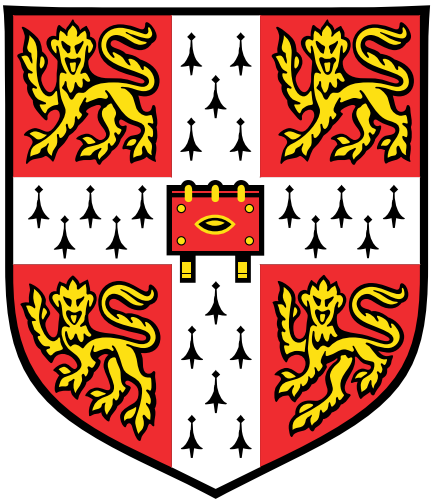
BEYOND THE RARE BLOOD GROUP VEL, UNCOVERING THE FUNCTIONS OF SMIM1 IN BLOOD AND IN OTHER ORGANS



Ana Rita Ramalho Tomé

Peterhouse

Department of Haematology

University of Cambridge

This dissertation is submitted for the degree of Doctor of Philosophy

December 2018

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 in the Preface 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 in the Preface and specified in the text. It does not exceed the prescribed word limit for the relevant Degree Committee.

Signed: Date:

Beyond the rare blood group Vel, uncovering the functions of SMIM1 in blood and in other organs

Ana Rita Ramalho Tomé

ABSTRACT

Vel is a universal antigen present on the red blood cells (RBCs) membrane, which defines the Vel-blood group system. The identification of Vel-encoding gene, *SMIM1*, suggested that this is a regulator of erythropoiesis. Its role in RBCs and potential role in other cells, however, remains largely unknown.

Here, I aimed to characterise the role of SMIM1 in haematopoietic lineages and other organs by using two main approaches (*in vivo* and *in vitro*) and three models (human volunteers, mouse and cell lines). First, using available datasets, I showed that *SMIM1* is expressed in megakaryocytes (MKs), platelets, neutrophils and naïve, memory and class-switched B-lymphocytes. I further demonstrated that SMIM1 localisation and the type of multimers it forms in some of those cells are cell type dependent.

Second, I assessed the effect of the absence of SMIM1 in qualitative and quantitative blood traits in two human cohorts (Vel-negative blood donors recruited in this study and Vel-negative/weak individuals of the 400,000 UK Biobank study) and in a *Smim1* mutant mouse model. Small but significant alterations were observed in RBC, platelets and neutrophils counts in Vel-negative individuals and *Smim1* mutant mice, however, the directionality of these changes and their significance was model specific.

Third, I investigated the role of SMIM1 in platelets formation and function. I showed that SMIM1 is phosphorylated during platelet activation in humans and its ablation in male mouse platelets leads to a reduction of P-selectin surface expression on resting platelets. These findings suggest a role for SMIM1 during platelet activation. Furthermore, I described a novel metabolic phenotype both in human and mouse *SMIM1* knockouts. I showed that SMIM1 absence leads to an increase in body weight (and consequently in body mass index (BMI)) in male Vel-negative donors and female *Smim1*$^{-/-}$ mice.

Altogether, these findings provide new insights into the role of SMIM1 and the biological processes it is involved, which may have the potential to reveal unknown clinical phenotypes related with SMIM1 ablation and to inform their prevention and treatment.

ABSTRACT

Vel is a universal antigen present on the red blood cells (RBCs) membrane, which defines the Vel-blood group system. The identification of Vel-encoding gene, *SMIM1*, suggested that this is a regulator of erythropoiesis. Its role in RBCs and potential role in other cells, however, remains largely unknown.

Here, I aimed to characterise the role of SMIM1 in haematopoietic lineages and other organs by using two main approaches (*in vivo* and *in vitro*) and three models (human volunteers, mouse and cell lines). First, using available datasets, I showed that *SMIM1* is expressed in megakaryocytes (MKs), platelets, neutrophils and naïve, memory and class-switched B-lymphocytes. I further demonstrated that SMIM1 localisation and the type of multimers it forms in some of those cells are cell type dependent.

Second, I assessed the effect of the absence of SMIM1 in qualitative and quantitative blood traits in two human cohorts (Vel-negative blood donors recruited in this study and Vel-negative/weak individuals of the 400,000 UK Biobank study) and in a *Smim1* mutant mouse model. Small but significant alterations were observed in RBC, platelets and neutrophils counts in Vel-negative individuals and *Smim1* mutant mice, however, the directionality of these changes and their significance was model specific.

Third, I investigated the role of SMIM1 in platelets formation and function. I showed that SMIM1 is phosphorylated during platelet activation in humans and its ablation in male mouse platelets leads to a reduction of P-selectin surface expression on resting platelets. These findings suggest a role for SMIM1 during platelet activation. Furthermore, I described a novel metabolic phenotype both in human and mouse *SMIM1* knockouts. I showed that SMIM1 absence leads to an increase in body weight (and consequently in body mass index (BMI)) in male Vel-negative donors and female *Smim1*$^{-/-}$ mice.

Altogether, these findings provide new insights into the role of SMIM1 and the biological processes it is involved, which may have the potential to reveal unknown clinical phenotypes related with SMIM1 ablation and to inform their prevention and treatment.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisors, Professor Willem H. Ouwehand and Dr Mattia Frontini, for this opportunity, and their encouragement, support and guidance throughout my PhD. I am also thankful to Landsteiner Foundation for Blood Transfusion Research and Peterhouse, for funding my PhD, and to all the blood donors that participated in this study.

I would like to thank all the past and present members of my lab for their support and friendship. Thanks to: Ms Isabel Rosa, Ms Tadbir Bariana, Ms Samantha Farrow, Dr Luigi Grassi, Mr Roman Kreuzhuber, Dr Romina Petersen, Dr John Lambourne, Mr Luca Stefanucci, Dr Fizzah Choudry, Dr Stephanie Maiwald, Ms Carly Kempster, Ms Harriet McKinney and Mr Patrick Thomas.

I would also like to acknowledge all the collaborators that have contributed to this study, for their scientific support and availability. Thanks to: Professor Wendy Erber, Professor Gary Hoffman, Professor Nick Morrell, Professor Kathleen Freson, Dr Cedric Ghevaert, Dr Karola Rehnstrom, Dr Jose Guererro, Professor Edwin Chilvers, Professor Antonio Vidal-Puig and their research groups, including Dr Thomas Moreau, Dr Amanda Evans, Dr Holly Foster, Dr Jose Ballester-Beltran, Ms Louisa Mayer, Dr Vian Azzu, Dr Alexi Crosby and Dr Stephen Moore. Thanks to Welcome Trust Sanger Institute for creating *Smim1* mutant mice and all the staff members at CBS for taking care of the mice. Special thanks to: Dr Annett Müller and Dr Cavan Bennett, for their friendship, time, patience and invaluable support in teaching me all about the new world of mice; Dr Michele Vacca for sharing with me his knowledge about liver and for his dedicated and enthusiastic approach to science, which is just beautiful and inspiring; Ms Katie Bashant for doing the mice blood deformability experiments; Mr Gregory Strachan for the long hours helping me with HALO analysis; Ms Frances Burden for her friendship and help with RNA libraries; Dr William Astle for his help and advice with statistics and Dr Denis Seyres for performing RNA-sequencing analysis and for making my life easier by helping me with R.

Many thanks to all the colleagues and friends that I have made during this journey at NHSBT, which all at their one way contribute to this beautiful ride.

Many thanks to all my friends, who are always there for me: Fátima, Joana Mota, Joana Reis, Carolina Martinho, Marina, Filipa, Francisca and Canhão. Thanks too to the Birdlife team, Ricardo, Steffen, Martin, Antonio and Anna for their friendship and for sharing their adventures fighting for conservation with me. Thanks to my dear housemates, Edo and Myles, for making 1 Warkworth Street an even more beautiful home.

I could not thank enough my Cambridge family, who kept me sane through the years and helped me to focus on what matters in life. Thanks for being such inspiring people and for making this journey with me. Obrigada do Fundo do coração a todos vós! To the ladies, Maria, Mariana F., Mariana D., Carolina, Sara, Joana B., Liliana, Chiara and the gentlemen Tiago, João, Simão e David.

Last but not least, a huge and warm thank you, OBRIGADA, to my family, Mãe Judite, Pai Tomé, mano João e Pedro, Filipa e Vasquinho. Thanks for always being supportive and encourage me to continue. Thank you so much for all your Love.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 5 | Characterisation of SMIM1 role in megakaryopoiesis and platelet function

Chapter 8 | Appendix

LIST OF ABBREVIATIONS AND ACRONYMS

2D – two dimensions

3D – three dimensions

ACD – acid citrate dextrose

ACTN1 – actinin 1

ADP – adenosine diphosphate

Afdn – adherens junction formation factor

Akt – protein kinase B

Arrdc3 – arrestin domain containing 3

ATP – adenosine triphosphate

BFU - burst-forming unit

BMI – body mass index

BMP4 – bone morphogenetic protein 4

bp – base pair

BSA – bovine serum albumin

cAMP – cyclic adenosine monophosphate

CAS9n – CRISPR-associated protein 9 nickase

Cdc42 – cell division cycle 42

CFU – colony-forming unit

cGMP – cyclic guanosine monophosphate

CLP – common lymphoid progenitor

CMP – common myeloid progenitor

CRISPR – clustered regularly interspaced short palindromic repeats

CRLF3 – cytokine receptor-like factor 3

CRP-XL – cross-linked collagen related peptide

CXCL4 – C-X-C motif ligand 4

CXCL7 – C-X-C motif ligand 7

D-PBS – Dulbecco's phosphate-buffered saline

DAPI – 4',6-diamidino-2-phenylindole

Dcdc2a – doublecortin domain containing 2

DMEM – Dulbecco's modified Eagle medium

DMS – demarcation membrane system

DMSO – dimethyl sulfoxide

DNA – deoxyribonucleic acid

DTS – dense tubular system

DTT – dithiothreitol

EB – erythroblast

EB3 – end-binding protein 3

EBs – embryoid bodies

EDTA – ethylenediaminetetraacetic acid

EPO – erythropoietin

EPOR – EPO receptor

FACS – Fluorescence activated cell sorting

FBS – foetal bovine serum

FGF – fibroblast growth factor

FLI1 – friend leukaemia integration 1

FOG1 – friend of GATA1

GAPDH – glyceraldehyde 3- phosphate dehydrogenase

GATA – GATA binding protein

gDNA – genomic DNA

GFI1b – growth factor independent 1B transcriptional repressor

GFP – green fluorescent protein

GMP – granulocyte-macrophage progenitor

Gng4 – guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-4

GP – glycoprotein

GPRP – Gly-Pro-Arg-Pro

GSK3-β – glycogen synthase kinase 3 beta

H&E – haematoxylin and eosin

HBS – HEPES buffered saline

HCT – haematocrit

HEPES – 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid

HGB – haemoglobin concentration

Hb – haemoglobin

hiPSC – human induced pluripotent stem cell

HMGB1 – high motility group box-1 protein

HSC – hematopoietic stem cell

HSPCs – hematopoietic stem and progenitor cells

ICAM-1 – intercellular Adhesion Molecule 1

IgG – immunoglobin G

IL – interleukin

IPF – immature platelet fraction

IRF – immature reticulocyte fraction

ITGA2B – integrin subunit alpha 2b

IVC – inferior vena cava

JAK2 – Janus Kinase 2

Kb – kilobases

LB – Luria Bertani medium

LIF – leukaemia-inhibitor factor

Lrrc41 – leucine rich repeat containing 41

MAPK – mitogen-activated protein kinase

MCH – mean cell haemoglobin

MCHC – mean cell haemoglobin concentration

MCV– mean cell volume

MEF – mouse embryonic fibroblast

MEP – megakaryocyte-erythrocyte progenitor

MK – megakaryocyte

MK-p – megakaryocyte progenitor

MOI – multiplicity of infection

MPL – myeloproliferative leukaemia protein

MPV – mean platelet volume

mRNA – messenger ribonucleic acid

MYH9 – myosin heavy chain 9

MYO1C – myosin IC

Myo1g – myosin IG

Nacad – NAC alpha domain containing

NETs – neutrophil extracellular traps

NEUT – neutrophil count

NFE2 – nuclear factor erythroid 2

NHSBT – National Health System Blood and Transplant

NK – natural killer

NO – nitric oxide

OCS – open canalicular system

P-LCR – platelet larger cell ratio

PAR – protease-activated receptor

PAS – periodic acid-Schiff

PBMCs – peripheral blood mononuclear cells

PCR – polymerase chain reaction

PCT – plateletcrit

PDI – protein disulphide isomerase

PDK1 – 3-phosphoinositide-dependent protein kinase 1

PDW – platelet distribution width

PF4 – platelet factor 4

PFA – paraformaldehyde

PGE1 – prostaglandin E1

PGI2 – prostaglandin I2

PI3K – phosphatidylinositide-3-kinase

PKA – protein kinase A

PKC – protein kinase C

PKG – protein kinase G

PLC – phospholipase C

PLT – platelet count

PRP – platelet rich plasma

PSGL-1 – P-selectin glycoprotein ligand-1

PtdIns(3,4,5)P3 – phosphatidylinositol 3,4,5- trisphosphate

Puro – puromycin

PVDF – polyvinylidene fluoride

QC – quality control

qPCR – quantitative polymerase chain reaction

RAGE – receptor for advance glycation endproducts

RBC – red blood cell

RDW-CV – red blood cell distribution width - coefficient of variation

RDW-SD – red blood cell distribution width - standard deviation

RET – reticulocyte count

Rh – rhesus

RhAG – Rh-associated glycoprotein

RhoA – Ras homolog gene family, member A

RNA – ribonucleic acid

RT – room temperature

RT-DC - real-time deformability cytometry

RUNX1 – runt-related transcription factor 1

S1P – sphingosine-1-phosphate

S1pr1 – sphingosine-1-phosphate receptor 1

SCF – stem cell factor

SD – standard deviation

Sfi – SFI1 centrin binding protein

sgRNA – single guide RNA

Slc38a6 – solute carrier family 38 member 6

SMIM1 – small integral membrane protein 1

SNARE– soluble NSF [N-ethylmaleimide-sensitive factor] attachment protein receptors

STAT – signal transducer and activator of transcription

TAL1 – T-cell acute lymphocytic leukaemia protein 1

TAP – tandem affinity purification

TBE – tris-borate-EDTA

TEM – transmission electron microscopy

TEV – tobacco etch virus

TGF-β – transforming growth factor-beta

THBS1 – thrombospondin-1

TLR – toll-like receptor

Tnfaip2 – TNF Alpha Induced Protein 2

TP – thromboxane receptor

TPO – thrombopoietin

TRAP-6 – thrombin-receptor agonist peptide-6

TUBB1 – beta-1-tubulin

TxA_2 – thromboxane A2

VAMP – vesicle-associated membrane protein

VEGF – vascular endothelial growth factor

VTN – vitronectin

VWF – von Willebrand factor

WBC – white blood cell

WT – wildtype

CHAPTER 1 | INTRODUCTION

1.1 An overview of haematopoiesis

Haematopoiesis is the developmental process that gives rise to blood cells. These represent approximately one-quarter of all the human cells and, every second of our life, millions of blood cells are replaced to maintain our haematopoietic system. All these cells are produced from hematopoietic stem cells (HSCs), which have the capacity to self-renew and to generate all the blood lineages through a series of increasingly committed progenitors (Figure 1.1). This process is highly conserved between vertebrates (reviewed in [1]).

In the developing vertebrate embryo, haematopoiesis occurs in two discrete waves, termed primitive and definitive. The primitive wave takes place during early embryonic development in the yolk sac and generates transitory haematopoietic cell populations (HSC independent) of primitive nucleated red blood cells (RBCs, also named erythrocytes), macrophages and megakaryocytes (MKs, which are more proliferative and have lower ploidy level (modal ploidy class at 4n) than adult MKs)[2-6]. Primitive RBCs facilitate tissue oxygenation during embryo development, while macrophages and MKs are important for tissue remodelling [7, 8]. The definitive wave originates later in development when multipotent HSCs capable of self-renewing and generating all the different blood lineages of the adult organism are produced [7, 9]. HSCs are generated between day 27 and day 40 of gestation in the aorta-gonad-mesonephros in an evolutionarily conserved process [9, 10]. These cells sequentially colonize the foetal liver, the thymus and finally the bone marrow, where adult HSCs live and give rise to all blood lineages [9, 10]. Additionally, in most organisms, between these two waves, there is a transient wave of definitive haematopoiesis, called the pro-definitive wave. During this wave, erythroid-myeloid progenitors emerge in the yolk sac and then migrate to the foetal liver [11]. Also, a subset of B-lymphocytes (B-1a cells) and T-lymphocytes have been detected during this wave [8, 12].

In the classical model of haematopoiesis, adult HSCs go through a multipotent progenitor stage that can commit into common lymphoid progenitor (CLP) and common myeloid progenitor (CMP). CLPs undergo differentiation into B-lymphocytes, T-lymphocytes and natural killer (NK) cells. While committed CMPs can be differentiated into granulocyte-macrophage progenitors (GMPs) or megakaryocyte-erythrocyte progenitors (MEPs) (Figure 1.1). Following a series of specification steps, GMPs give

rise to neutrophils, eosinophils, basophils and monocytes, whereas MEPs are differentiated into progenitor cells, named erythroblasts (EBs) and megakaryoblasts, which are then matured into RBCs and MKs, respectively. Mature MKs will later give rise to platelets [13]. The differentiation of CMPs towards MEPs or GMPs is regulated by two major transcription factors, GATA1 and PU.1. The increased expression of GATA1 together with the decreased expression of PU.1 lead to the specification of CMPs towards MEPs [14]. Since the megakaryocyte and the erythrocyte lineages derive from a common precursor (MEP), they share several transcription factors that are critical for their development, such as GATA1, FOG1 (GATA1 cofactor), TAL1 and GFI1b [15, 16].

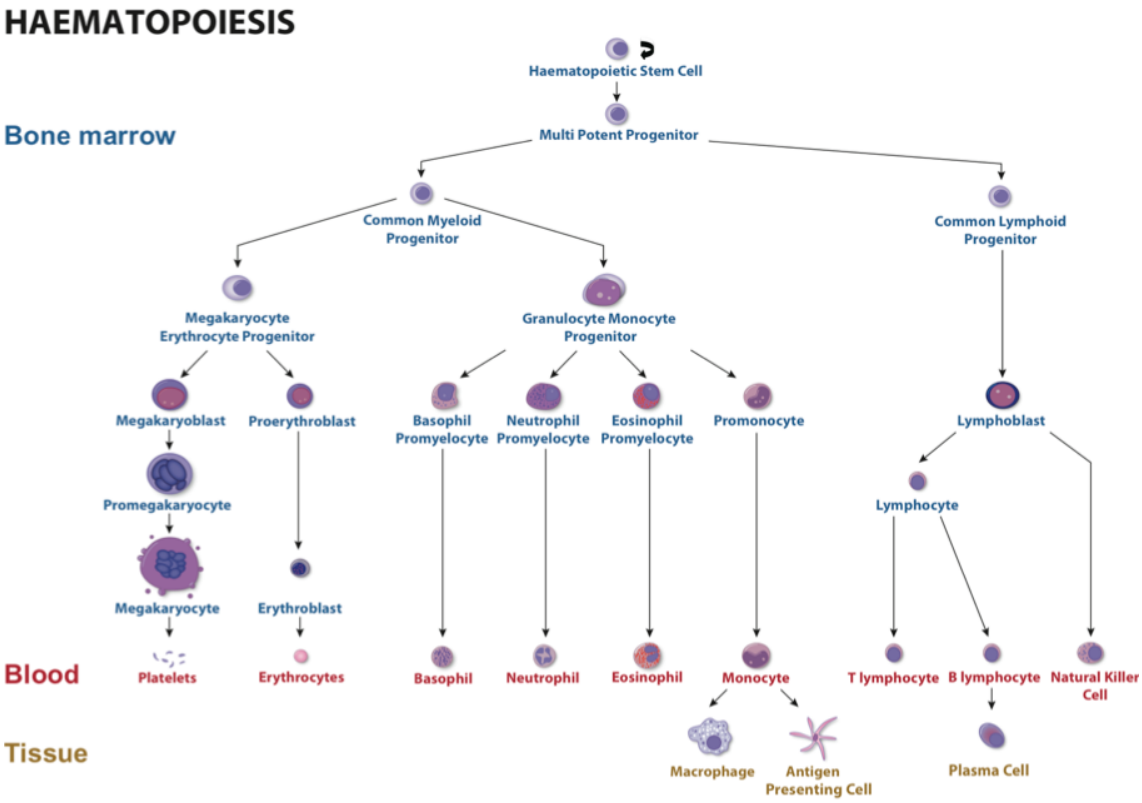


Figure 1.1 – The classic model of haematopoiesis. Schematic representation of the blood cells progeny (Figure from Prof. Owehand's group).

Several more recent studies have shown that the HSC population is highly heterogeneous. This heterogeneity results in a number of HSC subtypes that have different functional properties, including different self-renewal abilities [17, 18], cell cycling status [19] and multilineage differentiation output [20-22]. These studies have led to the formulation of several alternatives and more complex hierarchical models of haematopoiesis [23-25]. However, Velten *et al*. have recently proposed a new model

that challenges the classic hierarchical model, suggesting that HSC differentiation instead, occurs through a mechanism of continuous lineage priming. This model suggests that CLOUD-HSPCs, which is a pool of uncommitted hematopoietic stem and progenitor cells (HSPCs) in transitory states, progressively acquire transcriptomic unilineage priming and directly differentiate in unilineage progenitors and, ultimately, into distinct mature blood lineages [26].

1.2 Megakaryopoiesis

Megakaryopoiesis is the process that leads to the production of MKs, whose function is the production of platelets (section 1.3). MKs are amongst the rarest and largest nucleated cells in the bone marrow. They represent 0.01% of the total blood nucleated cells and have 50-100 μm diameter in the human and 20-30 μm diameter in the mouse [27-29].

Megakaryopoiesis is a complex process involving the commitment of MK lineage, MK progenitor (MK-p) proliferation and, finally, MK maturation (Figure 1.2 – A). As described above MK-p are generated through the commitment of HSCs to CMP and then to MEP [30]. However, recent studies demonstrated that MKs can also be generated by an alternative unipotent pathway, in which MK-biased HSCs directly differentiate into MK-p [31]. The physiological role of this alternative pathway is not fully understood but it is thought that it might play an important physiological role during platelet acute situations [31]. Subsequently, MK-p undergoes a maturation process including polyploidisation and cytoplasmic maturation and expansion, which results in the production of 1,000-5,000 platelets per MK (reviewed in [32]). During this process, MKs undergo endomitosis, initiating multiple cycles of DNA replication without completing cytokinesis [33], resulting in a geometric increase of ploidy that ranges up to 128n. The production of platelets can occur at any ploidy level; however, increased platelets production is correlated with higher ploidy levels. It has been demonstrated that steady-state platelet production in the bone marrow corresponds to 16n ploidy [3, 34]. Polyploidisation results in gene amplification, which is thought to be essential to increase protein and lipid synthesis and to the cell enlargement required for platelets formation and function [35]. At the same time, cytoplasm maturation and expansion take place, with the acquisition of all the necessary components for platelet formation and function. During the final maturation stages, an increase in non-specific

structures and organelles has been observed, such as endoplasmic reticulum, Golgi bodies, mitochondria, vesicles, microtubules and actin filaments. As well as MK/platelet-specific organelles, such as alpha- (α-) granules and dense- (δ-) granules in the cytosol (reviewed in [32]). Additionally, an increase of heterochromatin has been observed (reviewed in [32]) and the dispersion throughout the cytoplasm of the demarcation membrane system (DMS) [36]. During MKs terminal differentiation at the bone marrow sinusoids, mature MKs form cytoplasmic projections, named proplatelets, that are extended into the bloodstream and subsequently fragmented into platelets (this process is called thrombopoiesis, and it is described in section 1.3, Figure 1.2 - A-B). Although megakaryopoiesis occurs mainly in the bone marrow, it has been suggested that MKs also circulate to other locations, such as lungs and spleen [37, 38].

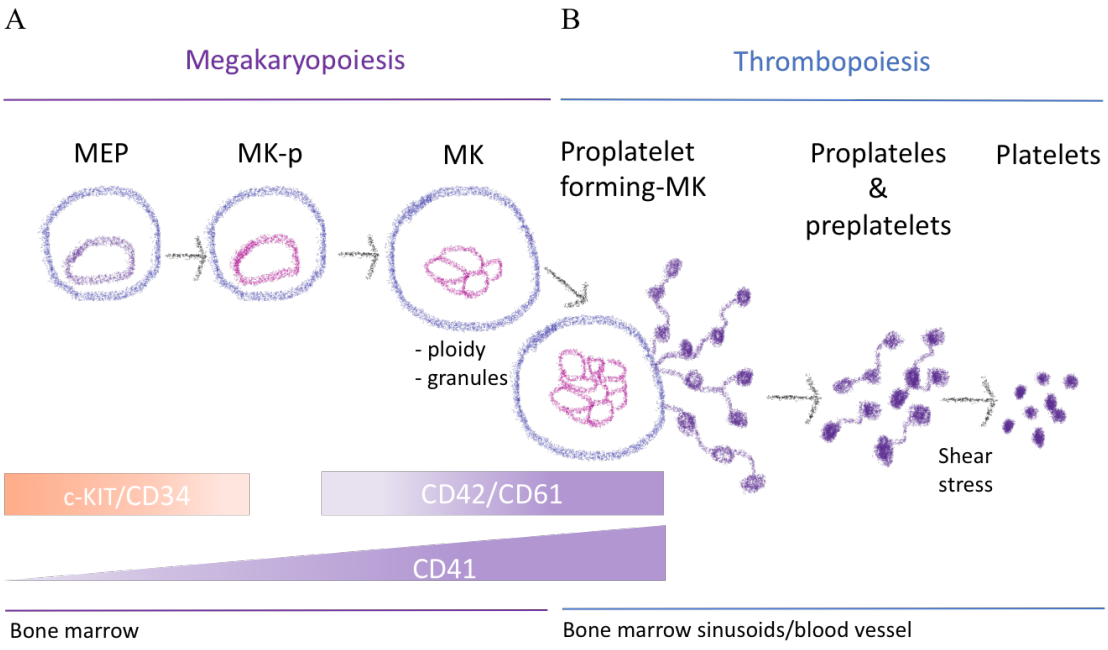


Figure 1.2 - Megakaryopoiesis and thrombopoiesis. Schematic representation of megakaryopoiesis and thrombopoiesis. **A** – Schematic representation of megakaryopoiesis. The differentiation of MEPs into mature MKs is followed by a decrease in surface expression of c-KIT and CD34, and an increase in surface expression of MK-specific markers (CD41, CD61 and CD42), which is accompanied by the rise of ploidy and synthesis of granular content. **B** – Schematic representation of thrombopoiesis. Fully mature MKs form and release proplatelets into the bone marrow sinusoids. Preplatelets, which are smaller fragments originated by proplatelet fragment fission, are further fragmented by shear stress to form mature platelets. MEP – megakaryocyte-erythrocyte-progenitor; MK – megakaryocyte; MK-p – MK progenitor.

1.2.1 Regulation of megakaryopoiesis

Megakaryopoiesis is regulated by several transcription factors such as GATA1, RUNX1, TAL1, NFE2 and FLI1, and by the bone marrow niche, including, chemokines and cytokines, such as interleukin-3 (IL-3), IL-6, IL-11, leukaemia-inhibitor factor (LIF), stem-cell factor (SCF) erythropoietin (EPO) and thrombopoietin (TPO) [39-51]. TPO is the primary growth factor of the megakaryocytic lineage, and is required for proliferation and differentiation of MKs [39]. TPO also induces transcription factors (GATA1 and NFE2) to drive MK development and thrombopoiesis [52]. This factor is produced in the liver and binds to the myeloproliferative leukaemia protein (MPL; c-Mpl), a receptor on MK and platelet membranes [53, 54]. Binding of TPO to c-Mpl activates it and induces its conformational change that initiates a signal transduction via activation of Janus Kinase 2 (JAK2). JAK2 phosphorylates the receptor's tyrosine residues, activating downstream signal pathways and signal transducers and activators of transcription such as mitogen-activated protein kinases (MAPKs), phosphoinositol-3-kinase (PI3K) and signal transducer and activator of transcription (STATs), important for cell survival and proliferation [46]. TPO regulates MKs development through a concentration-dependent compensatory response: increased levels of free TPO stimulate MKs maturation and proplatelet formation [55]. c-Mpl or TPO-deficient mice, however, still produce platelets, which suggests a compensatory mechanism from other regulators [54, 56, 57]. In fact, Nishimura *et al.* have demonstrated that platelet release from MKs could occur by an alternative mechanism, triggered by IL-1α, which leads to higher platelet number. It has been proposed that this mechanism occurs during acute platelet requirement, which explains how MKs can maintain platelet equilibrium and quickly restore platelet numbers under acute needs [58]. Furthermore, it has been shown that TPO is also essential for the regulation of HSC production and functions [59]. Notably, two studies demonstrated that MKs regulate HSCs quiescence through a C-X-C motif ligand 4 (CXCL4) or through the transforming growth factor-beta (TGF-β) [60], and HSCs proliferation through fibroblast growth factor 1 (FGF1)) [61].

1.2.2 Megakaryocytes: cell-surface markers

During megakaryopoiesis, several surface receptors and other surface proteins important for MK development and platelet formation/function are produced. Early in MK lineage commitment and progenitor maturation, a down-regulation of CD34 (cell-cell adhesion factor expressed on haematopoietic progenitor cells), c-KIT (SCF receptor), CD150

(glycoprotein expressed in HCSs), Tie-2 (angiopoietin receptor) and CD45 (receptor-linked protein tyrosine phosphatase) markers is observed. Simultaneously, a progressive up-regulation of MK-specific cell surface markers is observed; such as c-Mpl (TPO receptor), glycoprotein IIb/IIIa (GPIIb/IIIa – also known αIIbβ3 integrin or CD41a/CD61 complex, a fibrinogen receptor crucial for MK migration and adhesion) and GPIb-V-IX complex (GPIbα, GPIbβ, GPIX and GPV also known as CD42a, b, c and d, which is a von Willebrand factor (VWF) receptor important for platelet function). The expression of CD42 is initiated at a later differentiation step and therefore is used a final maturation marker. CXCR4, the receptor of stromal derived factor-1 (SDF-1), is expressed during all stages of differentiation [32, 46, 62].

1.3 Thrombopoiesis

Thrombopoiesis is the name of the process that gives rise to platelets, which play a fundamental role in maintaining haemostasis and thrombosis but also have roles in angiogenesis [63], inflammation [64], infection [65], tissue repair and regeneration [66], wound healing and in cancer biology [67-69]. Platelets are small, discoid, anucleate cells with 1.5-3.0 μm of diameter in human and 0.5 μm of diameter in the mouse. They are the second most abundant blood cell type with normal counts around 150-400 x 10^9 platelets per litre of blood in humans and 1000-1500 x 10^9 platelets per litre of blood in mice [29, 70]. Platelets have a short lifespan, they circulate in the blood for 7 to 10 days in humans and 4 to 5 days in mice, after which they are removed in the spleen or liver [27, 29, 71, 72]. Platelet count in an adult human is maintained by the daily production of approximately 10^{11} platelets in steady-state though platelet production can increase more than 10-fold under states of stress such as traumatic blood loss [73].

Platelet formation has been studied for many years and several theories on how this process works have been proposed, which includes i) platelet budding from the MK surface, ii) cytoplasmic fragmentation of MK via the DMS/"platelet fields" and iii) proplatelet formation [74]. The latter is currently the most widely accepted theory, which postulates that at the bone marrow sinusoids mature MKs undergo drastic cytoskeleton reorganisation to produce long, branching pseudopods cytoplasmic projections, called proplatelets. Each MK forms 10-20 proplatelets, which virtually carry all the intracellular content of MKs (cytoplasm and organelles, including pre-packaged granules) [27]. After proplatelet formation, the bare multilobed nucleus of the

MK is eventually extruded and degraded [75]. Proplatelets are formed by multiple platelet-sized swellings (which contain the organelles and granules essential for their haemostatic function) connected by thin cytoplasmic bridges [76]. After proplatelets are released in the peripheral blood they are fragmented to produce mature platelets. The fission of these large proplatelet fragments results in smaller fragments, named preplatelets, which undergo further fission to form mature platelets. It is thought that this process is facilitated by blood flow–induced shear stress [77, 78] (Figure 1.2 - B). This is supported by several observations such as: i) in the lung capillary bed, the proplatelet counts are higher in pre pulmonary vessels than in post pulmonary vessels [79], while platelet counts are higher in the latter and ii) *in vitro* platelet formation is accelerated by shear stress [80, 81].

The understanding and characterisation of platelet formation has greatly improved with the study of inherited disorders related to impaired platelet formation, as well as with the advance of methodologies to study platelet biogenesis, such as the development of *in vitro* megakaryocyte culture systems and, more recently, *in vivo* imaging of bone marrow using intravital two-photon microscopy [82-85]. Nevertheless, the signals that trigger proplatelet formation in MKs, as well as the final stages of proplatelet maturation and platelet release, remain poorly understood.

Proplatelet formation and platelet release are mediated by the drastic reorganisation of the cytoskeletal proteins, including microtubules, actin and spectrin [75, 86]. Platelet size is also limited by microtubule bundling, elastic bending, and actin-myosin-spectrin cortex forces [70]. Studies of impaired proplatelet formation in congenital thrombocytopenia have identified several genes as regulators of this process, such as *MYH9* [87, 88], *ACTN1* [89], *TUBB1* [90, 91], *RUNX1*[92] (for review [83, 84]). Recently, the polarization of MK and the proplatelet release have been shown to be regulated by the crosstalk of GPIbα receptor with the two small GTPases, cell division cycle 42 (Cdc42) and RhoA. The functional mutation of Cdc42 or GPIbα receptor results in impairment of proplatelet formation, while in contrast the ablation of RhoA results in increased Cdc42 activity and transmigration of MKs into the sinusoids lumen [93]. Moreover, it has been shown that a fine balance of GPIbα receptor and filamin A expression is important to regulate platelet size. Ablation of filamin A in MKs leads to a

premature release of large platelets, which is associated with a decrease of GPIbα receptor expression [94].

It has been shown that platelet formation is also regulated by the microenvironment in the bone marrow: the osteoblastic and the vascular niches. The osteoblastic niche plays a role in MK differentiation and maturation, and suppresses proplatelet formation by the binding of collagen to the GPIa/IIa (also known as α2β1 integrin) [95, 96]. The migration of mature MKs from the osteoblastic niche to the vascular niche, has been suggested to be promoted by a gradient of the chemokine SDF-1 and by its receptor CXCR4 on MKs membrane [97, 98]. A gradient of transendothelial sphingosine 1-phosphate (S1P) has been also suggested to guide the directional formation of proplatelets into the bone marrow sinusoids [99]. However, a recent study suggests that, contrary to what was previously thought, there is a vessel-biased MK pool within the bone marrow and that these MKs have a reduced capacity of migration. Therefore, it has been proposed that thrombopoiesis is mainly regulated by the bone marrow vasculature [100].

In the vascular niche, the interaction of MKs with extracellular proteins such as fibrinogen [101], fibronectin [102] and VWF [103], seems to promote the later stage of MK maturation and proplatelet formation. Moreover, high S1P concentrations in the blood, as well as its receptor S1pr1, seem to be important for platelet release. *S1pr1* knockout mice have a drastic reduction of proplatelet formation and release. It has been further shown that activation of S1pr1 directly stimulated the release of new platelets through activation of the Gi/Rac GTPase signalling cascade [99].

Several observations (such as the ones described above - observation of circulating MKs and platelets release in the lung) suggest that apart from the bone marrow sinusoids, the lung capillary bed is also a primary site of platelet release. Recently, mice MKs were shown to release proplatelets in the lung, and it was suggested that as much as 50% of platelet formation in mice could occur in the lungs [104-108].

1.3.1 Platelet structure

As described above, platelets acquire most of their RNA, proteins and numerous organelles including α-granules, δ-granules, peroxisomes, lysosomes, and mitochondria

from the MKs, but while platelets mature, they also endocytose some proteins from the environment, such as fibrinogen and vascular endothelial growth factor (VEGF) [109-111]. Furthermore, during platelet formation, platelets develop defining features such as: i) a delimited plasma membrane; ii) a dynamic and bipolar marginal coiled microtubule band, which is responsible for its discoid shape at resting state; iii) an open canalicular system (OCS), which is formed by invaginations of the surface membrane (the OCS is a reservoir of membrane for platelet spreading and enables the uptake and release of granule contents) [112]; iv) a dense tubular system (DTS), which is an endoplasmic reticulum membrane system that is important for the regulation of platelet activation; v) a spectrin-based membrane skeleton and an actin-based cytoskeletal network, both also important for platelet shape change and spreading during platelet activation [113].

The resting platelet can be divided into three zones: i) the peripheral zone, which regulates platelet adhesion and aggregation. This zone includes the plasma membrane that expresses receptors important for platelet activation and signalling; a glycoprotein coat that is concentrated in the inner part of the membrane at rest and cytoskeleton; ii) the sol-gel zone, which facilitates contraction and supports the microtubule network. This zone comprises the OCS and the DTS; and iii) the organelle zone, which contains all the platelet organelles and molecules present in the cytoplasm, such as calcium, platelet factor 4 (PF4), glycogen granules, α-granules and lysosomal granules [113, 114].

Platelet secretory granules

Platelets contain four types of secretory granules: α-granules, δ-granules, lysosomes and T-granules, which play critical roles in platelet function.

i) α- granules

α- granules are the most abundant, with 50 to 80 α-granules per platelet, and also the largest, with a diameter ranging from 200 to 500 nm. α-granules contain hundreds of proteins including adhesive and healing proteins that mediate different processes, such as: i) haemostasis and coagulation (e.g. VWF, fibrinogen, fibronectin, thrombospondin-1 (THBS1) and prothrombin (inactive precursor of thrombin) factor V, XI and XIII); ii) inflammation (e.g. PF4, β-thromboglobulin (CXCL7) and IL-8 [115, 116]) and iii) angiogenesis (e.g. VEFG (pro-angiogenic regulators) and endostatin (anti-angiogenic

regulator)). In their membrane they express several proteins that are also expressed on the platelet membrane such as GPIb-V-IX, GPVI, GPIIb/IIIa and P-selectin [117, 118]. The formation of α-granules is initiated in MKs, but its maturation continues in platelets [119]. During platelet activation α-granules are secreted via a complex process, which involves SNARE (soluble NSF [N-ethylmaleimide-sensitive factor] attachment protein receptors), SNARE-associated proteins, and membrane lipids. In this process their membranes fuse with the platelet membrane, enabling the release of their constituents to the extracellular environment and the expression of other constituents such as P-selectin on the platelet membrane [117, 120]. It has been shown that platelets contain distinct subpopulations of α-granules that undergo differential secretion during platelet activation [121]. Nevertheless, both the α-granule biogenesis and its selective secretion is not yet fully understood.

ii) δ-granules

δ-granules are the second most abundant secretory granule, with three to eight δ-granules per platelet, and have a diameter of approximately 150 nm. Their name results from the characteristic dense/dark appearance on electron microscopy. δ-granules contain mainly small molecules such as ADP, ATP, serotonin and calcium that are released upon platelet activation and are important for homeostasis [117, 120]. Moreover, δ-granules also contain lysosomal membrane proteins such as CD63 (LAMP-3) and LAMP-1/2, and membrane proteins that are mainly expressed in α-granules, such as P-selectin and GPIb [122]. Upon platelet activation, these membrane proteins are translocated to the platelet membrane. The secretion mechanism of δ-granules is comparable to those in α-granules, but differences have been reported [123]. The biogenesis of δ-granules is initiated in MKs but only completed in platelets, although the underlying mechanism is also not fully understood [117, 120].

iii) Lysosomes

Lysosomes are less abundant, with a maximum of three per platelet, and with a diameter ranging from 200 to 250 nm. They are not platelet-specific and contain hydrolytic enzymes, such as proteases and glycosidases, which digest cellular components. Like δ-granules they are characterised by CD63 and LAMP-1/2 on their membrane. Lysosome function in platelets is not fully characterised. It is thought that they are released in a similar fashion to the other granules but during greater platelet activation. It has been suggested that they play a role in the regulation of thrombus formation and in remodelling the extracellular matrix [117, 120].

iv) T-granules

T-granules were recently described. They have a tubular morphology and contain toll-like receptor-9 (TLR9), vesicle-associated membrane protein-7/8 (VAMP-7/8) and protein disulphide isomerase (PDI) [124]. PDI has been shown to be released from activated platelets and plays a role in thrombus formation [125].

1.3.2 Platelet function

The primary function of platelets is maintaining homeostasis and thrombus formation at vascular injury sites. Circulating platelets are in a resting state, which is continuously mediated by the vascular endothelium through several mechanisms, which include: i) a physical barrier of endothelial cells in the blood vessels; ii) the endothelial surface expression of ectonucleotidases (which degrade ATP and ADP) [126] and thrombomodulin (which inactivates thrombin) [127]; and iii) the release of inhibitory factors from endothelial cell such as nitric oxide (NO, which increases intracellular levels of cyclic GMP (cGMP) and consequently activates PKG) and prostaglandin I_2 (PGI_2, which binds to platelet IP receptor and leads to an increase of intracellular cyclic adenosine monophosphate (cAMP) levels and consequently activates PKA). The release of both NO and PGI_2 increases the threshold for platelet activation by the mediated mechanisms described above [128-130]. The exposure of the subendothelial extracellular matrix at the site of vascular injury leads to platelet activation and thrombus formation, through a series of sequential steps. These steps can be divided into: platelet adhesion, platelet activation and aggregation, and thrombus stabilization [114, 131, 132].

<u>Platelet adhesion</u>

Platelet adhesion is initially mediated via the binding of GPIb platelet receptor (which is part of GPIb-V-IX complex) to activated VWF immobilized on exposed collagen at high shear rates. This initial and loose adhesion is important to slow down platelet velocity, which subsequently enables stable platelet adhesion [133]. At low shear rates a stable adhesion is mediated via the direct and synergistic binding of GPVI and activated GPIa/IIa platelet receptors to collagen [131, 134, 135]. This stable adhesion leads to platelet activation.

Platelet activation

Collagen binding to GPIV and GPIa/IIa platelet receptors stimulates the activation of multiple intracellular signalling pathways, which are regulated by important signalling molecules such as phospholipase C (PLC), protein kinase C (PKC), PI3K. The activation of this intricate network of positive and negative regulatory signals, which involves signalling adapters and regulators, known as platelet signalosome, leads to an increase of the intracellular level of calcium, and underpins drastic platelet shape change, spreading and secretion [136].

The change of platelet shape from discoid to spiny spheres, as well as their spreading, results from the reorganisation of the microtubule band and by an increase of microtubule polymerisation, which produces filopodia as the platelet spreads [137]. This reorganisation of the cytoskeleton is calcium-dependent and mediated by the activation of myosin light chain kinase via small GTPase-Rho [136]. It has been also shown that platelet spreading is mediated by distinct isoforms of PKC proteins which have differential effects on platelet spreading [138].

Upon activation, platelets undergo degranulation and secretion/release of their granule's constituents and other secondary mediators, such as thromboxane A_2 (TxA_2). This enables the recruitment and stimulation of new platelets and leads to subsequent activation of the coagulation pathway at the platelet surface, which is critical for clot growth and stabilization [139].

As described above, platelet granules have a high concentration of factors that are important to mediate/amplify platelet activation, such as ADP and prothrombin. Also, TxA_2 is a strong platelet activator, which is synthesised upon platelet activation through the conversion of arachidonic acid by several enzymes including cyclooxygenase-1 (which is the target of aspirin). These secondary mediators act by binding to G-protein-coupled receptors, which activates PLCβ and in turn amplifies platelet activation [136]. ADP binds to the transmembrane domain of $P2Y_1$ and $P2Y_{12}$ purinergic platelet receptors, while TxA_2 binds to thromboxane receptor (TP) α and TPβ [140].

Platelet activation also stimulates the exposure of phosphatidylserine on the platelet plasma membrane. This enables the assembly of the prothrombinase complex (part of

the coagulation pathway) on the platelet surface, which catalyses the transformation of prothrombin into thrombin [141]. Thrombin is a potent platelet activator that can stimulate and amplify thrombus formation via its binding to protease-activated receptor (PAR) 1 and PAR4 platelet receptors in humans and PAR3 and PAR4 in mouse [142, 143]. Furthermore, thrombin production can also be mediated via tissue factor. Tissue factor is a membrane protein constitutively expressed on several cells of the vessel wall. At the injury site tissue factor gets exposed and forms a complex with the activated coagulation factor VII present in the bloodstream. The formation of this complex activates factor IX, which initiates a coagulation cascade involving several blood coagulation factors that leads to thrombin formation and consequently fibrin formation [144].

Platelet aggregation

During platelet activation, the GPIIb/IIIa platelet receptor undergoes conformational changes that mediate its activation, which in turn supports platelet aggregation. The activated GPIIb/IIIa receptor binds to adhesive proteins such as fibrinogen and VWF, which form bridges between two GPIIb/IIIa receptors on adjacent platelets. The formation of these bridges allows platelets to aggregate and form a haemostatic plug [145, 146].

Thrombus growth and stabilisation

Platelet aggregates and their secreted/produced secondary mediators, together with the vascular endothelium, other blood cells and coagulation factors, promote the growth and the stabilisation of the thrombus through signalling events and formation of a fibrin mesh [131].

The stabilisation of the growing thrombus is mainly mediated by the formation of a fibrin mesh around platelets. This fibrin mesh is formed by the accumulation of fibrin which is produced through the cleavage of circulating and platelet-secreted fibrinogen by thrombin [144]. Platelet-neutrophil interactions mediate another important mechanism for thrombus growth and stabilisation. Upon platelet activation, P-selectin gets exposed on platelets membrane and mediates platelet-neutrophil interaction via its binding to P-selectin glycoprotein ligand-1 (PSGL-1) on neutrophils [147]. In turn, this interaction facilitates neutrophil activation through the binding of the transmembrane

receptor RAGE on neutrophils to the high motility group box-1 protein (HMGB-1) exposed on platelet membrane [148]. Neutrophil activation is followed by a complex mechanism, named NETosis, which culminate with the formation of neutrophil extracellular traps (NETs). NETs are networks of extracellular chromatin fibres surround by histones and bound to granular and cytoplasmic proteins. It has been suggested that these NETs mediate platelet activation, directly through their histones and by the direct stimulation of the coagulation cascade [149-152] (see [153] for review). Hence, platelet-neutrophil interaction enhances thrombus growth and stabilisation. Furthermore, thrombus growth and stabilisation are controlled by a negative feedback loop, which involves compensatory molecules and signalling events that prevent undesirable platelet and fibrin accumulation, and thus pathological thrombosis [131].

1.4 Erythropoiesis

Erythropoiesis is the process that leads to the production of RBCs, whose primary function is tissue oxygenation. RBCs are anucleate, circular, biconcave and small, with a diameter of 6-8 μm [154]. The lack of nucleus allows an increased storage of haemoglobin (Hb) and, consequently, more oxygen transported. The biconcave shape increases the surface area for the diffusion of oxygen, and their small size and flexibility (deformability capability) enables the RBCs to pass through small blood vessels. RBC shape and deformability are dependent on the two-dimensional network of spectrin (cytoskeleton) to which their phospholipid bilayer membrane is bound [155]. RBCs are the most abundant cells in the blood (~5 million RBCs per microliter of blood) and remain in the bloodstream in average 120 days until they are removed by macrophages in the spleen and liver [156]. The number of RBCs remains relatively constant through the balance between their continuous loss and production and it has been estimated that approximately 2.4 million RBCs are produced per second [154].

As mentioned above in section 1.1, during development there are primitive and definitive erythropoiesis. Although it is still debatable whether both processes have completely separate origins, it is well established that both give rise to erythropoietic cells which can be distinguished by their morphology, surface markers, cytokine responsiveness, growth kinetics, gene expression and Hb [157]. Three different types of Hb have been described: embryonic (Hb-Grower I, Hb-Grower II and Hb-Portland;

from the primitive wave), foetal (HbF; from the definitive wave) and adult (HbA and HbA_2; from the definitive wave) [158, 159]. Here, I will briefly describe definitive erythropoiesis.

Erythropoiesis occurs mainly in red bone marrow. Under erythroid stress conditions, however, extramedullary erythropoiesis can happen in the spleen [160]. RBCs, like MKs, are generated through a similar series of commitment steps, HSC, CMP, and MEP. Then, erythroid progenitors undergo proliferation and differentiation through a series of stages until they become mature RBCs [154]. The different stages of erythroid lineage differentiation are shown in Figure 1.3. The presence of two progenitor types characterises early stages of erythropoiesis: initially, the burst-forming unit (BDU-e) and, later, the colony-forming unit (CFU-e). During later stages, from the differentiation of pro-erythroblast to reticulocyte, there is an increase of membrane and cytoskeletal proteins production (including Band3, Duffy, GPA and ankyrin-1) and the cells experience a reduction in size, an increase of cytoplasm, and a reduction of the nucleus, which is then expelled, forming reticulocytes. Reticulocytes contain residual RNA and mitochondria, their maturation is initiated in the bone marrow but as they mature, they lose adhesion receptors and become more deformable, which enables their release from the bone marrow and subsequently final maturation into RBCs in the bloodstream [154, 157, 161].

Figure of erythropoiesis removed for copyright reasons. Copyright holder is Cold Spring Harbor Perspectives in Medicine.

Figure 1.3 - Erythropoiesis. Expression of most commonly used cell surface markers to identify the various stages is indicated by the bars. Cells at the CFU-E and pro-erythroblast stage are the most sensitive to, and dependent on, the presence of EPO. Gray - low expression; black - high expression; HSC - hematopoietic stem cell; CMP - common myeloid progenitor; MEP - megakaryocyte-erythroid progenitor; BFU-e - burst-forming unit, erythroid; CFU-e - colony-forming unit, erythroid (Figure adapted from [154]).

1.4.1 Regulation of erythropoiesis

Erythropoiesis is dependent on transcription factors (e.g. GATA2, GATA1, FOG1, GPI1b and EKLF) [162-164], growth factors, hormones and cytokines, which are mainly induced by the microenvironment of stromal and immunoregulatory cells (fibroblasts, fat cells, endothelial cells, smooth muscle cells, monocytes, macrophages and lymphocytes) [157]. The early progenitors are responsive to cytokines, such as TPO, IL-3, IL-11 and SCF [154]. In later stage erythroid progenitors, SCF acts synergistically with EPO [165]. EPO is secreted by renal Epo-producing cells (REP cells) [166] and has been described as the most critical cytokine for erythropoiesis [167-169], regulating proliferation [170], differentiation [171] and inhibition of apoptosis [172] during this process. In response to low levels of oxygen, kidneys release EPO, which triggers erythropoiesis by binding to EPO receptor (EPOR) and by signalling to committed erythroid progenitors [173]. Binding of EPO to its receptor induces conformational changes of the receptor and subsequently activation of JAK2 and downstream activation of signal pathways and signal transducers and activators of transcription [157, 174, 175]. EPO expression is regulated by hypoxia-inducible factors (HIFs), which are the key transcription factors for oxygen-dependent gene regulation [176].

1.4.2 Erythroid cell: cell-surface markers

Erythroid progenitors and precursors express distinct cell-surface proteins, which reflect their different signalling programmes. The common cell-surface markers used to distinguish the different erythroid precursors are: CD34 (adhesion factor expressed early in haematopoiesis) and c-KIT (SCF receptor), which are present until the formation of BFU-Es; EPOR (EPO receptor), which is expressed in low levels in BFU-Es and in high levels in CFU-Es and pro-erythroblast; TfR (transferrin receptor, CD71; responsible for iron uptake), which is highly expressed in cells that are actively synthesizing Hb and then its expression levels decrease during terminal differentiation, being absent in mature RBCs; adhesion molecules, intercellular adhesion molecule 1 (ICAM-1, a member of the immunoglobulin superfamily) and α4β1 integrin (CD49d/CD29) that interacts with fibronectin, are highly expressed in the early precursors and, then, lost as maturation occurs; toward the later stages of erythroid differentiation several important surface proteins are expressed such as GPA (CD235a, which is used as a surface marker), rhesus (Rh) protein and Rh-associated glycoprotein (RhAG) [157, 161, 177].

Most of the proteins on the RBC membrane are glycosylated and belong to a blood group system.

1.5 Blood group systems

Until the beginning of the last century, it was thought that blood was identical among individuals. However, with haemolytic disease of newborns as well as frequent severe haemolytic reactions and fatal transfusions, it was possible to determine the presence of different antigens on RBCs and later establish different blood groups [178]. A blood group system is described by the International Society of Blood Transfusion, as one or more antigens encoded at a single gene locus, or by two or more very closely linked homologous genes with small or no observable recombination between them [179]. Although the ABO and Rhesus-D are the most critical and characterised blood systems, so far more than 300 antigens in RBCs have been described, and most of these are clustered in 36 blood group systems by the International Society of Blood transfusion [180] (http://goo.gl/56j3Np). One of these antigens is the Vel-antigen, which was recently defined as the Vel-blood group system.

1.5.1 Vel-blood group system

The Vel-antigen was first recognised in 1952 when Sussman and Miller reported a Vel-negative blood type phenotype in a patient, named Mrs Vel [181, 182]. Mrs Vel suffered a severe haemolytic reaction after a transfusion with Vel-positive blood. It was subsequently shown that Mrs Vel had an antibody in her serum that agglutinated the blood cells of all but four of 10,000 individuals [181, 182].

It has been shown that Vel-negative blood type is inherited as an autosomal recessive trait [183, 184]. Overall, the frequency of Vel-negative blood type, as estimated in several studies and different populations, is approximately 0.04%, even though regional differences exist [184-187]. Estimated frequency in Europe is 0.025%, while in northern Scandinavia is 0.06% [184-187]. This low frequency makes Vel-negative a rare blood group system and therefore of high clinical relevance. Maintaining Vel-negative blood stocks is challenging due to the reduced number of Vel-negative donors, but it is absolutely required for safe transfusion of individual carrying this rare blood group. The locus encoding the Vel antigen and the genetic variation responsible for the Vel-negative phenotype has remained unknown for more than six decades. Recently, using

different approaches, three research groups were able to identify the gene that underlies the Vel antigen, *Small integral membrane protein 1* (*SMIM1),* and to determine the polymorphism responsible for the Vel-negative phenotype [182, 186, 188].

1.5.1.1 *SMIM1* underlies the Vel blood group

SMIM1 is a small gene composed of four exons and is located in chromosome 1p36.32 in humans. This gene encodes three protein-coding transcripts. Two small transcripts of 5 and 37 amino acids and a longer transcript of 78 amino acids encodes a small integral membrane protein 1 (SMIM1, Figure 1.4).

According to the most recent data available at the Human Gene Organization (HUGO - https://www.genenames.org/) and Gene Cards® – Human Gene Database (https://www.genecards.org/), SMIM1 belongs to a collection of 44 unrelated SMIMs that have a single predicted transmembrane domain and low theoretical molecular weight (ranging from 5.8 kDa up to 20 kDa). Most of these proteins are of unknown function (Table 8.1 – Chapter 8). Moreover, SMIM5 and SMIM7, are highly expressed at the mRNA level in platelets (Figure 8.1 – Chapter 8).

SMIM1 coding exons are highly conserved among several vertebrate species as shown in the alignment in Figure 1.4 [186, 188]. In mice, *Smim1* gene is located in chromosome 4 and has 13 protein-coding transcripts (https://goo.gl/8kvnAs). The mouse *Smim1* longer transcript encodes a protein with 78 amino acids, which has 81% homology with the human longer isoform.

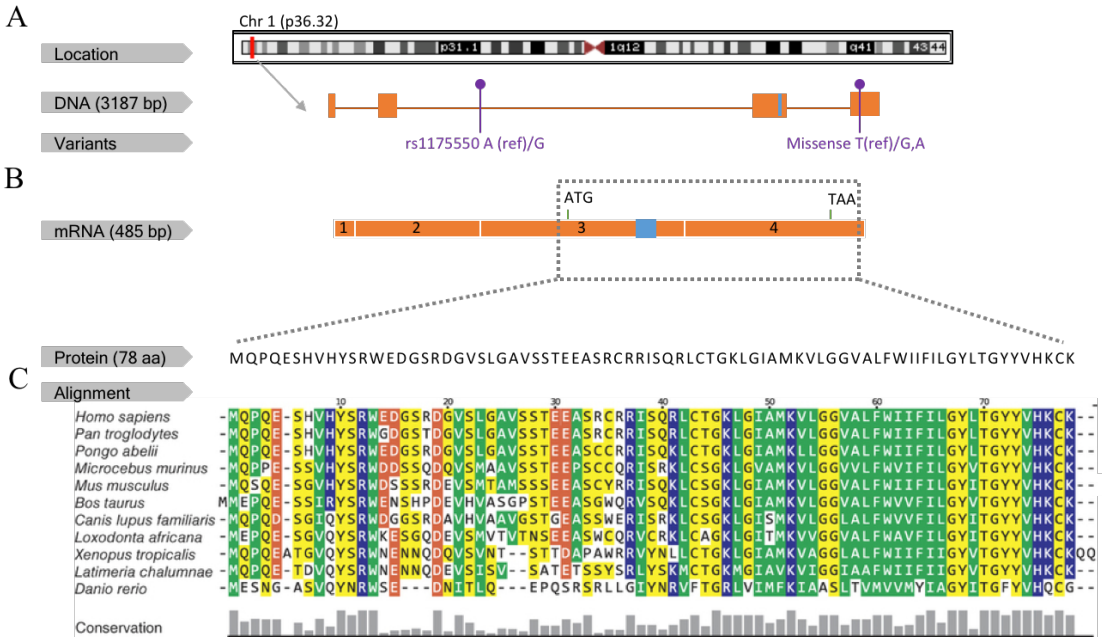


Figure 1.4 – *SMIM1* gene underlies the Vel blood group. **A** - Schematic representation of *SMIM1* location, highlighted in red, in chromosome 1. *SMIM1* comprises four exons that are represented by orange boxes and the 17-bp frameshift deletion in exon three, marked in blue, has been identified as the cause of the Vel-negative phenotype. In purple are represented two common variants responsible for the Vel-weak phenotype; **B** - *SMIM1* gene encodes a 78 amino acid protein. **C** - Alignment of human SMIM1 and vertebrate's orthologues, shows a high level of amino acid conservation between species. Conserved amino acids are coloured according to their properties (green, hydrophobic nonpolar; yellow, polar uncharged; orange, acidic; blue, basic), and non-conserved amino acids are uncoloured. Bars indicate the degree of conservation (Figure adapted from [188]).

