*ND4*	NADH dehydrogenase, subunit 4 (complex I)	*MT-RNR2*	Mitochondrially Encoded 16S RNA
4	*TMSB4X*	thymosin beta 4, X-linked(TMSB4X)	*ND4*	NADH dehydrogenase, subunit 4 (complex I)
5	*MT-RNR2*	Mitochondrially Encoded 16S RNA	*TMSB4X*	thymosin beta 4, X-linked(TMSB4X)
6	*MTND2*	MTND2(ND2)	*COX1*	cytochrome c oxidase subunit I(COX1)
7	*COX1*	cytochrome c oxidase subunit I(COX1)	*ND2*	MTND2(ND2)
8	*CYTB*	cytochrome b(CYTB)	*CYTB*	cytochrome b(CYTB)
9	*COX3*	cytochrome c oxidase III(COX3)	*COX3*	cytochrome c oxidase III(COX3)
10	*B2M*	beta-2-microglobulin(B2M)	*B2M*	beta-2-microglobulin(B2M)
11	*PPBP*	pro-platelet basic protein(PPBP)	*PPBP*	pro-platelet basic protein(PPBP)
12	*ACTB*	actin beta(ACTB)	*COX2*	cytochrome c oxidase subunit II(COX2)
13	*NRGN*	neurogranin(NRGN)	*ND6*	NADH dehydrogenase, subunit 6 (complex I)
14	*PF4*	platelet factor 4(PF4)	*ATP6*	ATP synthase F0 subunit 6(ATP6)
15	*ATP6*	ATP synthase F0 subunit 6(ATP6)	*NRGN*	neurogranin(NRGN)
16	*COX2*	cytochrome c oxidase subunit II(COX2)	*MT-RNR1*	Mitochondrially Encoded 12S RNA
17	*ND6*	NADH dehydrogenase, subunit 6 (complex I)	*PF4*	platelet factor 4(PF4)
18	*MT-RNR1*	Mitochondrially Encoded 12S RNA	*ACTB*	actin beta(ACTB)
19	*FLNA*	filamin A(FLNA)	*SDPR*	serum deprivation response(SDPR)
20	*TLN1*	Talin 1	*FLNA*	filamin A(FLNA)
21	*SDPR*	serum deprivation response(SDPR)	*TUBB1*	tubulin beta 1 class VI(TUBB1)
22	*TUBB1*	tubulin beta 1 class VI(TUBB1)	*TLN1*	Talin 1
23	*F13A1*	coagulation factor XIII A chain(F13A1)	*F13A1*	coagulation factor XIII A chain(F13A1)
24	*MYH9*	myosin heavy chain 9(MYH9)	*TAGLN2*	transgelin 2(TAGLN2)
25	*TAGLN2*	transgelin 2(TAGLN2)	*ND3*	NADH dehydrogenase, subunit 3 (complex I)
26	*HLA-E*	major histocompatibility complex, class I, E	*MAP3K7CL*	MAP3K7 C-terminal like(MAP3K7CL)
27	*SPARC*	secreted protein acidic and cysteine rich	*MYH9*	myosin heavy chain 9(MYH9)
28	*ND3*	NADH dehydrogenase, subunit 3 (complex I)	*SPARC*	secreted protein acidic and cysteine rich
29	*VCL*	vinculin(VCL)	*FTH1*	ferritin heavy chain 1(FTH1)
30	*FTH1*	ferritin heavy chain 1(FTH1)	*HLA-E*	major histocompatibility complex, class I, E
31	*MAP3K7CL*	MAP3K7 C-terminal like(MAP3K7CL)	*MIR1244-1*	microRNA 1244-1(MIR1244-1)
32	*MIR1244-1*	microRNA 1244-1(MIR1244-1)	*HBA2*	hemoglobin subunit alpha 2(HBA2)
33	*YWHAZ*	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein zeta	*VCL*	vinculin(VCL)
34	*ITGA2B*	integrin subunit alpha 2b(ITGA2B)	*RGS18*	regulator of G-protein signaling 18
35	*NAP1L1*	nucleosome assembly protein 1 like 1	*YWHAZ*	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein zeta
36	*MTPN*	myotrophin(MTPN)	*NAP1L1*	nucleosome assembly protein 1 like 1
37	*CLU*	Clusterin	*MTPN*	myotrophin(MTPN)
38	*RGS18*	regulator of G-protein signaling 18(RGS18)	*KIF2A*	kinesin family member 2A(KIF2A)
39	*KIF2A*	kinesin family member 2A(KIF2A)	*GNAS*	GNAS complex locus(GNAS)
40	*GNAS*	GNAS complex locus(GNAS)	*GRAP2*	GRB2-related adaptor protein 2(GRAP2)
41	*HBA2*	hemoglobin subunit alpha 2(HBA2)	*CLU*	Clusterin
42	*GRAP2*	GRB2-related adaptor protein 2(GRAP2)	*PTGS1*	prostaglandin-endoperoxide synthase 1
43	*PTGS1*	prostaglandin-endoperoxide synthase 1	*ITGA2B*	integrin subunit alpha 2b(ITGA2B)
44	*PRKAR2B*	protein kinase cAMP-dependent type II regulatory subunit beta	*PRKAR2B*	protein kinase cAMP-dependent type II regulatory subunit beta(PRKAR2B)
45	*PKM*	pyruvate kinase, muscle(PKM)	*TSC22D1*	TSC22 domain family member 1
46	*PF4V1*	platelet factor 4 variant 1(PF4V1)	*RN7SK*	RNA, 7SK small nuclear(RN7SK)
47	*MYL9*	myosin light chain 9(MYL9)	*SOD2*	Superoxide Dismutase 2
48	*TGFB1*	transforming growth factor beta 1(TGFB1)	*TGFB1*	transforming growth factor beta 1(TGFB1)
49	*SOD2*	Superoxide Dismutase 2	*PKM*	pyruvate kinase, muscle(PKM)
50	*C6orf25*	chromosome 6 open reading frame 25	*OST4*	oligosaccharyltransferase complex subunit 4, non-catalytic(OST4)
51	*RN7SK*	RNA, 7SK small nuclear(RN7SK)	*ND4L*	NADH dehydrogenase, subunit 4L (complex I)
52	*LIMS1*	LIM zinc finger domain containing 1	*LIMS1*	LIM zinc finger domain containing 1
53	*ITM2B*	integral membrane protein 2B(ITM2B)	*HIST1H2AC*	histone cluster 1 H2A family member c(HIST1H2AC)
54	*SH3BGRL3*	SH3 domain binding glutamate rich protein like 3	*ITM2B*	integral membrane protein 2B(ITM2B)
55	*MBNL1*	muscleblind like splicing regulator 1	*MBNL1*	muscleblind like splicing regulator 1
56	*HIST1H2AC*	histone cluster 1 H2A family member c(HIST1H2AC)	*EIF1*	eukaryotic translation initiation factor 1
57	*THBS1*	thrombospondin 1(THBS1)	*PF4V1*	platelet factor 4 variant 1(PF4V1)
58	*OST4*	oligosaccharyltransferase complex subunit 4, non-catalytic(OST4)	*FTL*	ferritin light chain(FTL)
59	*TSC22D1*	TSC22 domain family member 1	*MYL9*	myosin light chain 9(MYL9)
60	*SH3BGRL2*	SH3 domain binding glutamate rich protein like 2(SH3BGRL2)	*SH3BGRL3*	SH3 domain binding glutamate rich protein like 3
61	*ITGB1*	integrin subunit beta 1(ITGB1)	*RANGRF*	RAN guanine nucleotide release factor
62	*TPM4*	tropomyosin 4(TPM4)	*RAB11A*	RAB11A, member RAS oncogene family
63	*SNN*	stannin(SNN)	*SNN*	stannin(SNN)
64	*WIPF1*	WAS/WASL interacting protein family member 1(WIPF1)	*C6orf25*	chromosome 6 open reading frame 25
65	*EIF1*	eukaryotic translation initiation factor 1	*SH3BGRL*	SH3 domain binding glutamate rich protein like(SH3BGRL)
66	*ND4L*	NADH dehydrogenase, subunit 4L (complex I)(ND4L)	*GPX1*	glutathione peroxidase 1(GPX1)
67	*GPX1*	glutathione peroxidase 1(GPX1)	*MAX*	MYC associated factor X(MAX)

68	*CALM3*	calmodulin 3(CALM3)	*CALM3*	calmodulin 3(CALM3)
69	*FTL*	ferritin light chain(FTL)	*CA2*	carbonic anhydrase 2(CA2)
70	*ACTG1*	actin gamma 1(ACTG1)	*SH3BGRL2*	SH3 domain binding glutamate rich protein like 2
71	*MAX*	MYC associated factor X(MAX)	*WIPF1*	WAS/WASL interacting protein family member 1
72	*LOC101928834*	uncharacterized LOC101928834	*ITGB1*	integrin subunit beta 1(ITGB1)
73	*RUFY1*	RUN and FYVE domain containing 1	*RAP1B*	RAP1B, member of RAS oncogene family
74	*RSU1*	Ras suppressor protein 1(RSU1)	*RUFY1*	RUN and FYVE domain containing 1
75	*SH3BGRL*	SH3 domain binding glutamate rich protein like(SH3BGRL)	*C2orf88*	chromosome 2 open reading frame 88
76	*RANGRF*	RAN guanine nucleotide release factor	*RSU1*	Ras suppressor protein 1(RSU1)
77	*ACTN1*	actinin alpha 1(ACTN1)	*THBS1*	thrombospondin 1(THBS1)
78	*RAB11A*	RAB11A, member RAS oncogene family	*H3F3A*	H3 histone family member 3A(H3F3A)
79	*C2orf88*	chromosome 2 open reading frame 88	*MTURN*	maturin, neural progenitor differentiation regulator homolog
80	*PTPRJ*	protein tyrosine phosphatase, receptor type J	*TPM4*	tropomyosin 4(TPM4)
81	*MTURN*	maturin, neural progenitor differentiation regulator homolog(MTURN)	*ACTN1*	actinin alpha 1(ACTN1)
82	*RAP1B*	RAP1B, member of RAS oncogene family	*ENSG0000023 3968*	uncharacterized LOC101928834
83	*HLA-C*	major histocompatibility complex, class I, C	*ACTG1*	actin gamma 1(ACTG1)
84	*H3F3A*	H3 histone family member 3A(H3F3A)	*PIP4K2A*	phosphatidylinositol-5-phosphate 4-kinase type 2 alpha
85	*ANO6*	anoctamin 6(ANO6)	*ARHGDIB*	Rho GDP dissociation inhibitor beta
86	*ARHGDIB*	Rho GDP dissociation inhibitor beta	*PTPRJ*	protein tyrosine phosphatase, receptor type J
87	*OAZ1*	ornithine decarboxylase antizyme 1(OAZ1)	*HLA-C*	major histocompatibility complex, class I, C
88	*CA2*	carbonic anhydrase 2(CA2)	*RNF11*	ring finger protein 11(RNF11)
89	*RNF11*	ring finger protein 11(RNF11)	*OAZ1*	ornithine decarboxylase antizyme 1(OAZ1)
90	*CD226*	CD226 molecule(CD226)	*PGRMC1*	progesterone receptor membrane component 1
91	*CXCL5*	C-X-C motif chemokine ligand 5(CXCL5)	*PCGF5*	polycomb group ring finger 5(PCGF5)
92	*HIST2H2BE*	histone cluster 2 H2B family member e(HIST2H2BE)	*CD226*	CD226 molecule(CD226)
93	*GNG11*	G protein subunit gamma 11(GNG11)	*GNG11*	G protein subunit gamma 11(GNG11)
94	*WDR1*	WD repeat domain 1(WDR1)	*HIST2H2BE*	histone cluster 2 H2B family member e(HIST2H2BE)
95	*FERMT3*	fermitin family member 3(FERMT3)	*PTPN18*	Protein Tyrosine Phosphatase, Non-Receptor Type 18
96	*PGRMC1*	progesterone receptor membrane component 1	*ANO6*	anoctamin 6(ANO6)
97	*CD99*	CD99 molecule(CD99)	*CXCL5*	C-X-C motif chemokine ligand 5(CXCL5)
98	*PCGF5*	polycomb group ring finger 5(PCGF5)	*YWHAE*	Tyrosine 3-Monooxygenase/Tryptophan 5-Monooxygenase Activation Protein Epsilon
99	*PIP4K2A*	phosphatidylinositol-5-phosphate 4-kinase type 2 alpha	*WDR1*	WD repeat domain 1(WDR1)
100	*CTSA*	cathepsin A(CTSA)	*YPEL5*	Yippee Like 5

Table 5.11 The 100 most highly expressed transcripts in platelets from patients with AMI compared to control

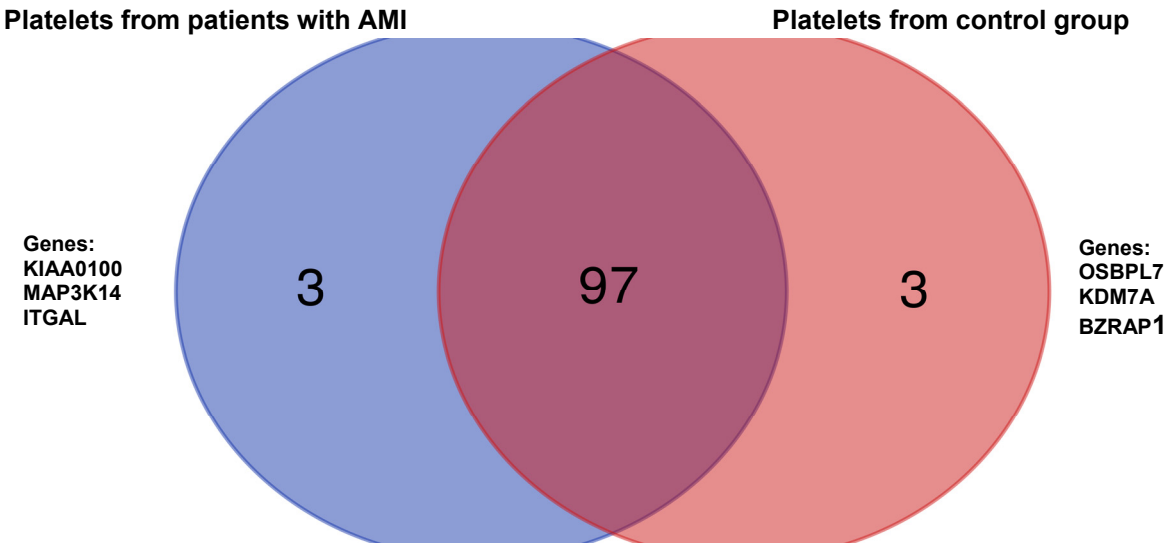


Figure 5.24 Venn diagram to show common expression of abundantly expressed genes in platelets from patients with AMI and from a control group

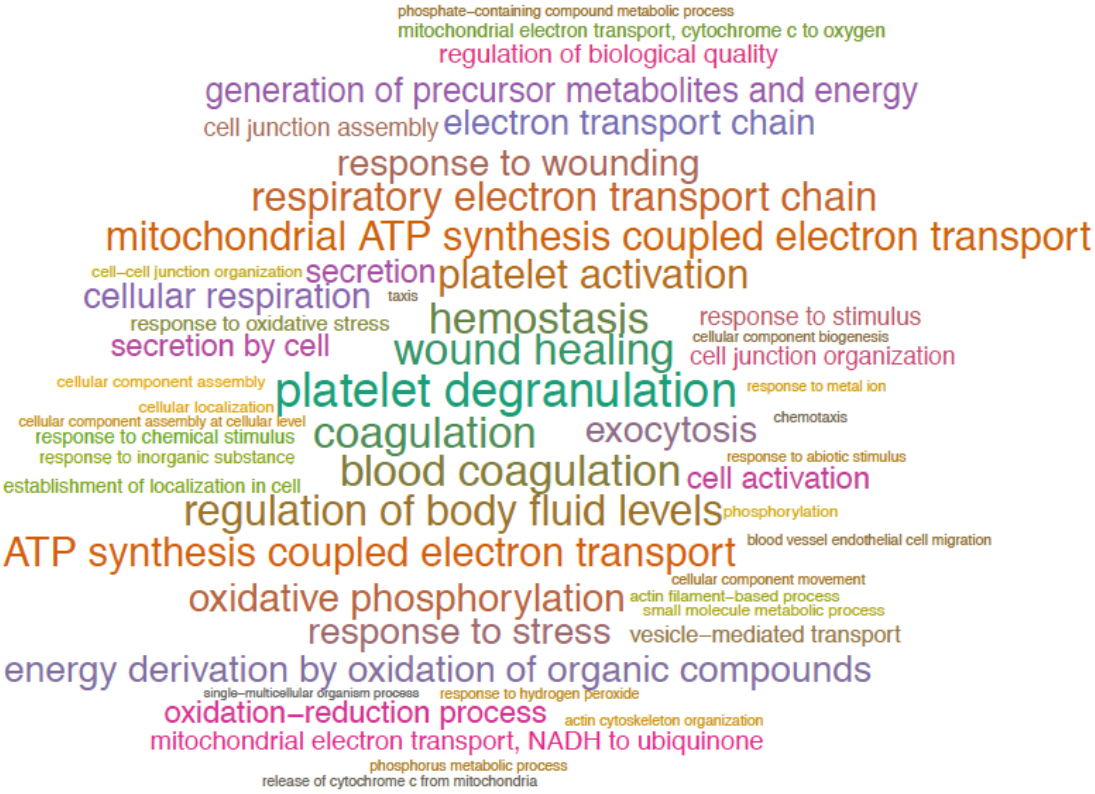


Figure 5.25 Over-represented gene ontologies for abundantly expressed genes in platelets from AMI and control groups that have common expression.
The full GO analysis may be found in Appendix 5, Table S5.7. The word cloud is a visual representation of the over-represented GO biological processes with their size an indicator of significance.

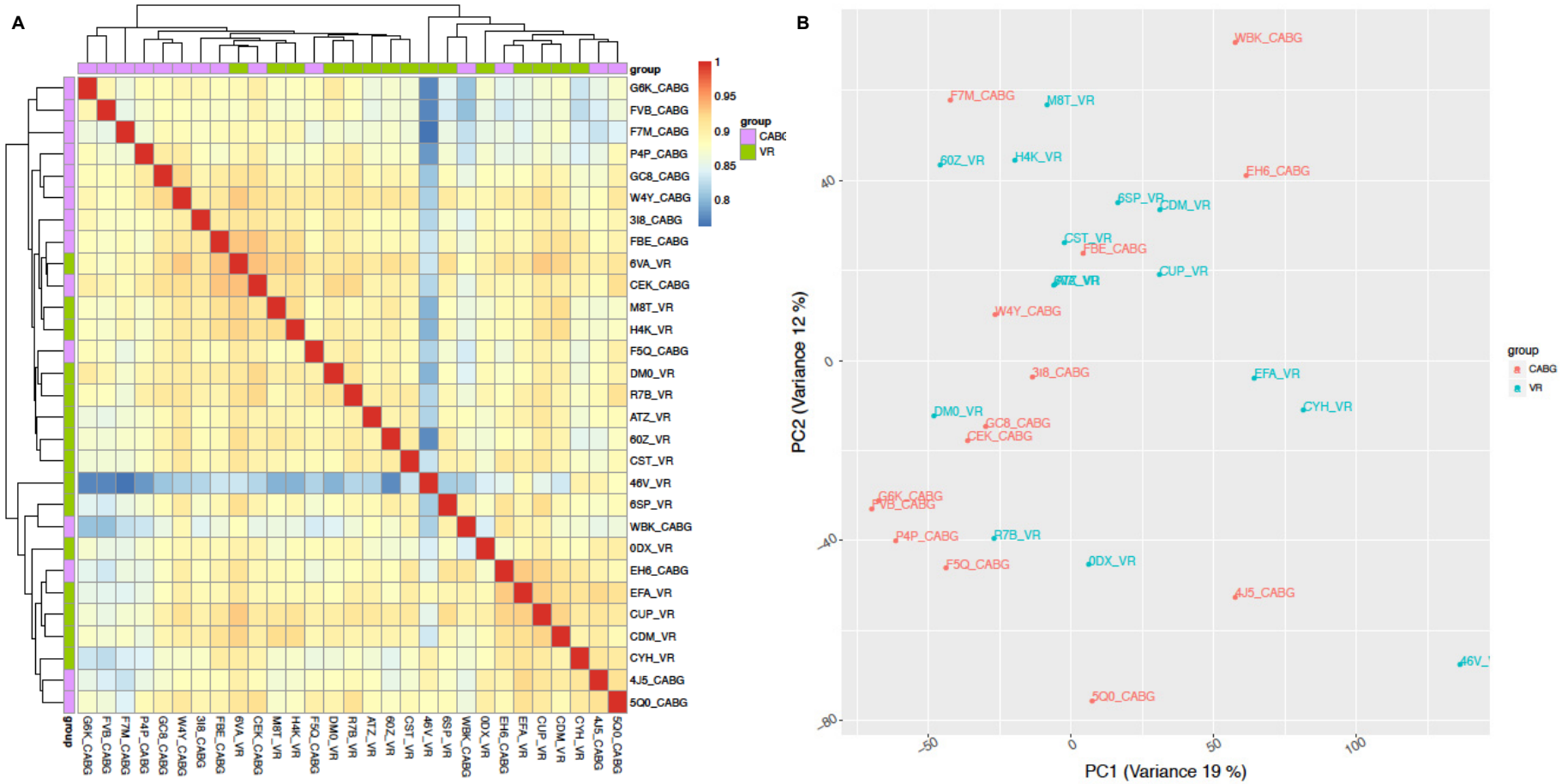


Figure 5.26 Clustering of platelet transcriptome data by gene expression

A. Heatmap showing hierarchical clustering of gene expression correlation between samples (mean-centred normalized and corrected log-expression values). Dendrograms are formed by hierarchical clustering on the Euclidean distances between genes. Coloured bars on both axes indicates whether the sample is from the AMI or control group (CABG denotes the AMI group: recent infarct with severe coronary disease requiring coronary artery bypass grafting – CABG; VR denotes the control group who has valve replacement cardiac surgery – VR). B. PCA plot constructed from normalized log-expression values of correlated highly variable genes, where each point represents a sample (labelled with patient ID and group: CABG: AMI group; VR: control group). First and second principal components are shown, along with the percentage of variance explained.

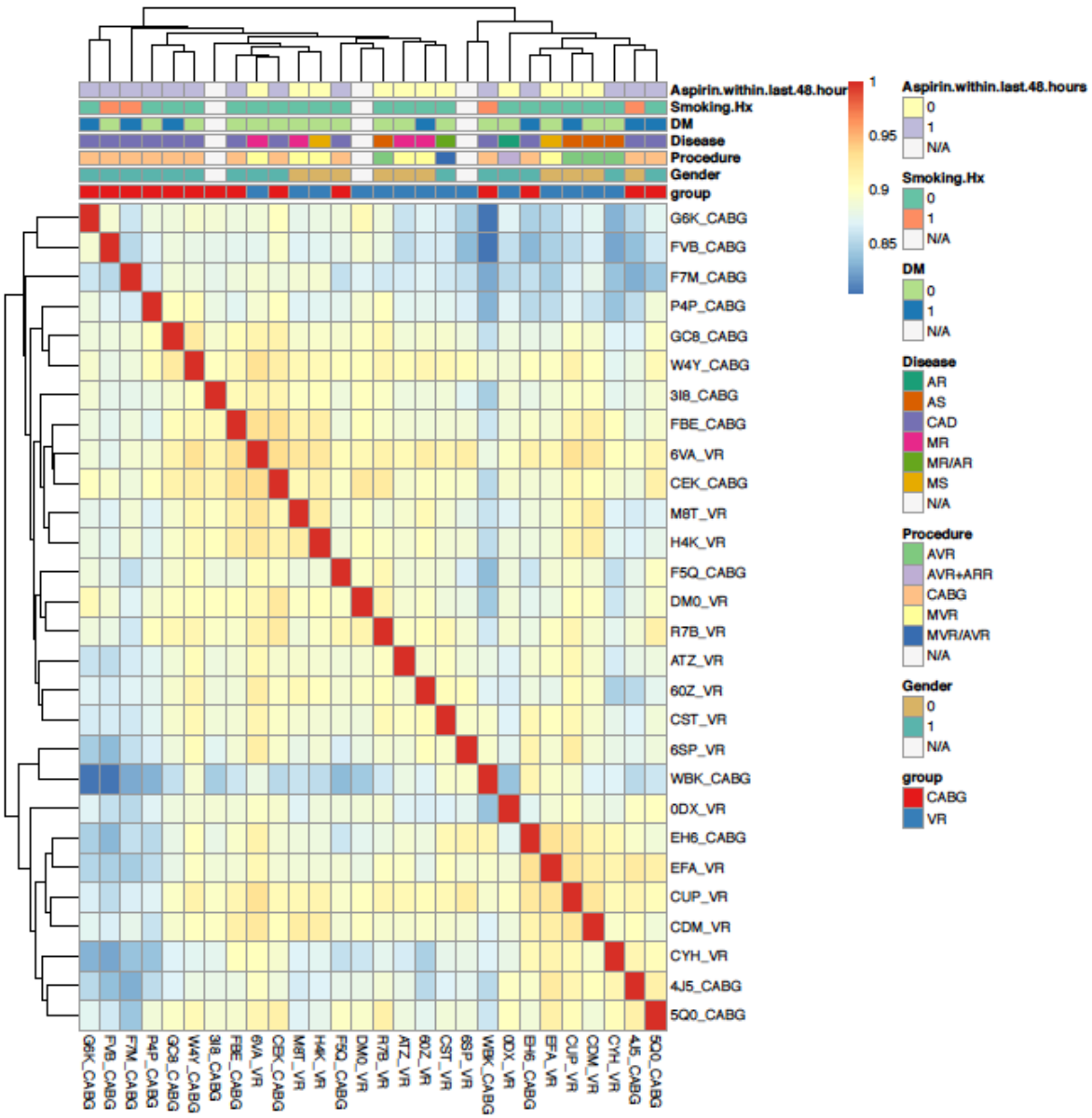


Figure 5.27 Clustering of platelet transcriptome data by gene expression, association with patient factors.

A. Heatmap showing hierarchical clustering of gene expression correlation between samples (mean-centred normalized and corrected log-expression values). Dendrograms are formed by hierarchical clustering on the Euclidean distances between genes. Coloured bars on top axis indicates categorical clinical data pertaining to individual patients: Aspirin within last 48 hours; smoking history; DM: diabetes; disease: AR: aortic regurgitation, AS: aortic stenosis, CAD: coronary artery disease, MR: mitral regurgitation, MS: mitral stenosis; procedure: AVR: aortic vave replacement, ARR: aortic root replacement, CABG: coronary artery bypass grafting, MVR: mitral valve replacement. In terms of group: CABG denotes the AMI group: recent infarct with severe coronary disease requiring coronary artery bypass grafting (CABG), VR denotes the control group who has valve replacement cardiac surgery (VR).

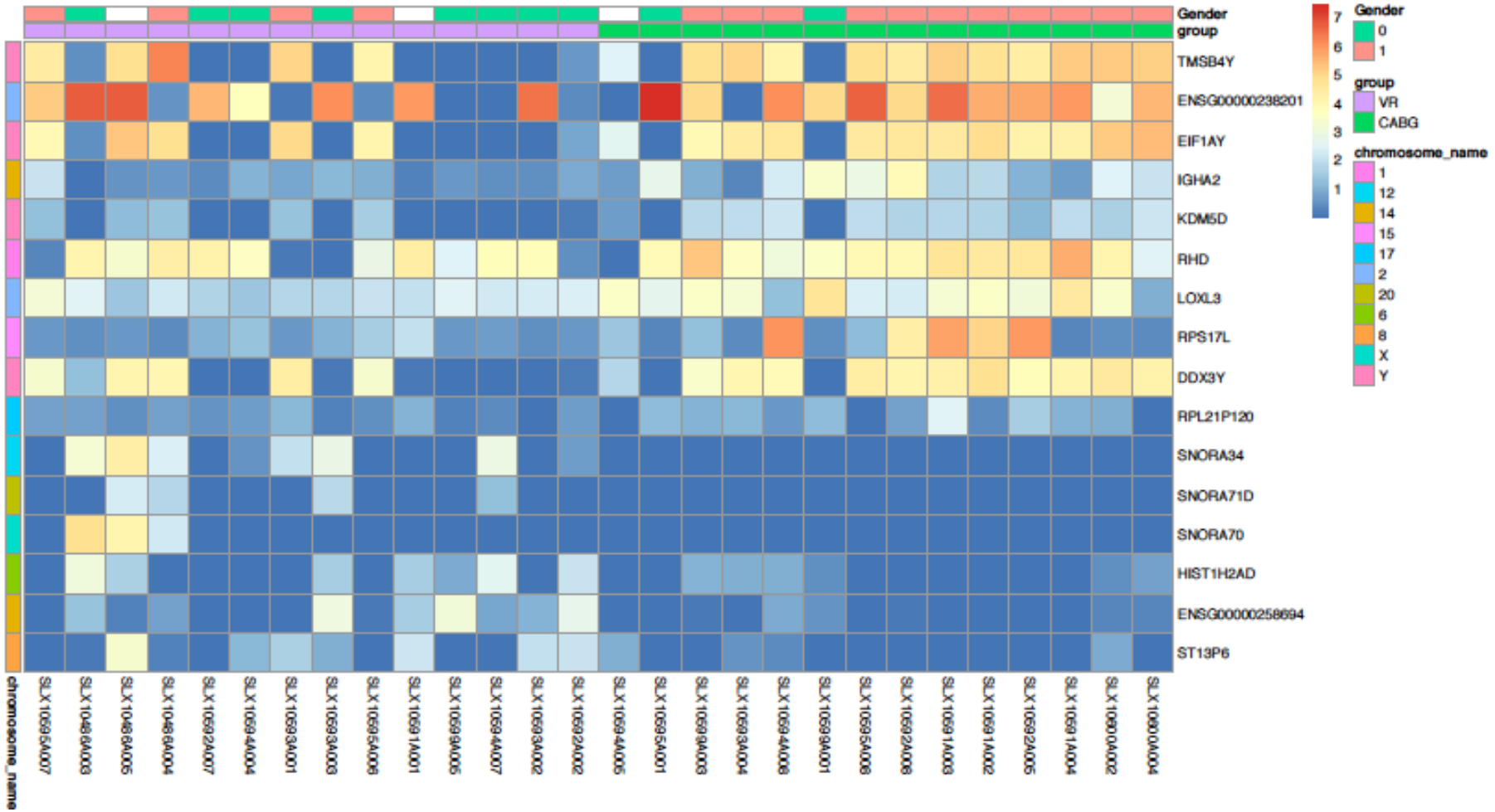


Figure 5.28 Differential gene expression between platelets from patients with recent AMI and control group

Heatmap of mean-centred normalized and corrected log-expression values for differentially expressed genes between the AMI and control groups. Columns are individual samples with column colours representing group and gender. Rows are differentially expressed genes with HGNC ID and row colours represent chromosome name. CABG denotes the AMI group: recent infarct with severe coronary disease requiring coronary artery bypass grafting (CABG), VR denotes the control group who has valve replacement cardiac surgery (VR).

5.4 Discussion

The aim of this chapter was to investigate platelet and megakaryocyte phenotype and transcriptomes in the setting of recent acute coronary thrombosis causing AMI. The most common pathophysiology is acute atherosclerotic plaque rupture combined with platelet hyperactivity. This occurs on the background of a chronic inflammatory process of atherosclerosis that is characterised by abnormal lipid metabolism, endothelial cell dysfunction and inflammatory cell activation. The final common pathway of platelet thrombosis is the target of the acute management of AMI with antiplatelet agents: COX1 inhibition using aspirin, P2Y12 inhibition using clopidogrel, prasugrel or ticagrelor, GP2b/3a inhibition using antagonists such as eptifibatide, abciximab or tirofiban and reperfusion with pPCI. There is now accumulating clinical observational data to show that in the acute phase MPV as well as platelet turnover measured by reticulated platelet count (IPF) are increased despite the use of contemporary antiplatelet agents (Sansanayudh et al. 2014). It is unclear if this observation is a cause or consequence of the acute event, however it has been linked to prognosis thereafter (re-infarction and mortality).

Whether MPV itself is related to the development of coronary artery disease or AMI at the population level has recently been challenged in an imputational GWAS analysis of 173,480 individuals (Astle et al. 2016). However, the implication of the clinical observation in the acute phase of infarct, whether cause or consequence, is that there is a pathological cycle of production of more reactive platelets or accelerated thrombopoiesis, the control of which if better understood could offer further therapeutic opportunities including novel agents or information that may guide choice and dosing strategy of current agents (Funck-Jensen et al. 2013; Armstrong et al. 2017; Bernlochner et al. 2015).

Based on this clinical observation, the setting of acute coronary thrombosis was used as a model for accelerated thrombopoiesis with the aim to gain further insight into drivers of megakaryopoiesis and platelet production, as well as the pathways that might be upregulated in this disease group. To do this, the phenotype and transcriptome of platelets and megakaryocytes was analysed in patients with severe coronary disease with recent acute coronary thrombosis and a control group without coronary artery disease.

The primary objective of the study was to measure megakaryocyte data which required bone marrow sampling. Therefore, the clinical study was designed around this primary aim. The disease population (AMI group) were patients who had a recent AMI and required coronary artery bypass grafting for severe coronary disease. The control population were also undergoing cardiac surgery (aortic or mitral valve replacement) but had no evidence of coronary artery disease on coronary angiography. Megakaryocytes were isolated from bone marrow taken as a scraping at the time of median sternotomy and blood samples for platelet

parameters, function and transcriptome analysis was taken on the day of surgery. Blood and bone marrow was sampled from 46 individuals for megakaryocyte and platelet phenotype data. Megakaryocyte transcriptomes were sequenced from 15 individuals and platelet transcriptomes sequenced from 29 individuals. As expected, there were significantly more males in the AMI group (p=0.013), with a significantly higher incidence of hypertension (p=0.01560), statin use (p=0.0156) and smoking (p=0.0009). There was also a significantly higher level of residual antiplatelet activity in the AMI group at the time of sampling. Interestingly, there was also a significantly lower platelet count in the control group.

In this study, no significant difference was found between MPV in the AMI group compared to control, in keeping with this there was also no difference in the reticulated platelet population between the groups. These two parameters were shown to be highly correlated in line with previous studies (Grove, Hvas, and Kristensen 2008). Despite this, overall platelet count at the time of sampling was significantly increased in the AMI group with a correspondingly greater circulating platelet mass; another platelet marker that has been shown to be increased in AMI and have prognostic value (Ergelen and Uyarel 2014). These findings suggest that there is a difference in platelet production between the two groups. Furthermore, when resting P-selectin levels were measured as a marker of baseline platelet activation, there was a significantly higher surface expression in platelets from the AMI group compared with the controls and a healthy volunteer cohort. This is a finding that corroborates previous evidence to show increased P-selectin expression on platelets after AMI compared with controls (Ault et al. 1999; Shimomura et al. 1998; Michelson 2004). In platelet function testing however, platelets from the AMI group showed significantly lower responses to ADP shown to be due to residual P2Y12 inhibitor activity in the patients who had taken these drugs in the last 5 days. There were no other significant differences between the AMI and control group in terms of platelet function.

One explanation for finding no difference in MPV is the possibility of chronically lower platelet count and increased MPV in patients with valve disease which could have ameliorated any difference seen between the two groups. The pre-operative platelet counts in the controls (taken approximately 2 weeks prior to planned surgery) were seen to be significantly lower than the AMI group, which could have led to a higher than expected MPV. The length of time prior to sampling that this platelet count was recorded may reflect chronicity. Increased MPV has been reported in a number of small studies in mitral valve prolapse (Icli et al. 2013), mitral stenosis (Varol et al. 2009) and aortic stenosis (Varol et al. 2011). Another explanation for this observation could be length of time after AMI that the samples were taken. The need for bone marrow at the same time as blood samples meant that the time of sampling was at best in the sub-acute phase. These were patients who had been admitted with AMI some of whom had been stented acutely and were scheduled to undergo either inpatient or early outpatient coronary artery bypass grafting. The mean number of days from infarction was 39 days. It is

plausible that if indeed raised MPV is a consequence to acute systemic inflammation at the time of acute thrombosis, it may have reached its peak and started to decline. However, there is evidence to show that MPV remains elevated for up to 4 months post AMI (Martin et al. 1983; Cameron et al. 1983; Kishk, Trowbridge, and Martin 1985). Furthermore, patients in the AMI group all had significant atherosclerotic disease burden to warrant coronary bypass surgery and raised MPV and reticulated platelet population has also been demonstrated in this cohort of patients (Guthikonda et al. 2008; Cesari et al. 2008).

No differences in terms of frequency or level of polyploidisation were found between megakaryocytes of patients with AMI compared with control. Assessment of megakaryocyte ploidy has previously been made in healthy individuals and in patients with chronic thrombocytopaenia (Mazur et al. 1988; Tomer et al. 1989). While in healthy bone marrow the mean megakaryocyte ploidy is 16N, in accelerated thrombopoietic states where there is either acute or chronic platelet need, MPV is seen to increase as well as an increase in mean megakaryocyte ploidy to 32N. The hypothesis at the beginning of this project was that acute myocardial infarction may be analogous these states of increased haemostatic demand as the reticulated platelet population is known to be increased, which is the only clinically feasible marker of megakaryocyte activity. Therefore, it was hypothesised that the escalation in platelet production is related to a transition in megakaryocyte phenotype which might be an increase in ploidy, which was found not to be the case. Megakaryocytes from both the AMI and control groups were found have a mean ploidy of 16N as has been shown previously in healthy individuals.

In order to assess if changes in megakaryocyte activity and platelet production were due to changes in gene expression, transcriptome analysis of megakaryocytes from the AMI and control groups was performed. The most highly expressed genes in megakaryocytes from patients with AMI were highly comparable with those from the control group with an overlap of 73 genes. Similar to the results in Chapter 4, the over-represented GO categories were cellular metabolism with a high concentration of mitochondrial reads, coagulation, haemostasis, defense and translation. There were a number of platelet function related transcripts highly expressed in both groups. *THBS1*, the gene encoding thrombospondin 1 and *CD47* for which thrombospondin is a ligand were highly expressed in megakaryocytes from the AMI group but not seen in the most highly expressed gene in the control group. *S100A9* that encodes the S100 calcium binding protein A9 (*MRP14*) is a gene that has been found to be differentially expressed in platelet transcriptomes of patients with STEMI compared with controls (Healy et al. 2006). Here in megakaryocytes it was found to be expressed at very high levels in both AMI and control groups.

A differential expression analysis between megakaryocytes from AMI and control group revealed 810 transcripts, demonstrating considerable differences in megakaryocytes from

these groups at the transcript level. 670 of these transcripts were significantly downregulated in patients with AMI with the major ontological bias being related to translational processes, protein localisation and immune response. 140 transcripts were upregulated in the AMI group. Although there were no GO biological processes that were significantly enriched, 11 out of the 140 genes encoded for immunoglobulin domains. In comparison, only four immunoglobulin domains were significantly downregulated despite a number of downregulated immune system and defense GO categories. Immunoglobulin domains have been seen to be expressed throughout the bone marrow megakaryocyte sequencing although the implications are unclear. There are also 36 unannotated transcripts that are significantly upregulated.

Interestingly of the upregulated genes there are three that are directly involved in platelet function. *PPBP* is significantly upregulated in megakaryocytes from AMI patients. It is highly specific to the megakaryocyte/platelet lineage and is located on chromosome 4. It encodes a platelet-derived growth factor also known as CXCL7. It is packaged in the alpha granule and released on platelet activation with its main role as a chemoattractant for neutrophils although it also stimulates various other cellular processes including mitosis, glycolysis, prostaglandin E2 secretion, and secretion of plasminogen activator by synovial cells.

THBS1 which was seen in the most highly expressed gene list for AMI patients is also significantly upregulated in this group. It encodes thrombospondin 1 which is a glycoprotein involved in cell-cell and cell-matrix interactions. It is located on chromosome 15 and like PPBP is released from the platelet alpha granule upon activation. It binds to a number of surface receptors including CD47 which was also highly expressed in the AMI group. It also binds fibrinogen, fibronectin, laminin, type V collagen and CD36. Through its binding to CD47 and CD36, it is known to be a critical regulator of angiogenesis (Isenberg et al. 2006; Isenberg et al. 2007). It may also have a regulatory role in terms of a negative feedback mechanism on megakaryopoiesis via its CD36 receptor through which it can lead to megakaryocyte apoptosis when platelet need has been met (M. Yang et al. 2003, 2016). Thrombospondin and its CD47 receptor have also been linked to cardiovascular disease and is known to be upregulated in endothelial cells in atherosclerosis (Rogers et al. 2014). In fact, polymorphism in THBS1 has been linked with premature AMI (Topol et al. 2001).

The upregulation of both *PPBP* and *THSB1* in the megakaryocytes of patients with recent AMI compared with controls suggests an increased protein expression within the platelets in line with increased platelet activity and consequently pathogenicity.

RAP1B is another gene significantly upregulated in megakaryocytes from AMI patients. It is located on chromosome 12 and encodes a small GTPase from the RAS family involved in a common critical step in platelet activation through a number of pathways. RAP1B is

stimulated downstream through both collagen binding to GP6 (Gilio et al. 2009) and ADP binding to P2Y12 receptors (Larson et al. 2003). In turn, it is then plays a critical role in regulating GPIIb/IIIa activation and its affinity state (Crittenden et al. 2004; Bernardi et al. 2006). It is essential to normal platelet function and RAP1B deficiency in animal models leads to a mild bleeding disorder, with *RAP1B* knockout mice exhibiting a more dramatic bleeding phenotype but are protected from thrombosis (Chrzanowska-Wodnicka et al. 2005). Since RAP1B is involved in a number of platelet activation pathways it has been proposed a novel therapeutic target for antiplatelet agents (Steinhubl et al. 2008). The fact that in the megakaryocyte transcriptome of patients post AMI, *RAP1B* is significantly upregulated despite some residual P2Y12 inhibition suggests that it may have a pathological role in this setting.

Other upregulated genes in megakaryocytes from AMI patients include a number of transmembrane proteins. One of these is *GRIA1*, located on chromosome 5 encodes an AMPA glutamate receptor that is known to be transcribed in platelets and megakaryocytes (Morrell et al. 2008). Glutamate is released from platelet dense granules and plasma levels are increased in cerebral thrombosis (Mallolas et al. 2006) and have also been linked to coronary atherosclerosis and myocardial ischaemia (Djekic et al. 2015; Liu et al. 2010). Antagonism of glutamate signalling through the AMPA receptor leads to resistance to thrombosis *in vitro* (Morrell et al. 2008) and glutamate NMDA receptor antagonism in megakaryocyte cell lines has shown defective megakaryopoiesis (Genever et al. 1999). This suggests that upregulation of glutamate signalling through the GRIA1 receptor in megakaryocytes in the setting of AMI may stimulate megakaryopoiesis and lead to accelerated platelet production. Enhanced thrombopoiesis by stimulation of megakaryocyte maturation has recently been described in the setting of atherosclerosis. Pathogen-associated molecular patterns released by the atherosclerotic plaque stimulate TLR2 on the megakaryocyte surface leading to enhanced megakaryopoiesis (Beaulieu et al. 2011). Defective cholesterol efflux that occurs in atherosclerosis is also thought to enhance megakaryopoiesis via the cholesterol transporter ABCG4 as its deficiency leads to enhanced TPO signalling and platelet production in an animal model (Murphy et al. 2013, 2014). Probably the most compelling evidence to suggest that enhanced megakaryopoiesis may be an important pathogenic driver in the setting of atherosclerosis and myocardial infarction is from the positive results of the recent CANTOS trial (Ridker et al. 2017). IL1B is known to be increased in the setting of athersclerosis (Hayashi et al. 2010) and is also an important driver of megakaryopoiesis both *in vitro* and *in vivo* (Kimura et al. 1990; Jiang et al. 1994; Beaulieu et al. 2014) as well as playing a regulatory role in HSCs (Espín-Palazón et al. 2014). The CANTOS trial demonstrated an improvement in cardiovascular outcome in patients with previous myocardial infarction through the administeration of Canakinumab, a therapeutic monoclonal antibody targeting IL1B. While the data in this chapter does not show differential expression of IL1B or its receptors in patients with AMI compared to control, it could be

argued that the trial result may be in part related to modulation of megakaryopoiesis and thereby thrombopoiesis. The upregulation of *GRIA1* encoding the AMPA glutamate receptor seen in the AMI group could be another mechanism through which megakaryopoiesis is enhanced which may be a causal factor in acute coronary thrombosis.

The megakaryocyte gene signatures from the AMI group and from the control group were then superimposed on the Monocle plot clustering the single cell HSCs from Chapter 3 according to differentiation trajectories (Figure 5.21). While both HSC clusters 1 and 4 were hypothesised to be megakaryocyte-biased based on their similarity with gene signatures from previously published cultured megakaryocyte data, cluster 1 was hypothesised to lead to megakaryopoiesis in the steady state while cluster 4 was hypothesised to be mobilised towards the megakaryocyte lineage in states of accelerated thrombopoiesis. In this chapter the aim was to use the setting of acute coronary thrombosis as a model for accelerated thrombosis. Superimposing the megakaryocyte transcriptome data on the Monocle 2 plot showed that both disease and control gene signatures were highly expressed in cluster 1 HSCs with low levels of expression in all other HSC clusters. This finding would fit with the discussion above that the setting of AMI may not be an ideal model for increased haemostatic demand and stress megakaryopoiesis. However, there are a number of limitations of this study to bear in mind. Firstly, as discussed above the patient groups may not optimal to address the specific aims here. Due to need for bone marrow sampling the AMI group were patients with AMI mean of 39 days prior to recruitment so it could be argued this was not the acute phase. Furthermore, the control group did not consist of healthy volunteers rather patients with chronic valve disease and often impaired ventricular function. Secondly, due to technical challenges relating to overwhelming mitochondrial transcripts within the megakaryocyte RNA-seq libraries, samples were selected for sequencing on the basis of expression of *MT-RNR2* and *PPBP* using qPCR. This may well have introduced additional biases in the resulting transcriptome data.

In contrast with the megakaryocyte transcriptome data, analysis of the platelet transcriptome, showed only 10 differentially upregulated genes in the AMI compared to the controls. Many of these were Y-linked and therefore a likely result of the significantly higher numbers of males in the AMI group. In the top 100 abundantly expressed genes in both groups there was an overlap of 95 demonstrating a marked similarity in transcriptome between the two groups. To add confidence to this observation no pattern was identified on unsupervised clustering of samples even with other clinical parameters accounted for. This is in contrast to previously described gene expression data in coronary disease vs control cohorts. 54 transcripts were found to be differentially expressed in a comparison between STEMI patient and controls in an early study using microarrays (Healy et al. 2006). This study identified *CD69* and *S100A9* as important markers of disease. While both of these genes were switched on in our dataset there was no significant difference between the two groups. In another expression array study

GP1BA encoding for the VWF receptor was differentially expressed in platelets from NSTEMI patients compared to controls (Colombo et al. 2011). Again, while GP1BA was highly expressed in both the AMI and control group there was no significant differential expression found. In a GWAS analysis *COMMD7* and *LRRFIP1* were identified as being associated with the development of myocardial infarction (Goodall et al. 2010). Both of these genes were found equally expressed in both the AMI and control cohorts. Recently, platelet transcriptome data has been reported comparing STEMI and NSTEMI patients. Similarly, no significant differences were shown between these two clinical groups. This data corroborated by the finding presented in this thesis questions the significance of the platelet transcriptome despite evidence that platelets do possess translational machinery to form proteins. While acute coronary thrombosis is shown to be related to megakaryocyte transcriptional changes, platelet transcriptome in the setting of thrombosis and control show no differences. Therefore, it is possible that upregulation of specific megakaryocyte transcripts would lead to their translation within the megakaryocyte and thereby upregulation in the platelet proteome rather than effecting change within the platelet transcriptome.

This study found that while there were phenotypic differences in platelets from patients with recent AMI, namely increased platelet mass and resting platelet P-selectin levels, the widely reported difference in MPV levels were not reproduced. Furthermore, megakaryocyte ploidy was similar in both groups. Platelet transcriptome data from patients with recent AMI compared with controls demonstrated no significant differences. However, at the megakaryocyte level a number of genes that are directly associated with platelet activation were significantly upregulated. Importantly, *RAP1B* was found to be significantly upregulated, encoding a small GTPase that is vital in GPIIb/IIIa activation and has been implicated as a target for novel antiplatelet therapeutics. Furthermore, *GRIA1* encoding the AMPA receptor for glutamate was also upregulated. Systemic thrombosis has been shown to increase plasma glutamate levels and which have been previously shown to have an additional regulatory role in megakaryopoiesis. Therefore, it could be hypothesised that myocardial thrombosis and ischaemia through this AMPA receptor can affect megakaryopoiesis, promoting increased platelet development, which would warrant further investigation.