A homozygous 17-bp frameshift deletion in the third exon of *SMIM1* gene (c.64_80del encoding p.S22Qfs, represented in blue in Figure 1.4), responsible for the abolishment of SMIM1 protein expression, was identified as the cause for the Vel-negative phenotype [182, 186, 188]. Moreover, it was also shown that Vel-antigen expression could vary between strong, weak and negative [186]. The variation of Vel expression has been attributed to different genetic variants. In particular, the weak expression of Vel was characterised by a heterozygous 17-bp frameshift deletion (c.64_80del encoding p.S22Qfs) or by a heterozygous missense mutation at the amino acid Met51 (two different missense mutations at Met51 were described, c.152T>A encoding p.Met51Lys and c.152T>G encoding p.Met51Arg) (Figure 1.4) [186]. The heterozygous missense mutation c.152T>G has a dominant negative effect on Vel expression [182]. This missense mutation is close to the predicted transmembrane domain of SMIM1 and may potentially lead to the formation of a non-functional protein without the ability to incorporate the membrane, or to the alteration of the protein's

epitope, decreasing significantly the binding of anti-Vel [186]. Furthermore, it has been shown in an expression quantitative trait locus (eQTL) that rs1175550 (minor allele frequency, MAF= 0.23), a common variant in intron 2 of *SMIM1*, also contributes to the variable expression of Vel antigen [186]. Here, the major allele A was associated with decreased *SMIM1* expression levels when compared with the minor allele G [186]. This common variant is in an erythroid-specific regulatory element, to which several transcription factors such as GATA-1, GATA-2 and TAL1 bind [186, 188-190]. Although this common variant does not disrupt any transcription factor binding motif, it seems to affect *SMIM1* expression by altering DNA-shape characteristics, that consequently interfere with transcription factor binding affinity. GATA-1 and TAL-1 binding seems to be favour by minor allele G [189, 190].

The Vel blood group is clinically relevant because of: i) severe haemolytic reactions that occur in Vel-negative individuals transfused with Vel-positive RBCs, and ii) haemolytic disease of newborns of Vel-negative mothers [183, 188, 191-194]. As the biological function of SMIM1 remains unknown, additional phenotypes may yet be uncovered. Currently, the two challenges that health care providers face when dealing with Vel-negative individuals are the ability to maintain blood stocks for safe transfusions and to perform the correct serologic typing. The latter is particularly challenging due to the high variability of Vel expression across individuals.

1.6 Thesis aims and overview

Even though *SMIM1* has been identified as the gene underlying the Vel-blood group, its biological functions remain largely unknown. The overall objective of my thesis work was to improve the current knowledge of the role of SMIM1. This knowledge may be essential to reveal unknown clinical phenotypes related to SMIM1 absence and to inform their prevention and treatment. The approaches I have taken in this thesis were: i) to characterise the role of SMIM1 in haematopoietic lineages and ii) to investigate the role of SMIM1 in other organs.

This thesis is divided into three results chapters:

In **Chapter 3**, I characterised *SMIM1* gene expression throughout the haematopoietic lineage and SMIM1 protein localisation in some blood cell types.

In **Chapter 4**, using two cohorts of human *SMIM1* naturally occurring knockouts and an *Smim1* mutant mouse model, I investigated the effect of SMIM1 absence in different blood traits and other organs.

In **Chapter 5**, I further assessed the role of SMIM1 in megakaryopoiesis and platelet function *in vitro*.

CHAPTER 2 | MATERIALS AND METHODS

This chapter is divided into three main sections in accordance with the three models used in this study: human, mice and cell lines. However, some methods were used across the different models.

2.1 Human samples

2.1.1 Vel-negative and control donors recruitment and sample collection

Recruitment and sample collection of Vel-negative donors were carried out under the project "*A BluePrint of Blood Cells",* ethically approved by Cambridge University Hospitals NHS Trust research ethics committee (REC number 12/EE/0040). Vel-negative blood donors living in the United Kingdom (UK) that participated in a previous Vel study (n=39) [186], were contacted to take part in this study. Out of these, 27 accepted and signed the consent form. The donor recruitment was coordinated by Dr Karola Rehnstrom, Study Coordinator for the Ouwehand group. Blood samples were collected by a research nurse and the blood tubes were processed accordingly based on the test to be performed. A 2 mL ethylenediaminetetraacetic acid (EDTA) blood tube was collected to re-test the Vel status by serologic test (test performed by NHS Blood and Transplant (NHSBT)) and for genomic DNA extraction to confirm the null allele by Sanger sequencing. Other samples were collected, processed and analysed as described below.

The control blood donors used in this study were members of the NIHR Cambridge BioResource (http://www.cambridgebioresource.org.uk/) with informed consent (REC 12/EE/0040) at NHSBT, Cambridge. Details of the control donor cohort used for full blood count comparison with the Vel-negative donors are described in [195]. Vel-negative and weak individuals in the original control donor cohort were removed for this study. Full blood count was performed as described below for Vel-negative donors.

2.1.1.1 Human full blood counts and blood smears

Full blood counts were performed in blood collected in EDTA tubes using Sysmex XE-5000 hematology analyzer. Blood smears of blood collected in EDTA were performed by spreading 2.5 µl of whole blood across a microscope slide. The slides were then stained with Romanowsky staining at Haematopathology and Oncology Diagnostic

Service, Haematology Department, Cambridge University Hospitals NHS Foundation Trust.

2.1.2 UK Biobank cohort

UK Biobank is a prospective cohort study of approximately 500,000 participants aged between 40 and 69 years, which were recruited across the UK between 2006 and 2010. Each participant answered an extensive questionnaire focused on questions of health and lifestyle and have undergone physical measurements and provided blood, urine and saliva samples for analysis. This cohort, however, is not representative of the general population. UK Biobank participants lived within 25 miles of one the assessment centres and were more likely to be female, older, and to live in less socioeconomically deprived areas than the non-participants. Moreover, the participants in this cohort represent a healthier population when comparing with the general population (https://www.ukbiobank.ac.uk).

In this thesis, an association analysis of 26 blood traits with Vel-negative European ancestry individuals (part of the release 400,000 participants of UK Biobank) was performed by Dr William Astle, Lecturer in Haematological Genomics at University of Cambridge. The phenotypes were prepared according to the quality control (QC) outlined described in section "Phenotype Measurement, QC, and Processing" in the supplementary text in Astle *et al.* [196]. The genotype QC was performed in the Cardiovascular Epidemiology Unit at the University of Cambridge according to the technical report in Chapter 8 – Figure 8.2. The QC included removal of sex mismatches, removal of participants with a low genotyping call rate and with outlying levels of heterozygosity. The subset of European ancestry individuals was then used for statistical analysis fitting the following linear regressions:

a) $T = \alpha + s\gamma + x\beta + \varepsilon$

where T is the trait value, s indicates the sex of the individual (0=female, 1=male), x is a 0,1,2 variable counting the number of copies of the deletion the individual has and ε represents normally distributed noise. β can be interpreted as the averaged additive effect of the of the deletion allele (i.e. the average effect across all individuals in the UK Biobank sample) estimated by EFFECT_average. P_average corresponds to a two tailed t-test a $\beta$is non-zero.

b) $T = \alpha' + s\gamma' + x\beta' + sx\theta + \varepsilon'$

where, again, T is the trait value, s indicates the sex of the individual (0=female, 1=male), x is a 0,1,2 variable counting the number of copies of the deletion the individual carries and ε' represents normally distributed noise. β can be interpreted as the additive effect of the deletion allele in females (estimated by EFFECT - female) and $\beta + \theta$ can be the additive effect of the deletion allele in males (estimated by EFFECT - male). P_sex_interaction corresponds to a *t*-test that theta is non-zero.

2.1.3 Characterisation of *SMIM1* transcription profile in multiple blood lineages

Transcription profile of *SMIM1* in multiple blood cells was assessed using the available blueprint datasets, generated as part of the Blueprint project (https://blueprint.haem.cam.ac.uk/mRNA/) [197]. Briefly RNA-sequencing was performed in highly pure samples of progenitor and precursor blood cells (including HSC, CLP, CMP, GMP, MEP, MK, EB, platelet, neutrophil, monocyte, macrophage M0, macrophage M1, T helper cell, central memory T cell, effector memory T cell, cytotoxic (killer) T cell, terminally differentiated effector memory T cell, regulatory T cell, natural killer cell, mesenchymal stem cell, blood outgrowth endothelial cell and human umbilical vein endothelial cell). EBs and MKs were generated by differentiation of $CD34^+$ HSCs, as previously described [186]. HSCs and other progenitor and precursor cells were purified from the mononuclear cell fraction of human cord or venous blood by fluorescent-activated cell sorting (FACS) using cell-type specific antibodies, as previously described [186]. Purified samples (with 20,000 to 100,000 cells) were collected into TRIzol® (Invitrogen) and total RNA was extracted using the RNeasy mini kit (Qiagen). RNA-sequencing libraries were obtained using SMARTer Ultra Low Input RNA for Illumina sequencing and Advantage 2 PCR kit (Clontech) using 100 pg of total RNA as input. Paired-end sequencing was performed on a HiSeq 2000 system with TruSeq reagents (Illumina), according to the manufacturer's protocol. Sequencing results were analysed as previously described and the gene expression was quantified using Cufflinks v.1.3.0 [186].

2.1.4 Characterisation of SMIM1 localisation

2.1.4.1 Human platelet-rich plasma and washed platelets

Blood anticoagulated in sodium citrate was centrifuged at 150 g for 15 minutes (acceleration and brake 0) to obtain platelet-rich plasma (PRP). Two volumes acid citrate dextrose (ACD, Sigma-Aldrich) and 1 μM prostaglandin E1 (PGE1, Sigma-Aldrich) were added to 1/3 of PRP and gently inverted to mix. PRP was centrifuged at 1000 g for 15 minutes. Platelet pellet was washed with platelet washing buffer (36 mM citric acid (Sigma-Aldrich), 103 nM sodium chloride (Sigma-Aldrich), 5 mM potassium chloride (Fisher Scientific), 5 mM EDTA (Sigma-Aldrich) and 5.6 mM D-glucose (VWR)) supplemented with 0.1 μM PGE1 to prevent platelet activation. After centrifugation at 1000 g for 10 minutes the platelet pellet was washed again and then resuspended as required.

2.1.4.2 Human platelet fractionation

Platelet fractionation was performed by Louisa Mayer, a colleague in the laboratory. In brief, washed platelets were lysed in lysis buffer supplemented with protease inhibitor through consecutives freeze-thaw cycles. Lysate was then placed on a sucrose gradient of 30% to 60%, at 5% increments and centrifuged for 90 minutes at 200.000 g at 4 °C. The different gradient fractions were carefully collected in individual tubes and protein was precipitated from the sucrose gradient by incubation with 1:10 of 100% trichloroacetic acid on ice for 2 hours, following washing with cold acetone. Protein pellets were resuspended with 1X NuPAGE® LDS sample buffer and 1X NuPAGE® reducing agent (Invitrogen) and then processed as described in section 2.2.8.

2.1.4.3 Human RBCs isolation and lysis for western blot

For RBCs isolation, blood anticoagulated in EDTA was centrifuged as mentioned above and PRP and peripheral blood mononuclear cells (PBMCs) removed. RBCs were then diluted with 1 volume of D-PBS and centrifuged at 2000 g for 10 minutes. RBCs were washed three times with D-PBS by centrifugation at 2000 g for 10 minutes and stored at -20 °C. RBCs were then lysed by successive washes with cold RBC buffer (5 mM sodium phosphate dibasic (Sigma-Aldrich), 1 mM EDTA (Sigma-Aldrich), pH 8) supplemented with complete EDTA-free protease inhibitor (Roche) followed by ultra-centrifugations at 4 °C until the supernatant was completely clear. RBCs pellet was resuspended in RBC buffer and processed as describe in western blot section 2.2.8.

2.1.4.4 Whole blood flow cytometry

1% whole blood diluted in HEPES buffered saline (HBS, Table 2.9), PRP or PBMCs was incubated with anti-SMIM1 at room temperature (RT) for 45 minutes. Cells were washed once with D-PBS-1% bovine serum albumin (BSA) at 600 g for 6 minutes and then incubated with secondary antibody and when adequate with conjugated-cell-type specific antibodies, in the dark for 25 minutes (Table 2.1). Cells were washed with D-PBS-1% BSA and centrifuged at 600 g for 6 minutes. After the third wash, cells were fixed with 0.2% formyl saline (Table 2.9) and analysed. Flow cytometry was performed on FC-500 cytometer and analysed using Kaluza Analysis Software v.1.5a (Beckman Coulter).

Table 2.1 - Antibodies used for flow cytometry of whole blood.

Antibody	Manufacturer	Dilution
Human anti-human SMIM1 [SpG213Dc human monoclonal antibody – recognises the C-terminal of SMIM1]	Gift from Yannic Danger, LNPRM, FR	10 µg/ml
Goat anti-human IgG H&L - FITC	Ab98524	1:50

2.1.4.5 Differentiation of human $CD34^+$ cells into MKs

CD34 progenitor cells were isolated from apheresis cones in two steps: first mononuclear cells were isolated by a density gradient centrifugation and then CD34 progenitor cells were enriched by positive selection using the CD34 MicroBead Kit human (Miltenyl Biotec). In brief, blood was diluted up to 100 ml with buffer 1 (D-PBS (Sigma-Aldrich), 13 mM trisodium citrate (Sigma-Aldrich), and 0.2% HAS (Seralab)). 25 ml of diluted blood was carefully layered on top of 12.5 ml of Ficoll-Paque™ (GE Healthcare, 17-5442-03). Sample was centrifuged at 800 g for 15 minutes (acceleration 8 and brake 0). PBMC layer was carefully collected into a 50 ml tube and washed with buffer 1. After centrifugation at 600 g for 6 minutes (at 4 °C, acceleration 5 and brake 3), supernatant was removed, and cells resuspended up to 50 ml with cold buffer 4 (D-PBS (Sigma-Aldrich), 2 mM EDTA (Sigma-Aldrich), and 0.2% HAS (Seralab)). An aliquot was used for determination of cell count in a haemocytometer using trypan blue solution (Sigma-Aldrich, T8154) and the remaining cells were pelleted at 600 g for 6 minutes (at 4 °C, acceleration 5 and brake 3). Pelleted cells were resuspended in 150 µl of buffer 4 per 10^8 cells. 50 µl of FcR bloking reagent and 50 µl of CD34 microbeads per 10^8 cells were added and sample was incubated at 4 °C. After 30 minutes, 20 ml of

buffer 4 was added and sample was centrifuged at 300 g for 6 minutes (at 4 °C, acceleration 5 and brake 3). Cells were then resuspended in 500 µl of buffer 4 per 10^8 cells and purified on the AutoMacs machine following manufacture's protocol. Enriched sample was collected directly on CellGRO SCGM (CellGenix). Cells were counted as described above and purity was determined by flow cytometry. Cells were cultured with a density of $1x10^5$ per ml in MK medium (CellGRO SCGM (CellGenix) supplemented with 100 ng/ml TPO (CellGenix) and 10 ng/ml IL1β (R&D) in 12-well plates. Cells were examined regularly and when needed (up to day 5) split or topped up with fresh medium with the correct concentration of cytokines for total volume. From day 6 cell's maturity was monitored by looking at CD41 and CD42 expression by flow cytometry. When the percentage of double positive cells was higher than 70%, MKs were used for proplatelet assay, western blot (section 2.1.4.6 and 2.2.8) and immunostaining. For immunostaining, $30\text{-}50x10^4$ MKs were spun on microscope slides at 400 rpm for 5 minutes using a Shandon Cytospin 4 machine and then stained as described in section 2.1.4.7. For flow cytometry, $3x10^3$ unstained cells and stained with anti-CD34-FITC (to confirm sample purity - 1:10, BD Pharmingen), anti-CD41a-APC and anti-CD42a-FITC (to confirm MK maturity – 1:10, BD Pharmingen) where incubated in the dark for 20 minutes. Fixed samples with 600 µl 0.2% formyl saline (Table 2.9) were then processed using a FC-500 cytometer and analysed using Kaluza Analysis v.1.5a (Beckman Coulter).

2.1.4.6 Proplatelet assay

Sterile glass coverslips were coated in 200 µg/ml fibrinogen (F3879**,** Sigma-Aldrich) in D-PBS overnight at 4 °C. Coverslips were washed with D-PBS and MKs diluted $0.5x10^6$ MKs/ml in CellGRO SCGM. 300 µl or 150 µl of MKs were seeded per well of a 24-well plate or 12 µm well chamber (ibidi), respectively. Cells were examined regularly and when proplatelet formation was seen they were fixed in 1% paraformaldehyde (PFA) in D-PBS for 10 minutes at RT. Cells were washed three times with D-PBS and the plate kept at 4 °C until staining as described in section 2.1.4.7.

2.1.4.7 Immunostaining

Cells were quenched with 50µM ammonium chloride in D-PBS (Sigma-Aldrich) for 5 minutes at RT. Then cells were permeabilised three times for 5 minutes with D-PBS containing, 0.1% w/v saponin and 0.2 w/v gelatin (both form Sigma-Aldrich). Cells were incubated at RT with primary antibody diluted in saponin-gelatin-PBS for 1 hour,

followed by three times 5 minutes washes with saponin-gelatin-PBS. Then cells were incubated in the dark at RT with secondary antibody in saponin-gelatin-PBS for 1 hour (Table 2.2). Cells were again washed twice with saponin-gelatin-PBS and incubated for 10 minutes with 1 µg/ml 4',6-diamidino-2-phenylindole (DAPI) or Hoechst 33342 (Thermo Fisher Scientific) followed by two washes with PBS. Cells were mounted with hydromount (National Diagnostics) overnight in the dark. Images were acquired using a Leica DMi8 fluorescent microscope with DFC7000T camera or Leica Sp5 confocal microscope (Leica microsystems). Images captured were analysed using ImageJ software.

Table 2.2 - Antibodies used for immunostaining.

Antibody	clone	Manufacturer	Dilution
Rabbit anti-SMIM1	Polyclonal – recognises the N-terminal of SMIM1	HPA069088 - Atlas antibodies - Sigma	1:200
Mouse anti-β Actin	AC15	Ab6276 - Abcam	1:200
Phalloidin-A555		A34055 – Thermo Fisher Scientific	1:20
Mouse anti-THBS1	[A6.1]	Ab1823 - Abcam	1:200
Mouse anti-Lamp3-PECy5		BD561982	1:50
Mouse anti-Lamp-2	[H4B4]	Sc-18822	1:20
Mouse anti-58K-Golgi		Abcam	1:50
Mouse anti-IgG-A633		Thermo Fisher Scientific	1:500
Rabbit anti-IgG-A555		Thermo Fisher Scientific	1:500
Rabbit anti-IgG-A488		Thermo Fisher Scientific	1:500-1:1000

2.1.5 Human platelet activation and SMIM1 analysis by flow cytometry and western blot

This protocol was optimised in order to control for platelet activation during sample manipulation. Blood was collected in sodium citrate and 1 µM PGE1, 0.5 U/ml apyrase and 100 µM aspirin was immediately added to the negative control sample. The PRP was then prepared as described in section 2.1.4.1 and 2 ml samples of the obtained PRP were either inhibited (1 µM PGE1, 0.5 U/ml apyrase and 10 % ACD v/v (Sigma-Aldrich)) or stimulated with thrombin-receptor agonist peptide-6 (TRAP-6, Tocris) and centrifuged at 800 g for 10 minutes. Samples were then washed twice with platelet washing buffer (section 2.1.4.1) and centrifuged at 800 g for 10 minutes. Negative control sample was always washed with platelet washing buffer supplemented with (1 µM PGE1 and 0.25 U/ml apyrase). 5 µl of each sample was collected for analysis by

flow cytometry and remaining sample was pellet for lysis and western blot as described in section 2.2.8. For Western blot the following alterations were applied: lysed samples were divided and treated with: complete EDTA-free protease inhibitor (Roche) or protease and phosphatase inhibitors (Thermo Scientific). Samples treated with protease inhibitor were then incubated at 30 °C for 1 hour with Lambda protein phosphatase (P0753S, NEB) following manufacture's protocol. Platelet activation was assessed by flow cytometry, measuring the level of P-selectin exposure. Flow samples were incubated for 20 minutes in the dark with anti-CD62-PE-Cy5 (AK-4, 551142, BD Pharmingen) and IgG1- PE-Cy5 isotype control (555750, BD Pharmingen). The expression level of SMIM1 was determined as described in section 2.1.4.4. Fixed samples in 0.2% formyl saline (Table 2.9) were processed using a FC-500 flow cytometer and analysed using Kaluza Analysis Software v.1.5a (Beckman Coulter).

2.2 Mouse model

2.2.1 Mouse husbandry

Smim1 mutant colony was established using heterozygous (*Smim1*$^{+/-}$) mice generated by Wellcome Trust Sanger Institute as part of the International Mouse Phenotyping Consortium (http://mousephenotype.org/). Mice used in all experiments were bred and housed at Central Biomedical Services - University of Cambridge, a specific pathogen-free facility. Mice had *ad libitum* access to food and water. Normal diet was used, unless otherwise specified. General mouse care and ear notches for identification and genotyping were performed by in-house animal technicians. All mice were genotyped after weaning (at four to six weeks of age). Experimental protocols used in this study were performed in accordance to the United Kingdom Home Office regulations under the project licence 70/8850.

2.2.2 Generation of *Smim1* mutant mice

Smim1 mutant mice were generated in C57BL/6N genetic background using CRISPR/Cas9 technology at Wellcome Trust Sanger Institute. Briefly, four guide RNAs (Table 2.3) were used to mediate a deletion of critical exons, present in all *Smim1* transcripts. The resulting deletion has 865 bp, and its genomic location is at chromosome 4:154023124 – 154023988 (Figure 2.1 – genomic region highlighted in light grey).

Table 2.3 - gRNAs used to generate *Smim1* mutant mice. gRNAs sequence and chromosome location.

gRNA ID	Sequence	Chr	Chr Start	Chr End
1	CCTGTGTGACTCTCGCTGGTGGC	4	154023132	154023154
2	CTTGGGTCTACCCTGTGTACTGG	4	154023176	154023198
3	CCTGACCCTCGGGAATGTAAGTA	4	154024001	154024023
4	CCCTCGGGAATGTAAGTAAGCCT	4	154024006	154024028

The deletion was confirmed using two genotyping techniques: by end-point PCR and by loss of wildtype allele qPCR assay. Genotyping end-point PCR was done through a combination of two separate PCR reactions that detect the *smim1* wildtype allele and the mutant allele. PCR was carried out with 50-100 ng of genomic DNA, 10 μM of primers pairs (Table 2.4) and Platinum Taq (Invitrogen), with the following amplification conditions: 94 °C for 5 min, then cycled 34 times between 94 °C for 30 seconds, 58 °C for 30 seconds and 72 °C for 1:30 seconds followed by a 72 °C 5-min final extension step. Reactions were then analysed on agarose gels. Genotyping by loss of wildtype allele qPCR assay was done using smim1 probe (Table 2.4, Invitrogen) and FAM-labelled custom qPCR TaqMan® assay, VIC® labelled endogenous control assay following manufactory's protocol (Table 2.4, Invitrogen).

Table 2.4 - Primers used to genotype *Smim1* wildtype and mutant alleles at Wellcome Trust Sanger Institute.

Primer name	Fwd /rev	Sequence (5' - 3')	Target/assay
Smim1_DF1	fwd	TGGCAGCTACACTTTCCTGG	*Smim1* wildtype– PCR
Smim1_ER1	rev	ACTAGCTCAGGGTCACATCAG	*Smim1* wildtype– PCR
Smim1_DR1	rev	CTCAGCTGGAGCTTAGCCTAGT	*Smim1* mutant– PCR
Smim1_qPCR_1	fwd	GGAGTTACCTATCTGCTCAAAGG	LoA qPCR assay
Smim1_qPCR_2	rev	CCAGCAGCTCCCACTGA	LoA qPCR assay
Smim1_qPCR_p	probe	CTGGGCTTGAACAAAG	LoA qPCR assay

Nevertheless, CRISPR/Cas9 gRNAs off-target analysis was not performed. Therefore, one cannot exclude the possibility that other genes may also be affected.

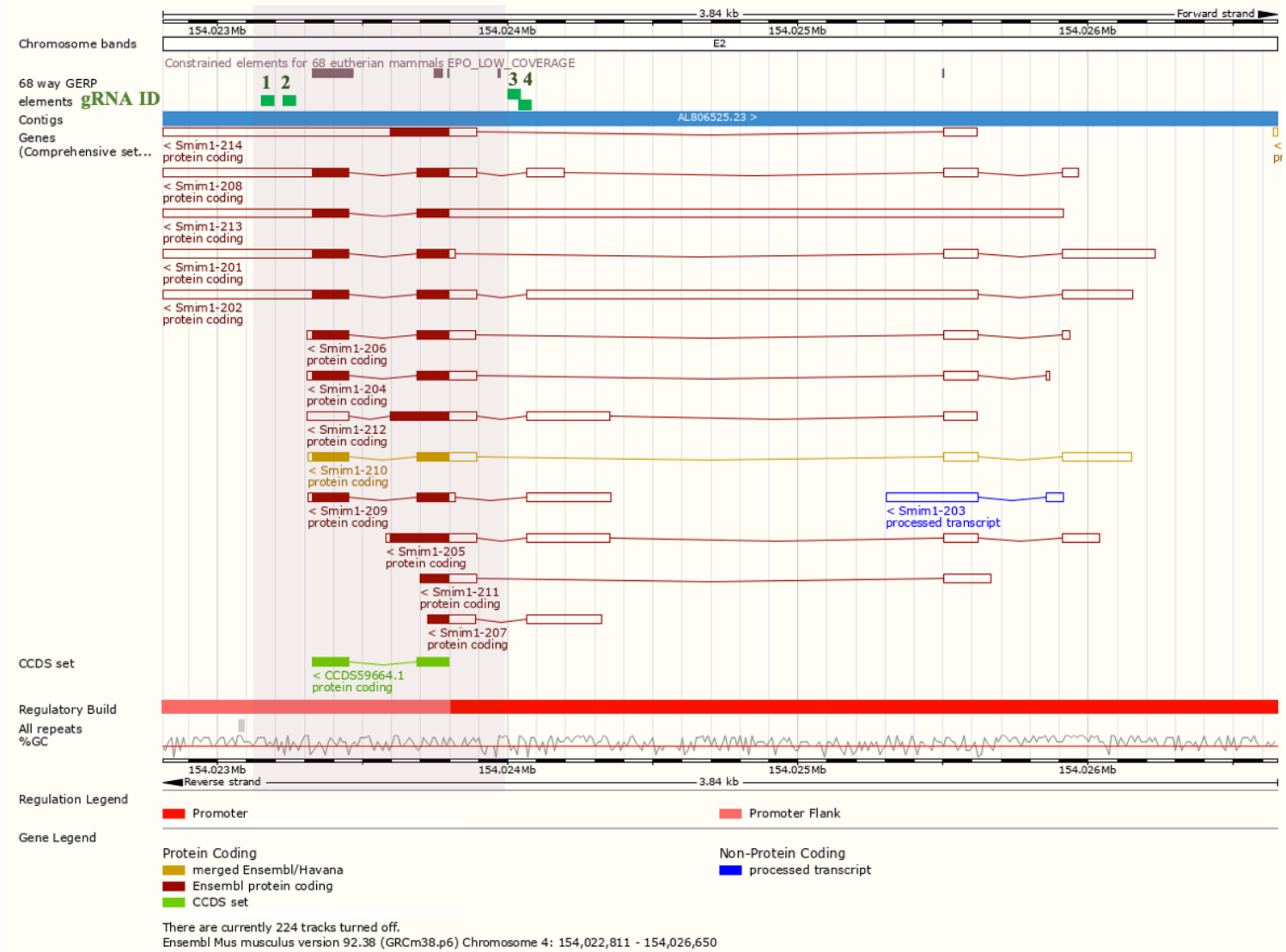


Figure 2.1 - Highlighting deleted region in *Smim1* mutant mice. Overview of mouse *Smim1* transcripts in chromosome 4 and schematic representation of *Smim1* deletion generated by CRISPR/Cas9 system in *Smim1* mutant mice. Four gRNAs (binding site represented in green) were used to delete *Smim1* region highlighted in grey (chr 4:154023124 – 154023988) (Figure/data obtain from ensemble **http://www.ensembl.org/Mus_musculus/Location/View?r=4:154022811-154026650;mr=4:154023124-154023988**) on 2018.11.20.

2.2.3 Mouse genotyping

Genotyping of mice was performed by combination of two separate PCR reactions using 0.2 µM of specific primers (Table 2.5) that detected *Smim1* wildtype and mutant allele. Genomic DNA was extracted from ear notches using PureLink Genomic DNA Kit (Invitrogen), following manufacturer's protocol. PCR was carried out with KAPA HiFi HotStart Ready mix (Biosystems), with the following amplification conditions: 95 °C for 5 min, then cycled 35 times between 98 °C for 20 seconds, 68 °C for 30 seconds and 72 °C for 15 seconds followed by a 72 °C 5-min final extension step. PCR products were resolved by electrophoresis as described in section 2.3.5.6.

Table 2.5 - Primers used to genotype *Smim1* wildtype and mutant alleles in this study.

Primer name	Fwd /rev	Sequence (5' - 3')	Target/assay
Smim1_F_U	fwd	CCTGCACAGAAAGATTTGAGGCCC	*Smim1* universal– PCR
Smim1_R_wt	rev	TCCCCATCTCCAGTACACAGGGTA	*Smim1* wildtype– PCR
Smim1_R_ko	rev	CAGAGGCCCCAGTTGTAAATGCAG	*Smim1* mutant– PCR

2.2.4 Mouse blood collection

Mouse blood was collected from inferior vena cava, directly into anticoagulant. For inferior vena cava bleeds, mice were sacrificed by exposure to increasing volumes of carbon dioxide and blood collected using a 25G x 1" needle (Terumo) and 1ml soft-ject® syringe (Henke Sass Wolf) into 100 µl 0.5 M EDTA (CIMR) or 100 µl acid citrate dextrose (ACD; Table 2.9) to prevent coagulation.

2.2.5 Mouse full blood count and blood smears

Mouse full blood counts were done using a Scil Vet ABC Plus animal blood analyser (scil animal care company), using blood collected into EDTA. Blood smears were done by spreading 2.5 µl of whole blood across a microscope slide. Dry blood smears were then stained with Rapid Romanowsky staining kit (TCS biosciences) following the manufacturer's protocol.

2.2.6 Mouse histology

Mouse was dissected, tissues collected and fixed immediately in 10% neutral formalin solution (Sigma-Aldrich) for 24h-48h at RT. Fixed hind legs bones were washed with

D-PBS and decalcified with 10% EDTA (pH 7.4) for 7-10 days. Tissues were then dehydrated through series of graded ethanol baths and paraffin-embedded. 4 μm sections were cut and stained with haemotoxylin and eosin (H&E) and periodic acid-Schiff (PAS). Mouse histology was performed by James Warner at histology core - Wellcome Trust-MRC Institute of Metabolic Science. Tissue morphology and specific stains were evaluated using Leica DMi8 fluorescent microscope with DFC7000T camera (Leica microsystems) or imaged using Zeiss Axioscan Z1 Slidescanner and analysed using HALO™ image analysis software. Professor Gary Hoffman (University of Western Australia), kindly gave me his feedback regarding general tissue morphology.

2.2.6.1 Assessment of glycogen content in mouse liver using PAS/DPAS staining

PAS/DPAS staining was done in two consecutive liver slices. In brief, one of the two slides was pre-treated with diastase, an enzyme that specifically breaks down glycogen, and then both slides were stained with PAS (untreated - PAS and treated-DPAS). All mouse liver samples for comparison were treated and stained at the same time. Slides were imaged using Zeiss Axioscan Z1 Slidescanner and analysed using HALO™ image analysis software with a convolution neural network classifier. Three classifiers (tissue, vessels and background) and a fixed margin were used to avoid blood vessel and margin effects, and to specifically detect areas of PAS stain. DPAS sample was used to set up the threshold of a weak stained area and the PAS sample was used to define strongly stained areas. Steatosis areas were also quantified, and its areas subtracted to total tissue area. The same analysis settings were applied to all samples. PAS stained areas were calculated in a ratio of tissue intensity area to tissue area, and then expressed as a percentage. Results were presented in total PAS stained area, weak PAS stained area, moderate PAS stained area and strong PAS stained area. To specifically quantify glycogen area, the quantification of the PAS stain area in the DPAS sample was subtracted from the quantified PAS area in the PAS sample. Slides scanning and HALO™ analysis was performed at Wellcome Trust-MRC Institute of Metabolic Science in collaboration with Professor Toni Vida-Puig, Dr Michele Vacca and Dr Vian Azzu, and with technical support of Mr Gregory Strachan.

2.2.6.2 Quantification of megakaryocytes in mouse bone marrow

Sections of mouse hind legs bones stained with H&E were imaged using Leica DMi8 fluorescent microscope with DFC7000T camera (Leica microsystems). Quantification was performed manually in ten fields per mouse, using ImageJ software.

2.2.7 Mouse platelet isolation for western blot and platelet spreading

Mouse whole blood was collected from the inferior vena cava (IVC) and anticoagulated with EDTA or ACD. Blood was diluted with 1.5 volumes of Tyrode's buffer (Table 2.9) and PRP was carefully collected after centrifugation at 100 g for 10 minutes (with acceleration and brake zero). PRP was washed with Tyrode's buffer containing 1 μM of PGE1 and 0.5 U/ml of apyrase to prevent platelet activation and centrifuged at 1000 g for 10 minutes. Platelet pellets were washed with the Tyrode's buffer containing 1 μM of PGE1 and 0.5 U/ml of apyrase and centrifuged at 1000 g for 10 minutes, and then resuspended in Tyrode's buffer for spreading assay or lysis buffer (RIPA buffer with protease inhibitors – section 2.2.8) for western blot.

2.2.8 Western blot

Cells were lysed using RIPA buffer (150 mM sodium chloride,60 mM Tris pH8, 0.6% w/v sodium deoxycholate, 0.1% w/v sodium dodecyl sulfate, 1% v/v NP-40) with complete EDTA-free protease inhibitor (Roche) for 15 to 30 minutes at 4 °C. Cell lysates were then centrifuged at 13000 rpm for 15 minutes at 4 °C. Supernatant proteins were quantified by Bradford protein assay (Bio-Rad) following the manufacturer's protocol. Protein samples were prepared with 1X NuPAGE® LDS sample buffer and 1X NuPAGE® reducing agent (Invitrogen), following heating at 70 °C for 10 minutes. Protein samples and PageRuler prestained protein ladder (Thermo Scientific) were loaded onto a NuPAGE® 4-12 % Bis-Tris gel and run in NuPAGE® MES or MOPS SDS running buffer with NuPAGE® Antioxidant (Invitrogen) at 200 V. Proteins were transferred onto PVDF membrane (Millipore) by XCell II blot module (Invitrogen) at 30 V for 1 hour, following the manufacturer's protocol. Membranes were blocked with 5 % non-fat dry milk (Sigma-Aldrich) in TBS-T buffer (1X Tris buffered saline with 0.1 % Tween20) or Odyssey Blocking Buffer (LI-COR®) at RT for 1 hour. Then, membranes were incubated with primary antibodies (Table 2.6) diluted in 3 % non-fat dry milk (Sigma-Aldrich) prepared in TBS-T or Odyssey Blocking Buffer, overnight at 4 °C. After primary antibody incubation, membranes were washed three times in TBS-T

or Odyssey Blocking Buffer for 10 minutes and then incubated with secondary antibody (Table 2.6) in 1.5 % non-fat dry milk (Sigma-Aldrich) prepared in TBS-T or Odyssey Blocking Buffer supplemented with 0.1% Tween20 for 1 hour at RT. Then, the membranes were washed three times in TBS-T and developed. Antibody incubations and washes were performed with gentle rotation and rocking. For secondary-HRP antibodies, membranes were developed with Pierce™ ECL Plus Western Blotting Substrate (Thermo Scientific) following the manufacturer's protocol and exposed to photographic film in the dark room. The films were developed and the image was digitalized. When using LI-COR® Odyssey imaging system, washed membranes were imaged with Odyssey Fc OFC-1095 version 1.0.36 imaging system (LI-COR; IRDye. 680RD and IRDye. 800CW conjugated). When needed the membranes were stripped with stripping buffer (2% SDS, 63 mM Tris-HCl pH 6.8, 114 mM β-Mercaptoethanol) for 45 minutes at 50 °C, washed several times with TBS-T at RT, and then blocked and processed as mentioned above. Stripping of membranes used with LI-COR® Odyssey system, was performed according manufacture's protocol.

Table 2.6 - Antibodies used for western blot.

Antibody	clone	Manufacturer	Dilution
Rabbit anti-SMIM1	Polyclonal – recognises the N-terminal of SMIM1	HPA069088 - Atlas antibodies	1:2000
Rabbit anti-CD62P	Polyclonal	Ab59738 - Abcam	1:2000
Rabbbit anti-VEGFA	[VG-1]	Ab1316 - Abcam	1:5000
Mouse anti- THBS1	[A6.1]	Ab1823 - Abcam	1:5000
Rabbit anti-VWF	Polyclonal	A0082 - Dako	1:5000
Rabbit anti- CD42b	Polyclonal	Sc-292722 Santa Cruz	1:1000
Rabbit anti-human CRLF3	Polyclonal	Atlas antibodies	1:1000
Rabbit anti-human GAPDH	14C10	Cell Signalling Techonology	1:1000
Mouse anti- β Actin	AC15	Ab6276 - Abcam	1:10000
Pierce goat anti-rabbit IgG Poly - HRP		32260 – Thermo Scientific	1:10000
Pierce goat anti-mouse IgG Poly-HRP		32230 – Thermo Scientific	1:10000
IRDye 800CW Donkey anti-rabbit IgG		LI-COR®	1:10000
IRDye 800CW Goat anti-mouse IgG		LI-COR®	1:10000
IRDye 680LT Goat anti-rabbit IgG		LI-COR®	1:10000
IRDye 680LT Donkey anti-mouse IgG		LI-COR®	1:10000

2.2.9 Mouse platelet spreading

Fibrinogen coated coverslips were prepared in advance by incubation with 200 µg/ml of fibrinogen (Sigma, F3879) overnight at 4°C. Coverslips were then washed twice with D-PBS and blocked with D-PBS-1 % BSA (Sigma-Aldrich) for at least one hour at RT. After washing twice with D-PBS, coverslips were washed with Tyrode's buffer supplemented with 2 mM calcium chloride. Washed platelets were diluted in Tyrode's buffer supplemented with 2 mM calcium chloride and adjusted to a concentration of $25x10^3$ platelets/µl. Adjusted platelets, after incubation for 30 minutes at 37°C, were supplemented with 0.01 U/ml of thrombin to induce platelet activation/spreading, and immediately added to the coverslip previously coated with fibrinogen. After incubation for 30 minutes at 37°C, unbound platelets were removed by washing twice with PBS. Spreading platelets were fixed with 4 % PFA for 15 minutes at RT and then washed twice with D-PBS. Platelet immunostaining was performed as described in section 2.1.4.7, using Alexa Fluor® 555 phalloidin (1:20 dilution - Invitrogen, A34055). Ten pictures per coverslip were taken in Leica DMi8 microscope with DFC7000T camera (Leica microsystems). Percentage of attached and spread platelets was determined manually using ImageJ software.

2.2.10 Mouse platelet transmission electron microscopy (TEM)

Blood collected from IVC was diluted in 1.5 volumes of 0.9% sodium chloride (Sigma-Aldrich) and PRP obtained after centrifugation at 100 g for 10 minutes. Washed PRP in 0.9% sodium chloride (Sigma-Aldrich) was centrifuged at 1000 g for 10 minutes. After centrifugation, the obtained platelet pellet was resuspended in primary fixation buffer (0.5% glutaraldehyde, in 0.1M sodium cacodylate buffer, pH 7.4 – gift from Cambridge Advanced Imaging Centre) and incubated overnight at 4 °C. Fixed platelets were then processed by Dr James McMillan at Cambridge Advanced Imaging Centre. In brief, platelets were resuspended in secondary fixative buffer (2% glutaraldehyde, 2% formaldehyde in 0.05 M sodium cacodylate buffer containing 2 mmol/L calcium chloride, pH 7.4) and incubated overnight at 4 °C. Then, fixed platelets were incubated serially in buffers containing 1% osmium tetroxide and 1.5% potassium ferricyanide, 1% thiocarbohydrazide, 2% osmium tetroxide and finally 2% uranyl acetate in 0.05 M maleate buffer, pH 5.5. Samples were then dehydrated through washes in ascending series of ethanol solutions (50%, 70%, 90% and 100%), followed by treatment with dry acetone and dry acetonitrile, and infiltration with Quetol 651 epoxy resin. Sections were

cut at 60 nm using a diamond knife with a Leica UCT ultramicrotome. Sections were imaged using a FEI Tecnai G2 TEM, AMT XR60B camera and Deben software. Professor Kathleen Freson (University of KU LEUVEN), kindly gave me her feedback regarding platelet ultrastructure.

2.2.11 Mouse platelet activation assay

Mouse platelet response to different agonists was assessed by flow cytometry, measuring the level of P-selectin exposure and fibrinogen binding. ACD anticoagulated blood diluted in 200 µl Tyrode's buffer (Table 2.9) was diluted 1:10 with HBS (Table 2.9) supplemented with an antibody mix and either no agonist, 10 mM EDTA, cross-linked collagen related peptide (CRP-XL, gift from Dr Richard Farndale, University of Cambridge) or thrombin (Sigma-Aldrich). Thrombin stimulated samples were treated with 1.2 mM of Gly-Pro-Arg-Pro (GPRP, Sigma) to inhibit thrombin-induced fibrin clot formation and platelet aggregation [198]. Samples were mixed and incubated for 20 minutes at RT in the dark. The reaction was stopped by fixing the samples with 1:10 of 0.2% formyl saline (Table 2.9) for 10 minutes in the dark. Samples were further diluted to 1:30 with 0.2% formyl saline and processed using Gallios flow cytometer and analysed using Kaluza Analysis Software version 1.5a (Beckman-Coulter), within 1 hour of fixation. Blood collection and sample manipulation were performed carefully in order to avoid platelet activation.

The samples were incubated either without antibody, with antibody mix or with an isotype. The antibody mix was composed by: anti-human fibrinogen-FITC (1:25, F0111, Dako), anti-mouse P-selectin- PE (1:10, clone Wug.E9 emfret) and anti-mouse CD41-APC (1:25, clone eBioMWReg30, eBioscience). And the isotypes used were: anti-human IgG-FITC (1:25, F0215, Dako), IgG-PE isotype control (1:10, P190-2 emfret) and IgG1-APC (1:25, clone eBRG1, eBioscience). AbC Total Compensation Capture beads were used for compensation (A10497, Thermo Fisher Scientific).

Flow cytometry analysis was performed using logarithmic scale for platelet detection and differentiation in the light-scatter and fluorescence channels. Platelets were identified by their light-scatter profile and by the binding of anti-CD41-APC, which binds to platelet glycoprotein IIb regardless of platelet activation and doesn't affect fibrinogen binding. Five thousand individual platelets per sample were acquired.

Untreated platelets stained with IgG-PE isotype were used to set up P-selectin-PE gate, and CD41-positive platelets treated with EDTA, were used to set up the fibrinogen-FITC gate, both according to pick shape. The results were expressed as the percentage of P-selectin and fibrinogen-positive platelets for untreated (basal activation) and treated samples. Representation of gating strategy and flow cytometry results is shown in Figure **2.2**.

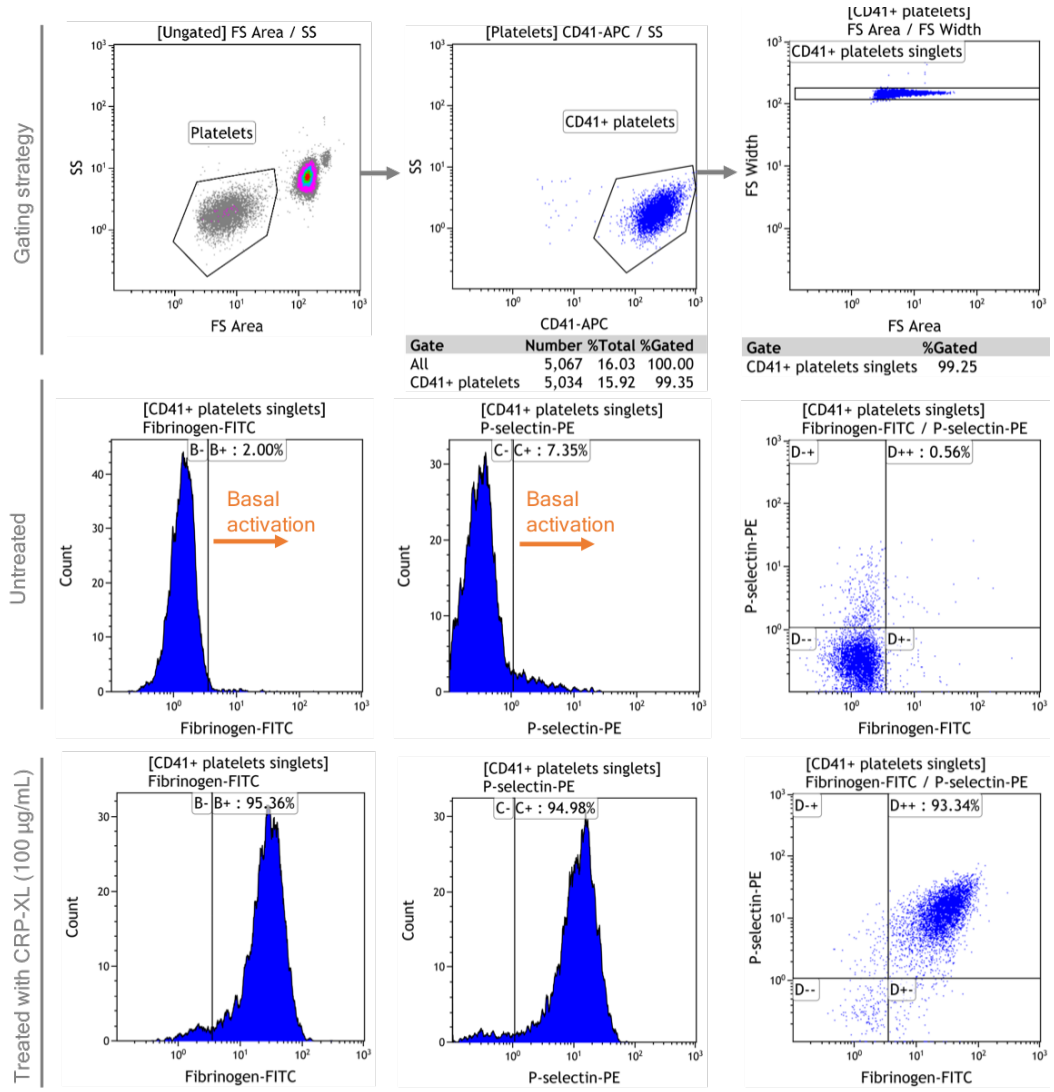


Figure 2.2 - Flow cytometry analysis of mice platelet activation. Mouse whole blood was treated with different agonists and platelet activation determined by P-selectin exposure and fibrinogen binding. Platelets were identified by their size and granularity distribution using side light scatter (SS) vs. forward light scatter (FS) area (first gate – platelets) and by expression of glycoprotein IIb - positive staining with CD41-APC (second gate –$CD41^+$ platelets). The percentage of P-selectin exposure and fibrinogen binding was determined in individual CD41-positive platelets (third gate –$CD41^+$ platelets singlets), in all the samples - untreated (resting platelets) and treated with the different agonists.

2.2.12 Mouse platelet isolation for RNA extraction

EDTA anticoagulated whole blood was diluted with 1.5 volumes of Tyrode's buffer (Table 2.9) containing 1 μM of PGE1 and 2 mM EDTA, to prevent platelet activation. Blood was centrifuged at 100 g for 10 minutes (acceleration and brake 0) to obtain PRP. PRP was diluted 1:1 with the same buffer and centrifuged at 100 g for 10 minutes to remove white blood cells excess. PRP was further purified by incubation, for 20 minutes at RT, with 50μl of Dynabeads (Invitrogen, 11035) previously prepared with 0.75 μg anti-TER-119 (BD Pharmingen, 550565) and 1.5 μg anti-CD45 (BD553086), to deplete residual red blood cells and leukocytes, respectively. PRP was magnetised twice and purified platelets were quantified by animal blood analyser (Scil Vet ABC Plus). The purified platelet pellet was lysed in TRIzol® (Invitrogen) and stored at -80 °C until RNA extraction. Platelet lysates of two to three mice per genotype were pooled to obtain enough RNA to perform RNA-Sequencing.

2.2.13 RNA extraction

Platelet lysate was transfer to a 2 ml heavy phase lock tube (5prime) and homogenised with 0.2 ml of chloroform per 1 ml of TRIzol® (Invitrogen) used. Post incubation for 3 minutes at RT, the sample was centrifuge at 12,000 g for 15 minutes at 4 °C. The upper aqueous phase containing the RNA was collected and RNA was co-precipitated with 10 μg of glycogen by adding 0.5 ml of isopropanol per 1 ml of TRIzol® (Invitrogen). After incubation at RT for 10 minutes, the sample was centrifuged at 12,000 g for 10 minutes at 4 °C. RNA pellet was washed twice with 1 ml of ice cold 75 % ethanol per 1 ml of TRIzol® (Invitrogen), following centrifugation at 8000 g for 10 minutes at 4 °C. The air-dry RNA pellet was then resuspended in RNAse free water (Qiagen). RNA was quantified using Qubit RNA HS Assay (Invitrogen) or Qubit RNA BR Assay (Invitrogen), following manufacturer's protocol. RNA quality was accessed on the Agilent 2100 Bioanalyzer, using Agilent RNA 6000 Pico Kit (Agilent Technologies).

2.2.14 RNA-sequencing and analysis

Mouse platelet RNA libraries were prepared by Ms Frances Burden, Research Associate in Prof. Ouwehand group. Libraries were made using KAPA Stranded RNA-Seq Kit with RiboErase (Human/Mouse/Rat) Illumina®plataform (KAPABiosystems), with the following modifications to the manufacturer's protocol: step 15 – library amplification clean-up was performed with 0.9X SPRI® clean-up beads instead of 1X, and libraries

were resuspended in water. Libraries were quantified by RTqPCR using the KAPA Library Quantification Kit (KAPABiosystems). Sequencing was performed by Genomics core facility at Cancer Research UK Cambridge Institute using HiSeq4000. RNA-sequencing data processing and quality control was performed by Dr Denis Seyres, Research Associate in Prof. Ouwehand group. In brief, mouse GRCm38.71(mm10) reference genome was used to build a Kallisto transcriptome index and a pseudoalignment of transcripts was done using Kallisto [199]. Then, transcript abundance estimates were summarized to gene level using Tximport (R package), which is a more relevant and stable method to compare replicates and experiments. Finally, differential analysis was performed with DESeq2 using gene counts and at a 5% false discovery rate. Principal component analysis (PCA) was performed using the transcript per million (TPM) data obtain with Kallisto.

2.2.15 Mouse RBCs real-time deformability cytometry (RT-DC)

EDTA anticoagulated blood was collected from IVC and processed within 30 minutes. Sample processing and analysis was performed by Ms Katie Bashant, PhD student at Professor Dr. Edwin Chilvers laboratory, Department of Medicine, University of Cambridge. In summary, 25 µl of whole blood was diluted 1:20 in PBS containing 0.5% (w/v) methylcellulose, which increases sample density and viscosity, hence, reduces cell sedimentation during the experiment and allow higher shear forces at lower flow velocities, respectively. Samples were loaded into a microfluidic chip (flic20 – normally used for human whole blood analysis, which did affect the quality of the measurements) and passed through a 20 µm microfluid constriction channel under pressure gradients and shear stresses that induce cell deformation. Cells were captured with a high-speed camera and mechanical proprieties were determined using a cell-tracing algorithm and offline ShapeOut software. Data acquisition was carried out at the rear part of the constriction channel, where the cell shape reaches a steady state. The following mechanical proprieties were quantified: area, deformability and surface smoothness. A cell-tracing algorithm was applied, which uses cell contour to derive cell cross-sectional area (A) and perimeter (l), and then, calculates cell circularity ($c, c = 2\sqrt{\pi A}/l$), which is used to determine cell-deformability. Deformation was determined by cell deviation from circularity when constricted in the channel ($Deformation = 1 - c$) (product of cell size and stiffness). A perfect circular object has c =1 so, a deformable object $c < 1$. [200]. Area ratio, which corresponds to surface smoothness, was calculated from the

ratio of convex hull area (cell size (μm^2)) and cell area. Three different RT-DC gating strategies were used: 5/80/80 to analyse red blood cells, (constant flow rate for 1 minute); 3/80/80, for leukocytes (constant flow rate for 7 minute); and 1/30/30 for platelets (constant flow rate for 5 minutes). These basic gates indicate to the RT-DC camera which size of cell to capture and refer to (minimum cross-sectional length/maximum horizontal cross-section length/maximum vertical cross-section length) 0/80/80 is a completely open gate. As reference, the circularity of non-deformed cells was determined in the reservoir (30 seconds x 3) in all the experiments. Minimum of 5000 RBCs were analysed using the gating strategy referred above. RBCs were identified on the basis of size and brightness and 1-1.07 area ratio was used to calculate size and deformation in smooth cells, and 1-1.2 area ratio was used to calculate roughness. These area ratio maximum gates eliminated red blood cell doublets from analysis.

2.3 Cell lines

2.3.1 Generation, maintenance and manipulation of human induced pluripotent stem cells (hiPSCs)

In this study two hiPSC lines were used: S4-SF5 and A1ATD1-c (referred as S4 and Bobc for simplicity, respectively). Both cell lines were a gift from Dr Cedric Ghevaert group at Department of Haematology, University of Cambridge. These hiPSC lines were derived from human skin fibroblasts through a non-integrating method by using Sendai virus vectors (SeVs; encoding for Yamanaka's factors OCT4, SOX2, KLF4 and MYC) [201, 202]. Bobc line was derived from a patient with a mutation in the alpha1 anti-trypsin gene, which was corrected as described by Yusa *et al.* [202].

2.3.1.1 Maintenance of hiPSCs

hiPSCs were maintained on vitronectin-coated plates (Invitrogen, A14700). When in single-cell state, hiPSCs were seeded onto irradiated mouse embryonic fibroblast (MEF) feeder plates to avoid differentiation. The cells were kept in an incubator at 37 °C, 5% CO_2 and daily fed with AE8 medium (AE6 medium supplemented with FGF2 and Activin A (Cambridge Stem Cell Institute, Tissue culture facilities)) or with Essential 8 medium (E8, Gibco). hiPSCs were observed daily under an inverted light microscope (Leica) and photographed using EVOS® FL Cell Imaging System (Life Technologies).

2.3.1.2 Preparation of vitronectin and feeder plates

The vitronectin-coating plates were prepared using 5 μg/ml rhVTN (Invitrogen, A14700) prepared in D-PBS for 1 hour at RT. To prepare MEFs feeders, the wells were firstly coated with 0.1% gelatin prepared in water from embryo transfer (G2500, W1503 - Sigma-Aldrich) for 25 minutes at RT. Then, a vial of D4 IRR MEFs (Cambridge Stem Cell Institute, Tissue Culture Facilities) was thawed in water bath at 37 °C and transferred drop-wise to a tube containing pre-warmed MEF medium (150 ml Dulbecco's modified Eagle medium (DMEM)/F12 (Life technologies), 1.66 ml L-Glutamine (Invitrogen), 16.6 ml fetal bovine serum, 1.16 μl β-mercaptoethanol). The cells were centrifuged at 200 g (acceleration and brake 5) for 5 minutes and then the pellet was resuspended in MEFs medium. Approximately $0.15x10^5$ cells/well were seeded in a 6-well plate. MEF feeder plates were used within one week to maintain hiPSCs in an undifferentiated state.

2.3.1.3 Passaging, freezing and thawing of hiPSCs

hiPSCs were passaged or frozen when 70-90% of confluence was reached to avoid cell differentiation. Cells were dissociated with pre-warmed D-PBS/ EDTA (5 mM EDTA in D-PBS) for 5 minutes at RT. After removing the D-PBS/EDTA, cells were washed and gently resuspended with pre-warmed wash medium (Table 2.9). Cells were then centrifuged at 200 g (acceleration and brake 5) for 5 minutes. When cell splitting was required, hiPSCs pellets were gently resuspended with pre-warmed AE8 medium and seeded in a vitronectin-coated plate (ratio 1/3 to 1/20). When preparing hiPSC for freezing, hiPSCs pellets were resuspended in pre-warmed freezing medium (Table 2.9) and placed into cryovials which were then transferred to a Mr. Frosty™ (Fisher Scientific) and placed at -80 °C or -150 °C for long-term storage. Culture of cryopreserved hiPSCs was performed by partially thawing cryopreserved hiPSCs in a water bath at 37 °C and transferring drop-wise into a tube with pre-warmed wash medium. Wash medium (table 2.9) was then added drop-wise into cryovials to recover the remaining cells. The cells were centrifuged at 200 g (acceleration and brake 5) for 5 minutes and then gently resuspended with pre-warmed AE8 medium and seeded in a vitronectin-coated plate.

2.3.2 Generation of *SMIM1* knockout hiPSC cell line using CRISPR/Cas9n

SMIM1 knockout line was created using CRISPR/Cas9n technology in S4 hiPSC line.

2.3.2.1 Design and cloning of CRISPR-RNA guides

CRISPR (Clustered regulatory interspaced short palindromic repeats) guide sequence consists of a 20-bp oligonucleotide located immediately upstream of a 5'-protospacer adaptor motif (PAM, 5'-NGG motif) [203]. All the gRNA sequences were identified in the human genome (GRCh37) via online CRISPR design tools from the Wellcome Trust Sanger Institute, Cambridge, UK (http://www.sanger.ac.uk/htgt/wge/crispr), DNA2.0, Inc. (https://www.dna20.com/products/crispr) and Zhang Lab, MIT (http://crispr.mit.edu/). The results of the different online tools were compared and the best-rated guide sequences/pairs, according to off-targets, offset (0 to 20 bp) and also binding site, were chosen (**Table 2.7**). To clone the gRNAs, a forward (Fwd) and reverse (Rev) oligonucleotide containing the restriction site of BbsI endonuclease was designed. A guanine (G) base was added to the beginning of each gRNA sequence to improve its expression under U6-promoter. The oligonucleotides were designed without the PAM sequence (Table 2.10). gRNA annealing and cloning were done as previously described [204-206]. Briefly, gRNA oligonucleotides were phosphorylated and annealed in a 10 µl reaction (10 µM of each oligonucleotide, 1x T4 DNA ligase buffer 10x (NEB, B0202S) and 10 U of T4 polynucleotide kinase (NEB, M0201S)) by a thermal cycler using the following parameters: 37 °C for 30 minutes; 95 °C for 5 minutes; ramp down to 25 °C at 5 °C minutes^{-1}. Then, GRNAs were diluted 1:200 in water at RT and cloned into pSpCas9n(BB)-2A-Puro (PX462) (Addgene plasmid #48141, gift from Feng Zhang). A simultaneous digestion and ligation were performed to clone RNA guide into PX462. A 20-µl reaction (100 ng of PX462, 2 µl of diluted gRNA, 1x Tango buffer 10x (Fermentas/Thermo Scientific, BY5), 0.5 mM dithiothreitol (DTT, Thermo Scientific, R0861), 1 µl FastDigest BbsI (Fermentas/Thermo Scientific, FD1014)) was placed into a thermocycler with the following parameters: 6 cycles of 37 °C for 5 minutes and 21 °C for 5 minutes. Then, 11 µl of ligation was treated with PlasmidSafe exonuclease (Epicentre, E3101K) in a 15-µl reaction (1x Plasmid buffer 10x, 1 mM ATP, 10 U PlasmidSafe exonuclease) at 37 °C for 30 minutes, followed by 30 minutes at 70 °C [206]. Finally, 2 µl of treated plasmid was used to transform *E.coli DH5α* high-competent cells (NEB) as described in section

2.3.5.1. All cloned gRNAs were expanded. Sequence and orientation were confirmed by Sanger sequencing, using U6-Fwd and/or hGATA4-Rev primers (Table 2.10).

Table 2.7 - Guide RNA used to target *SMIM1* in hiPSC. The reverse sequence is represented for guide sequences targeting the bottom strand. The sequence in bold corresponds to the PAM motif. The target region refers to the localization of the guide sequence in GRCh37 human genome.

Guide name	Target strand	Sequence (5'-3')	Target region
SMIM1_CR_1	Bottom	ATAGTGGACGTGGCTCTCCT-**GGG**	1:3691945-3691967
SMIM1_CR_2	Top	CAGGGACGGAGTCAGCCTAG-**GGG**	1:3691988-3692010
SMIM1_CR_3	Bottom	CGTCCTCCCACCTACTATAG-**TGG**	1:3691961-3691983
SMIM1_CR_4	Top	GCAGGGACGGAGTCAGCCTA-**GGG**	1:3691987-3692009
SMIM1_CR_5	Bottom	CCCACCTACTATAGTGGACG-**TGG**	1:3691955-3691977

2.3.2.2 Nucleofection, colony picking and screening of hiPSC

S4 cells were fed with pre-warmed AE8 medium complemented with 10 µM ROCK inhibitor (Y-27632, Sigma-Aldrich) 12-24h before nucleofection. hiPSCs were washed twice with D-PBS and then incubated at 37 °C for 5 minutes with cell dissociation solution (StemPro® Accutase®, Life Technologies, A1110501). Cells were gently resuspended to form 3-4 cell clumps and collected into wash medium; then, cells were centrifuged for 5 minutes at 200 g (acceleration and brake 5) and resuspended in pre-warmed AE8 medium complemented with 10 µM ROCK inhibitor. Approximately $1.5x10^6$ S4 cells were resuspended with a mix of 100 µl of Human Stem Cell Nucleofector® Kit 2 (Lonza, VPH-5022), and 2 µg, 10 µg or 20 µg of sgRNA/Cas9n-enconding plasmids and donor vector (equimolar concentrations of each sgRNAs and donor vector). Cells were promptly electroporated using Amaxa™ Nucleofector™ II – program B-016 (Lonza). Immediately after nucleofection, 600 µl AE8 medium complemented with 10-µM ROCK inhibitor was added and cells transferred onto 3 wells of a 6-well MEFs-coated plate with AE8 medium complemented with 10 µM ROCK inhibitor and incubated. 24h post-nucleofection (day 1), nucleofection efficiency was determined by monitoring green fluorescent protein (GFP) expression using EVOS® FL Cell Imaging System. Antibiotic selection with 1 µg/ml puromycin (Clontech 631305) was also initiated. Selection was carried out for 2 days to eliminate all cells without sgRNA/Cas9n-plasmid. Small colonies started to appear 8-12 days after nucleofection. On days 14-18, colonies were manually picked into vitronectin-

coated 24-well plates, expanded for genomic DNA (gDNA) extraction (described in section 2.3.5.2) and culture was continued.

Identification of targeted clones was carried out by PCR screening of genomic DNA (gDNA), using specific primers (Table 2.10) and MyTaq (Bioline) (described in section 2.3.5.4). PCR products were analysed by agarose gel electrophoresis (described in section 2.3.5.6). Sanger sequencing of the PCR product was undertaken to access the specific mutation of each positive clone. When required, the PCR product of each clone was cloned into pGEM-T easy vector (Promega), according to the manufacturer's instructions, and several colonies were sequenced using Sp6 and T7 primers (Table 2.10).

2.3.3 Forward programming of hiPSC to megakaryocytes

Forward programming of hiPSCs into MKs was undertaken following the method published by Moreau *et al* [207]. However, alterations have been made to improve the published protocol by Dr Tomas Moreau, Research Associate at Department of Haematology, University of Cambridge. Therefore, two different models of forward programming were used, according to the cell line used. A 3D model was used to forward programming S4 hiPSCs and a 2D model for Bobc hiPSCs.

In brief, in the 3D model on day 0 an Aggrewell™400 plate (Stemcell technologies), was prepared by washing the aggrewells with 500 µl of Aggrewell Rinse solution (Stemcell technologies). Following plate centrifugation at 2000 g for 5 minutes it was added 500 µl of transduction medium, EB1 medium (AE6 media supplemented with 10 µM ROCK inhibitor Y-27632 (Sigma-Aldrich), 10 ng/ml bone morphogenetic protein 4 (BMP4, R&D) and 10 µg/ml protamine sulfate (Sigma-Aldrich)) per aggrewell and the plate was incubated at 37 °C, 5% CO_2. A sub-confluent hiPSC culture was dissociated to single cell with TrypLE (Thermo Fisher Scientific) and $1x10^6$ cells per aggrewell were resuspended in 200 µl of EB1 medium supplemented with GATA1, FLI1 and TAL1 lentiviral vectors (multiplicity of infection (MOI) 20, Vectalys). Plate was centrifuged at 100 g for 3 minutes and incubated at 37 °C, 5% CO_2. 24 hours after transduction (day 1), the embryoid bodies (EBs) formed in the aggrewells were collected and washed in washing medium. After centrifugation at 20g for 1 minute, EBs were resuspended in 2 ml of mesoderm medium, EB2 (AE6 medium supplemented with

10 ng/ml BMP4 and 5 ng/ml FGF), seeded in 1 well of a 6 well ultra low binding plate (Corning) and incubated. 24 hours later (day 2), EBs were washed in 10 ml D-PBS and, after centrifugation at 20g for 1 minute, resuspended in MK1 medium (CellGRO SCGM (CellGenix) supplemented with 100 ng/ml TPO (Bio-Techne) and 25 ng/ml SCF (Gibco)) and, split in 2 wells of a 6 well ultra low binding plate and incubate at 37 °C, 5% CO_2. Half of media volume was replaced every two days with fresh 2x MK1 medium (day 5 and 8). Ten days after transduction, EBs were collected, washed in 10 ml D-PBS at 300 g for 5 minutes and then dissociated with 1 ml of 1:1 1 mg/ml Collagenase IV (Life Technologies) and Dispase II (Life Technologies) by gently pipetting and incubated for 45 minutes at 37 °C. Cells were then washed with 10 ml D-PBS and resuspended (mixed every 2 minutes up to 10 minutes) with 300 µl of cell dissociation buffer (Life technologies), to generate a single cell suspension. Cells were washed with CellGRO SCGM (CellGenix) at 300 g for 5 minutes, resuspended with MK2 medium (CellGRO SCGM supplemented with 20 ng/ml TPO (Bio-Techne) and 50 ng/ml SCF (Gibco)), cells were then cultured in suspension on tissue culture plates or flasks. Half of the MK2 medium was changed every two days until cells were used for further studies (maximum 25 days post transduction).

For the 2D system, a sub-confluent hiPSC culture was dissociated to single cell with TrypLE (Thermo Fisher Scientific) and $2.5x10^5$ cells resuspended in AE8 medium supplemented with 10 µM ROCK inhibitor Y-27632 (Sigma-Aldrich) were seeded onto vitronectin coated 6 well plates and incubated at 37 °C, 5% CO_2. After 24 hours, cells were washed with D-PBS and transduced with GATA1, FLI1 and TAL1 lentiviral vectors (MOI 20, Vectalys) suspended in transduction media (AE6 supplemented with 20 ng/ml of FGF, 10 ng/ml BMP4 and 10 µg/ml protamine sulphate). 4 hours post-transduction (day 1), medium was removed, cells were washed with D-PBS and mesoderm medium (AE6 medium supplemented with 10 ng/ml BMP4 and 20 ng/ml FGF) was added. After 24 hours (day 2), cells were washed with D-PBS and fed with MK medium (CellGRO SCGM supplemented with 20 ng/ml TPO (Bio-Techne) and 25 ng/ml SCF (Gibco)). On days 5 and 8 half of media volume was replaced by new 2x MK medium. On day 10, cells in suspension were collected and the adherent cell layer was split with TrypLE (Thermo Fisher Scientific) at 37 °C for 10 minutes. All cells were pooled and washed with 10 ml D-PBS at 300 g for 5 minutes. Cells were resuspended in MK medium (CellGRO SCGM supplemented with 20 ng/ml TPO (Bio-

Techne), 25 ng/ml SCF (Gibco)) and cultured in suspension tissue culture plates or flasks. Half of MK medium was changed every two days until cells were used for further studies (maximum 25 days post-transduction). For both 3D and 2D systems, MK maturation was monitored by flow cytometry from day 10 up to day 25.

2.3.3.1 MK maturation analysis by flow cytometry

Cells were directly stained with a mix of CD235a, CD41a and CD42a (APC, PE and FITC) for 20 minutes in the dark at RT. Cells were washed in flow buffer (D-PBS (Sigma-Aldrich), 0.5% BSA and 2mM EDTA) and resuspended in flow buffer supplemented with 0.5% PFA and 0.1 µg/ml Hoechst 33342 (Thermo Fisher Scientific) to gate viable cells. An unstained mix and isotype controls were used to visualise cells and set gates. Single stained AbC Total Compensation Capture beads were used for compensation (A10497, Thermo Fisher Scientific). Flow cytometry was performed using Beckman Coulter Gallios Cytometer and analysed with Kaluza Analysis v.1.5a (Beckman Coulter). Results were expressed as the percentage of positive cells for each marker and/or double positive cells. Gating strategy representation and flow cytometry analysis are shown in Figure 2.3.

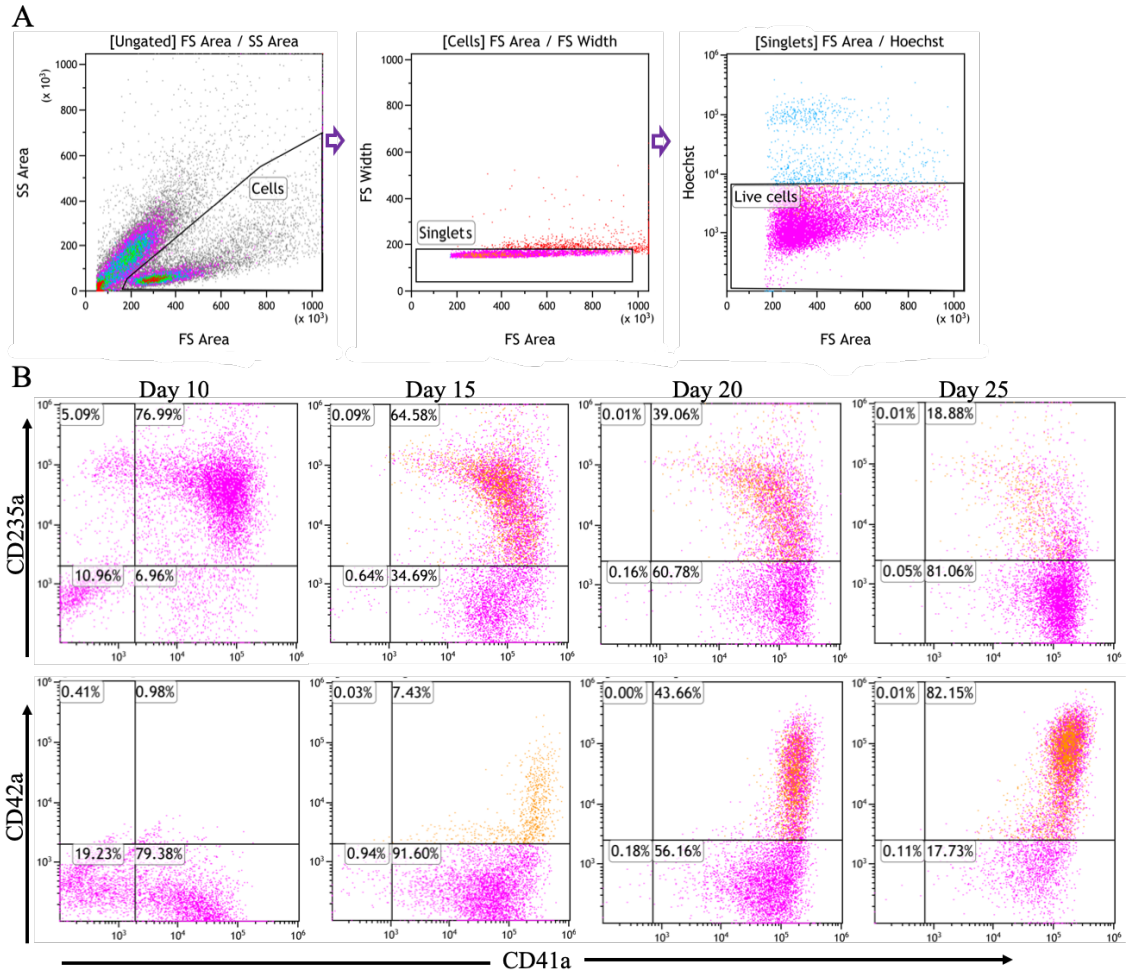


Figure 2.3 - Flow cytometry analysis of hiPSC forward programming into MKs. **A**- Gating strategy used for the analysis of forward programming. First, cells were gated according to their size and granularity using forward side scatter. Then, singlets were gated using forward scatter area and width. Finally, live singlets were gated using Hoechst staining. Expression (%) of MK markers was then determined in live singlets. **B** - Representative time course of 2D forward programming of Bobc-TAP-SMIM1 into MKs, showing MK progenitor (%CD41a$^+$/CD235a$^+$ cells) and MK maturation (%CD42a$^+$ cells) from whole culture.

Table 2.8 -Antibodies used for iMKs flow cytometry.

Antibody	clone	Manufacturer	Dilution
Mouse anti-Human CD41a-APC	HIP8	559777, BD Pharmingen	1:20
Mouse anti-Human CD42a - PE	ALMA.16	558819, BD Pharmingen	1:20
Mouse anti-Human CD42b – PE/APC	HIP1	555473, 551061 BD Pharmingen	1:20
Mouse anti-Human CD235a –PE	GA-R2	555570, BD Pharmingen	1:50
Mouse anti-Human CD235a –FITC	GA-R2	559943, BD Pharmingen	1:50
Mouse IgG2b, κ isotype -FITC	27-35	555742, BD Pharmingen	1:20
Mouse IgG1, κ isotype -APC	MOPC-21	555751, BD Pharmingen	1:20
Mouse IgG1, κ isotype -PE	MOPC-21	555749, BD Pharmingen	1:20

2.3.3.2 Ploidy analysis by flow cytometry

Fixed cells with 4% PFA for 10 minutes at RT were stained with CD41a and CD42a as described above and then incubated in PBS 0.1% Tween with 1 µg/ml Hoechst 33342 (Thermo Fisher Scientific) for 15 min at RT. Samples were processed on Beckman Coulter Gallios Cytometer and analysed with Kaluza Analysis v.1.5a (Beckman Coulter). Ploidy was assessed by gating positive cells for Hoechst and double positive for CD41 and CD42.

2.3.4 Overexpression of SMIM1 with a tandem affinity purification (TAP) in human hiPSCs

2.3.4.1 Cloning of TAP-SMIM1

TAP-SMIM1 insert was synthesized by IDT® and digested with NdeI and PacI. Digested insert was cloned into pWPI-TAP vector digested with NdeI and PacI (gift from Dr Jose Guerrero, Research Associate in Prof. Ouwehand group). Ligation reaction (section 2.3.5.3) was transformed in DH5α competent cells (section 2.3.5.1). Plasmid sequence was confirmed by sanger sequencing (section 2.3.5.7) using specific primers (Table 2.10), and the plasmid was expanded as described in section 2.3.5.2.

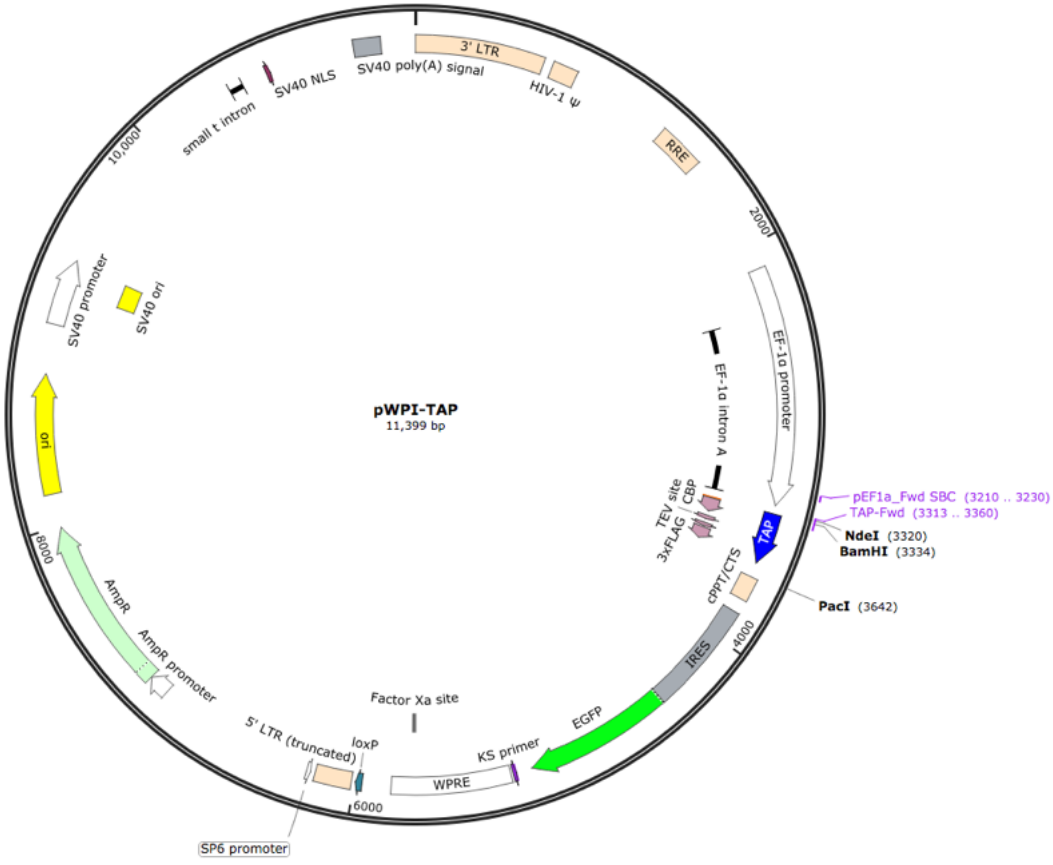


Figure 2.4 - Plasmid used for overexpression of TAP-SMIM1.

2.3.4.2 TAP-SMIM1 lentivirus production and transduction in human hiPSCs

Lentivirus production was undertaken in HEK293T cell line (laboratory stock). In brief, $4x10^6$ HEK293T cells were cultured in complete medium (DMEM (D6429, Sigma-Aldrich) supplemented with 10% foetal bovine serum (FBS) heat inactivated (Sigma-Aldrich) and 1% penicillin-streptomycin (Thermo Fisher Scientific)). Four hours after, 20 µg of pWPI-TAP-SMIM1 or pWPI-TAP were individually mixed with packaging system (8 µg PAX2 and 2.5 µg VSVG) in 1 ml of DMEM (D6429, Sigma-Aldrich) and 75 µl 1 mg/ml PEI (Sigma-Aldrich) and incubated at RT for 20 minutes. The mixture was spread drop wise over the cells and cells were incubated at 37 °C, 5% CO_2. 24 hours post-transduction the medium was changed. Virus were collected 48 and 72 hours post-transduction and concentrated using Lenti-X™ Concentrator (Clontech). Virus were resuspended in AE8 medium and stored at -80 °C. 60%-70% confluent hiPSCs were transduced with different volumes of virus, expanded and sorted for GFP positive cells using the BD FACSDiva 8.0.1 from NIHR Cambridge BRC Cell Phenotyping Hub.

2.3.5 Preparation, manipulation and analysis of DNA

2.3.5.1 Bacterial culture and transformation

Bacteria were cultured at 37 °C for 9–18 hours either in shaking liquid culture (Luria Bertani (LB) medium) or solid culture (streaking on LB agar). When required, cultures were supplemented with 100 µg/ml of ampicillin.

Chemically *DH5α* high-competent cells (NEB, C2987H) were incubated with DNA on ice for 20 minutes. Cells were heat-shocked for 45 s at 42 °C and returned to ice for 2 minutes. Then, 600 µl LB was added and bacteria were cultured for 60 minutes in a shaker at 37 °C, before plating onto LB agar with appropriate antibiotic.

2.3.5.2 Isolation of plasmid and genomic DNA

Bacterial cells from liquid culture were pelleted and plasmid DNA using QIAprep Spin Miniprep Kit (Qiagen, 27104) or with the PureYield™ Plasmid Maxiprep System (Promega, A2393), according to the manufacturer's instructions. Genomic DNA of hiPSCs was isolated and purified using Wizard® SV Genomic DNA Purification Kit (Promega, A2360), according to manufacturer's protocol.

2.3.5.3 Restriction endonuclease digestion and DNA ligation

DNA was digested at 37 °C for 1-3 h with restriction enzymes (NEB and Fermentas) in the provided buffers, according to the manufacturer's instructions. Typically, 10 units of enzyme were used per 1 μg of DNA. Cut and uncut fragments were analysed on an agarose gel. DNA fragments were ligated as required with 3:1 or 7:1 molar ratio of fragment to vector using T4/T7 DNA ligase (NEB, M0202) or Quick Ligase™ kit (NEB, M2200). Reactions were left for 30 min at 16 ºC when used T4/T7 DNA ligase and 10 min at RT when used Quick Ligase™ Kit.

2.3.5.4 Polymerase chain reaction and colony PCR

For a typical PCR reaction, 10 μM of forward and reverse primer were used along with 2x MyTaq (Bioline, BIO-25041) or Phusion PCR kit (NEB, E0553S). PCR reactions were heated to 95 °C for 2 min, then cycled between 95 °C for 10 s, T_m -5 °C for 30 s and 72 °C for 30 s/kb followed by a 5-min final extension step on a thermocycler Veriti® Thermal Cycler (AppliedBiosystems). Reactions were then analysed on agarose gels. More specifically, to avoid non-specific amplifications a touchdown PCR was used for the amplification of *SMIM1* gene: 95 °C for 3 min, then 15 cycles of 95 °C for 30 s, 70 °C for 1 min Touchdown 1 degree/cycle and 72 °C for 30 s followed by 25 cycles of 95 °C for 30 s, 55 °C for 1 min and 72 °C for 30 s and final extension step of 72 °C for 5 min. To screen bacterial colonies, individual colonies were picked and placed in 100 μl miliQ water. PCR was carried out as above with ~5 μl colony suspension. The remaining ~95 μl from the positive colonies were used to inoculate liquid culture.

2.3.5.5 Purification and gel extraction of DNA

As required, DNA was purified by selective binding to a silica membrane using QIAGEN QIAquick PCR Purification Kit® (Qiagen, 28104) or QIAquick Gel Extraction Kit® (Qiagen, 28704), according to the manufacturer's instructions. Resolved DNA bands were visualised through an orange filter under 470-nm light in a transilluminator. DNA bands were excised and agarose melted at 50 °C according to the manufacturer's instructions.

2.3.5.6 Agarose gel electrophoresis

Nucleic acids were mixed with loading buffer (NEB, B7024S), resolved on 1% or 1.5% agarose in 1X Tris-Borate-EDTA (TBE) buffer and supplemented with SYBR® Safe DNA Gel Stain (Life Technologies, S33102). GeneRuler 1 kb DNA ladder (Life

Technologies, SM0312) was used. Typically, gels were run at 90 V in 1X TBE. Nucleic acid was visualised with trans-UV light and photographed.

2.3.5.7 Determining the concentration of nucleic acid and sequencing

Nucleic acid concentration was determined by its A_{260} with an ND-1000 spectrophotometer (NanoDrop Technologies, Inc.). An extinction coefficient of 50 was used for double strand DNA. Purity was determined by the profile of the sample on an agarose gel and/or the $A_{230:260:280}$. Sanger sequencing with appropriate primers was done by Source Bioscience service. Sequence files were blasted in the NCBI tool (www.blast.ncbi.nlm.nih.gov/) and were aligned with the multiple alignment server clustalW2 (www.ebi.ac.uk/Tools/msa/clustalw2/) or the SNAPgene software using human genome GRCh37 as the reference genome.

2.4 Solutions and primers

2.4.1 Solutions

Solutions were autoclaved or filtered with a 0.22 μM PES corning filter as required. Solutions were stored at RT or 4 °C (*) (Table 2.9).

Table 2.9 - List of solutions.

Name	Composition
MEF Medium*	DMEM (Invitrogen, 11965118), 10% v/v FBS (Hyclone, SH30070.01), 2 mM L-Glutamine (Invitrogen, 25030156)
AE6 *	DMEM/F12 (Invitrogen, 11330057), 2X Insulin–Transferrin-Selenium (Invitrogen 41400-045), 0.054% $NaHCO_3$, 7.5% (Life technologies, 25080-094), 320 μg/ml L-Ascorbic acid 2-phosphate sesquimagnesium salt hydrate (Sigma-Aldrich, A8960-5G)
AE8 *	AE6, 5 μg/ml bFGF2 (Cambridge Stem Cell Institute, Tissue culture facilities), 10 μg/ml hActivin A (Cambridge Stem Cell Institute, Tissue culture facilities)
Wash Medium*	Advanced DMEM/F12 (Life Techonologies)
Freezing Medium*	90% v/v KOSR (Invitrogen, 10828028), 10% v/v dimethyl sulfoxide (DMSO, Sigma-Aldrich, D-2650)
HBS*	150 mM sodium chloride, 5 mM potassium chloride, 1 mM magnesium sulphate, 10 mM HEPES, pH 7.4 – all from Sigma-Aldrich.
Tyrode's buffer*	134 mM sodium chloride, 2.9 mM potassium chloride, 0.34 mM sodium phosphate dibasic, 12 mM sodium bicarbonate, 20 mM HEPES, 1mM magnesium chloride and 5 mM glucose (VWR), pH 7.3 – all the other from Sigma-Aldrich
ACD*	111mM glucose (VWR), 71mM citric acid, 116mM sodiumcitrate (Sigma-Aldrich)
0.2 % formyl saline	0.2% formaldehyde, 0.85% sodium chloride (Sigma-Aldrich)
TBE 1x	89 mM Tris base, 89 mM boric acid 2 mM EDTA

2.4.2 Primers

Purified primers (Sigma-Aldrich) were resuspended to 100-µM stock concentrations and stored at -20 °C. Typically, 10 µM primers were used per reaction (Table 2.10).

Table 2.10 -Primers used in this study.

Primer name	Fwd/ rev	Sequence (5' - 3')	Target/used for
SMX1F2	fwd	CCCTTAGTGCCCCTCTCCTA	SMIM1 exon 3 - PCR/Sequencing
SMIM1_4b-R	rev	GTGTGCCCAGCTGTTTGTAGGT	SMIM1 exon 4 - PCR/Sequencing
SMX1R1 _3d-R	rev	GTTTGCTGAGCTGCTGGAC	SMIM1 exon 3 - PCR/Sequencing
SMIM1_CR_1_Bottom	fwd	AAACAGGAGAGCCACGTCCACTATC	Guide seq. +BbsI
SMIM1_CR_1_Bottom	rev	CACCGATAGTGGACGTGGCTCTCCT	Guide seq. +BbsI
SMIM1_CR_2_Top	fwd	CACCGCAGGGACGGAGTCAGCCTAG	Guide seq. +BbsI
SMIM1_CR_2_Top	rev	AAACCTAGGCTGACTCCGTCCCTGC	Guide seq. +BbsI
SMIM1_CR_3_Bottom	fwd	AAACCTATAGTAGGTGGGAGGACGC	Guide seq. +BbsI
SMIM1_CR_3_Bottom	rev	CACCGCGTCCTCCCACCTACTATAG	Guide seq. +BbsI
SMIM1_CR_4_Top	fwd	CACCGGCAGGGACGGAGTCAGCCTA	Guide seq. +BbsI
SMIM1_CR_4_Top	rev	AAACTAGGCTGACTCCGTCCCTGCC	Guide seq. +BbsI
SMIM1_CR_5_Bottom	fwd	AAACCGTCCACTATAGTAGGTGGGC	Guide seq. +BbsI
SMIM1_CR_5_Bottom	rev	CACCGCCCACCTACTATAGTGGACG	Guide seq. +BbsI
U6-Forward	fwd	GAGGGCCTATTTCCCATGATTCC	Sequencing PX462
hGATA4 -Rev	rev	ATTGTGGATGAATACTGCC	Sequencing PX462
TAP_Fwd	fwd	GGATATCATATGCCCGGGCAAGGATCCAT GGAGAAGAGGCGGTGGAAG	Sequencing pWPI
pEF1a_Fwd	fwd	ATTTGCCCTTTTTGAGTTTGG	Sequencing pWPI
T7	fwd	TAATACGACTCACTATAGGG	PCR/Sequencing
Sp6	rev	ATTTAGGTGACACTATAGAA	PCR/Sequencing

2.5 Statistics

Data are presented as mean ± standard deviation (SD). Statistical analysis was performed using GraphPad Prism 7 scientific, R (version: 3.4.4, packages: carData and caTools). Data normality was tested using D'Agostino & Pearson and/or Shapiro-Wilk test and/or diagnostic Q-Q plots as required. For data following normal distribution, one-way ANOVA with Tukey's HSD test was used to compare the differences between three experimental groups. Studies with two experimental groups were evaluated using unpaired two-way Student's t-test. Non-normal distributed data were analysed by non-parametic tests: Kruskal-Wallis with Dunn’s multiple comparison test, when comparing three groups and Mann-Whitney, when comparing two groups. When required multivariable analysis was performed using additive models. Statistical analysis of unbalanced additive models was performed using Anova type II. Results were considered statistically significant if P-value < 0.05.

CHAPTER 3 | MOLECULAR CHARACTERISATION OF SMIM1 GENE EXPRESSION AND PROTEIN LOCALISATION

3.1 Introduction

SMIM1 is a small transmembrane protein of unknown function that was identified as the gene encoding the Vel antigen, a universal antigen on the surface of RBCs [182, 186, 188]. The SMIM1 protein is formed by 78 amino acids, has a theoretical molecular weight of 8,7 kDa (Figure 3.1) and its preliminary characterisation has been performed in human RBC cell membrane and erythroid cell lines by genetic, proteomic and biochemistry studies.

Proteomic and biochemistry studies suggested that SMIM1 is a type II membrane protein, in which the transmembrane domain functions as a non-cleavable signal peptide and is located towards the C-terminus [182, 208]. The N-terminus of SMIM1 was initially predicted to reside in the extracellular domain [188]. However, recent biochemical studies described by Arnaud *et al.* have shown that the predicted SMIM1 C-terminal (67 to 76 amino acid) is the extracellular domain that forms the Vel antigen (Figure 3.1) [182, 188, 208]. Flow cytometry analysis of erythroleukemia cells expressing N-or C-terminally Flag-tagged SMIM1, as well as a C-terminal mutant of SMIM1 protein and a KELVEL chimera confirmed that the C-terminus was crucial for Vel-antigen exposure [208]. Moreover, proteomic and biochemistry studies showed that SMIM1 N-terminal undergoes post-translational modifications. Acetylation and phosphorylation of different residues in the N-terminal of SMIM1 were also identified (Figure 3.1) [208-211]. However, studies performed with the KELVEL chimera and with different N-terminal SMIM1 mutants suggested that the SMIM1 N-terminal is not crucial for Vel-antigen exposure [208].

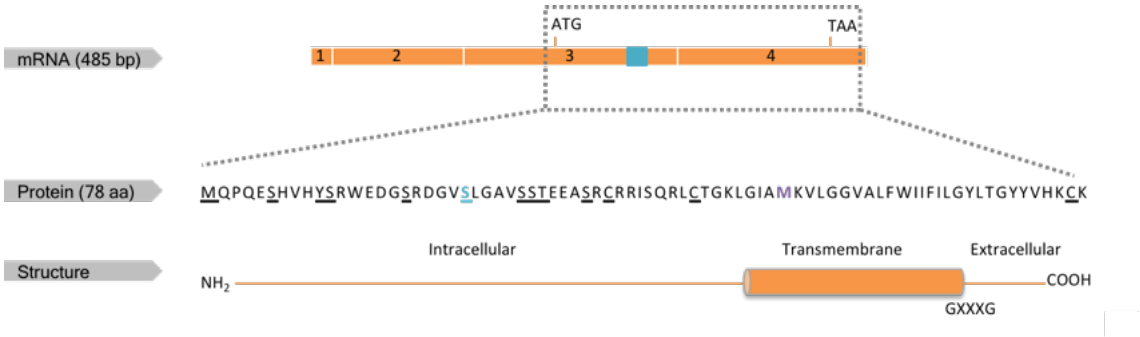


Figure 3.1 – Schematic representation of SMIM1 at mRNA and protein level. The four exons by which it is composed are represented by orange boxes and the coding region is delineated by dash lines. SMIM1 protein sequence with underline of 1) three-cysteine residues potentially involved in complex formation; 2) serine, tyrosine and threonine residues subjected to phosphorylation; and 3) methionine subject to acetylation. Serine and methionine highlighted in blue and purple respectively, represents the amino acids where the 17-bp frameshift deletion (blue) and the missense mutation (purple) responsible for the Vel-negative and Vel-weak phenotype occur. Representation of protein structure displaying transmembrane domain and GXXXG motif at amino acid position 67 to 71 (predicted to mediate the transmembrane dimerization) (Adapted from [188]).

Despite disagreement on SMIM1 molecular weight and membrane orientation among the literature, its ability to form oligomers is well established [182, 188, 208, 212]. In the literature, it has been suggested that SMIM1 has the ability to form oxidation-dependent molecular complexes (homodimers or others) through the three cysteine residues present in SMIM1 (Cys35 and Cys43 in the N-terminal half, and Cys77 in the C-terminal) [182, 188, 208]. Furthermore, it has been also suggested that oxidation-independent dimerization might be mediated by the GxxxG motif in the transmembrane domain via helix-helix association [188].

Different multiprotein complexes have been identified in the RBC membrane, such as the Band 3 complex and the Rh complex. These multiprotein complexes are functionally regulated by a hierarchy of protein expression dependencies [213]. Since SMIM1 is a membrane protein, Haer-Wigman *et al.* investigated if Vel/SMIM1 expression on RBC membrane was dependent on the expression of other important RBC membrane proteins and vice-versa [187]. However, analysis of SMIM1 expression level on RBCs of individuals lacking different blood group systems and analysis of different important RBC membrane proteins in Vel-negative, weak and positive individuals, revealed that SMIM1 expression level on RBCs is not dependent and does not affect the expression levels of the RBC membrane proteins tested so far [187].

As mentioned above, SMIM1 was first described as a protein on RBC membrane, therefore, the biochemical characterisation of SMIM1 performed so far was done in RBC cell membrane or in erythroid cell line models. Nevertheless, RNA-sequencing of human blood cell progenitors and precursors showed that *SMIM1* transcripts are highly expressed in other blood cell types: MEP and MK [186]. However, *SMIM1* is also expressed in non-haematopoietic tissues as shown in Figure 3.2 [188]. Testis, atrial appendage of heart and pituitary are the human tissues with highest expression levels when examining the data present in GTEx portal (https://gtexportal.org/home/). SMIM1 expression in non-haematopoietic tissues and other blood cells is still poorly characterised.

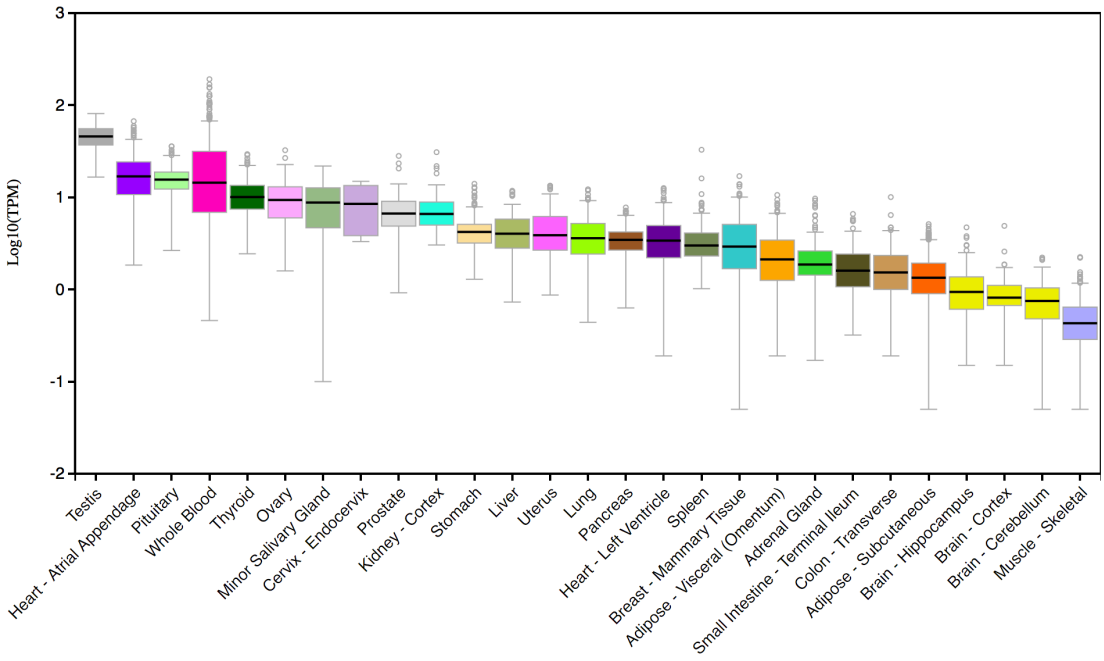


Figure 3.2 – *SMIM1* expression in different human tissues. Representation of Log10 TPM- transcripts per kilobase million of *SMIM1* across different human tissues. This data/figure was obtained from https://gtexportal.org/home/.on 2018.10.25.

3.2 Chapter aims and overview

In this chapter I had two aims: i) to characterise *SMIM1* gene expression in the haematopoietic lineages and ii) further investigate SMIM1 protein subcellular localisation in other blood cells expressing SMIM1.

First, I have described *SMIM1* expression throughout the haematopoietic lineages and showed that SMIM1 is specifically expressed in naïve, memory and class-switched B-lymphocytes, neutrophils, EB, MKs and platelets by RNA-sequencing analysis and

western blot. Using flow cytometry and immunofluorescence microscopy, I have further shown that SMIM1 subcellular localisation in the white blood cell (WBC), RBC, MKs and platelets seems to differ.

3.3 Results

3.3.1 *SMIM1* is expressed in different cell-types

Using RNA-sequencing data of primary human blood precursor and progenitor cells obtained as part of the Blueprint project [197], I investigated *SMIM1* expression level across the haematopoietic system. Firstly, I have observed that human *SMIM1* is expressed at a low level throughout the haematopoietic tree (Figure 3.3). An increase of *SMIM1* expression level was observed towards the myeloid branch. Additionally, analysis of expression patterns of *SMIM1* and other genes encoding membrane proteins that are important for MK/platelet (glycoprotein IV - *GP6* and Integrin subunit alpha 2b - *ITGA2B* (CD41b) [214]*)* and EBs/RBC (glycoprotein A, GYPA [215]) biological functions, suggested the relevance of SMIM1 in MK and EB specification and therefore possibly in the biological function of these cells (Figure 3.3).

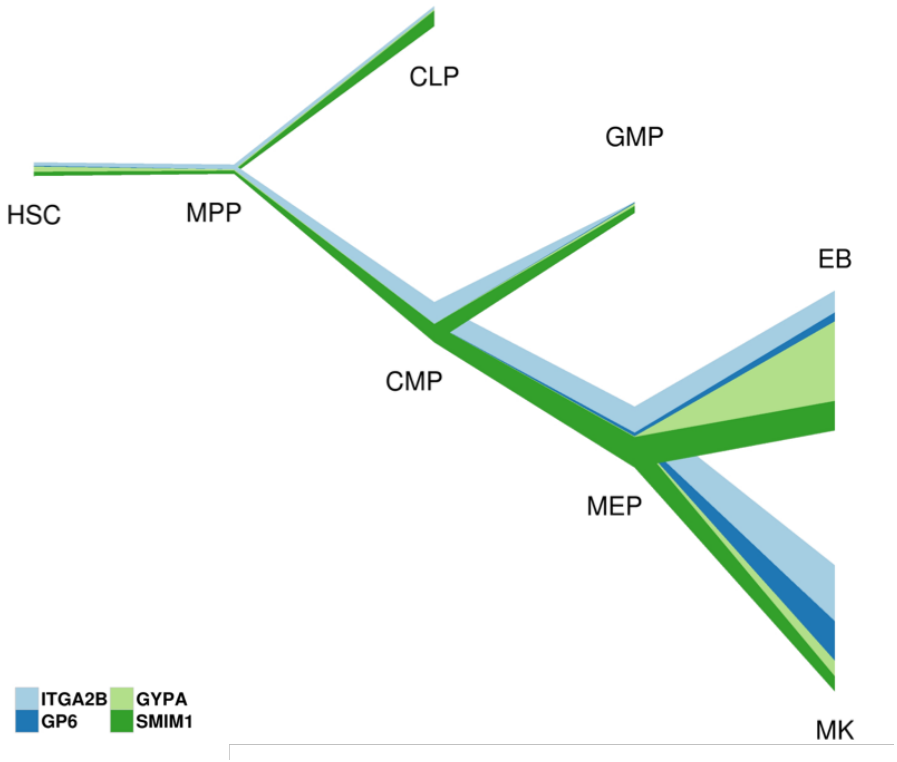


Figure 3.3 – Overview of *SMIM1* expression in haematological lineages. Expression of *SMIM1* (dark green), *ITGA2B* (light blue), *GP6*(dark blue), and *GYPA* (light green) transcripts in the haematological lineage using RNA-sequencing data from the Blueprint project [197]. *SMIM1* is expressed throughout blood lineages but is highly expressed in myeloid lineage, specifically in MEP, as well as in EBs. Thickness of line represents the expression level. HSC- haematopoietic stem cell; MPP- multi-potent progenitor CMP- common myeloid progenitor; CLP- common lymphoid progenitor; GMP- granulocyte-macrophage progenitor; MEP- megakaryocyte-erythroid progenitor; EB- erythroblast and MK- megakaryocyte. Data obtained through https://blueprint.haem.cam.ac.uk using tools-river plot.

Secondly, I performed a detailed analysis of *SMIM1* expression level in the different blood cell types. Here, I observed that among other cells, EB, MK and platelets were the cells with higher *SMIM1* expression (Figure 3.4). *SMIM1* expression was also detected in naïve, memory and class-switched B-lymphocytes, and neutrophils.

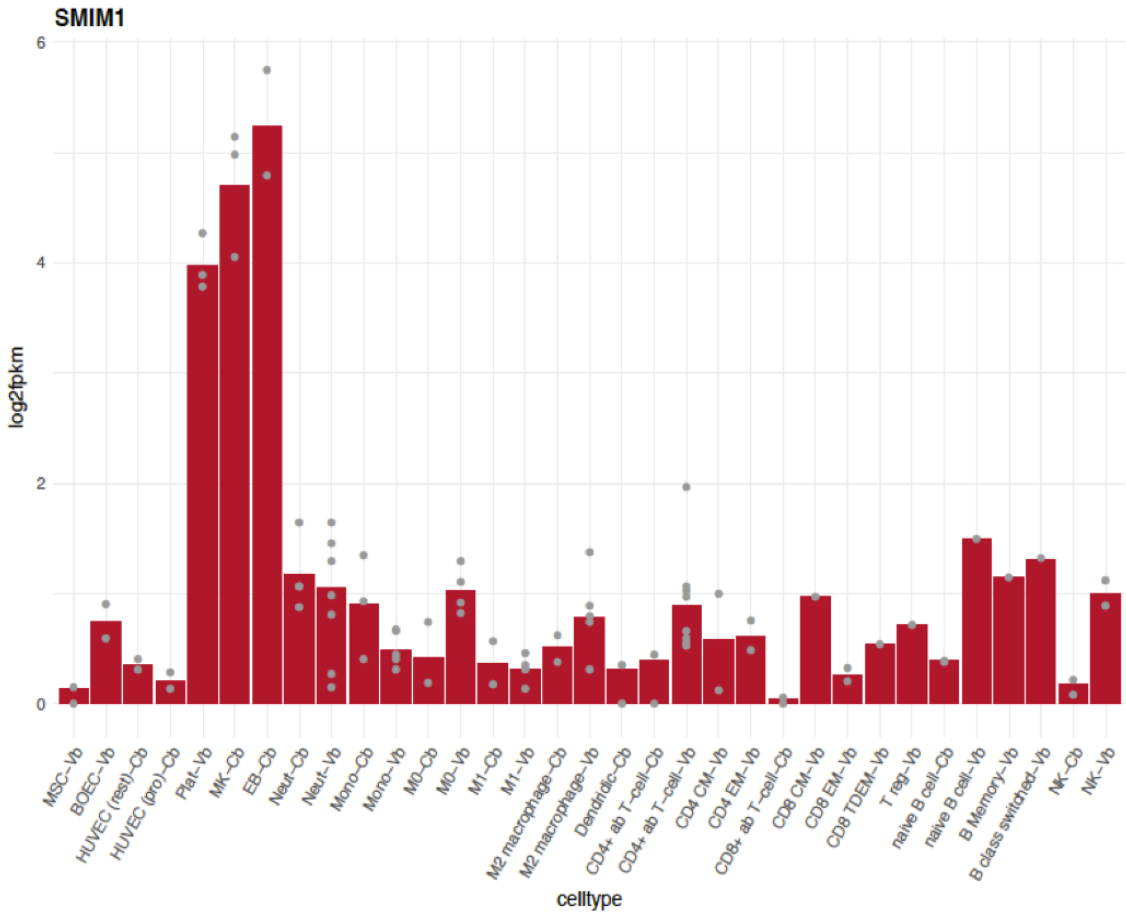


Figure 3.4 - *SMIM1* expression level in blood cells. Expression levels of *SMIM1* larger transcript in different blood cell types based on RNA-sequencing data from the Blueprint project [197]. *SMIM1* is expressed in neutrophils, and naïve, memory and class-switched B-lymphocytes, but it is highly expressed in erythroblasts, megakaryocytes and platelets. Grey dots- biologic replicates for each cell type; Log2 (Fpkm+1)- fragments per kilobase million; Vb- venous blood; Cb- cord blood; MSC- mesenchymal stem cells; BOEC- blood outgrowth endothelial cells; HUVEC- human umbilical vein endothelial cells; Plat- platelet; MK- megakaryocyte; EB- erythroblast; Neut- neutrophil; Mono- monocytes; M0- macrophage; M1- macrophage M1; CD4+ T-cell- T helper cell; CM- central memory T cells; EM- effector memory T cell; CD8+ T cell- cytotoxic (killer) T cell; CD8 TDEM- terminally differentiated effector memory T cell ; T-reg- regulatory T cell; NK-natural killer cell. Data obtained through Blood RNAexpress- https://blueprint.haem.cam.ac.uk/mRNA/.

As *SMIM1* encodes three transcripts (described in section 1.5.1.1), I then characterised the expression of the three different *SMIM1* transcripts in the blood cells with higher *SMIM1* expression (EB, MKs and platelets). The expression of the three *SMIM1* transcripts was detected in EB, MK and platelets (Figure 3.4). However, the larger transcript, encoding the 78 amino acid protein, was the most abundant across the three cell types (Figure 3.5 - A).

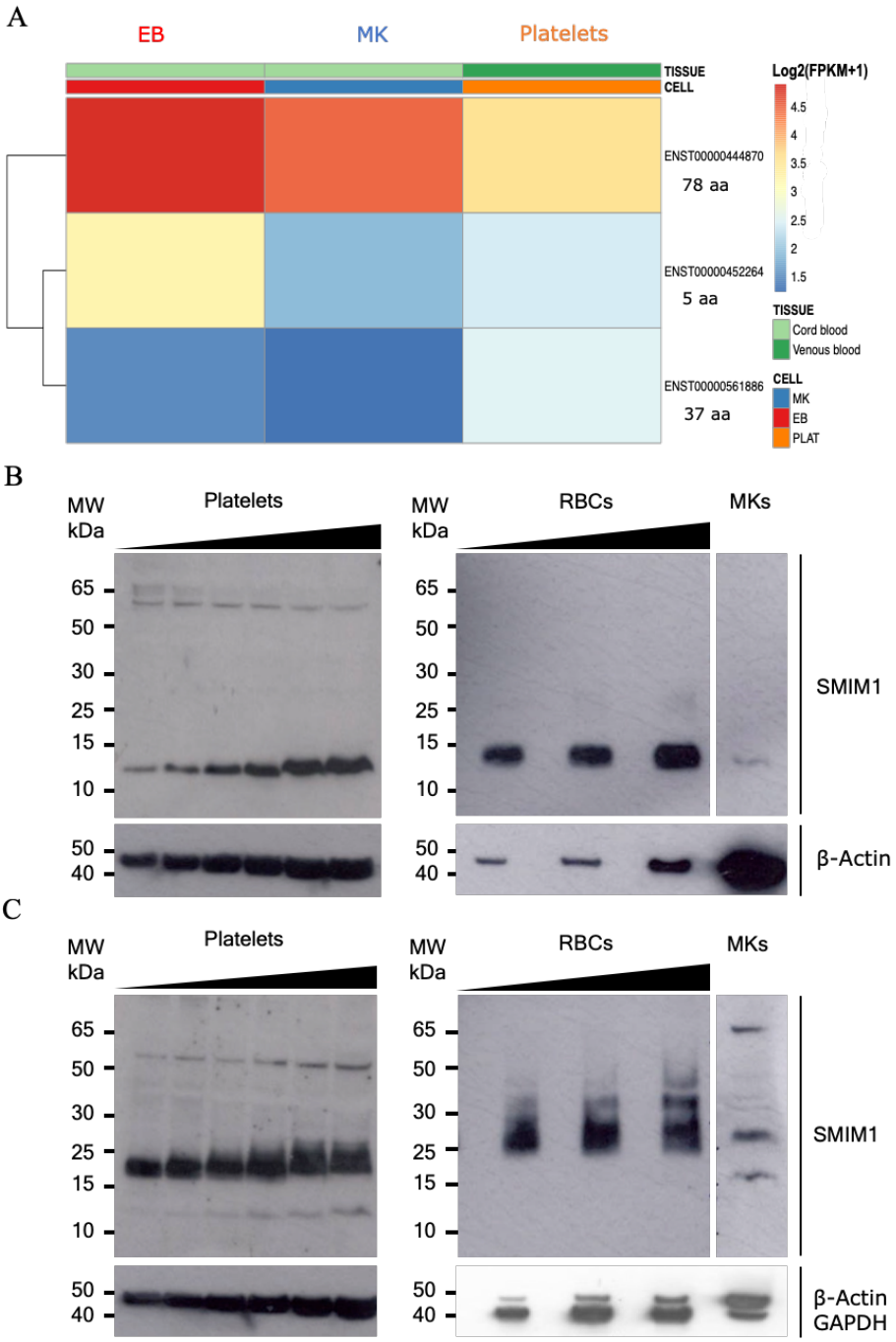


Figure 3.5 - Expression of SMIM1 at mRNA and protein level. A - Heat map of expression level (Log2(fpkm+1)) of the three human *SMIM1* transcripts in erythroblasts (EB), megakaryocytes (MK) and platelets using RNA-sequencing data from Blueprint project [197] (heat map made by Dr Luigi Grassi – Research associate in Prof. Ouwehand group). Larger transcript, encoding 78 amino acid protein, is the highest transcript among the three cell types. Fpkm - fragments per kilobase million. **B** - Western blot in reduced conditions (lysates treated with DTT) of SMIM1 polyclonal and β-actin antibodies in MK (29 µg), platelet (increased concentration: 5, 10, 20, 30, 40 and 50 µg) and RBC lysates (increased concentration 10, 20 and 28 µg). **C** - Western blot in non-reduced conditions (lysates not treated with DTT) of SMIM1 polyclonal, GAPDH and β-actin antibodies in MK (29 µg), platelet (increased concentration: 5, 10, 20, 30, 40 and 50 µg) and RBC lysates (increased concentration 10, 20 and 28 µg). In non-reduced conditions SMIM1 monomer is present in platelets and MKs but not in RBCs. Multimers are present across cell types.

SMIM1 protein expression was characterised by western blot. RBC, MK and platelet lysates from Vel-positive blood donors were resolved by SDS-PAGE followed by blotting in a membrane and incubation with polyclonal anti-SMIM1 antibody. MKs were obtained through *in vitro* differentiation of CD34-positive cells towards MKs. Platelet and RBC protein lysates were loaded in increasing concentrations, while MK lysate was loaded only at maximum protein concentration obtained. In agreement with *SMIM1* expression results, SMIM1 protein was detected in RBCs, MKs and platelets (Figure 3.5 - B-C). Under reducing conditions, a band of approximately 14 kDa was identified in the three cell types. This 14 kDa band corresponds to the monomer form of SMIM1 protein, which presents a higher molecular weight than its theoretical molecular weight. RBCs also presented a less abundant band of approximately 25 kDa (Figure 3.5 - B). Interestingly, under non-reducing-conditions distinct forms and distribution of SMIM1 between the three cell types was observed, suggesting formation of cell-specific protein complexes. In the platelet lysate, as seen under reducing conditions, a faint band at approximately 14 kDa was detected (Figure 3.5 - C). Additionally, higher molecular weight and more intense bands were observed between the 18 and 26 kDa range, which possibly correspond to post-translational modifications of the monomer, homodimers or other protein complexes. Several bands with different intensities between 25 kDa and 48 kDa, were detected in RBCs. On the other hand, intense bands at approximately 18 kDa, 26 kDa and 65 kDa were detected in MKs, suggesting the presence of similar molecular complexes found in platelets and RBCs, and a MK-specific molecular complex. Also, two other faint bands at approximately 30 and 35 kDa were detected in MKs (Figure 3.5 - C). The pattern of RBC western blot detected in both reducing and non-reducing conditions resembles the pattern described in previous studies, though with subtle differences in molecular weights. In summary, SMIM1 disulfide-bond-dependent complexes were detected in RBCs, MKs and platelets. However, these complexes seem to be different in the three cell types.

3.3.2 SMIM1 localisation is cell-type dependent

SMIM1 subcellular localisation in RBCs is well established [188, 208, 212]. However, its subcellular localisation in other blood SMIM1-expressing cells has not been characterised so far. Because SMIM1 has been described as a membrane protein, I started to investigate SMIM1 membrane exposure in platelets and WBCs by flow cytometry. As expected, whole blood of Vel-positive donors and of wildtype mice

incubated with the human monoclonal anti-SMIM1 antibody (that recognises the C-terminal of SMIM1 – SpG213Dc human mAb [212]) revealed SMIM1 exposure on RBC membrane (Figure 3.6 – A-B). SMIM1 membrane expression was also detected in WBCs, in both human and mice. To further explore which WBCs were expressing SMIM1, different WBC purification strategies and WBC markers were used. However, due to a high unspecific binding of the secondary antibody on purified WBCs, I was unable to obtain a conclusive result (data not shown). Interestingly, SMIM1 was not detected on the platelet's membrane by flow cytometry (Figure 3.6), suggesting that SMIM1 does not migrate to the plasma membrane in these cells or that the epitope is not recognized due to allosteric constraints.

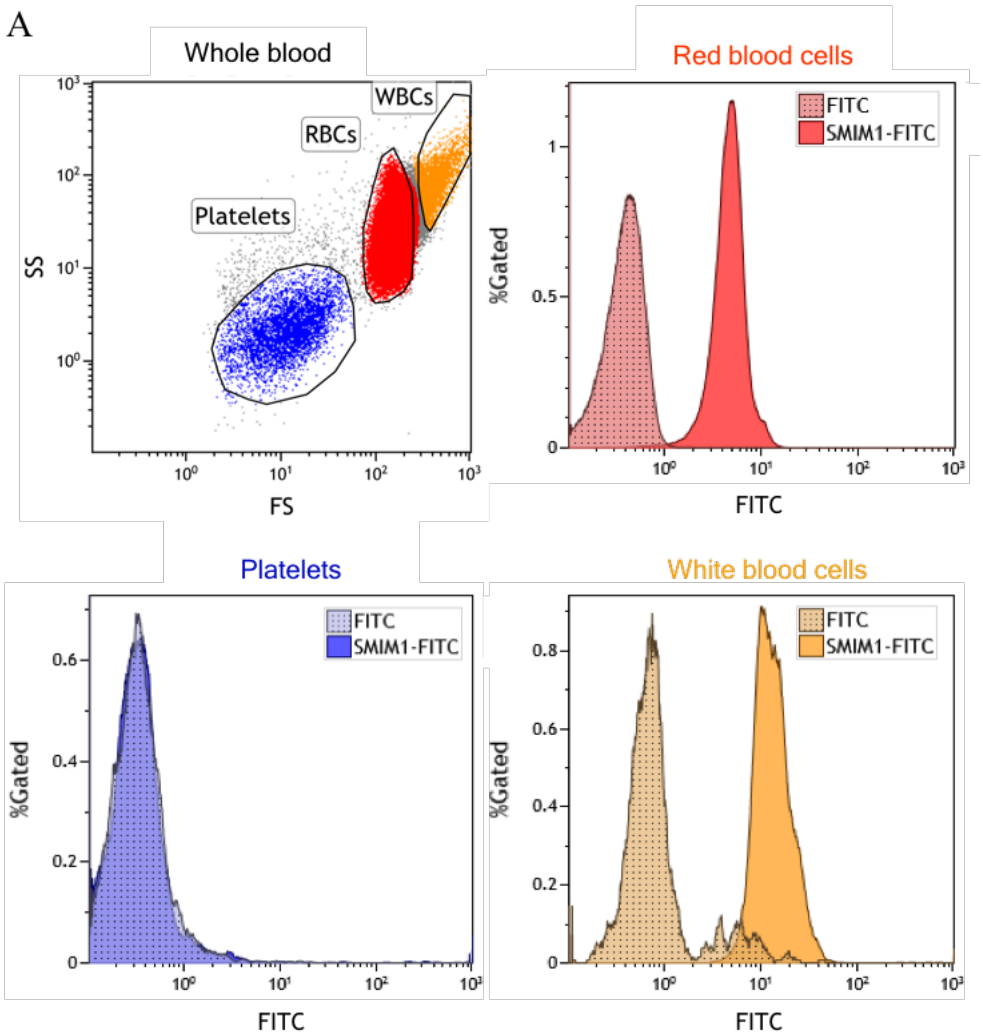


Figure 3.6 (A) - Localisation of SMIM1 in whole blood. Representative flow cytometry dot plot for forward and side scatter and histograms of SMIM1 expression for red blood cells (RBCs), platelets and white blood cells (WBCs), against secondary FITC (negative control – dotted histograms) in human whole blood.