CHAPTER 6
Conclusions

In this chapter the main findings from my thesis have been summarised and future work based on these findings is discussed.

The novel findings presented in this thesis are:

1. One of the first demonstrations of transcriptionally distinct subpopulations within the human bone marrow HSC compartment, using single cell RNA-seq. Two megakaryocyte-biased subsets within the human bone marrow HSC compartment, with potentially differing functional relevance.
2. The first interrogation of the human bone marrow megakaryocyte transcriptome demonstrating regulation of gene expression according to ploidy level. Increasing megakaryocyte ploidy is transcriptionally associated with the downregulation of platelet specific genetic programs and the upregulation of translation and protein localisation as well as expression of a number of transmembrane receptors which might have functional relevance.
3. The first comparison of megakaryocyte transcriptome data in patients with AMI and severe coronary disease compared with a control group. There was an upregulation of platelet activation related gene expression in megakaryocytes isolated from patients with AMI and severe coronary artery disease.

6.1 The first demonstration of transcriptionally distinct subpopulations within the human bone marrow HSC compartment, using single cell RNA-seq.

The concept of functional heterogeneity within the HSC compartment has existed for over 50 years (Siminovitch, McCulloch, and Till 1963; Till, McCulloch, and Siminovitch 1964). My PhD work represents one of the first times single cell RNA-seq has been used to identify transcriptionally distinct cell populations within the human bone marrow HSC population. Based on gene expression, single HSCs clustered into 5 discrete subpopulations, using two independent clustering methods. Two of these subpopulations, Cluster 1 and Cluster 4 which accounted for 56 of 119 single HSCs, were identified as megakaryocyte-biased. This is on the basis of marker genes expressed by each cluster that distinguish these two clusters from each other and from the other three HSC clusters. HSCs in these two clusters differentially expressed genes that have been shown to be expressed highly in the megakaryocyte-platelet lineage according to previous expression array and transcriptome studies of platelets and CD34+ derived megakaryocytes.

Many of the HSC Cluster 1 marker genes encoded proteins relating to HSC function and megakaryopoiesis, with fewer directly related to platelet function. Furthermore, Cluster 1 also included genes previously shown to be expressed HSC, MPP, CMP, GMP and MEP (L. Chen et al. 2014). This suggests that cluster 1 HSCs are primed to produce megakaryocytes through the classical megakaryocyte lineage differentiation pathway. Interestingly, out of 313 marker genes identified for cluster 1, 77 were related to ribosomal subunits and GO functional enrichment analysis shows that translation and protein localisation have an important functional role in megakaryopoiesis from the HSC. By comparison, the majority of cluster 4 marker genes encoded proteins directly related to platelet activation. Moreover, the genes expressed were more specific to the megakaryocyte lineage than cluster 1 marker genes, with little expression of these genes observed in previous datasets in the megakaryocyte progenitors. This observation suggests that while Cluster 1 may represent HSCs that differentiate along the classical pathway, Cluster 4 represents HSCs with direct megakaryocyte lineage potential, independent of intermediate progenitors. Furthermore, HSCs in cluster 4 have already turned on genetic programs that are required for platelet function required post thrombopoiesis from the terminally differentiated megakaryocyte and are therefore primed for thrombopoiesis. Based on this, I hypothesise that while megakaryocytes originating from cluster 1 HSCs are those produced during steady state haematopoiesis (Figure 1.7), megakaryocytes originating from cluster 4 HSCs are those that are released in accelerated thrombopoiesis or stress thrombopoiesis which occurs in states of either physiological or pathological increased haemostatic demand where emergency platelet production is required (Martin et al. 2012; Haas et al. 2015).

The cells sorted and defined as HSCs for transcriptome analysis were functionally validated to be HSCs by engraftment and multilineage repopulation when transplanted into sub-lethally irradiated NSG mice. Further functional validation is required to investigate the lineage biases of clusters 1 and 4 using either *in vitro* clonal assays or transplant-based tracking of their reconstitution activity. This however, is reliant on prospective identification of cells belonging to these HSC clusters on the basis of surface marker expression. This could either be using a combination of the surface marker used to identify HSCs or with addition cell surface markers. Sorting greater numbers of cells for further transcriptome analysis would also give additional statistical power to the observations made thus far.

6.2 The first interrogation of the human bone marrow megakaryocyte transcriptome demonstrating regulation of gene expression according to ploidy level.

The transcriptome of human bone marrow megakaryocytes has not previously been profiled. This important in gaining further understanding into platelet production as well as into the origin of both thrombotic and bleeding disorders. There remains a lack of understanding of the

process of megakaryocyte maturation, its association with polyploidisation and platelet production which means that the number of platelets produced per megakaryocyte *in vitro* is >1000-fold lower than *in vivo* (Moreau et al. 2016; Feng et al. 2014; Nakamura et al. 2014; Y. Liu et al. 2015).

In this study, I have successfully obtained transcriptome data from human bone marrow megakaryocytes for the first time. Amongst the most abundantly expressed genes were those encoding for mitochondrial proteins and other proteins involved in cellular metabolism. This observation is indicative of the high levels of energy required to maintain polyploid cells with a large mature cytoplasm to support thrombopoiesis. When comparison was made with cultured megakaryocytes, *ex-vivo* bone marrow megakaryocytes also showed similar high levels of expression of megakaryocyte and platelet lineage genes. However, this was a limited comparison due to the inconsistencies of experimental procedures.

The ability to perform low input and single cell RNA-seq on human bone marrow megakaryocytes allowed for the investigation of changes in gene expression with increasing ploidy class. This study has for the first time shown that with increasing ploidy up to 32N there is a downregulation of genetic programs that are related to platelet functionality although these transcripts are still switched on and remain expressed at high levels within the higher ploidy megakaryocytes. Instead there is a marked upregulation of transcripts with increasing ploidy level encoding ribosomal subunits, and those associated with translation and protein localisation. Not only does this show that gene expression is indeed regulated with increasing ploidy and maturation, it reveals a clear transcriptional difference between the higher ploidy megakaryocytes which are most abundant *in vivo* and those grown *in vitro*. While high levels of platelet specific gene transcription occur in the lower ploidy classes as further polyploidisation occurs enough transcript has been generated and the focus is redirected to translation of mRNA to protein at the ribosome and appropriate localisation of the newly generated proteins in preparation for thrombopoiesis. The high translational requirement of bone marrow megakaryocytes, particularly those of higher ploidy was one of the most important observations, highlighting a focus on platelet protein production and packaging into alpha granules, dense granules and other vesicles, ready for thrombopoiesis and platelet activation. It also explains the high levels of mitochondrial and other transcripts involved in cellular metabolism since the production of ribosomes consumes large amounts of energy (Lane and Martin 2010). Further work would include differential expression analysis between individual pairs of ploidy classes to identify marker genes for each ploidy group, at the level of both multicellular pools and single cells. Deeper sequencing of megakaryocytes of different ploidy would also support the lowput/single cell data presented here but would rely on larger numbers of isolated cells or pooling samples from a number of individuals.

A number of genes upregulated with increasing ploidy were identified as transmembrane receptors, which upon binding to their ligands, might modulate functional effects such as megakaryocyte chemotactic migration from the osteoclastic niche to the sinusoidal space or indeed initiate platelet production. Future work on these transmembrane receptors and testing appropriate ligands to investigate their role in megakaryocyte maturation might ultimately help to improve *in vitro* platelet production by simulating the bone marrow niche.

6.3 The first comparison of megakaryocyte transcriptome data in patients with AMI and severe coronary disease compared with a control group.

In order to assess if there are transcriptional differences in megakaryocytes and platelets in accelerated and steady state thrombopoiesis, the clinical setting of AMI was used as a model, as acute changes in platelet parameters have been widely documented which may be cause or effect of acute platelet thrombosis (Martin et al. 2012). While there was no difference in platelet and megakaryocytes physical traits between patients with AMI and control there were significant megakaryocyte transcriptional differences. A number of genes upregulated in patients with acute myocardial infarction were directly related to platelet activation and proteins released by the alpha granule. The observation that *RAP1B* was upregulated in megakaryocytes in patients with AMI is interesting. *RAP1B* encodes a protein which is involved in a number of platelet activation pathways, it is stimulated by collagen binding to GP6 and ADP binding to P2Y12 and plays a critical role in modulating the affinity state of GP2b/3a (Crittenden et al. 2004; Bernardi et al. 2006). Importantly it has also been identified as a potential therapeutic target in AMI (Steinhubl et al. 2008). Another important finding was that the gene encoding the AMPA glutamate receptor, *GRIA1*, is also upregulated in megakaryocytes from patients with AMI. As glutamate serum levels are increased in thrombosis (Djekic et al. 2015; Z. Liu et al. 2010) and interruption of glutamate binding to the NMDA glutamate receptor in megakaryocyte cell lines has impaired megakaryopoiesis (Genever et al. 1999), this observation raises the possibility of a positive feedback mechanism of glutamate signalling leading to increased platelet production perpetuating a prothrombotic state.

No significant difference in platelet transcriptome were observed between patients with AMI compared with control, despite differences in megakaryocyte transcriptome. I therefore postulated that translation of differentially expressed megakaryocyte transcripts occurs within the megakaryocyte itself with abundance of protein production depending on transcript expression. Accordingly, proteins would already be packaged within the megakaryocyte, ready for platelet activation before the onset of platelet release, which would explain finding no significant transcriptional differences in patients with platelet thrombosis compared to controls. It is however possible that the transcripts found within the platelets may play a role in

post-translational protein modification. Further study of platelet proteomics in the two clinical groups would be required to investigate this hypothesis.

Ploidy and AMI megakaryocyte signatures are overlaid onto the monocle differentiation trajectories showing HSC clusters 1-5. Both upregulated and downregulated megakaryocyte signatures for ploidy and AMI were expressed highly in cluster 1 HSCs. This would suggest that the bone marrow megakaryocytes that were sequenced were all generated via the classical megakaryocyte lineage commitment pathway. None of the megakaryocyte genes regulated by ploidy or disease aligned to cluster 4 HSCs. This was proposed to be the HSC subpopulation that was primed for emergency or stress megakaryopoiesis and platelet production. It is not clear from the bone marrow megakaryocyte transcriptome data presented here what the functional relevance of this HSC is. Future work would include HSC transcriptome analysis in the setting of AMI or accelerated thrombopoiesis to ascertain if cluster 4 HSCs are indeed expanded. It is also reasonable to suggest that a different clinical setting may be a better model for accelerated thrombopoiesis or increased haemostatic demand such as immune thrombocytopaenia where increased megakaryocyte ploidy has been shown with increased MPV and IPF, or immediately following cardiopulmonary bypass.

6.4 Summary of future work

1. Prospective identification of HSC clusters 1 and 4 for functional validation (*in vitro* clonal assays and xenotransplantation) and further transcriptome analysis to increase statistical power.
2. Pair-wise differential expression analysis of megakaryocytes of different ploidy classes to identify transcriptional signatures for individual ploidy classes.
3. Functional work on transmembrane receptors upregulated in high ploidy megakaryocytes to further understand their role in megakaryocyte maturation and platelet production.
4. Functional work on megakaryocyte genes upregulated in the setting of AMI to further understand their role in megakaryocyte maturation and platelet production.
5. Proteomic study of platelets in the setting of AMI with the assumption that this is more reflective of platelet prothrombotic potential than platelet transcriptome.
6. HSC single cell transcriptome analysis in setting of AMI to compare transcriptionally distinct subpopulations between AMI and controls.
7. Use of a different model for accelerated thrombopoiesis.

REFERENCES

Adolfsson, J, O J Borge, D Bryder, K Theilgaard-Mönch, I Astrand-Grundström, E Sitnicka, Y Sasaki, and S E Jacobsen. 2001. “Upregulation of Flt3 Expression within the Bone Marrow Lin(-)Sca1(+)c-Kit(+) Stem Cell Compartment Is Accompanied by Loss of Self-Renewal Capacity.” *Immunity* 15 (4):659–69.

Adolfsson, J, R Månsson, N Buza-Vidas, A Hultquist, K Liuba, CT Jensen, D Bryder, et al. 2005. “Identification of Flt3+lympho-Myeloid Stem Cells Lacking Erythro-Megakaryocytic Potential: A Revised Road Map for Adult Blood Lineage Commitment.” *Cell* 121 (2):295–306.

Akashi, K., X He, J Chen, H Iwasaki, C Niu, B Steenhard, J Zhang, J Haug, and L Li. 2003. “Transcriptional Accessibility for Genes of Multiple Tissues and Hematopoietic Lineages Is Hierarchically Controlled during Early Hematopoiesis.” *Blood* 101 (2):383–89.

Akashi, K, D Traver, T Miyamoto, and IL Weissman. 2000. “A Clonogenic Common Myeloid Progenitor That Gives Rise to All Myeloid Lineages.” *Nature* 404 (6774):193–97.

Aksu, H, O Ozer, H Unal, G Hobikoglu, T Norgaz, A Buturak, O Soylu, and A Narin. 2009. “Significance of Mean Platelet Volume on Prognosis of Patients with and without Aspirin Resistance in Settings of Non-ST-Segment Elevated Acute Coronary Syndromes.” *Blood Coagul Fibrinolysis* 20 (8):686–93.

Alexander, W S, A W Roberts, N A Nicola, R Li, and D Metcalf. 1996. “Deficiencies in Progenitor Cells of Multiple Hematopoietic Lineages and Defective Megakaryocytopoiesis in Mice Lacking the Thrombopoietic Receptor c-Mpl.” *Blood* 87 (6):2162–70.

Arai, F, A Hirao, M Ohmura, H Sato, S Matsuoka, K Takubo, K Ito, G Y Koh, and T Suda. 2004. “Tie2/Angiopoietin-1 Signaling Regulates Hematopoietic Stem Cell Quiescence in the Bone Marrow Niche.” *Cell* 118 (2):149–61.

Arinobu, Y, S Mizuno, Y Chong, H Shigematsu, T Iino, H Iwasaki, T Graf, et al. 2007. “Reciprocal Activation of GATA-1 and PU.1 Marks Initial Specification of Hematopoietic Stem Cells into Myeloerythroid and Myelolymphoid Lineages.” *Cell Stem Cell* 1 (4):416–27.

Armstrong, P C, T Hoefer, R B Knowles, A T Tucker, M A Hayman, P M Ferreira, M V Chan, and T D Warner. 2017. “Newly Formed Reticulated Platelets Undermine Pharmacokinetically Short-Lived Antiplatelet Therapies.” *Arteriosclerosis, Thrombosis, and Vascular Biology* 37 (5):949–56.

Aslam, R, E R Speck, M Kim, A R Crow, K W A Bang, F P Nestel, H Ni, A H Lazarus, J Freedman, and J W Semple. 2006. “Platelet Toll-like Receptor Expression Modulates Lipopolysaccharide-Induced Thrombocytopenia and Tumor Necrosis Factor- Production in Vivo.” *Blood* 107 (2):637–41.

Astle, WJ., H Elding, T Jiang, D Allen, D Ruklisa, AL Mann, D Mead, et al. 2016. “The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex

Disease." *Cell* 167 (5):1415–1429.e19.

Ault, K A, C P Cannon, J Mitchell, J McCahan, R P Tracy, W F Novotny, J D Reimann, and E Braunwald. 1999. "Platelet Activation in Patients after an Acute Coronary Syndrome: Results from the TIMI-12 Trial." *Journal of the American College of Cardiology* 33 (3).

Ault, KA, and C Knowles. 1995. "In Vivo Biotinylation Demonstrates That Reticulated Platelets Are the Youngest Platelets in Circulation." *Exp Haematol* 23 (9):996.

Ault, K A, H M Rinder, J Mitchell, M B Carmody, C P Vary, and R S Hillman. 1992. "The Significance of Platelets with Increased RNA Content (Reticulated Platelets). A Measure of the Rate of Thrombopoiesis." *Am J Clin Path* 98 (6):637.

Auton, A, G R Abecasis, D M Altshuler, R M Durbin, G R Abecasis, D R Bentley, A Chakravarti, et al. 2015. "A Global Reference for Human Genetic Variation." *Nature* 526 (7571). Nature Publishing Group:68–74.

Avecilla, S T, K Hattori, B Heissig, R Tejada, F Liao, K Shido, D K Jin, et al. 2004. "Chemokine-Mediated Interaction of Hematopoietic Progenitors with the Bone Marrow Vascular Niche Is Required for Thrombopoiesis." *Nat Med* 10 (1):64–71.

Azab, B, E Torbey, J Singh, M Akerman, G Khoueiry, J T Mcginn Jr, W D Widmann, and J Lafferty. 2011. "Mean Platelet Volume/platelet Count Ratio as a Predictor of Long-Term Mortality after Non-ST-Elevation Myocardial Infarction." *Platelets*,0:1–10.

Azcoitia, V, M Aracil, C Martínez-A, and M Torres. 2005. "The Homeodomain Protein Meis1 Is Essential for Definitive Hematopoiesis and Vascular Patterning in the Mouse Embryo." *Developmental Biology* 280 (2):307–20.

Babraham Institute. 2014. "Trim Galore!" http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.

Bain, B J. 1996. "Ethnic and Sex Differences in the Total and Differential White Cell Count and Platelet Count." *Journal of Clinical Pathology* 49 (8):664–66.

Balduini, A, A Malara, A Pecci, S Badalucco, V Bozzi, I Pallotta, P Noris, M Torti, and C L Balduini. 2009. "Proplatelet Formation in Heterozygous Bernard-Soulier Syndrome Type Bolzano." *Journal of Thrombosis and Haemostasis* 7 (3):478–84.

Bartley, T D, J Bogenberger, P Hunt, Y S Li, H S Lu, F Martin, M S Chang, et al. 1994. "Identification and Cloning of a Megakaryocyte Growth and Development Factor That Is a Ligand for the Cytokine Receptor Mpl." *Cell* 77 (7). UNITED STATES:1117–24.

Bath, P M. 1993. "The Routine Measurement of Platelet Size Using Sodium Citrate Alone as the Anticoagulant." *Thrombosis and Haemostasis* 70 (4):687–90.

Beaulieu, L M, E Lin, K M Morin, K Tanriverdi, and J E Freedman. 2011. "Regulatory Effects of TLR2 on Megakaryocytic Cell Function." *Blood* 117 (22). American Society of Hematology:5963–74.

Beaulieu, L M, E Lin, E Mick, M Koupenova, E O Weinberg, C D Kramer, C A Genco, et al. 2014. "Interleukin 1 Receptor 1 and Interleukin 1β Regulate Megakaryocyte Maturation, Platelet Activation, and Transcript Profile during Inflammation in Mice and Humans." *Arteriosclerosis, Thrombosis, and Vascular Biology* 34 (3):552–64.

Behnke, O, and A Forer. 1998. “From Megakaryocytes to Platelets: Platelet Morphogenesis Takes Place in the Bloodstream.” *Eur J Haematol Suppl* 61:3–23.

Behrens, W E von. 1972. “Evidence of Phylogenetic Canalisation of the Circulating Platelet Mass in Man.” *Thrombosis et Diathesis Haemorrhagica* 27 (1):159–72.

Bernardi, B, G F Guidetti, F Campus, J R Crittenden, A M Graybiel, C Balduini, and M Torti. 2006. “The Small GTPase Rap1b Regulates the Cross Talk between Platelet Integrin 2beta1 and Integrin IIbbeta3.” *Blood* 107 (7):2728–35.

Bernlochner, I, A Goedel, C Plischke, S Schupke, B Haller, C Schulz, K Mayer, et al. 2015. “Impact of Immature Platelets on Platelet Response to Ticagrelor and Prasugrel in Patients with Acute Coronary Syndrome.” *European Heart Journal* 36 (45):3202–10.

Bertin, A, M C Mahaney, L A Cox, J Rogers, J L VandeBerg, C Brugnara, and O S Platt. 2007. “Quantitative Trait Loci for Peripheral Blood Cell Counts: A Study in Baboons.” *Mammalian Genome* 18 (5):361–72.

Bertozzi, C C, P R Hess, and M L Kahn. 2010. “Platelets: Covert Regulators of Lymphatic Development.” *Arteriosclerosis, Thrombosis, and Vascular Biology* 30 (12):2368–71.

Bessis, M. 1956. *Cytology of the Blood and Blood-Forming Organs*. Grune & Stratton New York.

Bessis, M. 1973. “The Thrombocytic Series.” In *Living Cells, Their Ultrastructure*, edited by M Bessis, 370. Springer, Berlin.

Bessman, J D. 1984. “The Relation of Megakaryocyte Ploidy to Platelet Volume.” *Am J Hematol* 16 (2):161–70.

Biino, G, C L Balduini, L Casula, P Cavallo, S Vaccargiu, D Parracciani, D Serra, L Portas, F Murgia, and M Pirastu. 2011. “Analysis of 12,517 Inhabitants of a Sardinian Geographic Isolate Reveals That Predispositions to Thrombocytopenia and Thrombocytosis Are Inherited Traits.” *Haematologica* 96 (1):96–101.

Bilen, E, I H Tanboga, M Kurt, U Kocak, H Ayhan, T Durmaz, and E Bozkurt. 2012. “Mean Platelet Volume Is Increased in Patients with Bicuspid Aortic Valve.” *Clinical and Applied Thrombosis/Hemostasis* 18 (4):351–55.

Bluteau, O., T. Langlois, P. Rivera-Munoz, F. Favale, P. Rameau, G. Meurice, P. Dessen, et al. 2013. “Developmental Changes in Human Megakaryopoiesis.” *Journal of Thrombosis and Haemostasis* 11 (9):1730–41.

Bouchard, B A, C S Catcher, B R Thrash, C Adida, and P B Tracy. 1997. “Effector Cell Protease Receptor-1, a Platelet Activation-Dependent Membrane Protein, Regulates Prothrombinase-Catalyzed Thrombin Generation.” *Journal of Biological Chemistry* 272 (14):9244–51.

Boudreau, L H, A Duchez, N Cloutier, D Soulet, N Martin, J Bollinger, A Paré, et al. 2014. “Platelets Release Mitochondria Serving as Substrate for Bactericidal Group IIA-Secreted Phospholipase A2 to Promote Inflammation.” *Blood* 124 (14):2173–83.

Boughton, B J, A Macwhannell, A Simpson, and R Hawker. 1990. “Platelet Size and Adenine Nucleotides in Patients Undergoing Bone Marrow Ablation: A Useful Model for Studying

Platelet Ageing." *Platelets* 1 (1):21–24.

Bray, P F, R A Mathias, N Faraday, L R Yanek, M D Fallin, J E Herrera-Galeano, A F Wilson, L C Becker, and D M Becker. 2007. "Heritability of Platelet Function in Families with Premature Coronary Artery Disease." *Journal of Thrombosis and Haemostasis : JTH* 5 (8):1617–23.

Breiman, L. 2001. "Random Forests." *Machine Learning* 45 (1). Kluwer Academic Publishers:5–32.

Briddell, R A, J E Brandt, J E Straneva, E F Srour, and R Hoffman. 1989. "Characterization of the Human Burst-Forming Unit-Megakaryocyte." *Blood* 74 (1):145–51.

Briggs, C, I Mellors, A Roderick, A Ward, C O'Malley, J Barker, B De La Salle, P McTaggart, K Hyde, and S J MacHin. 2009. "Quality Counts: New Parameters in Blood Cell Counting." *International Journal of Laboratory Hematology* 31 (3):277–97.

Briggs, C, S Kunka, D Hart, S Oguni, and S J Machin. 2004. "Assessment of an Immature Platelet Fraction (IPF) in Peripheral Thrombocytopenia." *British Journal of Haematology* 126 (1):93–99.

Brinkmann, V, U Reichard, C Goosmann, B Fauler, Y Uhlemann, D S Weiss, Y Weinrauch, and A Zychlinsky. 2004. "Neutrophil Extracellular Traps Kill Bacteria." *Science* 303 (5663):1532–35.