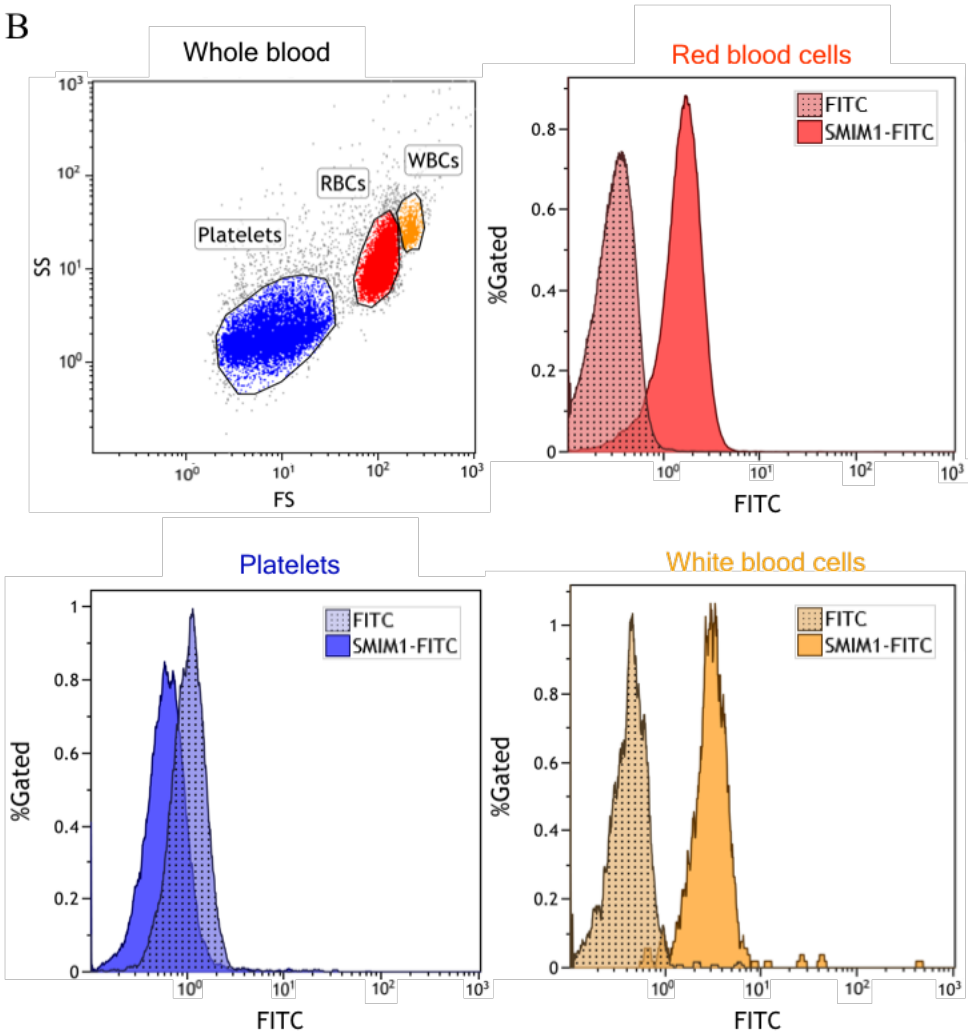


Figure 3.6 (B) - Localisation of SMIM1 in whole blood. Representative flow cytometry dot plot for forward and side scatter and histograms of SMIM1 expression for red blood cells (RBCs), platelets and white blood cells (WBCs), against secondary FITC (negative control – dotted histograms) in mouse whole blood.

Given that SMIM1 was not detected on the platelet's membrane, I then investigated SMIM1 exposure in platelet precursor cells by flow cytometry and immunofluorescence microscopy. Flow cytometry analysis of MKs derived from CD34-positive cells stained with the human monoclonal anti-SMIM1 antibody (SpG213Dc human mAb – that recognises the C-terminal of SMIM1) shown that, as in platelets, SMIM1 is not detected on MKs surface (Figure 3.7 - A). Immunofluorescence microscopy of MKs stained with the polyclonal anti-SMIM1 antibody (that recognises the N-terminal of SMIM1 - Sigma), revealed that SMIM1 is broadly expressed in MK cytoplasm and forms some cytoplasmic accumulations near to the nucleus (Figure 3.7 - B). To further investigate SMIM1 cytoplasmic accumulations and localisation in MKs, I performed co-localization studies of SMIM1 with α-granule´s proteins (THBS1) and δ-

granules/lysosomes (LAMP-3, LAMP-2) and Golgi protein (58K-Golgi). Immunofluorescence microscopy of these co-localisation studies indicated that SMIM1 does not seem to specifically co-localise with any of the analysed proteins (Figure 3.7 - C).

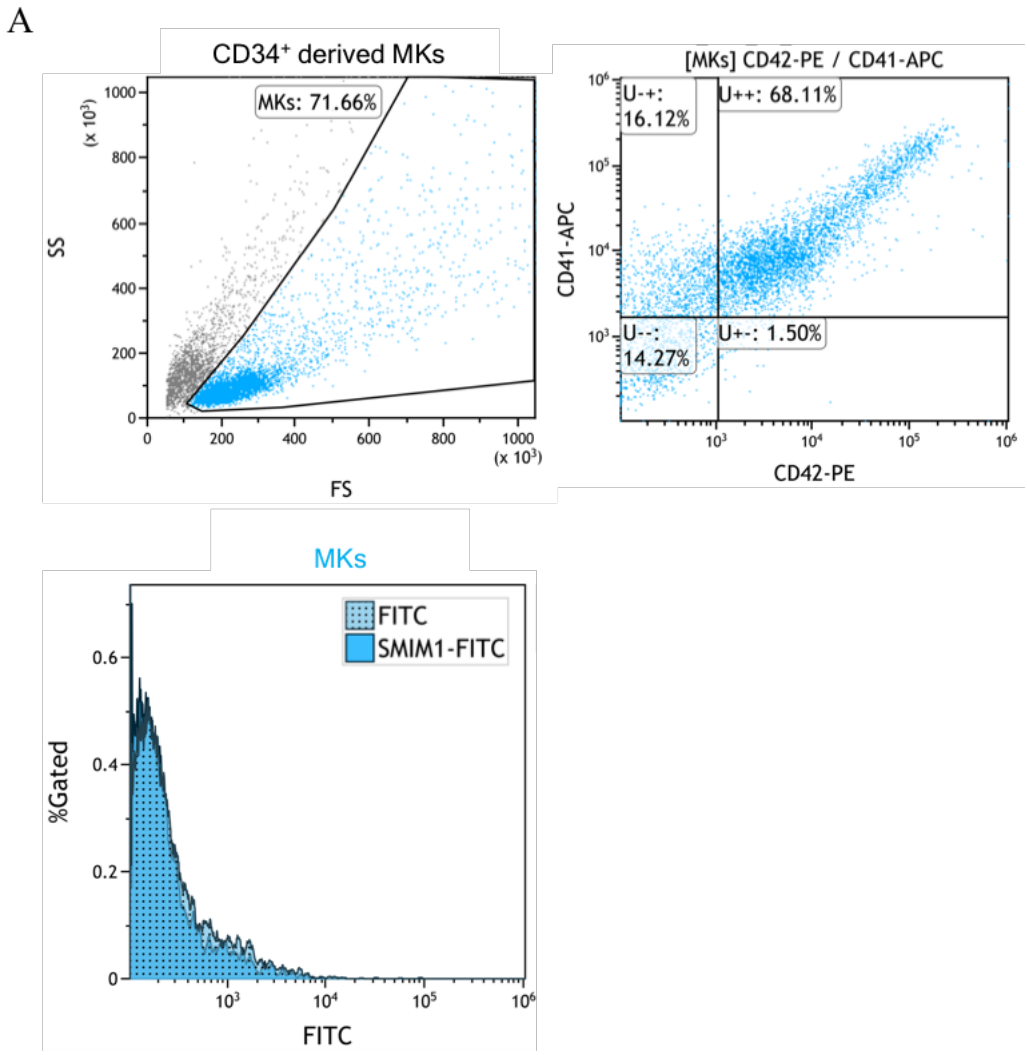


Figure 3.7 (A) – Characterisation of SMIM1 subcellular localisation in MKs derived from CD34⁺ cells. Flow cytometry analysis of day 10 MKs. MKs were gated according to cell size and granularity using forward and side scatter. 68 % of the MKs were double positive for maturation markers, CD41 conjugated with APC and CD42 conjugated with PE. SMIM1 expression was assessed in mature MKs using a human monoclonal anti-SMIM1 antibody that recognises the C-terminal of SMIM1. Overlay plot of the control sample (mature MKs stained with secondary anti-FITC antibody) and test sample (mature MKs stained with anti-SMIM1 and secondary anti-FITC) showed that SMIM1 expression was not detected on MKs membrane.

B

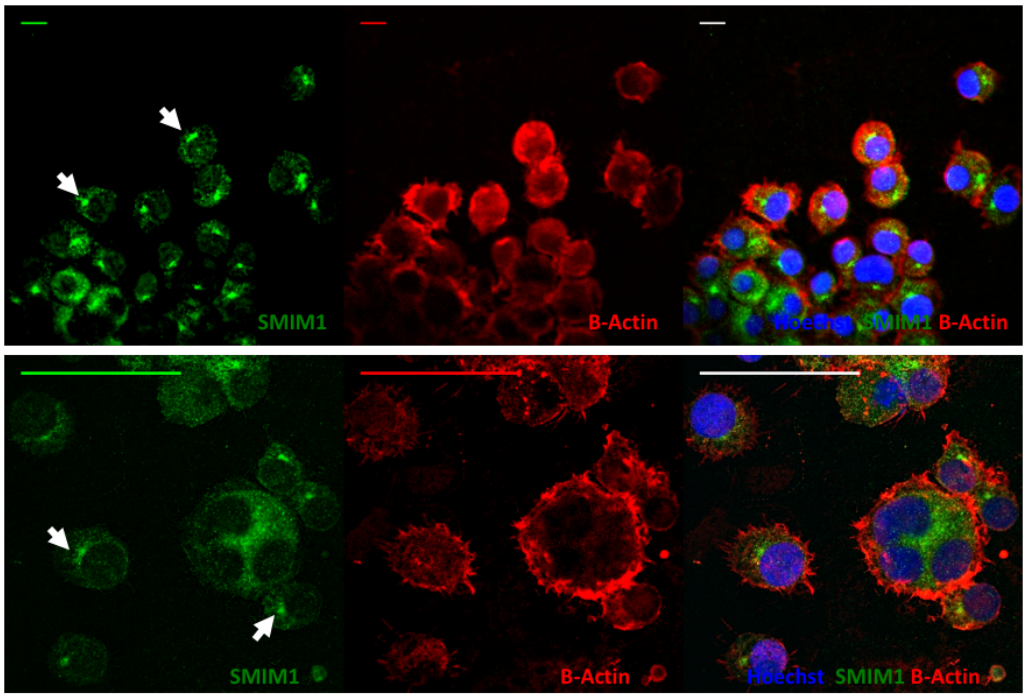


Figure 3.7 (B) – Characterisation of SMIM1 subcellular localisation in MKs derived from CD34^{+} cells. Immunofluorescence microscopy of cytospin MKs stained with the polyclonal anti-SMIM1 and secondary A488 antibody (green), β-actin and secondary A633 antibody (red) and Hoechst 33342 (DNA staining, blue), showing cytoplasmic expression of SMIM1. First panel – wide-field fluorescence microscopy; second panel - confocal microscopy. Scale bar – 50 μm.

C

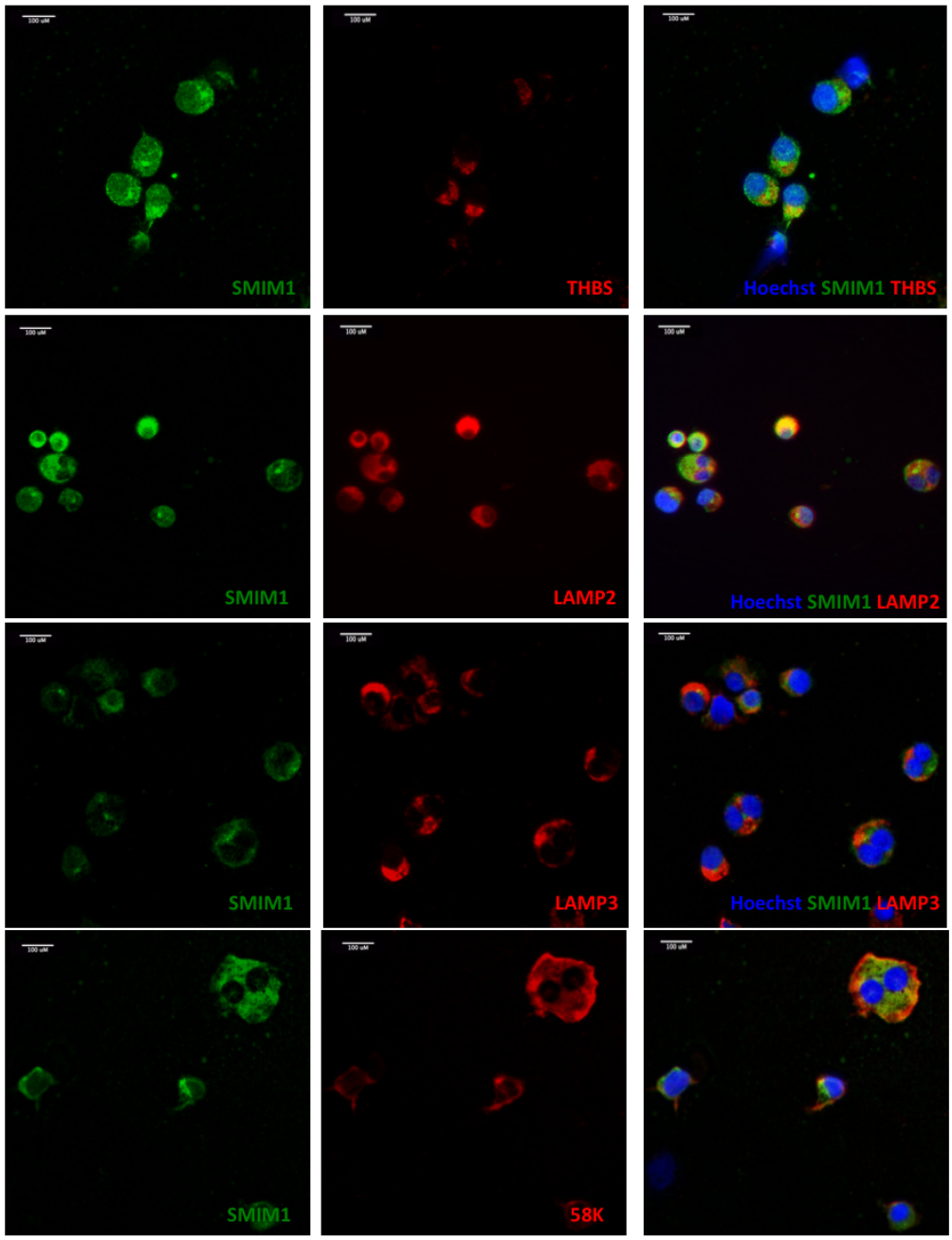


Figure 3.7 (C) – Characterisation of SMIM1 subcellular localisation in MKs derived from CD34$^+$ cells. Immunofluorescence microscopy of cytospin MKs co-stained with polyclonal anti-SMIM1-A488 (green) and THBS1-A633 (red); polyclonal anti-SMIM1-A488 (green) and LAMP 2-A633 (red); polyclonal anti-SMIM1-A488 (green) and LAMP3-PECy5 (red); and polyclonal anti-SMIM1-A488 (green) and 58K-A633 (red). Scale bar – 100 µm.

Platelets are formed through the extension of cytoplasmic protrusions in mature MKs, called proplatelets, into the bone marrow sinusoids [75]. During proplatelet development, several crucial structural alterations occur, including packing of granules and organelles on the distal end of the proplatelet. These structural rearrangements provide the newly-formed platelet with all the components necessary for its biological

function [75]. Possible alteration of SMIM1's localisation upon proplatelet formation was examined by immunofluorescence microscopy of MKs seeded onto fibrinogen for 48 hours to induce proplatelet formation. Although previous studies have shown that fibrinogen is one of the strongest proplatelet inducers, regulating proplatelet formation and release *in vitro* [101], I was unable to induce a strong proplatelet formation of CD34 derived MKs. The reduced proplatelet formation observed probably results from the concentration of fibrinogen used. Nevertheless, confocal analysis of fixed fibrinogen-induced proplatelets, shown that SMIM1 is located in the cytoplasm (Figure 3.8- white arrows) and was also detected in the distal end of the proplatelet, suggesting its migration towards that region. However, due to poor proplatelet induction, this observation may not be representative and additional experiments should be performed to confirm this observation.

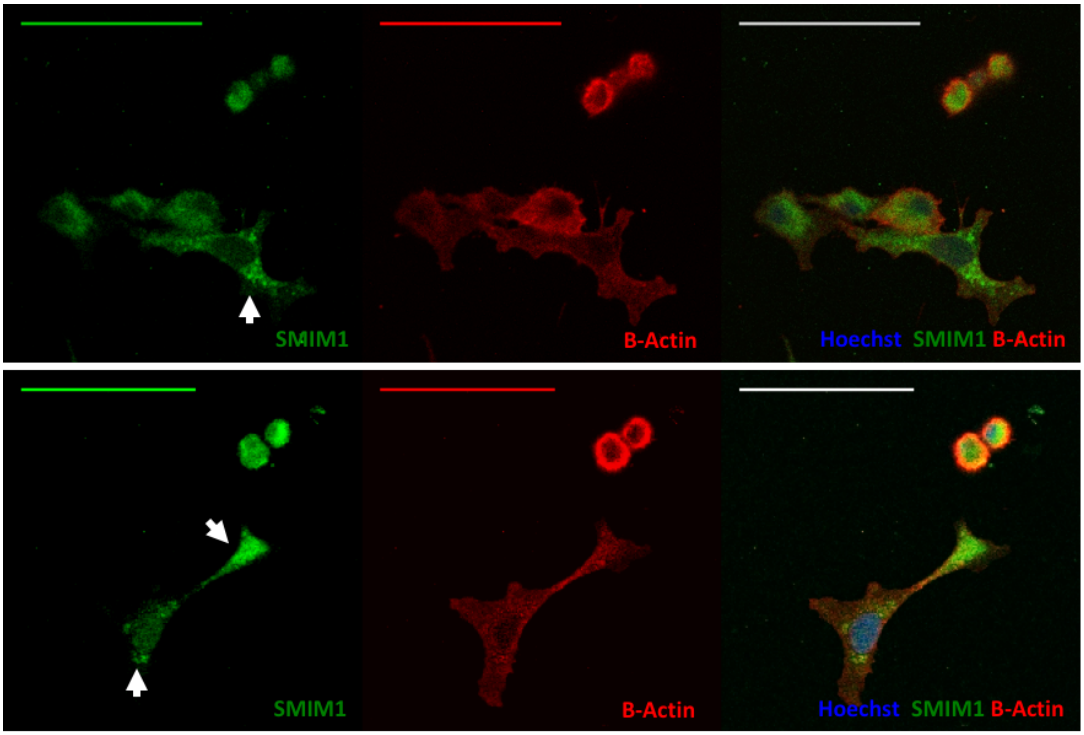


Figure 3.8 – Confocal analysis of SMIM1 localisation in proplatelets. Fibrinogen-induced proplatelets stained with polyclonal anti-SMIM1 and secondary A488 antibody (green), β-actin and secondary A633 antibody (red) and Hoechst 33342 (DNA staining, blue). Scale bar – 50 μm.

I further investigated the localisation of SMIM1 in platelets by subcellular fractionation of human platelets in a sucrose gradient. Subcellular fractionation studies attempting to correlate the localisation of different proteins with specific organelles have been reported in the literature and although this is not the most precise methodology to access

protein localisation, may provide important insights in protein localisation. Here using a membrane of subcellular fractionation of human platelets performed by Louisa Mayer – PhD student in Prof. Ouwehand group, I showed that the first fractions (Figure 3.9, fractions 1 to 5) were enriched for cytoskeleton proteins such as α-tubulin, β-actin and cytokine receptor-like factor 3 (CRLF3), and for the platelet plasma membrane protein, glycoprotein Iba (CD42b – larger form of approximately 200kDa). Proteins present in α-granules, such as THBS1 and CD42b (short form of approximately 130 kDa – present on the alpha granule membrane) were enriched in fractions 7 and 8 (Figure 3.9). Interestingly, SMIM1 co-localised in the fractions enriched for α-granule´s proteins, suggesting that SMIM1 has an intracellular distribution similar to these proteins.

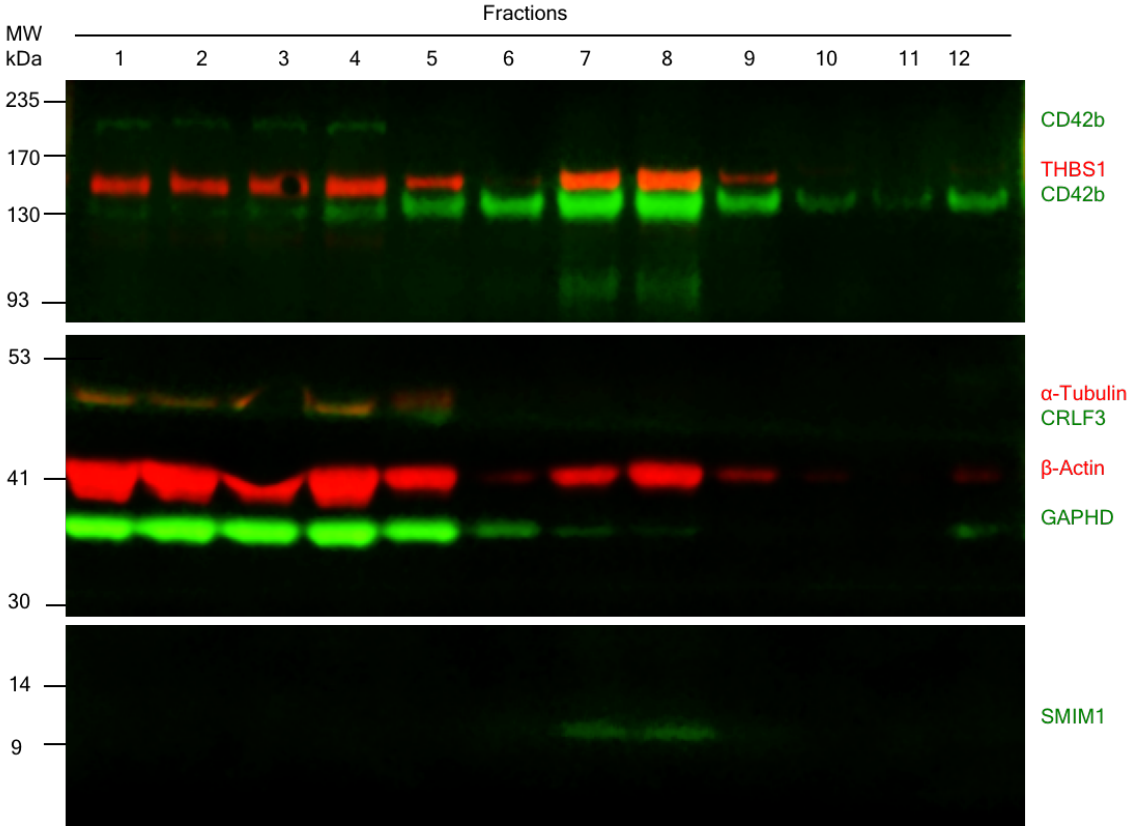


Figure 3.9 - Localization of SMIM1 in platelets. SDS-PAGE separated fractions 1 to 12 of human platelet fractionation by sucrose gradient centrifugation (performed by Louisa Mayer). Western blot of SMIM1 (polyclonal antibody that recognises the N-terminal) and other specific antibodies: CD42a, THBS1, tubulin, CRLF3, actin, GAPDH. SMIM1 is expressed in fractions 7 and 8 of platelet fractionation, like CD42b and THBS1.

3.4 Discussion

In this chapter, I have characterised *SMIM1* expression throughout the haematopoietic lineage and showed that *SMIM1* gene is expressed in naïve, memory and class-switched B-lymphocytes, neutrophils, RBCs, MKs and platelets. Moreover, I have demonstrated that SMIM1 protein subcellular localisation and form is cell-type dependent. RBCs and WBCs express SMIM1 on the membrane, whereas platelets and MKs seem to express SMIM1 only in the cytoplasm. RBCs, MKs and platelets present different molecular complexes of SMIM1. Altogether, these findings may indicate that SMIM1 function is cell-type dependent.

Previous studies were mainly focused in SMIM1 characterisation on RBCs, where the protein was originally identified. However, with the public availability of gene expression data of different cells and tissues, it is now known that *SMIM1* is not only expressed in different blood cell types, but also in non-haematopoietic tissues.

In this chapter, I have shown for the first time that *SMIM1* is expressed in WBC cell types, more specifically in naïve, memory and class-switched B-lymphocytes and neutrophils. Analysis of WBCs according to cell size and granularity (forward and side scatter) by flow cytometry of whole blood incubated with SMIM1 antibody, suggested that neutrophils and B- lymphocytes, like RBCs, express SMIM1 on the plasma membrane. Storry *et al.* previously suggested that SMIM1 is not expressed at the protein level in non-erythroid bone marrow sorted cells, including neutrophils and B-lymphocytes [188]. The discrepancy of our and Storry *et al*. findings may be explained by the fact that SMIM1 is expressed at low level in these cells and consequently its detection in the presence of high abundance proteins is more challenging. Low level of loaded protein extract, accordingly to the faint bands detected in the loading control, as well as the low detection of SMIM1 in CD235a-positive cells (marker expressed by erythroid precursors and erythrocytes) in the western blot present in Storry *et al.* suggests that SMIM1 expression was below the detection threshold [188].

Striking differences in *SMIM1* expression throughout the haematopoietic lineage, with an increase in expression towards myeloid lineage (MEP, EB and MKs), may suggest a possible role of *SMIM1* in EB and MK specification. MEP specification is highly regulated at the transcriptional level through a fine balance between two key

transcription factors, GATA1 and PU.1. Interestingly, *SMIM1* expression is also regulated by GATA1 [190]. On the other hand, high levels of *SMIM1* expression observed in both EB and MKs, may also suggest a possible role of SMIM1 in erythropoiesis and megakaryopoiesis respectively. These observations were previously described by Cvejic *et al.* and the former hypothesis was investigated in the same study using zebrafish as the animal model [186]. This hypothesis and the previous studies are further explored in Chapters 4 and 5.

It has been reported that SMIM1 forms detected in RBCs are related to either post-translational modifications, disulfide-independent complexes and disulfide-dependent complexes [182, 188, 208]. In this chapter, I showed for the first time that SMIM1 is expressed at the protein level in MKs and platelets and that SMIM1 in these cells also form disulfide-dependent complexes (Figure 3.5 – B-C). Under reducing conditions, all the bands detected in platelets and MKs collapse into a single band of approximately 14 kDa, corresponding to the monomeric form of the protein. Analysis under non-reducing conditions suggest that SMIM1 can form homodimers and other multimers in RBCs, MKs and platelets. Contrary to RBCs, the monomer form of SMIM1 is also detected under non-reducing conditions in MKs and platelets. Moreover, the presence and abundance of different bands in the three cell types suggests that SMIM1 undergoes different modifications and different protein-complexes are formed in RBCs, MKs and platelets. As reported in different studies, I also observed that SMIM1 does not migrate at its theoretical molecular weight. Observed differences in molecular weight could be explained by the hydrophobic nature of SMIM1 transmembrane domain (Figure 3.5 – B-C), since hydrophobic transmembrane domains have reduced affinity to SDS molecules, the complete denaturation of proteins with transmembrane domains is impaired and therefore the migration of membrane proteins often deviate from their theoretical molecular weight [216]. Moreover, it has been shown that SMIM1 is subjected to post-translation modifications, such as acetylation and phosphorylation. The four phosphorylation sites identified in SMIM1 N-terminal (Ser6, Ser17, Ser22 and Ser27), rather than the predicted O-glycosylation site, seem to contribute for the observed difference in molecular weight [182, 188, 208, 212]. Additionally, the conditions (gel type and running buffer) used in the SDS-PAGE may contribute to different migration patterns observed in the different studies.

Considering the small size of SMIM1 together with its ability to form different complexes detected by western blot, I hypothesise that SMIM1 interacts with different proteins in RBCs and MKs/platelets, which in turn are specific and important for SMIM1 function in those cells. To date, the investigation performed in RBCs was unable to identify the possible partners of SMIM1. Haer-Wigman *et al*. showed that SMIM1 membrane exposure was not affected in RBCs lacking membrane proteins that are part of different blood group systems (e.g. Rh, Kell, RHAH and others). Moreover, they also showed that Vel-negativity did not impair exposure of other important membrane proteins (such as GPA, GPB, CD71 and others) [187]. Arnaud *et al*. also observed that the N-terminal of SMIM1, which is subject to post-translational modifications (acetylation and phosphorylation), is not required for SMIM1 membrane exposure [208]. Although SMIM1 exposure is not affected and does not impair expression of these membrane proteins, one cannot exclude that SMIM1 presence is important for the biological function of those complexes or that SMIM1 is part of a different complex that has not yet been identified. The identification of SMIM1 interactors will be important in order to dissect SMIM1 function. SMIM1 interactors will be probably cell-type specific, and therefore this investigation should be performed for each cell type.

Interestingly, exposure of SMIM1 on MK and platelet membrane was not detected by flow cytometry. Using immunofluorescence microscopy, I was able to detect SMIM1 in MK cytoplasm, as well as the presence of some accumulations, suggesting a possible interaction of SMIM1 with an organelle or structure present in MK's cytoplasm. On the other hand, using human platelet fractionation, I have shown that SMIM1 in platelets has an intracellular distribution similar to α-granule´s proteins. While these findings have shed new light on SMIM1 localisation in MKs and platelets, an obvious drawback of the investigation I described in this chapter is the inability to identify the precise localisation of SMIM1 in platelets. Co-localization studies in platelets using immunofluorescence microscopy are quite challenging due to the small size of platelets. Therefore, an alternative experiment to overcome this problem and determine the precise localisation of SMIM1 in platelets would be the analysis of immunogold labelled platelets with SMIM1 antibody by transmission electron microscopy.

Altogether, these findings suggest that SMIM1 may be important for platelet function and raises new questions regarding whether Vel-negative individuals present a not yet

recognised haematological feature and/or disease. This hypothesis is further investigated in Chapters 4 and 5. On the other hand, the fact that SMIM1 is not detected on the platelet's membrane indicates that SMIM1 expression in platelets will not be clinically relevant for platelet transfusion medicine, contrary to what occurs with RBCs.

CHAPTER 4 | INVESTIGATION OF THE EFFECTS OF *SMIM1* KNOCKOUT IN BLOOD LINEAGES AND OTHER ORGANS

4.1 Introduction

The Vel-negative blood group is rare but clinically relevant, as individuals with this blood type are at risk of haemolytic transfusion reactions and rarely of haemolytic disease of newborns. No other clinical phenotypes have been ascribed to the absence of the Vel antigen, or its gene, *SMIM1*. However, the *SMIM1* coding sequence is highly conserved among vertebrate [186, 188] and the gene is highly expressed in multiple blood cells and non-haematopoietic tissues, as shown in the previous chapter. This suggests that SMIM1 may have additional, yet undiscovered, importance.

In the existing literature, there are only a few studies aimed at the characterisation of SMIM1 biological role, these have shown strong evidence indicating that SMIM1 has an effect on RBC formation. Meta-analysis of a genome-wide association study of six RBC parameters in approximately 72,000 individuals (62,553 people of European ancestry and 9,308 people of South Asian ancestry) showed that the major allele of the common variant rs1175550, responsible for decreased of Vel/SMIM1 expression [186], is also associated with decreased mean haemoglobin concentration in RBCs (MCHC, $p = 8.6 \times 10^{-15}$) [217]. At the time of this meta-analysis *SMIM1* had not been annotated and consequently *CCDC27* and *LRRC48*, located nearby in the genome, were proposed candidate genes [217]. Further statistical analysis of this study showed that rs1175550 is also associated with a decrease in related parameters: RBC count (RBC, $p=0.005$), haemoglobin concentration (HGB, $p=0.001$) and mean RBC volume (MCV, $p=4.5 \times 10^{-6}$) [186]. A subsequent genome-wide association analysis of 36 blood parameters in 173,480 individuals of European ancestry in two large studies (UK Biobank and INTERVAL), showed that the major allele of rs1175550 is associated with the decrease of both mature and immature RBC parameters (RBC count, MCHC, reticulocyte count and percentage, immature reticulocyte fraction, and high light scatter reticulocyte count and percentage) [196]. This study also revealed that the 17-bp deletion responsible for the Vel-negative phenotype (homozygous and heterozygous individuals – rs566629828, MAF=0.014), is positively associated with red blood cell distribution width (RDW) [196]. The evolutionary importance of *smim1* is supported by studies of morpholino knockdown in zebrafish, that show a reduction in RBC formation [186]. Interestingly, knockdown of *smim1* in CD41-transgenic zebrafish led to a reduction of thrombocyte formation (equivalent to platelets in humans), suggesting for the first time that SMIM1

also regulates megakaryopoiesis, and by inference, may have a role in human platelet formation (unpublished data; Prof. Ouwehand's group).

Quantitative and qualitative alterations in blood cells traits are critical indicators of change in physiological processes such as differentiation and maturation and/or cell function. Quantitative alterations of blood parameters occur due to altered production, clearance or cell redistribution [218]. These variations can be mild, and within accepted 'normal ranges', nevertheless, they are representative of physiological differences. More extreme alterations in blood parameters indicate underlying pathologies such as: anaemia, polycythaemia, disorders of thrombosis and haemostasis, immunodeficiencies and cancer (e.g. leukaemia and lymphoma). These variations can also be associated with predisposition to different conditions such as metabolic and cardiovascular diseases [196, 219-221].

The alteration in RBC parameters associated with *SMIM1* variants suggests that SMIM1 affects erythropoiesis, although even in homozygous Vel-negative humans no overt pathology has been detected. However, the rarity of Vel-negative individuals may have limited the ability of these studies to identify phenotypic alterations. Further investigation is required to examine the mechanisms by which SMIM1 influences RBC physiology, and whether platelet number or function are influenced in humans, as in zebrafish. Furthermore, the role of SMIM1 in non-haematopoietic tissues is entirely unknown.

4.2 Chapter aims and overview

In this chapter I had two main aims:

i) To examine the effect of SMIM1 absence on anthropometric parameters and tissue morphology;
ii) To investigate the role of SMIM1 in the different blood lineages by assessing qualitative and quantitative haematological traits in a:
    a. Cohort of Vel-negative blood donors – recruited in this study;
    b. Cohort of Vel-negative/weak individuals – that are a subset of European ancestry individuals of the 400,000 participants part of the UK Biobank;
    c. *Smim1* mutant (heterozygous and homozygous) mouse model.

Initially, I investigated the effect of the absence of SMIM1 in other tissues, by performing analyses of organ to body weight ratio and tissue histology in *Smim1* mutant mice. I showed that body weight is affected by SMIM1 ablation as well as spleen and liver body weight ratios in a sex-dependent manner. Moreover, histological analysis revealed morphologic alterations in the liver of male *Smim1* mutant mice. Analysis of body mass index (BMI) in male Vel-negative blood donors also indicated that the absence of SMIM1 is associated with higher BMI.

Using two human cohorts and a *Smim1* mouse model, I then characterised the importance of SMIM1 in blood cells morphology and parameters. While blood smear examination did not reveal striking abnormalities, the automated full blood count analysis showed that SMIM1 plays a small, but significant role, in the biology of different blood cell types: RBCs, WBCs and platelets. However, this role seems to be sex and model dependent. Using the mouse model, I further characterised SMIM1 role in the mechanical properties of RBCs by real-time deformability cytometry (RT-DC). I showed that RBCs obtained from *Smim1*$^{-/-}$ male mice have reduced deformability.

4.3 Results

It is well established that fundamental differences between male and female influence the incidence, course and severity of the majority of common diseases in humans [222]. In fact, sexual dimorphism has been described in several haematological traits such as HGB levels, where women present mean levels approximately 12% lower than men [223]. However, despite the evidence, many studies (mainly when using the mouse model) are still designed and analysed without taking sexual dimorphism into account. A large study performed in mice, stressed once more the importance of sex consideration by showing that approximately 53% of the quantitative traits analysed in wildtype mice were influenced by sex, and that sex modified the genotype effect in approximately 17.7% of the quantitative data sets analysed [224]. Therefore, the analyses of males and females were performed independently. Additionally, when possible and appropriate to control for confounders such as age and body weight or BMI, these were used as covariates in the statistical model.

4.3.1 Investigation of the effects of SMIM1 absence in BMI and body weight

4.3.1.1 The Vel-negative donor cohort

27 Vel-negative European ancestry NHSBT blood donors (12 female and 15 male), who had participated in a previous study [186], were recruited by Dr Karola Rehnstrom, Study Coordinator at Prof. Ouwehand group - University of Cambridge. Sanger sequencing of the *SMIM1* locus in these individuals showed that 2/27 donors were heterozygous for the 17-bp deletion. These Vel-weak donors had been erroneously typed as Vel-negative possibly because of an extremely low level of SMIM1 and were thereby excluded from further analysis. In-house data from 178 European ancestry control blood donors (99 female and 79 male), which were genotyped as described in [195], were used as controls.

Male Vel-negative blood donors in this study presented a mean age of 49.7±14.3 years and were significantly younger than the control donors (p=0.0276), which had a mean age of 58.2±10.3 years (Figure 4.1 – A). Female Vel-negative blood donors had a mean age of 54.5±13.4 years, whereas the female control donors showed a mean age of 58.3±12.2 years (Figure 4.1 – A).

4.3.1.2 SMIM1 absence is associated with higher BMI in male blood donors

To determine whether the absence of SMIM1 affects BMI, I compared the BMI of Vel-negative blood donors with that of control donors by using a linear model with age as a covariate (Figure 4.1 – B). Male Vel-negative blood donors presented a mean BMI of 29.5±8.3 Kg/m^2, which was significantly higher (p=0.0156) than the mean BMI (26.2±3 Kg/m^2) of male control donors (Figure 4.1 – B). 25% and 42% of the male Vel-negative blood donors were classified as overweight (BMI=25.0-29.9 Kg/m^2) and obese (BMI ≥ 30 Kg/m^2), respectively. While 48% and 11% of the male control donors were classified as overweight and obese, respectively. As the prevalence of higher BMI is associated with increasing age [225], and the male Vel-negative blood donors were considerably younger, the effect of SMIM1 in BMI may become more evident with increasing age. Female Vel-negative blood donors had a mean BMI of 27.8±6.6 Kg/m^2, while the female control donors showed a mean BMI of 26.6±6.2 Kg/m^2 (Figure 4.1 – B). Although non-significant, a trend towards an increase in BMI was also observed in

female Vel-negative blood donors. 18% and 36% of the female Vel-negative blood donors and 33% and 21% of the female control donors were classified as overweight and obese, respectively.

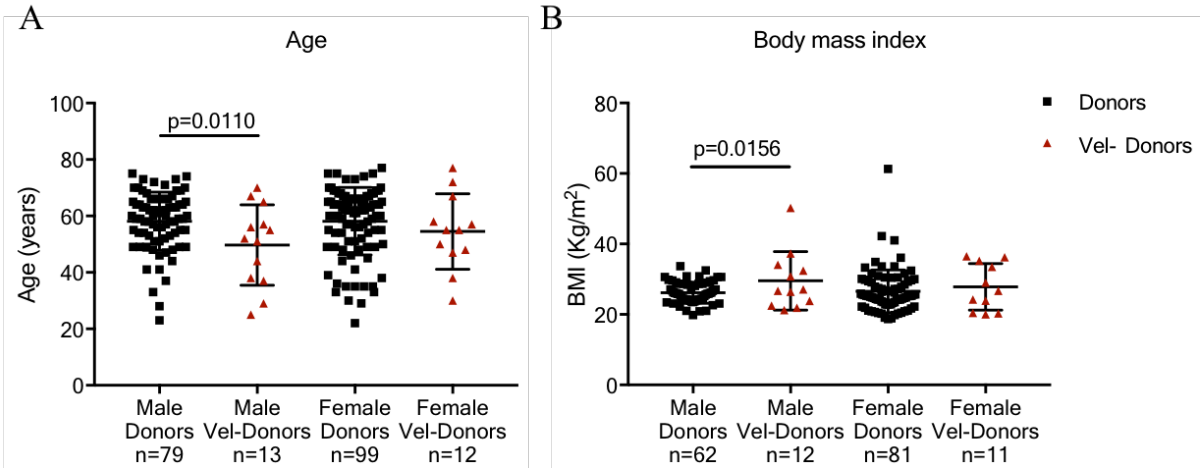


Figure 4.1 – Demographic and anthropometric parameters of blood donors. **A** - Representation of the age range; **B** - Representation of the body mass index (BMI, Kg/m^2). Black squares – control donors and red triangles – Vel-negative blood donors. Data represent mean ± standard deviation. Age analysis was performed using the Mann Whitney test and BMI analysis was performed using a linear model with age as a covariant. Only significant differences are represented.

4.3.1.3 The *Smim1* mutant mice

Smim1 mutant mice used in this study were generated by the Wellcome Trust Sanger Institute as part of the International Mouse Phenotyping Consortium (http://mousephenotype.org/). These mice were created in C57BL/6N genetic background using CRISPR/Cas9 technology that specifically disrupted two critical exons present in all *Smim1* transcripts (Figure 2.1 - Chapter 2). Heterozygous mutant mice (*Smim1*$^{+/-}$) received from the Wellcome Trust Sanger Institute were used to establish a colony of *Smim1* mutant mice in the mouse facility at Central Biomedical Services – University of Cambridge. No abnormalities in breeding, development or behaviour were observed in *Smim1* mutant mice, showing that *Smim1* is not critical for normal mouse development or reproduction. The genotype of all mice used in this study was confirmed by two PCR reactions using gDNA extracted from ear notches. The PCRs were designed to amplify specifically the wildtype allele (using primer set 1 - P1) and the mutant allele (using primer set 2- P2) as represented in Figure 4.2 - A. Homozygous mice for the mutant allele did not originate a PCR product when using P1, and originated a PCR product smaller than the wildtype when using P2 (Figure 4.2 - B).

Heterozygous mice for the mutant allele originated both a PCR product smaller than the wildtype and a PCR product similar to the wildtype when using P2 (Figure 4.2 - B). Additionally, a band of approximately 900 bp was also observed in the PCR of heterozygous mice when using P2. This 900 bp band might be a PCR heteroduplex, which results from the amplification of two allelic DNA segments with different sequences (wildtype and mutant). Moreover, Vel/SMIM1 expression was confirmed by flow cytometry of whole blood incubated with anti-SMIM1 (Figure 4.2 - C). As expected, SMIM1 was detected on WBCs and RBCs of wildtype mice but not in *Smim1*$^{-/-}$ mice (Figure 4.2 - C). Since SMIM1 is not exposed on platelet membranes, I assessed SMIM1 expression in platelets by RNA-sequencing (Figure 4.2 - D) and western blot (Figure 4.2 - E). Platelet RNA-sequencing showed that the deletion of the two crucial exons lead to a reduction of *Smim1* expression level in *Smim1*$^{+/-}$ mice and a vestigial expression in *Smim1*$^{-/-}$ mice when compared with wildtype mice. Western blot of platelet lysates confirmed the results obtained in the RNA-sequencing. I detected a band of approximately 14 kDa in the wildtype lane, a faint band of the same size in the *Smim1*$^{+/-}$ lane (green arrow, Figure 4.2 - E), and as expected no band was detected in the platelet extracts of *Smim1*$^{-/-}$ mice.

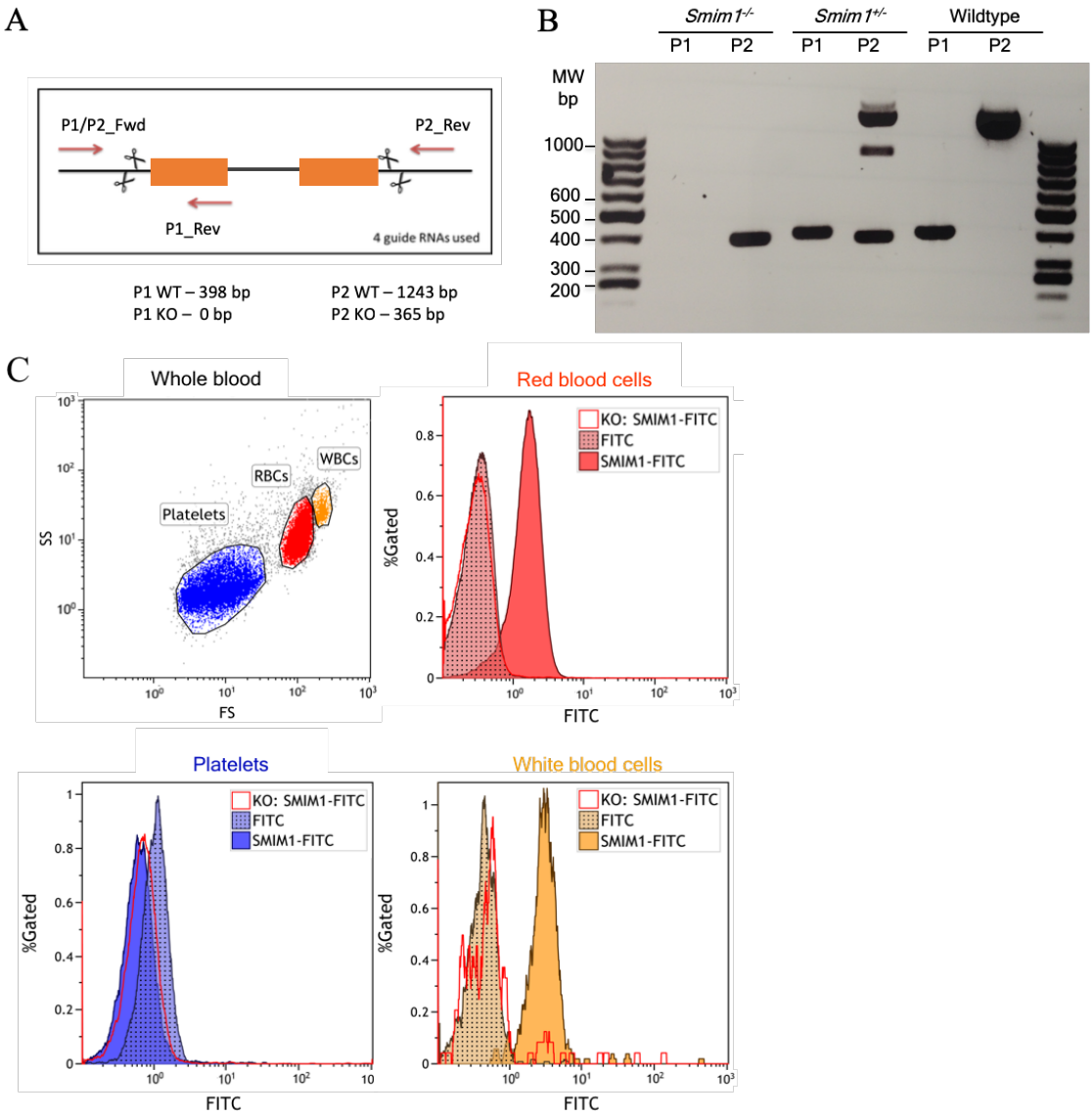


Figure 4.2 (A,B and C)- Genotyping of *Smim1* mutant mice. A - Schematic representation of CRISPR and genotyping strategy; **B** - Agarose gel of PCR products of gDNA amplified by PCR using specific primers; **C** - Representative flow cytometry dot plot for forward and side scatter and histograms of SMIM1 expression of wildtype and *Smim1* $^{-/-}$ mice whole blood incubated with the SpG213Dc human monoclonal anti-SMIM1 and or secondary FITC, red line histogram represent results for *Smim1* $^{-/-}$ mouse (KO: SMIM1-FITC).

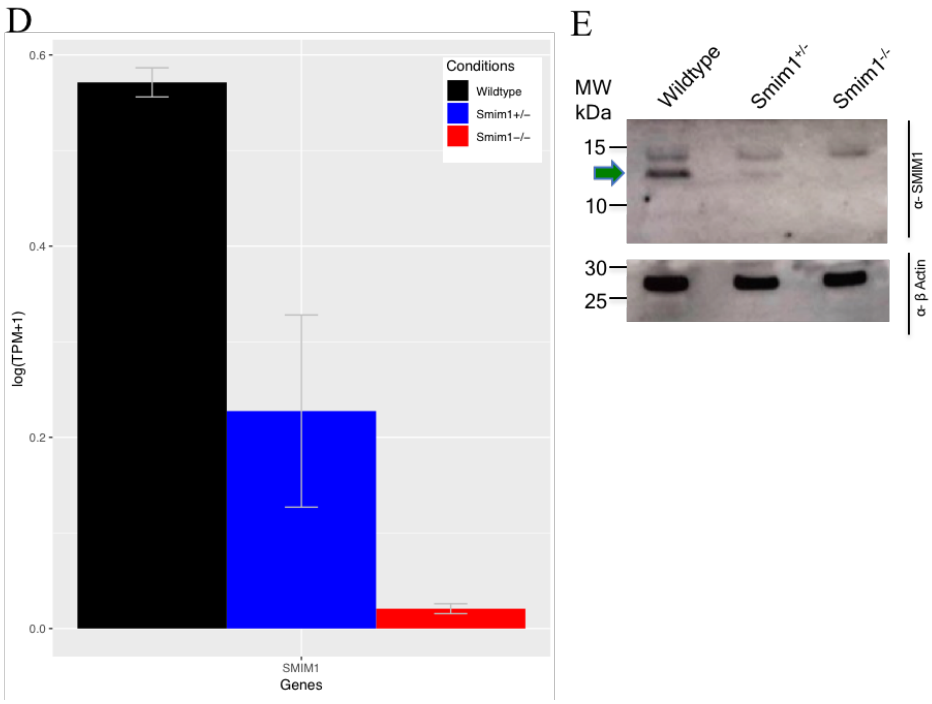


Figure 4.2 (D and E) - Genotyping of *Smim1* mutant mice. D – Platelet expression of *Smim1* in wildtype, *Smim1*$^{+/-}$ and *Smim1* $^{-/-}$ obtained by RNA-sequencing, TPM – transcripts per Kilobase million reads; **E** – Western blot of wildtype, *Smim1*$^{+/-}$ and *Smim1* $^{-/-}$ platelets using polyclonal SMIM1 and b-actin antibodies.

4.3.1.4 SMIM1 ablation leads to the alteration of body weight in age and sex-dependent manner in the mouse model

As the absence of SMIM1 is associated with an increase of BMI in male Vel-negative blood donors, I then investigated whether the ablation of SMIM1 in the mouse model would also affect body weight. Terminal body weight of *Smim1* mutant and wildtype mice was measured at two (nine to ten weeks), four (16 to 18 weeks) and six-seven (24 to 31 weeks) months of age (Table 4.1 and Figure 4.3), which corresponds to the range of ages used in the other types of experiments in this study. *Smim1*$^{-/-}$ female at two and six-seven months old showed increased of body weight when compared with wildtype animals (two months old – p=0.0425; six-seven months old – p= 0.0092; Table 4.1 and Figure 4.3 – B). This observation is in agreement with the observed trends in female Vel-negative blood donors described in section 4.3.1.2. Opposite to what was observed in females, two months old *Smim1*$^{-/-}$ males did not show significant difference in body weight and in fact, four months old *Smim1*$^{-/-}$ male mice presented a significant decrease in body weight when compared with wildtype and *Smim1*$^{+/-}$ male (*Smim1*$^{-/-}$ *vs* wildtype p=0.049; *Smim1*$^{-/-}$*vs Smim1*$^{+/-}$ p=0.0009; Table 4.1 and Figure 4.3 – A). However, at six-seven months old the trend was inverted and *Smim1*$^{-/-}$ male mice presented a non-

significant trend of body weight increase, as it is observed in male Vel-negative blood donors (Table 4.1, Figure 4.3 – A and Figure 4.1). In summary, SMIM1 seems to affect body weight in age and sex-dependent manner in mice.

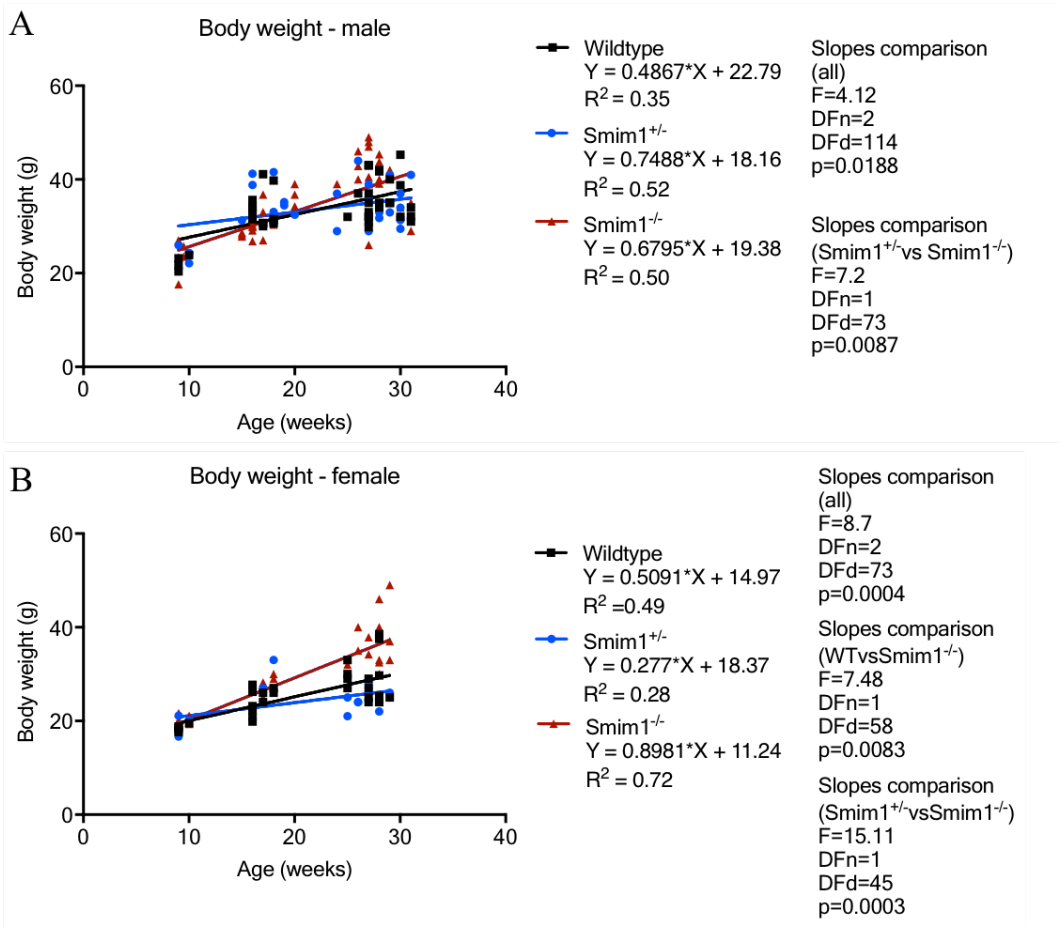


Figure 4.3 – Characterisation of body weight in *Smim1* mutant mice. A - Body weight of male mice at two, four and six-seven months old. Slopes comparison of Wildtype vs *Smim1*$^{+/-}$ vs *Smim1*$^{-/-}$- F=4.12, DFn=2, DFd=114, p=0.0188; slopes comparison of *Smim1*$^{+/-}$ vs *Smim1*$^{-/-}$- F=7.12, DFn=1, DFd=73, p=0.0087. **B** - Body weight of female mice at two, four and six-seven months old. Slopes comparison of Wildtype vs *Smim1*$^{+/-}$ vs *Smim1*$^{-/-}$- F=8.7, DFn=2, DFd=73, p=0.0004; slopes comparison of Wildtype vs *Smim1*$^{-/-}$- F=7.48, DFn=1, DFd=58, p=0.0083; slopes comparison of *Smim1*$^{+/-}$ vs *Smim1*$^{-/-}$- F=15.11, DFn=1, DFd=45, p=0.0003. Black squares – wildtype; blue circle - *Smim1*$^{+/-}$; red triangles - *Smim1*$^{-/-}$. The analysis was performed using linear regression. F-F test, DF – degree of freedom (n) -numerator (d) denominator.

4.3.2 Assessment of the effect of SMIM1 ablation in different organs in the mouse model

As shown in the previous chapter, *SMIM1* is highly expressed also in some non-haematopoietic tissues and/or its role in haematopoietic cells may also indirectly affect the function of other organs. To assess the direct and indirect effects of SMIM1 ablation in different organs in *Smim1* mutant mice, I weighed and morphologically characterised the organs that play a direct or indirect role in haematopoiesis: spleen, liver, kidneys, heart and bone marrow.

4.3.2.1 Investigation of the effect of SMIM1 ablation in the organ to body weight ratio

To investigate whether ablation of SMIM1 has an effect on the organ to body weight ratio, spleen, liver, kidneys and heart of *Smim1* mutant and wildtype mice, were dissected and weighed after terminal body weight measurement. The analysis was performed in two age groups: four and six-seven months old mice. Analysis of organ to body weight ratios showed non-significant differences between *Smim1*$^{-/-}$ and wildtype mice of both age groups (Table 4.1 and Figure 4.4).

However, some significant differences in the organ to body weight ratios were observed between *Smim1*$^{+/-}$ and wildtype and/or *Smim1* $^{-/-}$. Specifically, liver to body weight ratio of six-seven months old *Smim1*$^{+/-}$ male mice was significantly decreased when compared with wildtype and *Smim1* $^{-/-}$ mice (*Smim1*$^{+/-}$ *vs* wildtype p=0.0155; *Smim1*$^{+/-}$ *vs Smim1*$^{-/-}$ p=0.0045; Table 4.1 and Figure 4.4). In females, I observed a significant increase of spleen to body weight ratio in six-seven months old *Smim1*$^{+/-}$ mice when compared with wildtype and *Smim1* $^{-/-}$ mice (*Smim1*$^{+/-}$ *vs* wildtype p=0.0051; *Smim1*$^{+/-}$ *vs Smim1*$^{-/-}$ p=0.0001; Table 4.1 and Figure 4.4). I also observed a significant decrease in the heart to body weight ratio of six-seven months old *Smim1*$^{-/-}$ female mice when compared to *Smim1*$^{+/-}$ (p=0.0095; Table 4.1 and Figure 4.4). This alteration, however, was due to the significant increase of body weight in *Smim1*$^{-/-}$ female mice (Table 4.1).

In summary, these results indicate that SMIM1 ablation has no striking effect in organ to body weight ratios. Whereas, decreased levels of SMIM1 expression observed in *Smim1*$^{+/-}$ mice affects liver and spleen to body weight ratios, in males and females, respectively. Furthermore, the effects of SMIM1 ablation (*Smim1*$^{-/-}$ mice) and SMIM1

knockdown (*Smim1*$^{+/-}$ mice) on tissue to body weight ratio seem to be opposite. Together, this may indicate that SMIM1 levels affect liver and spleen characteristics/function, in males and females, respectively.

Table 4.1 – Effect of SMIM1 ablation in body weight and organ to body weight ratio in mice. Summary of body weight and organ to body weight ratio measurements in different age groups (two, four and six-seven months old). Data represent mean ± standard deviation (SD). The analysis was performed using Anova and Tukey's multiple comparisons test or Kruskal-Wallis and Dunn's multiple comparisons test. WT – wildtype.

Male	Wildtype			*Smim1*$^{+/-}$			*Smim1*$^{-/-}$			
Age (months)	**2**									
Parameter	**Mean**	**SD**	**Number**	**Mean**	**SD**	**Number**	**Mean**	**SD**	**Number**	**p-value**
Body (g)	22.30	1.56	4	24.20	1.91	3	23.30	3.24	6	ns
Spleen ratio	2.73E-03	1.54E-04	4	2.90E-03	2.23E-04	3	2.88E-03	2.73E-04	5	ns
Age (months)	**4**									
Parameter	**Mean**	**SD**	**Number**	**Mean**	**SD**	**Number**	**Mean**	**SD**	**Number**	**p-value**
Body (g)	33.32	3.58	13	36.56	4.06	7	30.23	2.50	16	p=0.0490 (WTvsSmim1$^{-/-}$) p=0.0009 (Smim$^{+/-}$vsSmim1$^{-/-}$)
Spleen ratio	2.65E-03	6.25E-04	11	2.22E-03	3.94E-04	4	3.05E-03	5.79E-04	15	p=0.0452 (Smim$^{+/-}$vsSmim1$^{-/-}$)
Heart ratio	4.89E-03	6.36E-04	11	3.79E-03	4.83E-04	2	5.02E-03	4.61E-04	13	ns
Liver ratio	5.15E-02	2.33E-03	7	NA	NA	NA	5.18E-02	5.45E-03	7	ns
Kidneys ratio	1.49E-02	1.31E-03	7	NA	NA	NA	1.43E-02	7.28E-04	8	ns
Age (months)	**6 to 7**									
Parameter	**Mean**	**SD**	**Number**	**Mean**	**SD**	**Number**	**Mean**	**SD**	**Number**	**p-value**
Body (g)	36.10	4.55	26	35.04	4.43	16	38.90	6.09	24	ns
Spleen ratio	2.83E-03	6.00E-04	26	2.95E-03	9.07E-04	16	2.65E-03	4.98E-04	23	ns
Heart ratio	4.68E-03	6.38E-04	18	4.43E-03	4.34E-04	9	4.38E-03	4.88E-04	15	ns
Liver ratio	5.76E-02	5.24E-03	22	5.12E-02	4.99E-03	13	5.88E-02	8.09E-03	19	p=0.0155 (WTvsSmim1$^{+/-}$) p=0.0045 (Smim$^{+/-}$vsSmim1$^{-/-}$)
Kidneys ratio	1.37E-02	1.10E-03	17	1.34E-02	2.51E-03	9	1.35E-02	1.56E-03	15	ns

Female	Wildtype			*Smim1*$^{+/-}$			*Smim1*$^{-/-}$			
Age (months)	**2**									
Parameter	**Mean**	**SD**	**Number**	**Mean**	**SD**	**Number**	**Mean**	**SD**	**Number**	**p-value**
Body (g)	18.48	0.73	5	19.00	1.71	5	20.27	0.83	7	p=0.0425 (WTvsSmim1$^{-/-}$)
Spleen ratio	3.72E-03	4.23E-04	5	3.75E-03	3.74E-04	5	3.43E-03	2.90E-04	7	ns
Age (months)	**4**									
Parameter	**Mean**	**SD**	**Number**	**Mean**	**SD**	**Number**	**Mean**	**SD**	**Number**	**p-value**
Body (g)	24.29	2.71	10	29.20	3.30	3	25.96	2.69	10	p=0.0354 (WTvsSmim1$^{+/-}$)
Spleen ratio	3.15E-03	4.58E-04	10	3.15E-03	3.91E-04	2	3.23E-03	3.49E-04	9	ns
Heart ratio	4.99E-03	6.38E-04	9	4.25E-03	2.95E-04	2	4.72E-03	4.82E-04	9	ns
Liver ratio	4.94E-02	3.94E-03	8	5.69E-02	7.32E-03	2	4.85E-02	1.01E-02	7	ns
Kidneys ratio	1.41E-02	1.44E-03	8	1.22E-02	3.04E-04	2	1.16E-02	4.54E-03	7	ns
Age (months)	**6 to 7**									
Parameter	**Mean**	**SD**	**Number**	**Mean**	**SD**	**Number**	**Mean**	**SD**	**Number**	**p-value**
Body (g)	28.53	4.67	15	24.78	2.28	9	35.84	6.10	15	p=0.0092 (WTvsSmim1$^{-/-}$) p=8.4E-05 (Smim$^{+/-}$vsSmim1$^{-/-}$)
Spleen ratio	3.29E-03	4.57E-04	10	3.96E-03	2.69E-04	6	2.99E-03	3.17E-04	10	p=0.0051 (WTvsSmim1$^{+/-}$) p=0.0001 (Smim$^{+/-}$vsSmim1$^{-/-}$)
Heart ratio	4.28E-03	3.63E-04	6	4.93E-03	5.86E-04	6	3.86E-03	6.26E-04	6	p=0.0095 (Smim$^{+/-}$vsSmim1$^{-/-}$)
Liver ratio	5.53E-02	4.83E-03	6	5.39E-02	5.96E-03	6	5.03E-02	3.13E-03	6	ns
Kidneys ratio	1.37E-02	1.90E-03	6	1.40E-02	1.38E-03	6	1.19E-02	2.13E-03	6	ns

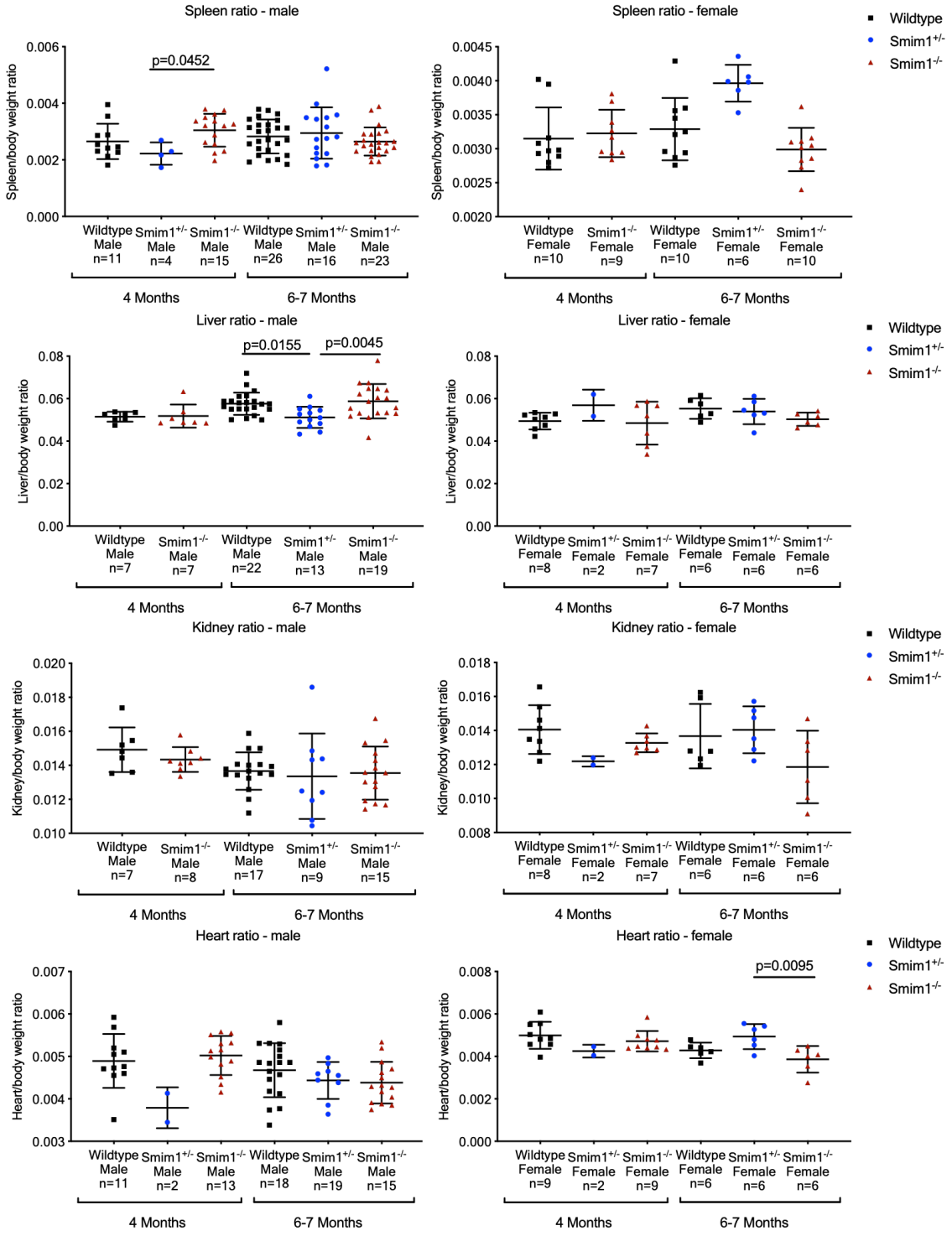


Figure 4.4 –Effect of SMIM1 ablation in organ to body weight ratio in mice. Black squares – wildtype mice, blue circles - *Smim1*$^{+/-}$ mice and red triangles – *Smim1*$^{-/-}$ mice. Data represent mean ± standard deviation. The analysis was performed using Anova and Tukey's multiple comparisons test or Kruskal-Wallis and Dunn's multiple comparisons test. Only significant differences are represented.

4.3.2.2 Investigation of the effect of SMIM1 ablation in tissue morphology

To determine whether the ablation of SMIM1 has an effect on tissue morphology, I performed a histological analysis of spleen, liver, kidneys, heart and bone marrow of six-seven months old *Smim1* mutant and wildtype mice. Following tissue dissection and processing, the histological sections and H&E staining were performed by Mr James Wraner, Chief research technician at Wellcome Trust-MRC Institute of Metabolic Science. The H&E staining is a common stain used to characterise cell morphology. Haematoxylin has a blue-purple colour and stains basophilic structures such as nucleic acids, while eosin has a pink colour and stains acidophilic structures such as proteins. Therefore, in tissue stained with H&E the cells present the nuclei in purple-blue and the cytoplasm and extracellular matrix in pink.

Examination of spleen, kidneys, heart and bone marrow histological sections of *Smim1* mutant mice showed no striking abnormalities in morphology or cellularity (data not shown). Liver histological sections of male *Smim1* mutant mice showed morphologic alterations in the cytoplasm of the hepatocytes (Figure 4.5). Hepatocytes of *Smim1*$^{+/-}$ male mice presented increased eosin staining in their cytoplasm when compared with hepatocytes from wildtype and *Smim1*$^{-/-}$ male mice (Figure 4.5). Whereas hepatocytes from *Smim1*$^{-/-}$ male mice showed reduced eosin staining in their cytoplasm when compared with hepatocytes from wildtype and *Smim1*$^{+/-}$ male mice (Figure 4.5). The reason for the reduction of eosin staining in hepatocytes from *Smim1*$^{-/-}$ male mice could be due to an accumulation of glycogen or lipid droplets, which are neutral and therefore not stained when using this technique. To characterise this morphologic alteration, it would be important to perform a specific staining to glycogen and/or lipids, such as Periodic Acid-Schiff and Oil Red O, respectively. The distinct morphologic alterations observed in the liver of *Smim1*$^{+/-}$ and *Smim1*$^{-/-}$ male mice are in agreement with the observations of the liver to body weight ratio alterations described above in section 4.3.2.1. Examination of liver morphology in female mice showed no differences between *Smim1*$^{-/-}$ and wildtype mice (Figure 4.5).

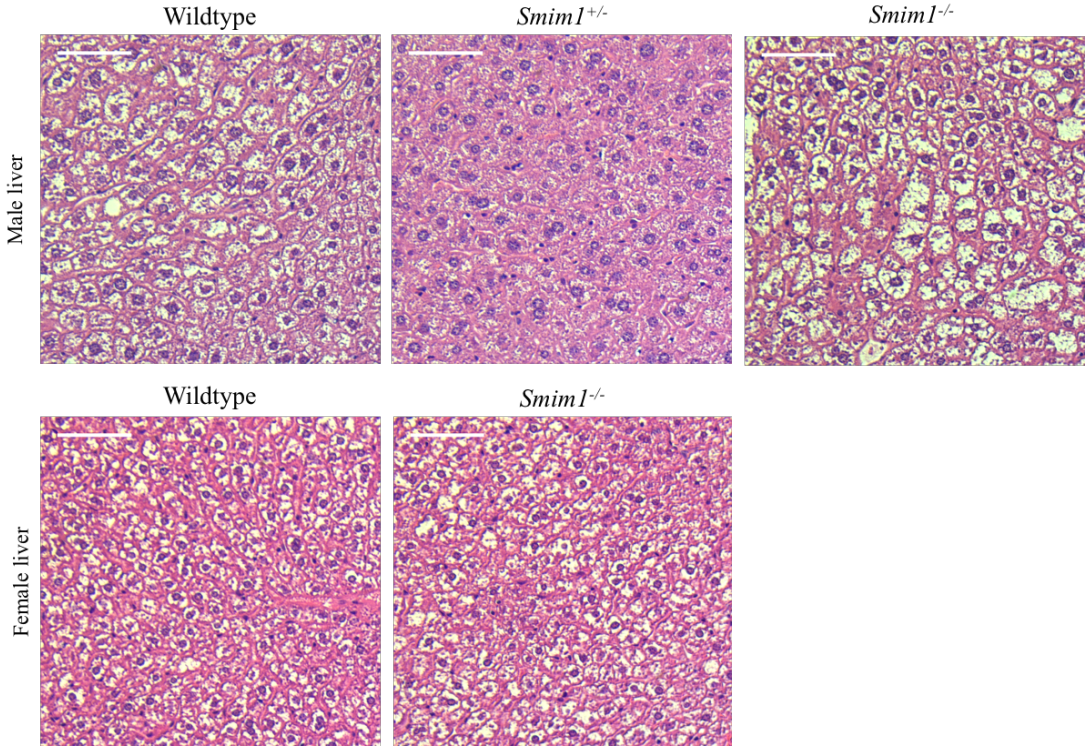


Figure 4.5 – Effect of SMIM1 ablation in the liver morphology of mice. Representative pictures of liver histological sections stained with H&E. Histology analysis was performed in four wildtype and *Smim1*$^{-/-}$ mice and three *Smim1*$^{+/-}$ male mice. Scale bar – 50 μm.

4.3.2.2.1 Liver morphological alterations of Smim1$^{-/-}$ *male mice are not caused by glycogen accumulation*

To further investigate whether the morphologic alteration observed in the liver of *Smim1*$^{-/-}$ male mice was caused by an accumulation of glycogen, I performed a histological analysis using PAS staining in a larger group of male mice. PAS staining is used to detect polysaccharides such as glycogen. However, PAS is specific for aldohexoses and therefore stains all the molecules containing aldohexoses. This staining results from the oxidation of carbon-carbon bounds in 1,2-glycols by the periodic acid, which leads to the formation of dialdehydes. These aldehydes then react with the Schiff reagent to originate a purple-magenta stain. The distinction between glycoproteins and glycogen is obtained through the comparison of PAS staining with and without treatment with diastase, an enzyme that digests specifically glycogen [226]. This, however, is a semi-quantitative method. To perform this analysis, 26 to 27 weeks-old male mice were weighted and dissected within 1h (to control for glycogen consumption). Their livers were weighted and immediately processed for histology as described in section 2.2.6. Semi-quantification of PAS/DPAS staining performed using

HALO software as described in section 2.2.6.1, has shown that glycogen content was not altered in *Smim1*$^{-/-}$ mice when compared with wildtype, while it was significantly increased in *Smim1*$^{+/-}$ mice when compared with wildtype and *Smim1*$^{-/-}$ mice. Moreover, this analysis has also shown that the reduction of H&E staining in the cytoplasm of hepatocytes from *Smim1*$^{-/-}$ mice is most likely due to an accumulation of lipid droplets, also called steatosis (Figure 4.6). To confirm this finding and to determine if the steatosis observed in *Smim1*$^{-/-}$ male mice is significantly different from the wildtype animals, it would be important to perform a lipid-specific stain as described above. In summary, this analysis shows that the morphologic alteration in the liver of *Smim1*$^{-/-}$ male mice does not result from glycogen accumulation.

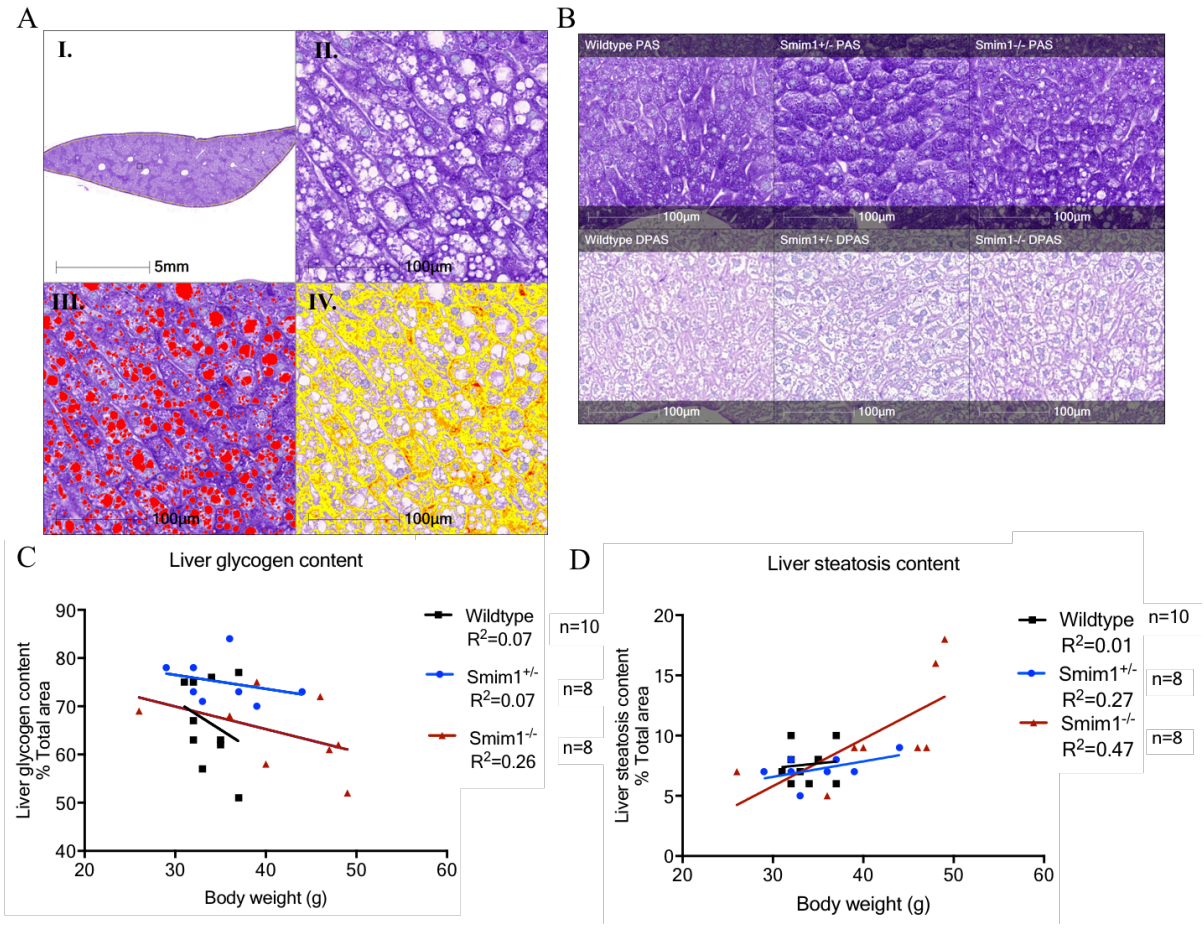


Figure 4.6 - Liver glycogen content is unaffected by SMIM1 ablation in mice. A – schematic representation of semi-quantitative PAS analysis using HALO software: I. Liver histological section analysed, II. PAS staining; III. Identification of steatosis in red; IV quantification of PAS staining – yellow/orange/red staining. **B** – representative liver histological sections of PAS and DPAS staining of wildtype, *Smim1*$^{+/-}$ and *Smim1*$^{-/-}$; **C** – graphic representation of relative glycogen area (%) vs body weight in wildtype, *Smim1*$^{+/-}$ and *Smim1*$^{-/-}$; wildtype - Y = -1.197*X + 107.1; *Smim1*$^{+/-}$ - Y = -0.2884*X + 85.17; *Smim1*$^{-/-}$ - Y = -0.4695*X + 84.05. Slopes comparison show non-significant differences. Intercept comparison: wildtype vs *Smim1*$^{-/-}$ - F = 0.1846, DFn = 1, DFd = 15, p=0.6736; wildtype vs *Smim1*$^{+/-}$ - F = 6.368, DFn = 1, DFd = 15 p=0.0234; *Smim1*$^{-/-}$ vs *Smim1*$^{+/-}$ - F = 5.43, DFn = 1, DFd = 13, p=0.0365; **D**- graphic representation of relative steatosis area (%) vs body weight in wildtype, *Smim1*$^{+/-}$ and *Smim1*$^{-/-}$; Wildtype - Y = 0.07692*X + 5 ; *Smim1*$^{+/-}$ - Y = 0.1285*X + 2.719 ; *Smim1*$^{-/-}$ - Y = 0.3909*X - 5.923; Black squares and trend line – wildtype mice, blue circles and trend line - *Smim1*$^{+/-}$ mice and red triangles and trend line – *Smim1*$^{-/-}$ mice.

4.3.3 Assessment of SMIM1 role in different blood cell lineages

To investigate the possible role of SMIM1 in different blood lineages, I assessed qualitative and quantitative traits of blood cells in a: i) Vel-negative blood donors' cohort, ii) Vel-negative/weak individuals' cohort (UK Biobank) and iii) *Smim1* mutant mice. Specifically, I performed a morphologic analysis of blood cells by examination of blood smears under the microscope, and a more detailed and quantitative analysis of the effect of SMIM1 ablation in 31 or 26 blood parameters in humans, and ten blood parameters in mice using automated haematology analysers.

4.3.3.1 Qualitative and quantitative analysis of blood cells parameters in Vel-negative blood donor cohort

Human blood smears and full blood counts were performed within 5 h after bleeding, as described in section 2.1.1.1. Examination of blood smears for assessment of morphological abnormalities was performed with the assistance of Professor Wendy Erber (University of Western Australia). This qualitative analysis did not reveal any striking phenotype in blood cells morphology as shown in Figure 4.7. Red blood cell size, shape and colour were unremarkable. Platelet number, size and granularity were unremarkable. 3/12 females showed a mild increase of platelet number with normal morphology, which may be attributable to intercurrent infection. Moreover, leucocytes were also normal.

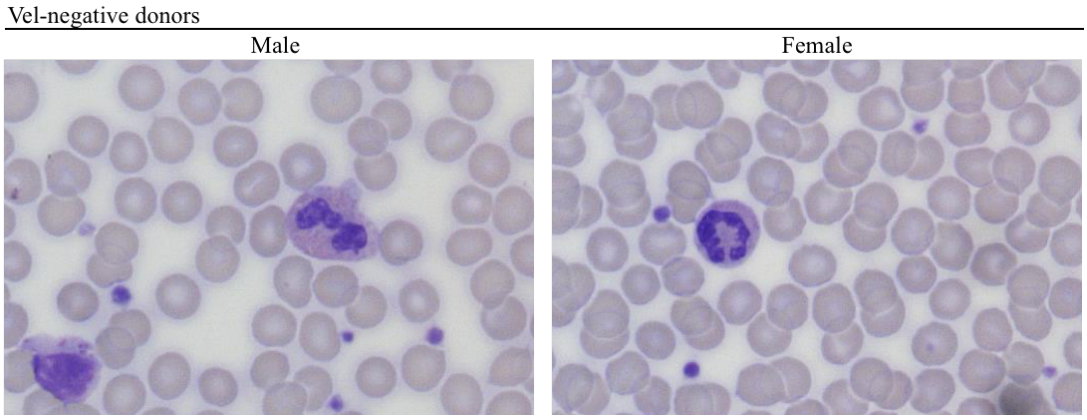


Figure 4.7 – Blood smear of Vel-negative blood donors. Representative blood smear of a male and female Vel-negative donor stained with Romanowsky stain (x400). The blood smears show red blood cells with little variation in size and shape; platelets with normal size and granularity; normal polymorphonuclear neutrophils and a lymphocyte in the blood smear of the male donor.

Then, I performed a quantitative analysis of SMIM1 effect in different blood lineages by assessment of full blood count using a Sysmex XE-5000 hematology analyzer. The statistical analysis was performed using a linear model with age and BMI as covariant and the results were summarised in Table 4.2.

Table 4.2 – Effect of SMIM1 absence in blood parameters of blood donors. Summary of full blood counts of control donors and Vel-negative blood donors obtained by Sysmex XE-5000 hematology analyzer. Data represent mean ± standard deviation (SD). The analysis was performed using a linear model (Parameter ~ cohort + Age + BMI) and Anova type II. Reference ranges were obtained in the INTERVAL study (Prof. Ouwehand, personal communication). ns-non-significant; na- non-applicable.

Gender	Male						Female					
Cohort	Control donors		Vel-negative donors			Reference range	Control donors		Vel-negative donors			Reference range
Number	79		13			27K	99		12			27K
Parameter	Mean	SD	Mean	SD	p-value		Mean	SD	Mean	SD	p-value	
RBC (10^12/L)	5.12	0.37	4.83	0.24	0.0098	3.5-5.6	4.66	0.31	4.45	0.25	0.0332	3.5-5.6
HGB (g/dL)	15.27	0.95	14.77	0.82	ns	10.7-16.2	13.76	0.94	13.18	1.16	ns	10.7-16.2
HCT (%)	44.60	2.61	42.90	2.41	ns	34.0-51.2	40.91	2.59	39.72	3.31	ns	34.0-51.2
MCV (fL)	87.28	3.87	88.87	4.31	ns	78.1-108.5	87.98	4.22	89.34	5.37	ns	78.1-108.5
MCH (pg)	29.90	1.46	30.61	1.68	ns	23.8-35.1	29.59	1.79	29.41	1.64	ns	23.8-35.1
MCHC (g/dL)	34.25	0.99	34.44	0.95	ns	28.1-35.0	33.62	0.86	33.18	0.65	ns	28.1-35.0
RET (10^12/L)	0.05	0.02	0.04	0.01	0.0038	na	0.04	0.01	0.03	0.01	0.0043	na
RET (%)	0.93	0.29	0.79	0.21	0.0233	na	0.84	0.24	0.66	0.21	0.0115	na
RET-He	33.87	1.99	34.79	1.81	ns	na	32.94	2.35	33.78	2.19	ns	na
IRF (%)	4.04	2.39	2.24	1.73	0.0024	na	3.16	2.19	1.76	3.25	0.0449	na
RDW-SD (%)	42.07	2.95	44.21	3.33	0.0081	na	43.69	3.23	45.60	3.29	0.0283	na
RDW-CV (%)	13.25	0.75	13.63	0.90	ns	na	13.63	1.24	13.95	0.95	ns	na
MicroR (%)	1.77	1.20	1.49	1.08	ns	na	2.10	2.43	1.87	1.57	ns	na
MacroR (%)	6.91	1.33	7.41	1.94	ns	na	6.54	1.24	7.65	3.75	ns	na
PLT (10^9/L)	201.72	44.44	219.15	51.19	ns	85.6-438.1	226.70	46.17	270.00	61.17	0.0003	85.6-438.1
MPV (fL)	10.59	0.94	11.15	0.90	0.0408	8.5-14.2	10.73	1.01	11.28	1.08	ns	8.5-14.2
PDW (fL)	12.59	2.07	13.69	1.93	ns	na	12.85	2.19	13.88	2.85	ns	na
PCT (%)	0.21	0.04	0.24	0.05	ns	na	0.24	0.05	0.30	0.06	6.4E-06	na
P-LCR (%)	29.65	7.93	34.96	7.84	0.0274	na	30.71	8.48	35.00	8.50	ns	na
IPF (%)	4.79	3.02	5.12	2.48	ns	na	4.24	2.56	6.89	3.69	0.0037	na
WBC (10^9/L)	5.59	1.36	6.67	2.09	0.0243	1.6-11.9	5.91	1.25	6.29	1.28	ns	1.6-11.9
NEUT (10^9/L)	3.20	1.01	3.93	1.34	0.0065	0 - 7.8	3.52	0.96	3.79	1.05	ns	0 - 7.8
LYMPH (10^9/L)	1.65	0.52	1.93	0.83	ns	0.3 - 3.9	1.76	0.45	1.84	0.43	ns	0.3 - 3.9
Mono (10^9/L)	0.53	0.16	0.59	0.31	ns	0.1 - 1.0	0.46	0.12	0.44	0.11	ns	0.1 - 1.0
Eo (10^9/L)	0.18	0.11	0.18	0.11	ns	0 - 0.6	0.14	0.09	0.19	0.11	ns	0 - 0.6
Baso (10^9/L)	0.03	0.02	0.04	0.02	ns	0 - 0.1	0.03	0.02	0.03	0.02	ns	0 - 0.1
NEUT (%)	56.71	7.85	59.20	7.99	0.0362	na	59.14	6.88	59.68	7.47	ns	na
LYMPH (%)	29.77	6.68	29.10	7.91	ns	na	30.14	6.48	29.48	5.98	ns	na
Mono (%)	9.74	2.31	8.58	2.11	ns	na	7.86	1.80	7.04	1.25	ns	na
Eo (%)	3.24	2.00	2.65	1.26	ns	na	2.37	1.39	3.23	2.35	ns	na
Baso (%)	0.54	0.36	0.58	0.36	ns	na	0.50	0.25	0.57	0.38	ns	na

RBC – red blood cell count; HGB – haemoglobin concentration; HCT – haematocrit; MCV – mean cell volume; MCH – mean cell haemoglobin; MCHC – mean cell haemoglobin concentration; RET – reticulocyte count, RET-He – reticulocyte haemoglobin equivalent; IRF – immature reticulocyte fraction; RDW-SD – red blood cell distribution width - standard deviation; RDW-CV – red blood cell distribution width - coefficient of variation; MicroR – microcytic RBC count; MacroR – macrocytic RBC count; PLT – platelet count; MPV – mean platelet volume; PDW – platelet distribution width; PCT – plateletcrit; P-LCR – platelet larger cell ratio; IPF – immature platelet fraction; WBC – white blood cell count; NEUT – neutrophil count; LYMPH – lymphocyte count; Mono – monocyte count; Eo – eosinophil count; Baso – basophil count.

4.3.3.1.1 Analysis of RBC parameters in Vel-negative blood donor cohort

Different parameters of mature and immature RBCs were investigated using the Sysmex XE-5000 hematology analyzer (Table 4.2). RBC count and RDW are measurements of number and size variation of fully mature RBC, respectively. Whereas reticulocyte count and percentage, as well as immature reticulocyte fraction, are measurements of anucleate immature RBCs. Reticulocyte absolute counts reflect the efficiency of erythropoiesis in the bone marrow, whereas their relative number is indicative of RBC life span [227]. Immature reticulocyte fraction is a relative measurement of immature reticulocytes that have higher RNA content, and it is an early and sensitive parameter for evaluating erythropoiesis [227]. Analysis of RBC parameters assessing mature and immature features of RBCs, showed that male and female Vel-negative blood donors when compared with control donors, had a weak but significantly decreased of: RBC count (RBC (10^{12}/L): M- 4.83±0.24 , p=0.0098; F- 4.45±0.25, p=0.0332), reticulocyte percentage (Ret (%): M- 0.79±0.21, p=0.0223; F- 0.66±0.21, p=0.0115), reticulocyte count (Ret (10^{12}/L): M- 0.04±0.01, p=0.0038; F- 0.03±0.01, p=0.0443) and immature reticulocyte fraction (IRF (%): M-2.24±1.73, p=0.0024; F- 1.76±3.25, p=0.0449). RDW-standard deviation (RDW-SD (%): M-44.21±3.33, p=0.0081; F- 45.60±3.29, p=0.0283), in turn, was slightly but significantly increased when compared with control donors (Table 4.2 and Figure 4.8). Although the alterations observed in this analysis were subtle and within the reference range, these are significant and therefore suggest that SMIM1 is important during erythropoiesis and possibly for the maturation process of RBCs and consequently, for RBC function.

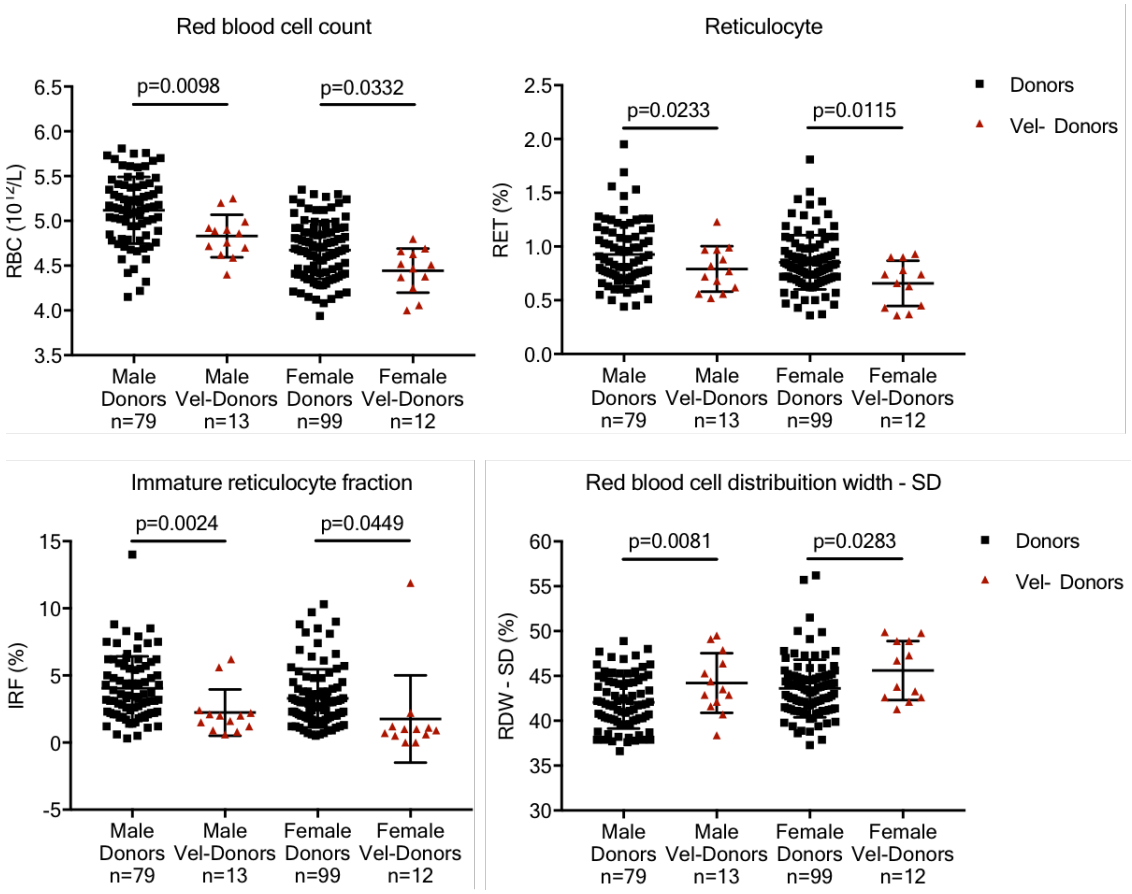


Figure 4.8 - Effect of SMIM1 absence in RBCs parameters of blood donors. Graphic representation of full blood count results for mature and immature RBCs parameters in control donors and Vel-negative blood donors. Black squares – control donors and red triangles – Vel-negative blood donors. Statistical analysis was performed using a linear model (Parameter ~ cohort + Age + BMI) and Anova type II. Only significant differences are represented.

4.3.3.1.2 Analysis of platelets parameters in Vel-negative blood donor cohort

Different parameters of mature and immature platelet features were also investigated using Sysmex XE-5000 hematology analyzer. Analysis of platelet parameters showed that male Vel-negative blood donors, when compared with the control donors exhibited a weak but significant increase of mean platelet volume (MPV (fL): M- 11.15±0.90, p=0.0408) and platelet large cell ratio (P-LCR (%): 34.96±7.84, p=0.0274) (Table 4.2 and Figure 4.9). On other hand, female Vel-negative blood donors presented a weak but significant increase of platelet count (PLT (10^9/L): F- 270±61.17, p=0.0003), plateletcrit (PCT (%): F- 0.30±0.06, p=6.4 x 10^{-6} – total platelet mass) and immature platelet fraction (IPF (%): F- 6.89±3.69, p=0.0037 – fraction of reticulated platelets with greater amount of residual RNA[228]), when compared with control donors (Table 4.2 and

Figure 4.9). These observations suggest that SMIM1 plays a role in the regulation of platelets traits such as production, count and volume.

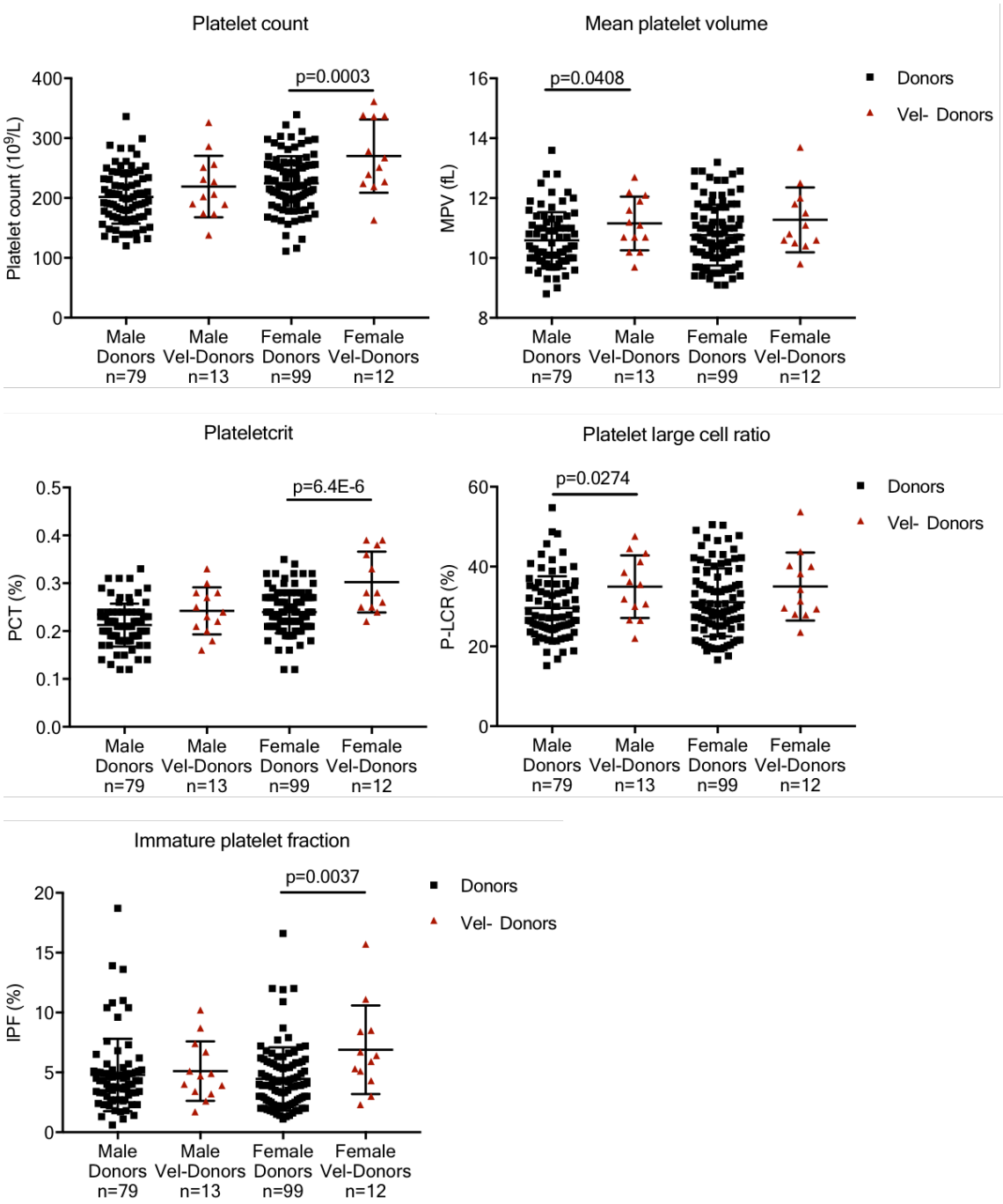


Figure 4.9 – Effect of SMIM1 absence in platelet parameters of blood donors. Graphic representation of full blood count results for platelet-related parameters in control donors and Vel-negative blood donors. Black squares – control donors and red triangles – Vel-negative blood donors. The analysis was performed using a linear model (Parameter ~ cohort + Age + BMI) and Anova type II. Only significant differences are represented.

4.3.3.1.3 Analysis of WBC parameters in Vel-negative blood donor cohort

Analysis of WBC parameters showed that male Vel-negative blood donors presented a weak but significant increase in WBC count (WBC (10^9/L): M- 6.67±2.09, p=0.0243), neutrophils count (NEUT (10^9/L): M- 3.93±1.34, p=0.0065) and neutrophils percentage (NEUT (%): M- 59.20±7.99, p=0.0362) (Table 4.2 and Figure 4.10). Female Vel-negative blood donors did not present any significant difference in WBC parameters when compared to control donors. These observations suggest that SMIM1 role in neutrophils is sex-dependent.

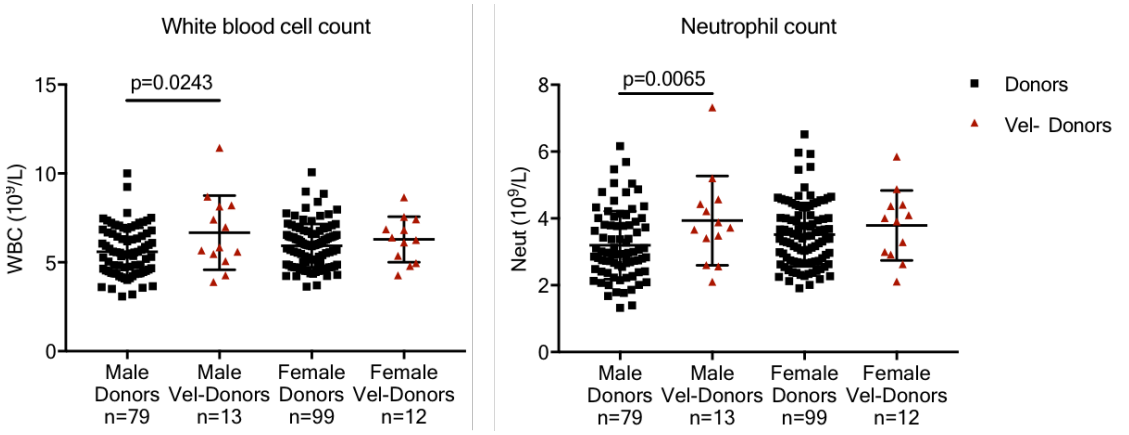


Figure 4.10 - Effect of SMIM1 absence in WBCs parameters of blood donors. Graphic representation of full blood count results for WBC and neutrophil counts in control donors and Vel-negative blood donors. Black squares – control donors and red triangles – Vel-negative blood donors. Data represent mean ± standard deviation. The analysis was performed using a linear model (Parameter ~ cohort + Age + BMI) and Anova type II. Only significant differences are represented.

As shown above, in section 4.3.1.2, SMIM1 affects BMI. However, it is still unknown if this phenotype is a direct or indirect effect of SMIM1 ablation. Therefore, in order not to disregard any phenotype, I performed the analysis of full blood count also without BMI as a covariate. Similar phenotypic changes were found when not accounting for BMI as a covariate (data not shown). However, some differences were observed in reticulocytes and neutrophils. Reticulocyte count and percentage, as well as the neutrophil percentage in male Vel-negative blood donors, were not significantly different when not using BMI as a covariate. In female Vel-negative blood donors, immature reticulocyte fraction was not significantly different when not using BMI as a covariate. These observations suggest that the phenotype observed in these parameters is also directly affected by BMI. Altogether, this analysis showed that SMIM1 is not

crucial but affects different parameters of RBC, platelets and WBCs. Moreover, the effect of SMIM1 in particular parameters seems to be sex-dependent.

4.3.3.2 Quantitative analysis of blood cell parameters in Vel-negative/weak individuals of the UK Biobank

Using the subset of individuals of European ancestry in the most recent data release of approximately 400,000 participants (244,135 female and 205,576 male) in the UK Biobank study, Dr William Astle, Lecturer in Haematological Genomics at University of Cambridge, performed an association analysis of *SMIM1* 17-bp deletion – rs566629828 (Vel-negative/weak individuals) with 26 blood parameters (Table 4.3, unpublished data – association analysis performed in individuals homozygous (n=79: female=40 and male=39, MAF=0.02%) and heterozygous (n=11998: female=6465, male=5533, MAF=2.7%) for the 17-bp deletion in *SMIM1*). This analysis showed a positive association of rs566629828 with RDW-CV (p=2.8 x 10^{-50}) and mean sphered corpuscular volume (MSCV, p=4.19 x 10^{-10}). Both parameters measure features correlated with RBC size. As described above, RDW measures the variability of RBCs sizes, whereas MSCV is the measurement of RBC volume in hypo-osmotic solution. The increased MSCV in individuals carrying the *SMIM1* 17-bp deletion indicates an increase in the ratio of cell surface area to cell volume. Moreover, a negative association of rs566629828 with IRF (p=1.3 x 10^{-10}), reticulocyte count *(*p=2.6 x 10^{-21}) and percentage (p=4.3 x 10^{-20}), HGB (p=0.0015), MCHC (p=7.0 x 10^{-8}), platelet count (p=0.0003) and plateletcrit (p=0.0002) was observed. These observations suggest that SMIM1 is important for erythropoiesis and platelet production. Differences in effect of SMIM1 on HGB, MCHC, platelet count and plateletcrit measurements were observed between the two studies, the UK Biobank (Vel-negative/weak individuals) and my study (Vel-negative blood donors). These differences may be due to the fact that: i) the individuals in my study are all Vel-negative (homozygous for the *SMIM1* 17-bp deletion), while the majority of the individuals in the UK Biobank are heterozygous for the *SMIM1* 17-bp deletion; and ii) the individuals in my study are active blood donors. Vel-negative blood donors are not only selected based on HGB levels (HGB level needs to be at least 125g/L for women and 135g/L for men) but are also subject to repeated blood donations that alter their iron homeostasis [229].

Table 4.3 – Association analysis of rs566629828 with blood parameters in UK Biobank individuals.

Parameter	P-values average	Effect avarege	Effect in female	Effect in male	P-value for sex interaction
RBC (10^12/L)	0.1086	-0.0148	-0.0155	-0.0139	0.9308
HGB (g/dL)	0.0016	-0.0291	-0.0382	-0.0185	0.2870
HCT (%)	0.2020	-0.0118	-0.0184	-0.0040	0.4353
MCV (fL)	0.2738	0.0101	0.0007	0.0210	0.2712
MCH (pg)	0.1480	-0.0134	-0.0232	-0.0020	0.2509
MCHC (g/dL)	7.0E-08	-0.0497	-0.0564	-0.0421	0.4405
RET (10^12/L)	2.7E-21	-0.0886	-0.0743	-0.1052	0.0993
RET (%)	4.3E-20	-0.0859	-0.0711	-0.1030	0.0888
IRF (%)	1.3E-10	-0.0601	-0.0483	-0.0737	0.1768
MSCV (fL)	4.2E-10	0.0584	0.0554	0.0620	0.7247
RDW-CV (%)	2.8E-50	0.1380	0.1623	0.1101	0.0049
PLT (10^9/L)	0.0003	-0.0337	-0.0355	-0.0317	0.8434
MPV (fL)	0.6810	0.0039	0.0075	-0.0002	0.6830
PDW (fL)	0.9659	-0.0004	0.0091	-0.0113	0.2805
PCT (%)	0.0002	-0.0356	-0.0354	-0.0360	0.9758
WBC (10^9/L)	0.2124	-0.0115	-0.0091	-0.0143	0.7776
NEUT (10^9/L)	0.3026	-0.0095	-0.0135	-0.0049	0.6392
LYMPH (10^9/L	0.4411	-0.0071	0.0039	-0.0198	0.2009
MONO (10^9/L)	0.1707	-0.0127	-0.0149	-0.0101	0.7970
BASO (10^9/L)	0.4636	-0.0068	-0.0111	-0.0018	0.6173
EO (10^9/L)	0.6585	0.0041	0.0094	-0.0021	0.5374
NEUT (%)	0.8288	-0.0020	-0.0136	0.0115	0.1744
EO (%)	0.4205	0.0074	0.0140	-0.0001	0.4478
LYMPH (%)	0.6572	0.0041	0.0145	-0.0080	0.2255
MONO (%)	0.5724	-0.0052	-0.0090	-0.0009	0.6616
BASO (%)	0.8594	-0.0016	-0.0065	0.0039	0.5749

Effect average: the average additive allelic effect size combining males and females; P-value average: P-value for test against the null hypothesis that average effect is zero; Effect in female: additive allelic effect size estimated within females; Effect in males: additive allelic effect size estimated within males; P-value for sex_interaction: P-value for test against null hypothesis that effect size is the same in males and females. For all effect sizes: the baseline allele is the wildtype allele and the units are standard deviations of the inverse rank normalised trait; RBC – red blood cell count; HGB – haemoglobin concentration; HCT – haematocrit; MCV – mean cell volume; MCH – mean cell haemoglobin; MCHC – mean cell haemoglobin concentration; RET – reticulocyte count or percentage; IRF – immature reticulocyte fraction; RDW-SD – red blood cell distribution width - standard deviation; RDW-CV – red blood cell distribution width - coefficient of variation; MicroR – microcytic RBC count; MacroR – macrocytic RBC count; PLT – platelet count; MPV – mean platelet volume; PDW – platelet distribution width; PCT – plateletcrit; P-LCR – platelet larger cell ratio; IPF – immature platelet fraction; WBC – white blood cell count; NEUT – neutrophil count or percentage; LYMPH – lymphocyte count or percentage; Mono – monocyte count or percentage; Eo – eosinophil count or percentage and Baso – basophil count or percentage.

4.3.3.3 Qualitative and quantitative analysis of blood cell parameters in *Smim1* mutant mice

To characterise the possible role of SMIM1 in different blood parameters in mice, I performed terminal blood smears and full blood counts in mice of two age groups: i) young - < six months (thirteen to twenty-one weeks), and ii) mature - ≥ six months (twenty-four to thirty-five weeks). Examination of blood smears did not show any striking phenotype as observed in Vel-negative blood donors (data not shown), whereas analysis of full blood counts showed that SMIM1 affected different blood lineages in a sex-age-dependent manner (Table 4.4 and 4.5). This analysis was performed using a linear model with age in weeks as a covariate. In addition, since SMIM1 affects body weight in mice, I also performed this analysis with both age in weeks and the body weight as covariates. Similar phenotypic changes were found in both analyses (with and without body weight as covariant - data not shown) in mice. Taking that into consideration and since the dataset without body weight was considerably larger, which improves the power of the statistic test, the analysis without body weight as covariant was preferred and is presented in Table 4.4 and 4.5. The dataset without body weight was considerably larger because I performed full blood counts in several mice that were not weighted.

4.3.3.3.1 Analysis of full blood counts in male Smim1 mutant mice

Analysis of full blood counts in male *Smim1* mutant mice showed that WBC count and different RBC parameters were affected (Table 4.4 and Figure 4.11). This analysis showed, however, that the effect of SMIM1 in RBC parameters is age-dependent, since opposite trends and effects were observed between age groups.

Young *Smim1*$^{-/-}$ male

Young *Smim1*$^{-/-}$ male mice showed a significant decrease of RBC count (9.03±0.50 10^6/µl, p=0.0391) and haematocrit (HCT: 42.66±2.46 %, p=0.0386) when compared with wildtype mice. Additionally, young *Smim1*$^{-/-}$ male mice showed a significant increase of RDW (16.15±0.56 %, p=0.0029) and MPV (5.21±0.14 µm^3, p=0.0189) when compared with *Smim1*$^{+/-}$ male mice (Table 4.4 and Figure 4.11).

Mature *Smim1*$^{-/-}$ male

Mature *Smim1*$^{-/-}$ male mice showed a significant increase of WBC count (8.74±1.63 10^3/µl, p=0.0470), RBC count (9.40±1.12 10^6/µl, p=0.0142), HGB (14.33±1.42 g/dL, p=0.0012), HCT (43.61±5.41 %, p=0.0044) and MCV (46.40±1 µm^3, p=0.0148) when compared with age-matched wildtype mice. Moreover, mature *Smim1*$^{-/-}$ male mice also showed a significant increase of RBC count (p=0.0012), HGB (p=0.0002), HCT (p=0.002) and MCV (p=0.0367) when compared with *Smim1*$^{+/-}$ male mice (Table 4.4 and Figure 4.11).

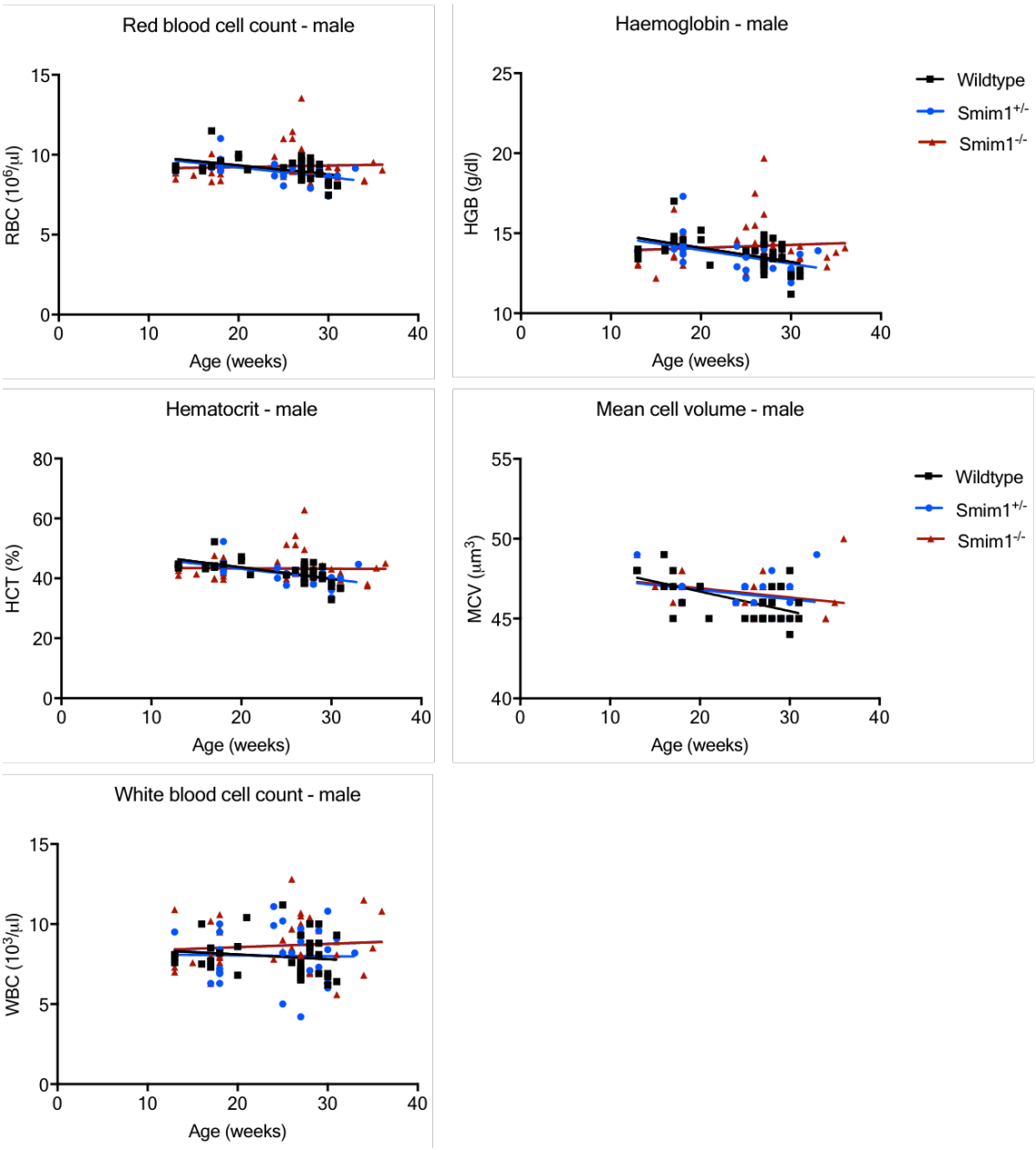


Figure 4.11 – Effect of SMIM1 ablation in blood parameters of male mice. Graphic representation of full blood count results for RBC parameters and WBC count. Black squares and trend line – wildtype mice, blue circles and trend line - *Smim1*$^{+/-}$ mice and red triangles and trend line – *Smim1*$^{-/-}$ mice.

Altogether, these results indicate that *Smim1*$^{-/-}$ and *Smim1*$^{+/-}$ male mice have distinct trends in some parameters, and that SMIM1 affects erythropoiesis and WBC production in male mice. Furthermore, the opposite trends of RBC parameters observed in the two age groups of *Smim1*$^{-/-}$ male mice, may indicate that the phenotype observed in mature mice results from a compensatory mechanism or is associated with a metabolic alteration, as suggested by body weight alteration.

Table 4.4 – Effect of SMIM1 ablation in blood parameters of male mice. Summary of full blood counts of wildtype, *Smim1*$^{+/-}$ and *Smim1*$^{-/-}$ young and mature mice, obtained by an automated haematology analyser. Data represent mean ± standard deviation. The analysis was performed using a linear model (Parameter ~ Genotype + Age (weeks)).

Male		<6 Months							>6 Months						
Parameter		Wildtype (n=12)		*Smim1*$^{+/-}$ (n=10)		*Smim1*$^{-/-}$ (n=15)		p-value	Wildtype (n=24)		*Smim1*$^{+/-}$ (n=20)		*Smim1*$^{-/-}$ (n=30)		p-value
		mean	SD	mean	SD	mean	SD		mean	SD	mean	SD	mean	SD	
WBC	(10^3/µl)	8.23	1.05	7.88	1.38	8.47	1.38	ns	7.85	1.37	7.94	1.78	8.74	1.63	p=0.0470 (WT*vsSmim1*$^{-/-}$)
RBC	(10^6/µl)	9.52	0.70	9.48	0.60	9.03	0.50	p=0.0391 (WT*vsSmim1*$^{-/-}$)	8.88	0.63	8.69	0.53	9.40	1.12	p=0.0142 (WT*vsSmim1*$^{-/-}$) p=0.0012 (*Smim1*$^{+/-}$ *vsSmim 1*$^{-/-}$)
HGB	(g/dl)	14.40	1.02	14.35	1.14	13.85	1.07	ns	13.37	0.89	13.23	0.73	14.33	1.42	p=0.0012 (WT*vsSmim1*$^{-/-}$) p=0.0002 (*Smim1*$^{+/-}$ *vsSmim 1*$^{-/-}$)
HCT	(%)	44.91	2.72	44.67	2.83	42.66	2.46	p=0.0386 (WT*vsSmim1*$^{-/-}$)	40.60	3.04	40.30	2.61	43.61	5.41	p=0.0044 (WT*vsSmim1*$^{-/-}$) p=0.0020 (*Smim1*$^{+/-}$ *vsSmim 1*$^{-/-}$)
MCV	(µm^3)	47.08	1.24	47.00	0.94	47.13	0.99	ns	45.71	0.91	46.35	1.09	46.40	1.00	p=0.0148 (WT*vsSmim1*$^{-/-}$) p=0.0367 (*Smim1*$^{+/-}$ *vsSmim1*$^{-/-}$)
MCH	(pg)	15.16	0.42	15.11	0.41	15.43	0.53	ns	15.07	0.38	15.24	0.49	15.32	0.72	ns
MCHC	(g/dl)	32.12	0.88	32.07	0.81	32.59	1.25	ns	32.96	0.76	32.34	2.49	32.96	1.63	ns
RDW	(%)	15.76	0.80	15.21	0.79	16.15	0.56	p=0.0029 (*Smim1*$^{+/-}$ *vsSmim 1*$^{-/-}$)	15.88	0.65	16.12	0.87	16.03	0.65	ns
PLT	(10^3/µl)	1150.00	162.87	1142.30	104.26	1151.39	329.27	ns	1479.04	295.02	1357.20	378.61	1334.67	309.18	ns
MPV	(µm^3)	5.16	0.13	5.08	0.18	5.21	0.14	p=0.0189 (*Smim1*$^{+/-}$ *vsSmim 1*$^{-/-}$)	5.21	0.11	5.24	0.15	5.23	0.15	ns

WBC – white blood cell count; RBC – red blood cell count; HGB – haemoglobin concentration; HCT – haematocrit; MCV – mean cell volume; MCH – mean cell haemoglobin; MCHC – mean cell haemoglobin concentration; RDW – red blood cell distribution width; PLT – platelet count; and MPV – mean platelet volume

4.3.3.3.2 Analysis of full blood counts in female Smim1 mutant mice

Analysis of full blood counts in female *Smim1* mutant mice showed that ablation of SMIM1 only affected MCV in young female mice and WBC parameters in mature female mice (Table 4.5 and Figure 4.12). The alteration of WBC parameters, however, did not seem to be age-dependent, since female *Smim1* mutant mice from both group ages presented the same trends. Young *Smim1*$^{-/-}$ female mice showed a significant increase of MCV (48.50±1.68 μm^3, p=0.0255) when compared with wildtype. However, this was the only RBC parameter that was altered, and no differences in MCV were observed in mature mice. Mature *Smim1*$^{-/-}$ female mice showed a significant increase of WBC count (6.99±1.10 10^3/μl, p=0.0054) when compared to wildtype (Table 4.5 and Figure 4.12), suggesting that SMIM1 is important for WBC production.