Brown, A S, Y Hong, A de Belder, H Beacon, J Beeso, R Sherwood, M Edmonds, J F Martin, and J D Erusalimsky. 1997. "Megakaryocyte Ploidy and Platelet Changes in Human Diabetes and Atherosclerosis." *Arterioscler Thromb Vasc Biol* 17 (4):802–7.

Bruns, I, D Lucas, S Pinho, J Ahmed, M P Lambert, Y Kunisaki, C Scheiermann, et al. 2014. "Megakaryocytes Regulate Hematopoietic Stem Cell Quiescence through CXCL4 Secretion." *Nature Medicine* 20 (11):1315–20.

Bugert, P, A Dugrillon, A Gunaydin, H Eichler, and H Kluter. 2003. "Messenger RNA Profiling of Human Platelets by Microarray Hybridization." *Thromb Haemost* 90 (4):738–48.

Bury, L, A Malara, P Gresele, and A Balduini. 2012. "Outside-in Signalling Generated by a Constitutively Activated Integrin AIIbB3 Impairs Proplatelet Formation in Human Megakaryocytes." *PLoS ONE* 7 (4):1–7.

Busch, K, K Klapproth, M Barile, M Flossdorf, T Holland-Letz, S M Schlenner, M Reth, T Höfer, and H Reimer Rodewald. 2015. "Fundamental Properties of Unperturbed Haematopoiesis from Stem Cells in Vivo." *Nature* 518 (7540):542–46.

Cabezas-Wallscheid, N, D Klimmeck, J Hansson, D B Lipka, A Reyes, Q Wang, D Weichenhan, et al. 2014. "Identification of Regulatory Networks in HSCs and Their Immediate Progeny via Integrated Proteome, Transcriptome, and DNA Methylome Analysis." *Cell Stem Cell* 15 (4):507–22.

Cameron, H A, R Phillips, R M Ibbotson, and P H Carson. 1983. "Platelet Size in Myocardial Infarction." *British Medical Journal (Clinical Research Ed.)* 287 (6390):449–51.

Cavazzana-Calvo, M, A Fischer, F D Bushman, E Payen, P Leboulch, et al. 2011. "Is Normal Hematopoiesis Maintained Solely by Long-Term Multipotent Stem Cells ?" *Blood* 117

(17):4420–24.

Cecchetti, L, N D Tolley, N Michetti, L Bury, A S Weyrich, and P Gresele. 2011. “Megakaryocytes Differentially Sort mRNAs for Matrix Metalloproteinases and Their Inhibitors into Platelets: A Mechanism for Regulating Synthetic Events.” *Blood* 118 (7):1903–11.

Cesari, F, R Marcucci, R Caporale, R Paniccia, E Romano, G F Gensini, R Abbate, and A M Gori. 2008. “Relationship between High Platelet Turnover and Platelet Function in High-Risk Patients with Coronary Artery Disease on Dual Antiplatelet Therapy.” *Thromb Haemost* 99 (5).

Cesari, F, R Marcucci, A M Gori, R Caporale, A Fanelli, G Casola, D Balzi, et al. 2013. “Reticulated Platelets Predict Cardiovascular Death in Acute Coronary Syndrome Patients. Insights from the AMI-Florence 2 Study.” *Thromb Haemost* 109 (5):846–53.

Chagraoui, H, M Kassouf, S Banerjee, N Goardon, K Clark, A Atzberger, A C Pearce, et al. 2011. “SCL-Mediated Regulation of the Cell-Cycle Regulator p21 Is Critical for Murine Megakaryopoiesis.” *Blood* 118 (3):723–35.

Chang, A N, A B Cantor, Y Fujiwara, M B Lodish, S Droho, J D Crispino, and S H Orkin. 2002. “GATA-Factor Dependence of the Multitype Zinc-Finger Protein FOG-1 for Its Essential Role in Megakaryopoiesis.” *Proceedings of the National Academy of Sciences* 99 (14):9237–42.

Chapman, L M, A A Aggrey, D J Field, K Srivastava, S Ture, K Yui, D J Topham, W M Baldwin, and C N Morrell. 2012. “Platelets Present Antigen in the Context of MHC Class I.” *Journal of Immunology (Baltimore, Md. : 1950)* 189 (2):916–23.

Chen, L, M Kostadima, J H A Martens, G Canu, S P Garcia, E Turro, K Downes, et al. 2014. “Transcriptional Diversity during Lineage Commitment of Human Blood Progenitors.” *Science* 345 (6204):1251033–1251033.

Chen, Q, F J De Sauvage, G Solar, and D L Eaton. 1998. “IL-3 Does Not Contribute to Platelet Production in c-Mpl-deficient Mice.” *Stem Cells* 16 (S1):31–36.

Choi, E S, J L Nichol, M M Hokom, A C Hornkohl, and P Hunt. 1995. “Platelets Generated in Vitro from Proplatelet-Displaying Human Megakaryocytes Are Functional.” *Blood* 85 (2):402–13.

Chrzanowska-Wodnicka, M, S S Smyth, S M Schoenwaelder, T H Fischer, and G C White. 2005. “Rap1b Is Required for Normal Platelet Function and Hemostasis in Mice.” *The Journal of Clinical Investigation* 115 (3):680–87.

Chu, H, W L Chen, C C Huang, H Y Chang, H Y Kuo, C M Gau, Y C Chang, and Y S Shen. 2011. “Diagnostic Performance of Mean Platelet Volume for Patients with Acute Coronary Syndrome Visiting an Emergency Department with Acute Chest Pain: The Chinese Scenario.” *Emerg Med J* 28 (7). England:569–74.

Chu, S G, R C Becker, P B Berger, D L Bhatt, J W Eikelboom, B Konkle, E R Mohler, M P Reilly, and J S Berger. 2010. “Mean Platelet Volume as a Predictor of Cardiovascular Risk: A Systematic Review and Meta-Analysis.” *Journal of Thrombosis and*

Haemostasis 8 (1):148–56.

Coban, E, M Ozdogan, G Yazicioglu, and F Akcit. 2005. “The Mean Platelet Volume in Patients with Obesity.” *International Journal of Clinical Practice* 59 (8):981–82.

Colombo, G, K Gertow, G Marenzi, M Brambilla, M De Metrio, E Tremoli, and M Camera. 2011. “Gene Expression Profiling Reveals Multiple Differences in Platelets from Patients with Stable Angina or Non-ST Elevation Acute Coronary Syndrome.” *Thrombosis Research* 128 (2):161–68.

Corash, L, H Y Chen, J Levin, G Baker, H Lu, and Y Mok. 1987. “Regulation of Thrombopoiesis: Effects of the Degree of Thrombocytopenia on Megakaryocyte Ploidy and Platelet Volume.” *Blood* 70 (1):177–85.

Corash, L, and J Levin. 1990. “The Relationship between Megakaryocyte Ploidy and Platelet Volume in Normal and Thrombocytopenic C3H Mice.” *Exp Hematol* 18 (9):985–89.

Cowley, D O, J A Rivera-Perez, M Schliekelman, Y J He, T G Oliver, L Lu, R O’Quinn, E D Salmon, T Magnuson, and T Van Dyke. 2009. “Aurora-A Kinase Is Essential for Bipolar Spindle Formation and Early Development.” *Mol Cell Biol* 29 (4):1059–71.

Crittenden, J R, W Bergmeier, Y Zhang, C L Piffath, Y Liang, D D Wagner, D E Housman, and A M Graybiel. 2004. “CalDAG-GEFI Integrates Signaling for Platelet Aggregation and Thrombus Formation.” *Nature Medicine* 10 (9):982–86.

D’Andrea, D, L Grassi, M Mazzapioda, and A Tramontano. 2013. “FIDEA: A Server for the Functional Interpretation of Differential Expression Analysis.” *Nucleic Acids Research* 41 (W1):W84–88.

Dale, G L, P Friese, L Hynes, and S Burstein. 1995. “Demonstration That Thiazole-Orange-Positive Platelets in the Dog Are Less than 24 Hours Old.” *Blood* 85 (7):1822–25.

Debili, N, L Coulombel, L Croisille, A Katz, J Guichard, J Breton-Gorius, and W Vainchenker. 1996. “Characterization of a Bipotent Erythro-Megakaryocytic Progenitor in Human Bone Marrow.” *Blood* 88 (4):1284–96.

Debili, N, C Robin, V Schiavon, R Letestu, F Pflumio, M T Mitjavila-Garcia, L Coulombel, and W Vainchenker. 2001. “Different Expression of CD41 on Human Lymphoid and Myeloid Progenitors from Adults and Neonates.” *Blood* 97 (7):2023–30.

Deloukas, P, S Kanoni, C Willenborg, M Farrall, T L Assimes, J R Thompson, E Ingelsson, et al. 2013. “Large-Scale Association Analysis Identifies New Risk Loci for Coronary Artery Disease.” *Nature Genetics* 45 (1):25–33.

Denis, M M, N D Tolley, M Bunting, H Schwertz, H Jiang, S Lindemann, C C Yost, et al. 2005. “Escaping the Nuclear Confines: Signal-Dependent Pre-mRNA Splicing in Anucleate Platelets.” *Cell* 122 (3):379–91.

Dickinson, C J, and J F Martin. 1987. “Megakaryocytes and Platelet Clumps as the Cause of Finger Clubbing.” *Lancet* 330 (8573):1434–35.

Ding, J, J E Swain, and G D Smith. 2011. “Aurora Kinase-A Regulates Microtubule Organizing Center (MTOC) Localization, Chromosome Dynamics, and Histone-H3 Phosphorylation in Mouse Oocytes.” *Mol Reprod Dev* 78 (2):80–90.

Ding, L, and S J Morrison. 2013. “Haematopoietic Stem Cells and Early Lymphoid Progenitors Occupy Distinct Bone Marrow Niches.” *Nature* 495 (7440):231–35.

Ding, L, T L Saunders, G Enikolopov, and S J Morrison. 2012. “Endothelial and Perivascular Cells Maintain Haematopoietic Stem Cells.” *Nature* 481 (7382):457–62.

Djekic, D, R Nicoll, M Novo, and M Henein. 2015. “Metabolomics in Atherosclerosis.” *IJC Metabolic & Endocrine* 8:26–30.

Dobin, A, C A Davis, F Schlesinger, J Drenkow, C Zaleski, S Jha, P Batut, M Chaisson, and T R Gingeras. 2013. “STAR: Ultrafast Universal RNA-Seq Aligner.” *Bioinformatics* 29 (1):15–21.

Dorsch, M, P D Fan, N N Danial, P B Rothman, and S P Goff. 1997. “The Thrombopoietin Receptor Can Mediate Proliferation without Activation of the Jak-STAT Pathway.” *The Journal of Experimental Medicine* 186 (12):1947–55.

Du, W, S Amarachintha, A F Wilson, and Q Pang. 2016. “Hyper-Active Non-Homologous End Joining Selects for Synthetic Lethality Resistant and Pathological Fanconi Anemia Hematopoietic Stem and Progenitor Cells.” *Scientific Reports* 6 (1):22167.

Dykstra, B, D Kent, M Bowie, L McCaffrey, M Hamilton, K Lyons, S Lee, R Brinkman, and C Eaves. 2007. “Long-Term Propagation of Distinct Hematopoietic Differentiation Programs In Vivo.” *Cell Stem Cell* 1 (2):218–29.

Edelstein, L C, L M Simon, R T Montoya, M Holinstat, E S Chen, A Bergeron, X Kong, et al. 2013. “Racial Differences in Human Platelet PAR4 Reactivity Reflect Expression of PCTP and miR-376c.” *Nature Medicine* 19 (12):1609–16.

Eicher, J D, Y Wakabayashi, O Vitseva, N Esa, Y Yang, J Zhu, J E Freedman, D D McManus, and A D Johnson. 2016. “Characterization of the Platelet Transcriptome by RNA Sequencing in Patients with Acute Myocardial Infarction.” *Platelets* 27 (3):230–39.

Engel, C, M Loeffler, H Franke, and S Schmitz. 1999. “Endogenous Thrombopoietin Serum Levels during Multicycle Chemotherapy.” *British Journal of Haematology* 105 (3):832–38.

Erdmann, J, A Großhennig, P S Braund, I R König, C Hengstenberg, A S Hall, P Linsel-Nitschke, et al. 2009. “New Susceptibility Locus for Coronary Artery Disease on Chromosome 3q22.3.” *Nature Genetics* 41 (3):280–82.

Ergelen, M, and H Uyarel. 2014. “Plateletcrit: A Novel Prognostic Marker for Acute Coronary Syndrome.” *International Journal of Cardiology* 177 (1). Elsevier:161.

Espín-Palazón, R, D L Stachura, C A Campbell, D García-Moreno, N Del Cid, A D Kim, S Candel, J Meseguer, V Mulero, and D Traver. 2014. “Proinflammatory Signaling Regulates Hematopoietic Stem Cell Emergence.” *Cell* 159 (5):1070–85.

Estévez-Loureiro, R, J Salgado-Fernández, R Marzoa-Rivas, E Barge-Caballero, A Pérez-Pérez, V Noriega-Concepción, R Calviño-Santos, J M Vázquez-Rodríguez, N Vázquez-González, and A Castro-Beiras. 2009. “Mean Platelet Volume Predicts Patency of the Infarct-Related Artery before Mechanical Reperfusion and Short-Term Mortality in Patients with ST-Segment Elevation Myocardial Infarction Undergoing Primary

Percutaneous Coronary Intervention." *Thromb Res* 124 (5):536–40.

Etzrodt, M, M Endele, and T Schroeder. 2014. "Quantitative Single-Cell Approaches to Stem Cell Research." *Cell Stem Cell* 15 (5):546–58.

Evans, D M, I H Frazer, and N G Martin. 1999. "Genetic and Environmental Causes of Variation in Basal Levels of Blood Cells." *Twin Research : The Official Journal of the International Society for Twin Studies* 2 (4):250–57.

Fabre, J E, M Nguyen, K Athirakul, K Coggins, J D McNeish, S Austin, L K Parise, G A FitzGerald, T M Coffman, and B H Koller. 2001. "Activation of the Murine EP3 Receptor for PGE2 Inhibits cAMP Production and Promotes Platelet Aggregation." *The Journal of Clinical Investigation* 107 (5):603–10.

Falk, E, M Nakano, J F Bentzon, A V Finn, and R Virmani. 2013. "Update on Acute Coronary Syndromes: The Pathologists' View." *European Heart Journal* 34 (10):719–28.

Feng, Q, N Shabrani, J N Thon, H Huo, A Thiel, K R Machlus, K Kim, et al. 2014. "Scalable Generation of Universal Platelets from Human Induced Pluripotent Stem Cells." *Stem Cell Reports* 3 (5):817–31.

Feng, W, M Madajka, B A Kerr, G H Mahabeleshwar, S W Whiteheart, and T V Byzova. 2011. "A Novel Role for Platelet Secretion in Angiogenesis: Mediating Bone Marrow-Derived Cell Mobilization and Homing." *Blood* 117 (14):3893–3902.

Ferkowicz, M. J. 2003. "CD41 Expression Defines the Onset of Primitive and Definitive Hematopoiesis in the Murine Embryo." *Development* 130 (18):4393–4403.

Fink, L, H Holschermann, G Kwapiszewska, J P Muyal, B Lengemann, R M Bohle, and S Santoso. 2003. "Characterization of Platelet-Specific mRNA by Real-Time PCR after Laser-Assisted Microdissection." *Thromb Haemost* 90 (4):749–56.

Fitzgerald, J R, T J Foster, and D Cox. 2006. "The Interaction of Bacterial Pathogens with Platelets." *Nature Reviews Microbiology* 4 (6). Nature Publishing Group:445–57.

Florell, S R, C M Coffin, J A Holden, J W Zimmermann, J W Gerwels, B K Summers, D A Jones, and S A Leachman. 2001. "Preservation of RNA for Functional Genomic Studies: A Multidisciplinary Tumor Bank Protocol." *Modern Pathology* 14 (2):116–28.

Forsberg, E C, E Passegué, S S Prohaska, A J Wagers, M Koeva, J M Stuart, and I L Weissman. 2010. "Molecular Signatures of Quiescent, Mobilized and Leukemia-Initiating Hematopoietic Stem Cells." Edited by Catherine M. Verfaillie. *PLoS ONE* 5 (1):e8785.

Forsberg, E C, T Serwold, S Kogan, I L Weissman, and E Passegué. 2006. "New Evidence Supporting Megakaryocyte-Erythrocyte Potential of flk2/flt3+ Multipotent Hematopoietic Progenitors." *Cell* 126 (2):415–26.

Fox, S B, C A Day, and K C Gatter. 1991. "Association between Platelet Microthrombi and Finger Clubbing." *Lancet* 3;338(8762):313-4.

Fritzsche, M, and C Mandenius. 2010. "Fluorescent Cell-Based Sensing Approaches for Toxicity Testing." *Analytical and Bioanalytical Chemistry* 398 (1):181–91.

Fuentes, R, Y Wang, J Hirsch, C Wang, L Rauova, G S Worthen, M A Kowalska, and M Poncz. 2010. "Infusion of Mature Megakaryocytes into Mice Yields Functional Platelets."

The Journal of Clinical Investigation 120 (11):3917–22.

Funck-Jensen, K L, J Dalsgaard, E L Grove, A M Hvas, and S D Kristensen. 2013. “Increased Platelet Aggregation and Turnover in the Acute Phase of ST-Elevation Myocardial Infarction.” *Platelets* 24 (7):528–37.

Gagnon, A W, D R Manning, L Catani, A Gewirtz, M Poncz, and L F Brass. 1991. “Identification of Gz Alpha as a Pertussis Toxin-Insensitive G Protein in Human Platelets and Megakaryocytes.” *Blood* 78 (5):1247–53.

Gainsford, T, A W Roberts, S Kimura, D Metcalf, G Dranoff, R C Mulligan, C G Begley, L Robb, and W S Alexander. 1998. “Cytokine Production and Function in c-Mpl-Deficient Mice: No Physiologic Role for Interleukin-3 in Residual Megakaryocyte and Platelet Production.” *Blood* 91 (8):2745–52.

Garner, C, T Tatu, J E Reittie, T Littlewood, J Darley, S Cervino, M Farrall, P Kelly, T D Spector, and S L Thein. 2000. “Genetic Influences on F Cells and Other Hematologic Variables: A Twin Heritability Study.” *Blood* 95 (1):342–46.

Garner, S F, A Furnell, B C Kahan, C I Jones, A Attwood, P Harrison, A M Kelly, A H Goodall, R Cardigan, and W H Ouwehand. 2017. “Platelet Responses to Agonists in a Cohort of Highly Characterised Platelet Donors Are Consistent over Time.” *Vox Sanguinis* 112 (1):18–24.

Gear, A RL, and D Camerini. 2010. “Platelet Chemokines and Chemokine Receptors: Linking Hemostasis, Inflammation, and Host Defense.” *Microcirculation* 10 (3–4):335–50.

Gekas, C, and T Graf. 2013. “CD41 Expression Marks Myeloid Biased Adult Hematopoietic Stem Cells and Increases with Age.” *Blood* 121 (22):4463–72.

Gekas, C, K E Rhodes, L M Gereige, H Helgadottir, R Ferrari, S K Kurdistani, E Montecino-Rodriguez, et al. 2009. “Mef2C Is a Lineage-Restricted Target of Scl/Tall and Regulates Megakaryopoiesis and B-Cell Homeostasis.” *Blood* 113 (15):3461–71.

Genecards.org. n.d. “HIPED (the Human Integrated Protein Expression Database).”

Genever, P G, D J Wilkinson, A J Patton, N M Peet, Y Hong, A Mathur, J D Erusalimsky, and T M Skerry. 1999. “Expression of a Functional N-Methyl-D-Aspartate-Type Glutamate Receptor by Bone Marrow Megakaryocytes.” *Blood* 93 (9):2876–83.

Gerrits, A, B Dykstra, O J Kalmykowa, K Klauke, E Verovskaya, M J C Broekhuis, G De Haan, and L V Bystrykh. 2010. “Cellular Barcoding Tool for Clonal Analysis in the Hematopoietic System.” *Blood* 115 (13):2610–18.

Gevaert, K, M Goethals, L Martens, J Van Damme, A Staes, G R Thomas, and J Vandekerckhove. 2003. “Exploring Proteomes and Analyzing Protein Processing by Mass Spectrometric Identification of Sorted N-Terminal Peptides.” *Nature Biotechnology* 21 (5):566–69.

Giammona, L M, P G Fuhrken, E T Papoutsakis, and W M Miller. 2006. “Nicotinamide (Vitamin B3) Increases the Polyploidisation and Proplatelet Formation of Cultured Primary Human Megakaryocytes.” *British Journal of Haematology* 135 (4):554–66.

Giammona, L M, S Panuganti, J M Kemper, P A Apostolidis, S Lindsey, E T Papoutsakis, and

W M Miller. 2009. “Mechanistic Studies on the Effects of Nicotinamide on Megakaryocytic Polyploidization and the Roles of NAD+ Levels and SIRT Inhibition.” *Experimental Hematology* 37 (11):1340–1352.e3.

Gibbins, Jonathan M. 2004. “Platelet Adhesion Signalling and the Regulation of Thrombus Formation.” *Journal of Cell Science* 117 (Pt 16):3415–25.

Gieger, C, A Radhakrishnan, A Cvejic, W Tang, E Porcu, G Pistis, J Serbanovic-Canic, et al. 2011. “New Gene Functions in Megakaryopoiesis and Platelet Formation.” *Nature* 480 (7376):201–8.

Giles, H, R E Smith, and J F Martin. 1994. “Platelet Glycoprotein IIb-IIIa and Size Are Increased in Acute Myocardial Infarction.” *Eur J Clin Invest* 24 (1):69–72.

Gilliland, D G, W Song, M G Sullivan, R D Legare, S Hutchings, X Tan, D Kufrin, et al. 1999. “Haploinsufficiency of CBFA2 Causes Familial Thrombocytopenia with Propensity to Develop Acute Myelogenous Leukaemia.” *Nature Genetics* 23 (2):166–75.

Gilio, K, I C A Munnix, P Mangin, J M E M Cosemans, M A H Feijge, P E J van der Meijden, S Olieslagers, et al. 2009. “Non-Redundant Roles of Phosphoinositide 3-Kinase Isoforms Alpha and Beta in Glycoprotein VI-Induced Platelet Signaling and Thrombus Formation.” *The Journal of Biological Chemistry* 284 (49):33750–62.

Gnatenko, D V, J J Dunn, S R McCorkle, D Weissmann, P L Perrotta, and W F Bahou. 2003. “Transcript Profiling of Human Platelets Using Microarray and Serial Analysis of Gene Expression.” *Blood* 101 (6):2285–93. https://doi.org/10.1182/blood-2002-09-2797.

Goncalves, S C, M Labinaz, M Le May, C Glover, M Froeschl, J F Marquis, E O’Brien, et al. 2011. “Usefulness of Mean Platelet Volume as a Biomarker for Long-Term Outcomes after Percutaneous Coronary Intervention.” *The American Journal of Cardiology* 107 (2):204–9.

Goodall, A H, P Burns, I Salles, I C Macaulay, C I Jones, D Ardissino, B De Bono, et al. 2010. “Transcription Profiling in Human Platelets Reveals LRRFIP1 as a Novel Protein Regulating Platelet Function.” *Blood* 116 (22):4646–56.

Goschnick, M W, L Lau, J L Wee, Y S Liu, P M Hogarth, L M Robb, M J Hickey, M D Wright, and D E Jackson. 2006. “Impaired "outside-in" Integrin AIIbB3 Signaling and Thrombus Stability in TSSC6-Deficient Mice.” *Blood* 108 (6):1911–18.

Graaf, C. A. de, M. Kauppi, T. Baldwin, C. D. Hyland, D. Metcalf, T. A. Willson, M. R. Carpinelli, G. K. Smyth, W. S. Alexander, and D. J. Hilton. 2010. “Regulation of Hematopoietic Stem Cells by Their Mature Progeny.” *Proceedings of the National Academy of Sciences* 107 (50):21689–94.

Greenbaum, A, Y M S Hsu, R B Day, L G Schuettpelz, M J Christopher, J N Borgerding, T Nagasawa, and D C Link. 2013. “CXCL12 in Early Mesenchymal Progenitors Is Required for Haematopoietic Stem-Cell Maintenance.” *Nature* 495 (7440):227–30.

Greenberg, S M, D S Rosenthal, T A Greeley, R Tantravahi, and R I Handin. 1988. “Characterization of a New Megakaryocytic Cell Line: The Dami Cell.” *Blood* 72 (6):1968–77.