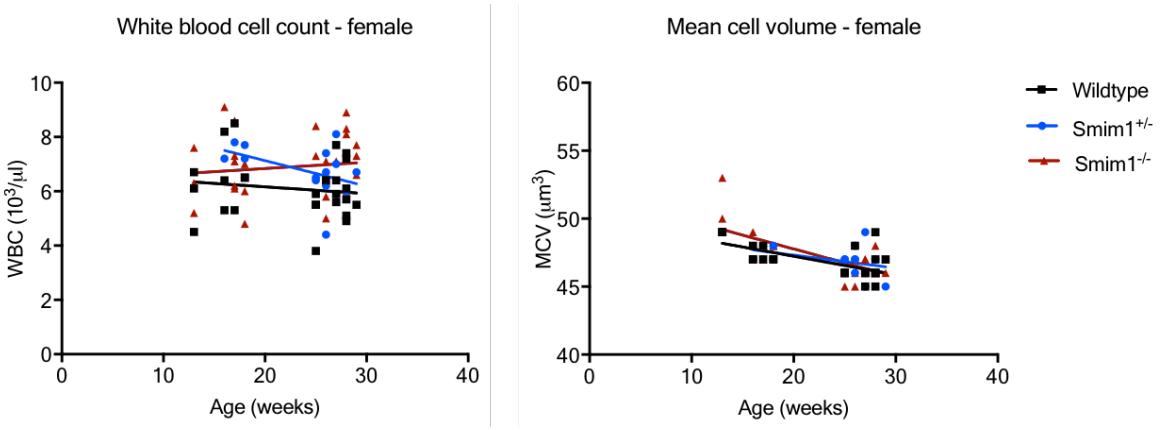


Figure 4.12 – Effect of SMIM1 ablation in blood parameters of female mice. Graphic representation of full blood count results for WBC count and MCV versus age in weeks. Black squares and trend line – wildtype mice, blue circles and trend line- *Smim1*$^{+/-}$ mice and red triangles and trend line– *Smim1*$^{-/-}$ mice.

In summary, SMIM1 affects different blood parameters in a sex-dependent manner in mice. SMIM1 seems to have a stronger effect on RBC parameters in males and WBCs parameters in females. Platelet parameters, however, did not show any significant differences in both sexes, as opposed to what was observed in humans.

Table 4.5 – Effect of SMIM1 ablation in blood parameters of female mice. Summary of full blood counts of wildtype, *Smim1*$^{+/-}$ and *Smim1*$^{-/-}$ young and mature mice, obtained by an automated haematology analyser. Data represent mean ± standard deviation. The analysis was performed using a linear model (Parameter ~ Genotype + Age (weeks)).

Female		<6 Months							>6 Months						
Parameter		Wildtype (n=10)		*Smim1*$^{+/-}$ (n=4)		*Smim1*$^{-/-}$ (n=12)		p-value	Wildtype (n=16)		*Smim1*$^{+/-}$ (n=11)		*Smim1*$^{-/-}$ (n=17)		p-value
		mean	SD	mean	SD	mean	SD		mean	SD	mean	SD	mean	SD	
WBC	(10^3/µl)	6.40	1.24	7.48	0.32	6.78	1.27	ns	5.91	0.98	6.48	0.94	6.99	1.10	p=0.0054 (WTvs*Smim1*$^{-/-}$)
RBC	(10^6/µl)	8.46	0.48	8.96	0.45	8.44	0.57	ns	8.99	0.70	8.69	0.69	9.29	1.07	ns
HGB	(g/dl)	13.40	0.72	14.20	0.71	13.48	0.82	ns	13.78	1.03	13.45	0.90	14.42	1.62	ns
HCT	(%)	40.40	2.35	42.58	2.20	40.80	3.13	ns	41.74	3.87	40.54	3.38	43.14	5.58	ns
MCV	(µm^3)	47.80	0.92	47.50	0.58	48.50	1.68	p=0.0255 (WT*vs*Smim1$^{-/-}$)	46.31	1.01	46.73	1.19	46.35	0.79	ns
MCH	(pg)	15.86	0.44	16.10	1.00	16.01	0.67	ns	15.36	0.48	15.48	0.55	15.54	0.54	ns
MCHC	(g/dl)	33.20	1.33	33.40	1.62	33.15	1.93	ns	33.08	1.04	33.25	1.17	33.51	1.14	ns
RDW	(%)	16.33	0.60	15.98	0.46	16.33	0.42	ns	15.71	0.45	15.73	0.58	15.66	0.56	ns
PLT	(10^3/µl)	1025.50	78.88	974.25	113.97	1005.42	174.39	ns	1077.50	229.09	1097.18	343.46	1089.65	170.65	ns
MPV	(µm^3)	5.17	0.09	5.23	0.10	5.18	0.07	ns	5.23	0.13	5.33	0.17	5.22	0.16	ns

WBC – white blood cell count; RBC – red blood cell count; HGB – haemoglobin concentration; HCT – haematocrit; MCV – mean cell volume; MCH – mean cell haemoglobin; MCHC – mean cell haemoglobin concentration; RDW – red blood cell distribution width; PLT – platelet count; and MPV – mean platelet volume.

4.3.3.3.3 Analysis of full blood counts of Smim1$^{-/-}$ mice of the International Mouse Phenotyping Consortium

Smim1 mutant mice were generated under the International Mouse Phenotyping Consortium and the SMIM1 phenotypic data from this consortium became recently available for the community. Therefore, I further analysed the available blood parameters of 16 weeks-old *Smim1* mutant mice (equivalent to the young mice group used in this study). In order to control for daily variation of the automated haematology analyser, I performed the analysis using the data of *Smim1* mutant mice (seven males and seven females) and day control wildtype mice. This analysis showed similar trends to the ones observed in mature mice described above. However, some differences were also observed. Specifically, *Smim1*$^{-/-}$ male mice showed a significant increase of RBC count (10.31±0.28 10^6/µl, p=0.01), HGB (14.79±0.39 g/dl, p=0.03), HCT (46.80±1.26 %, p=0.013), decrease of MCH (14.33±0.14 pg, p=0.031) and MCHC (31.53±0.22 g/dl, p=0.015) but not MCV, whereas *Smim1*$^{-/-}$ female mice showed an increase of WBC count (6.14±01.18 10^3/µl, p=0.0072), MCV (46.14±0.38 µm^3, p=0.0199) and also MCH (14.91±0.36 g/dl, p=0.0101) (Table 4.6). The differences observed in the two studies are related to alterations in MCV. These results indicate that the phenotypes observed in mature mice, were present earlier in the Sanger *Smim1* mutant mice, suggesting that the phenotype may be influenced by environmental factors that are related with different housing conditions and/or by the haematology analyser used.

Table 4.6 – Effect of SMIM1 ablation in blood parameters of mice of the International Mouse Phenotyping Consortium. Data represent mean ± standard deviation (SD). The analysis was performed using an unpaired two tailed T-test or Mann-Whitney test.

Parameter		Male					Female				
		Wildtype (n=35)		*Smim1*$^{-/-}$ (n=7)		p-value	Wildtype (n=34)		*Smim1*$^{-/-}$ (n=7)		p-value
		mean	SD	mean	SD		mean	SD	mean	SD	
WBC	(10^3/µl)	6.31	0.96	6.17	1.30	ns	5.05	0.87	6.14	1.18	0.0072
RBC	(10^6/µl)	9.89	0.39	10.31	0.28	0.010	9.98	0.28	10.01	0.36	ns
HGB	(g/dl)	14.35	0.48	14.79	0.39	0.030	14.64	0.40	14.91	0.53	ns
HCT	(%)	45.06	1.66	46.80	1.26	0.013	45.55	1.34	46.17	1.79	ns
MCV	(µm^3)	45.51	0.66	45.50	0.55	ns	45.59	0.56	46.14	0.38	0.0199
MCH	(pg)	14.52	0.22	14.33	0.14	0.031	14.67	0.17	14.91	0.36	0.0101
MCHC	(g/dl)	31.87	0.34	31.53	0.22	0.015	32.14	0.25	32.30	0.64	ns
RDW	(%)	14.72	0.36	14.47	0.25	ns	14.43	0.21	14.47	0.27	ns
PLT	(10^3/µl)	980.00	144.55	988.29	44.12	ns	848.47	63.34	895.71	65.95	ns
MPV	(µm^3)	5.07	0.17	4.91	0.11	0.016	5.06	0.17	5.03	0.17	ns

WBC – white blood cell count; RBC – red blood cell count; HGB – haemoglobin concentration; HCT – haematocrit; MCV – mean cell volume; MCH – mean cell haemoglobin; MCHC – mean cell haemoglobin concentration; RDW – red blood cell distribution width; PLT – platelet count; and MPV – mean platelet volume.

4.3.3.3.4 Assessment of RBCs deformability of Smim1 mutant mice by real-time deformability cytometry

RBCs morphologic and mechanical properties are critical for their role of oxygen transport from the lungs to the tissues [155, 230]. The deformability of RBCs, which is dependent on structures of the membrane cortex and viscosity of the cell cytoplasm [231], determines RBCs ability to pass through the narrowest capillaries in the tissues and therefore alterations in RBC deformability have been associated with various conditions such as hereditary disorders and metabolic disorders [155, 232-236]. Since SMIM1 is a membrane protein in RBCs and its expression is associated with alteration of parameters like RDW and MCV, I further investigated if the deformability, smoothness and area of *Smim1* mutant mice RBCs were affected by real-time deformability cytometry. This experiment was performed by Ms Katie Bashant – PhD student in Professor Dr Edwin Chilvers group – University of Cambridge, and a more detailed explanation of the technique is described in section 2.2.15. Analysis of whole blood of six months old mice by real-time deformability cytometry showed that SMIM1 did not affect RBC area and smoothness (Figure 4.13 - A) in either males or females, suggesting no alteration in skeletal integrity of the cell membrane. Moreover, analysis of RBCs that passed through a microfluid constriction channel under pressure gradients and shear stresses that induced cell deformation showed that SMIM1 did not affect female RBCs deformability but interestingly affected male RBCs deformability (Figure

4.13 - B). Male *Smim1*$^{-/-}$ RBCs were significantly less prone to deformation (p=0.0338), suggesting that the lack of SMIM1 increases RBC stiffness. However, this experiment was performed only in three male mice per genotype and therefore it would be important to increase the number of male mice analysed in order to draw further conclusions.

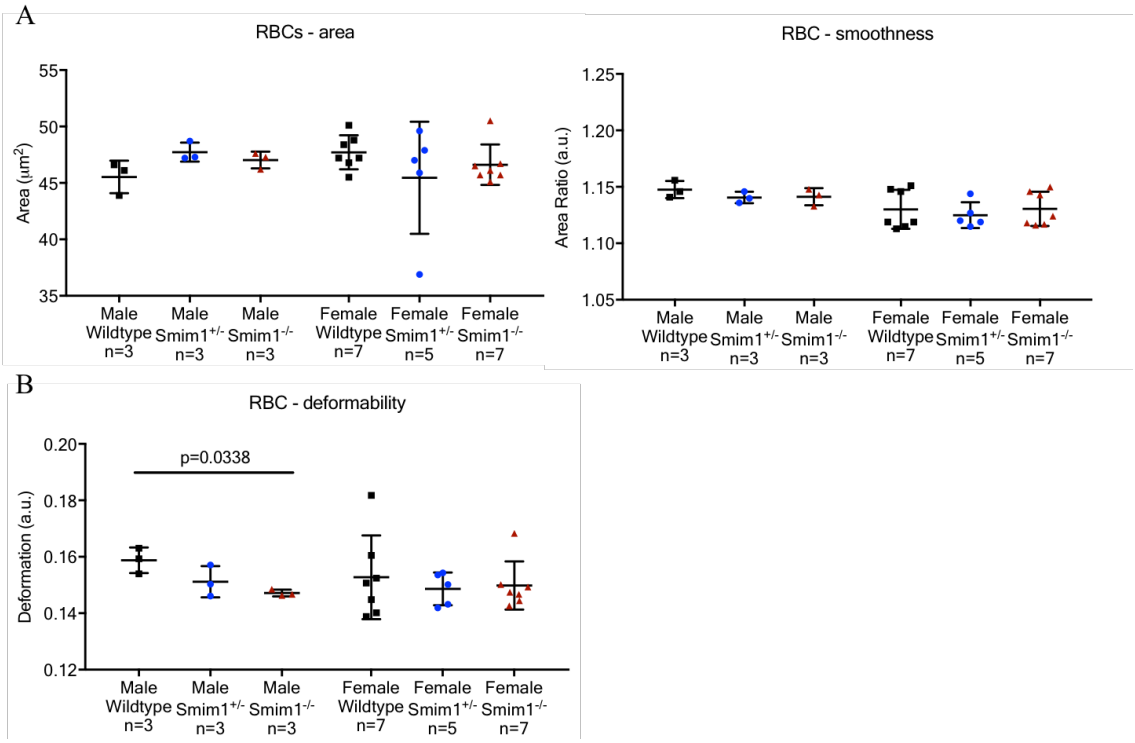


Figure 4.13 - Effect of SMIM1 ablation on morphologic and mechanical properties of RBCs of mice. Data represent mean ± standard deviation. The analysis was performed using Anova and Tukey's multiple comparisons test. Black squares – wildtype mice, blue circles - *Smim1*$^{+/-}$ mice and red triangles – *Smim1*$^{-/-}$ mice. Only significant differences are represented.

4.4 Discussion

Haemolytic reactions during transfusions and haemolytic disease in newborns are the only known clinical manifestations related to the absence of SMIM1. Nevertheless, the alteration of RBC parameters associated with SMIM1 common variant, rs1175550, and with the 17-bp deletion responsible for the Vel-negative phenotype, suggests that SMIM1 is important for RBC biology. Also, the high expression levels of *SMIM1* in other blood cells such as MKs and platelets, and in non-haematopoietic tissues (Chapter 3), may indicate that SMIM1 plays a role in MK and/or platelet physiology, as well as in other tissues. Despite this evidence, a deeper understanding of SMIM1 role in RBC biology and the investigation of its possible role in other blood cells and non-haematopoietic tissues is still missing.

In this chapter, using two human cohorts (Vel-negative blood donors recruited in this study and Vel-negative/weak individuals of the released 400,000 UK Biobank study) and a *Smim1* mutant mouse model, I assessed the effect of SMIM1 absence on: i) different blood traits; ii) BMI and body weight, iii) organ to body weight ratios and iv) tissue morphology. I have demonstrated that SMIM1 is associated with weak but significant alterations in several blood traits, specifically in RBC, platelets and neutrophils. However, the directionality of these alterations and their significance was model specific. Besides, some of the observed associations were sex-dependent in both human and mice, suggesting a possible hormonal regulation. Additionally, the male mouse model showed that some of the effects were also age-dependent. Here, I described how SMIM1 is associated with two metabolic phenotypes. Specifically, analysis of BMI and body weight in humans and mice, respectively, revealed that absence of SMIM1 leads to an increase in BMI of male Vel-negative blood donors, and increased body weight in *Smim1*$^{-/-}$ female mice. Histological and organ to body weight ratio analyses uncovered an alteration in liver morphology in male *Smim1* mutant mice, with potential clinical implications.

Analysis of blood traits in Vel-negative blood donors

i) RBCs

Examination of RBC traits in the Vel-negative blood donor cohort has revealed that the absence of SMIM1 results in an impairment in erythropoiesis in both female and male

blood donors (Table 4.7). The number of reticulocytes and immature reticulocytes, which are the immature RBCs released in the bloodstream few days before their maturation is complete, was decreased, and consequently, the RBC count was also reduced. Although the reduction of reticulocytes observed falls within the reference range, these observations suggest an inappropriate bone marrow response during RBC formation. This might be caused by an alteration of erythrocyte proliferation, maturation or rate of delivery into the peripheral blood. A previous study performed by Storry *et al*. showed that *in vitro* $CD34^{+}$ cells, as well as bone marrow cells cultured towards erythropoiesis, presented an increased *SMIM1* expression towards the end of the process, suggesting that *SMIM1* may play a role during erythroid maturation [188]. This observation supports the idea that the reduction of immature reticulocytes and reticulocytes in Vel-negative blood donors results from an impairment of erythroid maturation. Additionally, Cveijc *et al.* reported similar findings in *Smim1* knockdown zebrafish [186].

This analysis has also shown an increase in RBC size variability in Vel-negative blood donors, which has been previously described by Astle *et al*., [196]. Morphologic examination of blood smears, however, did not show any striking anisocytosis. This is not surprising since the observed alteration in RDW was subtle and falls within the reference range. The increased RBC size variability observed in Vel-negative blood donors, suggests that SMIM1 is important for the regulation of RBCs size. The size and shape of RBCs is largely dependent of its membrane complex structure. The membrane of RBCs comprises a phospholipid bilayer with membrane skeletal proteins and a complex network of cytoskeleton proteins (spectrin) that connect/interact with the bilayer. As SMIM1 is a membrane protein on RBCs surface, its absence in the phospholipid bilayer may disrupt this complex network, which may alter not only their size but also their deformability. RBC small size and flexibility are essential features of its biological function – those features enable RBC to pass through the narrowest capillaries and consequently delivery oxygen to the tissues. Together, these findings suggest that SMIM1 is important for erythroid maturation and RBC size. These phenotypic alterations, in turn, may impact RBC function, with physiological implications yet to be identified. For instance, one could speculate that absence of SMIM1 might be associated with higher incidence of different conditions related to an increase in RBC size variability, such as cardiovascular diseases and liver disease [237, 238].

ii) Neutrophils and platelets

In addition to its function in erythropoiesis, I observed that SMIM1 has a positive association with platelet and neutrophil production in blood donors (Table 4.7). Interestingly, this association seems to be sex-dependent. While there was no morphological alteration in neutrophils of male Vel-negative blood donors, I observed a weak but significant increase in their number. An increase of absolute neutrophil count, can result from: i) increased proliferation and differentiation of precursor cells in the bone marrow; ii) increased release of mature neutrophils from the bone marrow into the blood; iii) alteration in margination and egress of neutrophils from the blood into the tissues (where neutrophils play their function); and/or iv) decreased clearance and spontaneous apoptosis of mature neutrophils [218, 239]. These processes are regulated by a highly dynamic feedback system that is influenced by genetic and environmental factors (see for review [240]). Sexual-dimorphism of neutrophil biology has previously been described, as well as some of the responsible mechanisms. For example, the regulation of neutrophil apoptosis is mediated by female hormones via an oestradiol-receptor mechanism causing a delay in apoptosis. Both oestradiol receptors subtypes (ER-α and β) are present in male and female neutrophils, however, the response of ER-β to oestradiol is sex-dependent [241]. This regulatory mechanism seems to be associated with physiologic neutrophilia during pregnancy and labour [242]. Another example of neutrophil sexual-dimorphism is the stronger response of male neutrophils to chemokine CXCL6 in diverse acute inflammatory situations, resulting in increased stimulation and trafficking of bone marrow neutrophils [243]. Furthermore, several studies have shown that neutrophil sexual-dimorphism results in different prevalence, severity and mortality of disorders related with prolonged or excessive innate immune responses [244-246]. As previously described in Chapter 3, SMIM1 seems to be expressed on the neutrophil membrane. Thus, one could speculate that exposure of SMIM1 in the membrane of neutrophils may contribute to the retention mechanism within the bone marrow or to their clearance/apoptosis, which may be hormonally regulated. Moreover, Grassi *et al.* have shown that *SMIM1* is highly expressed in the initial steps of neutrophil differentiation, which are important for the establishment of neutrophil cytotoxic capability [247]. Altogether, these observations suggest that SMIM1 plays a role in the function of neutrophils.

Analysis of platelet traits of Vel-negative blood donors on an automated haematology analyser revealed that SMIM1 is associated with platelet production (Table 4.7). The effects of SMIM1 however, manifest in a different sex-dependent manner. In female blood donors, the absence of SMIM1 leads to a higher production of platelets, reflected in higher immature platelet fraction, platelet count and plateletcrit. Male donors lacking SMIM1, in turn, shown production of larger platelets, reflected by an increase of platelet large cell ratio and mean platelet volume. The alterations observed in platelets of female Vel-negative blood donors suggest an increased platelet turnover, while the alterations observed in males suggest an impairment in proplatelet/platelet formation. Immature and large platelets are enzymatically and metabolically more active than mature platelets [248, 249], and are associated with an increased platelet reactivity (increased aggregation, increased expression of adhesion molecules). Importantly, it has been reported that the increase of these traits is associated with an increased risk of myocardial infarction in healthy individuals and with adverse health outcomes in patients with cardiovascular disease [250-255].

Analysis of blood traits in Vel-negative/weak individuals of the UK Biobank

In this chapter, I described an association analysis of Vel-negative/weak individuals (part of 400,000 participants of UK Biobank) with several blood traits. UK Biobank participants were not selected for any specific trait. Though, they are not fully representative of the general population, as they live in less socioeconomically deprived areas and are "healthier" ("healthy participant" selection bias) than the general population (https://www.ukbiobank.ac.uk). The extrapolation of prevalence and incidence rate of Vel-negative blood type in this study is, therefore, not accurate. Nevertheless, the large size of this cohort allows the valid assessment of exposure-disease associations.

i) RBCs

Herein, I showed associations between *SMIM1*-17-bp deletion (rs566629828 – homozygous and heterozygous individuals) and a number of RBC traits, suggesting a role during erythropoiesis (Table 4.7). Specifically, I observed a negative association of rs566629828 with immature reticulocyte and reticulocytes; and a positive association with RDW. These findings are in agreement with observations in Vel-negative blood donors. Furthermore, rs566629828 presented a negative association with HGB and MCHC traits. Together, these observations suggest that SMIM1 may regulate

erythropoiesis and possibly HGB synthesis. However, SMIM1 role in the latter process is likely to be dependent on other factors, since Vel-negative blood donors did not show significant differences in HGB or MCHC. Intriguingly, given the reduced number of reticulocytes, the RBC count was unaffected.

i) Neutrophils and platelets

Contrary to the observations in Vel-negative blood donors, Vel-negative/weak individuals did not show differences in neutrophil count (Table 4.7). Moreover, rs566629828 was negatively associated with platelet count and plateletcrit in both male and female Vel-negative/weak individuals, as opposed to a positive association in Vel-negative donors. This suggests that SMIM1 is a regulator of platelet formation, but its role is influenced by other factors not yet identified. As previously mentioned, the differences observed in both cohorts (Vel-negative blood donors and Vel-negative/weak individuals) may be explained by i) different associations between heterozygous and homozygous individuals and/or ii) the alteration of iron homeostasis in blood donors by repeated blood donation [229]. Altogether, these findings indicate that SMIM1 is important for the regulation of platelet production, with a possible role during megakaryocyte maturation. Consequently, these alterations may impact platelet function, a hypothesis further investigated in Chapter 5.

Analysis of blood traits in *Smim1* mutant mice

Using the *Smim1* mutant mice model, I discovered that the effect of SMIM1 in qualitative and quantitative traits of blood cells in mice are less pronounced and with different trends than the observed in humans (Table 4.7). Mature *Smim1* mutant male mice had an increase of RBC, HGB, HCT and MCV, suggesting that SMIM1 is important for the regulation of erythroid differentiation. As HGB and MCV are associated with characteristics that are developed during erythroid maturation, this may suggest that SMIM1 is important for the regulation of erythroid maturation. Moreover, decreased RBC deformability without alteration of cell smoothness in *Smim1*$^{-/-}$ male mice, suggests that the lack of SMIM1 in the phospholipid bilayer may affect the spectrin-network binding, and in turn, RBC function. Young and mature *Smim1*$^{-/-}$ male mice presented opposite trends in RBC traits (i.e. a decrease of RBC traits of younger male mice). Interestingly, the same inversion of trends was observed in the body weight of these mice. Taking into consideration the opposite trends in RBC parameters, as well as, the alteration in liver morphology in mature mice, the increase observed in RBC

parameters of mature mice could be due to: i) a compensatory mechanism to overcome a deficit of functional RBCs, and/or ii) an impaired clearance of stressed RBCs. The increase of stressed RBCs could contribute to increased levels of tissue-damaging iron, heme and HGB in the circulation, and consequently induce oxidative stress in the liver, possibly explaining the alterations observed in the liver. On the other hand, the liver alterations may also result from an accumulation of lipid droplets, which may be related to increased body weight. In addition, SMIM1 seems to be important for regulation of WBC production and appears to have a stronger effect in female mice, suggesting that SMIM1 role is hormonally regulated. Female hormones have been shown to regulate not only apoptosis in neutrophils (as described above), but also apoptosis and activation of lymphocytes-B, specifically oestrogen protects lymphocyte-B receptor-mediated apoptosis [256].

The differences in the effect of SMIM1 in human and mice blood traits may indicate a distinct molecular mechanism behind its function in blood cells of mice and human. However, one cannot exclude the possibility that those differences could be age-dependent. The mice used in this study were considerably younger than the Vel-negative blood donors. In addition, these mice were not in reproductive senescence [257], while the average of Vel-negative blood donors taking part in this study falls into reproductive senescence. In the eventuality that the effects of SMIM1 are proven to be hormonally regulated, as some of the observations suggest, this can have an impact in the observed phenotypes.

Table 4.7 – Summary of full blood count parameters significantly altered in the different cohorts.

Cohort	Parameter/ alteration	Female		Male	
		Increase	Decrease	Increase	Decrease
Vel-negative blood donors	RBCs	RDW-SD (p=0.0283)	RBC (p=0.0332)	RDW-SD (p=0.0081)	RBC (p=0.0332)
		–	RET (p=0.0043)	–	RET (p=0.0038)
		–	RET % (p=0.0115)	–	RET % (p=0.0233)
		–	IRF (p=0.0449)	–	IRF (p=0.0024)
	Platelets	PLT (p=0.0003)	–	MPV (p=0.0408)	–
		PCT (p=6.4E-6)	–	P-LCR (p=0.0274)	–
		IPF (p=0.0037	–	–	–
	WBCs	–	–	WBC (p=0.0243)	–
		–	–	NEUT (p=0.0065)	–
		–	–	NEUT % (p=0.0362)	–
UK Biobank Vel-negative/weak individuals	RBCs	RDW-CV (p=2.8E-50)	RET (p=2.7E-21)	RDW-CV (p=2.8E-50)	RET (p=2.7E-21)
		MSCV (p=4.2E-10)	RET % (p=4.3E-20)	MSCV (p=4.2E-10)	RET % (p=4.3E-20)
		–	IRF (p=1.3E-10)	–	IRF (p=1.3E-10)
		–	MCHC (p=7.0E-8)	–	MCHC (p=7.0E-8)
		–	HGB (p=0.0016)	–	HGB (p=0.0016)
	Platelets	–	PLT (p=0.0003)	–	PLT (p=0.0003)
		–	PCT (p=0.0002)	–	PCT (p=0.0002)
	WBCs	–	–	–	–
Young (< 6 months) *Smim1-/-* mice	RBCs	MCV (p=0.0255)	–	–	RBC (p=0.0391)
		–	–	–	HCT (p=0.0386)
	Platelets	–	–	–	–
	WBCs	–	–	–	–
Mature (≥ 6 months) *Smim1-/-* mice	RBCs	–	–	RBC (p=0.0142)	–
		–	–	HCT (p=0.0044)	–
		–	–	HGB (p=0.0012)	–
		–	–	MCV (p=0.0148)	–
	Platelets	–	–	–	–
	WBCs	WBC (p=0.0054)	–	WBC (p=0.0470)	–

RBC – red blood cell count; HGB – haemoglobin concentration; HCT – haematocrit; MCV – mean cell volume; MCHC – mean cell haemoglobin concentration; Ret – reticulocyte count; Ret % – reticulocyte percentage; IRF – immature reticulocyte fraction; RDW – red blood cell distribution width; PLT – platelet count; MPV – mean platelet volume; PCT – plateletcrit; P-LCR – platelet larger cell ratio; IPF – immature platelet fraction; WBC – white blood cell count; NEUT – neutrophil count; NEUT % – neutrophil percentage.

Novel metabolic phenotype

The examination of BMI and body weight in Vel-negative blood donors and *Smim1* mutant mice, respectively, showed that the absence of SMIM1 results in an increased BMI in male Vel-negative blood donors, and in an increased body weight in *Smim1*$^{-/-}$ female mice. This implies some common features in both models, that may or may not be a direct consequence of SMIM1 effect. These findings suggest that directly or indirectly SMIM1 may be an important regulator of glucose and/or lipids metabolism and therefore, its ablation results in an increase of body weight. Given the distinct blood phenotypes in both models, this metabolic phenotype is most likely independent of the effect of SMIM1 in blood cells. A tissue-specific *Smim1* knockout in haematopoietic

cells or a bone marrow transplant of a *Smim1*$^{-/-}$ mouse into a wildtype mouse, would be important to clarify if the body weight increase, as well as, the liver alteration observed in male mice are a direct cause of *Smim1* ablation or a consequence of its ablation effect in blood cells. This investigation may have clinical implications not yet identified, since excess body weight is strongly associated with several chronic diseases such as diabetes, and cardiovascular and liver diseases [258].

An obvious pitfall of this study is the small size of both Vel-negative cohorts. It would be important to confirm these findings in a larger study, such as the 500,000 individuals - UK Biobank study and 50,0000 blood donors - INTERVAL study, which will be soon available for the research community. Additionally, it would be important to further investigate the consistency of these findings in a cohort representative of the general population. On the other hand, it would be important to further investigate the molecular mechanism underlying the phenotypes here identified and assess if other metabolic markers of both humans and mice are affected.

In conclusion, my work has shed new light on the role of SMIM1 on multiple blood cell types traits, and it has revealed a novel association with body composition and liver morphology. These findings may have important clinical implications, as several of the associations I described here are known to increase the risk for diseases such as cardiovascular and liver disease.

CHAPTER 5 |
CHARACTERISATION OF SMIM1 ROLE IN MEGAKARYOPOIESIS AND PLATELET FUNCTION

5.1 Introduction

The fact that SMIM1 displays high levels of gene and protein expression in MKs and platelets (Chapter 3), raised the possibility that SMIM1 may play a role in the development and function of these cell types. Indeed, the characterisation of blood traits in human natural SMIM1 knockouts in Chapter 4, showed that SMIM1 absence leads to an alteration of platelet production. Specifically, an increase in platelet counts in Vel-negative/weak donors and a decrease in platelet counts in Vel-negative individuals from UK Biobank was observed. These observations suggest that SMIM1 may be relevant for the regulation of important aspects of megakaryopoiesis and/or thrombopoiesis and also platelet function. The determination of the role of SMIM1 in those processes and its underlying molecular mechanism is yet to be uncovered. Uncovering the underlying molecular mechanism of SMIM1 role in these processes will lead to a better understanding of the biology. In turn, this may provide important information for the development of new therapies to treat conditions related with megakaryopoiesis/thrombopoiesis and to improve protocols for *in vitro* MK/platelet production.

Defects in megakaryopoiesis and thrombopoiesis can lead to quantitative and qualitative defects in platelet production, that may affect platelet function and/or clearance, and culminate in the findings noted in platelet inherited disorders (for review [83, 84, 259]). Defects in megakaryopoiesis and thrombopoiesis arise from alterations in the following stages: MK lineage commitment and differentiation, MK maturation and platelet release [84]. Defects in platelet function result from alterations in the different steps of this process, which are described below [83]. Resting platelets in the bloodstream respond to a blood vessel injury by adhering at the injury site, which then leads to a drastic change of their shape, secretion of their granule content, and aggregation to form a platelet clot [144]. Platelet response is tightly regulated by a complex network of inhibitory and activation signals [136]. Defects in platelet function can, therefore, be categorised as defects in: adhesion, receptors and signalling that control platelet secretion and aggregation, granules biogenesis and secretion, aggregation and procoagulant activity (for review [83, 119]). A dysregulation in platelet responses can either lead to a hyper-reactive or to a hypo-reactive state, which can increase the chance of arterial and venous thrombotic events (e.g. tissue ischemia or infarction) or can lead to bleeding disorders, respectively.

Models to study megakaryopoiesis/thrombopoiesis – forward programming of human iPSC (hiPSC)

The understanding of megakaryopoiesis and thrombopoiesis was greatly advanced with the discovery of mpl receptor and the function of its ligand, TPO, in megakaryopoiesis [39, 56, 260-262]. This discovery and subsequent cloning of TPO allowed the development of several culture systems to study megakaryopoiesis/thrombopoiesis[263-267]. These culture systems have been successfully developed using a variety of human stem cells, including cord blood and adult $CD34^+$-hematopoietic progenitors, embryonic stem cells and more recently iPSC, that all differentiate into MKs. While using $CD34^+$-hematopoietic progenitors has the advantage of allowing the recapitulation of definitive (adult) megakaryopoiesis, this system is donor-dependent and limited: first these cells cannot be maintained indefinitely in culture; second, its genetic manipulation is challenging. Embryonic stem cells and iPSCs thus constitute a good alternative to $CD34^+$-hematopoietic progenitors, as they are able to be maintained in culture and can be easily genetically manipulated. However, the differentiation potential of these cells is limited, because most of the derived-MKs recapitulate primitive (embryonic) megakaryopoiesis [268]. The majority of the culture systems described in the literature are based on a direct sequential differentiation using cytokines and/or stromal cells [263-265]. However, a new method was developed recently, in which the generation of MKs in chemically defined conditions is achieved by using a forward programming approach that relies on the exogenous expression of three transcription factors: GATA1, FLI1 and TAL1. This method gives higher cell yields and purity compared with a standard MK-directed differentiation approach [207].

5.2 Chapter aims and overview

In this chapter I describe two main aims: i) the characterisation of SMIM1 role during megakaryopoiesis/thrombopoiesis and ii) the investigation of SMIM1 role in platelet function.

Using a *SMIM1*- knockout and a *SMIM1*-overexpression hiPSC line, created by CRISPR/Cas9n technology and lentiviral overexpression, respectively, I have investigated the importance of SMIM1 during megakaryopoiesis. I assessed the ability of these hiPSC lines to form MKs and I characterised MK maturation markers and ploidy during this process. These experiments showed that SMIM1 is not crucial for

megakaryopoiesis in this *in vitro* model. Furthermore, analysis of MK number within the bone marrow of *Smim1* mutant mice, as well as analysis of platelet morphology by transmission electron microscopy (TEM), also showed no striking differences, suggesting that SMIM1 is not crucial for MK formation and platelet morphology in the mouse.

Next, I have investigated the role of SMIM1 in platelet function. Firstly, using platelets of Vel-positive donors, I characterised the surface expression of SMIM1 during platelet activation by flow cytometry and Western blot. I showed that upon human platelet activation SMIM1 does not become exposed on the platelet membrane, but interestingly it is phosphorylated, suggesting a function during platelet activation. To further investigate this hypothesis, I have performed different platelet activation assays using platelets of *Smim1* mutant mice. Whereas a decrease of basal P-selectin exposure on platelets of male *Smim1*$^{-/-}$ mice was detected, no difference in platelet response to different agonists or platelet spreading in both male and female *Smim1* mutant mice were observed.

5.3 Results

5.3.1 Assessment of SMIM1 importance during megakaryopoiesis

To investigate the relevance of SMIM1 during megakaryopoiesis, I created a *SMIM1*-knockout hiPSC line and a *SMIM1*-overexpression hiPSC line, which were forward programmed into MKs.

5.3.1.1 hiPSC *SMIM1* knockout using CRISPR/Cas9n

As a first step CRISPR/Cas9n target gene editing was used in hiPSC, to create a *SMIM1* knockout line. Three pairs of sgRNA were designed to target the coding region of the *SMIM1* genomic locus. More specifically, the guide sequences were designed to bind close to the N-terminal of the protein (start codon), increasing the possibility that the indel created would lead to a true null allele (Figure 5.1 – B). The different guide sequences were individually cloned into a plasmid containing a scaffold and the Cas9n (Figure 5.2 – C), as described in section 2.3.2. This strategy was preferentially used because it has been shown to have a reduced number of off-target effects [269]. The sequence of sgRNAs was confirmed by Sanger sequencing and the plasmid was expanded as described in section 2.3.2.2.

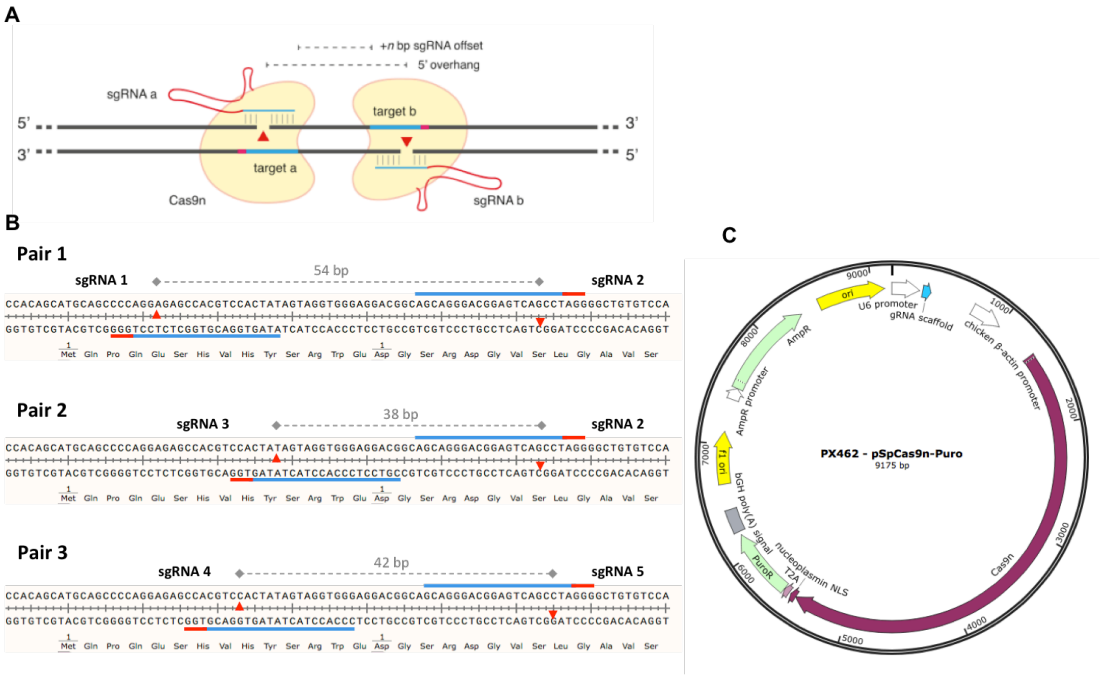
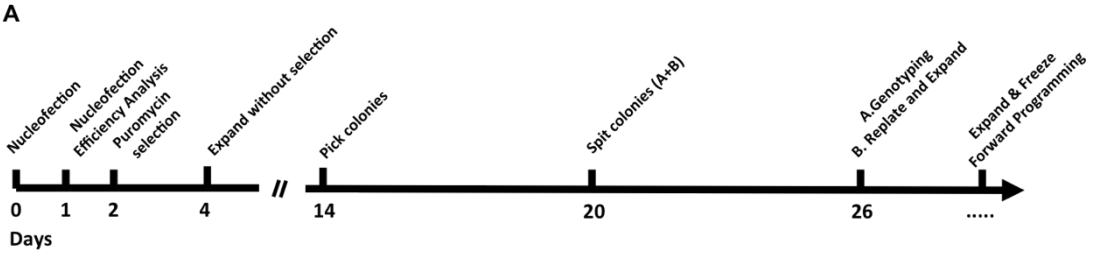


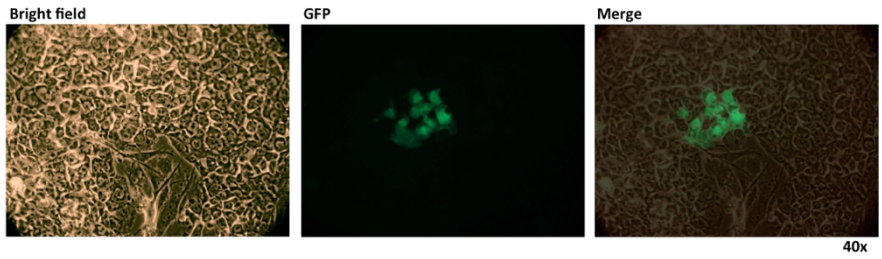
Figure 5.1 – Targeting *SMIM1* using CRISPR/Cas9n. **A** - Schematic representation of CRISPR/Cas9n strategy to target a genomic sequence. The Cas9n is targeted to genomic DNA by an sgRNA consisting of a 20-nt guide sequence (blue) and a scaffold (red). The guide sequence pairs with the complementary DNA target (DNA target - blue bar on opposite strand, upstream of a requisite PAM sequence in red). Cas9n mediates a cleavage of the DNA strand 3 bp upstream of the PAM (red triangle). Panel adapted from [269]; **B** - Representation of the binding pattern of different pairs of sgRNA used to target *SMIM1*; **C** - Map of the plasmid containing the Cas9n and scaffold, which was used to clone the guide sequences.

A previously characterised hiPSC line, S4, was used to perform the genomic editing of *SMIM1* by CRISPR/Cas9n [201]. hiPSCs were transfected by nucleofection with increasing concentrations (2 and 10 µg) of sgRNA/Cas9n pairs. Nucleofection efficiency was determined by fluorescent microscopy 24h after, and it was estimated to be approximately 10% (Figure 5.2 A and B). Several colonies for the three different pairs were obtained after puromycin selection. In all the three pairs, it was observed a higher number of colonies in the nucleofection with 10 µg of DNA (data not shown). Eight, four and six individual colonies of pair 1, 2 and 3, respectively, were manually picked and expanded. The gDNA of these colonies was extracted for screening and amplified as described in section 2.3.2.2. Sanger sequencing of the *SMIM1* locus has shown that CRISPR/Cas9n system was very effective in *SMIM1* editing. An indel in the target region was observed for all the analysed clones. In addition, the clone obtained in the negative control (PX462) had no mutation in the *SMIM1* locus, as expected. Moreover, it was observed that the heterozygous editing (n=17) was more frequent than homozygous editing (n=1).

A



B



C

							Positive control		Negative control
	P1	P1	P2	P2	P3	P3	PX462	pGFP	No DNA
DNA (µg)	10	2	10	2	10	2	5	2	

	P1	P2	P3	PX462	pGFP	No DNA
Total number of colonies	19	14	28	1	0	0
Number of gDNA Extraction colonies	8	4	6	1		
Genotyping by Sanger Sequence	8 heterozygous clones	1 homozygous clone 3 heterozygous clones	6 heterozygous clones	No mutation		

Figure 5.2 – CRISPR/Cas9n approach for *SMIM1* editing in hiPSCs. A - Timeline showing relevant time points for the targeting and screening of *SMIM1* editing; **B** - Determination of nucleofection efficiency through fluorescent microscopy of hiPSCs transfected with positive control (pGFP); **C** - Overview of the experiments and results obtained for each sgRNA/Cas9n pair.

Sanger sequencing of the homozygous clone obtained using pair 2, showed a homozygous 39-bp deletion (Figure 5.3). This deletion, however, has a conserved reading frame and even though the sequence of the protein is unaltered after the deletion, most likely it will originate a non-functional protein (Figure 5.3). To determine the different haplotypes in the heterozygous clones, I performed Sanger sequencing of several bacterial colonies obtained by cloning the PCR product of each clone. The analysis of two heterozygous clones, one from pair 1 and another from pair 3, demonstrated that the clone from pair 1 had two haplotypes, while the clone from pair 3 had three haplotypes, indicating that it was a mixed clone. For pair 1, the observed haplotypes were: i) a 46-bp deletion (this haplotype was observed in six colonies of the eleven analysed; 6/11) and ii) a 25-bp insertion plus 2-bp deletion (5/11). These haplotypes give rise to an early stop codon in the beginning of the protein and, most probably, the short protein produced is not functional and/or is degraded (Figure 5.3).

As *SMIM1* is not expressed in hiPSC, it was not possible to confirm SMIM1 ablation by western blot. The *SMIM1*-knockout hiPSC line created with pair 1, named S4-Cas9n-SMIM1, was used to investigate the impact of SMIM1 ablation during megakaryopoiesis.

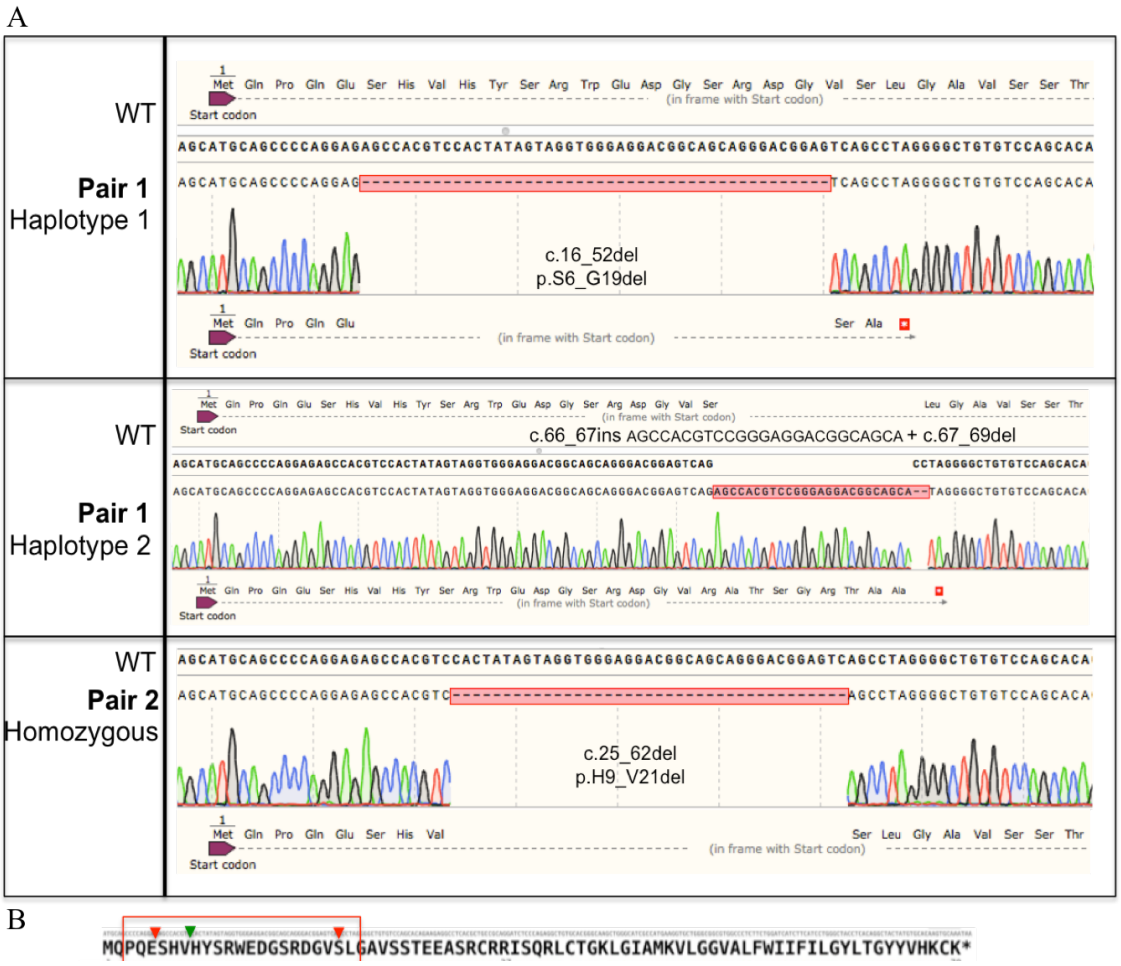


Figure 5.3 – *SMIM1* mutations obtained by CRISPR/cas9n technology. A - Alignment of *SMIM1* WT sequence with the *SMIM1* sequence obtained in the heterozygous clone of pair 1 (haplotypes 1 and 2) and in the homozygous clone of pair 2. The sequence of the resultant protein is also represented. B - Schematic representation of the SMIM1 protein structure with the initial targeting of each haplotype analysed. Arrows in red correspond to pair 1 (haplotype 1 and 2) and in green to pair 2.

5.3.1.2 *SMIM1* overexpression in hiPSCs

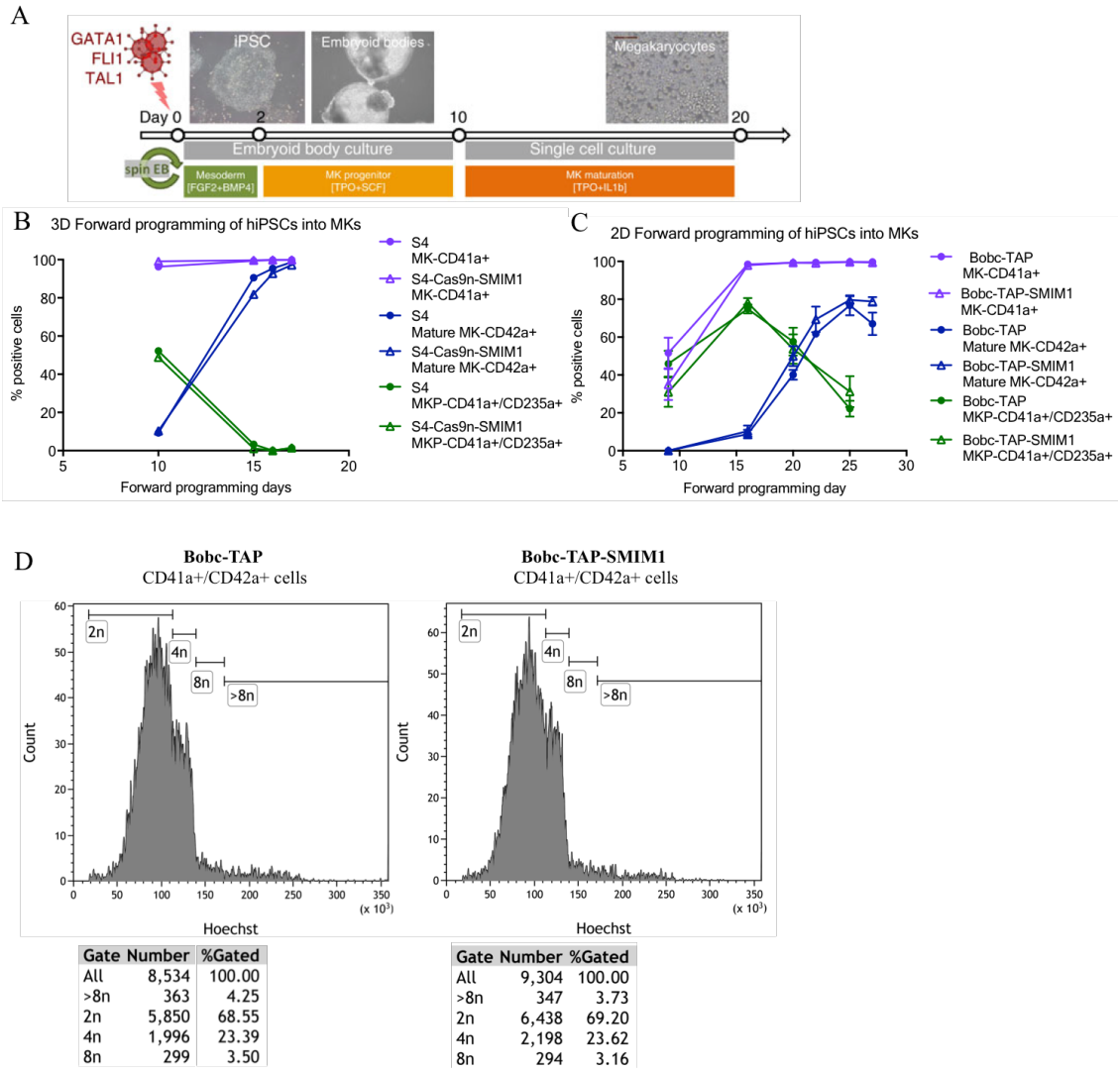
To further characterise the role of SMIM1 during megakaryopoiesis and determine SMIM1 interactome in MKs, an hiPSC line overexpressing tagged SMIM1 was created using a lentiviral plasmid. The lentiviral plasmid, carrying an in-frame fusion of tandem affinity purification (TAP)-tag at the N-terminal coding region of *SMIM1* gene, was created as described in section 2.3.4.1. The TAP-tag is composed by an N-terminal

calmodulin binding domain and a C-terminal 3x Flag repeated separated by a TEV cleavage site. This tag was chosen because its small size is expected to reduce possible allosteric constraints with SMIM1 interactors, and also because it allows a two-step purification, which reduces unspecific binding of interactors. The lentivirus of TAP and TAP-SMIM1 lentiviral plasmids were produced in HEK293T cells and used to transduce Bobc hiPSC as described in section 2.3.4.2. Transduced hiPSCs were then sorted by FACS based on GFP expression at NIHR Cambridge BRC Cell Phenotyping Hub, University of Cambridge and then expanded. At the time of writing, I have not been able to optimise the pull-down protocol to obtain the purity of protein required for proteomic studies. Therefore, the interactome of SMIM1 is still unknown. Bobc hiPSCs were chosen because of an improvement of the forward programming method to a 2D system by Dr Thomas Moreau. The 2D system improved the efficiency of MK production, which would be relevant for the pull-down experiment. The increased MK production, however, was cell-line dependent and not as efficient in S4 (personal communication of Dr Thomas Moreau).

5.3.1.3 Forward programming of SMIM1 knockout and SMIM1 overexpression hiPSC line towards MK

To investigate the effect of SMIM1 ablation during megakaryopoiesis, the S4-Cas9n-SMIM1 hiPSC line was differentiated towards MK using a 3D-forward programming method that was developed by Moreau *et al.* (see section 2.3.3 [207]). Briefly, the S4-control and S4-Cas9n-SMIM1 hiPSC lines were transduced with the transcription factors GATA1, TAL1 and FLI1, that together with medium containing BMP and FGF induced hiPSC into a mesoderm commitment. Then using MK differentiation medium (containing TPO and SCF) cells were differentiated into MK progenitors and finally towards mature MKs, in an approximately twenty-days (Figure 5.4 – A). The ability of S4-Cas9n-SMIM1 to form MKs was assessed by following the surface expression of CD235a, CD41a and CD42a overtime by flow cytometry (as described in section 2.3.3). It was observed that the S4-Cas9n-SMIM1 line, like the S4-control line, was able to form EBs and differentiate towards MKs. No differences in surface expression of CD235a, CD41a and CD42a were detected during the differentiation process of S4-Cas9n-SMIM1 line, when compared to S4-control line (Figure 5.4 – B). The role of SMIM1 in megakaryopoiesis was further examined by forward programming of SMIM1-overexpressing hiPSC line into MKs. Six rounds of Bobc-TAP-SMIM1 and Bobc-TAP forward programming were followed by flow cytometry as described above

and the cultures were collected for pull-down experiments when reached at least 60% of mature MKs (CD41a/CD42a positive cells). This experiment showed that SMIM1 overexpression also does not affect the expression level of CD235a, CD41a and CD42a (Figure 5.4 – C).



Gate	Number	%Gated
All	8,534	100.00
>8n	363	4.25
2n	5,850	68.55
4n	1,996	23.39
8n	299	3.50

Gate	Number	%Gated
All	9,304	100.00
>8n	347	3.73
2n	6,438	69.20
4n	2,198	23.62
8n	294	3.16

Figure 5.4 -SMIM1 does not affect megakaryopoiesis *in vitro*. A - Schematic representation of hiPSCs 3D forward programming into MKs [207]. **B** - Flow cytometry time course of 3D forward programming of S4 and S4-Cas9n-SMIM1 into MKs, showing percentage of MK progenitor (%CD41a+/CD235a+ cells - green), MK lineage committed (%CD41a+ cells - purple) and mature MK (%CD42a+ cells - blue) from whole culture (n=1, from day 9 to day 17 of forward programming). **C**- Flow cytometry time course of 2D forward programming of Bobc-TAP and Bobc-TAP-SMIM1 into MKs, showing percentage of MK progenitor (%CD41a+/CD235a+ cells - green), MK lineage committed (%CD41a+ cells - purple) and mature MK (%CD42a+ cells - blue) from whole culture (n=6, from day 10 to day 25 of forward programming). Analysis performed using Student T-test. **D** – Ploidy analysis of Bobc-TAP and Bobc-TAP-SMIM1 mature MKs (CD41/CD42+ cells) by flow cytometry. Histograms of DNA content were obtained by gating cells positive for Hoechst and double positive for CD41 and CD42. Representative results of 2 independent experiments.

DNA ploidy was also assessed as a marker for MK differentiation by flow cytometry. The DNA content of TAP-SMIM1 mature MKs (CD41/CD42 positive MKs) obtained in two independent experiments of forward programming, did not differ from the DNA content of TAP mature MKs (control, Figure 5.4 - D). It should be noted that both TAP-SMIM1 and TAP mature MKs, presented low levels of ploidy, which is in agreement with previous observations using the same method [207].

During maturation, MKs in the bone marrow migrate from the osteoblastic niche towards the sinusoids, where through a tight process of proplatelet formation they then release platelets to the bloodstream [75]. To determine whether proplatelet formation was affected by overexpression of SMIM1, cultures of forwarded programmed mature MKs were placed onto fibrinogen-coated coverslips for 48h. Very few MKs with proplatelet formation could be detected in both conditions, TAP-SMIM1 and TAP. This impaired any quantification of proplatelets and therefore halted any conclusion on the role of SMIM1 in proplatelet formation. This result was observed in two independent forward programming experiments, suggesting that the reduced level of proplatelets formation is most likely due to the maturation level of the MKs.

5.3.1.4 Bone marrow MK counts and platelet ultrastructure are unaffected in *Smim1* mutant mice

Megakaryopoiesis was also characterised in the *Smim1* mutant mouse model, through the examination of MK number in the bone marrow of six-month-old *Smim1* mutant mice. Sections of mouse hind legs bones were stained with H&E and MKs, were manually quantified (Figure 5.5 – A). As shown in Figure 5.5 – B, no differences were observed in the MK number present in the bone marrow in *Smim1* mutant mice when compared with the wildtype control.