Griesshammer, M, M Bangerter, T Sauer, R Wennauer, L Bergmann, and H Heimpel. 1999. “Aetiology and Clinical Significance of Thrombocytosis: Analysis of 732 Patients with an Elevated Platelet Count.” *Journal of Internal Medicine* 245 (3):295–300.

Grinenko, T, K Arndt, M Portz, N Mende, M Günther, K N Cosgun, D Alexopoulou, et al. 2014. “Clonal Expansion Capacity Defines Two Consecutive Developmental Stages of Long-Term Hematopoietic Stem Cells.” *The Journal of Experimental Medicine* 211 (2):209–15.

Grove, E L, A M Hvas, and S D Kristensen. 2008. “Immature Platelets in Patients with Acute Coronary Syndromes.” *Thrombosis and Haemostasis*, 101: 151–56.

Grover, A, A Sanjuan-Pla, S Thongjuea, J Carrelha, A Giustacchini, A Gambardella, I C Macaulay, et al. 2016. “Single-Cell RNA Sequencing Reveals Molecular and Functional Platelet Bias of Aged Haematopoietic Stem Cells.” *Nature Communications* 7: 11075.

Growney, J D, H Shigematsu, Z Li, B H Lee, J Adelsperger, R Rowan, D P Curley, et al. 2005. “Loss of Runx1 Perturbs Adult Hematopoiesis and Is Associated with a Myeloproliferative Phenotype.” *Blood* 106 (2):494–504.

Grozovsky, R, A J Begonja, K Liu, G Visner, J H Hartwig, H Falet, and K M Hoffmeister. 2015. “The Ashwell-Morell Receptor Regulates Hepatic Thrombopoietin Production via JAK2-STAT3 Signaling.” *Nature Medicine* 21 (1):47–54.

Guo, G, S Luc, E Marco, T W Lin, C Peng, M A Kerenyi, S Beyaz, et al. 2013. “Mapping Cellular Hierarchy by Single-Cell Analysis of the Cell Surface Repertoire.” *Cell Stem Cell* 13 (4):492–505.

Gurney, A L, K Carver-Moore, F J de Sauvage, and M W Moore. 1994. “Thrombocytopenia in c-Mpl-Deficient Mice.” *Science* 265 (5177):1445–47.

Guthikonda, S, C L. Alviar, M Vaduganathan, M Arikan, A Tellez, T DeLao, J F Granada, J F Dong, N S Kleiman, and E I Lev. 2008. “Role of Reticulated Platelets and Platelet Size Heterogeneity on Platelet Activity After Dual Antiplatelet Therapy With Aspirin and Clopidogrel in Patients With Stable Coronary Artery Disease.” *Journal of the American College of Cardiology* 52 (9):743–49.

Haas, S, J Hansson, D Klimmeck, D Loeffler, L Velten, H Uckelmann, S Wurzer, et al. 2015. “Inflammation-Induced Emergency Megakaryopoiesis Driven by Hematopoietic Stem Cell-like Megakaryocyte Progenitors.” *Cell Stem Cell* 17 (4):422–34.

Hamada, T, R Mohle, J Hesselgesser, J Hoxie, R L Nachman, M A Moore, and S Rafii. 1998. “Transendothelial Migration of Megakaryocytes in Response to Stromal Cell-Derived Factor 1 (SDF-1) Enhances Platelet Formation.” *J Exp Med* 188 (3):539–48.

Hamm, C W, H Möllmann, J P Bassand, and F Van de Werf. 2009. “The ESC Textbook of Cardiovascular Medicine: Second Edition.” *Oxford University Press*.

Han, J, D Choi, S Choi, B Kim, Y Ki, J Chung, Y Koh, K Chang, and S Hong. 2013. “Stroke or Coronary Artery Disease Prediction from Mean Platelet Volume in Patients with Type 2 Diabetes Mellitus.” *Platelets* 24 (5):401–6.

Hancock, V, J F Martin, and R Lelchuk. 1993. “The Relationship between Human Megakaryocyte Nuclear DNA Content and Gene Expression.” *British Journal of*

Haematology 85 (4):692–97.

Harker, L A, and C A Finch. 1969. “Thrombokinetics in Man.” *Journal of Clinical Investigation* 48 (6):963–74.

Harris, R A, and D G Penington. 1984. “The Effects of Low-Dose Vincristine on Megakaryocyte Colony-Forming Cells and Megakaryocyte Ploidy.” *Br J Haematol* 57 (1):37–48.

Hartley, S W, and J C Mullikin. 2015. “QoRTs: A Comprehensive Toolset for Quality Control and Data Processing of RNA-Seq Experiments.” *BMC Bioinformatics* 16 (1):224.

Haver, V M, and A R Gear. 1981. “Functional Fractionation of Platelets.” *J Lab Clin Med* 97 (2):187–204.

Hayashi, C, A G Madrigal, X Liu, T Ukai, S Goswami, C V Gudino, F C Gibson, and C A Genco. 2010. “Pathogen-Mediated Inflammatory Atherosclerosis Is Mediated in Part via Toll-Like Receptor 2-Induced Inflammatory Responses.” *Journal of Innate Immunity* 2 (4):334–43.

Healy, A M, M D Pickard, A D Pradhan, Y Wang, Z Chen, K Croce, M Sakuma, et al. 2006. “Platelet Expression Profiling and Clinical Validation of Myeloid-Related Protein-14 as a Novel Determinant of Cardiovascular Events.” *Circulation* 113 (19):2278–84.

Heazlewood, S Y, R J Neaves, B Williams, D N Haylock, T E Adams, and S K Nilsson. 2013. “Megakaryocytes Co-Localise with Hemopoietic Stem Cells and Release Cytokines That up-Regulate Stem Cell Proliferation.” *Stem Cell Research* 11 (2):782–92.

Heijnen, H F, N Debili, W Vainchencker, J Breton-Gorius, H J Geuze, and J J Sixma. 1998. “Multivesicular Bodies Are an Intermediate Stage in the Formation of Platelet Alpha-Granules.” *Blood* 91 (7):2313–25.

Hillmann, A G, S Harmon, S D Park, J O’brien, D C Shields, and D Kenny. 2006. “Comparative RNA Expression Analyses from Small-Scale, Single-Donor Platelet Samples.” *J Thromb Haemost* 4 (2):349–56.

Hicks, S C, M Teng, and R A Irizarry. 2015. “On the Widespread and Critical Impact of Systematic Bias and Batch Effects in Single-Cell RNA-Seq Data.” *bioRxiv*, August. Cold Spring Harbor Laboratory, 25528.

Hindorff, L A, P Sethupathy, H A Junkins, E M Ramos, J P Mehta, F S Collins, and T A Manolio. 2009. “Potential Etiologic and Functional Implications of Genome-Wide Association Loci for Human Diseases and Traits.” *Proceedings of the National Academy of Sciences of the United States of America* 106 (23):9362–67.

Hirayama, Y, S Sakamaki, T Matsunaga, T Kuga, H Kuroda, T Kusakabe, K Sasaki, et al. 1998. “Concentrations of Thrombopoietin in Bone Marrow in Normal Subjects and in Patients with Idiopathic Thrombocytopenic Purpura, Aplastic Anemia, and Essential Thrombocythemia Correlate with Its mRNA Expression of Bone Marrow Stromal Cells.” *Blood* 92 (1):46–52.

Hisa, T, S E Spence, R A Rachel, M Fujita, T Nakamura, J M Ward, D E Devor-Henneman, et al. 2004. “Hematopoietic, Angiogenic and Eye Defects in Meis1 Mutant Animals.” *The*

EMBO Journal 23 (2):450–59.

Hock, H., E Meade, S Medeiros, J W Schindler, P J M Valk, Y Fujiwara, and S H Orkin. 2004. "Tel/Etv6 Is an Essential and Selective Regulator of Adult Hematopoietic Stem Cell Survival." *Genes & Development* 18 (19):2336–41.

Hodohara, K, N Fujii, N Yamamoto, and K Kaushansky. 2000. "Stromal Cell-Derived Factor-1 (SDF-1) Acts Together with Thrombopoietin to Enhance the Development of Megakaryocytic Progenitor Cells (CFU-MK)." *Blood* 95 (3):769–75.

Hollander, J E, M R Muttreja, M R Dalesandro, and F S Shofer. 1999. "Risk Stratification of Emergency Department Patients with Acute Coronary Syndromes Using P-Selectin." *Journal of the American College of Cardiology* 34 (1):95–105.

Howson, J M M, W Zhao, D R Barnes, W Ho, R Young, D S Paul, L L Waite, et al. 2017. "Fifteen New Risk Loci for Coronary Artery Disease Highlight Arterial-Wall-Specific Mechanisms." *Nature Genetics* 49 (7):1113–19.

Hsu, H C, W H Tsai, M L Jiang, C H Ho, M L Hsu, C K Ho, and S Y Wang. 1999. "Circulating Levels of Thrombopoietic and Inflammatory Cytokines in Patients with Clonal and Reactive Thrombocytosis." *The Journal of Laboratory and Clinical Medicine* 134 (4):392–97.

Hu, M, D Krause, M Greaves, S Sharkis, M Dexter, C Heyworth, and T Enver. 1997. "Multilineage Gene Expression Precedes Commitment in the Hemopoietic System." *Genes & Development* 11 (6):774–85.

Huang, S. 2009. "Non-Genetic Heterogeneity of Cells in Development: More than Just Noise." *Development* 136 (23):3853–62.

Huang, T, J Hsieh, Y Wu, C Jen, Y Tsuang, S Chiou, J Partanen, et al. 2008. "Functional Network Reconstruction Reveals Somatic Stemness Genetic Maps and Dedifferentiation-Like Transcriptome Reprogramming Induced by GATA2." *Stem Cells* 26 (5):1186–1201.

Huang, X, Q Ruan, Y Fang, F Traganos, Z Darzynkiewicz, and W Dai. 2004. "Physical and Functional Interactions between Mitotic Kinases during Polyploidization and Megakaryocytic Differentiation." *Cell Cycle* 3 (7). United States:946–51.

Huang, Y Q, J J Li, and S Karpatkin. 2000. "Identification of a Family of Alternatively Spliced mRNA Species of Angiopoietin-1." *Blood* 95 (6):1993–99.

Huczek, Z, J Kochman, K J Filipiak, G J Horszczaruk, M Grabowski, R Piatkowski, J Wilczynska, A Zielinski, B Meier, and G Opolski. 2005. "Mean Platelet Volume on Admission Predicts Impaired Reperfusion and Long-Term Mortality in Acute Myocardial Infarction Treated with Primary Percutaneous Coronary Intervention." *Journal of the American College of Cardiology* 46 (2):284–90.

IBC 50K CAD Consortium. 2011. "Large-Scale Gene-Centric Analysis Identifies Novel Variants for Coronary Artery Disease." *PLoS Genetics* 7 (9):e1002260.

Ichikawa, N, F Ishida, S Shimodaira, T Tahara, T Kato, and K Kitano. 1996. "Regulation of Serum Thrombopoietin Levels by Platelets and Megakaryocytes in Patients with Aplastic

Anaemia and Idiopathic Thrombocytopenic Purpura." *Thrombosis and Haemostasis* 76 (2):156–60.

Icli, A, F Aksoy, A Dogan, A Arslan, I Ersoy, H Yucel, and O Gorgulu. 2013. "Mean Platelet Volume May Be Elevated in Mitral Valve Prolapse and Associated With the Severity of Prolapse." *Clinical and Applied Thrombosis/Hemostasis* 19 (6):608–12.

Ihara, K, E Ishii, M Eguchi, H Takada, A Suminoe, R A Good, and T Hara. 1999. "Identification of Mutations in the c-Mpl Gene in Congenital Amegakaryocytic Thrombocytopenia." *Proceedings of the National Academy of Sciences of the United States of America* 96 (6):3132–36.

Ilicic, T, J K Kim, A A Kolodziejczyk, F O Bagger, D J McCarthy, J C Marioni, and S A Teichmann. 2016. "Classification of Low Quality Cells from Single-Cell RNA-Seq Data." *Genome Biology* 17 (1):29.

Isenberg, J S, M J Romeo, M Abu-Asab, M Tsokos, A Oldenborg, L Pappan, D A Wink, W A Frazier, and D D Roberts. 2007. "Increasing Survival of Ischemic Tissue by Targeting CD47." *Circulation Research* 100 (5):712–20.

Isenberg, J S, L A Ridnour, J Dimitry, W A Frazier, D A Wink, and D D Roberts. 2006. "CD47 Is Necessary for Inhibition of Nitric Oxide-Stimulated Vascular Cell Responses by Thrombospondin-1." *Journal of Biological Chemistry* 281 (36):26069–80.

Ishibashi, T, Z M Ruggeri, L A Harker, and S A Burstein. 1986. "Separation of Human Megakaryocytes by State of Differentiation on Continuous Gradients of Percoll: Size and Ploidy Analysis of Cells Identified by Monoclonal Antibody to Glycoprotein IIb/IIIa." *Blood* 67 (5):1286–92.

Italiano, J E, and R A Shivdasani. 2003. "Megakaryocytes and beyond: The Birth of Platelets." *Journal of Thrombosis and Haemostasis* 1 (6):1174–82.

Italiano, J E., P Lecine, R A Shivdasani, and J H Hartwig. 1999. "Blood Platelets Are Assembled Principally at the Ends of Proplatelet Processes Produced by Differentiated Megakaryocytes." *Journal of Cell Biology* 147 (6):1299–1312.

Itoh, T, K Nakai, M Ono, and K Hiramori. 1995. "Can the Risk for Acute Cardiac Events in Acute Coronary Syndrome Be Indicated by Platelet Membrane Activation Marker P-Selectin?" *Coronary Artery Disease* 6 (8):645–50.

Iwasaki, H, S I Mizuno, Y Arinobu, H Ozawa, Y Mori, H Shigematsu, K Takatsu, D G Tenen, and K Akashi. 2006. "The Order of Expression of Transcription Factors Directs Hierarchical Specification of Hematopoietic Lineages." *Genes and Development* 20 (21):3010–21.

Jiang, S, J D Levine, Y Fu, B Deng, R London, J E Groopman, and H Avraham. 1994. "Cytokine Production by Primary Bone Marrow Megakaryocytes." *Blood* 84 (12):4151–56.

Johnson, A D, L R Yanek, M H Chen, N Faraday, M G Larson, G Tofler, S J Lin, et al. 2010. "Genome-Wide Meta-Analyses Identifies Seven Loci Associated with Platelet Aggregation in Response to Agonists." *Nat Genet* 42 (7):608–13.

Johnson C A, I Schulman, C F Abildgaard. 1971. “Functional Studies of Young versus Old Platelets in a Patient with Chronic Thrombocytopaenia.” *Blood* 37:163–71.

Josefsson, E C, C James, K J Henley, M A Debrincat, K L Rogers, M R Dowling, M J White, et al. 2011. “Megakaryocytes Possess a Functional Intrinsic Apoptosis Pathway That Must Be Restrained to Survive and Produce Platelets.” *J Exp Med* 208 (10):2017–31.

Junt, T, H Schulze, Z Chen, S Massberg, T Goerge, A Krueger, D D Wagner, et al. 2007. “Dynamic Visualization of Thrombopoiesis within Bone Marrow.” *Science* 317 (5845):1767–70.

Kalisky, T, P Blainey, and S R Quake. 2011. “Genomic Analysis at the Single-Cell Level.” *Annual Review of Genetics* 45 (1):431–45.

Karpatkin, S. 1978. “Heterogeneity of Human Platelets. VI. Correlation of Platelet Function with Platelet Volume.” *Blood* 51 (2):307–16.

Karpatkin, S, and A Charmatz. 1969. “Heterogeneity of Human Platelets: I. Metabolic and Kinetic Evidence Suggestive of Young and Old Platelets.” *J Clin Invest* 48 (6):1073.

Kaser, A, G Brandacher, W Steurer, S Kaser, F A Offner, H Zoller, I Theurl, et al. 2001. “Interleukin-6 Stimulates Thrombopoiesis through Thrombopoietin: Role in Inflammatory Thrombocytosis.” *Blood* 98 (9):2720–25.

Kathiresan, S, B F Voight, S Purcell, K Musunuru, D Ardissino, P M Mannucci, S Anand, et al. 2009. “Genome-Wide Association of Early-Onset Myocardial Infarction with Single Nucleotide Polymorphisms and Copy Number Variants.” *Nature Genetics* 41 (3):334–41.

Kaushansky, K. 2005. “Review Series The Molecular Mechanisms That Control Thrombopoiesis.” *Journal of Clinical Investigation* 115 (12):3339–47.

Kawasaki, E S. 2004. “Microarrays and the Gene Expression Profile of a Single Cell.” *Annals of the New York Academy of Sciences* 1020 (1):92–100.

Kent, D G, M R Copley, C Benz, S Wöhrer, B J Dykstra, E Ma, J Cheyne, et al. 2009. “Prospective Isolation and Molecular Characterization of Hematopoietic Stem Cells with Durable Self-Renewal Potential.” *Blood* 113 (25):6342–50.

Kerrigan, S W, M Gaur, R P Murphy, S J Shattil, and A D Leavitt. 2004. “Caspase-12: A Developmental Link between G-Protein-Coupled Receptors and Integrin AIIbB3 Activation.” *Blood* 104 (5):1327–34.

Khandekar, M M, A S Khurana, S D Deshmukh, A L Kakrani, A D Katdare, and A K Inamdar. 2006. “Platelet Volume Indices in Patients with Coronary Artery Disease and Acute Myocardial Infarction: An Indian Scenario.” *Journal of Clinical Pathology* 59 (2):146–49.

Kiel, M J, Ö H Yilmaz, T Iwashita, O H Yilmaz, C Terhorst, and S J Morrison. 2005. “SLAM Family Receptors Distinguish Hematopoietic Stem and Progenitor Cells and Reveal Endothelial Niches for Stem Cells.” *Cell* 121 (7):1109–21.

Kilbey, A, H Alzuherri, J McColl, C Calés, J Frampton, and C Bartholomew. 2005. “The Evi1 Proto-Oncoprotein Blocks Endomitosis in Megakaryocytes by Inhibiting Sustained Cyclin-Dependent Kinase 2 Catalytic Activity.” *British Journal of Haematology* 130 (6):902–11.

Kim, J, Y Jung, J Seoh, S Woo, J Seo, and H Kim. 2002. “Gene Expression Profile of Megakaryocytes from Human Cord Blood CD34 $^{+}$ Cells Ex Vivo Expanded by Thrombopoietin.” *Stem Cells* 20 (5):402–16.

Kimura, S, A W Roberts, D Metcalf, and W S Alexander. 1998. “Hematopoietic Stem Cell Deficiencies in Mice Lacking c-Mpl, the Receptor for Thrombopoietin.” *Proceedings of the National Academy of Sciences of the United States of America* 95 (3):1195–1200.

Kimura, H, T Ishibashi, Y Shikama, A Okano, Y Akiyama, T Uchida, and Y Maruyama. 1990. “Interleukin-1 Beta (IL-1 Beta) Induces Thrombocytosis in Mice: Possible Implication of IL-6.” *Blood* 76 (12):2493–2500.

King, M A. 2000. “Detection of Dead Cells and Measurement of Cell Killing by Flow Cytometry.” *Journal of Immunological Methods* 243 (1–2):155–66.

King, S M, and G L Reed. 2002. “Development of Platelet Secretory Granules.” *Seminars in Cell & Developmental Biology* 13 (4):293–302.

Kiselev, V Y, K Kirschner, M T Schaub, T Andrews, A Yiu, T Chandra, K N Natarajan, et al. 2017. “SC3: Consensus Clustering of Single-Cell RNA-Seq Data.” *Nature Methods* 14 (5):483–86.

Kishk, Y T, E A Trowbridge, and J F Martin. 1985. “Platelet Volume Subpopulations in Acute Myocardial Infarction: An Investigation of Their Homogeneity for Smoking, Infarct Size and Site.” *Clin Sci* 68 (4):419–25.

Klovaite, J, M Benn, S Yazdanyar, and B G Nordestgaard. 2011. “High Platelet Volume and Increased Risk of Myocardial Infarction: 39 531 Participants from the General Population.” *Journal of Thrombosis and Haemostasis* 9 (1):49–56.

Kondo, M, I L Weissman, and K Akashi. 1997. “Identification of Clonogenic Common Lymphoid Progenitors in Mouse Bone Marrow.” *Cell* 91 (5):661–72.

Kosaki, G, and J Kambayashi. 2011. “Thrombocytogenesis by Megakaryocyte; Interpretation by Protoplatelet Hypothesis.” *Proceedings of the Japan Academy. Series B, Physical and Biological Sciences* 87 (5):254–73.

Kosugi, S, Y Kurata, Y Tomiyama, T Tahara, T Kato, S Tadokoro, M Shiraga, S Honda, Y Kanakura, and Y Matsuzawa. 1996. “Circulating Thrombopoietin Level in Chronic Immune Thrombocytopenic Purpura.” *British Journal of Haematology* 93 (3):704–6.

Kraemer, B F, R A Campbell, H Schwertz, M J Cody, Z Franks, N D Tolley, W H A Kahr, et al. 2011. “Novel Anti-Bacterial Activities of β-Defensin 1 in Human Platelets: Suppression of Pathogen Growth and Signaling of Neutrophil Extracellular Trap Formation.” Edited by Frank R. DeLeo. *PLoS Pathogens* 7 (11):e1002355.

Kramer, P A, S Ravi, B Chacko, M S Johnson, and V M Darley-Usmar. 2014. “A Review of the Mitochondrial and Glycolytic Metabolism in Human Platelets and Leukocytes: Implications for Their Use as Bioenergetic Biomarkers.” *Redox Biology* 2:206–10.

Kristensen, S D, P M Bath, and J F Martin. 1990. “Differences in Bleeding Time, Aspirin Sensitivity and Adrenaline between Acute Myocardial Infarction and Unstable Angina.” *Cardiovasc Res* 24 (1):19–23.

Kroczek, R A, V Henn, J R Slupsky, M Gräfe, I Anagnostopoulos, R Förster, and G Müller-Berghaus. 1998. “CD40 Ligand on Activated Platelets Triggers an Inflammatory Reaction of Endothelial Cells.” *Nature* 391 (6667):591–94.

Krüger, A, C Ellerström, C Lundmark, C Christersson, and T Wurtz. 2002. “RP59, a Marker for Osteoblast Recruitment, Is Also Detected in Primitive Mesenchymal Cells, Erythroid Cells, and Megakaryocytes.” *Developmental Dynamics* 223 (3):414–18.

Kuter, D J, D L Beeler, and R D Rosenberg. 1994. “The Purification of Megapoietin: A Physiological Regulator of Megakaryocyte Growth and Platelet Production.” *Proceedings of the National Academy of Sciences of the United States of America* 91 (23):11104–8.

Kuter, D J, and R D Rosenberg. 1990. “Regulation of Megakaryocyte Ploidy in Vivo in the Rat.” *Blood* 75 (1):74–81.

Kuter, D J, and R D Rosenberg. 1995. “The Reciprocal Relationship of Thrombopoietin (c-Mpl Ligand) to Changes in the Platelet Mass during Busulfan-Induced Thrombocytopenia in the Rabbit.” *Blood* 85 (10):2720–30.

Kuvardina, O N, S Herkt, A Meyer, L Schneider, J Yillah, N Kohrs, H Bonig, E Seifried, C Müller-Tidow, and J Lausen. 2017. “Hematopoietic Transcription Factors and Differential Cofactor Binding Regulate PRKACB Isoform Expression.” *Oncotarget*, 8 (42):71685-71698.

Lai, A Y, and M Kondo. 2006. “Asymmetrical Lymphoid and Myeloid Lineage Commitment in Multipotent Hematopoietic Progenitors.” *The Journal of Experimental Medicine* 203 (8):1867–73.

Lane, N, and W Martin. 2010. “The Energetics of Genome Complexity.” *Nature* 467 (7318):929–34.

Larson, M K, H Chen, M L Kahn, A M Taylor, J Fabre, R M Mortensen, P B Conley, and L V Parise. 2003. “Identification of P2Y 12 -Dependent and -Independent Mechanisms of Glycoprotein VI–mediated Rap1 Activation in Platelets.” *Blood* 101:1409-1415.

Lecine, P, J E Italiano, S W Kim, J L Villeval, and R A Shivdasani. 2000. “Hematopoietic-Specific Beta 1 Tubulin Participates in a Pathway of Platelet Biogenesis Dependent on the Transcription Factor NF-E2.” *Blood* 96 (4):1366–73.

Leek, J T, R B Scharpf, H C Bravo, D Simcha, B Langmead, W E Johnson, D Geman, K Baggerly, and R A Irizarry. 2010. “Tackling the Widespread and Critical Impact of Batch Effects in High-Throughput Data.” *Nature Reviews Genetics* 11 (10):733–39.

Lefrançais, E, G Ortiz-Muñoz, A Caudrillier, B Mallavia, F Liu, D M Sayah, E E Thornton, et al. 2017. “The Lung Is a Site of Platelet Biogenesis and a Reservoir for Haematopoietic Progenitors.” *Nature* 544 (7648):105–9.

Legrand, A J, S Choul-Li, C Spriet, T Idziorek, D Vicogne, H Drobecq, F Dantzer, V Villeret, and M Aumercier. 2013. “The Level of Ets-1 Protein Is Regulated by Poly(ADP-Ribose) Polymerase-1 (PARP-1) in Cancer Cells to Prevent DNA Damage.” Edited by Rajesh Mohanraj. *PLoS ONE* 8 (2).

Lemarchandel, V, J Ghysdael, V Mignotte, C Rahuel, and P H Roméo. 1993. “GATA and Ets Cis-Acting Sequences Mediate Megakaryocyte-Specific Expression.” *Molecular and Cellular Biology* 13 (1):668–76.

Levin, J, and J D Bessman. 1983. “The Inverse Relation between Platelet Volume and Platelet Number. Abnormalities in Hematologic Disease and Evidence That Platelet Size Does Not Correlate with Platelet Age.” *J Lab Clin Med* 101 (2):295–307.

Levine, G N, E R Bates, J A Bittl, R G Brindis, S D Fihn, L A Fleisher, C B Granger, et al. 2016. “2016 ACC/AHA Guideline Focused Update on Duration of Dual Antiplatelet Therapy in Patients With Coronary Artery Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines.” *Journal of the American College of Cardiology* 68 (10):1082–1115.

Li, J J, Y Q Huang, R Basch, and S Karpatkin. 2001. “Thrombin Induces the Release of Angiopoietin-1 from Platelets.” *Thrombosis and Haemostasis* 85 (2):204–6.

Li, X, J Sipple, Q Pang, and W Du. 2012. “Salidroside Stimulates DNA Repair Enzyme Parp-1 Activity in Mouse HSC Maintenance.” *Blood* 119 (18):4162–73.

Libby, P. 2013. “Mechanisms of Acute Coronary Syndromes and Their Implications for Therapy.” *N Engl J Med* 368 (21):2004–13.

Liddington, R C, and M H Ginsberg. 2002. “Integrin Activation Takes Shape: Figure 1.” *The Journal of Cell Biology* 158 (5):833–39.

Lippi, G, L Filippozzi, G L Salvagno, M Montagnana, M Franchini, G C Guidi, and G Targher. 2009. “Increased Mean Platelet Volume in Patients with Acute Coronary Syndromes.” *Arch Pathol Lab Med* 133 (9):1441–43.

Liu, Y, Y Wang, Y Gao, J A Forbes, R Qayyum, L Becker, L Cheng, and Z Z Wang. 2015. “Efficient Generation of Megakaryocytes From Human Induced Pluripotent Stem Cells Using Food and Drug Administration-Approved Pharmacological Reagents.” *Stem Cells Translational Medicine* 4 (4):309–19.

Liu, Z, V Vuohelainen, M Tarkka, J Tenhunen, R S Lappalainen, S Narkilahti, T Paavonen, N Oksala, Z Wu, and A Mennander. 2010. “Glutamate Release Predicts Ongoing Myocardial Ischemia of Rat Hearts.” *Scandinavian Journal of Clinical and Laboratory Investigation* 70 (3):217–24.

Lok, S, K Kaushansky, R D Holly, J L Kuijper, C E Lofton-Day, P J Oort, F J Grant, et al. 1994. “Cloning and Expression of Murine Thrombopoietin cDNA and Stimulation of Platelet Production in Vivo.” *Nature* 369 (6481):565–68.

López, J A. 1994. “The Platelet Glycoprotein Ib-IX Complex.” *Blood Coagulation & Fibrinolysis : An International Journal in Haemostasis and Thrombosis* 5 (1):97–119.

López-Cuenca, A A, A Tello-Montoliu, V Roldán, P Pérez-Berbel, M Valdés, and F Marín. 2012. “Prognostic Value of Mean Platelet Volume in Patients with Non-ST-Elevation Acute Coronary Syndrome.” *Angiology* 63 (4):241–44.

Lordier, L, D Bluteau, A Jalil, C Legrand, J Pan, P Rameau, D Jouni, et al. 2012. “RUNX1-Induced Silencing of Non-Muscle Myosin Heavy Chain IIB Contributes to Megakaryocyte

Polyploidization." *Nat Commun* 3:717.

Lordier, L, Y Chang, A Jalil, F Aurade, L Garcon, Y Lecluse, F Larbret, et al. 2010. "Aurora B Is Dispensable for Megakaryocyte Polyploidization, but Contributes to the Endomitotic Process." *Blood* 116 (13):2345–55.

Lordier, L, A Jalil, F Aurade, F Larbret, J Larghero, N Debili, W Vainchenker, and Y Chang. 2008. "Megakaryocyte Endomitosis Is a Failure of Late Cytokinesis Related to Defects in the Contractile Ring and Rho/Rock Signaling." *Blood* 112 (8):3164–74.

Love, M I, W Huber, and S Anders. 2014. "Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2." *Genome Biology* 15 (12):550.

Loven, J, D A Orlando, A A Sigova, C Y Lin, P B Rahl, C B Burge, D L Levens, T I Lee, and R A Young. 2012. "Revisiting Global Gene Expression Analysis." *Cell* 151 (3):476–82.

Lu, R, N F Neff, S R Quake, and I L Weissman. 2011. "Tracking Single Hematopoietic Stem Cells in Vivo Using High-Throughput Sequencing in Conjunction with Viral Genetic Barcoding." *Nat Biotechnol* 29 (10):928–33.

Luc, S, K Anderson, S Kharazi, N Buza-Vidas, C Boiers, C T Jensen, Z Ma, L Wittmann, and S E W Jacobsen. 2008. "Down-Regulation of Mpl Marks the Transition to Lymphoid-Primed Multipotent Progenitors with Gradual Loss of Granulocyte-Monocyte Potential." *Blood* 111 (7):3424–34.

Ludlow, L B, B P Schick, M L Budarf, D A Driscoll, E H Zackai, A Cohen, and B A Konkle. 1996. "Identification of a Mutation in a GATA Binding Site of the Platelet Glycoprotein Ibβ Promoter Resulting in the Bernard-Soulier Syndrome." *Journal of Biological Chemistry* 271 (36):22076–80.

Lun, A T L, D J McCarthy, and J C Marioni. 2016. "A Step-by-Step Workflow for Low-Level Analysis of Single-Cell RNA-Seq Data." *F1000Research* 5 (August):2122.

Ly, H Q, A J Kirtane, S A Murphy, J Buros, C P Cannon, E Braunwald, C M Gibson, and TIMI Study Group. 2006. "Association of Platelet Counts on Presentation and Clinical Outcomes in ST-Elevation Myocardial Infarction (from the TIMI Trials)." *The American Journal of Cardiology* 98 (1):1–5.

Macaulay, I C, M R Tijssen, D C Thijssen-Timmer, A Gusnanto, M Steward, P Burns, C F Langford, et al. 2007. "Comparative Gene Expression Profiling of in Vitro Differentiated Megakaryocytes and Erythroblasts Identifies Novel Activatory and Inhibitory Platelet Membrane Proteins." *Blood* 109 (8):3260–69.

Macaulay, I C, W Haerty, P Kumar, Y I Li, T X Hu, M J Teng, M Goolam, et al. 2015. "G&T-Seq: Parallel Sequencing of Single-Cell Genomes and Transcriptomes." *Nature Methods* 12 (6):519–22.

Macaulay, I C, V Svensson, C Labalette, L Ferreira, F Hamey, T Voet, S A Teichmann, and A Cvejic. 2016. "Single-Cell RNA-Sequencing Reveals a Continuous Spectrum of Differentiation in Hematopoietic Cells." *Cell Reports* 14 (4):966–77.

Machlus, K R, and J E Italiano. 2013. "The Incredible Journey: From Megakaryocyte Development to Platelet Formation." *Journal of Cell Biology* 201 (6):785–96.

Machlus, K R, K E Johnson, R Kulenthirarajan, J A Forward, M D Tippy, T S Soussou, S H El-Husayni, et al. 2016. “CCL5 Derived from Platelets Increases Megakaryocyte Proplatelet Formation.” *Blood* 127 (7):921–26.

Mahaney, M C, C Brugnara, L R Lease, and O S Platt. 2005. “Genetic Influences on Peripheral Blood Cell Counts: A Study in Baboons.” *Blood* 106 (4):1210–14.

Mallolas, J, O Hurtado, M Castellanos, M Blanco, T Sobrino, J Serena, J Vivancos, et al. 2006. “A Polymorphism in the EAAT2 Promoter Is Associated with Higher Glutamate Concentrations and Higher Frequency of Progressing Stroke.” *The Journal of Experimental Medicine* 203 (3):711–17.

Mangalpally, K K R, A Siqueiros-Garcia, M Vaduganathan, J F Dong, N S Kleiman, and S Guthikonda. 2010. “Platelet Activation Patterns in Platelet Size Sub-Populations: Differential Responses to Aspirin in Vitro.” *Journal of Thrombosis and Thrombolysis* 30 (3):251–62.

Månsson, R, A Hultquist, S Luc, L Yang, K Anderson, S Kharazi, S Al-Hashmi, et al. 2007. “Molecular Evidence for Hierarchical Transcriptional Lineage Priming in Fetal and Adult Stem Cells and Multipotent Progenitors.” *Immunity* 26 (4):407–19.

Martin, J F, P M Bath, and M L Burr. 1991. “Influence of Platelet Size on Outcome after Myocardial Infarction.” *Lancet* 338 (8780):1409–11.

Martin, J F, T D Daniel, and E A Trowbridge. 1987. “Acute and Chronic Changes in Platelet Volume and Count after Cardiopulmonary Bypass Induced Thrombocytopenia in Man.” *Thromb Haemost* 57 (1):55–58.

Martin, J F, and D G Penington. 1983. “The Relationship between the Age and Density of Circulating 51-Cr Labelled Platelets in the Sub-Human Primate.” *Thromb Res* 30 (2):157–64.

Martin, J F, J Plumb, R S Kilbey, and Y T Kishk. 1983. “Changes in Volume and Density of Platelets in Myocardial Infarction.” *BMJ* 287 (6390):456–59.

Martin, J F, T Shaw, J Heggie, and D G Penington. 1983. “Measurement of the Density of Human Platelets and Its Relationship to Volume.” *Br J Haematol* 54 (3):337–52.

Martin, J F, E A Trowbridge, G L Salmon, and D N Slater. 1982. “The Relationship between Platelet and Megakaryocyte Volumes.” *Thromb Res* 28 (4):447–59.

Martin, J F, E A Trowbridge, G Salmon, and J Plumb. 1983. “The Biological Significance of Platelet Volume: Its Relationship to Bleeding Time, Platelet Thromboxane B2 Production and Megakaryocyte Nuclear DNA Concentration.” *Thromb Res* 32 (5):443–60.

Martin, J F, S D Kristensen, A Mathur, E L Grove, and F A Choudry. 2012. “The Causal Role of Megakaryocyte-Platelet Hyperactivity in Acute Coronary Syndromes.” *Nature Reviews Cardiology* 9 (11):658–70.

Mathur, A, M S C Robinson, J Cotton, J F Martin, and J D Erusalimsky. 2001. “Platelet Reactivity in Acute Coronary Syndromes: Evidence for Differences in Platelet Behaviour between Unstable Angina and Myocardial Infarction.” *Thrombosis and Haemostasis* 85 (6):989–94.

Mattia, G, F Vulcano, L Milazzo, A Barca, G Macioce, A Giampaolo, and H J Hassan. 2002. “Different Ploidy Levels of Megakaryocytes Generated from Peripheral or Cord Blood CD34+ Cells Are Correlated with Different Levels of Platelet Release.” *Blood* 99 (3):888–97.

Mazur, E M, D L Lindquist, P A de Alarcon, and J L Cohen. 1988. “Evaluation of Bone Marrow Megakaryocyte Ploidy Distributions in Persons with Normal and Abnormal Platelet Counts.” *J Lab Clin Med* 111 (2):194–202.

McCarthy, D J, K R Campbell, A T L Lun, and Q F Wills. 2017. “Scater: Pre-Processing, Quality Control, Normalization and Visualization of Single-Cell RNA-Seq Data in R.” *Bioinformatics* 33 (8):btw777.

McRedmond, J P, S D Park, D F Reilly, J A Coppinger, P B Maguire, D C Shields, and D J Fitzgerald. 2004. “Integration of Proteomics and Genomics in Platelets: A Profile of Platelet Proteins and Platelet-Specific Genes.” *Mol Cell Proteomics* 3 (2):133–44.

McVean, G A, D M Altshuler, R M Durbin, G R Abecasis, D R Bentley, A Chakravarti, A G Clark, et al. 2012. “An Integrated Map of Genetic Variation from 1,092 Human Genomes.” *Nature* 491 (7422):56–65.

Méndez-Ferrer, S, T V Michurina, F Ferraro, A R Mazloom, B D MacArthur, S A Lira, D T Scadden, A Ma’ayan, G N Enikolopov, and P S Frenette. 2010. “Mesenchymal and Haematopoietic Stem Cells Form a Unique Bone Marrow Niche.” *Nature* 466 (7308):829–34.

Michelson, A D. 2012. “Platelets.” In *Platelets*. Elsevier

Michelson, A D. 2004. “Platelet Function Testing in Cardiovascular Diseases.” *Circulation* 110 (19):e489–93.

Milner, P C, and J F Martin. 1985. “Shortened Bleeding Time in Acute Myocardial Infarction and Its Relation to Platelet Mass.” *British Medical Journal (Clinical Research Ed.)* 290 (6484):1767.

Miyamoto, T, H Iwasaki, B Reizis, M Ye, T Graf, I L Weissman, and K Akashi. 2002. “Myeloid or Lymphoid Promiscuity as a Critical Step in Hematopoietic Lineage Commitment.” *Developmental Cell* 3 (1):137–47.

Miyazaki, R, H Ogata, T Iguchi, S Sogo, T Kushida, T Ito, M Inaba, S Ikehara, and Y Kobayashi. 2000. “Comparative Analyses of Megakaryocytes Derived from Cord Blood and Bone Marrow.” *British Journal of Haematology* 108 (3):602–9.

Möhle, R, F Bautz, S Rafii, M A Moore, W Brugger, and L Kanz. 1998. “The Chemokine Receptor CXCR-4 Is Expressed on CD34+ Hematopoietic Progenitors and Leukemic Cells and Mediates Transendothelial Migration Induced by Stromal Cell-Derived Factor-1.” *Blood* 91 (12):4523–30.

Moignard, V, I C Macaulay, G Swiers, F Buettner, J Schütte, F J Calero-Nieto, S Kinston, et al. 2013. “Characterization of Transcriptional Networks in Blood Stem and Progenitor Cells Using High-Throughput Single-Cell Gene Expression Analysis.” *Nature Cell Biology* 15 (4):363–72.

Monaco, C, A Mathur, and J F Martin. 2005. “What Causes Acute Coronary Syndromes? Applying Koch’s Postulates.” *Atherosclerosis* 179 (1):1–15.

Moreau, T, A L Evans, L Vasquez, M R Tijssen, Y Yan, M W Trotter, D Howard, et al. 2016. “Large-Scale Production of Megakaryocytes from Human Pluripotent Stem Cells by Chemically Defined Forward Programming.” *Nature Communications* 7 (April):11208.

Moro, M A, R J Russell, S Cellek, I Lizasoain, Y Su, V M Darley-Usmar, M W Radomski, and S Moncada. 1996. “cGMP Mediates the Vascular and Platelet Actions of Nitric Oxide: Confirmation Using an Inhibitor of the Soluble Guanylyl Cyclase.” *Pharmacology* 93:1480–85.

Morrell, C N, A A Aggrey, L M Chapman, and K L Modjeski. 2014. “Emerging Roles for Platelets as Immune and Inflammatory Cells.” *Blood* 123 (18):2759–67.

Morrell, C N, H Sun, M Ikeda, J C Beique, A M Swaim, E Mason, T V Martin, et al. 2008. “Glutamate Mediates Platelet Activation through the AMPA Receptor.” *The Journal of Experimental Medicine* 205 (3):575–84.

Morrison, S J, and D T Scadden. 2014. “The Bone Marrow Niche for Haematopoietic Stem Cells.” *Nature* 505 (7483):327–34.

Morrow, D A, Y Wang, K Croce, M Sakuma, M S Sabatine, H Gao, A D Pradhan, et al. 2008. “Myeloid-Related Protein 8/14 and the Risk of Cardiovascular Death or Myocardial Infarction after an Acute Coronary Syndrome in the Pravastatin or Atorvastatin Evaluation and Infection Therapy: Thrombolysis in Myocardial Infarction (PROVE IT-TIMI 22) Trial.” *Am Heart J* 155 (1):49–55.

Muller-Sieburg, C E, R H Cho, L Karlsson, J F Huang, and H B Sieburg. 2004. “Myeloid-Biased Hematopoietic Stem Cells Have Extensive Self-Renewal Capacity but Generate Diminished Lymphoid Progeny with Impaired IL-7 Responsiveness.” *Blood* 103 (11):4111–18.

Müller-Sieburg, C E, R H Cho, M Thoman, B Adkins, and H B Sieburg. 2002. “Deterministic Regulation of Hematopoietic Stem Cell Self-Renewal and Differentiation.” *Blood* 100 (4):1302–9.

Muller-Sieburg, C E, H B Sieburg, J M Bernitz, and G Cattarossi. 2012. “Stem Cell Heterogeneity: Implications for Aging and Regenerative Medicine.” *Blood* 119 (17):3900–3907.

Muntean, A G, L Pang, M Poncz, S F Dowdy, G A Blobel, and J D. Crispino. 2007. “Cyclin D-Cdk4 Is Regulated by GATA-1 and Required for Megakaryocyte Growth and Polyploidization.” *Blood* 109 (12):5199–5207.

Murphy, A J, V Sarrazy, N Wang, N Bijl, S Abramowicz, M Westerterp, C B Welch, J D Schuetz, and L Yvan-Charvet. 2014. “Deficiency of ATP-Binding Cassette Transporter b6 in Megakaryocyte Progenitors Accelerates Atherosclerosis in Mice.” *Arteriosclerosis, Thrombosis, and Vascular Biology* 34 (4):751–58.

Murphy, A J, N Bijl, L Yvan-Charvet, C B Welch, N Bhagwat, A Reheman, Y Wang, et al. 2013. “Cholesterol Efflux in Megakaryocyte Progenitors Suppresses Platelet Production

and Thrombocytosis." *Nature Medicine* 19 (5):586–94.

Muscari, A, S De Pascalis, A Cenni, C Ludovico, N Castaldini, S Antonelli, G Bianchi, D Magalotti, and M Zoli. 2008. "Determinants of Mean Platelet Volume (MPV) in an Elderly Population: Relevance of Body Fat, Blood Glucose and Ischaemic Electrocardiographic Changes." *Thromb Haemost* 99 (6):1079–84.

Nagata, Y, Y Muro, and K Todokoro. 1997. "Thrombopoietin-Induced Polyploidization of Bone Marrow Megakaryocytes Is due to a Unique Regulatory Mechanism in Late Mitosis." *Journal of Cell Biology* 139 (2):449–57.

Nagata, Y, J Yoshikawa, A Hashimoto, M Yamamoto, A H Payne, and K Todokoro. 2003. "Proplatelet Formation of Megakaryocytes Is Triggered by Autocrine-Synthesized Estradiol." *Genes and Development* 17 (23):2864–69.

Nagl, W. 1978. "Endopolyploidy and Polyteny in Differentiation and Evolution: Towards an Understanding of Quantitative and Qualitative Variation of Nuclear DNA in Ontogeny and Phylogeny." *Amsterdam Etc.: North Holland Publishing Company Xi, 283p. Illus.. General (KR, 198000144)*.

Nakamura, S, N Takayama, S Hirata, H Seo, H Endo, K Ochi, K Fujita, et al. 2014. "Expandable Megakaryocyte Cell Lines Enable Clinically Applicable Generation of Platelets from Human Induced Pluripotent Stem Cells." *Cell Stem Cell* 14 (4). Elsevier:535–48.

Ng, A P, M Kauppi, D Metcalf, C D Hyland, E C Josefsson, M Lebois, J G. Zhang, et al. 2014. "Mpl Expression on Megakaryocytes and Platelets Is Dispensable for Thrombopoiesis but Essential to Prevent Myeloproliferation." *Proceedings of the National Academy of Sciences* 111 (16):5884–89.

Nichol, J L, M M Hokom, A Hornkohl, W P Sheridan, H Ohashi, T Kato, Y S Li, T D Bartley, E Choi, and J Bogenberger. 1995. "Megakaryocyte Growth and Development Factor. Analyses of in Vitro Effects on Human Megakaryopoiesis and Endogenous Serum Levels during Chemotherapy-Induced Thrombocytopenia." *Journal of Clinical Investigation* 95 (6):2973–78.

Nikpay, M, A Goel, H H Won, L M Hall, C Willenborg, S Kanoni, D Saleheen, T Kyriakou, et al. the CARDIoGRAMplusC4D Consortium. 2015. "A Comprehensive 1000 Genomes-Based Genome-Wide Association Meta-Analysis of Coronary Artery Disease." *Nature Genetics* 47 (10):1121–30.

Nieswandt, B, and S P Watson. 2003. "Platelet-Collagen Interaction: Is GPVI the Central Receptor?" *Blood* 102 (2):449–61.

Niswander, L M, A D Koniski, A Seraichick, S C Catherman, K H Fegan, P D Kingsley, L M Calvi, and J Palis. 2015. "Prostaglandin E2 Promotes the Sequential Recovery of Bone Marrow Vasculature and the Megakaryocyte Lineage Following Radiation Injury." *Blood* 126 (23).

North, T E, W Goessling, C R Walkley, C Lengerke, K R Kopani, A M Lord, G J Weber, et al. 2007. "Prostaglandin E2 Regulates Vertebrate Haematopoietic Stem Cell Homeostasis."

Nature 447 (7147):1007–11.

Notta, F, S Doulatov, E Laurenti, A Poeppl, I Jurisica, and J E Dick. 2011. “Isolation of Single Human Hematopoietic Stem Cells Capable of Long-Term Multilineage Engraftment.” *Science* 333 (6039):218–21.

Notta, F, S Zandi, N Takayama, S Dobson, O I Gan, G Wilson, K B Kaufmann, et al. 2016. “Distinct Routes of Lineage Development Reshape the Human Blood Hierarchy across Ontogeny.” *Science* 351 (6269):aab2116.

Novershtern, N, A Subramanian, L N Lawton, R H Mak, W N Haining, M E McConkey, N Habib, et al. 2011. “Densely Interconnected Transcriptional Circuits Control Cell States in Human Hematopoiesis.” *Cell* 144 (2):296–309.

Nurden, A T. 2011. “Platelets, Inflammation and Tissue Regeneration.” *Thromb Haemost* 105 Suppl:S13-33.

Odell, T T, and C W Jackson. 1968. “Polyploidy and Maturation of Rat Megakaryocytes.” *Blood* 32 (1):102–10.

Okonechnikov, K, A Conesa, and F García-Alcalde. 2015. “Qualimap 2: Advanced Multi-Sample Quality Control for High-Throughput Sequencing Data.” *Bioinformatics* 32 (2):btv566.

Olson, T S, A Caselli, S Otsuru, T J Hofmann, R Williams, P Paolucci, M Dominici, and E M Horwitz. 2013. “Megakaryocytes Promote Murine Osteoblastic HSC Niche Expansion and Stem Cell Engraftment after Radioablative Conditioning.” *Blood* 121 (6):5238–49.

Orkin, S H. 1992. “GATA-Binding Transcription Factors in Hematopoietic Cells.” *Blood* 80 (3):575–81.