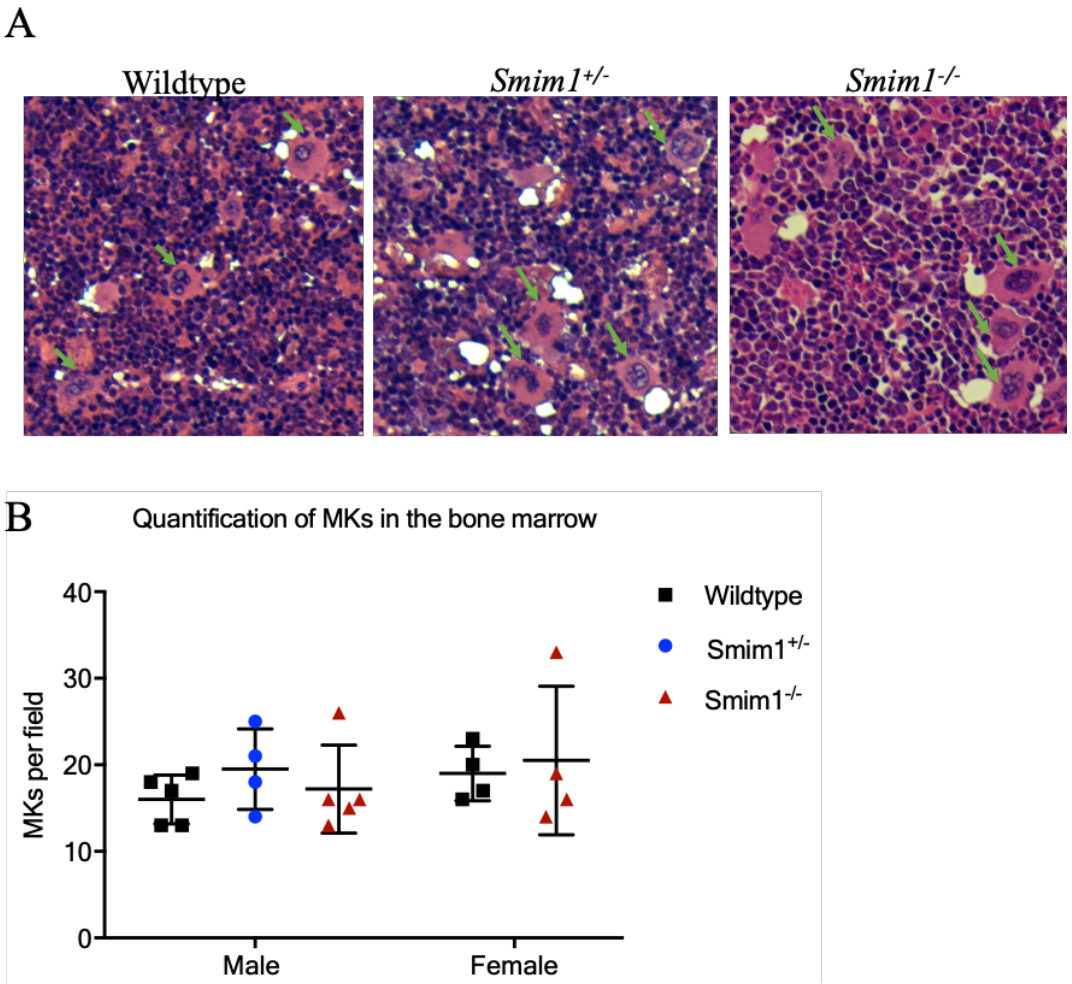


Figure 5.5 - Megakaryocyte number is not altered in *Smim1* mutant mice. A- Representative images of bone marrow sections of six-month old wildtype, *Smim1*$^{+/-}$ and *Smim1*$^{-/-}$ male mice that were stained with H&E. These images represent only a section of a field of view. Green arrows indicate MKs (x10). **B-** Quantification of MKs in the bone marrow was performed blinded by manual count of MKs in ten fields per mouse. Data show mean ± standard deviation. Black square – wildtype, blue circle – *Smim1*$^{+/-}$, red triangle – *Smim1*$^{-/-}$. Statistical analysis was performed using Anova for males and unpaired T-test for females (no significant differences were observed).

To investigate whether SMIM1 was relevant for platelet ultrastructure, washed platelets of four months old *Smim1*$^{-/-}$ and wildtype mice were fixed and prepared for TEM. Fixed platelets were processed by Dr James McMillan at Cambridge Advanced Imaging Centre and, sections and imaging were performed by me with Dr James McMillan's support. Analysis of platelet ultrastructure by TEM did not reveal abnormalities in *Smim1*$^{-/-}$ platelets structure as shown in Figure 5.6.

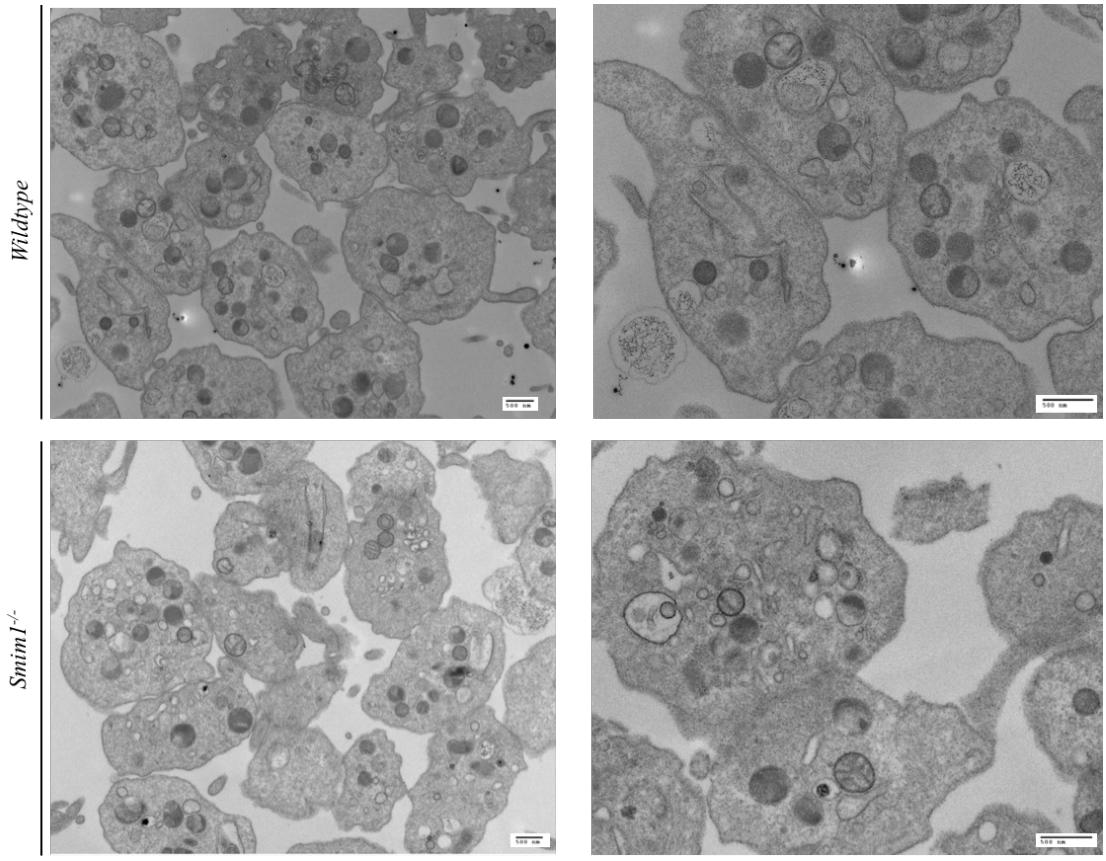


Figure 5.6 - SMIM1 does not affect ultrastructure of mice platelets. Platelets were isolated and fixed for TEM as described in section 2.2.7 and 2.2.10, respectively. TEM representative pictures of wildtype and *Smim1*$^{-/-}$ mice platelets. Scale bar - 500 nm.

5.3.1.5 RNA-sequencing of *Smim1* mutant mice platelets

Although platelets are anucleate cells, they have a rich repertoire of RNAs, including messenger RNAs, structural and catalytic RNAs, and regulatory RNAs, that are obtained from MKs and are crucial for platelet function. Thus, alterations in the platelet transcriptome may give key clues to understand the role of SMIM1 not only in platelet function but also in MKs and platelet formation [270]. To explore this hypothesis, I have assessed the platelet transcriptome of *Smim1*$^{-/-}$ mice by RNA-sequencing. Platelet pellets of three to four six-month-old male mice were pooled together, to obtain enough high-quality RNA for the RNA-sequencing analysis. The sequencing and analysis were performed in four and three independent pools of wildtype and *Smim1*$^{-/-}$ mouse platelets, respectively. This strategy not only improved the quality of the RNA but also reduced the inter-mouse variability in gene expression, therefore making it easier to identify the effects caused by SMIM1 ablation. After RNA extraction, four wildtype and three *Smim1*$^{-/-}$ RNA-sequencing libraries were prepared by Ms Frances Burden, Research Associate in Prof. Ouwehand group. The differential expression analysis at a 5% false discovery rate performed by Dr Denis Seyres, Research Associate in Prof. Ouwehand group, showed that the transcriptome of wildtype and *Smim1*$^{-/-}$ platelets was similar (Figure 5.7 - A), with only 12 genes being differentially expressed (four down-regulated and eight up-regulated) (Figure 5.7- B, Table 5.1). Manual annotation of the differentially expressed genes showed no particular enrichment for a biological process or pathway, and most of the genes were uncharacterised in platelets (http://pantherdb.org/geneListAnalysis.do). Five genes encode for binding proteins (*Arrdc3, Tnfaip2, Gng4, Lrrc41* and *Afdn*), two genes encode for transporter proteins (*Slc38a6* and *Nacad*) and three genes encoding for motor and cytoskeleton proteins (*Dcdc2a, Sfi1* and *Myo1g)* (http://pantherdb.org/geneListAnalysis.do). The *Unconventional myosin-Ig (Myo1g)* gene, which was up-regulated, encodes a predicted actin/ATP binding protein that has been described as part of phosphatidylinositol 3,4,5-trisphosphate (PtdIns(3,4,5)P3) interactome/signalosome in platelets, which is important for platelet activation [271].

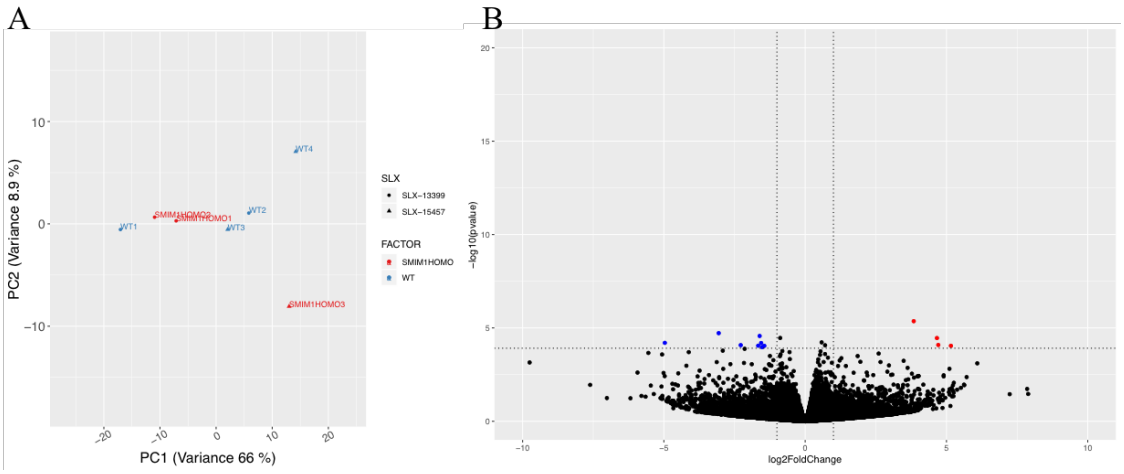


Figure 5.7 – Differential gene expression of *Smim1*$^{-/-}$ platelets. A - Principal component analysis showing the variance between wildtype (WT-blue) and *Smim1*$^{-/-}$ (SMIM1Homo - red) platelets. **B** - Volcano plot showing the genes differentially expressed between wildtype and *Smim1*$^{-/-}$ platelets at a 5% false discovery rate. Up-regulated genes are highlighted in blue and down-regulated genes are highlighted in red.

Table 5.1 – Genes differentially expressed in *Smim1*$^{-/-}$ platelets. Differential expression analysis at a 5% false discovery rate, up-regulated genes in blue and down-regulated genes in red.

Gene symbol	Log2-fold change	*p-value*
Nacad	3,84	2,3E-05
Gng4	2,60	1,2E-05
Tcp11l1	0,54	4,1E-05
Arrdc3	-0,55	3,7E-06
Myo1g	-0,81	6,9E-05
Slc38a6	-0,88	1,2E-05
Sfi1	-1,52	2,0E-05
Tnfaip2	-1,66	1,5E-05
Afdn	-2,14	6,7E-05
Lrrc41	-2,28	1,4E-05
Harbi1	-3,06	2,9E-05
Dcdc2a	-5,06	1,2E-04

5.3.2 Investigation of SMIM1 role in platelet function

Although the observation described above suggests that SMIM1 does not affect megakaryopoiesis nor platelet ultrastructure, it cannot be excluded a possible role of SMIM1 in platelet function. Therefore, I then characterised the SMIM1 role in platelet function using different assays that assessed distinct features of activated platelets.

As the Vel-negative donors that took part of this study were recruited across the UK, and platelet function needs to be performed within 30 minutes after blood collection [272], it was not possible to perform platelet function studies in these donors. Therefore, to investigate the role of SMIM1 in platelet function, I have used Vel-positive donors and taken advantage of *Smim1* mutant mice model. Mice constitute a good model to study platelet function, not only due to its platelet function similarity with humans [273-275], but also because it provides access to a large number of samples with an identical genetic background, which reduces the interindividual variability in platelet function observed in human subjects [276-279], therefore making it easier to discriminate the effects, if any, caused by the absence of SMIM1.

5.3.2.1 SMIM1 is phosphorylated upon platelet activation in humans

In Chapter 3, it was shown that SMIM1 has a cytoplasmic localisation in human platelets and co-localises in the fractions enriched for α-granule´s proteins, suggesting that SMIM1 has an intracellular distribution similar to these proteins. To address whether SMIM1 localisation would change upon platelet activation, as is observed for P-selectin (an α-granule protein), whole blood of Vel-positive donors was stimulated with TRAP. Platelet response to the agonist and SMIM1 localisation was determined by P-selectin and SMIM1 surface expression using flow cytometry. Flow cytometry analysis showed that SMIM1 is not detected on the platelet membrane upon platelet activation (Figure 5.8 - A). To further characterise SMIM1 during platelet activation, washed platelets were stimulated with TRAP, and SMIM1 was detected by western blot. Western blot of stimulated platelet lysates showed an increase of SMIM1 molecular weight upon activation (Figure 5.8 – B-C). As it has been shown that SMIM1 undergoes to post-translation modifications, such as phosphorylation [182, 188, 208, 212], to determine if the changes of electrophoretic mobility observed upon platelet activation were due to phosphorylation, platelet lysates were then treated with either phosphatase Lambda or with a phosphatase inhibitor. Stimulated platelet lysate treated

with phosphatase Lambda showed a reduction of SMIM1 molecular weight when compared with the stimulated platelet lysate treated with the phosphatase inhibitor (Figure 5.8 – B-C). The molecular weight of stimulated platelet lysate treated with phosphatase Lambda was similar to the molecular weight observed in resting platelets lysates treated with phosphatase Lambda or with a phosphatase inhibitor (Figure 5.8 - C). Thus, indicating that SMIM1 is phosphorylated upon platelet activation.

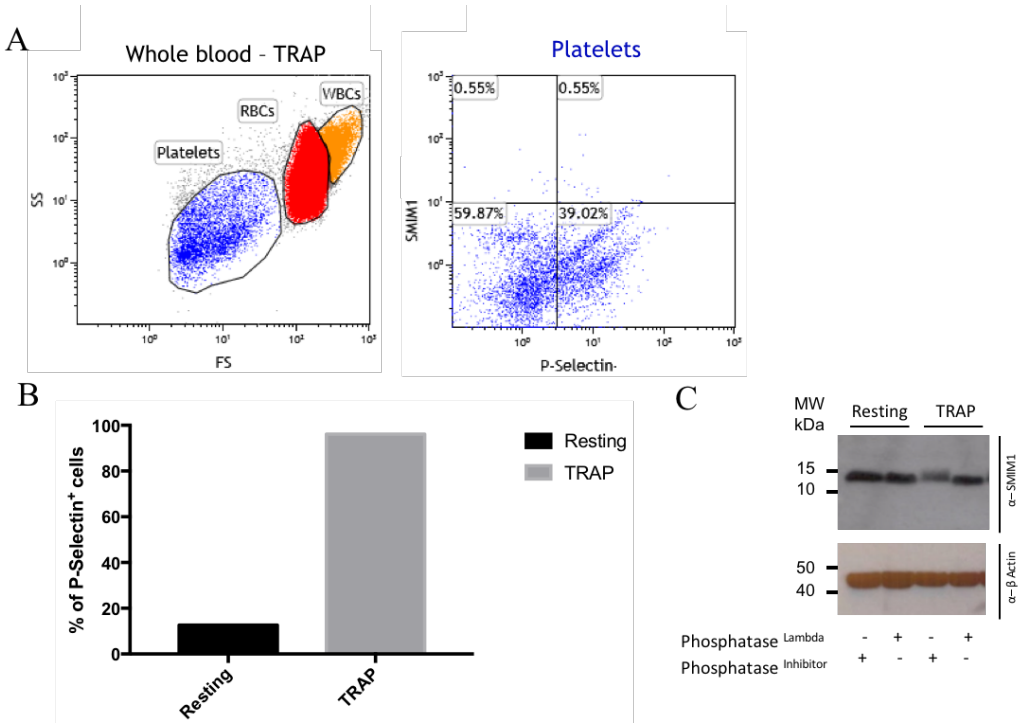


Figure 5.8 - SMIM1 is not expressed on the platelet membrane but is phosphorylated upon platelet activation. A - SMIM1 expression during platelet activation was assessed by flow cytometry of whole blood stimulated with TRAP agonist (3µM) and stained with anti-CD62-PE-Cy5 (which detects P-selectin), anti-SMIM1 (SpG213Dc human monoclonal) and FITC-secondary antibody against SMIM1. Dot plot of gated platelets shows no expression of SMIM1 and 39% of P-selectin expression on platelets after stimulation with TRAP agonist. **B** - Washed platelets were prepared as described in section 2.1.3.1 and treated with either platelet inhibitors (Resting) or agonist (TRAP – 5µM). Platelet response to agonist was assessed by P-selectin (CD62P-APC) expression on platelet membrane (% of P-selectin^{+}cells) using flow cytometry. **C** - Western blot of platelets lysates obtained in (B) were then treated either with phosphatase Lambda or phosphatase inhibitor. Detection of SMIM1 using the polyclonal anti-SMIM1 antibody (Sigma) shows a reduction of SMIM1 molecular weight in activated platelets treated with phosphatase lambda, indicating that SMIM1 is phosphorylated upon activation. β- actin was used as loading control and the blot is representative of 3 donors.

5.3.2.2 Basal levels of P-Selectin surface exposure are decreased in male *Smim1* mutant mice but platelet activation is unaffected

During platelet activation, platelets secrete their granule contents, this consequently leads to the activation of more platelets and contribute to platelet aggregation and clot formation [144]. During this process, various phenomena can be observed. Among those, P-selectin, an α-granule protein, becomes exposed on the platelet membrane and the GPIIbIIIa complex (CD41/CD61 complex, fibrinogen receptor) is activated inducing fibrinogen binding and thus promoting platelet aggregation. To investigate whether ablation of *Smim1* affected platelet activation, whole blood from four-month-old *Smim1* mutant and wildtype mice was stimulated with a range of concentrations of cross-linked collagen-related peptide (CRP-XL) and thrombin. Platelet response to the different agonists was determined by flow cytometry, measuring the expression of P-selectin (detected by anti-mouse P-selectin- PE antibody) and fibrinogen binding (detected by anti-human fibrinogen-FITC) (analysis strategy is described in section 2.2.11).

In the resting state, platelets from *Smim1* mutant male mice showed a significant reduction in the basal level of P-selectin exposure (*Smim1*$^{+/-}$ - p=0.0134, *Smim1*$^{-/-}$ - p=0.0176) and a trend, albeit not significant, of reduction in the basal level of fibrinogen binding when compared to wildtype platelets (Figure 5.9 A). In contrast, no significant differences were observed in the basal level of P-selectin exposure and fibrinogen binding between platelets of *Smim1*$^{-/-}$ and wildtype female mice (Figure 5.9 – B). Upon platelet activation, platelets from *Smim1*$^{-/-}$ male mice treated with low concentrations of thrombin, showed a significant difference in fibrinogen binding between wildtype platelets (*Smim1*$^{-/-}$: Thr (0.01 U/ml) - p=0.0204, Thr (0.1 U/ml) - p=0.0432) (Figure 5.9 – D). The expression levels of P-selectin and fibrinogen binding in these conditions, however, are similar to the expression levels observed in the resting state (Figure 5.9 A and D). This suggests that the observed difference is due to a reduced basal level of fibrinogen binding and not to platelet response to thrombin. The difference observed between the two conditions, resting and thrombin low concentrations can be explained by the smaller number of mice analysed in the latter condition, which occurred due to a technical experimental problem when preparing the samples. In platelets of female mice, no significant differences in platelet response were observed between *Smim1*$^{-/-}$ and wildtype platelets (Figure 5.9 E-F). A trend, albeit not significant, of P-selectin and

fibrinogen binding reduction was observed in *Smim1*$^{-/-}$ platelets of female mice stimulated with CRP (Figure 5.9 -E). In summary, SMIM1 ablation does not affect platelet response to agonists *in vitro*.

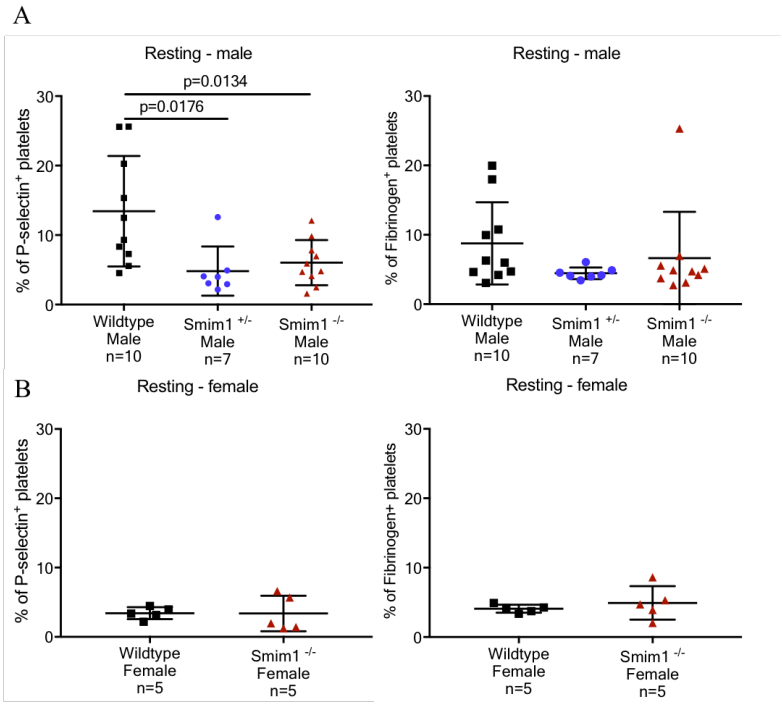


Figure 5.9 - Basal level of P-Selectin surface expression is decreased in platelets of *Smim1* mutant males, but platelet activation is unaffected. Platelet response to no agonist, CRP and thrombin was assessed by flow cytometry of whole blood incubated with antibodies against P-selectin, fibrinogen and CD41. **A-B** - percentage of CD41^{+} platelets expressing P-selectin (left panel) and fibrinogen (right panel) in male (**A**) and female (**B**) mice whole blood not treated (without agonist). Data represents mean± standard deviation. Statistical analysis was performed using Student T-test and ANOVA-Tukey for females and males, respectively.

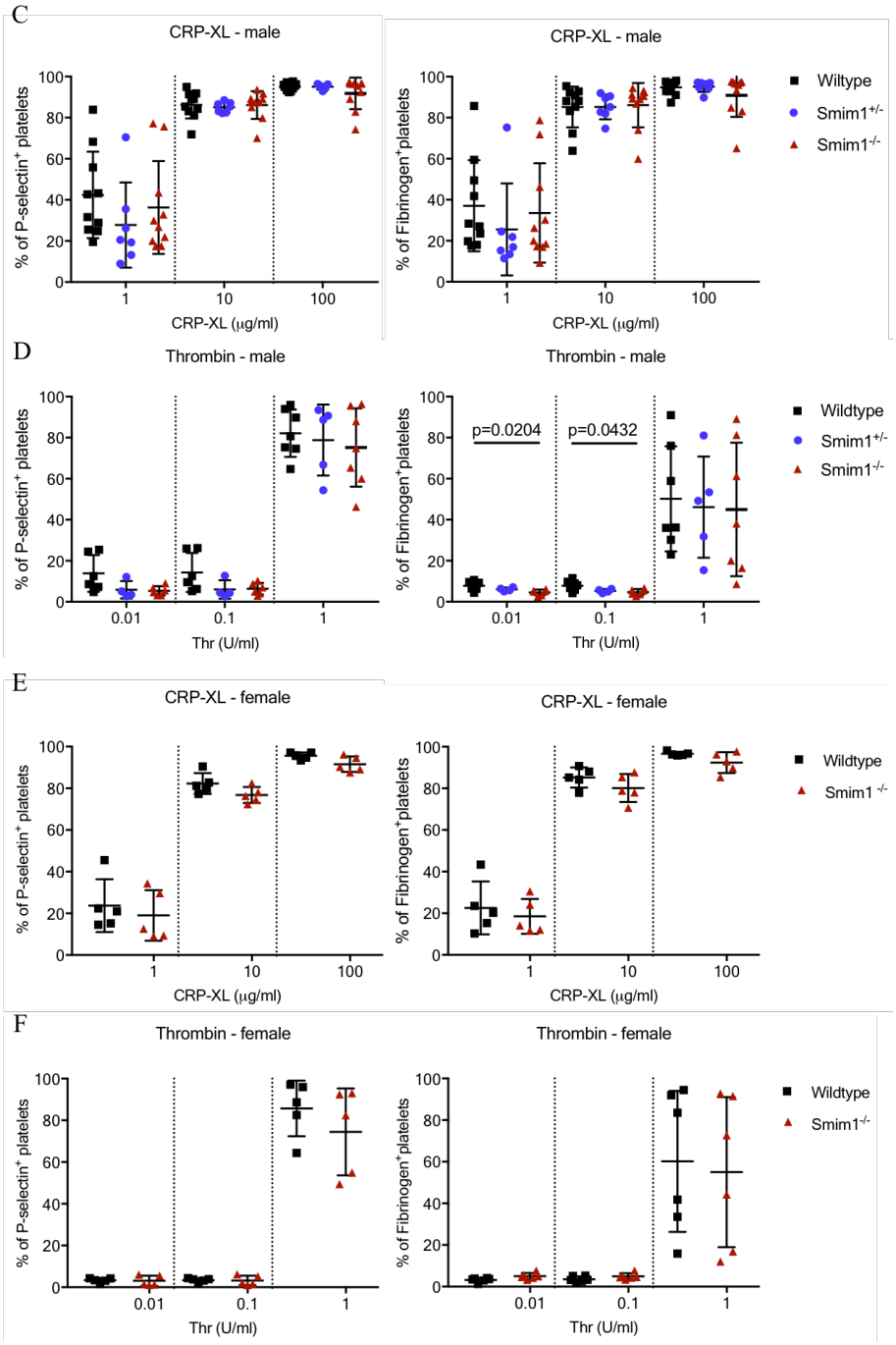


Figure 5.9 - Basal level of P-Selectin surface expression is decreased in platelets of *Smim1* mutant males, but platelet activation is unaffected. C-E - percentage of $CD41^+$ platelets expressing P-selectin (left panel) and fibrinogen (right panel) in male (C) and female (E) mice whole blood treated with CRP-XL. **D-F** - percentage of $CD41^+$ platelets expressing P-selectin (left panel) and fibrinogen (right panel) in male (**D**) and female (**F**) mice whole blood treated with thrombin. Data represents mean± standard deviation. Statistical analysis was performed using Student T-test and ANOVA-Tukey for females and males, respectively.

5.3.2.3 Platelet spreading is unaffected in *Smim1* mutant mice

Upon platelet attachment at the injury site, platelets undergo a drastic change in their shape, from a small discoid form to a spread and larger form, which results from the reorganisation of their cytoskeleton. To investigate whether SMIM1 ablation affected the ability of platelets to spread, washed platelets from four-month-old *Smim1*$^{-/-}$ and wildtype male mice were incubated on fibrinogen-coated coverslips for 30 minutes at 37 °C. Fixed platelets were stained with phalloidin-A555 and both attached and spread platelets per field were quantified. No differences in platelet spreading were detected between *Smim1*$^{-/-}$ and wildtype mice (Figure 5.10).

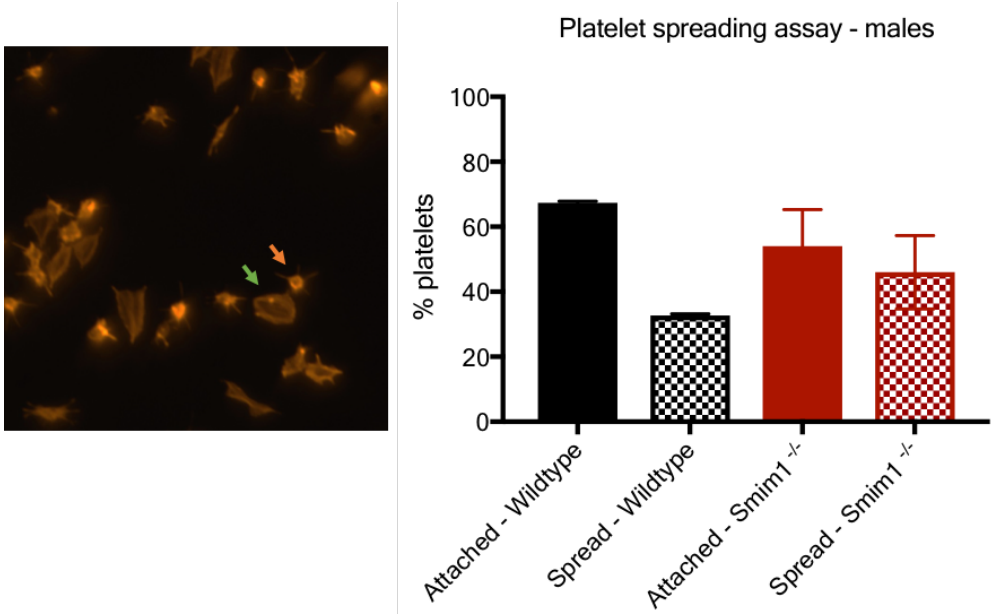


Figure 5.10 - SMIM1 does not affect mice platelet attachment and spreading. Washed platelets of wildtype and *Smim1*$^{-/-}$ male mice were prepared as described in section 2.2.9 and incubated onto fibrinogen-coated coverslips for 30 minutes at 37°C. Fixed platelets were stained with phalloidin-A555 and imaged using a wide-field fluorescence microscope. Attached (orange arrow) and spread (green arrow) platelets were manually counted in ten pictures per mouse using ImageJ software. Data represent the mean ± standard deviation of % platelets attached and spread per mouse (n=3). Statistical analysis was performed using the Mann-Whitney test.

5.3.2.4 Assessing expression level of α-granules proteins in platelets of *Smim1*$^{-/-}$ mice

As evidence has pointed to a reduction in the basal level of P-selectin exposure, I looked at expression level of P-selectin and other α-granules proteins, such as VWF, THBS1 and VEGFA, in platelets of *Smim1*$^{-/-}$ and wildtype male mice by western blot. Platelet lysates of four months *Smim1*$^{-/-}$ and wildtype male mice were prepared as described in section 2.2.7 and immunoblotted with specific antibodies for the different α-granule proteins (described in section 2.2.8). No differences in protein level of P-selectin and the other α-granule proteins were detected between *Smim1*$^{-/-}$ and wildtype platelets (Figure 5.11).

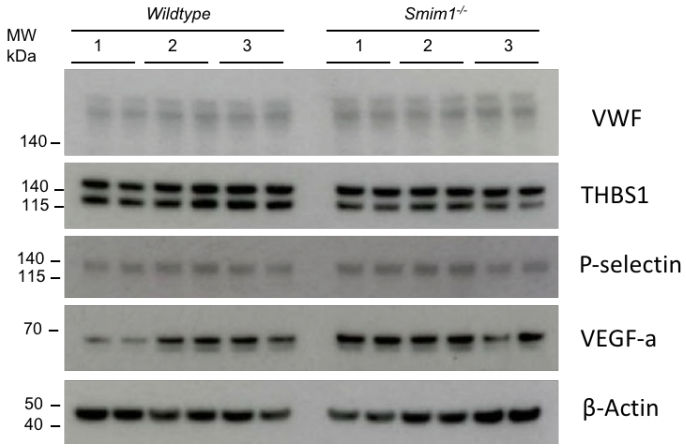


Figure 5.11 – SMIM1 does not affect the expression level of P-selectin, VWF, THBS1 and VEGF-a. Western blot of platelet lysates from wildtype and *Smim1*$^{-/-}$ male mice show no differences in expression level of VWF, THBS1, P-selectin and VEGF-a. Platelet lysates of three animals per genotype were loaded in duplicate (20 µg/ lane) and β-actin was used as protein loading control.

5.4 Discussion

SMIM1 is expressed in MKs and platelets and affects platelet count in humans (Chapter 3 and 4). However, the molecular mechanism through which SMIM1 affects platelet formation, be it either an alteration in megakaryopoiesis, thrombopoiesis and/or platelet function/clearance, remains unknown. The identification of SMIM1 role in these processes and the characterisation of its molecular mechanism will improve the general knowledge of these processes. In turn, these findings may be important for the development of novel therapies for MK/platelets disorders and to prevent/treat disorders that Vel-negative individuals may acquire that are not yet characterised.

In the present chapter, I have investigated the role of SMIM1 in megakaryopoiesis and in platelet activation *in vitro.* Firstly, I have characterised the role of SMIM1 in megakaryopoiesis using two models, hiPSC and mouse. The characterisation of megakaryopoiesis in *SMIM1*-knockout and *SMIM1*-overexpressing hiPSCs differentiated into MKs by forward programming, showed that SMIM1 does not alter the differentiation, maturation and ploidy in this model. This suggests that the alteration of platelet number in Vel-negative individuals, observed in the previous chapter, is not due to an impairment of megakaryopoiesis. As the forward programming model has several limitations regarding the maturity of the formed MKs and platelet production [207], it would be important to confirm these findings and further characterise the role of SMIM1 in proplatelet formation using a different model, such as Vel-negative $CD34^+$ cells differentiated towards MKs.

In the mouse model, bone marrow MK counts did not differ between *Smim1*$^{-/-}$ and wildtype mice, and the ultrastructure of *Smim1*$^{-/-}$ platelets was unaffected. Not surprisingly, given the absence of striking phenotypes in mouse MK and platelets, the characterisation of the platelet transcriptome in mice showed small differences in the transcription profile of *Smim1*$^{-/-}$ platelets. This resulted in a short list of genes differentially expressed. Among the identified genes, two of them seem to be involved in platelet function: *Myo1g* and *Guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-4 (Gng4)*. *Myo1g*, which was up-regulated, encodes an unconventional non-muscle myosin. These types of myosin are important motor proteins that are thought to regulate exocytosis in different cell types [280]. It has been shown that MYO1C is part of PtdIns (3,4,5)P3 signalosome, which plays an important role in

platelet activation and thrombus formation [271]. The *Gng4* gene, which was down-regulated, encodes for a protein with GTPase activity predicted to be a transducer in various transmembrane signalling pathways important in platelet activation, such as ADP signalling through P2Y purinoceptor 1 and 12, prostacyclin signalling through prostacyclin receptor and also PI3K-Akt signalling (http://pantherdb.org/genes/gene.do?acc=MOUSE%7CMGI%3DMGI%3D102703%7CUniProtKB%3DP50153). Even though most of the genes identified were uncharacterised in platelets, some of those genes are involved or predicted to be involved in cytoskeleton structure (*Dcdc2a, Sfi1* and *Myo1g)* and in signalling transduction. As platelet formation and function involves rearrangements of the cytoskeleton, this could suggest that SMIM1 has a role in proplatelet formation and/or platelet function. However, as platelet count and MPV in six-months old *Smim1*$^{-/-}$ mice was shown to be unaffected in the previous chapter, together these findings suggest that ablation of SMIM1 in mouse does not affect megakaryopoiesis and by inference thrombopoiesis.

In this chapter, I have also investigated the role of SMIM1 upon platelet activation. To this end, I used two models, human and mice, and I have assessed different features of activated platelets through a range of assays. The assessment of *Smim1*$^{-/-}$ platelets' ability to spread was performed in static conditions and showed that *Smim1*$^{-/-}$ platelets spreading ability is unaffected. RNA-sequencing showed that some genes differentially expressed are implicated with cytoskeleton structure (*Dcdc2a, Sfi1* and *Myo1g*), and given that the cytoskeleton constitutes a highly dynamic structure during proplatelet/platelet formation and in both resting and activated platelets, [137], it would be important to further characterise whether SMIM1 ablation affects these cellular dynamics *in vivo*. To assess proplatelet and platelet formation, two-photon intravital microscopy of the bone marrow of *Smim1*$^{-/-}$ mice could be performed as described by Dütting *et al.*[93]. While to specifically assess the cytoskeleton dynamics of resting and activated platelets, time-lapse microscopy of *Smim1*$^{-/-}$ platelets expressing end-binding protein 3(EB3)- GFP (EB3, is a protein that regulates microtubules dynamics) it could be performed as described by Patel-Hett *et al.* [137].

The assessment of platelet response to different agonists by flow cytometry showed that the ablation of SMIM1 does not affect platelet response. Platelets of female *Smim1*$^{-/-}$ mouse stimulated with CRP, however, showed a trend of reduced platelet response

(Figure 5.9 - E). Interestingly, it was also observed that SMIM1 ablation in resting platelets of male mice leads to a significant reduction of P-selectin exposure and a reduction of fibrinogen binding (not significant). These observations indicate that SMIM1 has a subtle effect on platelet activation. In support of this, I have shown that SMIM1 is phosphorylated upon human platelet activation *in vitro,* which suggests that SMIM1 may require phosphorylation in order to be functional. The molecular mechanism behind these findings and its possible implications are still unclear. One could speculate that these trends of reduction in platelet response may result from subtle alterations in the cytoskeleton rearrangement, which may affect granule secretion.

Platelet homeostasis and thrombus formation are tightly controlled by a complex network of inhibitory and activation signalling pathways, which are regulated by several protein kinases [136]. As SMIM1 is phosphorylated in serine residues [182, 188, 208, 212], it would be important to investigate which serine/threonine protein kinase is responsible for SMIM1 phosphorylation, in order to better understand SMIM1 role during platelet activation. Several serine/threonine protein kinases have important roles during platelet activation, such as protein kinase C (PKC), which has been shown to mediate platelet secretion and thrombus formation [281] and 3-phosphoinositide-dependent protein kinase 1 (PDK1), which has been shown to activate serine/threonine kinase protein kinase B (Akt) and inhibit glycogen synthase kinase 3 beta (GSK3β) thus enhancing platelet aggregation, clot retraction, platelet spreading and thrombin formation [282]. Identification of the serine/threonine protein kinase responsible for SMIM1 phosphorylation, could be performed by assessing SMIM1 phosphorylation state upon platelet activation either in wildtype/Vel-positive platelets treated with different protein kinase inhibitors, or in platelets of genetic models in which the expression of a single kinase is affected. These experiments would give important clues to further explore the molecular mechanism of SMIM1 role in platelets.

After platelet activation and aggregation, a stable thrombus is formed. In this chapter, I have shown that SMIM1 ablation does not affect platelet spreading and response to different agonists. However, I have not addressed whether SMIM1 ablation affects thrombus formation and degradation. It still remains to be investigated if SMIM1 ablation affects thrombus formation and stability *in vivo*. To that end, a ferric chloride or laser injury thrombosis intravital microscopy model could be used [283-285]. In

these models, an injury is induced in the vessel wall and thrombus development is imaged in real-time using intravital video-microscopy [283]. In the ferric chloride-injury model thrombosis is induced with an oxidative injury, which severely damages the vessel wall, exposing the subendothelial matrix. While in the laser-injury model thrombosis is induced with a heat injury, which does not damage the vessel wall. This injury is similar to an injury caused by inflammation and this technique allows a temporal and spatial resolution of the thrombus [283]. These models allow the characterisation of the different cell types playing a role in thrombus formation, including platelets, neutrophils, blood coagulation proteins, endothelium, and the vessel wall [283]. As SMIM1 is expressed in neutrophils (Chapter 3 and 4), and those like platelets have an important role for thrombus formation [286, 287], both models may give important clues on SMIM1 role in platelets and neutrophils during thrombus formation and stability.

In conclusion, in this chapter, I showed that SMIM1 does not affect megakaryopoiesis in the models used and that it plays a subtle role during platelet activation. However, the molecular mechanisms involved and therefore the biological relevance of these findings, remain unclear.

CHAPTER 6 | CONCLUSIONS AND FINAL REMARKS

The advance of sequencing technologies enabled the mapping of the gene that underpins the Vel-blood group system. The discovery of *SMIM1*, which had remained unknown for more than 60 years, has opened new avenues to investigate this blood group system [182, 186, 188]. SMIM1 protein has been preliminarily characterised on the RBC membrane [182, 186, 188, 208, 212] but its function in RBCs and its potential function in other cells remains largely unknown. For this reason, the molecular characterisation of SMIM1 will lead to a better understanding of the physiological processes in which this gene may be involved. In turn, these studies may have the potential to reveal unknown clinical phenotypes and to inform their prevention and treatment.

In my thesis, I have characterised the role of SMIM1 in haematopoiesis and other organs through two main approaches, *in vivo* and *in vitro*, and using three models (human subjects, mouse and cell lines). This work has led to four main findings:

i) detection of *SMIM1* expression in other mature blood cell types;
ii) association of SMIM1 absence with hematologic alterations in RBC, platelet and neutrophil traits.
iii) observation of molecular evidence of SMIM1 role in platelet activation;
iv) discovery of a metabolic phenotype in human and mouse *SMIM1* knockouts.

These findings are important clues of novel roles of SMIM1 in multiple blood cells and possibly in other non-hematopoietic cells. They also open new questions regarding the role of SMIM1 in these cells and its clinical relevance.

6.1 Molecular characterisation of SMIM1

The molecular characterisation of SMIM1 was performed in blood cells of Vel-positive blood donors. I have specifically shown that *SMIM1* is expressed in RBCs, as well as in neutrophils, naïve, memory and class-switched B-lymphocytes, MKs and platelets, through gene expression analysis of multiple blood lineages (Chapter 3). It was further observed that both the localisation of SMIM1 and the molecular weight of its multimers are cell-dependent through biochemical and flow cytometry assays (Chapter 3). These findings suggest that SMIM1 may play a role in the development and/or function of these cells, potentially in a cell-dependent manner.

6.2 Investigation of the role of SMIM1 in blood traits

To further investigate the role of SMIM1 in blood cells, I performed a quantitative and qualitative comparison of haematological traits between Vel-negative blood donors, Vel-negative/weak individuals of UK Biobank (European individuals of the UK Biobank 400,000 individuals) and *Smim1* mutant mice (Chapter 4). These analyses have shed new light on SMIM1's association with traits in RBC, platelets and neutrophils.

6.2.1 Investigation of the role of SMIM1 in blood traits in humans

The analysis of the Vel-negative blood donors cohort showed that SMIM1 absence is associated with a small but significant decrease in the number of RBC and reticulocytes, with an increase of the percentage of RDW, with an increase in the number and size of platelets, and with an increase in the number of neutrophils. Furthermore, it was shown that the alteration of platelets and neutrophil traits was sex-dependent, suggesting a possible hormonal regulation. Interestingly, some of these alterations in blood traits did not overlap with those observed in the UK Biobank cohort. A decreased number of platelets in both sexes and no alteration of RBC and neutrophil counts were observed in this study. Moreover, Vel-negative/weak individuals in the UK Biobank cohort have also shown a negative association with HGB and MCHC that suggest a possible role of SMIM1 in iron metabolism. Together this suggests that other factors not yet identified may potentially influence the association of SMIM1 with some blood traits.

It is important to note that the differences observed between Vel-negative blood donors and Vel-negative/weak individuals of UK Biobank may be due to: i) distinct associations in Vel-negative and Vel-weak individuals (which are the majority of the individuals in the UK Biobank), ii) the result of a positive selection for healthier individuals in the UK Biobank cohort, and iii) the result of a positive selection for individuals able to withstand several blood donations, and/or with an alteration in iron metabolism in the blood donors cohort. Recent results from the INTERVAL study have shown that active blood donors (that give regular donations) present altered iron metabolism resulting in decreased HGB and ferritin concentrations [229]. Although the association of frequent donations with platelet traits have not been investigated in this study, other studies have shown that altered iron metabolism is associated with increased platelet counts [288-293].

It would be important to further explore these differences through a follow-up study in a larger population of unselected individuals and blood donors. It would be particularly relevant to confirm these findings and to assess whether SMIM1 absence affects metabolic parameters related to iron metabolism. The complete data set of the UK Biobank (500,000 individuals), which was not available during the writing-up of this project, and the INTERVAL study (50,000 blood donors) represent two potential avenues for this next step. However, as the UK Biobank is not representative of the general population, a follow-up study in a larger population representative of the general population should also be prioritised.

6.2.2 Investigation of the role of SMIM1 in blood traits in mice

There were also phenotypical differences between the Vel-negative blood donors, the Vel-negative/weak individuals of the UK Biobank and the *Smim1* mutant mice. The ≥ six-months old *Smim1*$^{-/-}$ male mice had an increased number of RBC and no alteration in the number of platelets. Since the equivalent age of this observation was not represented in the human cohorts, the age-dependency of the SMIM1's effect on the RBC number in humans would have to be investigated. This could be done either by performing this analysis in a younger human cohort or in an older mice model.

6.3 Characterisation of SMIM1 role in megakaryopoiesis and thrombopoiesis

The investigation of the association of SMIM1 with platelet traits in both human cohorts (Vel-negative blood donors and Vel-negative/weak individuals of UK Biobank) have shown that SMIM1 is a regulator of platelet production. For this reason, I then focused on the investigation of the role of SMIM1 in megakaryopoiesis, thrombopoiesis and platelet function. I showed that SMIM1 does not affect MK differentiation and maturation by characterising two models of hiPSC line forward programmed into MKs: a *SMIM1* knockout model and a *SMIM1* overexpression model.

Testing an alternative *SMIM1* knockout hiPSC lines would enable further confirmation of this phenotype. It is also important to note that the differentiation model adopted in this study only enables a restricted level of MKs maturity, which consequently limits the assessment of thrombopoiesis. The analysis of thrombopoiesis would be feasible in a differentiation model such as the Vel-negative CD34^{+} cells differentiated into MKs

through a proplatelet formation assay of human *SMIM1* knockout MKs. Since some of the genes showing differential expression in platelets of *Smim1*$^{-/-}$ mice seem to be involved in the structure of the cytoskeleton (*Dcdc2a, Sfi1* and *Myo1g*, Chapter 5), it would be particularly interesting to analyse by immunofluorescence other cytoskeleton proteins (e.g. tubulin, myosin and actin) to detect if these display any alterations.

As the association of SMIM1 absence with alteration of platelet traits was only observed in humans, the study of thrombopoiesis in a human model would be highly relevant. However, as the recruitment of Vel-negative individuals might be potentially challenging. Another way to overcome the technical limitation of the model adopted in this study would be the assessment of thrombopoiesis in the *Smim1* mutant mouse model. Similarly to the alternative proposed human model, the *Smim1*$^{-/-}$ mouse CD34^{+} cells could be differentiated into MKs and enable the study of thrombopoiesis through proplatelet formation assay. This mouse model could also allow the characterisation of platelet formation *in vivo*, in which the microenvironment is maintained (as suggested in Chapter 5).

6.4 Characterisation of SMIM1 role in platelet function

The effect of SMIM1 ablation in platelet function was also assessed. Two main observations suggest the involvement of SMIM1 in platelet activation: i) a reduced basal level of P-selectin surface expression in platelets of *Smim1*$^{-/-}$ male mice compared with wildtype mice; and ii) SMIM1 phosphorylation in human platelets during platelet activation *in vitro* (Chapter 5). The molecular mechanism behind these findings and its possible implications are still unclear.

Further insight into the role of SMIM1 in platelets will require further clarification of molecular mechanisms underpinning the observed phenotype. As suggested in Chapter 5, it would be particularly interesting to i) assess if the detected SMIM1 phosphorylation is indeed involved in platelet activation, ii) to determine the protein kinase responsible for this phosphorylation pattern and, more broadly, iii) to characterise SMIM1's interactome, as described below.

The characterisation of SMIM1's interactome may provide important clues to understand the molecular mechanisms in which it is involved. I complemented the

aforementioned results by taking steps towards the characterisation of SMIM1's interactome. The strategy I adopted for this section of the study assumes that, since platelets acquire their RNA and proteins from MKs, the interactome of these two cell types should extensively overlap. Indeed, this extrapolation is also adopted in transcriptomic studies [270]. I created a hiPSC line overexpressing SMIM1-tagged with a TAP-tag (Bobc-TAP-SMIM1), which was then forward programmed into MKs. The mature MKs were collected for the intended immunoprecipitation and mass spectrometry studies. The time scale of this project did not allow for the optimisation of the immunoprecipitation protocol, which would have enabled the purification of protein required for proteomics studies.

Beyond the experiments proposed above and the study of other aspects of the molecular characterisation of SMIM1, it would also be important to perform functional assays assessing thrombus formation. Since platelets and neutrophils are involved in thrombus formation, the *in vivo* assessment of this process could provide new insights into the function of SMIM1 in these two cell types (as suggested in Chapter 5).

6.5 Assessment of SMIM1 role in erythropoiesis and RBC function

Alterations in the maturity, number, size and deformability of SMIM1 knockout RBCs were observed in the different models (Vel-negative blood donors, Vel-negative/weak individuals of UK Biobank and *Smim1* mutant mice; Chapter 4). This suggests that SMIM1 may be important for the formation and/or function of RBCs. The primary function of this cell type is the transportation of oxygen from the lungs to the tissues, which involves their mobility through the narrowest capillaries and for which morphologic and mechanical properties such as size and deformability are critical. Alterations in RBC traits such as RDW are associated with a higher incidence of cardiovascular diseases and liver disease [237, 238].

The role of SMIM1 during erythropoiesis could be further characterised through the adoption of an erythroid culture system of iPSC or $CD34^+$ cells generated/isolated from Vel-negative individuals differentiated into RBC. In addition, it would be interesting to assess if the decreased membrane deformability detected in the RBCs of *Smim1*$^{-/-}$ male mice is also observed in Vel-negative individuals.

6.6 A novel metabolic phenotype in human and mouse *SMIM1* knockouts

The characterisation of the Vel-negative donor cohort and *Smim1* mutant mouse model in this study identified two metabolic signatures: i) an increase of BMI and body weight in male Vel-negative blood donors and female *Smim1*$^{-/-}$ mice, respectively, and ii) an increase of lipid droplets or glycogen accumulation in the liver of *Smim1* mutant mice. These findings suggest the existence of a metabolic phenotype, which is probably independent of SMIM1 role in blood cells. In turn, this might have an important impact on the health of Vel/SMIM1-negative individuals, since these phenotypes may be associated with conditions such as non-alcoholic liver disease, insulin resistance and cardiovascular diseases.

This idea is further supported by two observations indicative of alterations in glucose metabolism, described by the International Mouse Phenotyping Consortium (http://mousephenotype.org/): i) *Smim1*$^{-/-}$ mice show increased levels of circulating fructosamine (p=7.12 x 10^{-5}, which is a glycated protein that reflects average glycemia level over the previous 2-3 weeks). Increased levels of circulating fructosamine have been associated with coronary artery atherosclerosis [294, 295] and prospectively associated with risk of morbidity and mortality of cardiovascular disease in patients with and without diabetes [296, 297]. Furthermore, this trait is an independent predictor of vascular outcomes [298]; and ii) the knockout mouse model of the *Nacad* gene, which was down-regulated in *Smim1*$^{-/-}$ platelets (p=2.57 x 10^{-5}), shows a decreased level of circulating glucose.

6.7 Investigation of SMIM1 role in metabolism

Since SMIM1 does not seem to be expressed in liver and adipose subcutaneous and visceral tissue (https://gtexportal.org/home/gene/SMIM1), these phenotypes are most likely an indirect effect of SMIM1 absence in other cells. To advance the understanding of SMIM1 association with metabolism it would be important to i) confirm the increase of BMI in Vel-negative individuals in a larger cohort such as the complete UK Biobank study (500,000 individuals); ii) assess whether metabolic parameters associated with glucose metabolism and liver damage (e.g. glucose, insulin and alanine aminotransferase) are affected by SMIM1 ablation; iii) assess whether the liver morphologic alteration (possibly steatosis) observed in mice is also present in Vel-

negative individuals (e.g. by examination of the liver with MRI in the complete UK Biobank or in the Vel-negative blood donors); iv) assess whether the metabolic phenotype is caused by SMIM1 absence in blood cells. For the latter study, a conditional *Smim1* knockout in haematopoietic cells or a bone marrow transplant of an *Smim1*$^{-/-}$ mouse into a wildtype mouse would be necessary. These experiments would clarify if the observed metabolic phenotype is a direct effect of SMIM1 ablation in blood cells but would not enable the determination of the cell type responsible for the observed phenotypes.

If the absence of SMIM1 is proven to induce metabolic syndrome, it will be important to determine the molecular mechanisms causing this phenotype. This may be particularly relevant for the treatment of these conditions, as well as for the prevention and treatment of Vel/SMIM1-negative individuals. Obesity, diabetes and liver-related diseases have a tremendous societal and economic impact in both developed and developing countries [299]. More specifically, these epidemic conditions represent a significant burden on national health care providers. Only in the UK, the national health system (NHS) estimates that £9.7 billion will be required to address overweight and obesity by 2050 [300].

CHAPTER 7 | REFERENCES

1. Jagannathan-Bogdan, M. and L.I. Zon, *Hematopoiesis.* Development, 2013. **140**(12): p. 2463-2467.
2. Xu, M.J., et al., Evidence for the presence of murine primitive megakaryocytopoiesis in the early yolk sac. Blood, 2001. **97**(7): p. 2016-22.
3. Mattia, G., et al., Different ploidy levels of megakaryocytes generated from peripheral or cord blood CD34+ cells are corre- lated with different levels of platelet release. Blood, 2002. **99**: p. 888-897.
4. Palis, J. and M.C. Yoder, *Yolk-sac hematopoiesis: the first blood cells of mouse and man.* Experimental Hematology, 2001. **29**(8): p. 927-936.
5. Palis, J., et al., Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. Development, 1999. **126**(22): p. 5073-84.
6. Tober, J., et al., The megakaryocyte lineage originates from hemangioblast precursors and is an integral component both of primitive and of definitive hematopoiesis. Blood, 2007. **109**(4): p. 1433-41.
7. Orkin, S.H. and L.I. Zon, *Hematopoiesis: An Evolving Paradigm for Stem Cell Biology.* Cell, 2008. **132**(4): p. 631-644.
8. Dzierzak, E. and A. Bigas, *Blood Development: Hematopoietic Stem Cell Dependence and Independence.* Cell Stem Cell, 2018. **22**(5): p. 639-651.
9. Dzierzak, E. and N.A. Speck, Of lineage and legacy: the development of mammalian hematopoietic stem cells. Nat Immunol, 2008. **9**(2): p. 129-36.
10. Tavian, M. and B. Peault, *Embryonic development of the human hematopoietic system.* Int J Dev Biol, 2005. **49**(2-3): p. 243-50.
11. McGrath, K.E., et al., A transient definitive erythroid lineage with unique regulation of the beta-globin locus in the mammalian embryo. Blood, 2011. **117**(17): p. 4600-8.
12. Cumano, A., C. Furlonger, and C.J. Paige, Differentiation and characterization of B-cell precursors detected in the yolk sac and embryo body of embryos beginning at the 10- to 12-somite stage. Proc Natl Acad Sci U S A, 1993. **90**(14): p. 6429-33.
13. Akashi, K., et al., A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. Nature, 2000. **404**(6774): p. 193-197.
14. Arinobu, Y., et al., Reciprocal activation of GATA-1 and PU.1 marks initial specification of hematopoietic stem cells into myeloerythroid and myelolymphoid lineages. Cell Stem Cell, 2007. **1**(4): p. 416-27.
15. Hamlett, I., et al., Characterization of megakaryocyte GATA1-interacting proteins: the corepressor ETO2 and GATA1 interact to regulate terminal megakaryocyte maturation. Blood, 2008. **112**(7): p. 2738-2749.
16. Foudi, A., et al., Distinct, strict requirements for Gfi-1b in adult bone marrow red cell and platelet generation. The Journal of Experimental Medicine, 2014. **211**(5): p. 909-927.
17. Muller-Sieburg, C.E., et al., Deterministic regulation of hematopoietic stem cell self-renewal and differentiation. Blood, 2002. **100**(4): p. 1302-9.
18. Ema, H., et al., Quantification of self-renewal capacity in single hematopoietic stem cells from normal and Lnk-deficient mice. Dev Cell, 2005. **8**(6): p. 907-14.
19. Wilson, A., et al., Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. Cell, 2008. **135**(6): p. 1118-29.
20. Yamamoto, R., et al., Clonal analysis unveils self-renewing lineage-restricted progenitors generated directly from hematopoietic stem cells. Cell, 2013. **154**(5): p. 1112-1126.
21. Verovskaya, E., et al., Heterogeneity of young and aged murine hematopoietic stem cells revealed by quantitative clonal analysis using cellular barcoding. Blood, 2013. **122**(4): p. 523-32.