Orkin, S H, and L I Zon. 2008. “Hematopoiesis: An Evolving Paradigm for Stem Cell Biology.” *Cell* 132 (4):631–44.

Orlowski, E, R Chand, J Yip, C Wong, M W Goschnick, M D Wright, L K Ashman, and D E Jackson. 2009. “A Platelet Tetraspanin Superfamily Member, CD151, Is Required for Regulation of Thrombus Growth and Stability *in Vivo*.” *Journal of Thrombosis and Haemostasis* 7 (12):2074–84.

Osawa, M, K Hanada, H Hamada, and H Nakauchi. 1996. “Long-Term Lymphohematopoietic Reconstitution by a Single CD34-Low/negative Hematopoietic Stem Cell.” *Science* 273 (5272):242–45.

Osborne, G W. 2011. “Recent Advances in Flow Cytometric Cell Sorting.” In *Methods in Cell Biology*, 102:533–56.

Pan, Q, O Shai, L J Lee, B J Frey, and B J Blencowe. 2008. “Deep Surveying of Alternative Splicing Complexity in the Human Transcriptome by High-Throughput Sequencing.” *Nat Genet* 40 (12):1413–15.

Pang, L, H H Xue, G Szalai, X Wang, Y Wang, D K Watson, W J Leonard, G A Blobel, and M Poncz. 2006. “Maturation Stage-Specific Regulation of Megakaryopoiesis by Pointed-Domain Ets Proteins.” *Blood* 108 (7):2198–2206.

Pang, W W, S L Schrier, and I L Weissman. 2017. “Age-Associated Changes in Human

Hematopoietic Stem Cells." *Seminars in Hematology* 54 (1):39–42.

Pang, W W, E Price, D Sahoo, I Beerman, W J Maloney, D J Rossi, S L Schrier, and I L Weissman. 2011. "Human Bone Marrow Hematopoietic Stem Cells Are Increased in Frequency and Myeloid-Biased with Age." *Proceedings of the National Academy of Sciences* 108 (50):20012–17.

Parker, R I, M E Rick, and H R Gralnick. 1985. "Effect of Calcium on the Availability of Platelet von Willebrand Factor." *The Journal of Laboratory and Clinical Medicine* 106 (3):336–42.

Parker, R I, B C Shafer, and H R Gralnick. 1987. "Platelet Density-Dependent Partition of Platelet-von Willebrand Factor between Alpha Granule and Non-Alpha Granule Pools." *Thrombosis and Haemostasis* 58 (3):911–14.

Patel, S R, J L Richardson, H Schulze, E Kahle, N Galjart, K Drabek, R A Shivdasani, J H Hartwig, and J E Italiano. 2005. "Differential Roles of Microtubule Assembly and Sliding in Proplatelet Formation by Megakaryocytes." *Blood* 106 (13):4076–85.

Paulus, J M. 1975. "Platelet Size in Man." *Blood* 46 (3):321–36.

Paulus, J M, J Bury, and J C Grosdent. 1979. "Control of Platelet Territory Development in Megakaryocytes." *Blood Cells* 5 (1):59.

Pedersen, N T. 1978. "Occurrence of Megakaryocytes in Various Vessels and Their Retention in the Pulmonary Capillaries in Man." *Scand J Haematol* 21 (5):369–75.

Penington, D G, K Streatfield, and A E Roxburgh. 1976. "Megakaryocytes and the Heterogeneity of Circulating Platelets." *Br J Haematol* 34 (4):639–53.

Pereg, D, T Berlin, and M Mosseri. 2010. "Mean Platelet Volume on Admission Correlates with Impaired Response to Thrombolysis in Patients with ST-Elevation Myocardial Infarction." *Platelets* 21 (2):117–21.

Pignatelli, P, L De Biase, L Lenti, G Tocci, A Brunelli, R Cangemi, S Riondino, S Grego, M Volpe, and F Violi. 2005. "Tumor Necrosis Factor- as Trigger of Platelet Activation in Patients with Heart Failure." *Blood* 106 (6):1992–94.

Picelli, S, O R Faridani, Å K Björklund, G Winberg, S Sagasser, and R Sandberg. 2014. "Full-Length RNA-Seq from Single Cells Using Smart-seq2." *Nature Protocols* 9 (1):171–81.

Pitchford, S C, T Lodie, and S M Rankin. 2012. "VEGFR1 Stimulates a CXCR4-Dependent Translocation of Megakaryocytes to the Vascular Niche, Enhancing Platelet Production in Mice." *Blood* 120 (14):2787–95.

Podolak-Dawidziak, M, V Hancock, R Lelchuk, S Kotlarek-Haus, and J F Martin. 1995. "The Expression of mRNA for Fibrinogen in Megakaryocytes Isolated from Patients with T-Cell Lymphoma." *Br J Haematol* 91 (2):362–66.

Pronk, C J H, D J Rossi, R Månsson, J L Attema, G L Norddahl, C K F Chan, M Sigvardsson, I L Weissman, and D Bryder. 2007. "Elucidation of the Phenotypic, Functional, and Molecular Topography of a Myeloerythroid Progenitor Cell Hierarchy." *Cell Stem Cell* 1 (4):428–42.

Psaila, B, N Barkas, D Iskander, A Roy, S Anderson, N Ashley, V S Caputo, et al. 2016.

"Single-Cell Profiling of Human Megakaryocyte-Erythroid Progenitors Identifies Distinct Megakaryocyte and Erythroid Differentiation Pathways." *Genome Biology* 17 (1). Genome Biology:1–19.

Qian, H, N Buza-Vidas, C D Hyland, C T Jensen, J Antonchuk, R Månsson, L A Thoren, M Ekblom, W S Alexander, and S E W Jacobsen. 2007. "Critical Role of Thrombopoietin in Maintaining Adult Quiescent Hematopoietic Stem Cells." *Cell Stem Cell* 1 (6):671–84.

Qiu, X, Q Mao, Y Tang, L Wang, R Chawla, H Pliner, and C Trapnell. 2017. "Reversed Graph Embedding Resolves Complex Single-Cell Developmental Trajectories." *bioRxiv*, February. Cold Spring Harbor Laboratory, 110668.

Radley, J M, and G Scurfield. 1980. "The Mechanism of Platelet Release." *Blood* 56 (6). American Society of Hematology:996.

Ranjith, M P, R Divya, V K Mehta, M G Krishnan, R KamalRaj, and A Kavishwar. 2009. "Significance of Platelet Volume Indices and Platelet Count in Ischaemic Heart Disease." *Journal of Clinical Pathology* 62 (9):830–33.

Raslova, H, A Kauffmann, D Sekkaï, H Ripoche, F Larbret, T Robert, D Tronik Le Roux, et al. 2007. "Interrelation between Polyploidization and Megakaryocyte Differentiation: A Gene Profiling Approach." *Blood* 109 (8):3225–34.

Raslova, H, L Roy, C Vourc, J P Le Couedic, O Brison, D Metivier, et al. 2003. "Megakaryocyte Polyploidization Is Associated with a Functional Gene Amplification." *Gene Expression* 101 (2):541–44.

Ratnoff, O D. 1987. "The Evolution of Hemostatic Mechanisms." *Perspectives in Biology and Medicine* 31 (1):4–33.

Reddi, A H, R Gay, S Gay, and E J Miller. 1977. "Transitions in Collagen Types during Matrix-Induced Cartilage, Bone, and Bone Marrow Formation." *Proceedings of the National Academy of Sciences of the United States of America* 74 (12):5589–92.

Reya, T, S J Morrison, M F Clarke, and I L Weissman. 2001. "Stem Cells, Cancer, and Cancer Stem Cells." *Nature* 414 (6859):105–11.

Rhodes, K E, C Gekas, Y Wang, C T Lux, C S Francis, D N Chan, S Conway, S H Orkin, M C Yoder, and H K A Mikkola. 2008. "The Emergence of Hematopoietic Stem Cells Is Initiated in the Placental Vasculature in the Absence of Circulation." *Cell Stem Cell* 2 (3):252–63.

Ridker, P M, B M Everett, T Thuren, J G MacFadyen, W H Chang, C Ballantyne, F Fonseca, et al. 2017. "Antiinflammatory Therapy with Canakinumab for Atherosclerotic Disease." *New England Journal of Medicine* 377 (12):1119–31.

Ritchie, M E, B Phipson, D Wu, Y Hu, C W Law, W Shi, and G K. Smyth. 2015. "Limma Powers Differential Expression Analyses for RNA-Sequencing and Microarray Studies." *Nucleic Acids Research* 43 (7):e47–e47.

Robin, C, K Ottersbach, and J C Boisset. 2011. "CD41 Is Developmentally Regulated and Differentially Expressed on Mouse Hematopoietic Stem Cells." *Blood* 117 (19):5088–92.

Rodrigues, N P, V Janzen, R Forkert, D M Dombkowski, A S Boyd, S H Orkin, T Enver, P

Vyas, and D T Scadden. 2005. “Haploinsufficiency of GATA-2 Perturbs Adult Hematopoietic Stem-Cell Homeostasis.” *Blood* 106 (2):477–84.

Rogers, N M, M Sharifi-Sanjani, G Csányi, P J Pagano, and J S Isenberg. 2014. “Thrombospondin-1 and CD47 Regulation of Cardiac, Pulmonary and Vascular Responses in Health and Disease.” *Matrix Biology* 37 (July):92–101.

Romeo, P, M Prandini, V Joulin, V Mignotte, M Prenant, W Vainchenker, G Marguerie, and G Uzan. 1990. “Megakaryocytic and Erythrocytic Lineages Share Specific Transcription Factors.” *Nature* 344 (6265):447–49.

Rouleau, M, A Patel, M J Hendzel, S H Kaufmann, and G G Poirier. 2010. “PARP Inhibition: PARP1 and beyond.” *Nature Reviews Cancer* 10 (4):293–301.

Rousseeuw, P J. 1987. “Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis.” *Journal of Computational and Applied Mathematics* 20:53–65.

Rowley, J W, A J Oler, N D Tolley, B N Hunter, E N Low, D A Nix, C C Yost, G A Zimmerman, and A S Weyrich. 2011. “Genome-Wide RNA-Seq Analysis of Human and Mouse Platelet Transcriptomes.” *Blood* 118 (14).

Ru, Y, S Dong, H Liang, and S Zhao. 2016. “Platelet Production of Megakaryocyte: A Review with Original Observations on Human *in Vivo* Cells and Bone Marrow.” *Ultrastructural Pathology* 40 (4):163–70. https://doi.org/10.3109/01913123.2016.1170744.

Ru, Y, S Zhao, S Dong, Y Yang, and B Eyden. 2015. “On the Maturation of Megakaryocytes: A Review with Original Observations on Human In Vivo Cells Emphasizing Morphology and Ultrastructure.” *Ultrastructural Pathology* 39 (2):79–87.

Ruf, A, and H Patscheke. 1995. “Flow Cytometric Detection of Activated Platelets: Comparison of Determining Shape Change, Fibrinogen Binding, and P-Selectin Expression.” *Seminars in Thrombosis and Hemostasis* 21 (2):146–51.

Sabri, S, M Jandrot-Perrus, J Bertoglio, R W Farndale, V Mansat-De Mas, N Debili, and W Vainchenker. 2004. “Differential Regulation of Actin Stress Fiber Assembly and Proplatelet Formation by α2β1 Integrin and GPVI in Human Megakaryocytes.” *Blood* 104 (10):3117–25.

Saliba, A, A J Westermann, S A Gorski, and J Vogel. 2014. “Single-Cell RNA-Seq: Advances and Future Challenges.” *Nucleic Acids Research* 42 (14):8845–60.

Samani, N J, J Erdmann, A S Hall, C Hengstenberg, M Mangino, B Mayer, R J Dixon, et al. 2007. “Genomewide Association Analysis of Coronary Artery Disease.” *New England Journal of Medicine* 357 (5):443–53.

Sanjuan-Pla, A, I C Macaulay, C T Jensen, P S Woll, T C Luis, A Mead, S Moore, et al. 2013. “Platelet-Biased Stem Cells Reside at the Apex of the Haematopoietic Stem-Cell Hierarchy.” *Nature* 502 (7470):232–36.

Sansanayudh, N, T Anothaisintawee, D Muntham, M McEvoy, J Attia, A Thakkinstian. 2014. “Mean Platelet Volume and Coronary Artery Disease: A Systematic Review and Meta-Analysis.” *International Journal of Cardiology* 175 (3):433–40.

Santimone, I, A di Castelnuovo, A de Curtis, M Spinelli, D Cugino, F Gianfagna, F Zito, et al.

2011. “White Blood Cell Count, Sex and Age Are Major Determinants of Heterogeneity of Platelet Indices in an Adult General Population: Results from the MOLI-SANI Project.” *Haematologica* 96 (8):1180–88.

Sauvage, F J de, P E Hass, S D Spencer, B E Malloy, A L Gurney, S A Spencer, W C Darbonne, et al. 1994. “Stimulation of Megakaryocytopoiesis and Thrombopoiesis by the c-Mpl Ligand.” *Nature* 369 (6481):533–38.

Savage, B, P R McFadden, S R Hanson, and L Harker. 1986. “The Relation of Platelet Density to Platelet Age: Survival of Low- and High-Density 111indium-Labeled Platelets in Baboons.” *Blood* 68:386–93.

Sawai, C M, S Babovic, S Upadhaya, M Merad, C J Eaves, and B Reizis. 2016. “Hematopoietic Stem Cells Are the Major Source of Multilineage Hematopoiesis in Adult Animals.” Immunity 45 (3): 597-609.

Saxena, S, A D Cramer, J M Weiner, and R Carmel. 1987. “Platelet Counts in Three Racial Groups.” *Am J Clin Pathol* 88 (1).:106–9.

Schroeder, Timm. 2010. “Hematopoietic Stem Cell Heterogeneity: Subtypes, Not Unpredictable Behavior.” *Cell Stem Cell* 6 (3):203–7.

Schunkert, H, I R König, S Kathiresan, M P Reilly, T L Assimes, H Holm, M Preuss, et al. 2011. “Large-Scale Association Analysis Identifies 13 New Susceptibility Loci for Coronary Artery Disease.” *Nature Genetics* 43 (4):333–40.

Schwertz, H, S Köster, W H A Kahr, N Michetti, B F Kraemer, D A Weitz, R C Blaylock, et al. 2010. “Anucleate Platelets Generate Progeny.” *Blood* 115 (18):3801–9.

Segal, J B, and A R Moliterno. 2006. “Platelet Counts Differ by Sex, Ethnicity, and Age in the United States.” *Ann Epidemiol* 16 (2):123–30.

Senaran, H, M Ileri, A Altinbas, A Kosar, E Yetkin, M Ozturk, Y Karaaslan, and S Kirazli. 2001. “Thrombopoietin and Mean Platelet Volume in Coronary Artery Disease.” *Clin Cardiol* 24 (5):405–8.

Shim, M, A Hoover, N Blake, J G Drachman, and J A Reems. 2004. “Gene Expression Profile of Primary Human CD34+CD38lo Cells Differentiating along the Megakaryocyte Lineage.” *Experimental Hematology* 32 (7):638–48.

Shimomura, H, H Ogawa, H Arai, Y Moriyama, K Takazoe, N Hirai, K Kaikita, et al. 1998. “Serial Changes in Plasma Levels of Soluble P-Selectin in Patients with Acute Myocardial Infarction.” *The American Journal of Cardiology* 81 (4):397–400.

Shin, J Y, W Hu, M Naramura, and C Y Park. 2014. “High c-Kit Expression Identifies Hematopoietic Stem Cells with Impaired Self-Renewal and Megakaryocytic Bias.” *The Journal of Experimental Medicine* 211 (2):217–31.

Shinjo, K, A Takeshita, S Nakamura, K Naitoh, M Yanagi, T Tobita, K Ohnishi, and R Ohno. 1998. “Serum Thrombopoietin Levels in Patients Correlate Inversely with Platelet Counts during Chemotherapy-Induced Thrombocytopenia.” *Leukemia* 12 (3):295–300.

Shivdasani, R A, Y Fujiwara, M A McDevitt, and S H Orkin. 1997. “A Lineage-Selective Knockout Establishes the Critical Role of Transcription Factor GATA-1 in

Megakaryocyte Growth and Platelet Development." *EMBO Journal* 16 (13):3965–73.

Sieburg, H B, R H Cho, B Dykstra, N Uchida, C J Eaves, and C E Muller-Sieburg. 2006. "The Hematopoietic Stem Compartment Consists of a Limited Number of Discrete Stem Cell Subsets." *Blood* 107 (6):2311–16.

Siminovitch, L, E A McCulloch, and J E Till. 1963. "The Distribution of Colony-Forming Cells among Spleen Colonies." *Journal of Cellular and Comparative Physiology* 62 (3):327–36.

Silveri, F, R De Angelis, F Argentati, D Brecciaroli, S Muti, and C Cervini. 1996. "Hypertrophic Osteoarthropathy: Endothelium and Platelet Function." *Clin Rheumatol* 15 (5):435–39.

Singh, R K, and T A Cooper. 2012. "Pre-mRNA Splicing in Disease and Therapeutics." *Trends Mol Med* 18 (8):472–82.

Slater, D N, E A Trowbridge, and J F Martin. 1983. "The Megakaryocyte in Thrombocytopenia: A Microscopic Study Which Supports the Theory That Platelets Are Produced in the Pulmonary Circulation." *Thromb Res* 31 (1):163–76.

Slavka, G, T Perkmann, H Haslacher, S Greisenegger, C Marsik, O F Wagner, and G Endler. 2011. "Mean Platelet Volume May Represent a Predictive Parameter for Overall Vascular Mortality and Ischemic Heart Disease." *Arteriosclerosis, Thrombosis, and Vascular Biology* 31 (5):1215–18.

Smolina, K, F L Wright, M Rayner, and M J Goldacre. 2012. "Determinants of the Decline in Mortality from Acute Myocardial Infarction in England between 2002 and 2010: Linked National Database Study." *Bmj* 344 (jan25 2):d8059–d8059.

Smyth, D W, J F Martin, L Michalis, C A Bucknall, and D E Jewitt. 1993. "Influence of Platelet Size before Coronary Angioplasty on Subsequent Restenosis." *Eur J Clin Invest* 23 (6):361–67.

Soranzo, N, A Rendon, C Gieger, C I Jones, N A Watkins, S Menzel, A Döring, et al. 2009. "Anovel Variant on Chromosome 7q22.3 Associated with Mean Platelet Volume, Counts, and Function." *Blood* 113 (16):3831–37.

Sportoletti, P, E Varasano, R Rossi, O Bereshchenko, D Cecchini, I Gionfriddo, N Bolli, et al. 2013. "The Human NPM1 Mutation A Perturbs Megakaryopoiesis in a Conditional Mouse Model." *Blood* 121 (17):3447–58.

Steg, P G, S K James, D Atar, L P Badano, C B Lundqvist, M A Borger, C Di Mario, et al. 2012. "ESC Guidelines for the Management of Acute Myocardial Infarction in Patients Presenting with ST-Segment Elevation." *European Heart Journal* 33 (20):2569–2619.

Stegle, O, S A Teichmann, and J C Marioni. 2015. "Computational and Analytical Challenges in Single-Cell Transcriptomics." *Nature Reviews Genetics* 16 (3):133–45.

Steinhubl, S R, R C Becker, G Goodman, M T Roe, A Kuliopulos, D J Moliterno, P A French, S S Smyth, et al. 2008. "Coupled Receptors as Signaling Targets for Antiplatelet Therapy − G-Protein Brief Review G-Protein–Coupled Receptors as Signaling Targets for Antiplatelet Therapy for the 2008 Platelet Colloquium Participants*." *Arterioscler Thromb Vasc Biol* 29 (20):449–57.

Stenberg, P E, and J Levin. 1989. “Ultrastructural Analysis of Acute Immune Thrombocytopenia in Mice: Dissociation between Alterations in Megakaryocytes and Platelets.” *J Cell Physiol* 141 (1):160–69.

Stenberg, P E, J Levin, G Baker, Y Mok, and L Corash. 1991. “Neuraminidase-Induced Thrombocytopenia in Mice: Effects on Thrombopoiesis.” *J Cell Physiol* 147 (1):7–16.

Stoffel, R, A Wiestner, and R C Skoda. 1996. “Thrombopoietin in Thrombocytopenic Mice: Evidence against Regulation at the mRNA Level and for a Direct Regulatory Role of Platelets.” *Blood* 87 (2):567–73.

Sun, J, A Ramos, B Chapman, J B Johnnidis, L Le, Y J Ho, A Klein, O Hofmann, and F D Camargo. 2014. “Clonal Dynamics of Native Haematopoiesis.” *Nature* 514 (7522):322–27.

Sun, L, J R Gorospe, E P Hoffman, and A Koneti Rao. 2007. “Decreased Platelet Expression of Myosin Regulatory Light Chain Polypeptide (MYL9) and Other Genes with Platelet Dysfunction and CBFA2/RUNX1 Mutation: Insights from Platelet Expression Profiling.” *Journal of Thrombosis and Haemostasis* 5 (1):146–54.

Sun, W, and J R Downing. 2004. “Haploinsufficiency of AML1 Results in a Decrease in the Number of LTR-HSCs While Simultaneously Inducing an Increase in More Mature Progenitors.” *Blood* 104 (12):3565–72.

Sungaran, R, B Markovic, and B H Chong. 1997. “Localization and Regulation of Thrombopoietin mRNa Expression in Human Kidney, Liver, Bone Marrow, and Spleen Using in Situ Hybridization.” *Blood* 89 (1):101–7.

Suter, D M, N Molina, D Gatfield, K Schneider, U Schibler, and F Naef. 2011. “Mammalian Genes Are Transcribed with Widely Different Bursting Kinetics.” *Science* 332 (6028):472–74.

Taglieri, N, F Saia, C Rapezzi, C Marrozzini, M L Bacchi Reggiani, T Palmerini, P Ortolani, et al. 2011. “Prognostic Significance of Mean Platelet Volume on Admission in an Unselected Cohort of Patients with Non ST-Segment Elevation Acute Coronary Syndrome.” *Thromb Haemost* 106 (1):132–40.

Takano, H, H Ema, K Sudo, and H Nakauchi. 2004. “Asymmetric Division and Lineage Commitment at the Level of Hematopoietic Stem Cells.” *The Journal of Experimental Medicine* 199 (3):295–302.

Takayama, M, R Fujita, M Suzuki, R Okuyama, S Aiba, H Motohashi, and M Yamamoto. 2010. “Genetic Analysis of Hierarchical Regulation for *Gata1* and *NF-E2 p45* Gene Expression in Megakaryopoiesis.” *Molecular and Cellular Biology* 30 (11):2668–80.

Tang, F, K Lao, and M A Surani. 2011. “Development and Applications of Single-Cell Transcriptome Analysis.” *Nature Methods* 8 (4 Suppl):S6-11.

Tanum, G, and A Engeset. 1983. “Low Ploidy Megakaryocytes in Steady-State Rat Bone Marrow.” *Blood* 62 (1):87–91.

Tavassoli, M, and M Aoki. 1989. “Localization of Megakaryocytes in the Bone Marrow.” *Blood Cells* 15 (1):3–14.

Thompson, C B, K A Eaton, S M Princiotta, C A Rushin, and C R Valeri. 1982. “Size Dependent Platelet Subpopulations: Relationship of Platelet Volume to Ultrastructure, Enzymatic Activity, and Function.” *Br J Haematol* 50 (3):509–19..

Thompson, C B, J A Jakubowski, P G Quinn, D Deykin, and C R Valeri. 1983. “Platelet Size as a Determinant of Platelet Function.” *J Lab Clin Med* 101 (2):205–13.

Thompson, C B, D G Love, P G Quinn, and C R Valeri. 1983. “Platelet Size Does Not Correlate with Platelet Age.” *Blood* 62 (2):487–94.

Thon, J N, A Montalvo, S Patel-Hett, M T Devine, J L Richardson, A Ehrlicher, M K Larson, et al. 2010. “Cytoskeletal Mechanics of Proplatelet Maturation and Platelet Release.” *The Journal of Cell Biology* 191 (4):861–74.

Thon, J N, and J E Italiano. 2012. “Does Size Matter in Platelet Production?” *Blood* 120 (8):1552–61.

Thorén, L A, K Liuba, D Bryder, J M Nygren, C T Jensen, H Qian, J Antonchuk, and S W Jacobsen. 2008. “Kit Regulates Maintenance of Quiescent Hematopoietic Stem Cells.” *Journal of Immunology (Baltimore, Md. : 1950)* 180 (4):2045–53.