22. Crisan, M. and E. Dzierzak, *The many faces of hematopoietic stem cell heterogeneity.* Development, 2016. **143**(24): p. 4571-4581.
23. Carrelha, J., et al., Hierarchically related lineage-restricted fates of multipotent haematopoietic stem cells. Nature, 2018. **554**(7690): p. 106-111.
24. Notta, F., et al., Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. Science, 2016. **351**(6269): p. aab2116.
25. Eaves, C.J., Hematopoietic stem cells: concepts, definitions, and the new reality. Blood, 2015. **125**(17): p. 2605-13.
26. Velten, L., et al., Human haematopoietic stem cell lineage commitment is a continuous process. Nat Cell Biol, 2017. **19**(4): p. 271-281.
27. Machlus, K.R. and J.E. Italiano, Jr., *The incredible journey: From megakaryocyte development to platelet formation.* J Cell Biol, 2013. **201**(6): p. 785-96.
28. Nakeff, A. and B. Maat, Separation of megakaryocytes from mouse bone marrow by velocity sedimentation. Blood, 1974. **43**(4): p. 591-5.
29. Schmitt, A., et al., Of mice and men: comparison of the ultrastructure of megakaryocytes and platelets. Exp Hematol, 2001. **29**(11): p. 1295-302.
30. Klimchenko, O., et al., A common bipotent progenitor generates the erythroid and megakaryocyte lineages in embryonic stem cell-derived primitive hematopoiesis. Blood, 2009. **114**: p. 1506-1517.
31. Nishikii, H., et al., Unipotent Megakaryopoietic Pathway Bridging Hematopoietic Stem Cells and Mature Megakaryocytes. Stem Cells, 2015. **33**: p. 2196-2207.
32. Ru, Y.X., et al., On the maturation of megakaryocytes: a review with original observations on human in vivo cells emphasizing morphology and ultrastructure. Ultrastruct Pathol, 2015. **39**(2): p. 79-87.
33. Lordier, L., et al., Megakaryocyte endomitosis is a failure of late cytokinesis related to defects in the contractile ring and Rho/Rock signaling. Blood, 2008. **112**: p. 3164-3164.
34. Tomer, A., L.A. Harker, and S.A. Burstein, *Flow cytometry analysis of normal humam megakaryocytes.* Blood, 1988. **71**: p. 1244-1252.
35. Raslova, H., et al., Megakaryocyte polyploidization is associated with a functional gene amplification. Blood, 2003. **101**: p. 541-544.
36. Schulze, H., et al., Characterization of the megakaryocyte demarcation membrane system and its role in thrombopoiesis. Blood, 2006. **107**(10): p. 3868-3875.
37. R.M., K., et al., *Origin of pulmonary megakaryocytes.* Blood, 1965. **25**: p. 767-775.
38. Yamada, E., The fine structure of the megakaryocyte in the mouse spleen. Acta Anat (Basel), 1957. **29**(3): p. 267-90.
39. de Sauvage, F.J., et al., Stimulation of megakaryocytopoiesis and thrombopoiesis by the c-Mpl ligand. Nature, 1994. **369**: p. 533 - 538.
40. Teramura, M., et al., Interleukin-11 enhances human megakaryocytopoiesis in vitro. Blood, 1992. **79**(2): p. 327-331.
41. Burstein, S.A., et al., Leukemia inhibitory factor and interleukin-11 promote maturation of murine and human megakaryocytes in vitro. Blood, 1992. **153**(2): p. 305-312.
42. Briddell, R.A., et al., *Effect of c-kit ligand on in vitro human megakaryocytopoiesis.* Blood, 1991. **78**(11): p. 2854 - 2859.
43. Ishibashi, T., et al., Human interleukin 6 is a direct promoter of maturation of megakaryocytesin vitro. Proc Natl Acad Sci USA, 1989. **86**(15): p. 5953-5957.
44. Ishibashi, T., J.A. Koziol, and S.A. Burstein, Human recombinant erythropoietin promotes differentiation of murine megakaryocytes in vitro. J Clin Invest, 1987. **79**(1): p. 286-289.

45. Ishibashi, T. and S.A. Burstein, Interleukin 3 promotes the differentiation of isolated single megakaryocytes. 167, 1986. **5**: p. 1512-1514.
46. Deutsch, V.R. and A. Tomer, Advances in megakaryocytopoiesis and thrombopoiesis: from bench to bedside. Br J Haematol, 2013. **161**(6): p. 778-93.
47. Shivdasani, R.A., et al., A lineage-selective knockout establishes the critical role of transcription factor GATA-1 in megakaryocyte growth and platelet development. The EMBO Journal, 1997. **16**(13): p. 3965-3973.
48. Lordier, L., et al., RUNX1-induced silencing of non-muscle myosin heavy chain IIB contributes to megakaryocyte polyploidization. Nat Commun, 2012. **3**: p. 717.
49. McCormack, M.P., et al., A critical role for the transcription factor Scl in platelet production during stress thrombopoiesis. Blood, 2006. **108**(7): p. 2248-56.
50. Takayama, M., et al., Genetic Analysis of Hierarchical Regulation for Gata1 and NF-E2 p45 Gene Expression in Megakaryopoiesis. Molecular and Cellular Biology, 2010. **30**(11): p. 2668-2680.
51. Tijssen, M.R., et al., Genome-wide analysis of simultaneous GATA1/2, RUNX1, FLI1, and SCL binding in megakaryocytes identifies hematopoietic regulators. Dev Cell, 2011. **20**(5): p. 597-609.
52. Nagata, Y., et al., Thrombopoietin Induces Megakaryocyte Differentiation in Hematopoietic Progenitor FDC-P2 Cells. Journal of Biological Chemistry, 1995. **270**(34): p. 19673-19675.
53. Bartley, T.D., et al., Identification and cloning of a megakaryocyte growth and development factor that is a ligand for the cytokine receptor MpI. Cell, 1994. **77**(7): p. 1117-1124.
54. Alexander, W.S., et al., Deficiencies in progenitor cells of multiple hematopoietic lineages and defective megakaryocytopoiesis in mice lacking the thrombopoietic receptor c-Mpl. Vol. 87. 1996. 2162-2170.
55. Deutsch, V.R. and A. Tomer, *Megakaryocyte development and platelet production.* British Journal of Haematology, 2006. **134**(5): p. 453-466.
56. Gurney, A.L., et al., *Thrombocytopenia in c-mpl-deficient mice.* Science, 1994. **265**(5177): p. 1445-1447.
57. de Sauvage, F.J., et al., *Physiological regulation of early and late stages of megakaryocytopoiesis by thrombopoietin.* The Journal of Experimental Medicine, 1996. **183**(2): p. 651-656.
58. Nihimura, S., et al., IL-1α induces thrombopoiesis through megakaryocyte rupture in response to accute platelets needs. J. Cell Biol., 2015.
59. Kimura, S., et al., Hematopoietic stem cell deficiencies in mice lacking c-Mpl, the receptor for thrombopoietin. Proc Natl Acad Sci USA, 1998. **95**(3): p. 1195-1200.
60. Bruns, I., et al., Megakaryocytes regulate hematopoietic stem cell quiescence via Cxcl4 secretion. Nature Medicine, 2014. **20**(11): p. 1315-1320.
61. Zhao, M., et al., Megakaryocytes maintain homeostatic quiescence and promote post-injury regeneration of hematopoietic stem cells. Nat Med, 2014. **20**(11): p. 1321-1326.
62. Chang, Y., et al., *From hematopoietic stem cells to platelets.* J Thromb Haemost, 2007. **5 Suppl 1**: p. 318-27.
63. Walsh, T.G., P. Metharom, and M.C. Berndt, *The functional role of platelets in the regulation of angiogenesis.* Platelets, 2015. **26**(3): p. 199-211.
64. Semple, J.W., J.E. Italiano, Jr., and J. Freedman, *Platelets and the immune continuum.* Nat Rev Immunol, 2011. **11**(4): p. 264-74.
65. Deppermann, C. and P. Kubes, Start a fire, kill the bug: The role of platelets in inflammation and infection. Innate Immun, 2018. **24**(6): p. 335-348.
66. Gawaz, M. and S. Vogel, Platelets in tissue repair: control of apoptosis and interactions with regenerative cells. Blood, 2013. **122**(15): p. 2550-4.

67. Golebiewska, E.M. and A.W. Poole, *Platelet secretion: From haemostasis to wound healing and beyond.* Blood Rev, 2015. **29**(3): p. 153-62.
68. Leslie, M., *Beyond Clotting: The Powers of Platelets.* Science, 2010. **328**(5978): p. 562-564.
69. Machlus, K.R. and J.E. Italiano, *The incredible journey: From megakaryocyte development to platelet formation.* The Journal of Cell Biology, 2013. **201**(6): p. 785-796.
70. Thon, J.N., et al., Microtubule and cortical forces determine platelet size during vascular platelet production. Nat Commun, 2012. **3**: p. 852.
71. Harker, L.A. and C.A. Finch, *Thrombokinetics in man.* J Clin Invest, 1969. **48**(6): p. 963-74.
72. Grozovsky, R., K.M. Hoffmeister, and H. Falet, *Novel clearance mechanisms of platelets.* Curr Opin Hematol, 2010. **17**(6): p. 585-9.
73. Kaushansky, K., *The molecular mechanisms that control thrombopoiesis.* J Clin Invest, 2005. **115**(12): p. 3339-47.
74. Michelson, A.D., *Platelets*. 3rd ed. Megakaryocyte development and platelet formation ed. J.E.I.J.J.H. Hartwig. 2013, London ; Waltham, MA: Academic Press. xliv, 1353 p.
75. Patel, S.R., J.H. Hartwig, and J.E. Italiano, Jr., *The biogenesis of platelets from megakaryocyte proplatelets.* J Clin Invest, 2005. **115**(12): p. 3348-54.
76. Italiano, J.E., Jr., et al., Blood platelets are assembled principally at the ends of proplatelet processes produced by differentiated megakaryocytes. J Cell Biol, 1999. **147**(6): p. 1299-312.
77. Thon, J.N., et al., Cytoskeletal mechanics of proplatelet maturation and platelet release. J Cell Biol, 2010. **191**(4): p. 861-74.
78. Thon, J.N. and J.E. Italiano, Jr., *Does size matter in platelet production?* Blood, 2012. **120**(8): p. 1552-61.
79. Handagama, P.J., et al., Circulating proplatelets: isolation and quantitation in healthy rats and in rats with induced acute blood loss. Am J Vet Res, 1987. **48**(6): p. 962-5.
80. Junt, T., et al., *Dynamic visualization of thrombopoiesis within bone marrow.* Science, 2007. **317**(5845): p. 1767-70.
81. Dunois-Larde, C., et al., Exposure of human megakaryocytes to high shear rates accelerates platelet production. Blood, 2009. **114**(9): p. 1875-83.
82. Nieswandt, B. and S. Stritt, *Megakaryocyte rupture for acute platelet needs.* J Cell Biol, 2015. **209**(3): p. 327-8.
83. Bianchi, E., et al., Genomic landscape of megakaryopoiesis and platelet function defects. Blood, 2016. **127**(10): p. 1249-59.
84. Eto, K. and S. Kunishima, Linkage between the mechanisms of thrombocytopenia and thrombopoiesis. Blood, 2016. **127**(10): p. 1234-1241.
85. Schulze, H. and D. Stegner, *Imaging platelet biogenesis in vivo.* Res Pract Thromb Haemost, 2018. **2**(3): p. 461-468.
86. Patel, S.R., et al., Differential roles of microtubule assembly and sliding in proplatelet formation by megakaryocytes. Blood, 2005. **106**(13): p. 4076-85.
87. Kelley, M.J., et al., Mutation of MYH9, encoding non-muscle myosin heavy chain A, in May-Hegglin anomaly. Nat Genet, 2000. **26**(1): p. 106-8.
88. Seri, M., et al., Mutations in MYH9 result in the May-Hegglin anomaly, and Fechtner and Sebastian syndromes. The May-Heggllin/Fechtner Syndrome Consortium. Nat Genet, 2000. **26**(1): p. 103-5.
89. Kunishima, S., et al., *ACTN1 mutations cause congenital macrothrombocytopenia.* Am J Hum Genet, 2013. **92**(3): p. 431-8.

90. Kunishima, S., et al., Mutation of the beta1-tubulin gene associated with congenital macrothrombocytopenia affecting microtubule assembly. Blood, 2009. **113**(2): p. 458-61.
91. Schwer, H.D., et al., A lineage-restricted and divergent beta-tubulin isoform is essential for the biogenesis, structure and function of blood platelets. Curr Biol, 2001. **11**(8): p. 579-86.
92. Bluteau, D., et al., Dysmegakaryopoiesis of FPD/AML pedigrees with constitutional RUNX1 mutations is linked to myosin II deregulated expression. Blood, 2012. **120**(13): p. 2708-18.
93. Dutting, S., et al., A Cdc42/RhoA regulatory circuit downstream of glycoprotein Ib guides transendothelial platelet biogenesis. Nat Commun, 2017. **8**: p. 15838.
94. Kanaji, T., et al., GPIbalpha regulates platelet size by controlling the subcellular localization of filamin. Blood, 2012. **119**(12): p. 2906-13.
95. Sabri, S., et al., Differential regulation of actin stress fiber assembly and proplatelet formation by alpha2beta1 integrin and GPVI in human megakaryocytes. Blood, 2004. **104**(10): p. 3117-25.
96. Zou, Z., et al., Negative regulation of activated alpha-2 integrins during thrombopoiesis. Blood, 2009. **113**(25): p. 6428-39.
97. Avecilla, S.T., et al., Chemokine-mediated interaction of hematopoietic progenitors with the bone marrow vascular niche is required for thrombopoiesis. Nat Med, 2004. **10**(1): p. 64-71.
98. Pitchford, S.C., T. Lodie, and S.M. Rankin, VEGFR1 stimulates a CXCR4-dependent translocation of megakaryocytes to the vascular niche, enhancing platelet production in mice. Blood, 2012. **120**(14): p. 2787-95.
99. Zhang, L., et al., A novel role of sphingosine 1-phosphate receptor S1pr1 in mouse thrombopoiesis. J Exp Med, 2012. **209**(12): p. 2165-81.
100. Stegner, D., et al., Thrombopoiesis is spatially regulated by the bone marrow vasculature. Nat Commun, 2017. **8**(1): p. 127.
101. Larson, M.K. and S.P. Watson, Regulation of proplatelet formation and platelet release by integrin alpha IIb beta3. Blood, 2006. **108**(5): p. 1509-14.
102. Matsunaga, T., et al., Potentiated activation of VLA-4 and VLA-5 accelerates proplatelet-like formation. Ann Hematol, 2012. **91**(10): p. 1633-43.
103. Takahashi, R., N. Sekine, and T. Nakatake, Influence of monoclonal antiplatelet glycoprotein antibodies on in vitro human megakaryocyte colony formation and proplatelet formation. Blood, 1999. **93**(6): p. 1951-8.
104. Levine, R.F., et al., Circulating megakaryocytes: delivery of large numbers of intact, mature megakaryocytes to the lungs. Eur J Haematol, 1993. **51**(4): p. 233-46.
105. Zucker-Franklin, D. and C.S. Philipp, Platelet production in the pulmonary capillary bed: new ultrastructural evidence for an old concept. Am J Pathol, 2000. **157**(1): p. 69-74.
106. Lefrancais, E., et al., The lung is a site of platelet biogenesis and a reservoir for haematopoietic progenitors. Nature, 2017. **544**(7648): p. 105-109.
107. Kaufman, R.M., et al., Circulating megakaryocytes and platelet release in the lung. Blood, 1965. **26**(6): p. 720-31.
108. Pedersen, N.T., Occurrence of megakaryocytes in various vessels and their retention in the pulmonary capillaries in man. Scand J Haematol, 1978. **21**(5): p. 369-75.
109. Harrison, P., et al., Uptake of plasma fibrinogen into the alpha granules of human megakaryocytes and platelets. J Clin Invest, 1989. **84**(4): p. 1320-4.
110. Handagama, P., et al., Endocytosis of fibrinogen into megakaryocyte and platelet alpha-granules is mediated by alpha IIb beta 3 (glycoprotein IIb-IIIa). Blood, 1993. **82**(1): p. 135-8.

111. Klement, G.L., et al., *Platelets actively sequester angiogenesis regulators.* Blood, 2009. **113**(12): p. 2835-42.
112. White, J.G. and C.C. Clawson, The surface-connected canalicular system of blood platelets--a fenestrated membrane system. Am J Pathol, 1980. **101**(2): p. 353-64.
113. Thon, J.N. and J.E. Italiano, *Platelets: production, morphology and ultrastructure.* Handb Exp Pharmacol, 2012(210): p. 3-22.
114. Cimmino, G. and P. Golino, *Platelet biology and receptor pathways.* J Cardiovasc Transl Res, 2013. **6**(3): p. 299-309.
115. Labelle, M., S. Begum, and R.O. Hynes, *Platelets guide the formation of early metastatic niches.* Proc Natl Acad Sci U S A, 2014. **111**(30): p. E3053-61.
116. Deuel, T.F., et al., *Platelet factor 4 is chemotactic for neutrophils and monocytes.* Proc Natl Acad Sci U S A, 1981. **78**(7): p. 4584-7.
117. Rumbaut, R.E. and P. Thiagarajan, General Characteristics of Platelets, in Platelet-Vessel Wall Interactions in Hemostasis and Thrombosis. 2010: San Rafael (CA).
118. Maynard, D.M., et al., *Proteomic analysis of platelet alpha-granules using mass spectrometry.* J Thromb Haemost, 2007. **5**(9): p. 1945-55.
119. Blair, P. and R. Flaumenhaft, *Platelet alpha-granules: basic biology and clinical correlates.* Blood Rev, 2009. **23**(4): p. 177-89.
120. Heijnen, H. and P. van der Sluijs, *Platelet secretory behaviour: as diverse as the granules ... or not?* J Thromb Haemost, 2015. **13**(12): p. 2141-51.
121. Italiano, J.E., Jr. and E.M. Battinelli, *Selective sorting of alpha-granule proteins.* J Thromb Haemost, 2009. **7 Suppl 1**: p. 173-6.
122. Israels, S.J., et al., Platelet dense granule membranes contain both granulophysin and P-selectin (GMP-140). Blood, 1992. **80**(1): p. 143-52.
123. Chen, D., et al., Molecular mechanisms of platelet exocytosis: role of SNAP-23 and syntaxin 2 in dense core granule release. Blood, 2000. **95**(3): p. 921-9.
124. Thon, J.N., et al., T granules in human platelets function in TLR9 organization and signaling. J Cell Biol, 2012. **198**(4): p. 561-74.
125. Cho, J., et al., A critical role for extracellular protein disulfide isomerase during thrombus formation in mice. J Clin Invest, 2008. **118**(3): p. 1123-31.
126. Marcus, A.J., et al., The endothelial cell ecto-ADPase responsible for inhibition of platelet function is CD39. J Clin Invest, 1997. **99**(6): p. 1351-60.
127. Dittman, W.A. and P.W. Majerus, *Structure and function of thrombomodulin: a natural anticoagulant.* Blood, 1990. **75**(2): p. 329-36.
128. Swieringa, F., et al., Targeting platelet receptor function in thrombus formation: the risk of bleeding. Blood Rev, 2014. **28**(1): p. 9-21.
129. Smolenski, A., *Novel roles of cAMP/cGMP-dependent signaling in platelets.* J Thromb Haemost, 2012. **10**(2): p. 167-76.
130. Naseem, K.M. and W. Roberts, *Nitric oxide at a glance.* Platelets, 2011. **22**(2): p. 148-52.
131. Rivera, J., et al., Platelet receptors and signaling in the dynamics of thrombus formation. Haematologica, 2009. **94**(5): p. 700-11.
132. Rumbaut, R.E. and P. Thiagarajan, Platelet Adhesion to Vascular Walls, in Platelet-Vessel Wall Interactions in Hemostasis and Thrombosis. 2010: San Rafael (CA).
133. Savage, B., F. Almus-Jacobs, and Z.M. Ruggeri, Specific synergy of multiple substrate-receptor interactions in platelet thrombus formation under flow. Cell, 1998. **94**(5): p. 657-66.
134. Gibbins, J.M., Platelet adhesion signalling and the regulation of thrombus formation. J Cell Sci, 2004. **117**(Pt 16): p. 3415-25.

135. Sarratt, K.L., et al., GPVI and alpha2beta1 play independent critical roles during platelet adhesion and aggregate formation to collagen under flow. Blood, 2005. **106**(4): p. 1268-77.
136. Bye, A.P., A.J. Unsworth, and J.M. Gibbins, *Platelet signaling: a complex interplay between inhibitory and activatory networks.* J Thromb Haemost, 2016. **14**(5): p. 918-30.
137. Patel-Hett, S., et al., Visualization of microtubule growth in living platelets reveals a dynamic marginal band with multiple microtubules. Blood, 2008. **111**(9): p. 4605-16.
138. Harper, M.T. and A.W. Poole, Diverse functions of protein kinase C isoforms in platelet activation and thrombus formation. J Thromb Haemost, 2010. **8**(3): p. 454-62.
139. Heemskerk, J.W., E.M. Bevers, and T. Lindhout, *Platelet activation and blood coagulation.* Thromb Haemost, 2002. **88**(2): p. 186-93.
140. Murugappan, S., H. Shankar, and S.P. Kunapuli, *Platelet receptors for adenine nucleotides and thromboxane A2.* Semin Thromb Hemost, 2004. **30**(4): p. 411-8.
141. Lentz, B.R., Exposure of platelet membrane phosphatidylserine regulates blood coagulation. Prog Lipid Res, 2003. **42**(5): p. 423-38.
142. Kahn, M.L., et al., Protease-activated receptors 1 and 4 mediate activation of human platelets by thrombin. J Clin Invest, 1999. **103**(6): p. 879-87.
143. Ishihara, H., et al., Antibodies to protease-activated receptor 3 inhibit activation of mouse platelets by thrombin. Blood, 1998. **91**(11): p. 4152-7.
144. Furie, B. and B.C. Furie, *Mechanisms of thrombus formation.* N Engl J Med, 2008. **359**(9): p. 938-49.
145. Niiya, K., et al., Increased surface expression of the membrane glycoprotein IIb/IIIa complex induced by platelet activation. Relationship to the binding of fibrinogen and platelet aggregation. Blood, 1987. **70**(2): p. 475-83.
146. Goto, S., et al., Distinct mechanisms of platelet aggregation as a consequence of different shearing flow conditions. J Clin Invest, 1998. **101**(2): p. 479-86.
147. Ma, A.C. and P. Kubes, Platelets, neutrophils, and neutrophil extracellular traps (NETs) in sepsis. J Thromb Haemost, 2008. **6**(3): p. 415-20.
148. Maugeri, N., et al., Activated platelets present high mobility group box 1 to neutrophils, inducing autophagy and promoting the extrusion of neutrophil extracellular traps. J Thromb Haemost, 2014. **12**(12): p. 2074-88.
149. Brill, A., et al., Neutrophil extracellular traps promote deep vein thrombosis in mice. J Thromb Haemost, 2012. **10**(1): p. 136-44.
150. Gould, T.J., et al., Neutrophil extracellular traps promote thrombin generation through platelet-dependent and platelet-independent mechanisms. Arterioscler Thromb Vasc Biol, 2014. **34**(9): p. 1977-84.
151. Stark, K., et al., Disulfide HMGB1 derived from platelets coordinates venous thrombosis in mice. Blood, 2016. **128**(20): p. 2435-2449.
152. Semeraro, F., et al., Extracellular histones promote thrombin generation through platelet-dependent mechanisms: involvement of platelet TLR2 and TLR4. Blood, 2011. **118**(7): p. 1952-61.
153. de Bont, C.M., W.C. Boelens, and G.J.M. Pruijn, *NETosis, complement, and coagulation: a triangular relationship.* Cell Mol Immunol, 2018.
154. Dzierzak, E. and S. Philipsen, *Erythropoiesis: Development and Differentiation.* Cold Spring Harbor Perspectives in Medicine, 2013. **3**(4).
155. Diez-Silva, M., et al., Shape and Biomechanical Characteristics of Human Red Blood Cells in Health and Disease. MRS Bull, 2010. **35**(5): p. 382-388.
156. Dean, L., *Blood Groups and Red Cell Antigens*, ed. B. (MD). 2005, US: National Center for Biotechnology Information.

157. Higgs, D.R. and W.G. Wood, *Erythropoiesis*, in *Postgraduate Haematology*. 2010, Wiley-Blackwell. p. 12-25.
158. Brinkman, R., A. Wildschut, and A. Wittermans, *On the occurrence of two kinds of hæmoglobin in normal human blood.* The Journal of Physiology, 1934. **80**(4): p. 377-387.
159. Wilber, A., A.W. Nienhuis, and D.A. Persons, Transcriptional regulation of fetal to adult hemoglobin switching: new therapeutic opportunities. Vol. 117. 2011. 3945-3953.
160. Socolovsky, M., *Molecular insights into stress erythropoiesis.* Curr Opin Hematol, 2007. **14**(3): p. 215-24.
161. Chen, K., et al., Resolving the distinct stages in erythroid differentiation based on dynamic changes in membrane protein expression during erythropoiesis. Proc Natl Acad Sci U S A, 2009. **106**(41): p. 17413-8.
162. Suzuki, M., et al., GATA factor switching from GATA2 to GATA1 contributes to erythroid differentiation. Genes Cells, 2013. **18**(11): p. 921-33.
163. Vassen, L., et al., Growth factor independence 1b (gfi1b) is important for the maturation of erythroid cells and the regulation of embryonic globin expression. PLoS One, 2014. **9**(5): p. e96636.
164. Hodge, D., et al., A global role for EKLF in definitive and primitive erythropoiesis. Blood, 2006. **107**(8): p. 3359-3370.
165. Menon, M.P., et al., Signals for stress erythropoiesis are integrated via an erythropoietin receptor-phosphotyrosine-343-Stat5 axis. Journal of Clinical Investigation, 2006. **116**(3): p. 683-694.
166. Obara, N., et al., Repression via the GATA box is essential for tissue-specific erythropoietin gene expression. Vol. 111. 2008. 5223-5232.
167. Krantz, S.B., *Erythropoietin.* Blood, 1991. **77**(3): p. 419-434.
168. Jacobs, K., et al., Isolation and characterization of genomic and cDNA clones of human erythropoietin. Nature, 1985. **313**(6005): p. 806-10.
169. Beru, N., et al., *Expression of the erythropoietin gene.* Molecular and Cellular Biology, 1986. **6**(7): p. 2571-2575.
170. Kirby, S.L., et al., Proliferation of multipotent hematopoietic cells controlled by a truncated erythropoietin receptor transgene. Proceedings of the National Academy of Sciences, 1996. **93**(18): p. 9402-9407.
171. Liboi, E., et al., *Erythropoietin receptor signals both proliferation and erythroid-specific differentiation.* Proceedings of the National Academy of Sciences, 1993. **90**(23): p. 11351-11355.
172. Kelley, L.L., et al., Survival or death of individual proerythroblasts results from differing erythropoietin sensitivities: a mechanism for controlled rates of erythrocyte production. Vol. 82. 1993. 2340-2352.
173. Koury, M.J. and M.C. Bondurant, *The molecular mechanism of erythropoietin action.* European Journal of Biochemistry, 1992. **210**(3): p. 649-663.
174. Witthuhn, B.A., et al., JAK2 associates with the erythropoietin receptor and is tyrosine phosphorylated and activated following stimulation with erythropoietin. Cell, 1993. **74**(2): p. 227-236.
175. Constantinescu, S.N., S. Ghaffari, and H.F. Lodish, *The Erythropoietin Receptor: Structure, Activation and Intracellular Signal Transduction.* Trends Endocrinol Metab, 1999. **10**(1): p. 18-23.
176. Suzuki, N., Erythropoietin Gene Expression: Developmental-Stage Specificity, Cell-Type Specificity, and Hypoxia Inducibility. The Tohoku Journal of Experimental Medicine, 2015. **235**(3): p. 233-240.

177. van den Akker, E., et al., Investigating the key membrane protein changes during in vitro erythropoiesis of protein 4.2 (-) cells (mutations Chartres 1 and 2). Haematologica, 2010. **95**(8): p. 1278-86.
178. Dean, L., *Blood group antigens are surface markers on the red blood cell membrane*, in *Blood Groups and Red Cell Antigens* B. (MD), Editor. 2005, National Center for Biotechnology Information (US): [Internet].
179. Daniels, G.L., et al., Blood group terminology 2004: from the International Society of Blood Transfusion committee on terminology for red cell surface antigens. Vox Sang, 2004. **87**(4): p. 304-16.
180. Storry, J.R., et al., International Society of Blood Transfusion Working Party on red cell immunogenetics and blood group terminology: Cancun report (2012). Vox Sang, 2014. **107**(1): p. 90-6.
181. Sussman, L.N. and E.B. Miller, *New blood factor: Vel.* Hemat, 1952. **7**(3): p. 368-371.
182. Ballif, B.A., et al., *Disruption of SMIM1 causes the Vel- blood type.* EMBO Mol Med, 2013. **5**(5): p. 751-61.
183. Levine, P., J.A. White, and M. Stroup, *Seven Ve-a (Vel) negative members in three generations of a family.* Transfusion, 1961. **1**(111-115).
184. Daniels, G., *Human Blood Groups*. 2002, Oxford: Blackwell Science.
185. Cedergren, B., C.M. Giles, and E.W. Ikin, *The Vel blood group in northern Sweden.* Vox Sang, 1976. **31**: p. 344-355.
186. Cvejic, A., et al., SMIM1 underlies the Vel blood group and influences red blood cell traits. Nat Genet, 2013. **45**(5): p. 542-5.
187. Haer-Wigman, L., et al., *Impact of genetic variation in the SMIM1 gene on Vel expression levels.* Transfusion, 2015. **55**(6 Pt 2): p. 1457-66.
188. Storry, J.R., et al., Homozygosity for a null allele of SMIM1 defines the Vel-negative blood group phenotype. Nat Genet, 2013. **45**(5): p. 537-41.
189. Christophersen, M.K., et al., SMIM1 variants rs1175550 and rs143702418 independently modulate Vel blood group antigen expression. Sci Rep, 2017. **7**: p. 40451.
190. Ulirsch, J.C., et al., Systematic Functional Dissection of Common Genetic Variation Affecting Red Blood Cell Traits. Cell, 2016. **165**(6): p. 1530-1545.
191. van Gammeren, A.J., et al., *Haemolytic disease of the newborn because of rare anti-Vel.* Transfus Med, 2008. **18**(3): p. 197-8.
192. Vucinovic, M., et al., [Haemolytic disease of the newborn--from a mother with anti-Kell, anti-E and anti-Vel anti-erythrocyte alloantibodies]. Z Geburtshilfe Neonatol, 2004. **208**(5): p. 197-202.
193. Becton, D.L. and T.R. Kinney, An infant girl with severe autoimmune hemolytic anemia: apparent anti-Vel specificity. Vox Sang, 1986. **51**(2): p. 108-11.
194. Ferrer, Z., et al., *A third example of haemolytic auto-anti-Vel.* Rev Fr Transfus Immunohematol, 1984. **27**(5): p. 639-44.
195. Chen, L., et al., Genetic Drivers of Epigenetic and Transcriptional Variation in Human Immune Cells. Cell, 2016. **167**(5): p. 1398-1414 e24.
196. Astle, W.J., et al., The Allelic Landscape of Human Blood Cell Trait Variation and Links to Common Complex Disease. Cell, 2016. **167**(5): p. 1415-1429 e19.
197. Adams, D., et al., *BLUEPRINT to decode the epigenetic signature written in blood.* Nat Biotechnol, 2012. **30**(3): p. 224-6.
198. Michelson, A.D., Platelet activation by thrombin can be directly measured in whole blood through the use of the peptide GPRP and flow cytometry: methods and clinical applications. Blood Coagul Fibrinolysis, 1994. **5**(1): p. 121-31.
199. Bray, N.L., et al., *Near-optimal probabilistic RNA-seq quantification.* Nat Biotechnol, 2016. **34**(5): p. 525-7.

200. Rosendahl, P., et al., *Real-time fluorescence and deformability cytometry.* Nat Methods, 2018. **15**(5): p. 355-358.
201. Rouhani, F., et al., Genetic Background Drives Transcriptional Variation in Human Induced Pluripotent Stem Cells. PLos Genetics, 2014. **10**(6): p. e1004432.
202. Yusa, K., et al., Targeted gene correction of alpha1-antitrypsin deficiency in induced pluripotent stem cells. Nature, 2011. **478**(7369): p. 391-4.
203. Jinek, M., et al., A Programmable Dual-RNA–Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science, 2012. **337**: p. 816-821.
204. Shalem, O., et al., Genome-scale CRISPR-Cas9 knockout screening in human cells. . Science, 2014. **343**: p. 83-87.
205. Sanjana, N.E., O. Shalem, and F. Zhang, *Improved lentiviral vectors and genome-wide libraries for CRISPR screening.* Nat Methods, 2014. **11**: p. 783–784.
206. Ran, F.A., et al., *Genome engineering using the CRISPR-Cas9 system.* Nat Protoc, 2013. **8**(11): p. 2281-2308.
207. Moreau, T., et al., Large-scale production of megakaryocytes from human pluripotent stem cells by chemically defined forward programming. Nat Commun, 2016. **7**: p. 11208.
208. Arnaud, L., et al., SMIM1 is a type II transmembrane phosphoprotein and displays the Vel blood group antigen at its carboxyl-terminus. FEBS Lett, 2015.
209. Mertins, P., et al., Proteogenomics connects somatic mutations to signalling in breast cancer. Nature, 2016. **534**(7605): p. 55-62.
210. Boeing, S., et al., *Multiomic Analysis of the UV-Induced DNA Damage Response.* Cell Rep, 2016. **15**(7): p. 1597-1610.
211. Bian, Y., et al., An enzyme assisted RP-RPLC approach for in-depth analysis of human liver phosphoproteome. J Proteomics, 2014. **96**: p. 253-62.
212. Danger, Y., et al., Characterization of a new human monoclonal antibody directed against the Vel antigen. Vox Sang, 2015.
213. Satchwell, T.J., et al., Critical band 3 multiprotein complex interactions establish early during human erythropoiesis. Blood, 2011. **118**(1): p. 182-191.
214. Watson, S.P., et al., *GPVI and integrin alphaIIb beta3 signaling in platelets.* J Thromb Haemost, 2005. **3**(8): p. 1752-62.
215. Chasis, J.A. and N. Mohandas, *Red blood cell glycophorins.* Blood, 1992. **80**(8): p. 1869-79.
216. Rath, A., et al., Detergent binding explains anomalous SDS-PAGE migration of membrane proteins. Proc Natl Acad Sci U S A, 2009. **106**(6): p. 1760-5.
217. van der Harst, P., et al., *Seventy-five genetic loci influencing the human red blood cell.* Nature, 2012. **492**(7429): p. 369-75.
218. Barbara J. Bain MBBS, F., FRCPath,, *Blood Cells: A Practical Guide*. Fifth Edition ed. 2015.
219. Schaffer, A., et al., Impact of red blood cells count on the relationship between high density lipoproteins and the prevalence and extent of coronary artery disease: a single centre study (vol 40, pg 61, 2015). Journal of Thrombosis and Thrombolysis, 2015. **40**(1): p. 69-69.
220. Agrawal, R., et al., Assessment of red blood cell deformability in type 2 diabetes mellitus and diabetic retinopathy by dual optical tweezers stretching technique. Scientific Reports, 2016. **6**.
221. Routes, J., et al., *ICON: The Early Diagnosis of Congenital Immunodeficiencies.* Journal of Clinical Immunology, 2014. **34**(4): p. 398-424.
222. Ober, C., D.A. Loisel, and Y. Gilad, *Sex-specific genetic architecture of human disease.* Nat Rev Genet, 2008. **9**(12): p. 911-22.
223. Murphy, W.G., The sex difference in haemoglobin levels in adults - mechanisms, causes, and consequences. Blood Rev, 2014. **28**(2): p. 41-7.

224. Karp, N.A., et al., Prevalence of sexual dimorphism in mammalian phenotypic traits. Nat Commun, 2017. **8**: p. 15475.
225. Droyvold, W.B., et al., Change in height, weight and body mass index: Longitudinal data from the HUNT Study in Norway. Int J Obes (Lond), 2006. **30**(6): p. 935-9.
226. Schaart, G., et al., A modified PAS stain combined with immunofluorescence for quantitative analyses of glycogen in muscle sections. Histochem Cell Biol, 2004. **122**(2): p. 161-9.
227. Piva, E., et al., Automated reticulocyte counting: state of the art and clinical applications in the evaluation of erythropoiesis. Clin Chem Lab Med, 2010. **48**(10): p. 1369-80.
228. Hoffmann, J.J., *Reticulated platelets: analytical aspects and clinical utility.* Clin Chem Lab Med, 2014. **52**(8): p. 1107-17.
229. Di Angelantonio, E., et al., Efficiency and safety of varying the frequency of whole blood donation (INTERVAL): a randomised trial of 45 000 donors. Lancet, 2017. **390**(10110): p. 2360-2371.
230. Mohandas, N. and J.A. Chasis, Red-Blood-Cell Deformability, Membrane Material Properties and Shape - Regulation by Transmembrane, Skeletal and Cytosolic Proteins and Lipids. Seminars in Hematology, 1993. **30**(3): p. 171-192.
231. Popescu, G., et al., *Imaging red blood cell dynamics by quantitative phase microscopy.* Blood Cells Molecules and Diseases, 2008. **41**(1): p. 10-16.
232. Mohandas, N. and E. Evans, *Mechanical-Properties of the Red-Cell Membrane in Relation to Molecular-Structure and Genetic-Defects.* Annual Review of Biophysics and Biomolecular Structure, 1994. **23**: p. 787-818.
233. Tomaiuolo, G., Biomechanical properties of red blood cells in health and disease towards microfluidics. Biomicrofluidics, 2014. **8**(5): p. 051501.
234. Da Costa, L., et al., Hereditary spherocytosis, elliptocytosis, and other red cell membrane disorders. Blood Rev, 2013. **27**(4): p. 167-78.
235. Hosseini, S.M. and J.J. Feng, *How Malaria Parasites Reduce the Deformability of Infected Red Blood Cells.* Biophysical Journal, 2012. **103**(1): p. 1-10.
236. Pivkin, I.V., et al., *Biomechanics of red blood cells in human spleen and consequences for physiology and disease.* Proceedings of the National Academy of Sciences of the United States of America, 2016. **113**(28): p. 7804-7809.
237. Li, N., H. Zhou, and Q.Z. Tang, Red Blood Cell Distribution Width: A Novel Predictive Indicator for Cardiovascular and Cerebrovascular Diseases. Disease Markers, 2017.
238. Hu, Z.D., et al., *Red blood cell distribution width is a potential prognostic index for liver disease.* Clinical Chemistry and Laboratory Medicine, 2013. **51**(7): p. 1403-1408.
239. Zini, G., *Abnormalities in leukocyte morphology and number*, in *Blood and Bone Marrow Pathology*, J.M.a.W.N.E. Anna Porwit, Editor. 2011, Churchill Livingstone. p. 247-261.
240. von Vietinghoff, S. and K. Ley, *Homeostatic regulation of blood neutrophil counts.* J Immunol, 2008. **181**(8): p. 5183-8.
241. Molero, L., et al., Expression of estrogen receptor subtypes and neuronal nitric oxide synthase in neutrophils from women and men: regulation by estrogen. Cardiovasc Res, 2002. **56**(1): p. 43-51.
242. Molloy, E.J., et al., Sex-specific alterations in neutrophil apoptosis: the role of estradiol and progesterone. Blood, 2003. **102**(7): p. 2653-9.
243. Madalli, S., et al., Sex-specific regulation of chemokine Cxcl5/6 controls neutrophil recruitment and tissue injury in acute inflammatory states. Biol Sex Differ, 2015. **6**: p. 27.

244. Whitacre, C.C., *Sex differences in autoimmune disease.* Nat Immunol, 2001. **2**(9): p. 777-80.
245. Ngo, S.T., F.J. Steyn, and P.A. McCombe, *Gender differences in autoimmune disease.* Front Neuroendocrinol, 2014. **35**(3): p. 347-69.
246. Casimir, G.J., et al., Sex and inflammation in respiratory diseases: a clinical viewpoint. Biol Sex Differ, 2013. **4**: p. 16.
247. Grassi, L., et al., Dynamics of Transcription Regulation in Human Bone Marrow Myeloid Differentiation to Mature Blood Neutrophils. Cell Rep, 2018. **24**(10): p. 2784-2794.
248. Martin, J.F., et al., The biological significance of platelet volume: its relationship to bleeding time, platelet thromboxane B2 production and megakaryocyte nuclear DNA concentration. Thromb Res, 1983. **32**(5): p. 443-60.
249. Thompson, C.B., et al., Size dependent platelet subpopulations: relationship of platelet volume to ultrastructure, enzymatic activity, and function. Br J Haematol, 1982. **50**(3): p. 509-19.
250. Martin, J.F., et al., The causal role of megakaryocyte-platelet hyperactivity in acute coronary syndromes. Nat Rev Cardiol, 2012. **9**(11): p. 658-70.
251. Klovaite, J., et al., High platelet volume and increased risk of myocardial infarction: 39,531 participants from the general population. J Thromb Haemost, 2011. **9**(1): p. 49-56.
252. Chu, S.G., et al., Mean platelet volume as a predictor of cardiovascular risk: a systematic review and meta-analysis. Journal of Thrombosis and Haemostasis, 2010. **8**(1): p. 148-156.
253. Grove, E.L., A.M. Hvas, and S.D. Kristensen, *Immature platelets in patients with acute coronary syndromes.* Thromb Haemost, 2009. **101**(1): p. 151-6.
254. Slavka, G., et al., Mean platelet volume may represent a predictive parameter for overall vascular mortality and ischemic heart disease. Arterioscler Thromb Vasc Biol, 2011. **31**(5): p. 1215-8.
255. Bath, P.M. and R.J. Butterworth, *Platelet size: measurement, physiology and vascular disease.* Blood Coagul Fibrinolysis, 1996. **7**(2): p. 157-61.
256. Grimaldi, C.M., et al., *Estrogen alters thresholds for B cell apoptosis and activation.* J Clin Invest, 2002. **109**(12): p. 1625-33.
257. Dutta, S. and P. Sengupta, *Men and mice: Relating their ages.* Life Sci, 2016. **152**: p. 244-8.
258. Knight, J.A., *Diseases and disorders associated with excess body weight.* Ann Clin Lab Sci, 2011. **41**(2): p. 107-21.
259. Guo, T., et al., Megakaryopoiesis and platelet production: insight into hematopoietic stem cell proliferation and differentiation. Stem Cell Investig, 2015. **2**: p. 3.
260. Kaushansky, K., et al., Thrombopoietin, the Mp1 ligand, is essential for full megakaryocyte development. Proc Natl Acad Sci U S A, 1995. **92**(8): p. 3234-8.
261. Kuter, D.J., D.L. Beeler, and R.D. Rosenberg, The purification of megapoietin: a physiological regulator of megakaryocyte growth and platelet production. Proc Natl Acad Sci U S A, 1994. **91**(23): p. 11104-8.
262. D., B.T., et al., Identification and cloning of a megakaryocyte growth and development factor that is a ligand for the cytokine receptor MpI. Cell, 1994. **77**(7): p. 1117-1124.
263. Feng, Q., et al., Scalable generation of universal platelets from human induced pluripotent stem cells. Stem Cell Reports, 2014. **3**(5): p. 817-31.
264. Matsunaga, T., et al., Ex vivo large-scale generation of human platelets from cord blood CD34+ cells. Stem Cells, 2006. **24**(12): p. 2877-87.

265. Gaur, M., et al., Megakaryocytes derived from human embryonic stem cells: a genetically tractable system to study megakaryocytopoiesis and integrin function. J Thromb Haemost, 2006. **4**(2): p. 436-42.
266. Guerriero, R., et al., Unilineage megakaryocytic proliferation and differentiation of purified hematopoietic progenitors in serum-free liquid culture. Blood, 1995. **86**(10): p. 3725-36.
267. Choi, E.S., et al., Platelets generated in vitro from proplatelet-displaying human megakaryocytes are functional. Blood, 1995. **85**(2): p. 402-13.
268. Bluteau, O., et al., *Developmental changes in human megakaryopoiesis.* J Thromb Haemost, 2013. **11**(9): p. 1730-41.
269. Ran, F.A., et al., Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Cell, 2013. **154**(6): p. 1380-1389.
270. Rowley, J.W., H. Schwertz, and A.S. Weyrich, *Platelet mRNA: the meaning behind the message.* Curr Opin Hematol, 2012. **19**(5): p. 385-91.
271. Durrant, T.N., et al., In-depth PtdIns(3,4,5)P3 signalosome analysis identifies DAPP1 as a negative regulator of GPVI-driven platelet function. Blood Adv, 2017. **1**(14): p. 918-932.
272. Harrison, P., et al., Guidelines for the laboratory investigation of heritable disorders of platelet function. Br J Haematol, 2011. **155**(1): p. 30-44.
273. Rowley, J.W., et al., Genome-wide RNA-seq analysis of human and mouse platelet transcriptomes. Blood, 2011. **118**(14): p. e101-11.
274. Jirouskova, M., A.S. Shet, and G.J. Johnson, *A guide to murine platelet structure, function, assays, and genetic alterations.* J Thromb Haemost, 2007. **5**(4): p. 661-9.
275. Nieswandt, B., et al., *Platelets in atherothrombosis: lessons from mouse models.* J Thromb Haemost, 2005. **3**(8): p. 1725-36.
276. Jones, C.I., et al., Mapping the platelet profile for functional genomic studies and demonstration of the effect size of the GP6 locus. J Thromb Haemost, 2007. **5**(8): p. 1756-65.
277. Yee, D.L., et al., Aggregometry detects platelet hyperreactivity in healthy individuals. Blood, 2005. **106**(8): p. 2723-9.
278. Panzer, S., L. Hocker, and D. Koren, Agonists-induced platelet activation varies considerably in healthy male individuals: studies by flow cytometry. Ann Hematol, 2006. **85**(2): p. 121-5.
279. Fontana, P., et al., Adenosine diphosphate-induced platelet aggregation is associated with P2Y12 gene sequence variations in healthy subjects. Circulation, 2003. **108**(8): p. 989-95.
280. Golebiewska, E.M. and A.W. Poole, Secrets of platelet exocytosis - what do we really know about platelet secretion mechanisms? Br J Haematol, 2013.
281. Konopatskaya, O., et al., Protein kinase C mediates platelet secretion and thrombus formation through protein kinase D2. Blood, 2011. **118**(2): p. 416-24.
282. Dangelmaier, C., et al., PDK1 selectively phosphorylates Thr(308) on Akt and contributes to human platelet functional responses. Thromb Haemost, 2014. **111**(3): p. 508-17.
283. Furie, B. and B.C. Furie, *Thrombus formation in vivo.* J Clin Invest, 2005. **115**(12): p. 3355-62.
284. Rosen, E.D., et al., Laser-induced noninvasive vascular injury models in mice generate platelet- and coagulation-dependent thrombi. Am J Pathol, 2001. **158**(5): p. 1613-22.
285. Bonnard, T. and C.E. Hagemeyer, Ferric Chloride-induced Thrombosis Mouse Model on Carotid Artery and Mesentery Vessel. J Vis Exp, 2015(100): p. e52838.

286. Swystun, L.L. and P.C. Liaw, *The role of leukocytes in thrombosis.* Blood, 2016. **128**(6): p. 753-62.
287. Martinod, K. and D.D. Wagner, *Thrombosis: tangled up in NETs.* Blood, 2014. **123**(18): p. 2768-76.
288. Evstatiev, R., et al., Iron deficiency alters megakaryopoiesis and platelet phenotype independent of thrombopoietin. Am J Hematol, 2014. **89**(5): p. 524-9.
289. Choi, S.I., J.V. Simone, and C.W. Jackson, *Megakaryocytopoiesis in Experimental Iron Deficiency Anemia.* Blood, 1974. **43**(1): p. 111-120.
290. Choi, S.I. and J.V. Simone, *Platelet Production in Experimental Iron Deficiency Anemia.* Blood, 1973. **42**(2): p. 219-228.
291. Park, M.J., et al., The relationship between iron parameters and platelet parameters in women with iron deficiency anemia and thrombocytosis. Platelets, 2013. **24**(5): p. 348-351.
292. Kuku, I., et al., *Platelet counts in adults with iron deficiency anemia.* Platelets, 2009. **20**(6): p. 401-5.
293. Kadikoylu, G., et al., *Platelet parameters in women with iron deficiency anemia.* Journal of the National Medical Association, 2006. **98**(3): p. 398-402.
294. Nichols, T.C., et al., Oxidized LDL and Fructosamine Associated with Severity of Coronary Artery Atherosclerosis in Insulin Resistant Pigs Fed a High Fat/High NaCl Diet. PLoS One, 2015. **10**(7): p. e0132302.
295. Selvin, E., et al., Nontraditional markers of glycemia: associations with microvascular conditions. Diabetes Care, 2011. **34**(4): p. 960-7.
296. Misciagna, G., G. De Michele, and M. Trevisan, *Non enzymatic glycated proteins in the blood and cardiovascular disease.* Curr Pharm Des, 2007. **13**(36): p. 3688-95.
297. Browner, W.S., et al., Association between serum fructosamine and mortality in elderly women: the study of osteoporotic fractures. Am J Epidemiol, 1999. **149**(5): p. 471-5.
298. Mittman, N., et al., Serum fructosamine versus glycosylated hemoglobin as an index of glycemic control, hospitalization, and infection in diabetic hemodialysis patients. Kidney Int Suppl, 2010(117): p. S41-5.
299. Organization, W.H. *Obesity and overweight*. 2018 [cited 2018 9 December 2018]; Available from: https://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight.
300. Gov.uk. *Health matters: obesity and the food environment*. 2018 9 December 2018]; Available from: https://www.gov.uk/government/publications/health-matters-obesity-and-the-food-environment/health-matters-obesity-and-the-food-environment--2.

CHAPTER 8 |APPENDIX

Table 8.1 – Collection of SMIM coding proteins. SMIM coding proteins described in the Human Gene Organization (HUGO - https://www.genenames.org/) and Gene Cards® – Human gene database (https://www.genecards.org/).

Gene symbol	Location (chromosome)	Protein size (KDa)	Function
SMIM1	1	8.7	Regulator of RBC formation
SMIM2	13	9.4	Unknown
SMIM3	5	6.5	Unknown
SMIM4	3	8.7	Unknown
SMIM5	17	8.5	Unknown
SMIM6	17	7	Unknown
SMIM7	19	8.6	Unknown
SMIM8	6	11	Unknown
SMIM9	X	10.8	Unknown
SMIM10	X	9.2	Unknown
SMIM10L2A	X	8.4	Unknown
SMIM10L1	X	7.4	Unknown
SMIM10L2B	X	8.4	Unknown
SMIM11A	21	6.9	Unknown
SMIM11B	21	7.9	Unknown
SMIM12	1	10.8	Unknown
SMIM13	6	10.4	Unknown
SMIM14	4	10.7	Unknown
SMIM15	5	8.6	Unknown
SMIM17	19	13.3	Unknown
SMIM18	8	11.1	Unknown
SMIM19	8	12.4	Unknown
SMIM20	4	7.7	Required for mitochondrial cytochrome c oxidase biogenesis
SMIM21	18	11.7	Unknown
SMIM22	16	14.6	Unknown
SMIM23	5	20	Unknown
SMIM24	19	14.9	Unknown
SMIM26	20	10.9	Unknown
SMIM27	9	6.4	Unknown
SMIM28	6	16.6	Unknown
SMIM29	6	11.6	Unknown
SMIM30	7	6.1	Unknown
SMIM31	4	8.4	Unknown
SMIM32	5	10.9	Unknown
SMIM33	5	14.1	Unknown
SMIM34A	21	15	Unknown
SMIM34B	21	15	Unknown
SMIM35	11	9.4	Unknown
SMIM36	17	10.1	Unknown
SMIM37/MTLN	2	6.5	Micropeptide regulator of Beta-oxidation
SMIM38	11	5.8	Unknown
SMIM39	2	6	Unknown
SMIM40	6	9	Unknown
SMIM41	12	9.3	Unknown

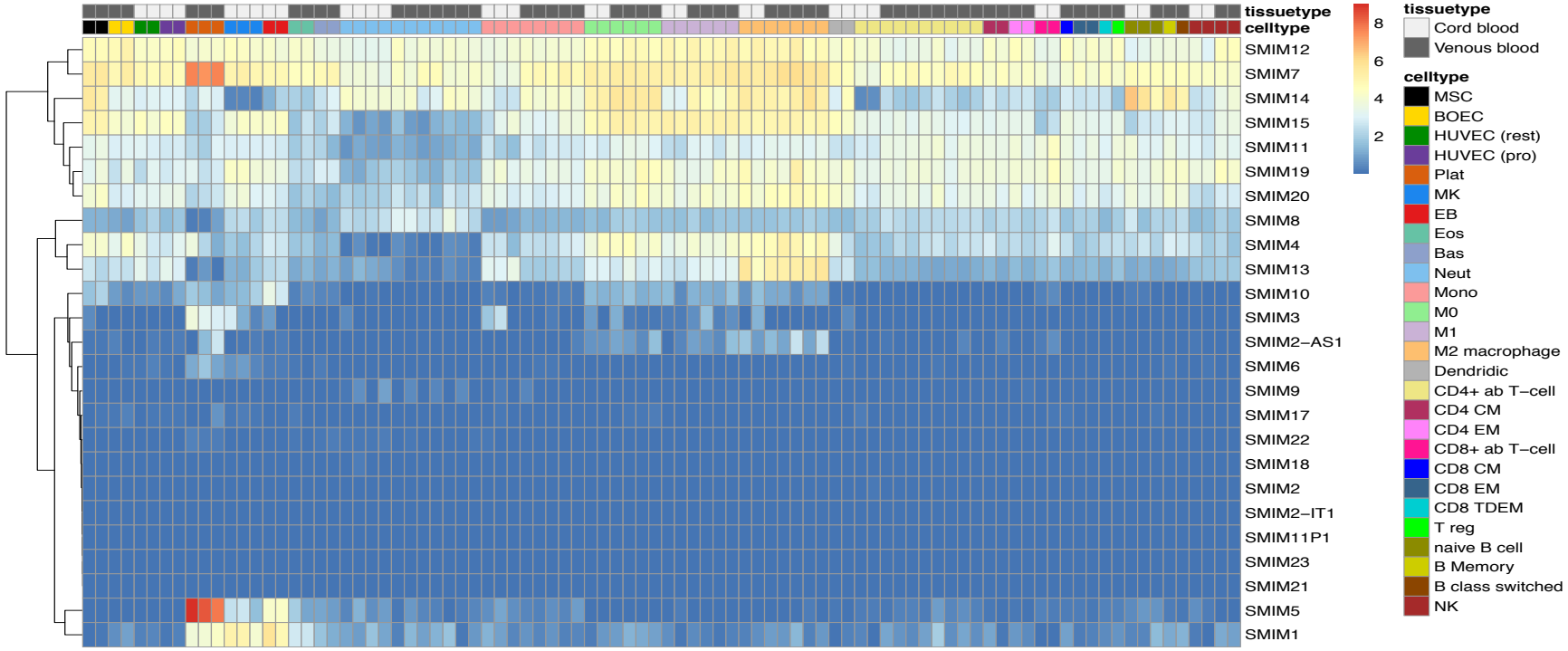


Figure 8.1 – SMIMs mRNA expression level in blood cells. Heat map of human SMIMs mRNA expression in different blood cell types based on RNA-Sequencing data from the Blueprint project [197]. *SMIM5* and *SMIM7* are highly expressed in platelets. Expression levels are color coded according with expression in Log2 (fpkm) – fragments per Kilobase million reads*;* Vb- venous blood; Cb- cord blood; MSC- mesenchymal stem cells; BOEC- blood outgrowth endothelial cells; HUVEC- human umbilical vein endothelial cells; Plat- platelet; MK- megakaryocyte; EB- erythroblast; Neut- neutrophil; Mono- monocytes; M0- macrophage; M1- macrophage M1; CD4+ T-cell- T helper cell; CM- central memory T cells; EM- effector memory T cell; CD8+ T cell- cytotoxic (killer) T cell; CD8 TDEM- terminally differentiated effector memory T cell ; T-reg- regulatory T cell; NK- natural killer cell. Data obtained through Blood RNAexpress- https://blueprint.haem.cam.ac.uk/mRNA/.

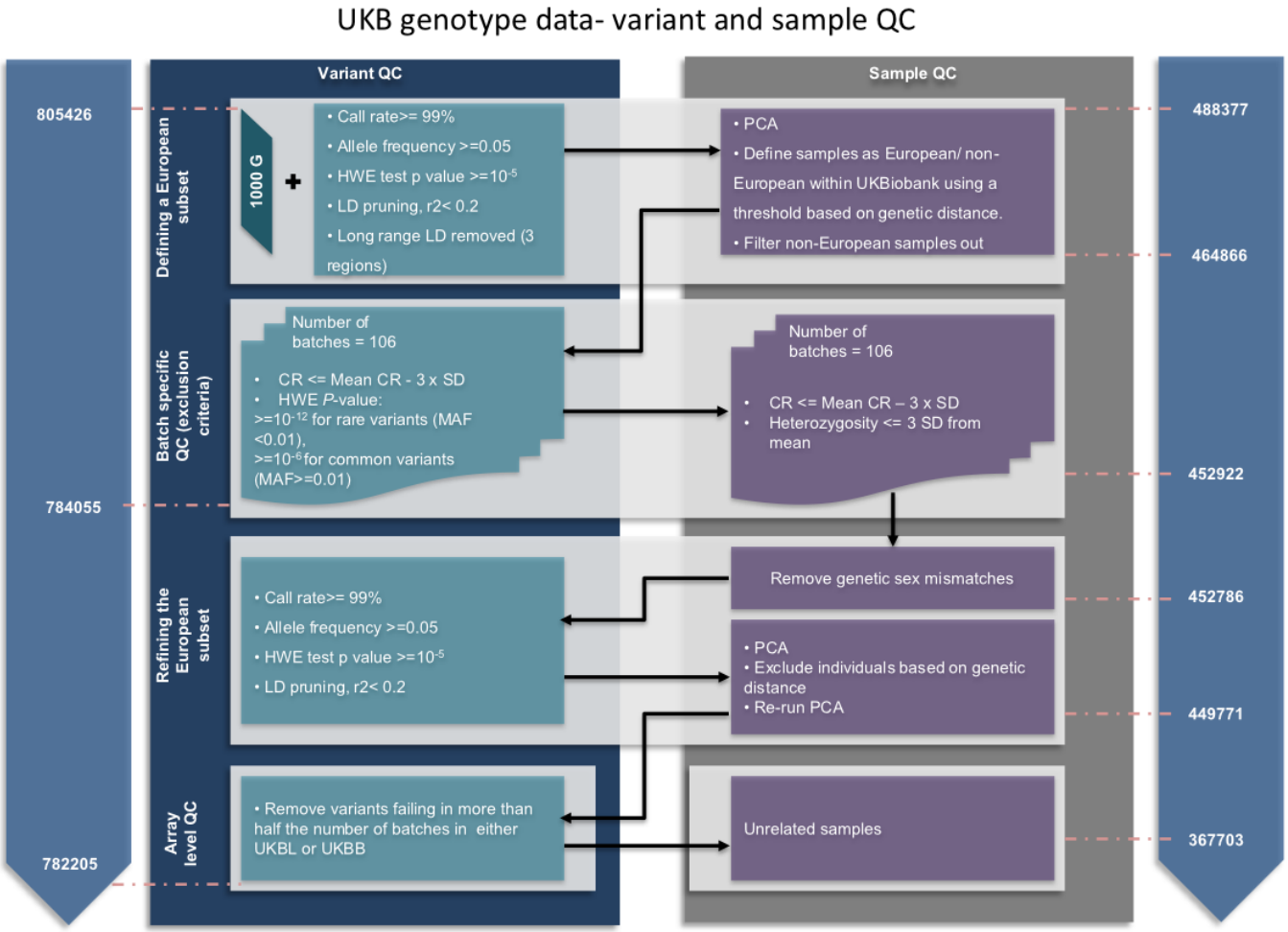


Figure 8.2 – Schematic representation of UK Biobank genotype data-variant and sample QC. Scheme adapted from the internal technical report from the Cardiovascular Epidemiology Unit at the University of Cambridge.