Tijssen, M R, A Cvejic, A Joshi, R L Hannah, R Ferreira, A Forrai, D C Bellissimo, et al. 2011. “Genome-Wide Analysis of Simultaneous GATA1/2, RUNX1, FLI1, and SCL Binding in Megakaryocytes Identifies Hematopoietic Regulators.” *Developmental Cell* 20 (5):597–609.

Till, J E, E A McCulloch, and L Siminovitch. 1964. “A Stochastic Model of Stem Cell Proliferation, based on the growth of Spleen Colony-forming Cells.” *Proceedings of the National Academy of Sciences of the United States of America* 51 (1). National Academy of Sciences:29–36.

Tiwari, S, J E Italiano, D C Barral, E H Mules, E K Novak, R T Swank, M C Seabra, and R A Shivdasani. 2003. “A Role for Rab27b in NF-E2-Dependent Pathways of Platelet Formation.” *Blood* 102 (12):3970–79.

Tomer, A, P Friese, R Conklin, W Bales, L Archer, L A Harker, and S A Burstein. 1989. “Flow Cytometric Analysis of Megakaryocytes from Patients with Abnormal Platelet Counts.” *Blood* 74 (2):594–601.

Tomer, A, L A Harker, and S A Burstein. 1988. “Flow Cytometric Analysis of Normal Human Megakaryocytes.” *Blood* 71 (5):1244–52.

Topol, E J, J McCarthy, S Gabriel, D J Moliterno, W J Rogers, L K Newby, M Freedman, et al. 2001. “Single Nucleotide Polymorphisms in Multiple Novel Thrombospondin Genes May Be Associated with Familial Premature Myocardial Infarction.” *Circulation* 104 (22):2641–44.

Trakala, M, S Rodriguez-Acebes, M Maroto, C E Symonds, D Santamaria, S Ortega, M Barbacid, J Mendez, and M Malumbres. 2015. “Functional Reprogramming of Polyploidization in Megakaryocytes.” *Developmental Cell* 32 (2):155–67.

Trapnell, C, D Cacchiarelli, J Grimsby, P Pokharel, S Li, M Morse, N J Lennon, K J Livak, T S Mikkelsen, and J L Rinn. 2014. “The Dynamics and Regulators of Cell Fate Decisions

Are Revealed by Pseudotemporal Ordering of Single Cells." *Nature Biotechnology* 32 (4):381–86.

Trowbridge, E A, and P J Harley. 1984. "A Computer Model of the Random Binary Sequential Division of Megakaryocyte Cytoplasm to Produce Platelets." *Phys Med Biol* 29:1477.

Trowbridge, E A, and J F Martin. 1987. "The Platelet Volume Distribution: A Signature of the Prethrombotic State in Coronary Heart Disease?" *Thromb Haemost* 58 (2):714–17.

Trowbridge, E A, J F Martin, and D N Slater. 1982. "Evidence for a Theory of Physical Fragmentation of Megakaryocytes, Implying That All Platelets Are Produced in the Pulmonary Circulation." *Thromb Res* 28 (4):461–75.

Trowbridge, E A, D N Slater, Y T Kishk, B W Woodcock, and J F Martin. 1984. "Platelet Production in Myocardial Infarction and Sudden Cardiac Death." *Thromb Haemost* 52 (2):167.

Ulich, TR, J del Castillo, S Yin, S Swift, D Padilla, G Senaldi, L Bennett, J Shutter, J Bogenberger, and D Sun. 1995. "Megakaryocyte Growth and Development Factor Ameliorates Carboplatin- Induced Thrombocytopenia in Mice." *Blood* 86 (3): 971-976.

Vannucchi, A M, F Paoletti, S Linari, C Cellai, R Caporale, P R Ferrini, M Sanchez, G Migliaccio, and A R Migliaccio. 2000. "Identification and Characterization of a Bipotent (Erythroid and Megakaryocytic) Cell Precursor from the Spleen of Phenylhydrazine-Treated Mice." *Blood* 95 (8):2559–68.

Varol, E, A Arslan, H Yucel, M Ozaydin, D Erdogan, and A Dogan. 2011. "Increased Mean Platelet Volume in Patients with Aortic Stenosis." *Clinical and Applied Thrombosis /Hemostasis* 17 (6):17–20.

Varol, E, M Ozaydin, Y Türker, and S Alaca. 2009. "Mean Platelet Volume, an Indicator of Platelet Activation, Is Increased in Patients with Mitral Stenosis and Sinus Rhythm." *Scandinavian Journal of Clinical and Laboratory Investigation* 69 (6):708–12.

Varol, E, F Aksoy, H A Bas, H Ari, and M Ozaydin. 2014. "Mean Platelet Volume Is Elevated in Patients With Low High-Density Lipoprotein Cholesterol." *Angiology* 65 (8):733–36.

Velten, L, S F Haas, S Raffel, S Blaszkiewicz, S Islam, B P Hennig, C Hirche, et al. 2017. "Human Haematopoietic Stem Cell Lineage Commitment Is a Continuous Process." *Nature Cell Biology* 19 (4).

Wang, E T, R Sandberg, S Luo, I Khrebtukova, L Zhang, C Mayr, S F Kingsmore, G P Schroth, and C B Burge. 2008. "Alternative Isoform Regulation in Human Tissue Transcriptomes." *Nature* 456 (7221):470–76.

Wang, R, R L Stone, J T Kaelber, R H Rochat, A M Nick, K V Vijayan, V Afshar-Kharghan, et al. 2015. "Electron Cryotomography Reveals Ultrastructure Alterations in Platelets from Patients with Ovarian Cancer." *Proceedings of the National Academy of Sciences of the United States of America* 112 (46):14266–71.

Wang, Z, Q Ma, Q Liu, H Yu, L Zhao, S Shen, and J Yao. 2008. "Blockade of SDF-1/CXCR4 Signalling Inhibits Pancreatic Cancer Progression in Vitro via Inactivation of Canonical Wnt Pathway." *British Journal of Cancer* 99 (10):1695–1703.

Watkins, N A, A Gusnanto, B De Bono, S De, D Miranda-Saavedra, D L Hardie, W G J Angenent, et al. 2009. “A HaemAtlas: Characterizing Gene Expression in Differentiated Human Blood Cells.” *Blood* 113 (19). https://doi.org/10.1182/blood-2008-06-162958.

Watson, S. P., J. M.J. Herbert, and A. Y. Pollitt. 2010. “GPVI and CLEC-2 in Hemostasis and Vascular Integrity.” *Journal of Thrombosis and Haemostasis* 8 (7):1456–67.

Weissman, I L, D J Anderson, and F Gage. 2001. “Stem and Progenitor Cells: Origins, Phenotypes, Lineage Commitments, and Transdifferentiations.” *Annual Review of Cell and Developmental Biology* 17 (1):387–403.

Weksberg, D C, S M Chambers, N C Boles, and M A Goodell. 2008. “CD150- Side Population Cells Represent a Functionally Distinct Population of Long-Term Hematopoietic Stem Cells.” *Blood* 111 (4):2444–51.

Wen, Q, B Goldenson, S J Silver, M Schenone, V Dancik, Z Huang, L Zhi Wang, et al. 2012. “Identification of Regulators of Polyploidization Presents Therapeutic Targets for Treatment of AMKL.” *Cell* 150 (3):575–89.

Weyrich, A S, H Schwertz, L W Kraiss, and G A Zimmerman. 2009. “Protein Synthesis by Platelets: Historical and New Perspectives.” *Journal of Thrombosis and Haemostasis* 7 (2):241–46.

White, M J, S M Schoenwaelder, E C Josefsson, K E Jarman, K J Henley, C James, M A Debrincat, S P Jackson, D C S Huang, and B T Kile. 2012. “Caspase-9 Mediates the Apoptotic Death of Megakaryocytes and Platelets, but Is Dispensable for Their Generation and Function.” *Blood* 119 (18):4283–90.

Wilson, N K, D G Kent, F Buettner, M Shehata, I C Macaulay, F J Calero-Nieto, M Sánchez Castillo, et al. 2015. “Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations.” *Cell Stem Cell* 16 (6):712–24.

Wolber, E M, J Fandrey, U Frackowski, and W Jelkmann. 2001. “Hepatic Thrombopoietin mRNA Is Increased in Acute Inflammation.” *Thrombosis and Haemostasis* 86 (6):1421–24.

Wong, K F, and J K C Chan. 2008. “Are ‘dysplastic’ and Hypogranular Megakaryocytes Specific Markers for Myelodysplastic Syndrome?” *British Journal of Haematology* 77 (4):509–14.

Woolthuis, C M, and C Y Park. 2016. “Hematopoietic Stem/progenitor Cell Commitment to the Megakaryocyte Lineage.” *Blood* 127 (10):1242–48.

Würtz, M, E L Grove, L N Wulff, A K Kaltoft, H H Tilsted, L O Jensen, A M Hvas, and S D Kristensen. 2010. “Patients with Previous Definite Stent Thrombosis Have a Reduced Antiplatelet Effect of Aspirin and a Larger Fraction of Immature Platelets.” *JACC: Cardiovascular Interventions* 3 (8):828–35.

Yamada, E. 1957. “The Fine Structure of the Megakaryocyte in the Mouse Spleen.” *Acta Anat (Basel)* 29 (3):267–90.

Yamamoto, R, Y Morita, J Ooehara, S Hamanaka, M Onodera, K L Rudolph, H Ema, and H Nakauchi. 2013. “Clonal Analysis Unveils Self-Renewing Lineage-Restricted Progenitors

Generated Directly from Hematopoietic Stem Cells." *Cell* 154 (5):1112–26.

Yamazaki, S, H Ema, G Karlsson, T Yamaguchi, H Miyoshi, S Shioda, M M Taketo, S Karlsson, A Iwama, and H Nakauchi. 2011. "Nonmyelinating Schwann Cells Maintain Hematopoietic Stem Cell Hibernation in the Bone Marrow Niche." *Cell* 147 (5):1146–58.

Yang, A, L Pizzulli, and B Lüderitz. 2006. "Mean Platelet Volume as Marker of Restenosis after Percutaneous Transluminal Coronary Angioplasty in Patients with Stable and Unstable Angina Pectoris." *Thrombosis Research* 117 (4):371–77.

Yang, L V, R H Nicholson, J Kaplan, A Galy, and L Li. 2001. "Hemogen Is a Novel Nuclear Factor Specifically Expressed in Mouse Hematopoietic Development and Its Human Homologue EDAG Maps to Chromosome 9q22, a Region Containing Breakpoints of Hematological Neoplasms." *Mechanisms of Development* 104 (1–2):105–11.

Yang, M, K Li, M H L Ng, P M P Yuen, T F Fok, C K Li, P J Hogg, and B H Chong. 2003. "Thrombospondin-1 Inhibits in Vitro Megakaryocytopoiesis via CD36." *Thrombosis Research* 109 (1):47–54.

Yang, M, E Liang, L Mao, B H Chong, and C Li. 2016. "Thrombospondin-1 Induces Apoptosis in Megakaryocytic Leukemia Via CD36 and Caspase-3 Signaling." *Blood* 128 (22).

Yavuz, B, D T Ertugrul, A A Yalcin, M Kucukazman, N Ata, and K Dal. 2009. "Increased Mean Platelet Volume in Rheumatic Mitral Stenosis: A Possible Factor for Thromboembolic Events." *Journal of Cardiology* 53 (2):204–7.

Yazici, H U, F Poyraz, N Sen, Y Tavil, M Turfan, M Tulmac, and A Abaci. 2011. "Relationship between Mean Platelet Volume and Left Ventricular Systolic Function in Patients with Metabolic Syndrome and ST-Elevation Myocardial Infarction." *Clin Invest Med* 34 (6). Canada:E330.

Yekutieli, D, and Y Benjamini. 2001. "The Control of the False Discovery Rate in Multiple Testing under Dependency under Dependency." *The Annals of Statistics* 29 (4):1165–88.

Yoshida, T, S Yao-Ming Ng, J C Zuniga-Pflucker, and K Georgopoulos. 2006. "Early Hematopoietic Lineage Restrictions Directed by Ikaros." *Nature Immunology* 7 (4):382–91.

Yoshihara, H, F Arai, K Hosokawa, T Hagiwara, K Takubo, Y Nakamura, Y Gomei, et al. 2007. "Thrombopoietin/MPL Signaling Regulates Hematopoietic Stem Cell Quiescence and Interaction with the Osteoblastic Niche." *Cell Stem Cell* 1 (6):685–97.

Youssefian, T, A Drouin, J Massé, J Guichard, E M Cramer, A Piccoli, L Totani, et al. 2002. "Host Defense Role of Platelets: Engulfment of HIV and Staphylococcus Aureus Occurs in a Specific Subcellular Compartment and Is Enhanced by Platelet Activation." *Blood* 99 (11):4021–29.

Youssefian, T, and E M Cramer. 2000. "Megakaryocyte Dense Granule Components Are Sorted in Multivesicular Bodies." *Blood* 95 (12):4004–7.

Zerbino, D R, P Achuthan, W Akanni, M R Amode, D Barrell, J Bhai, K Billis, et al. 2017. "Ensembl 2018." *Nucleic Acids Research*.

Zhao, M, J M Perry, H Marshall, A Venkatraman, P Qian, X C He, J Ahamed, and L Li. 2014. “Megakaryocytes Maintain Homeostatic Quiescence and Promote Post-Injury Regeneration of Hematopoietic Stem Cells.” *Nature Medicine* 20 (11):1321–26.

Zharikov, S, and S Shiva. 2013. “Platelet Mitochondrial Function: From Regulation of Thrombosis to Biomarker of Disease.” *Biochemical Society Transactions* 41 (1):118–23.

Zheng, C, R Yang, Z Han, B Zhou, L Liang, and M Lu. 2008. “TPO-Independent Megakaryocytopoiesis.” *Crit Rev Oncol Hematol* 65 (3):212–22.

Zimmet, J M, P Toselli, and K Ravid. 1998. “Cyclin D3 and Megakaryocyte Development: Exploration of a Transgenic Phenotype.” *Stem Cells* 16 Suppl 2 (1066–5099):97–106.

Zimmet, J, and K Ravid. 2000. “Polyploidy.” *Experimental Hematology* 28 (1):3–16.

APPENDIX 1

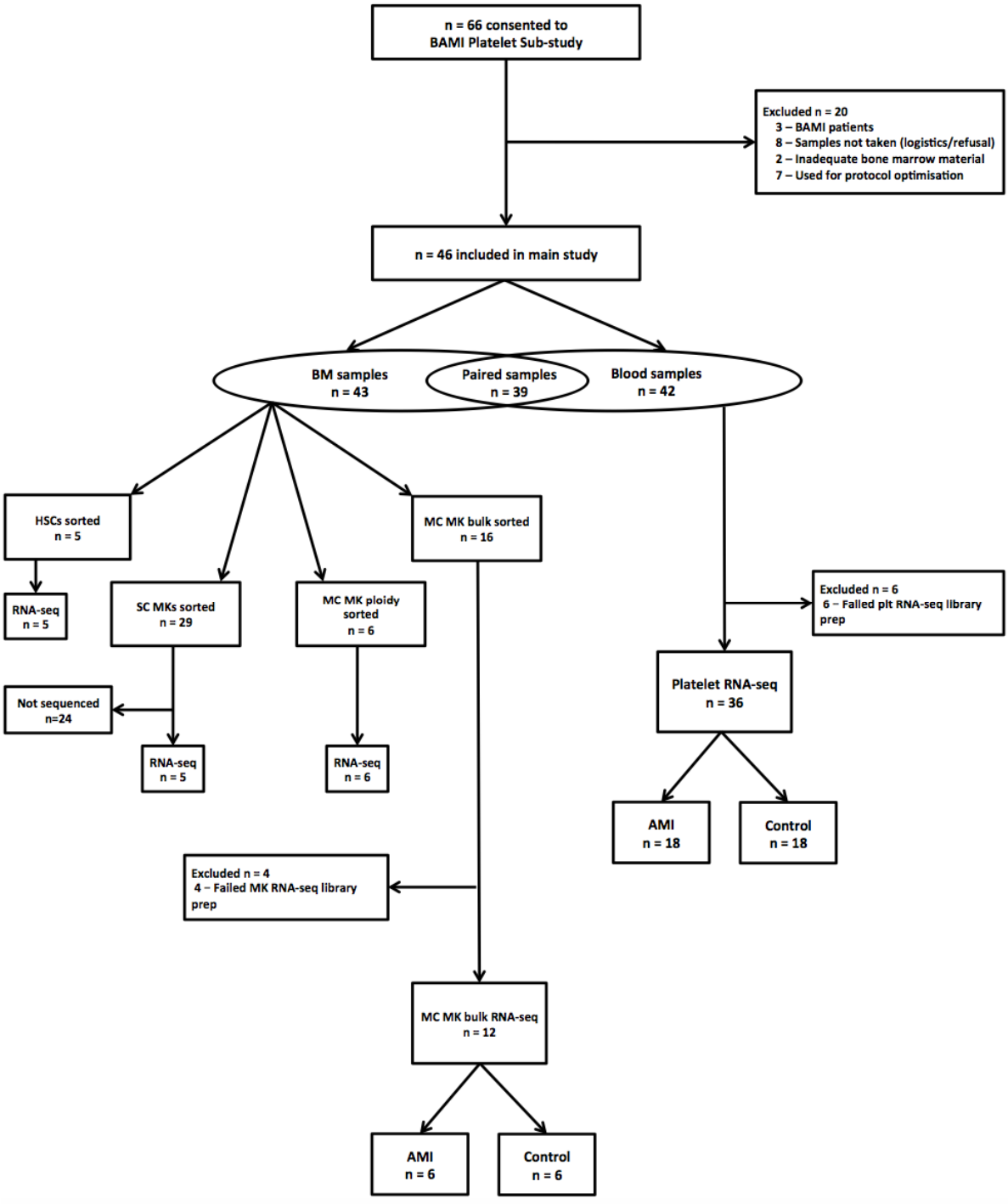


Figure S1.1 Study design
BM: bone marrow, HSC: haematopoietic stem cell, MK: megakaryocyte, Plt: platelet, SC: single cell, MC: multicell, AMI: acute myocardial infarction.

APPENDIX 2

Recruit no	BP ID	Recruitment date	Collection date	Group	Not included in main study	Sample not taken		HSCs		Megakaryocytes					Platelets	
						BM	Blood	HSC sorted	RNA-seq	MK parameters	Sorted: bulk	Sorted: ploidy	Sorted: SC	RNA-seq	Plt parameters	RNA-seq
1	S00QAH	10/06/14	10/06/14	AMI	Excluded											
5	S00VTX	22/09/14	23/09/14	AMI												
48	S01F9I	01/07/15	01/07/15	AMI												
13	S010YS	14/12/14	15/12/14	Ctrl	Failed		1									
16	S0122A	18/01/15	19/01/15	Ctrl			1									
2	-	30/06/14	-	AMI	Sample not taken	1	1									
12	-	07/12/14	-	Ctrl		1	1									
14	-	14/12/14	-	Ctrl		1	1									
17	-	20/01/15	-	Ctrl		1	1									
24	-	10/03/15	-	Ctrl		1	1									
38	-	31/05/15	-	Ctrl		1	1									
42	-	14/06/15	-	Ctrl		1	1									
51	-	08/07/15	-	Ctrl		1	1									
3	S000SM	08/09/14	15/10/14	Ctrl	Protocol optimisation		1									
4	S0014A	08/09/14	27/10/14	Ctrl			1									
6	S00VUV	22/09/14	23/09/14	Ctrl			1									
7	S00W53	07/10/14	08/10/14	Ctrl			1									
8	S000EA	16/11/14	17/11/14	Ctrl			1									
9	S0102I	25/11/14	26/11/14	Ctrl			1									
10	S0103G	30/11/14	01/12/14	Ctrl			1									
11	S010DX	07/12/14	08/12/14	Ctrl	Included					1		MC:ploidy	SC	1	1	1
15	S011FP	13/01/15	14/01/15	Ctrl						1		MC:ploidy	SC	1	1	
18	S01318	20/01/15	21/01/15	AMI						1		MC:ploidy	SC	1	1	1
19	S013XI	27/01/15	28/01/15	Ctrl						1	MC:bulk	MC:ploidy	SC	1	1	1
20	S0146V	03/02/15	04/02/15	Ctrl						1	MC:bulk				1	1
21	S014J5	10/02/15	11/02/15	AMI						1	MC:bulk			1	1	1
22	S01CST	10/02/15	18/05/15	Ctrl						1	MC:bulk		SC	1	1	1
23	S015Q0	09/03/15	10/03/15	AMI						1			SC		1	1
25	S015Y8	10/03/15	11/03/15	AMI						1			SC		1	1
26	S0160Z	10/03/15	11/03/15	Ctrl				HSC	1	1			SC	1	1	1
27	S0165P	15/03/15	16/03/15	Ctrl						1			SC		1	1
28	S016VA	22/03/15	23/03/15	Ctrl						1			SC		1	1
29	S017SC	30/03/15	31/03/15	Ctrl						1			SC		1	
30	S017TA	31/03/15	01/04/15	Ctrl						1			SC	1	1	
31	S01ATZ	05/05/15	06/05/15	Ctrl						1	MC:bulk			1	1	1
32	S01CEK	13/05/15	14/05/15	AMI						1	MC:bulk		SC	1	1	1
33	S01CDM	13/05/15	14/05/15	Ctrl						1			SC		1	1
34	S01CUP	17/05/15	18/05/15	Ctrl						1			SC	1	1	1

35	S01CYH	17/05/15	18/05/15	Ctrl						1			SC	1	1	1
36	S01DFE	26/05/15	27/05/15	AMI						1			SC		1	1
37	S01DGC	26/05/15	27/05/15	Ctrl						1			SC		1	
39	S01DM0	31/05/15	01/06/15	Ctrl						1			SC		1	1
40	S01DSP	02/06/15	03/06/15	Ctrl						1	MC:bulk		SC	1	1	1
41	S01EFA	14/06/15	15/06/15	Ctrl						1			SC		1	1
43	S01EH6	16/06/15	17/06/15	AMI						1	MC:bulk		SC	1	1	1
44	S01EMX	21/06/15	22/06/15	AMI						1	MC:bulk		SC	1	1	1
45	S01F4S	23/06/15	24/06/15	AMI						1			SC		1	1
46	S01F5Q	28/06/15	29/06/15	AMI						1	MC:bulk		SC		1	1
47	S01F7M	29/06/15	30/06/15	AMI						1			SC		1	1
49	S01FBE	02/07/15	03/07/15	AMI						1	MC:bulk		SC		1	1
50	S01FVB	06/07/15	07/07/15	AMI						1	MC:bulk		SC	1	1	1
52	S01FX7	09/07/15	10/07/15	Ctrl				HSC	1	1					1	
53	S01G0W	12/07/15	13/07/15	AMI						1					1	
54	S01G1U	13/07/15	14/07/15	Ctrl		1		HSC	1	1						
55	S01G6K	16/07/15	17/07/15	AMI						1			SC		1	1
56	S01GBA	19/07/15	20/07/15	Ctrl		1		HSC	1	1						
57	S01GC8	20/07/15	21/07/15	AMI						1					1	1
58	S01GNN	27/07/15	28/07/15	Ctrl		1				1						
59	S01GPJ	27/07/15	28/07/15	Ctrl		1		HSC	1	1						
60	S01H4K	02/08/15	03/08/15	Ctrl						1	MC:bulk	MC:ploidy		1	1	1
61	S01M8T	26/08/15	27/08/15	Ctrl						1	MC:bulk	MC:ploidy		1	1	1
62	S01P4P	06/10/15	07/10/15	AMI						1	MC:bulk			1	1	1
63	S01Q6H	12/10/15	13/10/15	Ctrl						1	MC:bulk			1	1	1
64	S01R7B	22/10/15	23/10/15	Ctrl	1										1	1
65	S01W4Y	01/02/16	02/02/16	AMI	1										1	1
66	S01WBK	14/02/16	15/02/16	AMI	1										1	1

Table S2.1 Log of patients recruited and sample use
BP ID: Blueprint ID (anonymised ID), BM: bone marrow, HSC: haematopoietic stem cell, MK: megakaryocyte, Plt: platelet, SC: single cell, MC: multicell, AMI: acute myocardial infarction, Ctrl: control.